Matrix Mathematics

Theory, Facts, and Formulas with Application to Linear Systems Theory

Dennis S. Bernstein

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Notes to Readers

This version was created on November 19, 2003.

I welcome and appreciate all comments, criticisms, and suggestions.

Some relevant points:

1. Chapter 12 is a work in progress. The index does not include Chapter 12.

2. I provide references for most of the nontrivial Facts. If you happen to know of additional relevant references, please let me know.

3. A few nontrivial facts lack a reference mainly because I have lost track of the original reference. I would like to find a reference or at least verify the correctness of the following facts:

Fact 5.9.25

Fact 9.8.26

4. About 60 problems are included. These problems concern extensions of known results or gaps in the literature. If you should know of any relevant literature (or solutions!), please advise.

5. A few more topics may be added such as: matrix pencils, matrices with block-tridiagonal or block-companion structure, and series (Fer-Magnus-Wei) representations of solutions of the matrix equation $\dot{X}(t) = A(t)X(t)$.

6. Please note errors of any kind.

7. Please feel free to suggest any additional facts or augmentations of existing facts.

Special Symbols

General Notation

π	$3.14159 \cdots$
e	$2.71828 \cdots$
$\triangleq$	equals by definition
$\binom{n}{m}$	$\frac{n!}{m!(n-m)!}$
$\lfloor a \rfloor$	largest integer less than or equal to a
δ_{ij}	1 if $i = j$, 0 if $i \neq j$ (Kronecker delta)
$\log$	logarithm with base e
$\operatorname{sign} \alpha$	1 if $\alpha > 0$, -1 if $\alpha < 0$, 0 if $\alpha = 0$
$\sinh x, \cosh x$	$\frac{1}{2}(e^x - e^{-x})$, $\frac{1}{2}(e^x + e^{-x})$

Chapter 1

$\{\ \}$	set (p. 2)
$\{\ \}_{\mathrm{m}}$	multiset (p. 2)
$\varnothing$	empty set (p. 2)
$\in$	is an element of (p. 2)
$\notin$	is not an element of (p. 2)
$\cap$	intersection (p. 2)
$\cup$	union (p. 2)
$\mathcal{S}^{\sim}$	complement of $\mathcal{S}$ (p. 2)
$\mathcal{S} \backslash \mathcal{S}'$	$\{x \in \mathcal{S}\colon\ x \notin \mathcal{S}'\}$ for sets $\mathcal{S}, \mathcal{S}'$ (p. 2)
$\subseteq$	is a subset of (p. 2)
$\subset$	is a proper subset of (p. 2)
$f\colon\ \mathcal{X} \mapsto \mathcal{Y}$	f is a function with domain $\mathcal{X}$ and codomain $\mathcal{Y}$ (p. 4)
$f^{-1}(\mathcal{S})$	inverse image of $\mathcal{S}$ (p. 4)
$f \bullet g$	composition of functions f and g (p. 4)

Chapter 2

$\mathbb{Z}$	integers (p. 13)
$\mathbb{N}$	nonnegative integers (p. 13)
$\mathbb{P}$	positive integers (p. 13)
$\mathbb{R}$	real numbers (p. 13)
$\mathbb{C}$	complex numbers (p. 13)
$\mathbb{F}$	$\mathbb{R}$ or $\mathbb{C}$ (p. 13)
$\jmath$	$\sqrt{-1}$ (p. 13)
$\overline{z}$	complex conjugate of $z \in \mathbb{C}$ (p. 13)
$\operatorname{Re} z$	real part of $z \in \mathbb{C}$ (p. 13)
$\operatorname{Im} z$	imaginary part of $z \in \mathbb{C}$ (p. 13)
$\lvert z \rvert$	absolute value of $z \in \mathbb{C}$ (p. 13)
CLHP	closed left half plane in $\mathbb{C}$ (p. 14)
OLHP	open left half plane in $\mathbb{C}$ (p. 14)
CRHP	closed right half plane in $\mathbb{C}$ (p. 14)
ORHP	open right half plane in $\mathbb{C}$ (p. 14)
$\jmath\mathbb{R}$	imaginary numbers (p. 14)
$\mathbb{R}^n$	$\mathbb{R}^{n\times 1}$ (real column vectors) (p. 14)
$\mathbb{C}^n$	$\mathbb{C}^{n\times 1}$ (complex column vectors) (p. 14)
$\mathbb{F}^n$	$\mathbb{R}^n$ or $\mathbb{C}^n$ (p. 14)
$x_{(i)}$	ith component of $x \in \mathbb{F}^n$ (p. 14)
$x \geq\geq y$	$x_{(i)} \geq y_{(i)}$ for all i ($x - y$ is nonnegative) (p. 14)
$x >> y$	$x_{(i)} > y_{(i)}$ for all i ($x - y$ is positive) (p. 14)
$\mathbb{R}^{n\times m}$	$n \times m$ real matrices (p. 15)
$\mathbb{C}^{n\times m}$	$n \times m$ complex matrices (p. 15)
$\mathbb{F}^{n\times m}$	$\mathbb{R}^{n\times m}$ or $\mathbb{C}^{n\times m}$ (p. 15)
$\mathrm{row}_i(A)$	ith row of A (p. 15)
$\mathrm{col}_i(A)$	ith column of A (p. 15)

$A_{(i,j)}$	(i,j) entry of A (p. 15)
$A \geq\geq B$	$A_{(i,j)} \geq B_{(i,j)}$ for all i,j ($A-B$ is nonnegative) (p. 16)
$A >> B$	$A_{(i,j)} > B_{(i,j)}$ for all i,j ($A-B$ is positive) (p. 16)
$A \overset{i}{\leftarrow} b$	matrix obtained from $A \in \mathbb{F}^{n\times m}$ by replacing $\mathrm{col}_i(A)$ with $b \in \mathbb{F}^n$ or $\mathrm{row}_i(A)$ with $b \in \mathbb{F}^{1\times m}$ (p. 16)
$\mathrm{d}_{\max}(A) \triangleq \mathrm{d}_1(A)$	largest diagonal entry of $A \in \mathbb{F}^{n\times n}$ having real diagonal entries (p. 16)
$\mathrm{d}_i(A)$	ith largest diagonal entry of $A \in \mathbb{F}^{n\times n}$ having real diagonal entries (p. 16)
$\mathrm{d}_{\min}(A) \triangleq \mathrm{d}_n(A)$	smallest diagonal entry of $A \in \mathbb{F}^{n\times n}$ having real diagonal entries (p. 16)
$[A,B]$	commutator $AB-BA$ (p. 18)
$\mathrm{ad}_A(X)$	adjoint operator $[A,X]$ (p. 18)
$x \times y$	cross product of vectors $x,y \in \mathbb{R}^3$ (p. 18)
$0_{n\times m}, 0$	$n\times m$ zero matrix (p. 18)
I_n, I	$n \times n$ identity matrix (p. 19)
$e_{i,n}, e_i$	$\mathrm{col}_i(I_n)$ (p. 19)
$E_{i,j,n\times m}, E_{i,j}$	$e_{i,n}e_{j,m}^{\mathrm{T}}$ (p. 20)
$1_{n\times m}, 1$	$n\times m$ ones matrix (p. 20)
$\hat{I}_n, \hat{I}$	$n\times n$ reverse identity matrix $\begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix}$ (p. 20)
A^{T}	transpose of A (p. 22)
$\mathrm{tr}\, A$	trace of A (p. 22)
$\overline{Z}$	complex conjugate of $Z \in \mathbb{C}^{n\times m}$ (p. 23)
A^*	$\overline{A}^{\mathrm{T}}$ conjugate transpose of A (p. 23)
$\mathrm{Re}\, A$	real part of $A \in \mathbb{F}^{n\times m}$ (p. 23)
$\mathrm{Im}\, A$	imaginary part of $A \in \mathbb{F}^{n\times m}$ (p. 23)
$\overline{\mathcal{S}}$	$\{\overline{Z}\colon\ Z \in \mathcal{S}\}$ or $\{\overline{Z}\colon\ Z \in \mathcal{S}\}_{\mathrm{m}}$ (p. 23)
$A^{\hat{\mathrm{T}}}$	$\hat{I}A^{\mathrm{T}}\hat{I}$ reverse transpose of A (p. 24)

$A^{\hat{*}}$	$\hat{I}A^*\hat{I}$ reverse conjugate transpose of A (p. 24)
$\operatorname{co} \mathcal{S}$	convex hull of $\mathcal{S}$ (p. 25)
$\operatorname{cone} \mathcal{S}$	conical hull of $\mathcal{S}$ (p. 25)
$\operatorname{coco} \mathcal{S}$	convex conical hull of $\mathcal{S}$ (p. 25)
$\operatorname{span} \mathcal{S}$	span of $\mathcal{S}$ (p. 25)
$\operatorname{aff} \mathcal{S}$	affine hull of $\mathcal{S}$ (p. 25)
$\dim \mathcal{S}$	dimension of $\mathcal{S}$ (p. 26)
$\operatorname{dcone} \mathcal{S}$	dual cone of $\mathcal{S}$ (p. 26)
$\mathcal{S}^\perp$	orthogonal complement of $\mathcal{S}$ (p. 26)
$\mathcal{R}(A)$	range of A (p. 29)
$\mathcal{N}(A)$	null space of A (p. 29)
$\operatorname{rank} A$	rank of A (p. 31)
$\operatorname{def} A$	defect of A (p. 31)
A^{L}	left inverse of A (p. 34)
A^{R}	right inverse of A (p. 34)
A^{-1}	inverse of A (p. 37)
$A^{-\mathrm{T}}$	$\left(A^{\mathrm{T}}\right)^{-1}$ (p. 38)
A^{-*}	$(A^*)^{-1}$ (p. 38)
$\det A$	determinant of A (p. 38)
$A_{[i,j]}$	submatrix of A obtained by deleting $\mathrm{row}_i(A)$ and $\mathrm{col}_j(A)$ (p. 41)
A^{A}	adjugate of A (p. 41)

Chapter 3

$\operatorname{diag}(a_1, \ldots, a_n)$	$\begin{bmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{bmatrix}$ (p. 79)
$\operatorname{revdiag}(a_1, \ldots, a_n)$	$\begin{bmatrix} 0 & & a_1 \\ & \cdot^{\cdot^{\cdot}} & \\ a_n & & 0 \end{bmatrix}$ (p. 79)
N_n, N	$n \times n$ standard nilpotent matrix (p. 78)

$\mathrm{diag}(A_1,\ldots,A_k)$	block-diagonal matrix $\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{bmatrix}$, where $A_i \in \mathbb{F}^{n_i \times m_j}$ (p. 79)
J_{2n}, J	$\left[\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right]$ (p. 81)
$\mathrm{gl}_{\mathbb{F}}(n)$, $\mathrm{pl}_{\mathbb{C}}(n)$, $\mathrm{sl}_{\mathbb{F}}(n)$, $\mathrm{u}(n)$, $\mathrm{su}(n)$, $\mathrm{so}(n)$, $\mathrm{sp}(n)$, $\mathrm{aff}_{\mathbb{F}}(n)$, $\mathrm{se}_{\mathbb{F}}(n)$, $\mathrm{trans}_{\mathbb{F}}(n)$	Lie algebras (p. 83)
$\mathrm{GL}_{\mathbb{F}}(n)$, $\mathrm{PL}_{\mathbb{F}}(n)$, $\mathrm{SL}_{\mathbb{F}}(n)$, $\mathrm{U}(n)$, $\mathrm{O}(n)$, $\mathrm{U}(n,m)$, $\mathrm{O}(n,m)$, $\mathrm{SU}(n)$, $\mathrm{SO}(n)$, $\mathrm{Sp}(n)$, $\mathrm{Aff}_{\mathbb{F}}(n)$, $\mathrm{SE}_{\mathbb{F}}(n)$, $\mathrm{Trans}_{\mathbb{F}}(n)$	groups (p. 84)

Chapter 4

$\mathbb{F}[s]$	polynomials with coefficients in $\mathbb{F}$ (p. 111)
$\deg p$	degree of $p \in \mathbb{F}[s]$ (p. 111)
$\mathrm{mroots}(p)$	multiset of roots of $p \in \mathbb{F}[s]$ (p. 112)
$\mathrm{roots}(p)$	set of roots of $p \in \mathbb{F}[s]$ (p. 112)
$\mathrm{m}_p(\lambda)$	multiplicity of λ as a root of $p \in \mathbb{F}[s]$ (p. 112)
$\mathbb{F}^{n\times m}[s]$	$n \times m$ matrices with entries in $\mathbb{F}[s]$ ($n \times m$ matrix polynomials with coefficients in $\mathbb{F}$) (p. 114)
$\mathrm{rank}\, P$	rank of $P \in \mathbb{F}^{n\times m}[s]$ (p. 115)
χ_A	characteristic polynomial of A (p. 120)
$\lambda_{\max}(A) \triangleq \lambda_1(A)$	largest eigenvalue of $A \in \mathbb{F}^{n\times n}$ having real eigenvalues (p. 120)
$\lambda_i(A)$	ith largest eigenvalue of $A \in \mathbb{F}^{n\times n}$ having real eigenvalues (p. 120)
$\lambda_{\min}(A) \triangleq \lambda_n(A)$	smallest eigenvalue of $A \in \mathbb{F}^{n\times n}$ having real eigenvalues (p. 120)

$\mathrm{am}_A(\lambda)$	algebraic multiplicity of $\lambda \in \mathrm{spec}(A)$ (p. 120)
$\mathrm{spec}(A)$	spectrum of A (p. 120)
$\mathrm{mspec}(A)$	multispectrum of A (p. 120)
$\mathrm{gm}_A(\lambda)$	geometric multiplicity of $\lambda \in \mathrm{spec}(A)$ (p. 125)
$\mathrm{spabs}(A)$	spectral abscissa of A (p. 126)
$\mathrm{sprad}(A)$	spectral radius of A (p. 126)
$\mathrm{In}(A)$	inertia of A (p. 126)
$\nu_-(A), \nu_0(A), \nu_+(A)$	number of eigenvalues of A counting algebraic multiplicity having negative, zero, and positive real part, respectively (p. 126)
μ_A	minimal polynomial of A (p. 127)
$\mathbb{F}(s)$	rational functions with coefficients in $\mathbb{F}$ (p. 129)
$\mathrm{reldeg}\, g$	relative degree of $g \in \mathbb{F}(s)$ (p. 129)
$\mathbb{F}^{n\times m}(s)$	$n \times m$ matrices with entries in $\mathbb{F}(s)$ (p. 129)
$\mathrm{rank}\, G$	rank of $G \in \mathbb{F}(s)$ (p. 129)
$B(p, q)$	Bezout matrix of $p, q \in \mathbb{F}[s]$ (p. 132, Fact 4.8.6)
$H(g)$	Hankel matrix of $g \in \mathbb{F}(s)$ (p. 134, Fact 4.8.7)

Chapter 5

$C(p)$	companion matrix for monic polynomial p (p. 152)
$\mathcal{H}_l(q)$	$l \times l$ or $2l \times 2l$ hypercompanion matrix (p. 156)
$\mathcal{J}_l(q)$	$l \times l$ or $2l \times 2l$ real Jordan matrix (p. 158)
$\mathrm{sig}(A)$	signature of A, that is, $\nu_+(A) - \nu_-(A)$ (p. 164)

$\mathrm{ind}_A(\lambda)$	index of λ with respect to A (p. 165)
$\mathrm{ind}\, A$	index of A, that is, $\mathrm{ind}_A(0)$ (p. 165)
$A_\perp$	complementary idempotent matrix or projector $I - A$ corresponding to the idempotent matrix or projector A (p. 167)
$\sigma_{\max}(A) \triangleq \sigma_1(A)$	largest singular value of $A \in \mathbb{F}^{n\times m}$ (p. 173)
$\sigma_i(A)$	ith largest singular value of $A \in \mathbb{F}^{n\times m}$ (p. 173)
$\sigma_{\min}(A) \triangleq \sigma_n(A)$	minimum singular value of $A \in \mathbb{F}^{n\times n}$ (p. 173)
$V(\lambda_1, \ldots, \lambda_n)$	Vandermonde matrix (p. 195, Fact 5.12.1)
$\mathrm{circ}(a_0, \ldots, a_{n-1})$	circulant matrix of $a_0, \ldots, a_{n-1} \in \mathbb{F}$ (p. 197, Fact 5.12.7)

Chapter 6

A^+	(Moore-Penrose) generalized inverse of A (p. 207)
$D\vert\mathcal{A}$	Schur complement of D with respect to $\mathcal{A}$ (p. 211)
A^{D}	Drazin generalized inverse of A (p. 211)
$A^{\#}$	group generalized inverse of A (p. 213)

Chapter 7

$\mathrm{vec}\, A$	vector formed by stacking columns of A (p. 225)
$\otimes$	Kronecker product (p. 226)
$P_{n,m}$	Kronecker permutation matrix (p. 228)
$\oplus$	Kronecker sum (p. 229)
$A \circ B$	Schur product of A and B (p. 230)
$A^{\{\alpha\}}$	Schur power of A, $\left(A^{\{\alpha\}}\right)_{(i,j)} = \left(A_{(i,j)}\right)^\alpha$ (p. 230)

Chapter 8

$\mathbf{H}^n$	$n \times n$ Hermitian matrices (p. 239)
$\mathbf{N}^n$	$n \times n$ nonnegative-semidefinite matrices (p. 239)
$\mathbf{P}^n$	$n \times n$ positive-definite matrices (p. 239)
$A \geq B$	$A - B \in \mathbf{N}^n$ (p. 239)
$A > B$	$A - B \in \mathbf{P}^n$ (p. 239)
$\langle A \rangle$	$(A^*A)^{1/2}$ (p. 254)
$A\#B$	geometric mean of A and B (p. 274, Fact 8.8.20)
$A{:}B$	parallel sum of A and B (p. 276, Fact 8.9.9)

Chapter 9

$\lvert x \rvert$	absolute value of $x \in \mathbb{F}^n$ (p. 303)
$\lvert A \rvert$	absolute value of $A \in \mathbb{F}^{n\times m}$ (p. 303)
$\lVert x \rVert_p$	Holder norm $\left[\sum_{i=1}^{n} \lvert x_{(i)}\rvert^p\right]^{1/p}$ (p. 304)
$\lVert A \rVert_{\mathrm{F}}$	Frobenius norm $\sqrt{\operatorname{tr} A^*A}$ (p. 308)
$\lVert A \rVert_p$	Holder norm $\left[\sum_{i,j=1}^{n,m} \lvert A_{(i,j)}\rvert^p\right]^{1/p}$ (p. 307)
$\lVert A \rVert_{\sigma p}$	Schatten norm $\left[\sum_{i=1}^{\operatorname{rank} A} \sigma_i^p(A)\right]^{1/p}$ (p. 309)
$\lVert A \rVert_{q,p}$	Holder-induced norm (p. 315)
$\lVert A \rVert_{\mathrm{col}}$	column norm $\lVert A \rVert_{1,1} = \max_{i\in\{1,\ldots,m\}} \lVert \mathrm{col}_i(A)\rVert_1$ (p. 317)
$\lVert A \rVert_{\mathrm{row}}$	row norm $\lVert A \rVert_{\infty,\infty} = \max_{i\in\{1,\ldots,n\}} \lVert \mathrm{row}_i(A)\rVert_1$ (p. 317)
$\ell(A)$	induced lower bound of A (p. 319)

$\ell_{q,p}(A)$	Holder-induced lower bound of A (p. 320)
$\|\cdot\|_{\mathrm{D}}$	dual norm (p. 326, Fact 9.7.8)

Chapter 10

$\mathbb{B}_\varepsilon(x)$	open ball of radius ε centered at x (p. 355)
$\mathbb{S}_\varepsilon(x)$	sphere of radius ε centered at x (p. 355)
$\operatorname{int} \mathcal{S}$	interior of $\mathcal{S}$ (p. 355)
$\operatorname{cl} \mathcal{S}$	closure of $\mathcal{S}$ (p. 355)
$\operatorname{int}_{\mathcal{S}'} \mathcal{S}$	interior of $\mathcal{S}$ relative to $\mathcal{S}'$ (p. 355)
$\operatorname{cl}_{\mathcal{S}'} \mathcal{S}$	closure of $\mathcal{S}$ relative to $\mathcal{S}'$ (p. 356)
$\operatorname{bd} \mathcal{S}$	boundary of $\mathcal{S}$ (p. 356)
$\operatorname{bd}_{\mathcal{S}'} \mathcal{S}$	boundary of $\mathcal{S}$ relative to $\mathcal{S}'$ (p. 356)
$\operatorname{vcone} \mathcal{D}$	variational cone of $\mathcal{D}$ (p. 359)
$\mathrm{D}_+ f(x_0;\xi)$	one-sided directional derivative of f at x_0 in the direction ξ (p. 359)
$\frac{\partial f(x_0)}{\partial x_{(i)}}$	partial derivative of f with respect to $x_{(i)}$ at x_0 (p. 359)
$f'(x)$	Frechet derivative of f at x (p. 360)
$\frac{\mathrm{d} f(x_0)}{\mathrm{d} x_{(i)}}$	$f'(x_0)$ (p. 360)
$f^{(k)}(x)$	kth Frechet derivative of f at x (p. 361)

Chapter 11

e^A or $\exp(A)$	matrix exponential (p. 372)
$\mathcal{S}_{\mathrm{s}}(A)$	asymptotically stable subspace of A (p. 389)
$\mathcal{S}_{\mathrm{u}}(A)$	unstable subspace of A (p. 389)
OUD	open unit disk in $\mathbb{C}$ (p. 395)
CUD	closed unit disk in $\mathbb{C}$ (p. 395)

Chapter 12

$\mathcal{L}$	Laplace transform (p. 434)
$\mathcal{U}(A, C)$	unobservable subspace of (A, C) (p. 436)
$\mathcal{O}(A, C)$	$\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$ (p. 437)
$\mathcal{C}(A, B)$	controllable subspace of (A, B) (p. 442)
$\mathcal{K}(A, B)$	$\begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix}$ (p. 443)

Conventions, Notation, and Terminology

When a word is defined, it is italicized.

The definition of a word, phrase, or symbol should always be understood as an "if and only if" statement, although for brevity "only if" is omitted. The symbol $\triangleq$ means equal by definition.

Analogous statements are written in parallel using the following style: If n is (even, odd), then $n+1$ is (odd, even).

i, j, k, l, m, n always denote integers. Hence, $k \geq 1$ denotes a positive integer, and the limit $\lim_{k\to\infty} A^k$ is taken over positive integers.

The prefix "non" means "not" in the words nonempty, nonzero, non-real, nonnegative, nonunique, nonsingular, nonpositive, nonconstant, and nonnormal. In some traditional usage, "non" may mean "not necessarily."

"Increasing" and "decreasing" indicate strict change for a change in the argument. The word "strict" is superfluous and thus is omitted. Nonincreasing means nowhere increasing, while nondecreasing means nowhere decreasing.

Multisets can have repeated elements so that $\{x\}_{\rm m}$ and $\{x, x\}_{\rm m}$ are different. Multisets help account for repeated eigenvalues. The listed elements α, β, γ of the conventional set $\{\alpha, \beta, \gamma\}$ are not necessarily distinct.

$\mathcal{S}_1 \subset \mathcal{S}_2$ means that $\mathcal{S}_1$ is a proper subset of $\mathcal{S}_2$, that is, $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\mathcal{S}_1 \neq \mathcal{S}_2$. $\mathcal{S}_1 \subseteq \mathcal{S}_2$ means that either $\mathcal{S}_1 \subset \mathcal{S}_2$ or $\mathcal{S}_1 = \mathcal{S}_2$. This notation is consistent with $<$ and $\leq$ for real numbers.

$1/\infty \triangleq 0$.

$0! \triangleq 1$.

$A^0 \triangleq I$ for all square matrices A. In particular, $0^0_{n\times n} = I_n$. With this convention, it is possible to write

$$\sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}$$

for all $-1 < \alpha < 1$. Of course, $\lim_{x\to 0^+} 0^x = 0$, $\lim_{x\to 0^+} x^0 = 1$, and $\lim_{x\to 0^+} x^x = 1$.

$\sqrt{-1}$ is always denoted by dotless $\jmath$. Although i is traditional in mathematics, this notation is common in electrical engineering.

The imaginary numbers are $\jmath\mathbb{R}$. Hence, 0 is both a real number and an imaginary number.

s always represents a complex scalar.

For the scalar ordering "$\le$", if $x \le y$, then $x < y$ if and only if $x \ne y$. For a vector or matrix ordering, $x \le y$ and $x \ne y$ do not imply that $x < y$.

Operations denoted by superscripts are applied before operations represented by preceding operators. For example, $\operatorname{tr}(A+B)^2$ means $\operatorname{tr}\big[(A+B)^2\big]$ and $\operatorname{cl}\mathcal{S}^\sim$ means $\operatorname{cl}(\mathcal{S}^\sim)$. This convention simplifies many formulas.

"Vector" means column vector. A vector is a matrix with one column.

Sets have elements, vectors have components, and matrices have entries. This terminology is traditional and has no mathematical consequence.

$A_{(i,j)}$ is the scalar (i,j) entry of A. $A_{i,j}$ or A_{ij} denotes a block or submatrix of A.

All matrices have nonnegative integral dimensions. If at least one of the dimensions of a matrix is zero, then the matrix is empty.

The entries of a submatrix $\hat{A}$ of a matrix A are the entries of A lying in specified rows and columns. $\hat{A}$ is a block of A if $\hat{A}$ is a submatrix of A whose entries are entries of adjacent rows and columns of A. Every matrix is both a submatrix and block of itself.

The determinant of a submatrix is a subdeterminant. (Some books use "minor.") The determinant of a matrix is also a subdeterminant of the matrix.

The dimension of the null space of a matrix is its defect. Some books use "nullity."

A block of a square matrix is diagonally located if the block is square and the diagonal entries of the block are also diagonal entries of the matrix; otherwise, the block is off-diagonally located. This terminology avoids confusion with a "diagonal block," which is block that is also a a square, diagonal submatrix.

$\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$ consistently in every result. For example, in Theorem 5.6.3, the three appearances of "$\mathbb{F}$" can be read as either all "$\mathbb{C}$" or all "$\mathbb{R}$."

If $\mathbb{F} = \mathbb{R}$, then $\overline{A}$ becomes A, A^* becomes A^{T}, "Hermitian" becomes "symmetric," "unitary" becomes "orthogonal," "unitarily" becomes "orthogonally," and "congruence" becomes "T-congruence." A square complex matrix A is symmetric if $A^{\mathrm{T}} = A$ and orthogonal if $A^{\mathrm{T}}A = I$.

The adjugate of $A \in \mathbb{F}^{n\times n}$ is denoted by A^{A}. The traditional notation is adj$A$. A^{A} is used in [523].

The diagonal entries of a matrix $A \in \mathbb{F}^{n\times n}$ all of whose diagonal entries are real are ordered as $\mathrm{d}_{\max}(A) = \mathrm{d}_1(A) \ge \mathrm{d}_2(A) \ge \cdots \ge \mathrm{d}_{\min}(A) = \mathrm{d}_n(A)$.

The eigenvalues of a matrix $A \in \mathbb{F}^{n\times n}$ all of whose eigenvalues are real are ordered as $\lambda_{\max}(A) = \lambda_1(A) \ge \lambda_2(A) \ge \cdots \ge \lambda_{\min}(A) = \lambda_n(A)$.

For $A \in \mathbb{F}^{n\times n}$, $\mathrm{am}_A(\lambda)$ is the number of copies of λ in the multispectrum of A, $\mathrm{gm}_A(\lambda)$ is the number of Jordan blocks of A associated with λ, and $\mathrm{ind}_A(\lambda)$ is the size of the largest Jordan block of A associated with λ.

An $n \times m$ matrix has exactly $\min\{n,m\}$ singular values, exactly $\operatorname{rank} A$ of which are positive.

The $\min\{n,m\}$ singular values of a matrix $A \in \mathbb{F}^{n\times m}$ are ordered as $\sigma_{\max}(A) \triangleq \sigma_1(A) \ge \sigma_2(A) \ge \cdots \ge \sigma_{\min\{n,m\}}(A)$. If $n = m$, then $\sigma_{\min}(A) \triangleq \sigma_n(A)$. The notation $\sigma_{\min}(A)$ is defined only for square matrices.

Nonnegative-semidefinite and positive-definite matrices are Hermitian.

A matrix that can be diagonalized by a similarity transformation is diagonalizable and thus semisimple since all of its eigenvalues are semisimple. If the matrix is real and all of its eigenvalues are real, then the matrix is diagonalizable over $\mathbb{R}$.

An idempotent matrix $A \in \mathbb{F}^{n\times n}$ satisfies $A^2 = A$, while a projector is a Hermitian, idempotent matrix. Some books use "projector" for idempotent and "orthogonal projector" for projector. A reflector is a Hermitian, involutory matrix. A projector is a normal matrix whose eigenvalues are 1 or 0, while a reflector is a normal matrix whose eigenvalues are 1 or -1.

An elementary matrix is a nonsingular matrix formed by adding an outer-product matrix to the identity matrix. An elementary reflector is a reflector exactly one of whose eigenvalues is -1. An elementary projector is a projector exactly one of whose eigenvalues is 0. Elementary reflectors are elementary matrices. However, elementary projectors are not elementary matrices since elementary projectors are singular.

The rank of a matrix polynomial or rational transfer function P is the maximum rank of $P(s)$ over $\mathbb{C}$. Some books call this "normal rank." We denote this quantity by $\operatorname{rank} P$ as distinct from $\operatorname{rank} P(s)$, which denotes the rank of the matrix $P(s)$, where $s \in \mathbb{C}$.

The symbol $\oplus$ denotes the Kronecker sum. (Some books use $\oplus$ to denote the direct sum of matrices.)

The Holder norms for vectors and matrices are denoted by $\|\cdot\|_p$. The matrix norm induced by $\|\cdot\|_q$ on the domain and $\|\cdot\|_p$ on the codomain is denoted by $\|\cdot\|_{p,q}$.

The Schatten norms for matrices are denoted by $\|\cdot\|_{\sigma p}$, and the Frobenius norm is denoted by $\|\cdot\|_{\mathrm{F}}$. Hence, $\|\cdot\|_{\sigma\infty} = \|\cdot\|_{2,2} = \sigma_{\max}(\cdot)$ and $\|\cdot\|_{\sigma 2} = \|\cdot\|_{\mathrm{F}}$.

Preface

The idea for this book began with the realization that at the heart of the solution to many problems in science, mathematics, and engineering often lies a "matrix fact," that is, an identity, inequality, or property of matrices that is crucial to the solution of the problem. Although there are numerous excellent books on linear algebra and matrix theory, no one book contains all or even most of the vast number of matrix facts that appear throughout the mathematical, scientific, and engineering literature. This book is an attempt to organize many of these facts into a reference source for users of matrix theory in diverse applications areas.

Matrix mathematics, which can be viewed as a significant extension of scalar mathematics, provides powerful tools for analyzing physical problems in science and engineering. Discretization of partial differential equations by means of finite differencing and finite elements yields linear algebraic or differential equations whose matrix structure reflects the nature of physical solutions [530]. Multivariate probability theory and statistical analysis use matrix methods to represent probability distributions, to compute moments, and to perform linear regression for data analysis [215, 249, 269, 387, 503]. The study of linear differential equations [281] depends heavily on matrix analysis, while linear systems theory and control theory are matrix-intensive areas of engineering [31,62,66,141,161,213,306,345,352,382,463,493,510,556, 572,615,632]. In addition, matrices are widely used in rigid body mechanics [11,344,399,432,449,562], structural dynamics [350,409,467], fluid dynamics [137,200,595], circuit theory [13], queuing and stochastic systems [265,436], graph theory [202], signal processing [569], statistical mechanics [7,69,574], demography [329], optics [226], and number theory [339].

In all applications involving matrices, computational techniques are essential for obtaining numerical solutions. The development of efficient and reliable algorithms for matrix computations is therefore an important area of research that has been extensively developed [44, 136, 169, 236, 280, 297, 309, 521, 522, 524, 554, 573, 596, 600, 601]. To facilitate the solution of matrix problems, entire computer packages have been developed using the language of matrices. However, this book is concerned with the analytical properties of matrices rather than their computational aspects.

This book encompasses a broad range of fundamental questions in matrix theory, which, in many cases can be viewed as extensions of related questions in scalar mathematics. A few such questions are:

> What are the basic properties of matrices? How can matrices be characterized, classified, and quantified?

> How can a matrix be decomposed into simpler matrices? A matrix decomposition may involve addition, multiplication, and partition. Decomposing a matrix into its fundamental components provides insight into its algebraic and geometric properties. For example, the polar decomposition states that every square matrix can be written as the product of a rotation and a dilation analogous to the polar representation of a complex number.

> Given a pair of matrices having certain properties, what can be inferred about the sum, product, and concatenation of these matrices? In particular, if a matrix has a given property, to what extent does that property change or remain unchanged if the matrix is perturbed by another matrix of a certain type by means of addition, multiplication, or concatenation? For example, if a matrix is nonsingular, how large can an additive perturbation to that matrix be without the sum becoming singular?

> How can properties of a matrix be determined by means of simple operations? For example, how can the location of the eigenvalues of a matrix be estimated directly in terms of the entries of the matrix?

> To what extent do matrices satisfy the formal properties of the real numbers? For example, while $0 \le a \le b$ implies that $a^r \le b^r$ for real numbers a, b and a positive integer r, when does $0 \le A \le B$ imply $A^r \le B^r$ for nonnegative-semidefinite matrices A and B and with the nonnegative-semidefinite ordering?

Questions of these types have occupied matrix theorists for at least a century, with motivation from diverse applications. The existing scope and depth of knowledge are enormous. Taken together, this body of knowledge provides a powerful framework for developing and analyzing models for scientific and engineering applications.

This book is intended to be useful for at least four groups of readers. Since linear algebra is a standard course in the mathematical sciences and engineering, graduate students in these fields can use this book to expand the scope of their linear algebra text. For instructors, many of the Facts can be used as exercises to augment standard material in matrix courses. For researchers in the mathematical sciences, including statistics, physics, and engineering, this book can be used as a general reference on matrix theory. Finally, for users of matrices in the applied sciences, this book will provide access to a large body of results in matrix theory. By collecting these results in a single source, it is my hope that this book will prove to be convenient and useful for a broad range of applications. The material in this book is thus intended to complement the large number of classical and modern texts and reference works on linear algebra and matrix theory [2, 214, 222, 223, 229, 244, 285, 383, 391, 395, 423, 440, 444, 466, 492, 509, 530].

After a review of mathematical preliminaries in Chapter 1, fundamental properties of matrices are described in Chapter 2. Chapter 3 summarizes the major classes of matrices and various matrix transformations. In Chapter 4 we turn to polynomial and rational matrices whose basic properties are essential for understanding the structure of constant matrices. Chapter 5 is concerned with various decompositions of matrices including the Jordan, Schur, and singular value decompositions. Chapter 6 provides a brief treatment of generalized inverses, while Chapter 7 describes the Kronecker and Schur product operations. Chapter 8 is concerned with the properties of nonnegative-semidefinite matrices. A detailed treatment of vector and matrix norms is given in Chapter 9, while formulas for matrix derivatives are given in Chapter 10. Next, Chapter 11 focuses on the matrix exponential and stability theory, which are central to the study of linear differential equations. In Chapter 12 we apply matrix theory to the analysis of linear systems, their state space realizations, and their transfer function representation. This chapter also includes a discussion of the matrix Riccati equation of control theory.

Each chapter provides a core of results with, in many cases, complete proofs. Sections at the end of each chapter provide a collection of Facts organized to correspond to the order of topics in the chapter. These Facts include corollaries and special cases of results presented in the chapter, as well as related results that go beyond the results of the chapter. In a few cases the Facts include open problems, illuminating remarks, and hints regarding proofs. The Facts are intended to provide the reader with a useful reference collection of matrix results as well as a gateway to the matrix theory literature.

The literature on matrix theory is enormous and includes numerous excellent textbooks and monographs as well as innumerable papers. The material in this book has been drawn from many sources, and these appear in the reference list. An attempt has been made to give appropriate credit wherever possible. However, there are surely omissions in this regard, and I regret all such oversights.

The author is indebted to many individuals who, over the years, provided helpful suggestions as well as material for this book.

Contents

Chapter One

Preliminaries

In this chapter we review some basic terminology and results concerning logic, sets, functions, and related concepts. This material is used throughout the book.

1.1 Logic and Sets

Let A and B be conditions. The *negation* of A is (not A), the *both* of A and B is (A and B), and the *either* of A and B is (A or B).

Let A and B be conditions. The *implication* or *statement* "if A is satisfied, then B is satisfied" or, equivalently, "A implies B," is written as $A \Longrightarrow B$, while $A \Longleftrightarrow B$ is equivalent to $[(A \Longrightarrow B)$ and $(A \Longleftarrow B)]$. Of course, $A \Longleftarrow B$ means $B \Longrightarrow A$.

Suppose $A \Longleftrightarrow B$. Then, A is satisfied *if and only if* B is satisfied. By convention, the implication $A \Longrightarrow B$ (the "only if" part) is *necessity*, while $B \Longrightarrow A$ (the "if" part) is *sufficiency*. The *converse* of $A \Longrightarrow B$ is $B \Longrightarrow A$. The statement $A \Longrightarrow B$ is equivalent to its *contrapositive* (not B) $\Longrightarrow$ (not A).

A *theorem* is a significant result, while a *proposition* is less significant. The primary role of a *lemma* is to support the proof of a theorem or proposition. Finally, a *corollary* is a direct consequence of a theorem or proposition.

Suppose that $A' \Longrightarrow A \Longrightarrow B \Longrightarrow B'$. Then, $A' \Longrightarrow B'$ is a corollary of $A \Longrightarrow B$.

Let A, B, and C be conditions, and assume that $A \Longrightarrow B$. Then, $A \Longrightarrow B$ is a *strengthening* of (A and C) $\Longrightarrow B$. If, in addition, $A \Longrightarrow C$, then the statement (A and C) $\Longrightarrow B$ has *redundant assumptions*.

Let $\mathcal{X} \triangleq \{x, y, z\}$ be a *set.* Then,

$$x \in \mathcal{X} \tag{1.1.1}$$

means that x is an *element* of $\mathcal{X}$. If w is not an element of $\mathcal{X}$, then we write

$$w \notin \mathcal{X}. \tag{1.1.2}$$

The set with no elements, denoted by $\varnothing$, is the *empty set.* If $\mathcal{X} \neq \varnothing$, then $\mathcal{X}$ is *nonempty.*

A set cannot have repeated elements. For example, $\{x, x\} = \{x\}$. However, a *multiset* is a collection of elements that allows for repetition. The multiset consisting of two copies of x is written as $\{x, x\}_{\mathrm{m}}$. However, we do not assume that the listed elements x, y of the conventional set $\{x, y\}$ are distinct.

There are two basic types of mathematical statements involving quantifiers. An *existential statement* is of the form

$$\text{there exists } x \in \mathcal{X} \text{ such that condition } Z \text{ is satisfied,} \tag{1.1.3}$$

while a *universal statement* has the structure

$$\text{condition } Z \text{ is satisfied for all } x \in \mathcal{X}. \tag{1.1.4}$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be sets. The *intersection* of $\mathcal{X}$ and $\mathcal{Y}$ is the set of common elements of $\mathcal{X}$ and $\mathcal{Y}$ given by

$$\mathcal{X} \cap \mathcal{Y} \triangleq \{x\colon\; x \in \mathcal{X} \text{ and } x \in \mathcal{Y}\} = \{x \in \mathcal{X}\colon\; x \in \mathcal{Y}\} \tag{1.1.5}$$
$$= \{x \in \mathcal{Y}\colon\; x \in \mathcal{X}\} = \mathcal{Y} \cap \mathcal{X}, \tag{1.1.6}$$

while the set of elements in either $\mathcal{X}$ or $\mathcal{Y}$ (the *union* of $\mathcal{X}$ and $\mathcal{Y}$) is

$$\mathcal{X} \cup \mathcal{Y} \triangleq \{x\colon\; x \in \mathcal{X} \text{ or } x \in \mathcal{Y}\} = \mathcal{Y} \cup \mathcal{X}. \tag{1.1.7}$$

The *complement* of $\mathcal{X}$ *relative* to $\mathcal{Y}$ is

$$\mathcal{Y} \backslash \mathcal{X} \triangleq \{x \in \mathcal{Y}\colon\; x \notin \mathcal{X}\}. \tag{1.1.8}$$

If $\mathcal{Y}$ is specified, then the *complement* of $\mathcal{X}$ is

$$\mathcal{X}^{\sim} \triangleq \mathcal{Y} \backslash \mathcal{X}. \tag{1.1.9}$$

If $x \in \mathcal{X}$ implies that $x \in \mathcal{Y}$, then $\mathcal{X}$ is *contained* in $\mathcal{Y}$ ($\mathcal{X}$ is a *subset* of $\mathcal{Y}$), which is written as

$$\mathcal{X} \subseteq \mathcal{Y}. \tag{1.1.10}$$

The statement $\mathcal{X} = \mathcal{Y}$ is equivalent to the validity of both $\mathcal{X} \subseteq \mathcal{Y}$ and $\mathcal{Y} \subseteq \mathcal{X}$. If $\mathcal{X} \subseteq \mathcal{Y}$ and $\mathcal{X} \neq \mathcal{Y}$, then $\mathcal{X}$ is a *proper subset* of $\mathcal{Y}$ and we write $\mathcal{X} \subset \mathcal{Y}$. The sets $\mathcal{X}$ and $\mathcal{Y}$ are *disjoint* if $\mathcal{X} \cap \mathcal{Y} = \varnothing$. A *partition* of $\mathcal{X}$ is a collection of pairwise disjoint subsets of $\mathcal{X}$ whose union is equal to $\mathcal{X}$.

The operations "$\cap$","$\cup$", and "$\backslash$" and the relations "$\subset$" and "$\subseteq$" extend directly to multisets. For example,

$$\{x,x\}_{\mathrm{m}} \cup \{x\}_{\mathrm{m}} = \{x,x,x\}_{\mathrm{m}}. \tag{1.1.11}$$

By ignoring repetitions, a multiset can be converted to a set, while a set can be viewed as a multiset with distinct elements.

1.2 Relations and Functions

The *Cartesian product* $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ of sets $\mathcal{X}_1,\ldots,\mathcal{X}_n$ is the set consisting of *ordered elements* of the form $(x_1,\ldots,x_n)$, where $x_i \in \mathcal{X}_i$ for all $i = 1,\ldots,n$. A *relation* $\mathcal{R}$ on a set $\mathcal{X}$ is a subset of $\mathcal{X} \times \mathcal{X}$. For convenience, $(x_1,x_2) \in \mathcal{R}$ is denoted by $x_1 \le x_2$, whereas $x_1 \not\le x_2$ denotes $(x_1,x_2) \notin \mathcal{R}$.

Definition 1.2.1. Let $\mathcal{R}$ be a relation on $\mathcal{X}$. Then, the following terminology is defined:

i) $\mathcal{R}$ is *reflexive* if $x \le x$ for all $x \in \mathcal{X}$.

ii) $\mathcal{R}$ is *antisymmetric* if $x_1 \le x_2$ and $x_2 \le x_1$ imply that $x_1 = x_2$.

iii) $\mathcal{R}$ is *symmetric* if $x_1 \le x_2$ implies that $x_2 \le x_1$.

iv) $\mathcal{R}$ is *transitive* if $x_1 \le x_2$ and $x_2 \le x_3$ imply that $x_1 \le x_3$.

v) $\mathcal{R}$ is *pairwise connected* if $x_1, x_2 \in \mathcal{X}$ implies that either $x_1 \le x_2$ or $x_2 \le x_1$.

vi) $\mathcal{R}$ is a *partial ordering* if it is reflexive, antisymmetric, and transitive.

vii) $\mathcal{R}$ is a *total ordering* if it is a pairwise connected partial ordering.

viii) $\mathcal{R}$ is an *equivalence relation* if it is reflexive, symmetric, and transitive.

For an equivalence relation $\mathcal{R}$, $x_1 \le x_2$ is denoted by $x_1 \equiv x_2$, whereas $x_1 \not\equiv x_2$ denotes $x_1 \not\le x_2$. If $\mathcal{R}$ is an equivalence relation and $x \in \mathcal{X}$, then the subset $\{y \in \mathcal{X}\colon\ y \equiv x\}$ of $\mathcal{X}$ is the *equivalence class of x induced by* $\mathcal{R}$.

Theorem 1.2.2. Let $\mathcal{R}$ be an equivalence relation on a set $\mathcal{X}$. Then, the collection of equivalence classes of $\mathcal{X}$ induced by $\mathcal{R}$ is a partition of $\mathcal{X}$. Conversely, given a partition of $\mathcal{X}$, the relation $\mathcal{R}$ defined by

$$(x,y) \in \mathcal{R} \Longleftrightarrow x \text{ and } y \text{ belong to the same partition subset of } \mathcal{X} \tag{1.2.1}$$

is an equivalence relation.

Proof. For $x \in \mathcal{X}$, let $\mathcal{S}_x$ denote the equivalence class of x induced by $\mathcal{R}$. Clearly, $\mathcal{X} = \bigcup_{x\in\mathcal{X}} \mathcal{S}_x$. It remains to be shown that if $x, y \in \mathcal{X}$, then either $\mathcal{S}_x = \mathcal{S}_y$ or $\mathcal{S}_x \cap \mathcal{S}_y = \varnothing$. Hence, let $x, y \in \mathcal{X}$, and suppose that $\mathcal{S}_x$ and $\mathcal{S}_y$ are not disjoint so that there exists $z \in \mathcal{S}_x \cap \mathcal{S}_y$. Thus, $(x, z) \in \mathcal{R}$ and $(z, y) \in \mathcal{R}$. Now, let $w \in \mathcal{S}_x$. Then, $(w, x) \in \mathcal{R}$, $(x, z) \in \mathcal{R}$, and $(z, y) \in \mathcal{R}$ imply that $(w, y) \in \mathcal{R}$. Hence, $w \in \mathcal{S}_y$, which implies that $\mathcal{S}_x \subseteq \mathcal{S}_y$. By a similar argument, $\mathcal{S}_y \subseteq \mathcal{S}_x$. Consequently, $\mathcal{S}_x = \mathcal{S}_y$. Finally, the proof of the second statement is immediate. □

Let $\mathcal{X}$ and $\mathcal{Y}$ be sets. Then, a *function* f that maps $\mathcal{X}$ into $\mathcal{Y}$ is a rule $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$ that assigns a unique element $f(x)$ (the *image* of x) of $\mathcal{Y}$ to every element x in $\mathcal{X}$. Equivalently, a function $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$ can be viewed as a subset $\mathcal{F}$ of $\mathcal{X} \times \mathcal{Y}$ such that, for all $x \in \mathcal{X}$, there exists $y \in \mathcal{Y}$ such that $(x, y) \in \mathcal{F}$ and, if $(x_1, y_1) \in \mathcal{F}$, $(x_2, y_2) \in \mathcal{F}$, and $x_1 = x_2$, then $y_1 = y_2$. In this case, $\mathcal{F} = \operatorname{graph}(f) \triangleq \{(x, f(x))\colon\ x \in \mathcal{X}\}$. The set $\mathcal{X}$ is the *domain* of f, while the set $\mathcal{Y}$ is the *codomain* of f. For $\mathcal{X}_1 \subseteq \mathcal{X}$, it is convenient to define $f(\mathcal{X}_1) \triangleq \{f(x)\colon\ x \in \mathcal{X}_1\}$. The set $f(\mathcal{X})$, which is denoted by $\mathcal{R}(f)$, is the *range* of f. If, in addition, $\mathcal{Z}$ is a set and $g\colon\ \mathcal{Y} \mapsto \mathcal{Z}$, then $g \bullet f\colon\ \mathcal{X} \mapsto \mathcal{Z}$ (the *composition* of g and f) is the function $(g \bullet f)(x) \triangleq g(f(x))$. If $x_1, x_2 \in \mathcal{X}$ and $f(x_1) = f(x_2)$ implies that $x_1 = x_2$, then f is *one-to-one*; if $\mathcal{R}(f) = \mathcal{Y}$, then f is *onto*. The function $I_{\mathcal{X}}\colon\ \mathcal{X} \mapsto \mathcal{X}$ defined by $I_{\mathcal{X}}(x) \triangleq x$ for all $x \in \mathcal{X}$ is the *identity* on $\mathcal{X}$.

Let $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$. Then, f is *left invertible* if there exists a function $g\colon\ \mathcal{Y} \mapsto \mathcal{X}$ (a *left inverse* of f) such that $g \bullet f = I_{\mathcal{X}}$, whereas f is *right invertible* if there exists a function $h\colon\ \mathcal{Y} \mapsto \mathcal{X}$ (a *right inverse* of f) such that $f \bullet h = I_{\mathcal{Y}}$. In addition, the function $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$ is *invertible* if there exists $f^{-1}\colon\ \mathcal{Y} \mapsto \mathcal{X}$ (the *inverse* of f) such that $f^{-1} \bullet f = I_{\mathcal{X}}$ and $f \bullet f^{-1} = I_{\mathcal{Y}}$. The *inverse image* $f^{-1}(\mathcal{S})$ of $\mathcal{S} \subseteq \mathcal{Y}$ is defined by

$$f^{-1}(\mathcal{S}) \triangleq \{x \in \mathcal{X}\colon\ f(x) \in \mathcal{S}\}. \tag{1.2.2}$$

Theorem 1.2.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, and let $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$. Then, the following statements hold:

i) f is left invertible if and only if f is one-to-one.

ii) f is right invertible if and only if f is onto.

Furthermore, the following statements are equivalent:

iii) f is invertible.

iv) f has a unique inverse.

v) f is one-to-one and onto.

vi) f is left invertible and right invertible.

vii) f has a unique left inverse.

viii) f has a unique right inverse.

Proof. To prove *i)*, suppose that f is left invertible with left inverse $g\colon\ \mathcal{Y} \mapsto \mathcal{X}$. Furthermore, suppose that $x_1, x_2 \in \mathcal{X}$ satisfy $f(x_1) = f(x_2)$. Then, $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$, which shows that f is one-to-one. Conversely, suppose that f is one-to-one so that, for all $y \in \mathcal{R}(f)$, there exists a unique $x \in \mathcal{X}$ such that $f(x) = y$. Hence, define the function $g\colon\ \mathcal{Y} \mapsto \mathcal{X}$ by $g(y) \triangleq x$ for all $y = f(x) \in \mathcal{R}(f)$ and by $g(y)$ arbitrary for all $y \in \mathcal{Y}\backslash\mathcal{R}(f)$. Consequently, $g(f(x)) = x$ for all $x \in \mathcal{X}$, which shows that g is a left inverse of f.

To prove *ii)*, suppose that f is right invertible with right inverse $g\colon\ \mathcal{Y} \mapsto \mathcal{X}$. Then, for all $y \in \mathcal{Y}$, it follows that $f(g(y)) = y$, which shows that f is onto. Conversely, suppose that f is onto so that, for all $y \in \mathcal{Y}$, there exists at least one $x \in \mathcal{X}$ such that $f(x) = y$. Selecting one such x arbitrarily, define $g\colon\ \mathcal{Y} \mapsto \mathcal{X}$ by $g(y) \triangleq x$. Consequently, $f(g(y)) = y$ for all $y \in \mathcal{Y}$, which shows that g is a right inverse of f. □

1.3 Facts on Logic, Sets, and Functions

Fact 1.3.1. Let A and B be conditions. Then, the following statements hold:

i) $(A \text{ or } B) \iff (\text{not } A \implies B)$.

ii) $(A \implies B) \iff (\text{not } A \text{ or } B)$.

iii) $[\text{not } (A \text{ or } B)] \iff (\text{not } A \text{ and not } B)$.

iv) $[\text{not } (A \implies B)] \iff (A \text{ and not } B)$.

Fact 1.3.2. The following statements are equivalent:

i) $A \implies (B \text{ or } C)$.

ii) $(A \text{ and not } B) \implies C$.

Fact 1.3.3. The following statements are equivalent:

i) $A \iff B$.

ii) $(A \text{ or not } B)$ and $[\text{not } (A \text{ and not } B)]$.

Fact 1.3.4. Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be subsets of a set $\mathcal{X}$. Then, the following identities hold:

i) $\mathcal{A} \cap \mathcal{A} = \mathcal{A} \cup \mathcal{A} = \mathcal{A}$.

ii) $(\mathcal{A} \cup \mathcal{B})^\sim = \mathcal{A}^\sim \cap \mathcal{B}^\sim$.

iii) $\mathcal{A}^\sim \cup \mathcal{B}^\sim = (\mathcal{A} \cap \mathcal{B})^\sim$.

iv) $[\mathcal{A} \backslash (\mathcal{A} \cap \mathcal{B})] \cup \mathcal{B} = \mathcal{A} \cup \mathcal{B}$.

v) $(\mathcal{A} \cup \mathcal{B}) \backslash (\mathcal{A} \cap \mathcal{B}) = (\mathcal{A} \cap \mathcal{B}^\sim) \cup (\mathcal{A}^\sim \cap \mathcal{B})$.

vi) $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$.

vii) $\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$.

viii) $(\mathcal{A} \cap \mathcal{B}) \backslash \mathcal{C} = (\mathcal{A} \backslash \mathcal{C}) \cap (\mathcal{B} \backslash \mathcal{C})$.

ix) $(\mathcal{A} \cup \mathcal{B}) \backslash \mathcal{C} = (\mathcal{A} \backslash \mathcal{C}) \cup (\mathcal{B} \backslash \mathcal{C})$.

x) $(\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{B}^\sim) = \mathcal{A}$.

xi) $(\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A}^\sim \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{B}^\sim) = \mathcal{A} \cap \mathcal{B}$.

Fact 1.3.5. Let $(x_1, y_1), (x_2, y_2) \in \mathbb{R} \times \mathbb{R}$. Then, the relation $(x_1, y_1) \le (x_2, y_2)$ defined by $x_1 \le x_2$ and $y_1 \le y_2$ is a partial ordering.

Fact 1.3.6. Let $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$ be invertible. Then,

$$(f^{-1})^{-1} = f.$$

Fact 1.3.7. Let $f\colon\ \mathcal{X} \mapsto \mathcal{Y}$ and $g\colon\ \mathcal{Y} \mapsto \mathcal{Z}$, and assume that f and g are invertible. Then, $g \bullet f$ is invertible and

$$(g \bullet f)^{-1} = f^{-1} \bullet g^{-1}.$$

Fact 1.3.8. Let $\mathcal{X}$ be a set, and let $\mathfrak{X}$ denote the class of subsets of $\mathcal{X}$. Then, "$\subset$" and "$\subseteq$" are transitive relations on $\mathfrak{X}$, and "$\subseteq$" is a partial ordering on $\mathfrak{X}$.

1.4 Facts on Scalar Inequalities

Fact 1.4.1. Let x be a positive number. Then,

$$x^\alpha \begin{cases} \le \alpha x + 1 - \alpha, & 0 \le \alpha \le 1, \\ \ge \alpha x + 1 - \alpha, & \alpha \le 0 \text{ or } \alpha \ge 1. \end{cases}$$

Fact 1.4.2. Let x and y be nonnegative numbers, and let $\alpha \in [0, 1]$. Then,

$$x^\alpha y^{1-\alpha} \le \alpha x + (1 - \alpha) y.$$

(Remark: See Fact 8.12.12 and Fact 8.12.13.)

Fact 1.4.3. Let x and y be real numbers, and let $\alpha \in [0,1]$. Then,

$$e^{\alpha x+(1-\alpha)y} \le \alpha e^x + (1-\alpha)e^y.$$

(Proof: Replace x and y by e^x and e^y, respectively, in Fact 1.4.2.) (Remark: This inequality is a convexity condition. See Definition 8.5.11 for the convexity of matrix-valued functions.)

Fact 1.4.4. Let x be a positive number. Then,

$$1 - x^{-1} \le \log x \le x - 1.$$

Furthermore, equality holds if and only if $x = 1$.

Fact 1.4.5. Let x and y be nonnegative numbers, and let $p, q \in [1, \infty)$ satisfy $1/p + 1/q = 1$. Then,

$$xy \le \frac{x^p}{p} + \frac{y^q}{q}.$$

(Remark: This result is *Young's inequality.* A matrix version is given by Fact 9.12.19.)

Fact 1.4.6. Let x and y be positive numbers, and let $0 \le p \le q$. Then,

$$\frac{x^p + y^p}{(xy)^{p/2}} \le \frac{x^q + y^q}{(xy)^{q/2}}.$$

(Remark: This inequality is a monotonicity property. See Fact 8.7.27.)

Fact 1.4.7. Let x and y be distinct positive numbers, and let p and q be real numbers such that $p < q$. Then,

$$\left(\frac{x^p + y^p}{2}\right)^{1/p} < \left(\frac{x^q + y^q}{2}\right)^{1/q}.$$

(Proof: See [375].) (Remark: This result is a *power mean inequality.* Letting $q = 1$ and $p \to 0$ yields the arithmetic-mean-geometric-mean inequality $\sqrt{xy} \le \frac{1}{2}(x + y)$.)

Fact 1.4.8. Let x and y be distinct positive numbers, let $1/3 \le p < 1 < q$. Then,

$$\sqrt{xy} < \frac{y - x}{\log y - \log x} < \left(\frac{x^p + y^p}{2}\right)^{1/p} < \frac{x + y}{2} < \left(\frac{x^q + y^q}{2}\right)^{1/q}.$$

(Proof: See [375].) (Remark: These inequalities are a refinement of the arithmetic-mean-geometric-mean inequality. Additional inequalities in n variables and related references are given in [619].)

Fact 1.4.9. Let $x_1,\ldots,x_n$ be nonnegative numbers. Then,

$$\left(\prod_{i=1}^n x_i\right)^{1/n} \le \tfrac{1}{n}\sum_{i=1}^n x_i.$$

Furthermore, equality holds if and only if $x_1 = x_2 = \cdots = x_n$. (Remark: This result is the *arithmetic-mean-geometric-mean inequality.* Several proofs are given in [119]. Bounds for the difference between these quantities are given in [12, 132, 558].)

Fact 1.4.10. Let $x_1,\ldots,x_n$ be nonnegative real numbers, let p be a real number, and define

$$M_p \triangleq \begin{cases} \left(\prod\limits_{i=1}^n x_i\right)^{1/n}, & p = 0,\\[2ex] \left(\frac{1}{n}\sum\limits_{i=1}^n x_i^p\right)^{1/p}, & p \ne 0.\end{cases}$$

Now, let p,q be real numbers such that $p \le q$. Then,

$$M_p \le M_q.$$

Furthermore, $p < q$ and at least two of the numbers $x_1,\ldots,x_n$ are distinct if and only if

$$M_p < M_q.$$

(Proof: See [117, p. 210] and [395, p. 105].) (Remark: If p and q are nonzero and $p \le q$, then,

$$\left(\sum_{i=1}^n x_i^p\right)^{1/p} \le \left(\tfrac{1}{n}\right)^{1/q-1/p}\left(\sum_{i=1}^n x_i^q\right)^{1/q},$$

which is a reverse form of Fact 1.4.13. (Remark: This result is a power mean inequality. $M_0 \le M_1$ is the arithmetic-mean-geometric-mean inequality given by Fact 1.4.9.)

Fact 1.4.11. Let $x_1,\ldots,x_n$ be nonnegative numbers, and let $\alpha_1,\ldots,\alpha_n$ be nonnegative numbers such that $\sum_{i=1}^n \alpha_i = 1$. Then,

$$\prod_{i=1}^n x_i^{\alpha_i} \le \sum_{i=1}^n \alpha_i x_i.$$

Furthermore, equality holds if and only if $x_1 = x_2 = \cdots = x_n$. (Remark: This result is the *weighted arithmetic-mean geometric-mean* inequality.) (Proof: Since $f(x) = -\log x$ is convex, it follows that $\log\prod_{i=1}^n x_i^{\alpha_i} = \sum_{i=1}^n \alpha_i \log x_i \le \log\sum_{i=1}^n \alpha_i x_i$. To prove the second statement, define $f\colon [0,\infty)^n \mapsto [0,\infty)$ by $f(\mu_1,\ldots,\mu_n) \triangleq \sum_{i=1}^n \alpha_i\mu_i - \prod_{i=1}^n \mu_i^{\alpha_i}$. Note that

$f(\mu, \ldots, \mu) = 0$ for all $\mu \geq 0$. If $x_1, \ldots, x_n$ minimizes f, then $\partial f/\partial \mu_i(x_1, \ldots, x_n) = 0$ for all $i = 1, \ldots, n$, which implies that $x_1 = x_2 = \cdots = x_n$.)

Fact 1.4.12. Let $x_1, \ldots, x_n$ be nonnegative numbers. Then,

$$1 + \left(\prod_{i=1}^{n} x_i\right)^{1/n} \leq \left[\prod_{i=1}^{n}(1 + x_i)\right]^{1/n}.$$

Furthermore, equality holds if and only if $x_1 = x_2 = \cdots = x_n$. (Proof: Use Fact 1.4.9.) (Remark: This inequality is used to prove Corollary 8.4.15.)

Fact 1.4.13. Let $x_1, \ldots, x_n$ be nonnegative real numbers, and let p, q be real numbers such that $p \leq q$. Then,

$$\left(\sum_{i=1}^{n} x_i^q\right)^{1/q} \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}.$$

Furthermore, the inequality is strict if and only if $p < q$ and at least two of the numbers $x_1, \ldots, x_n$ are nonzero. (Proof: See Proposition 9.1.5.) (Remark: This result is a *power sum inequality* or *Jensen's inequality*. See [117, p. 213]. The result implies that the Holder norm is a monotonic function of the exponent.)

Fact 1.4.14. Let $0 < x_1 < \cdots < x_n$, and let $\alpha_1, \ldots, \alpha_n \geq 0$ satisfy $\sum_{i=1}^{n} \alpha_i = 1$. Then,

$$\left(\sum_{i=1}^{n} \alpha_i x_i\right)\left(\sum_{i=1}^{n} \frac{\alpha_i}{x_i}\right) \leq \frac{(x_1 + x_n)^2}{4x_1 x_n}.$$

(Remark: This result is the *Kantorovich inequality*. See Fact 8.10.5 and [378].)

Fact 1.4.15. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be nonnegative numbers. Then,

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}\left(\sum_{i=1}^{n} y_i^2\right)^{1/2}.$$

Furthermore, equality holds if and only if $\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$ are linearly dependent. (Remark: This result is the *Cauchy-Schwarz inequality*.)

Fact 1.4.16. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be nonnegative numbers, and let $\alpha \in [0,1]$. Then,

$$\sum_{i=1}^n x_i^\alpha y_i^{1-\alpha} \le \left(\sum_{i=1}^n x_i\right)^\alpha \left(\sum_{i=1}^n y_i\right)^{1-\alpha}.$$

Now, let $p,q \in [1,\infty]$ satisfy $1/p + 1/q = 1$. Then, equivalently,

$$\sum_{i=1}^n x_i y_i \le \left(\sum_{i=1}^n x_i^p\right)^{1/p} \left(\sum_{i=1}^n y_i^q\right)^{1/q}.$$

Furthermore, equality holds if and only if $\begin{bmatrix} x_1^p & \cdots & x_n^p \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} y_1^q & \cdots & y_n^q \end{bmatrix}^{\mathrm{T}}$ are linearly dependent. (Remark: This result is *Holder's inequality.*) (Remark: Note the relationship between the *conjugate parameters* p,q and the *barycentric coordinates* $\alpha, 1-\alpha$. See Fact 8.15.23.)

Fact 1.4.17. Let $x_1,\ldots,x_n$ and $y_1,\ldots,y_n$ be nonnegative numbers. Then,

$$\left[\sum_{i=1}^n (x_i+y_i)^p\right]^{1/p} \begin{cases} \ge \left(\sum_{i=1}^n x_i^p\right)^{1/p} + \left(\sum_{i=1}^n y_i^p\right)^{1/p}, & 0 < p \le 1, \\ \le \left(\sum_{i=1}^n x_i^p\right)^{1/p} + \left(\sum_{i=1}^n y_i^p\right)^{1/p}, & p \ge 1. \end{cases}$$

Furthermore, equality holds if and only if either $p=1$ or $\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$ are linearly dependent. (Remark: This result is *Minkowski's inequality.*)

Fact 1.4.18. Let z be a complex scalar with complex conjugate $\overline{z}$, real part $\operatorname{Re} z$, and imaginary part $\operatorname{Im} z$. Then, the following statements hold:

i) $|\operatorname{Re} z| \le |z|$.

ii) If $z \ne 0$, then $z^{-1} = \overline{z}/|z|^2$.

iii) If $z \ne 0$, then $\operatorname{Re} z^{-1} = (\operatorname{Re} z)/|z|^2$.

iv) If $|z| = 1$, then $z^{-1} = \overline{z}$.

v) If $\operatorname{Re} z \ne 0$, then $\operatorname{Re} z^{-1} \ne 0$ $|z| = \sqrt{(\operatorname{Re} z)/(\operatorname{Re} z^{-1})}$.

vi) $|z^2| = |z|^2 = z\overline{z}$.

vii) $z^2 + \overline{z}^2 + 4(\operatorname{Im} z)^2 = 2|z|^2$.

viii) $z^2 + \overline{z}^2 + 2|z|^2 = 4(\operatorname{Re} z)^2$.

ix) $|z^2 + \overline{z}^2| \le 2|z|^2$.

x) $|e^z| \le e^{|z|}$.

Now, let z_1 and z_2 be complex scalars. Then, the following statements hold:

x) $|z_1 z_2| = |z_1||z_2|$.

xi) $|z_1 + z_2| \le |z_1| + |z_2|$.

xii) $|z_1 + z_2| = |z_1| + |z_2|$ if and only if there exists $\alpha \ge 0$ such that either $z_1 = \alpha z_2$ or $z_2 = \alpha z_1$.

(Remark: Matrix analogues of some of these results are given in [548].)

1.5 Notes

Most of the preliminary material in this chapter can be found in [434]. A related treatment of mathematical preliminaries is given in [484]. Reference works on inequalities include [70, 117–119, 149, 395, 400, 424].

Chapter Two

Basic Matrix Properties

In this chapter we provide a detailed treatment of the basic properties of matrices such as range, null space, rank, and invertibility. We also consider properties of convex sets, cones, and subspaces.

2.1 Matrix Algebra

The symbols $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{P}$ denote the sets of integers, nonnegative integers, and positive integers, respectively. The symbols $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex number fields, respectively, whose elements are *scalars*. Since $\mathbb{R}$ is a proper subset of $\mathbb{C}$, we state many results for $\mathbb{C}$. In other cases, it is either desirable to treat $\mathbb{R}$ and $\mathbb{C}$ separately or simply not to make a distinction. To do this efficiently, we use the symbol $\mathbb{F}$ to consistently denote either $\mathbb{R}$ or $\mathbb{C}$.

Let $x \in \mathbb{C}$. Then, $x = y + \jmath z$, where $y, z \in \mathbb{R}$ and $\jmath \triangleq \sqrt{-1}$. Define the *complex conjugate* $\overline{x}$ of x by

$$\overline{x} \triangleq y - \jmath z \tag{2.1.1}$$

and the real and imaginary parts $\operatorname{Re} x$ and $\operatorname{Im} x$ of x by

$$\operatorname{Re} x \triangleq \tfrac{1}{2}(x + \overline{x}) = y \tag{2.1.2}$$

and

$$\operatorname{Im} x \triangleq \tfrac{1}{2\jmath}(x - \overline{x}) = z. \tag{2.1.3}$$

Furthermore, the *absolute value* $|x|$ of x is defined by

$$|x| \triangleq \sqrt{x^2 + y^2}. \tag{2.1.4}$$

The *closed left half plane* (CLHP), *open left half plane* (OLHP), *closed right half plane* (CRHP), and *open right half plane* (ORHP) are the subsets of $\mathbb{C}$ defined by

$$\mathrm{CLHP} \triangleq \{s \in \mathbb{C}\colon\ \ \mathrm{Re}\, s \le 0\}, \tag{2.1.5}$$
$$\mathrm{OLHP} \triangleq \{s \in \mathbb{C}\colon\ \ \mathrm{Re}\, s < 0\}, \tag{2.1.6}$$
$$\mathrm{CRHP} \triangleq \{s \in \mathbb{C}\colon\ \ \mathrm{Re}\, s \ge 0\}, \tag{2.1.7}$$
$$\mathrm{ORHP} \triangleq \{s \in \mathbb{C}\colon\ \ \mathrm{Re}\, s > 0\}. \tag{2.1.8}$$

The imaginary numbers are represented by $\jmath\mathbb{R}$. Note that 0 is both a real number and an imaginary number.

The set $\mathbb{F}^n$ consists of *vectors* x of the form

$$x = \begin{bmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{bmatrix}, \tag{2.1.9}$$

where $x_{(1)}, \ldots, x_{(n)} \in \mathbb{F}$ are the *components* of x. Hence, the elements of $\mathbb{F}^n$ are *column vectors*. Since $\mathbb{F}^1 = \mathbb{F}$, it follows that every scalar is also a vector. If $x \in \mathbb{R}^n$ and every component of x is nonnegative, then x is *nonnegative*, which is written as $x \ge\ge 0$. If $x \in \mathbb{R}^n$ and every component of x is positive, then x is *positive*, which is written as $x >> 0$. If $x, y \in \mathbb{R}^n$, then $x \ge\ge y$ means that $x - y \ge\ge 0$, while $x >> y$ means that $x - y >> 0$.

Definition 2.1.1. Let $x, y \in \mathbb{R}^n$, and assume that $x_{(1)} \ge \cdots \ge x_{(n)}$ and $y_{(1)} \ge \cdots \ge y_{(n)}$. Then, the following terminology is defined:

i) y *weakly majorizes* x if, for all $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} x_{(i)} \le \sum_{i=1}^{k} y_{(i)}. \tag{2.1.10}$$

ii) y *strongly majorizes* x if y weakly majorizes x and

$$\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}. \tag{2.1.11}$$

Now, assume that x and y are nonnegative. Then, the following terminology is defined:

iii) y *weakly log majorizes* x if, for all $k = 1, \ldots, n$,

$$\prod_{i=1}^{k} x_{(i)} \le \prod_{i=1}^{k} y_{(i)}. \tag{2.1.12}$$

iv) y *strongly log majorizes* x if y weakly log majorizes x and

$$\prod_{i=1}^{n} x_{(i)} = \prod_{i=1}^{n} y_{(i)}. \tag{2.1.13}$$

If $\alpha \in \mathbb{F}$ and $x \in \mathbb{F}^n$, then $\alpha x \in \mathbb{F}^n$ is given by

$$\alpha x = \begin{bmatrix} \alpha x_{(1)} \\ \vdots \\ \alpha x_{(n)} \end{bmatrix}. \tag{2.1.14}$$

If $x, y \in \mathbb{F}^n$, then x and y are *linearly dependent* if there exists $\alpha \in \mathbb{F}$ such that either $x = \alpha y$ or $y = \alpha x$. Linear dependence for a set of two or more vectors is defined in Section 2.3. Furthermore, vectors add component by component, that is, if $x, y \in \mathbb{F}^n$, then

$$x + y = \begin{bmatrix} x_{(1)} + y_{(1)} \\ \vdots \\ x_{(n)} + y_{(n)} \end{bmatrix}. \tag{2.1.15}$$

Thus, if $\alpha, \beta \in \mathbb{F}$, then the *linear combination* $\alpha x + \beta y$ is given by

$$\alpha x + \beta y = \begin{bmatrix} \alpha x_{(1)} + \beta y_{(1)} \\ \vdots \\ \alpha x_{(n)} + \beta y_{(n)} \end{bmatrix}. \tag{2.1.16}$$

The vectors $x_1, \ldots, x_m \in \mathbb{F}^n$ placed side by side form the *matrix*

$$A \triangleq \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}, \tag{2.1.17}$$

which has n *rows* and m *columns*. The components of the vectors $x_1, \ldots, x_m$ are the *entries* of A. We write $A \in \mathbb{F}^{n \times m}$ and say that A has *size* $n \times m$. Since $\mathbb{F}^n = \mathbb{F}^{n \times 1}$, it follows that every vector is also a matrix. Note that $\mathbb{F}^{1 \times 1} = \mathbb{F}^1 = \mathbb{F}$. If $n = m$, then n is the *order* of A, and A is *square*. The ith row of A and the jth column of A are denoted by $\mathrm{row}_i(A)$ and $\mathrm{col}_j(A)$, respectively. Hence,

$$A = \begin{bmatrix} \mathrm{row}_1(A) \\ \vdots \\ \mathrm{row}_n(A) \end{bmatrix} = \begin{bmatrix} \mathrm{col}_1(A) & \cdots & \mathrm{col}_m(A) \end{bmatrix}. \tag{2.1.18}$$

The entry $x_{j(i)}$ of A in both the ith row of A and the jth column of A is denoted by $A_{(i,j)}$. Therefore, $x \in \mathbb{F}^n$ can be written as

$$x = \begin{bmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{bmatrix} = \begin{bmatrix} x_{(1,1)} \\ \vdots \\ x_{(n,1)} \end{bmatrix}. \tag{2.1.19}$$

Let $A \in \mathbb{F}^{n \times m}$. For $b \in \mathbb{F}^n$, the matrix obtained from A by replacing $\mathrm{col}_i(A)$ with b is denoted by

$$A \overset{i}{\leftarrow} b. \tag{2.1.20}$$

Likewise, for $b \in \mathbb{F}^{1\times m}$, the matrix obtained from A by replacing $\mathrm{row}_i(A)$ with b is denoted by (2.1.20).

Let $A \in \mathbb{R}^{n\times m}$. If every entry of A is nonnegative, then A is *nonnegative*, which is written as $A \geq\geq 0$. If $A \in \mathbb{R}^n$ and every entry of x is positive, then x is *positive*, which is written as $A >> 0$. If $A, B \in \mathbb{R}^{n\times m}$, then $A \geq\geq B$ means that $A - B \geq\geq 0$, while $A >> B$ means that $A - B >> 0$.

Let $A \in \mathbb{F}^{n\times m}$, and let $l \triangleq \min\{n, m\}$. Then, the entries $A_{(i,i)}$ for all $i = 1, \ldots, l$ and $A_{(i,j)}$ for all $i \neq j$ are the *diagonal entries* and *off-diagonal entries* of A, respectively. Moreover, for all $i = 1, \ldots, l-1$, the entries $A_{(i,i+1)}$ and $A_{(i+1,i)}$ are the *superdiagonal entries* and *subdiagonal entries* of A, respectively. In addition, the entries $A_{(i,l+1-i)}$ for all $i = 1, \ldots, l$ are the *reverse-diagonal entries* of A. If the diagonal entries $A_{(1,1)}, \ldots, A_{(l,l)}$ of A are real, then $\mathrm{d_{min}}(A)$ and $\mathrm{d_{max}}(A)$ denote the smallest and largest diagonal entries of A, respectively, and the diagonal entries of A are relabeled from largest to smallest as

$$\mathrm{d_{max}}(A) \triangleq \mathrm{d}_1(A) \geq \cdots \geq \mathrm{d_{min}}(A) \triangleq \mathrm{d}_l(A). \tag{2.1.21}$$

Partitioned matrices are of the form

$$\begin{bmatrix} A_{11} & \cdots & A_{1l} \\ \vdots & \mathinner{\cdot\vdots\cdot} & \vdots \\ A_{k1} & \cdots & A_{kl} \end{bmatrix}, \tag{2.1.22}$$

where, for all $i = 1, \ldots, k$ and $j = 1, \ldots, l$, the *block* A_{ij} of A is a matrix of size $n_i \times m_j$. If $n_i = m_j$ and the diagonal entries of A_{ij} lie on the diagonal of A, then the square matrix A_{ij} is a *diagonally located block*; otherwise, A_{ij} is an *off-diagonally located block*.

Matrices of the same size add entry by entry, that is, if $A, B \in \mathbb{F}^{n\times m}$, then, for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $(A + B)_{(i,j)} = A_{(i,j)} + B_{(i,j)}$. Furthermore, for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $(\alpha A)_{(i,j)} = \alpha A_{(i,j)}$ for all $\alpha \in \mathbb{F}$ so that $(\alpha A + \beta B)_{(i,j)} = \alpha A_{(i,j)} + \beta B_{(i,j)}$ for all $\alpha, \beta \in \mathbb{F}$. If $A, B \in \mathbb{F}^{n\times m}$, then A and B are *linearly dependent* if there exists $\alpha \in \mathbb{F}$ such that either $A = \alpha B$ or $B = \alpha A$.

Let $A \in \mathbb{F}^{n\times m}$ and $x \in \mathbb{F}^m$. Then, the matrix-vector product Ax is defined by

$$Ax \triangleq \begin{bmatrix} \mathrm{row}_1(A)x \\ \vdots \\ \mathrm{row}_n(A)x \end{bmatrix}. \tag{2.1.23}$$

It can be seen that Ax is a linear combination of the columns of A, that is,

$$Ax = \sum_{i=1}^{m} x_{(i)} \mathrm{col}_i(A). \tag{2.1.24}$$

The matrix A can be associated with the function $f\colon\ \mathbb{F}^m \mapsto \mathbb{F}^n$ defined by $f(x) \triangleq Ax$ for all $x \in \mathbb{F}^m$. The function $f\colon\ \mathbb{F}^m \mapsto \mathbb{F}^n$ is *linear* since, for all $\alpha, \beta \in \mathbb{F}$ and $x, y \in \mathbb{F}^m$, it follows that

$$f(\alpha x + \beta y) = \alpha Ax + \beta Ay. \tag{2.1.25}$$

The function $f\colon\ \mathbb{F}^m \mapsto \mathbb{F}^n$ defined by

$$f(x) \triangleq Ax + z, \tag{2.1.26}$$

where $z \in \mathbb{F}^n$, is *affine*.

Theorem 2.1.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, and define $f\colon\ \mathbb{F}^m \mapsto \mathbb{F}^n$ and $g\colon\ \mathbb{F}^l \mapsto \mathbb{F}^m$ by $f(x) \triangleq Ax$ and $g(y) \triangleq By$. Furthermore, define the composition $h \triangleq f \bullet g\colon\ \mathbb{F}^l \mapsto \mathbb{F}^n$. Then, for all $y \in \mathbb{R}^l$,

$$h(y) = (AB)y, \tag{2.1.27}$$

where, for all $i = 1, \ldots, n$ and $j = 1, \ldots, l$, $AB \in \mathbb{F}^{n\times l}$ is defined by

$$(AB)_{(i,j)} \triangleq \sum_{k=1}^{m} A_{(i,k)} B_{(k,j)}. \tag{2.1.28}$$

Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, $AB \in \mathbb{F}^{n\times l}$ is the *product* of A and B. The matrices A and B are *conformable*, and the product (2.1.28) defines *matrix multiplication*.

Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, AB can be written as

$$AB = \begin{bmatrix} A\mathrm{col}_1(B) & \cdots & A\mathrm{col}_l(B) \end{bmatrix} = \begin{bmatrix} \mathrm{row}_1(A)B \\ \vdots \\ \mathrm{row}_n(A)B \end{bmatrix}. \tag{2.1.29}$$

Thus, for all $i = 1, \ldots, n$ and $j = 1, \ldots, l$,

$$(AB)_{(i,j)} = \mathrm{row}_i(A)\mathrm{col}_j(B), \tag{2.1.30}$$

$$\mathrm{col}_j(AB) = A\mathrm{col}_j(B), \tag{2.1.31}$$

$$\mathrm{row}_i(AB) = \mathrm{row}_i(A)B. \tag{2.1.32}$$

As a special case, note that if $x \in \mathbb{F}^{1\times n}$ and $y \in \mathbb{F}^n = \mathbb{F}^{n\times 1}$, then the scalar $xy \in \mathbb{F}$ is given by

$$xy = \sum_{i=1}^{n} x_{(1,i)} y_{(i)}. \tag{2.1.33}$$

For conformable matrices A, B, C, the associative and distributive identities

$$(AB)C = A(BC), \tag{2.1.34}$$
$$A(B+C) = AB + AC, \tag{2.1.35}$$
$$(A+B)C = AC + BC \tag{2.1.36}$$

are valid. Hence, we write ABC for $(AB)C$ and $A(BC)$.

Let $A, B \in \mathbb{F}^{n\times n}$. Then, the *commutator* $[A, B] \in \mathbb{F}^{n\times n}$ of A and B is the matrix

$$[A, B] \triangleq AB - BA. \tag{2.1.37}$$

The *adjoint operator* $\mathrm{ad}_A\colon\ \mathbb{F}^{n\times n} \mapsto \mathbb{F}^{n\times n}$ is defined by

$$\mathrm{ad}_A(X) \triangleq [A, X]. \tag{2.1.38}$$

Let $x, y \in \mathbb{R}^3$. Then, the *cross product* $x \times y \in \mathbb{R}^3$ of x and y is defined by

$$x \times y \triangleq \begin{bmatrix} x_{(2)}y_{(3)} - x_{(3)}y_{(2)} \\ x_{(3)}y_{(1)} - x_{(1)}y_{(3)} \\ x_{(1)}y_{(2)} - x_{(2)}y_{(1)} \end{bmatrix}. \tag{2.1.39}$$

Multiplication of partitioned matrices is analogous to matrix multiplication with scalar entries. For example, for matrices with conformable blocks,

$$\begin{bmatrix} A & B \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = AC + BD, \tag{2.1.40}$$

$$\begin{bmatrix} A \\ B \end{bmatrix} C = \begin{bmatrix} AC \\ BC \end{bmatrix}, \tag{2.1.41}$$

$$\begin{bmatrix} A \\ B \end{bmatrix}\begin{bmatrix} C & D \end{bmatrix} = \begin{bmatrix} AC & AD \\ BC & BD \end{bmatrix}, \tag{2.1.42}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{bmatrix}. \tag{2.1.43}$$

The $n \times m$ *zero matrix*, all of whose entries are zero, is written as $0_{n\times m}$. If the dimensions are unambiguous, then we write just 0. Let $x \in \mathbb{F}^m$ and $A \in \mathbb{F}^{n\times m}$. Then, the zero matrix satisfies

$$0_{k\times m}x = 0_k, \tag{2.1.44}$$
$$A0_{m\times l} = 0_{n\times l}, \tag{2.1.45}$$
$$0_{k\times n}A = 0_{k\times m}. \tag{2.1.46}$$

Another special matrix is the *empty matrix*. For $n \in \mathbb{N}$, the $0 \times n$ empty matrix, which is written as $0_{0\times n}$, has zero rows and n columns, while the $n \times 0$ empty matrix, which is written as $0_{n\times 0}$, has n rows and zero columns. For $A \in \mathbb{F}^{n\times m}$, where $n, m \in \mathbb{N}$, the empty matrix satisfies the multiplication rules

$$0_{0\times n}A = 0_{0\times m} \tag{2.1.47}$$

and

$$A0_{m\times 0} = 0_{n\times 0}. \tag{2.1.48}$$

Although empty matrices have no entries, it is useful to define the product

$$0_{n\times 0}0_{0\times m} \triangleq 0_{n\times m}. \tag{2.1.49}$$

Also, we define

$$I_0 \triangleq \hat{I}_0 \triangleq 0_{0\times 0}. \tag{2.1.50}$$

For $n, m \in \mathbb{N}$, we define $\mathbb{F}^{0\times m} \triangleq \{0_{0\times m}\}$, $\mathbb{F}^{n\times 0} \triangleq \{0_{n\times 0}\}$, and $\mathbb{F}^0 \triangleq \mathbb{F}^{0\times 1}$. The empty matrix can be viewed as a useful device for matrices just as 0 is for real numbers and $\varnothing$ is for sets.

The $n \times n$ *identity matrix*, which has ones on the diagonal and zeros elsewhere, is denoted by I_n or just I. Let $x \in \mathbb{F}^n$ and $A \in \mathbb{F}^{n\times m}$. Then, the identity matrix satisfies

$$I_n x = x \tag{2.1.51}$$

and

$$AI_m = I_n A = A. \tag{2.1.52}$$

Let $A \in \mathbb{F}^{n\times n}$. Then, $A^2 \triangleq AA$ and, for all $k \in \mathbb{P}$, $A^k \triangleq AA^{k-1}$. We use the convention $A^0 \triangleq I$ even if A is the zero matrix. If $k \in \mathbb{N}$, then

$$A^{k\mathrm{T}} \triangleq \left(A^k\right)^{\mathrm{T}} = \left(A^{\mathrm{T}}\right)^k \tag{2.1.53}$$

and

$$A^{k*} \triangleq \left(A^k\right)^{*} = (A^*)^k. \tag{2.1.54}$$

The vector $e_{i,n} \in \mathbb{R}^n$, or just e_i, has 1 as its ith component and zeros elsewhere. Thus,

$$e_{i,n} = \mathrm{col}_i(I_n). \tag{2.1.55}$$

Let $A \in \mathbb{F}^{n\times m}$. Then, $e_i^{\mathrm{T}}A = \mathrm{row}_i(A)$ and $Ae_i = \mathrm{col}_i(A)$. Furthermore, the (i, j) entry of A can be written as

$$A_{(i,j)} = e_i^{\mathrm{T}}Ae_j = e_j^{\mathrm{T}}A^{\mathrm{T}}e_i. \tag{2.1.56}$$

The $n \times m$ matrix $E_{i,j,n\times m} \in \mathbb{R}^{n\times m}$, or just $E_{i,j}$, has 1 as its (i, j)

entry and zeros elsewhere. Thus,

$$E_{i,j,n\times m} = e_{i,n}e_{j,m}^{\mathrm{T}}. \tag{2.1.57}$$

Note that $E_{i,1,n\times 1} = e_{i,n}$ and

$$I_n = E_{1,1} + \cdots + E_{n,n} = \sum_{i=1}^{n} e_i e_i^{\mathrm{T}}. \tag{2.1.58}$$

Finally, the $n \times m$ *ones matrix*, all of whose entries are 1, is written as $1_{n\times m}$ or just 1. Thus,

$$1_{n\times m} = \sum_{i,j=1}^{n,m} E_{i,j,n\times m}. \tag{2.1.59}$$

Note that

$$1_{n\times 1} = \sum_{i=1}^{n} e_{i,n} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{2.1.60}$$

and

$$1_{n\times m} = 1_{n\times 1}1_{1\times m}. \tag{2.1.61}$$

The $n \times n$ *reverse identity matrix*, which has ones on the reverse diagonal and zeros elsewhere, is denoted by $\hat{I}_n$ or just $\hat{I}$. Left multiplication of $A \in \mathbb{F}^{n\times m}$ by $\hat{I}_n$ reverses the rows of A, while right multiplication of A by $\hat{I}_m$ reverses the columns of A.

2.2 Transpose and Inner Product

A fundamental vector and matrix operation is the transpose. If $x \in \mathbb{F}^n$, then the *transpose* x^{T} of x is defined to be the row vector

$$x^{\mathrm{T}} \triangleq \begin{bmatrix} x_{(1)} & \cdots & x_{(n)} \end{bmatrix} \in \mathbb{F}^{1\times n}. \tag{2.2.1}$$

Similarly, if $x = \begin{bmatrix} x_{(1,1)} & \cdots & x_{(1,n)} \end{bmatrix} \in \mathbb{F}^{1\times n}$, then

$$x^{\mathrm{T}} = \begin{bmatrix} x_{(1,1)} \\ \vdots \\ x_{(1,n)} \end{bmatrix} \in \mathbb{F}^{n\times 1}. \tag{2.2.2}$$

Let $x, y \in \mathbb{F}^n$. Then, $x^{\mathrm{T}}y \in \mathbb{F}$ is a scalar, and

$$x^{\mathrm{T}}y = \left(x^{\mathrm{T}}y\right)^{\mathrm{T}} = y^{\mathrm{T}}x = \sum_{i=1}^{n} x_{(i)}y_{(i)}. \tag{2.2.3}$$

Note that

$$x^{\mathrm{T}}x = \sum_{i=1}^{n} x_{(i)}^2. \tag{2.2.4}$$

Lemma 2.2.1. Let $x \in \mathbb{R}$. Then, $x^{\mathrm{T}}x = 0$ if and only if $x = 0$.

Let $x, y \in \mathbb{R}^n$. Then, $x^{\mathrm{T}}y \in \mathbb{R}$ is the *inner product* of x and y. Furthermore, x is *orthogonal* to y if $x^{\mathrm{T}}y = 0$.

Let $x \in \mathbb{C}^n$. Then, $x = y + \jmath z$, where $y, z \in \mathbb{R}^n$. Therefore, the transpose x^{T} of x is given by

$$x^{\mathrm{T}} = y^{\mathrm{T}} + \jmath z^{\mathrm{T}}. \tag{2.2.5}$$

The *complex conjugate* $\overline{x}$ of x is defined by

$$\overline{x} \triangleq y - \jmath z, \tag{2.2.6}$$

while the *complex conjugate transpose* x^* of x is defined by

$$x^* \triangleq \overline{x}^{\mathrm{T}} = y^{\mathrm{T}} - \jmath z^{\mathrm{T}}. \tag{2.2.7}$$

The vectors y and z are the *real* and *imaginary* parts $\operatorname{Re} x$ and $\operatorname{Im} x$ of x, respectively, which are denoted by

$$\operatorname{Re} x \triangleq \tfrac{1}{2}(x + \overline{x}) = y \tag{2.2.8}$$

and

$$\operatorname{Im} x \triangleq \tfrac{1}{2\jmath}(x - \overline{x}) = z. \tag{2.2.9}$$

Note that

$$x^*x = \sum_{i=1}^{n} \overline{x}_{(i)} x_{(i)} = \sum_{i=1}^{n} |x_{(i)}|^2 = \sum_{i=1}^{n} \left[y_{(i)}^2 + z_{(i)}^2 \right]. \tag{2.2.10}$$

If $w, x \in \mathbb{C}^n$, then $w^{\mathrm{T}}x = x^{\mathrm{T}}w$.

Lemma 2.2.2. Let $x \in \mathbb{C}^n$. Then, $x^*x = 0$ if and only if $x = 0$.

Let $x, y \in \mathbb{C}^n$. Then, $x^*y \in \mathbb{C}$ is the *inner product* of x and y, which is given by

$$x^*y = \sum_{i=1}^{n} \overline{x}_{(i)} y_{(i)}. \tag{2.2.11}$$

Furthermore, x is *orthogonal* to y if $x^*y = 0$.

Let $A \in \mathbb{F}^{n\times m}$. Then, the *transpose* $A^{\mathrm{T}} \in \mathbb{F}^{m\times n}$ of A is defined by

$$A^{\mathrm{T}} \triangleq \begin{bmatrix} [\mathrm{row}_1(A)]^{\mathrm{T}} & \cdots & [\mathrm{row}_n(A)]^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} [\mathrm{col}_1(A)]^{\mathrm{T}} \\ \vdots \\ [\mathrm{col}_m(A)]^{\mathrm{T}} \end{bmatrix}, \tag{2.2.12}$$

that is, $\mathrm{col}_i\big(A^{\mathrm{T}}\big) = [\mathrm{row}_i(A)]^{\mathrm{T}}$ for all $i = 1,\ldots,n$ and $\mathrm{row}_i\big(A^{\mathrm{T}}\big) = [\mathrm{col}_i(A)]^{\mathrm{T}}$ for all $i = 1,\ldots,m$. Hence, $\big(A^{\mathrm{T}}\big)_{(i,j)} = A_{(j,i)}$ and $\big(A^{\mathrm{T}}\big)^{\mathrm{T}} = A$. If $B \in \mathbb{F}^{m\times l}$, then

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}. \tag{2.2.13}$$

In particular, if $x \in \mathbb{F}^m$, then

$$(Ax)^{\mathrm{T}} = x^{\mathrm{T}}A^{\mathrm{T}}, \tag{2.2.14}$$

while if, in addition, $y \in \mathbb{F}^n$, then $y^{\mathrm{T}}Ax$ is a scalar and

$$y^{\mathrm{T}}Ax = \big(y^{\mathrm{T}}Ax\big)^{\mathrm{T}} = x^{\mathrm{T}}A^{\mathrm{T}}y. \tag{2.2.15}$$

If $B \in \mathbb{F}^{n\times m}$, then, for all $\alpha, \beta \in \mathbb{F}$,

$$(\alpha A + \beta B)^{\mathrm{T}} = \alpha A^{\mathrm{T}} + \beta B^{\mathrm{T}}. \tag{2.2.16}$$

Let $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^n$. Then, the matrix $xy^{\mathrm{T}} \in \mathbb{F}^{n\times m}$ is the *outer product* of x and y. The outer product xy^{T} is nonzero if and only if both x and y are nonzero.

The *trace* of a square matrix $A \in \mathbb{F}^{n\times n}$, denoted by $\mathrm{tr}\,A$, is defined to be the sum of its diagonal entries, that is,

$$\mathrm{tr}\,A \triangleq \sum_{i=1}^{n} A_{(i,i)}. \tag{2.2.17}$$

Note that

$$\mathrm{tr}\,A = \mathrm{tr}\,A^{\mathrm{T}}. \tag{2.2.18}$$

Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$. Then, AB and BA are square,

$$\mathrm{tr}\,AB = \mathrm{tr}\,BA = \mathrm{tr}\,A^{\mathrm{T}}B^{\mathrm{T}} = \mathrm{tr}\,B^{\mathrm{T}}A^{\mathrm{T}} = \sum_{i,j=1}^{n,m} A_{(i,j)}B_{(j,i)}, \tag{2.2.19}$$

and

$$\mathrm{tr}\,AA^{\mathrm{T}} = \mathrm{tr}\,A^{\mathrm{T}}A = \sum_{i,j=1}^{n,m} A_{(i,j)}^2. \tag{2.2.20}$$

Furthermore, if $n = m$, then, for all $\alpha, \beta \in \mathbb{F}$,

$$\mathrm{tr}(\alpha A + \beta B) = \alpha\,\mathrm{tr}\,A + \beta\,\mathrm{tr}\,B. \tag{2.2.21}$$

Lemma 2.2.3. Let $A \in \mathbb{R}^{n\times m}$. Then, $\operatorname{tr} A^{\mathrm{T}}A = 0$ if and only if $A = 0$.

Let $A, B \in \mathbb{R}^{n\times m}$. Then, the *inner product* of A and B is $\operatorname{tr} A^{\mathrm{T}}B$. Furthermore, A is *orthogonal* to B if $\operatorname{tr} A^{\mathrm{T}}B = 0$.

Let $C \in \mathbb{C}^{n\times m}$. Then, $C = A + \jmath B$, where $A, B \in \mathbb{R}^{n\times m}$. Therefore, the transpose C^{T} of C is given by

$$C^{\mathrm{T}} = A^{\mathrm{T}} + \jmath B^{\mathrm{T}}. \tag{2.2.22}$$

The *complex conjugate* $\overline{C}$ of C is

$$\overline{C} \triangleq A - \jmath B, \tag{2.2.23}$$

while the *complex conjugate transpose* C^* of C is

$$C^* \triangleq \overline{C}^{\mathrm{T}} = A^{\mathrm{T}} - \jmath B^{\mathrm{T}}. \tag{2.2.24}$$

Note that $\overline{C} = C$ if and only if $B = 0$, and that

$$\left(C^{\mathrm{T}}\right)^{\mathrm{T}} = \overline{\overline{C}} = (C^*)^* = C. \tag{2.2.25}$$

The matrices A and B are the real and imaginary parts $\operatorname{Re} C$ and $\operatorname{Im} C$ of C, respectively, which are denoted by

$$\operatorname{Re} C \triangleq \tfrac{1}{2}\left(C + \overline{C}\right) = A, \tag{2.2.26}$$

and

$$\operatorname{Im} C \triangleq \tfrac{1}{2\jmath}\left(C - \overline{C}\right) = B. \tag{2.2.27}$$

If C is square, then

$$\operatorname{tr} C = \operatorname{tr} A + \jmath \operatorname{tr} B. \tag{2.2.28}$$

If $\mathcal{S} \subseteq \mathbb{C}^{n\times m}$, then

$$\overline{\mathcal{S}} \triangleq \left\{\overline{A}\colon\ A \in \mathcal{S}\right\}. \tag{2.2.29}$$

If $\mathcal{S}$ is a multiset with elements in $\mathbb{C}^{n\times m}$, then

$$\overline{\mathcal{S}} = \left\{\overline{A}\colon\ A \in \mathcal{S}\right\}_{\mathrm{m}}. \tag{2.2.30}$$

Lemma 2.2.4. Let $A \in \mathbb{C}^{n\times m}$. Then, $\operatorname{tr} A^*A = 0$ if and only if $A = 0$.

Let $A, B \in \mathbb{C}^{n\times m}$. Then, the *inner product* of A and B is $\operatorname{tr} A^*B$. Furthermore, A is *orthogonal* to B if $\operatorname{tr} A^*B = 0$.

If $A, B \in \mathbb{C}^{n\times m}$, then, for all $\alpha, \beta \in \mathbb{C}$,

$$(\alpha A + \beta B)^* = \overline{\alpha} A^* + \overline{\beta} B^*, \tag{2.2.31}$$

while, if $A \in \mathbb{C}^{n\times m}$ and $B \in \mathbb{C}^{m\times l}$, then

$$\overline{AB} = \overline{A}\,\overline{B} \tag{2.2.32}$$

and

$$(AB)^* = B^*A^*. \tag{2.2.33}$$

In particular, if $A \in \mathbb{C}^{n\times m}$ and $x \in \mathbb{C}^m$, then

$$(Ax)^* = x^*A^*, \tag{2.2.34}$$

while if, in addition, $y \in \mathbb{C}^n$, then

$$y^*Ax = (y^*Ax)^{\mathrm{T}} = x^{\mathrm{T}}A^{\mathrm{T}}\overline{y} \tag{2.2.35}$$

and

$$(y^*Ax)^* = \left(\overline{y^*Ax}\right)^{\mathrm{T}} = \left(y^{\mathrm{T}}\overline{A}\overline{x}\right)^{\mathrm{T}} = x^*A^*y. \tag{2.2.36}$$

For $A \in \mathbb{F}^{n\times m}$ define the *reverse transpose* of A by

$$A^{\hat{\mathrm{T}}} \triangleq \hat{I}_m A^{\mathrm{T}} \hat{I}_n \tag{2.2.37}$$

and the *reverse complex conjugate transpose* of A by

$$A^{\hat{*}} \triangleq \hat{I}_m A^* \hat{I}_n. \tag{2.2.38}$$

For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^{\hat{\mathrm{T}}} = \begin{bmatrix} 6 & 3 \\ 5 & 2 \\ 4 & 1 \end{bmatrix}. \tag{2.2.39}$$

In general,

$$(A^*)^{\hat{*}} = \left(A^{\hat{*}}\right)^* = \left(A^{\mathrm{T}}\right)^{\hat{\mathrm{T}}} = \left(A^{\hat{\mathrm{T}}}\right)^{\mathrm{T}} = \hat{I}_n A \hat{I}_m \tag{2.2.40}$$

and

$$\left(A^{\hat{*}}\right)^{\hat{*}} = \left(A^{\hat{\mathrm{T}}}\right)^{\hat{\mathrm{T}}} = A. \tag{2.2.41}$$

Note that if $B \in \mathbb{F}^{m\times l}$, then

$$(AB)^{\hat{*}} = B^{\hat{*}}A^{\hat{*}} \tag{2.2.42}$$

and

$$(AB)^{\hat{\mathrm{T}}} = B^{\hat{\mathrm{T}}}A^{\hat{\mathrm{T}}}. \tag{2.2.43}$$

2.3 Convex Sets, Cones, and Subspaces

Let $\mathcal{S} \subseteq \mathbb{F}^n$. If $\alpha \in \mathbb{F}$, then $\alpha\mathcal{S} \triangleq \{\alpha x\colon\ x \in \mathcal{S}\}$ and, if $y \in \mathbb{F}^n$, then $y + \mathcal{S} = \{y + x\colon\ x \in \mathcal{S}\}$. We write $-\mathcal{S}$ for $(-1)\mathcal{S}$. The set $\mathcal{S}$ is *symmetric* if $\mathcal{S} = -\mathcal{S}$, that is, $x \in \mathcal{S}$ if and only if $-x \in \mathcal{S}$. For $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ define $\mathcal{S}_1 + \mathcal{S}_2 \triangleq \{x + y\colon\ x \in \mathcal{S}_1 \text{ and } y \in \mathcal{S}_2\}$.

If $x, y \in \mathbb{F}^n$ and $\alpha \in [0, 1]$, then $\alpha x + (1-\alpha)y$ is a *convex combination* of x and y with *barycentric coordinates* α and $1-\alpha$. $\mathcal{S} \subseteq \mathbb{F}^n$ is *convex* if, for all $x, y \in \mathcal{S}$, every convex combination of x and y is an element of $\mathcal{S}$.

Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then, $\mathcal{S}$ is a *cone* if, for all $x \in \mathcal{S}$ and all $\alpha > 0$, the vector αx is an element of $\mathcal{S}$. Now, assume that $\mathcal{S}$ is a cone. Then, $\mathcal{S}$ is *pointed* if $0 \in \mathcal{S}$, while $\mathcal{S}$ is *one-sided* if $x, -x \in \mathcal{S}$ implies that $x = 0$. Hence, $\mathcal{S}$ is one-sided if and only if $\mathcal{S} \cap -\mathcal{S} \subseteq \{0\}$. Finally, $\mathcal{S}$ is a *convex cone* if it is convex.

Let $\mathcal{S} \subseteq \mathbb{F}^n$ be nonempty. Then, $\mathcal{S}$ is a *subspace* if, for all $x, y \in \mathcal{S}$ and $\alpha, \beta \in \mathbb{F}$, the vector $\alpha x + \beta y$ is an element of $\mathcal{S}$. Note that if $\{x_1, \ldots, x_r\} \subset \mathbb{F}^n$, then the set $\{\sum_{i=1}^r \alpha_i x_i\colon\ \alpha_1, \ldots, \alpha_r \in \mathbb{F}\}$ is a subspace. In addition, $\mathcal{S}$ is an *affine subspace* if there exists $z \in \mathbb{F}^n$ such that $\mathcal{S} + z$ is a subspace. Affine subspaces $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ are *parallel* if there exists $z \in \mathbb{F}^n$ such that $\mathcal{S}_1 + z = \mathcal{S}_2$. If $\mathcal{S}$ is an affine subspace, then there exists a unique subspace parallel to $\mathcal{S}$. Trivially, the empty set is a convex cone, although it is neither a subspace nor an affine subspace. All of these definitions also apply to subsets of $\mathbb{F}^{n\times m}$.

Let $\mathcal{S} \subseteq \mathbb{F}^n$. The *convex hull* of $\mathcal{S}$, denoted by $\operatorname{co}\mathcal{S}$, is the smallest convex set containing $\mathcal{S}$. Hence, $\operatorname{co}\mathcal{S}$ is the intersection of all convex subsets of $\mathbb{F}^n$ that contain $\mathcal{S}$. The *conical hull* of $\mathcal{S}$, denoted by $\operatorname{cone}\mathcal{S}$, is the smallest cone in $\mathbb{F}^n$ containing $\mathcal{S}$, while the *convex conical hull* of $\mathcal{S}$, denoted by $\operatorname{coco}\mathcal{S}$, is the smallest convex cone in $\mathbb{F}^n$ containing $\mathcal{S}$. If $\mathcal{S}$ has a finite number of elements, then $\operatorname{co}\mathcal{S}$ is a *polytope* and coco is a *polyhedral convex cone*. The *span* of $\mathcal{S}$, denoted by $\operatorname{span}\mathcal{S}$, is the smallest subspace in $\mathbb{F}^n$ containing $\mathcal{S}$, while, if $\mathcal{S}$ is nonempty, then the *affine hull* of $\mathcal{S}$, denoted by $\operatorname{aff}\mathcal{S}$, is the smallest affine subspace in $\mathbb{F}^n$ containing $\mathcal{S}$. Note that $\mathcal{S}$ is convex if and only if $\mathcal{S} = \operatorname{co}\mathcal{S}$, while similar statements hold for $\operatorname{cone}\mathcal{S}$, $\operatorname{coco}\mathcal{S}$, $\operatorname{span}\mathcal{S}$, and $\operatorname{aff}\mathcal{S}$. Trivially, $\operatorname{co}\varnothing = \operatorname{cone}\varnothing = \operatorname{coco}\varnothing = \varnothing$, whereas, viewing $\varnothing \subset \mathbb{F}^n$, it follows that $\operatorname{span}\varnothing = \{0_{n\times 1}\}$. We define $\operatorname{aff}\varnothing \triangleq \{0_{n\times 1}\}$. All of these definitions also apply to subsets of $\mathbb{F}^{n\times m}$.

Let $x_1, \ldots, x_r \in \mathbb{F}^n$. Then, $x_1, \ldots, x_r$ are *linearly independent* if $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$ and

$$\sum_{i=1}^r \alpha_i x_i = 0, \tag{2.3.1}$$

imply that $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. Clearly, $x_1, \ldots, x_r$ is linearly independent if and only if $\overline{x_1}, \ldots, \overline{x_r}$ are linearly independent. If $x_1, \ldots, x_r$ are not linearly independent, then $x_1, \ldots, x_r$ are *linearly dependent*. Note that $\{0_{n\times 1}\}$ is linearly dependent.

Let $\mathcal{S} \subseteq \mathbb{F}^n$. If $\mathcal{S}$ is a subspace not equal to $\{0_{n\times 1}\}$, then there exist $x_1, \ldots, x_r \in \mathbb{F}^n$ such that $x_1, \ldots, x_r$ are linearly independent over $\mathbb{F}$ and such that $\operatorname{span}\{x_1, \ldots, x_r\} = \mathcal{S}$. The set of vectors $\{x_1, \ldots, x_r\}$ is a *basis* for $\mathcal{S}$. The positive integer r, which is the *dimension* of the subspace $\mathcal{S}$, is uniquely defined. The dimension of $\mathcal{S} = \{0_{n\times 1}\}$ is defined to be zero since $\operatorname{span}\varnothing = \{0_{n\times 1}\}$. The *dimension* of an arbitrary set $\mathcal{S} \subseteq \mathbb{F}^n$, denoted by $\dim \mathcal{S}$, is the dimension of the subspace parallel to $\operatorname{aff}\mathcal{S}$. We define $\dim \varnothing \triangleq -\infty$.

The following result is the *dimension theorem*.

Theorem 2.3.1. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then,

$$\dim(\mathcal{S}_1 + \mathcal{S}_2) + \dim(\mathcal{S}_1 \cap \mathcal{S}_2) = \dim \mathcal{S}_1 + \dim \mathcal{S}_2. \tag{2.3.2}$$

Proof. See [262, p. 227]. □

Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1$ and $\mathcal{S}_2$ are *complementary* if $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$ and $\mathcal{S}_1 + \mathcal{S}_2 = \mathbb{F}^n$. In this case, we say that $\mathcal{S}_1$ is complementary to $\mathcal{S}_2$, or vice versa.

Corollary 2.3.2. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1, \mathcal{S}_2$ are complementary if and only if $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$ and

$$\dim \mathcal{S}_1 + \dim \mathcal{S}_2 = n. \tag{2.3.3}$$

Let $\mathcal{S} \subseteq \mathbb{F}^n$ be nonempty. Then, the *orthogonal complement* $\mathcal{S}^\perp$ of $\mathcal{S}$ is defined by

$$\mathcal{S}^\perp \triangleq \{x \in \mathbb{F}^n\colon\; x^*y = 0 \text{ for all } y \in \mathcal{S}\}. \tag{2.3.4}$$

The orthogonal complement $\mathcal{S}^\perp$ of $\mathcal{S}$ is a subspace even if $\mathcal{S}$ is not.

Let $y \in \mathbb{F}^n$ be nonzero. Then, the subspace $\{y\}^\perp$, whose dimension is $n-1$, is a *hyperplane*. Furthermore, $\mathcal{S}$ is an *affine hyperplane* if there exists $z \in \mathbb{F}^n$ such that $\mathcal{S} + z$ is a hyperplane. The set $\{x \in \mathbb{F}^n\colon\; \operatorname{Re} x^*y \le 0\}$ is a *closed half space*, while the set $\{x \in \mathbb{F}^n\colon\; \operatorname{Re} x^*y < 0\}$ is an *open half space*. Finally, $\mathcal{S}$ is an *affine (closed, open) half space* if there exists $z \in \mathbb{F}^n$ such that $\mathcal{S} + z$ is a (closed, open) half space.

Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then,

$$\operatorname{dcone}\mathcal{S} \triangleq \{x \in \mathbb{F}^n\colon\; \operatorname{Re} x^*y \le 0 \text{ for all } y \in \mathcal{S}\} \tag{2.3.5}$$

is the *dual cone* of $\mathcal{S}$. Note that $\operatorname{dcone}\mathcal{S}$ is a pointed convex cone and that $\operatorname{dcone}\mathcal{S} = \operatorname{dcone}\operatorname{cone}\mathcal{S} = \operatorname{dcone}\operatorname{coco}\mathcal{S}$.

Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1$ and $\mathcal{S}_2$ are *orthogonally complementary* if $\mathcal{S}_1$ and $\mathcal{S}_2$ are complementary and $x^*y = 0$ for all $x \in \mathcal{S}_1$ and $y \in \mathcal{S}_2$.

Proposition 2.3.3. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1$ and $\mathcal{S}_2$ are orthogonally complementary if and only if $\mathcal{S}_1 = \mathcal{S}_2^\perp$.

For the next result, note that "$\subset$" indicates proper inclusion.

Lemma 2.3.4. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces such that $\mathcal{S}_1 \subseteq \mathcal{S}_2$. Then, $\mathcal{S}_1 \subset \mathcal{S}_2$ if and only if $\dim \mathcal{S}_1 < \dim \mathcal{S}_2$. Equivalently, $\mathcal{S}_1 = \mathcal{S}_2$ if and only if $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$.

The following result provides constructive characterizations of $\operatorname{co} \mathcal{S}$, $\operatorname{cone} \mathcal{S}$, $\operatorname{coco} \mathcal{S}$, $\operatorname{span} \mathcal{S}$, and $\operatorname{aff} \mathcal{S}$.

Theorem 2.3.5. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be nonempty. Then,

$$\operatorname{co} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i \colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{k} \alpha_i = 1 \right\} \tag{2.3.6}$$

$$= \left\{ \sum_{i=1}^{n+1} \alpha_i x_i \colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{n+1} \alpha_i = 1 \right\}, \tag{2.3.7}$$

$$\operatorname{cone} \mathcal{S} = \{\alpha x\colon\ x \in \mathcal{S} \text{ and } \alpha > 0\}, \tag{2.3.8}$$

$$\operatorname{coco} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i \colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{k} \alpha_i > 0 \right\} \tag{2.3.9}$$

$$= \left\{ \sum_{i=1}^{n} \alpha_i x_i \colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{n} \alpha_i > 0 \right\}, \tag{2.3.10}$$

$$\operatorname{span} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i \colon\ \alpha_i \in \mathbb{R} \text{ and } x_i \in \mathcal{S} \right\} \tag{2.3.11}$$

$$= \left\{ \sum_{i=1}^{n} \alpha_i x_i \colon\ \alpha_i \in \mathbb{R} \text{ and } x_i \in \mathcal{S} \right\}, \tag{2.3.12}$$

$$\operatorname{aff} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i \colon\ \alpha_i \in \mathbb{R},\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{k} \alpha_i = 1 \right\} \tag{2.3.13}$$

$$= \left\{ \sum_{i=1}^{n+1} \alpha_i x_i \colon\ \alpha_i \in \mathbb{R},\ x_i \in \mathcal{S},\ \text{and } \sum_{i=1}^{n+1} \alpha_i = 1 \right\}. \tag{2.3.14}$$

(2.3.15)

Now, let $\mathcal{S} \subseteq \mathbb{C}^n$. Then,

$$\operatorname{co} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i\colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{k} \alpha_i = 1 \right\} \tag{2.3.16}$$

$$= \left\{ \sum_{i=1}^{2n+1} \alpha_i x_i\colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{2n+1} \alpha_i = 1 \right\}, \tag{2.3.17}$$

$$\operatorname{cone} \mathcal{S} = \{\alpha x\colon\ x \in \mathcal{S} \text{ and } \alpha > 0\}, \tag{2.3.18}$$

$$\operatorname{coco} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i\colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{k} \alpha_i > 0 \right\} \tag{2.3.19}$$

$$= \left\{ \sum_{i=1}^{2n} \alpha_i x_i\colon\ \alpha_i \ge 0,\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{2n} \alpha_i > 0 \right\}, \tag{2.3.20}$$

$$\operatorname{span} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i\colon\ \alpha_i \in \mathbb{C} \text{ and } x_i \in \mathcal{S} \right\} \tag{2.3.21}$$

$$= \left\{ \sum_{i=1}^{n} \alpha_i x_i\colon\ \alpha_i \in \mathbb{C} \text{ and } x_i \in \mathcal{S} \right\}, \tag{2.3.22}$$

$$\operatorname{aff} \mathcal{S} = \bigcup_{k\in\mathbb{P}} \left\{ \sum_{i=1}^{k} \alpha_i x_i\colon\ \alpha_i \in \mathbb{C},\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{k} \alpha_i = 1 \right\} \tag{2.3.23}$$

$$= \left\{ \sum_{i=1}^{n+1} \alpha_i x_i\colon\ \alpha_i \in \mathbb{C},\ x_i \in \mathcal{S},\ \text{and}\ \sum_{i=1}^{n+1} \alpha_i = 1 \right\}. \tag{2.3.24}$$

Proof. Result (2.3.6) is immediate, while (2.3.7) is proved in [357, p. 17]. Furthermore, (2.3.8) is immediate. Next, note that, since $\operatorname{coco} \mathcal{S} = \operatorname{co} \operatorname{cone} \mathcal{S}$, it follows that (2.3.6) and (2.3.8) imply (2.3.10) with n replaced by $n+1$. However, every element of $\operatorname{coco} \mathcal{S}$ lies in the convex hull of $n+1$ points one of which is the origin. It thus follows that we can set $x_{n+1} = 0$, which yields (2.3.10). Similar arguments yield (2.3.12). Finally, note that all vectors of the form $x_1 + \beta(x_2 - x_1)$, where $x_1, x_2 \in \mathcal{S}$ and $\beta \in \mathbb{R}$, are elements of $\operatorname{aff} \mathcal{S}$. Forming the convex hull of these vectors yields (2.3.14). □

The following result shows that cones can be used to induce relations on $\mathbb{F}^n$.

Proposition 2.3.6. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be a cone and, for $x, y \in \mathbb{F}^n$, let $x \le y$ denote the relation $y - x \in \mathcal{S}$. Then, the following statements hold:

i) "$\le$" is reflexive if and only if $\mathcal{S}$ is a pointed cone.

ii) "$\le$" is antisymmetric if and only if $\mathcal{S}$ is a one-sided cone.

iii) "$\le$" is symmetric if and only if $\mathcal{S}$ is a symmetric cone.

iv) "$\le$" is transitive if and only if $\mathcal{S}$ is a convex cone.

Proof. The proofs of *i)*, *ii)* and *iii)* are immediate. To prove *iv)*, suppose that "$\le$" is transitive, and let $x, y \in \mathcal{S}$ so that $0 \le \alpha x \le \alpha x + (1-\alpha)y$ for all $\alpha \in [0,1]$. Hence, $\alpha x + (1-\alpha)y \in \mathcal{S}$ for all $\alpha \in [0,1]$, and thus $\mathcal{S}$ is convex. Conversely, suppose that $\mathcal{S}$ is a convex cone, and assume that $x \le y$ and $y \le z$. Then, $y - x \in \mathcal{S}$ and $z - y \in \mathcal{S}$ imply that $z - x = 2\big[\frac{1}{2}(y-x) + \frac{1}{2}(z-y)\big] \in \mathcal{S}$. Hence, $x \le z$, and thus "$\le$" is transitive. □

2.4 Range and Null Space

Two important features of a matrix $A \in \mathbb{F}^{n\times m}$ are its range and null space, denoted by $\mathcal{R}(A)$ and $\mathcal{N}(A)$, respectively. The *range* of A is defined by

$$\mathcal{R}(A) \triangleq \{Ax\colon\ x \in \mathbb{F}^m\}. \tag{2.4.1}$$

Note that $\mathcal{R}(0_{n\times 0}) = \{0_{n\times 1}\}$ and $\mathcal{R}(0_{0\times m}) = \{0_{0\times 1}\}$. Letting α_i denote $x_{(i)}$, it can be seen that

$$\mathcal{R}(A) = \left\{\sum_{i=1}^{m} \alpha_i \mathrm{col}_i(A)\colon\ \alpha_1, \ldots, \alpha_m \in \mathbb{F}\right\}, \tag{2.4.2}$$

which shows that $\mathcal{R}(A)$ is a subspace of $\mathbb{F}^n$. It thus follows from Theorem 2.3.5 that

$$\mathcal{R}(A) = \mathrm{span}\,\{\mathrm{col}_1(A), \ldots, \mathrm{col}_m(A)\}. \tag{2.4.3}$$

By viewing A as a function from $\mathbb{F}^m$ into $\mathbb{F}^n$, we can also write $\mathcal{R}(A) = A\mathbb{F}^m$.

The *null space* of $A \in \mathbb{F}^{n\times m}$ is defined by

$$\mathcal{N}(A) \triangleq \{x \in \mathbb{F}^m\colon\ Ax = 0\}. \tag{2.4.4}$$

Note that $\mathcal{N}(0_{n\times 0}) = \mathbb{F}^0 = \{0_{0\times 1}\}$ and $\mathcal{N}(0_{0\times m}) = \mathbb{F}^m$. Equivalently,

$$\mathcal{N}(A) = \Big\{x \in \mathbb{F}^m\colon\ x^{\mathrm{T}}[\mathrm{row}_i(A)]^{\mathrm{T}} = 0 \text{ for all } i = 1, \ldots, n\Big\} \tag{2.4.5}$$

$$= \Big\{[\mathrm{row}_1(A)]^{\mathrm{T}}, \ldots, [\mathrm{row}_n(A)]^{\mathrm{T}}\Big\}^{\perp}, \tag{2.4.6}$$

which shows that $\mathcal{N}(A)$ is a subspace of $\mathbb{F}^m$. Note that if $\alpha \in \mathbb{F}$ is nonzero, then $\mathcal{R}(\alpha A) = \mathcal{R}(A)$ and $\mathcal{N}(\alpha A) = \mathcal{N}(A)$. Finally, if $\mathbb{F} = \mathbb{C}$, then $\mathcal{R}(A)$ and $\mathcal{R}(\overline{A})$ are not necessarily identical. For example, let $A \triangleq \left[\begin{smallmatrix} \jmath \\ 1 \end{smallmatrix}\right]$.

Let $A \in \mathbb{F}^{n\times n}$, and let $\mathcal{S} \subseteq \mathbb{F}^n$ be a subspace. Then, $\mathcal{S}$ is an *invariant subspace* of A if $A\mathcal{S} \subseteq \mathcal{S}$. Note that $A\mathcal{R}(A) \subseteq A\mathbb{F}^m = \mathcal{R}(A)$ and $A\mathcal{N}(A) = \{0_n\} \subseteq \mathcal{N}(A)$. Hence, $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are invariant subspaces of A.

If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, then it is easy to see that

$$\mathcal{R}(AB) = A\mathcal{R}(B). \tag{2.4.7}$$

Hence, the following result is not surprising.

Lemma 2.4.1. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $C \in \mathbb{F}^{k\times n}$. Then,

$$\mathcal{R}(AB) \subseteq \mathcal{R}(A) \tag{2.4.8}$$

and

$$\mathcal{N}(A) \subseteq \mathcal{N}(CA). \tag{2.4.9}$$

Proof. Since $\mathcal{R}(B) \subseteq \mathbb{F}^m$, it follows that $\mathcal{R}(AB) = A\mathcal{R}(B) \subseteq A\mathbb{F}^m = \mathcal{R}(A)$. Furthermore, $y \in \mathcal{N}(A)$ implies that $Ay = 0$, and thus $CAy = 0$. □

Corollary 2.4.2. Let $A \in \mathbb{F}^{n\times n}$, and let $k \in \mathbb{P}$. Then,

$$\mathcal{R}\Big(A^k\Big) \subseteq \mathcal{R}(A) \tag{2.4.10}$$

and

$$\mathcal{N}(A) \subseteq \mathcal{N}\Big(A^k\Big). \tag{2.4.11}$$

Although $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$ for arbitrary conformable matrices A, B, we now show that equality holds in the special case $B = A^*$. This result, along with others, is the subject of the following basic theorem.

Theorem 2.4.3. Let $A \in \mathbb{F}^{n\times m}$. Then, the following identities hold:

i) $\mathcal{R}(A)^\perp = \mathcal{N}(A^*)$.

ii) $\mathcal{R}(A) = \mathcal{R}(AA^*)$.

iii) $\mathcal{N}(A) = \mathcal{N}(A^*A)$.

Proof. To prove *i*), we first show that $\mathcal{R}(A)^\perp \subseteq \mathcal{N}(A^*)$. Let $x \in \mathcal{R}(A)^\perp$. Then, $x^*z = 0$ for all $z \in \mathcal{R}(A)$. Hence, $x^*Ay = 0$ for all $y \in \mathbb{R}^m$. Equivalently, $y^*A^*x = 0$ for all $y \in \mathbb{R}^m$. Letting $y = A^*x$, it follows that $x^*AA^*x = 0$. Now, Lemma 2.2.2 implies that $A^*x = 0$. Thus, $x \in \mathcal{N}(A^*)$. Conversely, let us show that $\mathcal{N}(A^*) \subseteq \mathcal{R}(A)^\perp$. Letting $x \in \mathcal{N}(A^*)$, it follows that $A^*x = 0$, and, hence, $y^*A^*x = 0$ for all $y \in \mathbb{R}^m$. Equivalently, $x^*Ay = 0$ for all $y \in \mathbb{R}^m$. Hence, $x^*z = 0$ for all $z \in \mathcal{R}(A)$. Thus, $x \in \mathcal{R}(A)^\perp$, which proves *i*).

To prove *ii*), note that Lemma 2.4.1 with $B = A^*$ implies that $\mathcal{R}(AA^*) \subseteq \mathcal{R}(A)$. To show that $\mathcal{R}(A) \subseteq \mathcal{R}(AA^*)$, let $x \in \mathcal{R}(A)$, and suppose that

$x \notin \mathcal{R}(AA^*)$. Then, it follows from Proposition 2.3.3 that $x = x_1 + x_2$, where $x_1 \in \mathcal{R}(AA^*)$ and $x_2 \in \mathcal{R}(AA^*)^\perp$ with $x_2 \neq 0$. Thus, $x_2^*AA^*y = 0$ for all $y \in \mathbb{R}^n$, and setting $y = x_2$ yields $x_2^*AA^*x_2 = 0$. Hence, Lemma 2.2.2 implies that $A^*x_2 = 0$, so that, by *i*), $x_2 \in \mathcal{N}(A^*) = \mathcal{R}(A)^\perp$. Since $x \in \mathcal{R}(A)$, it follows that $0 = x_2^*x = x_2^*x_1 + x_2^*x_2$. However, $x_2^*x_1 = 0$ so that $x_2^*x_2 = 0$ and $x_2 = 0$, which is a contradiction. This proves *ii*).

To prove *iii*), note that *ii*) with A replaced by A^* implies that $\mathcal{R}(A^*A)^\perp = \mathcal{R}(A^*)^\perp$. Furthermore, replacing A by A^* in *i*) yields $\mathcal{R}(A^*)^\perp = \mathcal{N}(A)$. Hence, $\mathcal{N}(A) = \mathcal{R}(A^*A)^\perp$. Now, *i*) with A replaced by A^*A implies that $\mathcal{R}(A^*A)^\perp = \mathcal{N}(A^*A)$. Hence, $\mathcal{N}(A) = \mathcal{N}(A^*A)$, which proves *iii*). □

Result *i*) of Theorem 2.4.3 can be written equivalently as

$$\mathcal{N}(A)^\perp = \mathcal{R}(A^*), \tag{2.4.12}$$

$$\mathcal{N}(A) = \mathcal{R}(A^*)^\perp, \tag{2.4.13}$$

$$\mathcal{N}(A^*)^\perp = \mathcal{R}(A), \tag{2.4.14}$$

while replacing A by A^* in *ii*) and *iii*) of Theorem 2.4.3 yields

$$\mathcal{R}(A^*) = \mathcal{R}(A^*A), \tag{2.4.15}$$

$$\mathcal{N}(A^*) = \mathcal{N}(AA^*). \tag{2.4.16}$$

Using *ii*) of Theorem 2.4.3 and (2.4.15) it follows that

$$\mathcal{R}(AA^*A) = A\mathcal{R}(A^*A) = A\mathcal{R}(A^*) = \mathcal{R}(AA^*) = \mathcal{R}(A). \tag{2.4.17}$$

Letting $A \triangleq \begin{bmatrix} 1 & \jmath \end{bmatrix}$ shows that $\mathcal{R}(A)$ and $\mathcal{R}\big(AA^{\mathrm{T}}\big)$ are generally different.

2.5 Rank and Defect

The *rank* of $A \in \mathbb{F}^{n\times m}$ is defined by

$$\operatorname{rank} A \triangleq \dim \mathcal{R}(A). \tag{2.5.1}$$

It can be seen that the rank of A is equal to the number of linearly independent columns of A. Hence, $\operatorname{rank} A = \operatorname{rank} \overline{A}$, $\operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} A^*$, $\operatorname{rank} A \le m$, and $\operatorname{rank} A^{\mathrm{T}} \le n$. If $\operatorname{rank} A = m$, then A has *full column rank*, while if $\operatorname{rank} A^{\mathrm{T}} = n$, then A has *full row rank*. If A has either full column rank or full row rank, then A has *full rank*. Finally, the *defect* of A is

$$\operatorname{def} A \triangleq \dim \mathcal{N}(A). \tag{2.5.2}$$

The following result follows from Theorem 2.4.3.

Corollary 2.5.1. Let $A \in \mathbb{F}^{n\times m}$. Then, the following identities hold:

i) $\operatorname{rank} A^* + \operatorname{def} A = m$.

ii) $\operatorname{rank} A = \operatorname{rank} AA^*$.

iii) $\operatorname{def} A = \operatorname{def} A^*A$.

Proof. It follows from (2.4.12) and Proposition 2.3.2 that $\operatorname{rank} A^* = \dim \mathcal{R}(A^*) = \dim \mathcal{N}(A)^\perp = m - \dim \mathcal{N}(A) = m - \operatorname{def} A$, which proves *i)*. Results *ii)* and *iii)* follow from of *ii)* and *iii)* of Theorem 2.4.3. □

Replacing A by A^* in Corollary 2.5.1 yields

$$\operatorname{rank} A + \operatorname{def} A^* = n, \tag{2.5.3}$$

$$\operatorname{rank} A^* = \operatorname{rank} A^*A, \tag{2.5.4}$$

$$\operatorname{def} A^* = \operatorname{def} AA^*. \tag{2.5.5}$$

Furthermore, note that

$$\operatorname{def} A = \operatorname{def} \overline{A} \tag{2.5.6}$$

and

$$\operatorname{def} A^{\mathrm{T}} = \operatorname{def} A^*. \tag{2.5.7}$$

Lemma 2.5.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\operatorname{rank} AB \le \min\{\operatorname{rank} A, \operatorname{rank} B\}. \tag{2.5.8}$$

Proof. Since, by Lemma 2.4.1, $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$, it follows that $\operatorname{rank} AB \le \operatorname{rank} A$. Next, suppose that $\operatorname{rank} B < \operatorname{rank} AB$. Let $\{y_1, \ldots, y_r\} \subset \mathbb{F}^n$ be a basis for $\mathcal{R}(AB)$, where $r \triangleq \operatorname{rank} AB$, and, since $y_i \in A\mathcal{R}(B)$ for all $i = 1, \ldots, r$, let $x_i \in \mathcal{R}(B)$ be such that $y_i = Ax_i$ for all $i = 1, \ldots, r$. Since $\operatorname{rank} B < r$, it follows that $x_1, \ldots, x_r$ are linearly dependent. Hence, there exist $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$, not all zero, such that $\sum_{i=1}^r \alpha_i x_i = 0$, which implies that $\sum_{i=1}^r \alpha_i A x_i = \sum_{i=1}^r \alpha_i y_i = 0$. Thus, $y_1, \ldots, y_r$ are linearly dependent, which is a contradiction. □

Corollary 2.5.3. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{rank} A = \operatorname{rank} A^* \tag{2.5.9}$$

and

$$\operatorname{def} A = \operatorname{def} A^* + m - n. \tag{2.5.10}$$

If, in addition, $n = m$, then

$$\operatorname{def} A = \operatorname{def} A^*. \tag{2.5.11}$$

Proof. It follows from (2.5.8) with $B = A^*$ that $\operatorname{rank} AA^* \le \operatorname{rank} A^*$. Furthermore, *ii)* of Corollary 2.5.1 implies that $\operatorname{rank} A = \operatorname{rank} AA^*$. Hence,

$\operatorname{rank} A \le \operatorname{rank} A^*$. Interchanging A and A^* and repeating this argument yields $\operatorname{rank} A^* \le \operatorname{rank} A$. Hence, $\operatorname{rank} A = \operatorname{rank} A^*$. Next, using *i*) of Corollary 2.5.1, (2.5.9), and (2.5.3) it follows that $\operatorname{def} A = m - \operatorname{rank} A^* = m - \operatorname{rank} A = m - (n - \operatorname{def} A^*)$, which proves (2.5.10). □

Corollary 2.5.4. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{rank} A \le \min\{m, n\}. \tag{2.5.12}$$

Proof. By definition, $\operatorname{rank} A \le m$, while it follows from (2.5.9) that $\operatorname{rank} A = \operatorname{rank} A^* \le n$. □

The *fundamental theorem of linear algebra* is given by (2.5.13) in the following result.

Corollary 2.5.5. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{rank} A + \operatorname{def} A = m \tag{2.5.13}$$

and

$$\operatorname{rank} A = \operatorname{rank} A^*A. \tag{2.5.14}$$

Proof. The result (2.5.13) follows from *i*) of Corollary 2.5.1 and (2.5.9), while (2.5.14) follows from (2.5.4) and (2.5.9). □

Corollary 2.5.6. Let $A \in \mathbb{F}^{n\times n}$ and $k \in \mathbb{P}$. Then,

$$\operatorname{rank} A^k \le \operatorname{rank} A \tag{2.5.15}$$

and

$$\operatorname{def} A \le \operatorname{def} A^k. \tag{2.5.16}$$

Proposition 2.5.7. Let $A \in \mathbb{F}^{n\times n}$. If $\operatorname{rank} A^2 = \operatorname{rank} A$, then $\operatorname{rank} A^k = \operatorname{rank} A$ for all $k \in \mathbb{P}$. Equivalently, if $\operatorname{def} A^2 = \operatorname{def} A$, then $\operatorname{def} A^k = \operatorname{def} A$ for all $k \in \mathbb{P}$.

Proof. Since $\operatorname{rank} A^2 = \operatorname{rank} A$ and $\mathcal{R}(A^2) \subseteq \mathcal{R}(A)$, it follows from Lemma 2.3.4 that $\mathcal{R}(A^2) = \mathcal{R}(A)$. Hence, $\mathcal{R}(A^3) = A\mathcal{R}(A^2) = A\mathcal{R}(A) = \mathcal{R}(A^2)$. Thus, $\operatorname{rank} A^3 = \operatorname{rank} A$. Similar arguments yield $\operatorname{rank} A^k = \operatorname{rank} A$ for all $k \in \mathbb{P}$. □

We now prove *Sylvester's inequality*, which provides a lower bound for the rank of the product of two matrices.

Proposition 2.5.8. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\operatorname{rank} A + \operatorname{rank} B \le m + \operatorname{rank} AB. \tag{2.5.17}$$

Proof. Using (2.5.8) it follows that

$$\begin{aligned}\operatorname{rank} A + \operatorname{rank} B &\le \operatorname{rank}\begin{bmatrix} 0 & A \\ B & I \end{bmatrix}\\ &= \operatorname{rank}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}\begin{bmatrix} -AB & 0 \\ B & I \end{bmatrix}\\ &\le \operatorname{rank}\begin{bmatrix} -AB & 0 \\ B & I \end{bmatrix}\\ &\le \operatorname{rank}\begin{bmatrix} -AB & 0 \end{bmatrix} + \operatorname{rank}\begin{bmatrix} B & I \end{bmatrix}\\ &= \operatorname{rank} AB + m.\end{aligned}$$ □

Combining (2.5.8) with (2.5.17) yields the following result.

Corollary 2.5.9. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\operatorname{rank} A + \operatorname{rank} B - m \le \operatorname{rank} AB \le \min\{\operatorname{rank} A, \operatorname{rank} B\}. \tag{2.5.18}$$

2.6 Invertibility

Let $A \in \mathbb{F}^{n\times m}$. Then, A is *left invertible* if there exists $A^{\mathrm{L}} \in \mathbb{F}^{m\times n}$ such that $A^{\mathrm{L}}A = I_m$, while A is *right invertible* if there exists $A^{\mathrm{R}} \in \mathbb{F}^{m\times n}$ such that $AA^{\mathrm{R}} = I_n$. These definitions are consistent with the definitions of left and right invertibility given in Chapter 1 applied to the function $f\colon\ \mathbb{F}^m \mapsto \mathbb{F}^n$ given by $f(x) = Ax$.

Theorem 2.6.1. Let $A \in \mathbb{F}^{n\times m}$. Then, the following statements are equivalent:

i) A is left invertible.

ii) A is one-to-one.

iii) $\operatorname{def} A = 0$.

iv) $\operatorname{rank} A = m$.

v) A has full column rank.

The following statements are also equivalent:

vi) A is right invertible.

vii) A is onto.

viii) $\operatorname{def} A = m - n$.

ix) $\operatorname{rank} A = n$.

x) A has full row rank.

Note that A is left invertible if and only if A^* is right invertible.

The following result shows that the rank and defect of a matrix are not affected by either left multiplication by a left invertible matrix or right multiplication by a right invertible matrix.

Proposition 2.6.2. Let $A \in \mathbb{F}^{n\times m}$, and let $C \in \mathbb{F}^{k\times n}$ be left invertible and $B \in \mathbb{F}^{m\times l}$ be right invertible. Then,

$$\operatorname{rank} A = \operatorname{rank} CA = \operatorname{rank} AB \tag{2.6.1}$$

and

$$\operatorname{def} A = \operatorname{def} CA = \operatorname{def} AB + m - l. \tag{2.6.2}$$

Proof. Let C^{L} be a left inverse of C. Using both inequalities in (2.5.18) and the fact that $\operatorname{rank} A \le n$, it follows that

$$\operatorname{rank} A = \operatorname{rank} A + \operatorname{rank} C^{\mathrm{L}}C - n \le \operatorname{rank} C^{\mathrm{L}}CA \le \operatorname{rank} CA \le \operatorname{rank} A,$$

which implies that $\operatorname{rank} A = \operatorname{rank} CA$. A similar argument implies that $\operatorname{rank} A = \operatorname{rank} AB$. Next, (2.5.13) and (2.6.1) imply that $m - \operatorname{def} A = m - \operatorname{def} CA = l - \operatorname{def} AB$, which yields (2.6.2). □

In general, left and right inverses are not unique. For example, the matrix $A = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ is left invertible and has left inverses $\begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \end{bmatrix}$. In spite of this nonuniqueness, however, left inverses are useful for solving equations of the form $Ax = b$, where $A \in \mathbb{F}^{n\times m}$, $x \in \mathbb{F}^m$, and $b \in \mathbb{F}^n$. If A is left invertible, then one can formally (but not rigorously) solve $Ax = b$ by noting that $x = A^{\mathrm{L}}Ax = A^{\mathrm{L}}b$, where $A^{\mathrm{L}} \in \mathbb{R}^{m\times n}$ is a left inverse of A. However, it is necessary to determine beforehand whether or not there actually exists a vector x satisfying $Ax = b$. For example, if $A = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ and $b = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, then A is left invertible but there does not exist x satisfying $Ax = b$. The following result addresses the various possibilities that can arise. One interesting feature of this result is that if there exists a solution to $Ax = b$ and A is left invertible, then the solution is unique even if A does not have a unique left inverse. For this result, $\begin{bmatrix} A & b \end{bmatrix}$ denotes the $n \times (m+1)$ partitioned matrix formed from A and b. Note that $\operatorname{rank} A \le \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} \le m + 1$, while $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$ is equivalent to $b \in \mathcal{R}(A)$.

Theorem 2.6.3. Let $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$. Then, the following statements hold:

i) There does not exist $x \in \mathbb{F}^m$ satisfying $Ax = b$ if and only if $\operatorname{rank} A < \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$.

ii) There exists a unique $x \in \mathbb{F}^m$ satisfying $Ax = b$ if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} = m$. In this case, if $A^{\mathrm{L}} \in \mathbb{F}^{m \times n}$ is a left inverse of A, then the solution is given by $x = A^{\mathrm{L}} b$.

iii) There exist infinitely many $x \in \mathbb{F}^m$ satisfying $Ax = b$ if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} < m$. In this case, let $\hat{x} \in \mathbb{F}^m$ satisfy $A\hat{x} = b$. Then, the set of solutions of $Ax = b$ is given by $\hat{x} + \mathcal{N}(A)$.

iv) Assume that $\operatorname{rank} A = n$. Then, there exists at least one $x \in \mathbb{F}^m$ satisfying $Ax = b$. Furthermore, if $A^{\mathrm{R}} \in \mathbb{F}^{m \times n}$ is a right inverse of A, then $x = A^{\mathrm{R}} b$ satisfies $Ax = b$. If $n = m$, then $x = A^{\mathrm{R}} b$ is the unique solution of $Ax = b$. If $n < m$ and $\hat{x} \in \mathbb{F}^n$ satisfies $A\hat{x} = b$, then the set of solutions of $Ax = b$ is given by $\hat{x} + \mathcal{N}(A)$.

Proof. To prove *i*) note that $\operatorname{rank} A < \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$ is equivalent to the fact that b cannot be represented as a linear combination of columns of A, that is, $Ax = b$ does not have a solution $x \in \mathbb{F}^m$. To prove *ii*), suppose that $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} = m$ so that, by *i*), $Ax = b$ has a solution $x \in \mathbb{F}^m$. If $\hat{x} \in \mathbb{F}^m$ satisfies $A\hat{x} = b$, then $A(x - \hat{x}) = 0$. Since $\operatorname{rank} A = m$, it follows from Theorem 2.6.1 that A has a left inverse $A^{\mathrm{L}} \in \mathbb{F}^{m \times n}$. Hence, $x - \hat{x} = A^{\mathrm{L}} A(x - \hat{x}) = 0$, which proves that $Ax = b$ has a unique solution. Conversely, suppose that $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} = m$ and there exist $x, \hat{x} \in \mathbb{F}^m$, where $x \neq \hat{x}$, such that $Ax = b$ and $A\hat{x} = b$. Then, $A(x - \hat{x}) = 0$, which implies that $\operatorname{def} A \geq 1$. Therefore, $\operatorname{rank} A = m - \operatorname{def} A \leq m - 1$, which is a contradiction. This proves the first statement of *ii*). Assuming $Ax = b$ has a unique solution $x \in \mathbb{F}^m$, multiplying by A^{L} yields $x = A^{\mathrm{L}} b$. To prove *iii*) note that it follows from *i*) that $Ax = b$ has at least one solution $\hat{x} \in \mathbb{F}^m$. Hence, $x \in \mathbb{F}^m$ is a solution of $Ax = b$ if and only if $A(x - \hat{x}) = 0$, or, equivalently, $x \in \hat{x} + \mathcal{N}(A)$. To prove *iv*) note that since $\operatorname{rank} A = n$, it follows that $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$ and thus either *ii*) or *iii*) applies. $\square$

The set of solutions $x \in \mathbb{F}^m$ to $Ax = b$ is explicitly characterized by Proposition 6.1.7.

Let $A \in \mathbb{F}^{n \times m}$. Then, A is *nonsingular* if there exists $B \in \mathbb{F}^{m \times n}$, the *inverse* of A, such that $AB = I_n$ and $BA = I_m$, that is, B is both a left and right inverse for A. It follows from Theorem 2.6.1 that if A is nonsingular, then $\operatorname{rank} A = m$ and $\operatorname{rank} A = n$ so that $m = n$. Hence, only square matrices can be nonsingular. Furthermore, the inverse $B \in \mathbb{F}^{n \times n}$ of $A \in \mathbb{F}^{n \times n}$ is unique since, if $C \in \mathbb{F}^{n \times n}$ is a left inverse of A, then $C = CI_n = CAB = I_n B = B$, while if $D \in \mathbb{F}^{n \times n}$ is a right inverse of A, then $D = I_n D = BAD = BI_n = B$. The following result follows from similar arguments and Theorem 2.6.1. This result can be viewed as a specialization of Theorem 1.2.3 to the function $f\colon\ \mathbb{F}^n \mapsto \mathbb{F}^n$, where $f(x) = Ax$.

Corollary 2.6.4. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is nonsingular.

ii) A has a unique inverse.

iii) A is one-to-one.

iv) A is onto.

v) A is left invertible.

vi) A is right invertible.

vii) A has a unique left inverse.

viii) A has a unique right inverse.

ix) $\operatorname{rank} A = n$.

x) $\operatorname{def} A = 0$.

Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, the inverse of A, denoted by A^{-1}, is a unique $n \times n$ matrix with entries in $\mathbb{F}$. If A is not nonsingular, then A is *singular.*

The following result is a specialization of Theorem 2.6.3 to the case $n = m$.

Corollary 2.6.5. Let $A \in \mathbb{F}^{n\times n}$ and $b \in \mathbb{F}^n$. Then, the following statements hold:

i) A is nonsingular if and only if there exists a unique $x \in \mathbb{F}^n$ satisfying $Ax = b$. In this case, $x = A^{-1}b$.

ii) A is singular and $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$ if and only if there exist infinitely many $x \in \mathbb{R}^n$ satisfying $Ax = b$. In this case, let $\hat{x} \in \mathbb{F}^m$ satisfy $A\hat{x} = b$. Then, the set of solutions of $Ax = b$ is given by $\hat{x} + \mathcal{N}(A)$.

Proposition 2.6.6. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is nonsingular.

ii) $\overline{A}$ is nonsingular.

iii) A^{T} is nonsingular.

iv) A^* is nonsingular.

In this case,

$$(\overline{A})^{-1} = \overline{A^{-1}}, \tag{2.6.3}$$

$$\left(A^{\mathrm{T}}\right)^{-1} = \left(A^{-1}\right)^{\mathrm{T}}, \tag{2.6.4}$$

$$(A^*)^{-1} = \left(A^{-1}\right)^*. \tag{2.6.5}$$

Proof. Since $AA^{-1} = I$, it follows that $\left(A^{-1}\right)^* A^* = I$. Hence, $\left(A^{-1}\right)^* = (A^*)^{-1}$. □

We thus use $A^{-\mathrm{T}}$ to denote $\left(A^{\mathrm{T}}\right)^{-1}$ or $\left(A^{-1}\right)^{\mathrm{T}}$ and A^{-*} to denote $(A^*)^{-1}$ or $\left(A^{-1}\right)^*$.

Proposition 2.6.7. Let $A, B \in \mathbb{F}^{n\times n}$ be nonsingular. Then,

$$(AB)^{-1} = B^{-1}A^{-1}, \tag{2.6.6}$$

$$(AB)^{-\mathrm{T}} = A^{-\mathrm{T}}B^{-\mathrm{T}}, \tag{2.6.7}$$

$$(AB)^{-*} = A^{-*}B^{-*}. \tag{2.6.8}$$

Proof. Note that $ABB^{-1}A^{-1} = AIA^{-1} = I$, which shows that $B^{-1}A^{-1}$ is the inverse of AB. Similarly, $(AB)^*A^{-*}B^{-*} = B^*A^*A^{-*}B^{-*} = B^*IB^{-*} = I$, which shows that $A^{-*}B^{-*}$ is the inverse of $(AB)^*$. □

For a nonsingular matrix $A \in \mathbb{F}^{n\times n}$ and $r \in \mathbb{Z}$ we write

$$A^{-r} \triangleq (A^r)^{-1} = \left(A^{-1}\right)^r, \tag{2.6.9}$$

$$A^{-r\mathrm{T}} \triangleq (A^r)^{-\mathrm{T}} = \left(A^{-\mathrm{T}}\right)^r = (A^{-r})^{\mathrm{T}} = \left(A^{\mathrm{T}}\right)^{-r}, \tag{2.6.10}$$

$$A^{-r*} \triangleq (A^r)^{-*} = (A^{-*})^r = (A^{-r})^* = (A^*)^{-r}. \tag{2.6.11}$$

For example, $A^{-2*} = (A^{-*})^2$.

2.7 Determinants

One of the most important quantities associated with a square matrix is its determinant. In this section we develop some basic results pertaining to the determinant of a matrix.

The *determinant* of $A \in \mathbb{F}^{n\times n}$ is defined by

$$\det A \triangleq \sum_{\sigma} (-1)^{N_\sigma} \prod_{i=1}^{n} A_{(i,\sigma(i))}, \tag{2.7.1}$$

where the sum is taken over all n permutations $\sigma = (\sigma(1), \ldots, \sigma(n))$ of the column indices $1, \ldots, n$, and where N_σ is the minimal number of pairwise transpositions needed to transform $\sigma(1), \ldots, \sigma(n)$ to $1, \ldots, n$. The following result is an immediate consequence of this definition.

Proposition 2.7.1. Let $A \in \mathbb{F}^{n \times n}$. Then,

$$\det A^{\mathrm{T}} = \det A, \tag{2.7.2}$$

$$\det \overline{A} = \overline{\det A}, \tag{2.7.3}$$

$$\det A^* = \overline{\det A}, \tag{2.7.4}$$

and, for all $\alpha \in \mathbb{F}$,

$$\det \alpha A = \alpha^n \det A. \tag{2.7.5}$$

If, in addition, $B \in \mathbb{F}^{m \times n}$ and $C \in \mathbb{F}^{m \times m}$, then

$$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = (\det A)(\det C). \tag{2.7.6}$$

The following observations are immediate consequences of the definition of the determinant.

Proposition 2.7.2. Let $A, B \in \mathbb{F}^{n \times n}$. Then, the following statements hold:

i) If all of the off-diagonal entries of A are zero, then

$$\det A = \prod_{i=1}^{n} A_{(i,i)}. \tag{2.7.7}$$

In particular, $\det I_n = 1$.

ii) If A has a row or column consisting entirely of zeros, then $\det A = 0$.

iii) If A has two identical rows or two identical columns, then $\det A = 0$.

iv) If $x \in \mathbb{F}^n$ and $i \in \{1, \ldots, n\}$, then

$$\det\big(A + x e_i^{\mathrm{T}}\big) = \det A + \det\Big(A \overset{i}{\leftarrow} x\Big). \tag{2.7.8}$$

v) If $x \in \mathbb{F}^{1 \times n}$ and $i \in \{1, \ldots, n\}$, then

$$\det(A + e_i x) = \det A + \det\Big(A \overset{i}{\leftarrow} x\Big). \tag{2.7.9}$$

vi) If B is identical to A except that, for some $i \in \{1, \ldots, n\}$ and $\alpha \in \mathbb{F}$, $\mathrm{col}_i(B) = \alpha \mathrm{col}_i(A)$ or $\mathrm{row}_i(B) = \alpha \mathrm{row}_i(A)$, then $\det B = \alpha \det A$.

vii) If B is formed from A by interchanging two rows or two columns of A, then $\det B = -\det A$.

viii) If B is formed from A by adding a multiple of a (row, column) of A to another (row, column) of A, then $\det B = \det A$.

Statements *vi*)-*viii*) correspond, respectively, to multiplying the matrix A on the left or right by matrices of the form

$$I_n + (\alpha - 1)E_{i,i} = \begin{bmatrix} I_{i-1} & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & I_{n-i} \end{bmatrix}, \tag{2.7.10}$$

$$I_n + E_{i,j} + E_{j,i} - E_{i,i} - E_{j,j} = \begin{bmatrix} I_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{n-j} \end{bmatrix}, \tag{2.7.11}$$

where $i \neq j$, and

$$I_n + \beta E_{i,j} = \begin{bmatrix} I_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \beta & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_{n-j} \end{bmatrix}, \tag{2.7.12}$$

where $\beta \in \mathbb{F}$ and $i \neq j$. The matrices shown in (2.7.11) and (2.7.12) illustrate the case $i < j$. Since $I + (\alpha-1)E_{i,i} = I + (\alpha-1)e_ie_i^{\mathrm{T}}$, $I + E_{i,j} + E_{j,i} - E_{i,i} - E_{j,j} = I - (e_i - e_j)(e_i - e_j)^{\mathrm{T}}$, and $I + \beta E_{i,j} = I + \beta e_ie_j^{\mathrm{T}}$, it follows that all of these matrices are of the form $I - xy^{\mathrm{T}}$. If $\alpha \neq 0$ and $i \neq j$, then these are elementary matrices (see Definition 3.1.2).

Proposition 2.7.3. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\det AB = \det BA = (\det A)(\det B). \tag{2.7.13}$$

Proof. First note the identity

$$\begin{bmatrix} A & 0 \\ I & B \end{bmatrix} = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} -AB & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

The first and third matrices on the right-hand side of this identity add multiples of rows and columns of $\left[\begin{smallmatrix} -AB & 0 \\ 0 & I \end{smallmatrix}\right]$ to other rows and columns of $\left[\begin{smallmatrix} -AB & 0 \\ 0 & I \end{smallmatrix}\right]$. As already noted, these operations do not affect the determinant of $\left[\begin{smallmatrix} -AB & 0 \\ 0 & I \end{smallmatrix}\right]$. In addition, the fourth matrix on the right-hand side of this identity interchanges n pairs of columns of $\left[\begin{smallmatrix} 0 & A \\ B & I \end{smallmatrix}\right]$. Using (2.7.5), (2.7.6) and the fact that every interchange of a pair of columns of $\left[\begin{smallmatrix} 0 & A \\ B & I \end{smallmatrix}\right]$ entails a factor of -1, it thus follows that $(\det A)(\det B) = \det \left[\begin{smallmatrix} A & 0 \\ I & B \end{smallmatrix}\right] = (-1)^n \det \left[\begin{smallmatrix} -AB & 0 \\ 0 & I \end{smallmatrix}\right] = (-1)^n \det(-AB) = \det AB$. □

Corollary 2.7.4. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, $\det A \neq 0$ and

$$\det A^{-1} = (\det A)^{-1}. \tag{2.7.14}$$

Proof. Since $AA^{-1} = I_n$, it follows that $\det AA^{-1} = (\det A)\big(\det A^{-1}\big) = 1$. Hence, $\det A \neq 0$. In addition, $\det A^{-1} = 1/\det A$. □

Let $A \in \mathbb{F}^{n\times m}$. Then, a *submatrix* of A is formed by deleting rows and columns of A. By convention, A is a submatrix of A. If A is a partitioned matrix, then every block of A is a submatrix of A. A block is thus a submatrix whose entries are entries of adjacent rows and adjacent columns. The determinant of a square submatrix of A is a *subdeterminant* of A. By convention, the determinant of A is a subdeterminant of A.

Let $A \in \mathbb{F}^{n\times n}$. If like-numbered rows and columns of A are deleted, then the resulting square submatrix of A is a *principal submatrix* of A. If, in particular, rows and columns $j+1,\ldots,n$ of A are deleted, then the resulting $j\times j$ submatrix of A is a *leading principal submatrix* of A. Every diagonally located block is a principal submatrix. Finally, the determinant of a $j\times j$ (principal, leading principal) submatrix of A is a $j\times j$ (*principal, leading principal*) *subdeterminant* of A.

Let $A \in \mathbb{F}^{n\times n}$. Then, the *cofactor* of $A_{(i,j)}$, denoted by $A_{[i,j]}$, is the $(n-1)\times(n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. The following result provides a cofactor expansion of $\det A$.

Proposition 2.7.5. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $i = 1,\ldots,n$,

$$\det A = \sum_{k=1}^{n}(-1)^{i+k}A_{(i,k)}\det A_{[i,k]}. \tag{2.7.15}$$

Furthermore, for all $i,j = 1,\ldots,n$ such that $j \neq i$,

$$0 = \sum_{k=1}^{n}(-1)^{i+k}A_{(j,k)}\det A_{[i,k]}. \tag{2.7.16}$$

Proof. Identity (2.7.15) is an equivalent recursive form of the definition $\det A$, while the right-hand side of (2.7.16) is equal to $\det B$, where B is obtained from A by replacing $\mathrm{row}_i(A)$ by $\mathrm{row}_j(A)$. As already noted, $\det B = 0$. □

Let $A \in \mathbb{F}^{n\times n}$. To simplify (2.7.15) and (2.7.16) it is useful to define the *adjugate* of A, denoted by $A^{\mathrm{A}} \in \mathbb{F}^{n\times n}$, where, for all $i,j = 1,\ldots,n$,

$$\big(A^{\mathrm{A}}\big)_{(i,j)} \triangleq (-1)^{i+j}\det A_{[j,i]}. \tag{2.7.17}$$

Then, (2.7.15) and (2.7.16) imply that, for all $i = 1,\ldots,n$,

$$\big(AA^{\mathrm{A}}\big)_{(i,i)} = \big(A^{\mathrm{A}}A\big)_{(i,i)} = \det A \tag{2.7.18}$$

and, for all $i,j = 1,\ldots,n$ such that $j \neq i$,

$$\left(AA^{\mathrm{A}}\right)_{(i,j)} = \left(A^{\mathrm{A}}A\right)_{(i,j)} = 0. \tag{2.7.19}$$

Thus,

$$AA^{\mathrm{A}} = A^{\mathrm{A}}A = (\det A)I. \tag{2.7.20}$$

Consequently, if $\det A \neq 0$, then

$$A^{-1} = (\det A)^{-1}A^{\mathrm{A}}. \tag{2.7.21}$$

The following result provides the converse of Corollary 2.7.4 by using (2.7.21) to explicitly construct A^{-1} in terms of $(n-1)\times(n-1)$ subdeterminants of A.

Corollary 2.7.6. Let $A \in \mathbb{F}^{n\times n}$. Then, A is nonsingular if and only if $\det A \neq 0$. In this case, for all $i,j = 1,\ldots,n$, the (i,j) entry of A^{-1} is given by

$$\left(A^{-1}\right)_{(i,j)} = (-1)^{i+j}\frac{\det A_{[j,i]}}{\det A}. \tag{2.7.22}$$

Finally, the following result uses the nonsingularity of submatrices to characterize the rank of a matrix.

Proposition 2.7.7. Let $A \in \mathbb{F}^{n\times m}$. Then, $\operatorname{rank} A$ is the largest order of all nonsingular submatrices of A.

2.8 Properties of Partitioned Matrices

Partitioned matrices were used to state or prove several results in this chapter including Proposition 2.5.8, Theorem 2.6.3, Proposition 2.7.1, and Proposition 2.7.3. In this section we give several useful identities involving partitioned matrices.

Proposition 2.8.1. Let $A_{ij} \in \mathbb{F}^{n_i\times m_j}$ for all $i = 1,\ldots,k$ and $j = 1,\ldots,l$. Then,

$$\begin{bmatrix} A_{11} & \cdots & A_{1l} \\ \vdots & \cdot\vdots\cdot & \vdots \\ A_{k1} & \cdots & A_{kl} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} A_{11}^{\mathrm{T}} & \cdots & A_{k1}^{\mathrm{T}} \\ \vdots & \cdot\vdots\cdot & \vdots \\ A_{1l}^{\mathrm{T}} & \cdots & A_{kl}^{\mathrm{T}} \end{bmatrix} \tag{2.8.1}$$

and

$$\begin{bmatrix} A_{11} & \cdots & A_{1l} \\ \vdots & \cdot\vdots\cdot & \vdots \\ A_{k1} & \cdots & A_{kl} \end{bmatrix}^{*} = \begin{bmatrix} A_{11}^{*} & \cdots & A_{k1}^{*} \\ \vdots & \cdot\vdots\cdot & \vdots \\ A_{1l}^{*} & \cdots & A_{kl}^{*} \end{bmatrix}. \tag{2.8.2}$$

If, in addition, $k = l$ and $n_i = m_i$ for all $i = 1, \ldots, m$, then

$$\operatorname{tr} \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \vdots & \vdots \\ A_{k1} & \cdots & A_{kk} \end{bmatrix} = \sum_{i=1}^{k} \operatorname{tr} A_{ii} \tag{2.8.3}$$

and

$$\det \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ 0 & A_{22} & \cdots & A_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{kk} \end{bmatrix} = \prod_{i=1}^{k} \det A_{ii}. \tag{2.8.4}$$

Lemma 2.8.2. Let $B \in \mathbb{F}^{n \times m}$ and $C \in \mathbb{F}^{m \times n}$. Then,

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -B \\ 0 & I \end{bmatrix} \tag{2.8.5}$$

and

$$\begin{bmatrix} I & 0 \\ C & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -C & I \end{bmatrix}. \tag{2.8.6}$$

Let $A \in \mathbb{F}^{n \times n}$ and $D \in \mathbb{F}^{m \times m}$ be nonsingular. Then,

$$\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}. \tag{2.8.7}$$

Proposition 2.8.3. Let $A \in \mathbb{F}^{n \times n}$, $B \in \mathbb{F}^{n \times m}$, $C \in \mathbb{F}^{l \times n}$, and $D \in \mathbb{F}^{l \times m}$, and assume that A is nonsingular. Then,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \tag{2.8.8}$$

and

$$\operatorname{rank} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = n + \operatorname{rank}\big(D - CA^{-1}B\big). \tag{2.8.9}$$

If, furthermore, $l = m$, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = (\det A) \det\big(D - CA^{-1}B\big). \tag{2.8.10}$$

Proposition 2.8.4. Let $A \in \mathbb{F}^{n \times m}$, $B \in \mathbb{F}^{n \times l}$, $C \in \mathbb{F}^{l \times m}$, and $D \in \mathbb{F}^{l \times l}$, and assume that D is nonsingular. Then,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \tag{2.8.11}$$

and

$$\operatorname{rank} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = l + \operatorname{rank}\big(A - BD^{-1}C\big). \tag{2.8.12}$$

If, furthermore, $n = m$, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = (\det D) \det\bigl(A - BD^{-1}C\bigr). \tag{2.8.13}$$

Corollary 2.8.5. Let $A \in \mathbb{F}^{n \times m}$ and $B \in \mathbb{F}^{m \times n}$. Then,

$$\det \begin{bmatrix} I_n & A \\ -B & I_m \end{bmatrix} = \det(I_n + AB) = \det(I_m + BA). \tag{2.8.14}$$

Hence, $I_n + AB$ is nonsingular if and only if $I_m + BA$ is nonsingular.

Lemma 2.8.6. Let $A \in \mathbb{F}^{n \times n}$, $B \in \mathbb{F}^{n \times m}$, $C \in \mathbb{F}^{m \times n}$, and $D \in \mathbb{F}^{m \times m}$. If A and D are nonsingular, then

$$(\det A) \det(D - CA^{-1}B) = (\det D) \det\bigl(A - BD^{-1}C\bigr), \tag{2.8.15}$$

and thus $D - CA^{-1}B$ is nonsingular if and only if $A - BD^{-1}C$ is nonsingular.

Proposition 2.8.7. Let $A \in \mathbb{F}^{n \times n}$, $B \in \mathbb{F}^{n \times m}$, $C \in \mathbb{F}^{m \times n}$, and $D \in \mathbb{F}^{m \times m}$. If A and $D - CA^{-1}B$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\bigl(D - CA^{-1}B\bigr)^{-1}CA^{-1} & -A^{-1}B\bigl(D - CA^{-1}B\bigr)^{-1} \\ -\bigl(D - CA^{-1}B\bigr)^{-1}CA^{-1} & \bigl(D - CA^{-1}B\bigr)^{-1} \end{bmatrix}. \tag{2.8.16}$$

If D and $A - BD^{-1}C$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \bigl(A - BD^{-1}C\bigr)^{-1} & -\bigl(A - BD^{-1}C\bigr)^{-1}BD^{-1} \\ -D^{-1}C\bigl(A - BD^{-1}C\bigr)^{-1} & D^{-1} + D^{-1}C\bigl(A - BD^{-1}C\bigr)^{-1}BD^{-1} \end{bmatrix}. \tag{2.8.17}$$

If A, D, and $D - CA^{-1}B$ are nonsingular, then $A - BD^{-1}C$ is nonsingular and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \bigl(A - BD^{-1}C\bigr)^{-1} & -\bigl(A - BD^{-1}C\bigr)^{-1}BD^{-1} \\ -\bigl(D - CA^{-1}B\bigr)^{-1}CA^{-1} & \bigl(D - CA^{-1}B\bigr)^{-1} \end{bmatrix}. \tag{2.8.18}$$

The following result is the *matrix inversion lemma.*

Corollary 2.8.8. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, $C \in \mathbb{F}^{m\times n}$, and $D \in \mathbb{F}^{m\times m}$. If A, $D - CA^{-1}B$, and D are nonsingular, then, $A - BD^{-1}C$ is nonsingular and

$$\big(A - BD^{-1}C\big)^{-1} = A^{-1} + A^{-1}B\big(D - CA^{-1}B\big)^{-1}CA^{-1}. \tag{2.8.19}$$

If A and $I - CA^{-1}B$ are nonsingular, then $A - BC$ is nonsingular and

$$(A - BC)^{-1} = A^{-1} + A^{-1}B\big(I - CA^{-1}B\big)^{-1}CA^{-1}. \tag{2.8.20}$$

If $D - CB$, and D are nonsingular, then, $I_n - BD^{-1}C$ is nonsingular and

$$\big(I_n - BD^{-1}C\big)^{-1} = I_n + B(D - CB)^{-1}C. \tag{2.8.21}$$

If $I - CB$ is nonsingular, then $I - BC$ is nonsingular and

$$(I - BC)^{-1} = I + B(I - CB)^{-1}C. \tag{2.8.22}$$

Corollary 2.8.9. Let $A, B, C, D \in \mathbb{F}^{n\times n}$. If A, B, $C - DB^{-1}A$, and $D - CA^{-1}B$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - \big(C - DB^{-1}A\big)^{-1}CA^{-1} & \big(C - DB^{-1}A\big)^{-1} \\ -\big(D - CA^{-1}B\big)^{-1}CA^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{bmatrix}. \tag{2.8.23}$$

If A, C, $B - AC^{-1}D$, and $D - CA^{-1}B$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - A^{-1}B\big(B - AC^{-1}D\big)^{-1} & -A^{-1}B\big(D - CA^{-1}B\big)^{-1} \\ \big(B - AC^{-1}D\big)^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{bmatrix}. \tag{2.8.24}$$

If A, B, C, $B - AC^{-1}D$, and $D - CA^{-1}B$ are nonsingular, then $C - DB^{-1}A$ is nonsingular and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - A^{-1}B\big(B - AC^{-1}D\big)^{-1} & \big(C - DB^{-1}A\big)^{-1} \\ \big(B - AC^{-1}D\big)^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{bmatrix}. \tag{2.8.25}$$

If B, D, $A - BD^{-1}C$, and $C - DB^{-1}A$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \big(A - BD^{-1}C\big)^{-1} & \big(C - DB^{-1}A\big)^{-1} \\ -D^{-1}C\big(A - BD^{-1}C\big)^{-1} & D^{-1} - D^{-1}C\big(C - DB^{-1}A\big)^{-1} \end{bmatrix}. \tag{2.8.26}$$

If C, D, $A-BD^{-1}C$, and $B-AC^{-1}D$ are nonsingular, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A-BD^{-1}C\right)^{-1} & -\left(A-BD^{-1}C\right)^{-1}BD^{-1} \\ \left(B-AC^{-1}D\right)^{-1} & D^{-1}-\left(B-AC^{-1}D\right)^{-1}BD^{-1} \end{bmatrix}. \tag{2.8.27}$$

If B, C, D, $A-BD^{-1}C$, and $C-DB^{-1}A$ are nonsingular, then $B-AC^{-1}D$ is nonsingular and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A-BD^{-1}C\right)^{-1} & \left(C-DB^{-1}A\right) \\ \left(B-AC^{-1}D\right)^{-1} & D^{-1}-D^{-1}C\left(C-DB^{-1}A\right)^{-1} \end{bmatrix}. \tag{2.8.28}$$

Finally, if A, B, C, D, $A-BD^{-1}C$, and $B-AC^{-1}D$, are nonsingular, then $C-DB^{-1}A$ and $D-CA^{-1}B$ are nonsingular and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A-BD^{-1}C\right)^{-1} & \left(C-DB^{-1}A\right)^{-1} \\ \left(B-AC^{-1}D\right)^{-1} & \left(D-CA^{-1}B\right)^{-1} \end{bmatrix}. \tag{2.8.29}$$

Corollary 2.8.10. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and $I-A^{-1}B$ are nonsingular. Then, $A-B$ is nonsingular and

$$(A-B)^{-1} = A^{-1} + A^{-1}B\left(I-A^{-1}B\right)^{-1}A^{-1}. \tag{2.8.30}$$

If, in addition, B is nonsingular, then

$$(A-B)^{-1} = A^{-1} + A^{-1}\left(B^{-1}-A^{-1}\right)^{-1}A^{-1}. \tag{2.8.31}$$

2.9 Facts on Cones, Convex Hulls, and Subspaces

Fact 2.9.1. Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then, the following statements hold:

i) $\operatorname{coco}\mathcal{S} = \operatorname{co}\operatorname{cone}\mathcal{S} = \operatorname{cone}\operatorname{co}\mathcal{S}$.

ii) $\mathcal{S}^{\perp\perp} = \operatorname{span}\mathcal{S} = \operatorname{coco}(\mathcal{S}\cup -\mathcal{S})$.

iii) $\mathcal{S} \subseteq \operatorname{co}\mathcal{S} \subseteq (\operatorname{aff}\mathcal{S}\cap\operatorname{coco}\mathcal{S}) \subseteq \left\{\begin{matrix}\operatorname{aff}\mathcal{S}\\ \operatorname{coco}\mathcal{S}\end{matrix}\right\} \subseteq \operatorname{span}\mathcal{S}$.

iv) $\mathcal{S} \subseteq (\operatorname{co}\mathcal{S}\cap\operatorname{cone}\mathcal{S}) \subseteq \left\{\begin{matrix}\operatorname{co}\mathcal{S}\\ \operatorname{cone}\mathcal{S}\end{matrix}\right\} \subseteq \operatorname{coco}\mathcal{S} \subseteq \operatorname{span}\mathcal{S}$.

v) $\operatorname{dcone}\operatorname{dcone}\mathcal{S} = \operatorname{coco}\mathcal{S}$.

(Proof: See [79, p. 52] for the proof of *v*). Note that "pointed" in [79] means one-sided.)

Fact 2.9.2. Let $\mathcal{S} \subseteq \mathbb{F}^m$ and $A \in \mathbb{F}^{n\times m}$. If $\mathcal{S}$ is convex, then $A\mathcal{S}$ is convex. Conversely, if A is left invertible and $A\mathcal{S}$ is convex, then $\mathcal{S}$ is convex.

Fact 2.9.3. Let $\mathcal{S} \subset \mathbb{F}^n$. Then, $\mathcal{S}$ is an affine hyperplane if and only if there exist a nonzero vector $x \in \mathbb{F}^n$ and $\alpha \in \mathbb{R}$ such that $\mathcal{S} = \{x\colon\ \operatorname{Re} x^*y = \alpha\}$. Furthermore, $\mathcal{S}$ is an affine closed half space if and only if there exist a nonzero vector $x \in \mathbb{F}^n$ and $\alpha \in \mathbb{R}$ such that $\mathcal{S} = \{x \in \mathbb{F}^n\colon\ \operatorname{Re} x^*y \le \alpha\}$. Finally, $\mathcal{S}$ is an affine open half space if and only if there exist a nonzero vector $x \in \mathbb{F}^n$ and $\alpha \in \mathbb{R}$ such that $\mathcal{S} = \{x \in \mathbb{F}^n\colon\ \operatorname{Re} x^*y \le \alpha\}$. (Proof: Let $z \in \mathbb{F}^n$ satisfy $z^*y = \alpha$. Then, $\{x\colon\ x^*y = \alpha\} = \{y\}^\perp + z$.)

Fact 2.9.4. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be (cones, convex sets, convex cones, subspaces). Then, so are $\mathcal{S}_1 \cap \mathcal{S}_2$ and $\mathcal{S}_1 + \mathcal{S}_2$.

Fact 2.9.5. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be pointed convex cones. Then,

$$\operatorname{co}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathcal{S}_1 + \mathcal{S}_2.$$

Fact 2.9.6. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1 \cup \mathcal{S}_2$ is a subspace if and only if either $\mathcal{S}_1 \subseteq \mathcal{S}_2$ or $\mathcal{S}_2 \subseteq \mathcal{S}_1$.

Fact 2.9.7. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then,

$$\operatorname{span}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathcal{S}_1 + \mathcal{S}_2.$$

Fact 2.9.8. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1 \subseteq \mathcal{S}_2$ if and only if $\mathcal{S}_2^\perp \subseteq \mathcal{S}_1^\perp$. Furthermore, $\mathcal{S}_1 \subset \mathcal{S}_2$ if and only if $\mathcal{S}_2^\perp \subset \mathcal{S}_1^\perp$. (Remark: $\mathcal{S}_1 \subset \mathcal{S}_2$ denotes proper inclusion.)

Fact 2.9.9. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$. Then,

$$\mathcal{S}_1^\perp \cap \mathcal{S}_2^\perp \subseteq (\mathcal{S}_1 + \mathcal{S}_2)^\perp.$$

(Problem: Determine necessary and sufficient conditions under which equality holds.)

Fact 2.9.10. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then,

$$(\mathcal{S}_1 \cap \mathcal{S}_2)^\perp = \mathcal{S}_1^\perp + \mathcal{S}_2^\perp$$

and

$$(\mathcal{S}_1 + \mathcal{S}_2)^\perp = \mathcal{S}_1^\perp \cap \mathcal{S}_2^\perp.$$

Fact 2.9.11. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then, $\mathcal{S}_1, \mathcal{S}_2$ are complementary if and only if $\mathcal{S}_1^\perp, \mathcal{S}_2^\perp$ are complementary. (Remark: See Fact 3.5.15.)

Fact 2.9.12. Let $\mathcal{S}_1, \ldots, \mathcal{S}_k \subseteq \mathbb{F}^n$ be subspaces having the same dimension. Then, there exists a subspace $\hat{\mathcal{S}} \subseteq \mathbb{F}^n$ such that, for all $i = 1, \ldots, k$, $\hat{\mathcal{S}}$

and $\hat{\mathcal{S}}_i$ are complementary. (Proof: See [261, pp. 78, 79, 259, 260].)

Fact 2.9.13. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be subspaces. Then,

$$\begin{aligned}\dim(\mathcal{S}_1 \cap \mathcal{S}_2) &\le \min\{\dim \mathcal{S}_1, \dim \mathcal{S}_2\} \\ &\le \left\{\begin{matrix}\dim \mathcal{S}_1 \\ \dim \mathcal{S}_2\end{matrix}\right\} \\ &\le \max\{\dim \mathcal{S}_1, \dim \mathcal{S}_2\} \\ &\le \dim(\mathcal{S}_1 + \mathcal{S}_2) \\ &\le \min\{\dim \mathcal{S}_1 + \dim \mathcal{S}_2, n\}.\end{aligned}$$

2.10 Facts on Range, Null Space, Rank, and Defect

Fact 2.10.1. Let $n, m, k \in \mathbb{P}$. Then, $\operatorname{rank} 1_{n\times m} = 1$ and $1_{n\times n}^k = n^{k-1}1_{n\times n}$.

Fact 2.10.2. Let $A \in \mathbb{F}^{n\times n}$, $k \in \mathbb{P}$, and $l \in \mathbb{N}$. Then, the following identities hold:

i) $\mathcal{R}\Big[(AA^*)^k\Big] = \mathcal{R}\Big[(AA^*)^l A\Big]$.

ii) $\mathcal{N}\Big[(A^*A)^k\Big] = \mathcal{N}\Big[A(A^*A)^l\Big]$.

iii) $\operatorname{rank}\ (AA^*)^k = \operatorname{rank}\ (AA^*)^l A$.

iv) $\operatorname{def}\ (A^*A)^k = \operatorname{def}\ A(A^*A)^l$.

Fact 2.10.3. Let $A, B \in \mathbb{F}^{n\times n}$, and assume there exists $\alpha \in \mathbb{F}$ such that $\alpha A + B$ is nonsingular. Then, $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$. (Remark: The converse is not true. Let $A \triangleq \left[\begin{smallmatrix}1 & 0\\ 2 & 0\end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix}0 & 1\\ 0 & 2\end{smallmatrix}\right]$.)

Fact 2.10.4. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(A) \cap \mathcal{N}(A+B) = \mathcal{N}(A+B) \cap \mathcal{N}(B).$$

Fact 2.10.5. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$|\operatorname{rank} A - \operatorname{rank} B| \le \operatorname{rank}(A+B) \le \operatorname{rank} A + \operatorname{rank} B.$$

If, in addition, $\operatorname{rank} B \le k$, then

$$(\operatorname{rank} A) - k \le \operatorname{rank}(A+B) \le (\operatorname{rank} A) + k.$$

Fact 2.10.6. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $A^*B = 0$ and $BA^* = 0$. Then,

$$\operatorname{rank}(A+B) = \operatorname{rank} A + \operatorname{rank} B.$$

(Remark: This result is due to Hestenes. See [148].) (Proof: Use Fact 2.10.15 and Proposition 6.1.6.)

Fact 2.10.7. Let $A \triangleq \left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right]$. Then, $\operatorname{rank} AB = 1$ and $\operatorname{rank} BA = 0$.

Fact 2.10.8. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, the following statements hold:

i) $\operatorname{rank} AB + \operatorname{def} A = \dim[\mathcal{N}(A) + \mathcal{R}(B)]$.

ii) $\operatorname{rank} AB + \dim[\mathcal{N}(A) \cap \mathcal{R}(B)] = \operatorname{rank} B$.

iii) $\operatorname{def} AB + \operatorname{rank} A + \dim[\mathcal{N}(A) + \mathcal{R}(B)] = l + m$.

iv) $\operatorname{def} AB = \operatorname{def} B + \dim[\mathcal{N}(A) \cap \mathcal{R}(B)]$.

(Remark: $\operatorname{rank} B - \operatorname{rank} AB = \dim[\mathcal{N}(A) \cap \mathcal{R}(B)] \le \dim \mathcal{N}(A) = m - \operatorname{rank} A$ yields (2.5.17).)

Fact 2.10.9. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\max\{\operatorname{def} A + l - m, \operatorname{def} B\} \le \operatorname{def} AB \le \operatorname{def} A + \operatorname{def} B.$$

If, in addition, $m = l$, then

$$\max\{\operatorname{def} A, \operatorname{def} B\} \le \operatorname{def} AB.$$

(Remark: The first inequality is *Sylvester's law of nullity.*)

Fact 2.10.10. Let $\mathcal{S} \subseteq \mathbb{F}^m$, and let $A \in \mathbb{F}^{n\times m}$. Then, the following statements hold:

i) $\operatorname{rank} A + \dim \mathcal{S} - m \le \dim A\mathcal{S} \le \min\{\operatorname{rank} A, \dim \mathcal{S}\}$.

ii) $\dim(A\mathcal{S}) + \dim(\mathcal{N}(A) \cap \mathcal{S}) = \dim \mathcal{S}$.

iii) If A is left invertible, then $\dim A\mathcal{S} = \dim \mathcal{S}$.

(Proof: For *ii)*, see [484, p. 413].)

Fact 2.10.11. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{1\times m}$. Then, $\mathcal{N}(A) \subseteq \mathcal{N}(B)$ if and only if there exists $\lambda \in \mathbb{F}^n$ such that $B = \lambda^*A$.

Fact 2.10.12. Let $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$. Then, there exists $x \in \mathbb{F}^n$ satisfying $Ax = b$ if and only if $b^*\lambda = 0$ for all $\lambda \in \mathcal{N}(A^*)$. (Proof: Assume that $A^*\lambda = 0$ implies that $b^*\lambda = 0$. Then, $\mathcal{N}(A^*) \subseteq \mathcal{R}(b^*)$. Hence, $b \in \mathcal{R}(b) \subseteq \mathcal{R}(A)$.)

Fact 2.10.13. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times m}$. Then, $\mathcal{N}(B) \subseteq \mathcal{N}(A)$ if and only if there exists $C \in \mathbb{F}^{n\times l}$ such that $A = CB$. Now, let $A \in \mathbb{F}^{n\times m}$

and $B \in \mathbb{F}^{n\times l}$. Then, $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ if and only if there exists $C \in \mathbb{F}^{l\times m}$ such that $A = BC$.

Fact 2.10.14. Let $A, B \in \mathbb{F}^{n\times m}$, and let $C \in \mathbb{F}^{m\times l}$ be right invertible. If $\mathcal{R}(A) \subseteq \mathcal{R}(B)$, then $\mathcal{R}(AC) \subseteq \mathcal{R}(BC)$. Furthermore, $\mathcal{R}(A) = \mathcal{R}(B)$ if and only if $\mathcal{R}(AC) = \mathcal{R}(BC)$.

Fact 2.10.15. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $A^*B = 0$ and $BA^* = 0$. Then,

$$\operatorname{rank}(A + B) = \operatorname{rank} A + \operatorname{rank} B$$

if and only if there exists $C \in \mathbb{F}^{m\times n}$ such that $ACA = A$, $CB = 0$, and $BC = 0$. (Proof: See [148].)

Fact 2.10.16. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, $\operatorname{rank} AB = \operatorname{rank} A$ if and only if $\mathcal{R}(AB) = \mathcal{R}(A)$. (Proof: If $\mathcal{R}(AB) \subset \mathcal{R}(A)$ (note proper inclusion), then $\operatorname{rank} AB < \operatorname{rank} A$.)

Fact 2.10.17. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $C \in \mathbb{F}^{l\times k}$. If $\operatorname{rank} AB = \operatorname{rank} B$, then $\operatorname{rank} ABC = \operatorname{rank} BC$. (Proof: $\operatorname{rank} B^{\mathrm{T}}A^{\mathrm{T}} = \operatorname{rank} B^{\mathrm{T}}$ implies that $\mathcal{R}\big(C^{\mathrm{T}}B^{\mathrm{T}}A^{\mathrm{T}}\big) = \mathcal{R}\big(C^{\mathrm{T}}B^{\mathrm{T}}\big)$.)

Fact 2.10.18. Let $A \in \mathbb{F}^{n\times m}$. Then, $\operatorname{rank} A = 1$ if and only if there exist $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$ such that $x \neq 0$, $y \neq 0$, and $A = xy^{\mathrm{T}}$. In this case, $\operatorname{tr} A = y^{\mathrm{T}}x$.

Fact 2.10.19. Let $x, y \in \mathbb{F}^n$. Then,

$$\operatorname{rank}\big(xy^{\mathrm{T}} + yx^{\mathrm{T}}\big) \le 2.$$

Furthermore, $\operatorname{rank}\big(xy^{\mathrm{T}} + yx^{\mathrm{T}}\big) = 1$ if and only if there exists $\alpha \in \mathbb{F}$ such that $x = \alpha y \neq 0$.

Fact 2.10.20. Let $A \in \mathbb{F}^{n\times m}$, $x \in \mathbb{F}^n$, and $y \in \mathbb{F}^m$. Then,

$$(\operatorname{rank} A) - 1 \le \operatorname{rank}\big(A + xy^{\mathrm{T}}\big) \le (\operatorname{rank} A) + 1.$$

In addition, the following statements hold:

i) $\operatorname{rank}\big(A + xy^{\mathrm{T}}\big) = (\operatorname{rank} A) - 1$ if and only if there exist $\hat{x} \in \mathbb{F}^m$ and $\hat{y} \in \mathbb{F}^n$ such that $\hat{y}^{\mathrm{T}}A\hat{x} \neq 0$, $x = -\big(\hat{y}^{\mathrm{T}}A\hat{x}\big)^{-1}A\hat{x}$, and $y = A^{\mathrm{T}}\hat{y}$.

ii) If there exists $\hat{x} \in \mathbb{F}^m$ such that $x = A\hat{x}$ and $\hat{x}^{\mathrm{T}}y \neq -1$, then $\operatorname{rank}\big(A + xy^{\mathrm{T}}\big) = \operatorname{rank} A$.

iii) If $xy^{\mathrm{T}} \neq 0$, $A^*x = 0$, and $A\overline{y} = 0$, then $\operatorname{rank}\big(A + xy^{\mathrm{T}}\big) = (\operatorname{rank} A) + 1$.

(Proof: To prove *ii)*, note that $A + xy^{\mathrm{T}} = A\big(I + xy^{\mathrm{T}}\big)$ and $I + xy^{\mathrm{T}}$ is nonsingular. To prove *iii*) use Fact 2.10.21. See [297, p. 33] and [144].)

Fact 2.10.21. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{l\times n}$, $D \in \mathbb{F}^{l\times l}$, and assume that D is nonsingular. Then,

$$\operatorname{rank}\bigl(A - BD^{-1}C\bigr) = \operatorname{rank} A - \operatorname{rank} BD^{-1}C$$

if and only if there exist $X \in \mathbb{F}^{m\times l}$ and $Y \in \mathbb{F}^{l\times n}$ such that $B = AX$, $C = YA$, and $D = YAX$. (Proof: See [144].)

Fact 2.10.22. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\mathcal{R}\bigl(\begin{bmatrix} A & B \end{bmatrix}\bigr) = \mathcal{R}(A) + \mathcal{R}(B).$$

Fact 2.10.23. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\mathcal{R}(A) = \mathcal{R}(B)$$

if and only if

$$\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix}.$$

Fact 2.10.24. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\operatorname{rank} A + \operatorname{rank} B = \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} + \dim[\mathcal{R}(A) \cap \mathcal{R}(B)]$$

and

$$\operatorname{def} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{def} A + \operatorname{def} B + \dim[\mathcal{R}(A) \cap \mathcal{R}(B)].$$

Hence,

$$\max\{\operatorname{rank} A, \operatorname{rank} B\} \le \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} \le \operatorname{rank} A + \operatorname{rank} B$$

and

$$\operatorname{def} A + \operatorname{def} B \le \operatorname{def} \begin{bmatrix} A & B \end{bmatrix} \le \min\{l + \operatorname{def} A, m + \operatorname{def} B\}.$$

If, in addition, $A^*B = 0$, then

$$\operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank} A + \operatorname{rank} B$$

and

$$\operatorname{def} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{def} A + \operatorname{def} B.$$

(Proof: Use Fact 2.9.13. Assume $A^*B = 0$. Then,

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} &= \operatorname{rank} \begin{bmatrix} A^* \\ B^* \end{bmatrix} \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} A^*A & 0 \\ 0 & B^*B \end{bmatrix} \\ &= \operatorname{rank} A^*A + \operatorname{rank} B^*B = \operatorname{rank} A + \operatorname{rank} B.) \end{aligned}$$

Fact 2.10.25. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times m}$. Then,

$$\max\{\operatorname{rank} A, \operatorname{rank} B\} \le \operatorname{rank} \begin{bmatrix} A \\ B \end{bmatrix} \le \operatorname{rank} A + \operatorname{rank} B$$

and

$$\operatorname{def} A - \operatorname{rank} B \le \operatorname{def} \begin{bmatrix} A \\ B \end{bmatrix} \le \min\{\operatorname{def} A, \operatorname{def} B\}.$$

If, in addition, $AB^* = 0$, then

$$\operatorname{rank} \begin{bmatrix} A \\ B \end{bmatrix} = \operatorname{rank} A + \operatorname{rank} B$$

and

$$\operatorname{def} \begin{bmatrix} A \\ B \end{bmatrix} = \operatorname{def} A - \operatorname{rank} B.$$

(Proof: Use Fact 2.10.24 and Fact 2.9.13.)

Fact 2.10.26. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\left\{ \begin{array}{c} \max\{\operatorname{rank} A, \operatorname{rank} B\} \\ \\ \operatorname{rank}(A+B) \end{array} \right\} \le \left\{ \begin{array}{c} \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} \\ \\ \operatorname{rank} \begin{bmatrix} A \\ B \end{bmatrix} \end{array} \right\} \le \operatorname{rank} A + \operatorname{rank} B$$

and

$$\operatorname{def} A + \operatorname{def} B \le \left\{ \begin{array}{c} \operatorname{def} \begin{bmatrix} A & B \end{bmatrix} \\ \\ \operatorname{def} \begin{bmatrix} A \\ B \end{bmatrix} + m \end{array} \right\} \le \left\{ \begin{array}{c} m + \min\{\operatorname{def} A, \operatorname{def} B\} \\ \\ \operatorname{def}(A+B) + m \end{array} \right\}.$$

(Proof: $\operatorname{rank}(A+B) = \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} \left[\begin{smallmatrix} I \\ I \end{smallmatrix}\right] \le \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix}$, and $\operatorname{rank}(A+B) = \operatorname{rank} \begin{bmatrix} I & I \end{bmatrix} \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right] \le \operatorname{rank} \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]$.)

Fact 2.10.27. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{l\times k}$, and $C \in \mathbb{F}^{l\times m}$. Then,

$$\operatorname{rank} A + \operatorname{rank} B = \operatorname{rank} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \le \operatorname{rank} \begin{bmatrix} A & 0 \\ C & B \end{bmatrix}$$

and

$$\operatorname{rank} A + \operatorname{rank} B = \operatorname{rank} \begin{bmatrix} 0 & A \\ B & 0 \end{bmatrix} \le \operatorname{rank} \begin{bmatrix} 0 & A \\ B & C \end{bmatrix}.$$

Fact 2.10.28. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $C \in \mathbb{F}^{l\times k}$. Then,

$$\operatorname{rank} AB + \operatorname{rank} BC \le \operatorname{rank} \begin{bmatrix} 0 & AB \\ BC & B \end{bmatrix} = \operatorname{rank} B + \operatorname{rank} ABC.$$

Consequently,

$$\operatorname{rank} AB + \operatorname{rank} BC - \operatorname{rank} B \le \operatorname{rank} ABC.$$

(Remark: This result is *Frobenius' inequality.*) (Proof: Use Fact 2.10.27 and $\left[\begin{smallmatrix} 0 & AB \\ BC & B \end{smallmatrix}\right] = \left[\begin{smallmatrix} I & A \\ 0 & I \end{smallmatrix}\right] \left[\begin{smallmatrix} -ABC & 0 \\ 0 & B \end{smallmatrix}\right] \left[\begin{smallmatrix} I & 0 \\ C & I \end{smallmatrix}\right]$.) (Remark: See [398] for the case of equality.)

Fact 2.10.29. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{rank}\begin{bmatrix} A & B \end{bmatrix} + \operatorname{rank}\begin{bmatrix} A \\ B \end{bmatrix} \le \operatorname{rank}\begin{bmatrix} 0 & A & B \\ A & A & 0 \\ B & 0 & B \end{bmatrix}$$
$$= \operatorname{rank} A + \operatorname{rank} B + \operatorname{rank}(A+B).$$

(Proof: Use Frobenius' inequality with $A \triangleq C^{\mathrm{T}} \triangleq \begin{bmatrix} I & I \end{bmatrix}$ and with B replaced by $\left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$.)

Fact 2.10.30. Let $A \in \mathbb{F}^{n\times m}$, and let $B \in \mathbb{F}^{k\times l}$ be a submatrix of A. Then,

$$k + l - \operatorname{rank} B \le n + m - \operatorname{rank} A.$$

(Proof: See [57].)

2.11 Facts on Identities

Fact 2.11.1. Let $A \in \mathbb{F}^{2\times 2}$, assume that $\operatorname{tr} A + 2\sqrt{\det A} \ne 0$, and define $B \in \mathbb{F}^{2\times 2}$ by

$$B \triangleq \left(\operatorname{tr} A + 2\sqrt{\det A}\right)^{-1/2}\left(A + \sqrt{\det A} I\right).$$

Then, $B^2 = A$. (Proof: See [261, pp. 84, 266, 267].)

Fact 2.11.2. $\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}^3 = \left[\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right]^3 = I_2.$

Fact 2.11.3. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then, $AE_{i,j,m\times l}B = \operatorname{col}_i(A)\operatorname{row}_j(B)$.

Fact 2.11.4. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $C \in \mathbb{F}^{l\times n}$. Then,

$$\operatorname{tr} ABC = \sum_{i=1}^{n} \operatorname{row}_i(A) B \operatorname{col}_i(C).$$

Fact 2.11.5. Let $A \in \mathbb{F}^{n\times m}$. Then, $Ax = 0$ for all $x \in \mathbb{F}^m$ if and only if $A = 0$.

Fact 2.11.6. Let $x, y \in \mathbb{F}^n$. Then, $x^*x = y^*y$ and $\operatorname{Im} x^*y = 0$ if and only if $x - y$ is orthogonal to $x + y$.

Fact 2.11.7. Let $x, y \in \mathbb{R}^n$. Then, $xx^{\mathrm{T}} = yy^{\mathrm{T}}$ if and only if either

$x = y$ or $x = -y$.

Fact 2.11.8. Let $x, y \in \mathbb{R}^n$. Then, $xy^{\mathrm{T}} = yx^{\mathrm{T}}$ if and only if x and y are linearly dependent.

Fact 2.11.9. Let $x, y \in \mathbb{R}^n$. Then, $xy^{\mathrm{T}} = -yx^{\mathrm{T}}$ if and only if either $x = 0$ or $y = 0$. (Proof: If $x_{(i)} \neq 0$ and $y_{(j)} \neq 0$, then $x_{(j)} = y_{(i)} = 0$ and $0 \neq x_{(i)}y_{(j)} \neq x_{(j)}y_{(i)} = 0$.)

Fact 2.11.10. Let $x, y \in \mathbb{R}^n$. Then, $yx^{\mathrm{T}} + xy^{\mathrm{T}} = y^{\mathrm{T}}yxx^{\mathrm{T}}$ if and only if either $x = 0$ or $y = \frac{1}{2}y^{\mathrm{T}}yx$.

Fact 2.11.11. Let $x, y \in \mathbb{F}^n$. Then,

$$(xy^*)^r = (y^*x)^{r-1}xy^*.$$

Fact 2.11.12. Let $y \in \mathbb{F}^n$ and $x \in \mathbb{F}^m$. Then, there exists a matrix $A \in \mathbb{F}^{n\times m}$ such that $y = Ax$ if and only if either $y = 0$ or $x \neq 0$. If $y = 0$, then one such matrix is $A = 0$. If $x \neq 0$, then one such matrix is

$$A = (x^*x)^{-1}yx^*.$$

(Remark: See Fact 3.4.33.)

Fact 2.11.13. Let $A \in \mathbb{F}^{n\times m}$. Then, $A = 0$ if and only if $\operatorname{tr} AA^* = 0$.

Fact 2.11.14. Let $A, B \in \mathbb{F}^{n\times n}$, and define $\mathcal{A} \triangleq \left[\begin{smallmatrix} A & A \\ A & A \end{smallmatrix}\right]$ and $\mathcal{B} \triangleq \left[\begin{smallmatrix} B & -B \\ -B & B \end{smallmatrix}\right]$. Then,

$$\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = 0.$$

Fact 2.11.15. Let $A \in \mathbb{F}^{n\times n}$ and $k \in \mathbb{P}$. Then,

$$\operatorname{Re} \operatorname{tr} A^{2k} \leq \operatorname{tr} A^kA^{k*} \leq \operatorname{tr}\, (AA^*)^k.$$

(Remark: To prove the left-hand inequality consider $\operatorname{tr}\,(A^k - A^{k*})(A^{k*} - A^k)$. For the right-hand inequality when $k = 2$, consider $\operatorname{tr}\,(AA^* - A^*A)^2$.)

Fact 2.11.16. Let $A \in \mathbb{F}^{n\times n}$. Then, $\operatorname{tr} A^k = 0$ for all $k = 1, \ldots, n$ if and only if $A^n = 0$. (Proof: For sufficiency, Fact 4.10.2 implies that $\operatorname{spec}(A) = \{0\}$, and thus the Jordan form of A is a block-diagonal matrix each of whose diagonally located blocks is a standard nilpotent matrix. For necessity, see [629, p. 112].)

Fact 2.11.17. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{tr} A = 0$. If $A^2 = A$, then $A = 0$. If $A^k = A$, where $k \geq 4$ and $2 \leq n < p$, where p is the smallest prime divisor of $k-1$, then $A = 0$. (Proof: See [152].)

Fact 2.11.18. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $AB = 0$. Then, for all $k \in \mathbb{P}$,

$$\operatorname{tr}(A+B)^k = \operatorname{tr} A^k + \operatorname{tr} B^k.$$

Fact 2.11.19. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $AB + BA = \frac{1}{2}\big[(A+B)^2 - (A-B)^2\big]$.

ii) $(A+B)(A-B) = A^2 - B^2 - [A, B]$.

iii) $(A-B)(A+B) = A^2 - B^2 + [A, B]$.

iv) $A^2 - B^2 = \frac{1}{2}[(A+B)(A-B) + (A-B)(A+B)]$.

Fact 2.11.20. Let $A, B \in \mathbb{F}^{n\times n}$ and $k \in \mathbb{P}$. Then,

$$A^k - B^k = \sum_{i=0}^{k-1} A^i(A-B)B^{k-1-i}.$$

Fact 2.11.21. Let $\alpha \in \mathbb{R}$ and $A \in \mathbb{R}^{n\times n}$. Then, the matrix equation $\alpha A + A^{\mathrm{T}} = 0$ has a nonzero solution A if and only if $\alpha = 1$ or $\alpha = -1$.

2.12 Facts on Determinants

Fact 2.12.1. $\det \left[\begin{smallmatrix} 0 & I_n \\ I_m & 0 \end{smallmatrix}\right] = (-1)^{nm}$.

Fact 2.12.2. $\det \hat{I}_n = (-1)^{\lfloor n/2 \rfloor} = (-1)^{n(n-1)/2}$.

Fact 2.12.3. $\det(I_n + \alpha 1_{n\times n}) = 1 + \alpha n$.

Fact 2.12.4. Let $A \in \mathbb{F}^{n\times n}$, $x, y \in \mathbb{F}^n$, and $a \in \mathbb{F}$. Then,

$$\begin{bmatrix} A & x \\ y^{\mathrm{T}} & a \end{bmatrix} = \begin{cases} \begin{bmatrix} I & 0 \\ y^{\mathrm{T}}A^{-1} & 1 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & a - y^{\mathrm{T}}A^{-1}x \end{bmatrix} \begin{bmatrix} I & A^{-1}x \\ 0 & 1 \end{bmatrix}, & \det A \neq 0, \\ \begin{bmatrix} I & A^{-1}x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A - A^{-1}xy^{\mathrm{T}} & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} I & 0 \\ a^{-1}y^{\mathrm{T}} & 1 \end{bmatrix}, & a \neq 0. \end{cases}$$

(Remark: See Fact 6.4.24.)

Fact 2.12.5. Let $A \in \mathbb{F}^{n\times n}$, $x, y \in \mathbb{F}^n$, and $a \in \mathbb{F}$. Then,

$$\det \begin{bmatrix} A & x \\ y^{\mathrm{T}} & a \end{bmatrix} = a(\det A) - y^{\mathrm{T}}A^{\mathrm{A}}x.$$

Hence,

$$\det\begin{bmatrix} A & x \\ y^{\mathrm{T}} & a \end{bmatrix} = \begin{cases} (\det A)\big(a - y^{\mathrm{T}}A^{-1}x\big), & \det A \neq 0, \\ a\det\big(A - A^{-1}xy^{\mathrm{T}}\big), & a \neq 0, \\ -y^{\mathrm{T}}A^{\mathrm{A}}x, & a = 0. \end{cases}$$

In particular,

$$\det\begin{bmatrix} A & Ax \\ y^{\mathrm{T}}A & y^{\mathrm{T}}Ax \end{bmatrix} = 0.$$

Finally,

$$\det\big(A + xy^{\mathrm{T}}\big) = \det A + y^{\mathrm{T}}A^{\mathrm{A}}x = -\det\begin{bmatrix} A & x \\ y^{\mathrm{T}} & -1 \end{bmatrix}.$$

(Remark: See Fact 2.12.6 and Fact 2.13.3.)

Fact 2.12.6. Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, and $a \in \mathbb{R}$. Then,

$$\det\begin{bmatrix} A & b \\ b^{\mathrm{T}} & a \end{bmatrix} = a(\det A) - b^{\mathrm{T}}A^{\mathrm{A}}b.$$

In particular,

$$\det\begin{bmatrix} A & b \\ b^{\mathrm{T}} & a \end{bmatrix} = \begin{cases} (\det A)\big(a - b^{\mathrm{T}}A^{-1}b\big), & \det A \neq 0, \\ a\det\big(A - a^{-1}bb^{\mathrm{T}}\big), & a \neq 0, \\ -b^{\mathrm{T}}A^{\mathrm{A}}b, & a = 0. \end{cases}$$

(Remark: This identity is a specialization of Fact 2.12.5.)

Fact 2.12.7. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{rank}\begin{bmatrix} A & A \\ A & A \end{bmatrix} = \operatorname{rank}\begin{bmatrix} A & -A \\ -A & A \end{bmatrix} = \operatorname{rank} A,$$

$$\operatorname{rank}\begin{bmatrix} A & A \\ -A & A \end{bmatrix} = 2\operatorname{rank} A,$$

$$\det\begin{bmatrix} A & A \\ A & A \end{bmatrix} = \det\begin{bmatrix} A & -A \\ -A & A \end{bmatrix} = 0,$$

$$\det\begin{bmatrix} A & A \\ -A & A \end{bmatrix} = 2^n(\det A)^2.$$

(Remark: See Fact 2.12.8.)

Fact 2.12.8. Let $a, b, c, d \in \mathbb{F}$, let $A \in \mathbb{F}^{n\times n}$, and define $\mathcal{A} \triangleq \left[\begin{smallmatrix} aA & bA \\ cA & dA \end{smallmatrix}\right]$.

Then,

$$\operatorname{rank}\mathcal{A} = \left(\operatorname{rank}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)\operatorname{rank} A$$

and

$$\det\mathcal{A} = (ad-bc)^n(\det A)^2.$$

(Remark: See Fact 2.12.7.) (Proof: See Proposition 7.1.11 and Fact 7.4.20.)

Fact 2.12.9. Let $A\in\mathbb{F}^{n\times m}$, $B\in\mathbb{F}^{m\times n}$, and $m<n$. Then, $\det AB = 0$.

Fact 2.12.10. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times n}$, and $n \le m$. Then, $\det AB$ is equal to the sum of all $\binom{n}{m}$ products of pairs of subdeterminants of A and B formed by choosing n columns of A and the corresponding n rows of B. (Remark: This identity is the *Binet-Cauchy formula*, which yields Proposition 2.7.1 in the case $n = m$.)

Fact 2.12.11. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, and let $b \in \mathbb{F}^n$. Then, the solution $x \in \mathbb{F}^n$ of $Ax = b$ is given by

$$x = \begin{bmatrix} \dfrac{\det\left(A \overset{1}{\leftarrow} b\right)}{\det A} \\ \vdots \\ \dfrac{\det\left(A \overset{n}{\leftarrow} b\right)}{\det A} \end{bmatrix}.$$

(Proof: Note that $A\left(I \overset{i}{\leftarrow} x\right) = A \overset{i}{\leftarrow} b$. Since $\det\left(I \overset{i}{\leftarrow} x\right) = x_{(i)}$, it follows that $(\det A)x_{(i)} = \det\left(A \overset{i}{\leftarrow} b\right)$.) (Remark: This identity is *Cramer's rule.*)

Fact 2.12.12. Let $A \in \mathbb{F}^{n\times m}$ be right invertible, and let $b \in \mathbb{F}^n$. Then, a solution $x \in \mathbb{F}^m$ of $Ax = b$ is given by

$$x_{(i)} = \frac{\det\left[\left(A \overset{i}{\leftarrow} b\right)A^*\right] - \det\left[\left(A \overset{i}{\leftarrow} 0\right)A^*\right]}{\det(AA^*)},$$

for all $i = 1,\dots,m$. (Proof: See [349].)

Fact 2.12.13. Let A, B, C, D be conformable matrices with entries in $\mathbb{F}$. Then,

$$\begin{bmatrix} A & AB \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}\begin{bmatrix} A & 0 \\ C-CA & D-CB \end{bmatrix}\begin{bmatrix} I & B \\ 0 & I \end{bmatrix},$$

$$\det\begin{bmatrix} A & AB \\ C & D \end{bmatrix} = (\det A)\det(D-CB),$$

$$\begin{bmatrix} A & B \\ CA & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}\begin{bmatrix} A & B-AB \\ 0 & D-CB \end{bmatrix}\begin{bmatrix} I & B \\ 0 & I \end{bmatrix},$$

$$\det\begin{bmatrix} A & B \\ CA & D \end{bmatrix} = (\det A)\det(D-CB),$$

$$\begin{bmatrix} A & BD \\ C & D \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix}\begin{bmatrix} A-BC & 0 \\ C-DC & D \end{bmatrix}\begin{bmatrix} I & 0 \\ C & I \end{bmatrix},$$

$$\det\begin{bmatrix} A & BD \\ C & D \end{bmatrix} = \det(A-BC)\det D,$$

$$\begin{bmatrix} A & B \\ DC & D \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix}\begin{bmatrix} A-BC & B-BD \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ C & I \end{bmatrix},$$

$$\det\begin{bmatrix} A & B \\ DC & D \end{bmatrix} = \det(A-BC)\det\ D.$$

(Remark: See Fact 6.4.24.)

Fact 2.12.14. Let $A_1, A_2, B_1, B_2 \in \mathbb{F}^{n\times m}$, and define $\mathcal{A} \triangleq \left[\begin{smallmatrix} A_1 & A_2 \\ A_2 & A_1 \end{smallmatrix}\right]$ and $\mathcal{B} \triangleq \left[\begin{smallmatrix} B_1 & B_2 \\ B_2 & B_1 \end{smallmatrix}\right]$. Then,

$$\operatorname{rank}\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B} & \mathcal{A} \end{bmatrix} = \sum_{i=1}^{4}\operatorname{rank} C_i,$$

where $C_1 \triangleq A_1+A_2+B_1+B_2$, $C_2 \triangleq A_1+A_2-B_1-B_2$, $C_3 \triangleq A_1-A_2+B_1-B_2$, and $C_4 \triangleq A_1-A_2-B_1+B_2$. If, in addition, $n=m$, then

$$\det\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B} & \mathcal{A} \end{bmatrix} = \prod_{i=1}^{4}\det C_i.$$

(Proof: See [551].) (Remark: See Fact 3.11.3.)

Fact 2.12.15. Let $A, B, C, D \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{rank}\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] = n$. Then,

$$\det\begin{bmatrix} \det A & \det B \\ \det C & \det D \end{bmatrix} = 0.$$

Fact 2.12.16. Let $A, B, C, D \in \mathbb{F}^{n \times n}$. Then,

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{cases} \det(DA - CB), & AB = BA, \\ \det(AD - CB), & AC = CA, \\ \det(AD - BC), & DC = CD, \\ \det(DA - BC), & DB = BD. \end{cases}$$

(Remark: These identities are *Schur's formulas.* See [66, p. 11].) (Proof: If A is nonsingular, then

$$\begin{aligned} \det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= (\det A)\det\big(D - CA^{-1}B\big) = \det\big(DA - CA^{-1}BA\big) \\ &= \det(DA - CB). \end{aligned}$$

Alternatively, note the identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & DA - CB \end{bmatrix} \begin{bmatrix} I & BA^{-1} \\ 0 & A^{-1} \end{bmatrix}.$$

If A is singular, then replace A by $A + \varepsilon I$ and use continuity.) (Problem: Find a direct proof for the case in which A is singular.)

Fact 2.12.17. Let $A, B, C, D \in \mathbb{F}^{n \times n}$. Then,

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{cases} \det\big(DA^{\mathrm{T}} - CB^{\mathrm{T}}\big), & AB^{\mathrm{T}} = BA^{\mathrm{T}}, \\ \det\big(A^{\mathrm{T}}D - C^{\mathrm{T}}B\big), & A^{\mathrm{T}}C = C^{\mathrm{T}}A, \\ \det\big(AD^{\mathrm{T}} - BC^{\mathrm{T}}\big), & DC^{\mathrm{T}} = CD^{\mathrm{T}}, \\ \det\big(D^{\mathrm{T}}A - B^{\mathrm{T}}C\big), & D^{\mathrm{T}}B = B^{\mathrm{T}}D, \\ (-1)^{\operatorname{rank} B}\det\big(AD^{\mathrm{T}} + BC^{\mathrm{T}}\big), & AB^{\mathrm{T}} = -BA^{\mathrm{T}}, \\ (-1)^{\operatorname{rank} A}\det\big(A^{\mathrm{T}}D + C^{\mathrm{T}}B\big), & A^{\mathrm{T}}C = -C^{\mathrm{T}}A, \\ (-1)^{\operatorname{rank} C}\det\big(AD^{\mathrm{T}} + BC^{\mathrm{T}}\big), & DC^{\mathrm{T}} = -CD^{\mathrm{T}}, \\ (-1)^{\operatorname{rank} D}\det\big(DA^{\mathrm{T}} + BC^{\mathrm{T}}\big), & D^{\mathrm{T}}B = -B^{\mathrm{T}}D. \end{cases}$$

(Proof: If A is nonsingular and $AB^{\mathrm{T}} = BA^{\mathrm{T}}$, then

$$\begin{aligned} \det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= (\det A)\det\big(D - CA^{-1}B\big) \\ &= \det\big(DA^{\mathrm{T}} - CA^{-1}BA^{\mathrm{T}}\big) = \det\big(DA^{\mathrm{T}} - CB^{\mathrm{T}}\big). \end{aligned}$$

If A is singular, then a continuity argument can be used with B symmetrized by means of pre- and post-multiplication if necessary. If A is nonsingular and $AB^{\mathrm{T}} = -BA^{\mathrm{T}}$, then AB^{T} is skew symmetric, B has even rank, and $\det\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] = \det\left(DA^{\mathrm{T}} + CB^{\mathrm{T}}\right)$. See [393, 587].)

Fact 2.12.18. Let $A, B, C, D \in \mathbb{F}^{n\times n}$. Then,

$$\det\begin{bmatrix} A & B \\ C & D \end{bmatrix}^2$$

$$= \begin{cases} \det\left(A^2 + BC\right)\det\left(CB + D^2\right), & AB = -BD \text{ or } CA = -DC, \\ (-1)^n\det(AC + BA)\det(CD + DB), & AD = -B^2 \text{ or } C^2 = -DA, \\ (-1)^n\det(AB + BD)\det(CA + DC), & A^2 = -BC \text{ or } CB = -D^2, \\ \det\left(AD + B^2\right)\det\left(C^2 + DA\right), & AC = -BA \text{ or } CD = -DB, \\ \det\left(AA^{\mathrm{T}} + BB^{\mathrm{T}}\right)\det\left(CC^{\mathrm{T}} + DD^{\mathrm{T}}\right), & AC^{\mathrm{T}} = -BD^{\mathrm{T}} \text{ or } CA^{\mathrm{T}} = -DB^{\mathrm{T}}, \\ (-1)^n\det\left(AB^{\mathrm{T}} + BA^{\mathrm{T}}\right)\det\left(CD^{\mathrm{T}} + DC^{\mathrm{T}}\right), & AD^{\mathrm{T}} = -BC^{\mathrm{T}} \text{ or } CB^{\mathrm{T}} = -DA^{\mathrm{T}}, \\ \left[\det\left(AD^{\mathrm{T}} + BC^{\mathrm{T}}\right)\right]^2, & AB^{\mathrm{T}} = -BA^{\mathrm{T}} \text{ or } CD^{\mathrm{T}} = -DC^{\mathrm{T}}. \end{cases}$$

(Proof: Form $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]^2$, $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]\left[\begin{smallmatrix} C & D \\ A & B \end{smallmatrix}\right]$, etc.)

Fact 2.12.19. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\det\begin{bmatrix} A^*A & A^*B \\ B^*A & B^*B \end{bmatrix} = \begin{cases} \det(A^*A)\det[B^*B - B^*A(A^*A)^{-1}A^*B], & \operatorname{rank} A = m, \\ \det(B^*B)\det[A^*A - A^*B(B^*B)^{-1}B^*A], & \operatorname{rank} B = l, \\ 0, & n < m + l. \end{cases}$$

Fact 2.12.20. Let $A \in \mathbb{F}^{n\times n}$, and assume that either $A_{(i,j)} = 0$ for all i, j such that $i + j < n + 1$ or $A_{(i,j)} = 0$ for all i, j such that $i + j > n + 1$. Then,

$$\det A = (-1)^{\lfloor n/2\rfloor}\prod_{i=1}^{n} A_{(i,n+1-i)}.$$

(Remark: A is *lower reverse triangular*.)

Fact 2.12.21. Define $A \in \mathbb{R}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Then,

$$\det A = (-1)^{n+1}.$$

Fact 2.12.22. Let $a_1, \ldots, a_n \in \mathbb{F}$. Then,

$$\det \begin{bmatrix} 1+a_1 & a_2 & \cdots & a_n \\ a_1 & 1+a_2 & \cdots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & 1+a_n \end{bmatrix} = 1 + \sum_{i=1}^n a_i.$$

Fact 2.12.23. Let $a_1, \ldots, a_n \in \mathbb{F}$ be nonzero. Then,

$$\det \begin{bmatrix} \frac{1+a_1}{a_1} & 1 & \cdots & 1 \\ 1 & \frac{1+a_2}{a_2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{1+a_n}{a_n} \end{bmatrix} = \frac{1 + \sum_{i=1}^n a_i}{\prod_{i=1}^n a_i}.$$

Fact 2.12.24. Let $a, b, c_1, \ldots, c_n \in \mathbb{F}$, define $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} c_1 & a & a & \cdots & a \\ b & c_2 & a & \cdots & a \\ b & b & c_3 & \ddots & a \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ b & b & b & \cdots & c_n \end{bmatrix},$$

and let $p(x) = (c_1 - x)(c_2 - x)\cdots(c_n - x)$ and $p_i(x) = p(x)/(c_i - x)$ for all $i = 1, \ldots, n$. Then,

$$\det A = \begin{cases} \dfrac{bp(a) - ap(b)}{b-a}, & b \neq a, \\ a\displaystyle\sum_{i=1}^{n-1} p_i(a) + c_n p_n(a), & b = a. \end{cases}$$

In particular,

$$\det\begin{bmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \ddots & b \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix} = (a-b)^{n-1}[a+(n-1)b]$$

and

$$\det(aI_n + b1_{n\times n}) = a^{n-1}(a+bn).$$

(Remark: See Fact 4.10.11.) (Remark: The matrix $aI_n + b1_{n\times n}$ arises in combinatorics. See [114, 116].)

Fact 2.12.25. Let $A, B \in \mathbb{F}^{n\times n}$, and define $\mathcal{A} \in \mathbb{F}^{kn\times kn}$ by

$$\mathcal{A} \triangleq \begin{bmatrix} A & B & B & \cdots & B \\ B & A & B & \cdots & B \\ B & B & A & \ddots & B \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ B & B & B & \cdots & A \end{bmatrix}.$$

Then,

$$\det\mathcal{A} = [\det(A-B)]^{k-1}\det[A+(k-1)B].$$

If $k = 2$, then

$$\det\begin{bmatrix} A & B \\ B & A \end{bmatrix} = \det[(A+B)(A-B)] = \det\big(A^2 - B^2 - [A,B]\big).$$

(Proof: See [238].)

Fact 2.12.26. Define the tridiagonal matrix $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} a+b & ab & 0 & \cdots & 0 & 0 \\ 1 & a+b & ab & \cdots & 0 & 0 \\ 0 & 1 & a+b & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & a+b & ab \\ 0 & 0 & 0 & \cdots & 1 & a+b \end{bmatrix}.$$

Then,

$$\det A = \begin{cases} (n+1)a^n, & a = b, \\ \dfrac{a^{n+1} - b^{n+1}}{a-b}, & a \neq b. \end{cases}$$

(Proof: See [339, pp. 401, 621].)

2.13 Facts on Adjugates and Inverses

Fact 2.13.1. Let $x, y \in \mathbb{F}^n$. Then,

$$\big(I + xy^{\mathrm{T}}\big)^{\mathrm{A}} = \big(1 + y^{\mathrm{T}}x\big)I - xy^{\mathrm{T}}$$

and

$$\det\big(I + xy^{\mathrm{T}}\big) = \det\big(I + yx^{\mathrm{T}}\big) = 1 + x^{\mathrm{T}}y = 1 + y^{\mathrm{T}}x.$$

If, in addition, $x^{\mathrm{T}}y \neq -1$, then

$$\big(I + xy^{\mathrm{T}}\big)^{-1} = I - \big(1 + x^{\mathrm{T}}y\big)^{-1}xy^{\mathrm{T}}.$$

Fact 2.13.2. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, and let $x, y \in \mathbb{F}^n$. Then,

$$\det\big(A + xy^{\mathrm{T}}\big) = \big(1 + y^{\mathrm{T}}A^{-1}x\big)\det A$$

and

$$\big(A + xy^{\mathrm{T}}\big)^{\mathrm{A}} = \big(1 + y^{\mathrm{T}}A^{-1}x\big)(\det A)I - A^{\mathrm{A}}xy^{\mathrm{T}}.$$

Furthermore, $\det\big(A + xy^{\mathrm{T}}\big) \neq 0$ if and only if $y^{\mathrm{T}}A^{-1}x \neq -1$. In this case,

$$\big(A + xy^{\mathrm{T}}\big)^{-1} = A^{-1} - \big(1 + y^{\mathrm{T}}A^{-1}x\big)^{-1}A^{-1}xy^{\mathrm{T}}A^{-1}.$$

(Remark: This identity is the *Sherman-Morrison-Woodbury formula.*)

Fact 2.13.3. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, let $x, y \in \mathbb{F}^n$, let $a \in \mathbb{F}$, and assume that $y^{\mathrm{T}}A^{-1}x \neq a$. Then,

$$\begin{bmatrix} A & x \\ y^{\mathrm{T}} & a \end{bmatrix}^{-1} = \frac{1}{a - y^{\mathrm{T}}A^{-1}x}\begin{bmatrix} \big(a - y^{\mathrm{T}}A^{-1}x\big)A^{-1} + A^{-1}xy^{\mathrm{T}}A^{-1} & -A^{-1}x \\ -y^{\mathrm{T}}A^{-1} & 1 \end{bmatrix}$$

$$= \frac{1}{a\det A - y^{\mathrm{T}}A^{\mathrm{A}}x}\begin{bmatrix} \big[\big(a - y^{\mathrm{T}}A^{-1}x\big)I + A^{-1}xy^{\mathrm{T}}\big]A^{\mathrm{A}} & -A^{\mathrm{A}}x \\ -y^{\mathrm{T}}A^{\mathrm{A}} & 1 \end{bmatrix}.$$

(Problem: Find an expression for $\left[\begin{smallmatrix} A & x \\ y^{\mathrm{T}} & a \end{smallmatrix}\right]^{-1}$ in the case $\det A = 0$ and $y^{\mathrm{T}}A^{\mathrm{A}}x \neq 0$. See Fact 2.12.5.)

Fact 2.13.4. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $\left(\overline{A}\right)^{\mathrm{A}} = \overline{A^{\mathrm{A}}}$.

ii) $\left(A^{\mathrm{T}}\right)^{\mathrm{A}} = \left(A^{\mathrm{A}}\right)^{\mathrm{T}}$.

iii) $(A^*)^{\mathrm{A}} = \left(A^{\mathrm{A}}\right)^*$.

iv) If $\alpha \in \mathbb{F}$, then $(\alpha A)^{\mathrm{A}} = \alpha^{n-1}A^{\mathrm{A}}$.

v) $\det A^{\mathrm{A}} = (\det A)^{n-1}$.

vi) $\left(A^{\mathrm{A}}\right)^{\mathrm{A}} = (\det A)^{n-2}A$.

vii) $\det \left(A^{\mathrm{A}}\right)^{\mathrm{A}} = (\det A)^{(n-1)^2}$.

Fact 2.13.5. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\det(A + 1_{n\times n}) - \det A = 1_{1\times n}^{\mathrm{T}} A^{\mathrm{A}} 1 = \sum_{i=1}^{n} \det\!\left(A \overset{i}{\leftarrow} 1_{n\times 1}\right).$$

(Proof: See [99].) (Remark: See Fact 2.12.5, Fact 2.13.8, and Fact 10.8.13.)

Fact 2.13.6. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is singular. Then,

$$\mathcal{R}(A) \subseteq \mathcal{N}\left(A^{\mathrm{A}}\right).$$

Hence,

$$\operatorname{rank} A \le \operatorname{def} A^{\mathrm{A}}$$

and

$$\operatorname{rank} A + \operatorname{rank} A^{\mathrm{A}} \le n.$$

Furthermore, if $n \ge 2$, then $\mathcal{R}(A) = \mathcal{N}\left(A^{\mathrm{A}}\right)$ if and only if $\operatorname{rank} A = n - 1$.

Fact 2.13.7. Let $A \in \mathbb{F}^{n\times n}$ and $n \ge 2$. Then, the following statements hold:

i) $\operatorname{rank} A^{\mathrm{A}} = n$ if and only if $\operatorname{rank} A = n$.

ii) $\operatorname{rank} A^{\mathrm{A}} = 1$ if and only if $\operatorname{rank} A = n - 1$.

iii) $A^{\mathrm{A}} = 0$ if and only if $\operatorname{rank} A < n - 1$.

(Proof: See [466, p. 12].) (Remark: See Fact 4.10.3.)

Fact 2.13.8. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\left(A^{\mathrm{A}}B\right)_{(i,j)} = \det\!\left(A \overset{i}{\leftarrow} \operatorname{col}_j(B)\right).$$

(Remark: See Fact 10.8.13.)

Fact 2.13.9. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $(AB)^{\mathrm{A}} = B^{\mathrm{A}}A^{\mathrm{A}}$.

ii) If B is nonsingular, then $\left(BAB^{-1}\right)^{\mathrm{A}} = BA^{\mathrm{A}}B^{-1}$.

iii) If $AB = BA$, then $A^{\mathrm{A}}B = BA^{\mathrm{A}}$, $AB^{\mathrm{A}} = B^{\mathrm{A}}A$, and $A^{\mathrm{A}}B^{\mathrm{A}} = B^{\mathrm{A}}A^{\mathrm{A}}$.

Fact 2.13.10. Let $A, B, C, D \in \mathbb{F}^{n\times n}$ and $ABCD = I$. Then, $ABCD = DABC = CDAB = BCDA$.

Fact 2.13.11. Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathbb{F}^{2\times 2}$, where $ad - bc \neq 0$. Then,

$$A^{-1} = (ad-bc)^{-1}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Furthermore, if $A = \left[\begin{smallmatrix} a & b & c \\ d & e & f \\ g & h & i \end{smallmatrix}\right] \in \mathbb{F}^{3\times 3}$ and $\beta = a(ei-fh) - b(di-fg) + c(dh-eg) \neq 0$, then

$$A^{-1} = \beta^{-1}\begin{bmatrix} ei-fh & -(bi-ch) & bf-ce \\ -(di-fg) & ai-cg & -(af-cd) \\ ah-eg & -(ah-bg) & ae-bd \end{bmatrix}.$$

Fact 2.13.12. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $A+B$ is nonsingular. Then,

$$A(A+B)^{-1}B = B(A+B)^{-1}A = A - A(A+B)^{-1}A = B - B(A+B)^{-1}B.$$

Fact 2.13.13. Let $A, B \in \mathbb{F}^{n\times n}$ be nonsingular. Then,

$$A^{-1} + B^{-1} = A^{-1}(A+B)B^{-1}.$$

Furthermore, $A^{-1} + B^{-1}$ is nonsingular if and only if $A + B$ is nonsingular. In this case,

$$\begin{aligned} \left(A^{-1}+B^{-1}\right)^{-1} &= A(A+B)^{-1}B \\ &= B(A+B)^{-1}A \\ &= A - A(A+B)^{-1}A \\ &= B - B(A+B)^{-1}B. \end{aligned}$$

Fact 2.13.14. Let $A, B \in \mathbb{F}^{n\times n}$ be nonsingular, and assume that $A-B$ is nonsingular. Then,

$$\left(A^{-1}-B^{-1}\right)^{-1} = A - A(A-B)^{-1}A.$$

Fact 2.13.15. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$, and assume that $I + AB$ is nonsingular. Then, $I + BA$ is nonsingular and

$$(I_n + AB)^{-1}A = A(I_m + BA)^{-1}.$$

(Remark: This result is the *push-through identity.*) Furthermore,

$$(I + AB)^{-1} = I - (I + AB)^{-1}AB.$$

Fact 2.13.16. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $I+BA$ is nonsingular. Then,

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B.$$

Fact 2.13.17. Let $A \in \mathbb{F}^{n\times n}$, and assume that A and $A + I$ are nonsingular. Then,

$$(A + I)^{-1} + \left(A^{-1} + I\right)^{-1} = (A + I)^{-1} + (A + I)^{-1}A = I.$$

Fact 2.13.18. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$(I + AA^*)^{-1} = I - A(I + A^*A)^{-1}A^*.$$

Fact 2.13.19. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, let $B \in \mathbb{F}^{n\times m}$, let $C \in \mathbb{F}^{m\times n}$, and assume that $A + BC$ and $I + CA^{-1}B$ are nonsingular. Then,

$$(A + BC)^{-1}B = A^{-1}B\left(I + CA^{-1}B\right)^{-1}.$$

Fact 2.13.20. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that B is nonsingular. Then,

$$A = B\left[I + B^{-1}(A - B)\right].$$

Fact 2.13.21. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and $A + B$ are nonsingular. Then, for all $k \in \mathbb{N}$,

$$\begin{aligned}(A + B)^{-1} &= \sum_{i=0}^{k} A^{-1}\left(-BA^{-1}\right)^{i} + \left(-A^{-1}B\right)^{k+1}(A + B)^{-1}\\ &= \sum_{i=0}^{k} A^{-1}\left(-BA^{-1}\right)^{i} + A^{-1}\left(-BA^{-1}\right)^{k+1}\left(I + BA^{-1}\right)^{-1}.\end{aligned}$$

Fact 2.13.22. Let $A, B \in \mathbb{F}^{n\times n}$ and $\alpha \in \mathbb{F}$, and assume that A, B, $\alpha A^{-1} + (1 - \alpha)B^{-1}$, and $\alpha B + (1 - \alpha)A$ are nonsingular. Then,

$$\begin{aligned}&\alpha A + (1 - \alpha)B - \left[\alpha A^{-1} + (1 - \alpha)B^{-1}\right]^{-1}\\ &\qquad = \alpha(1 - \alpha)(A - B)[\alpha B + (1 - \alpha)A]^{-1}(A - B).\end{aligned}$$

Fact 2.13.23. Let $A \in \mathbb{F}^{n\times m}$. If $\operatorname{rank} A = m$, then $(A^*A)^{-1}A^*$ is a left inverse of A. If $\operatorname{rank} A = n$, then $A^*(AA^*)^{-1}$ is a right inverse of A. (Remark: See Fact 3.4.19, Fact 3.4.20, and Fact 3.5.3.) (Problem: If $\operatorname{rank} A = n$ and $b \in \mathbb{R}^n$, then, for every solution $x \in \mathbb{R}^m$ of $Ax = b$, does there exist a right inverse A^{R} of A such that $x = A^{\mathrm{R}}b$?)

Fact 2.13.24. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = m$. Then, $A^{\mathrm{L}} \in \mathbb{F}^{m\times n}$ is a left inverse of A if and only if there exists $B \in \mathbb{F}^{m\times n}$ such that BA is nonsingular and

$$A^{\mathrm{L}} = (BA)^{-1}B.$$

(Proof: For necessity, let $B = A^{\mathrm{L}}$.)

Fact 2.13.25. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, and assume that A and B are right invertible. Then, AB is right invertible. If, in addition, A^{R} is a right inverse of A and B^{R} is a right inverse of B, then $B^{\mathrm{R}}A^{\mathrm{R}}$ is a right inverse of AB.

Fact 2.13.26. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, and assume that A and B are left invertible. Then, AB is left invertible. If, in addition, A^{L} is a left inverse of A and B^{L} is a left inverse of B, then $B^{\mathrm{L}}A^{\mathrm{L}}$ is a left inverse of AB.

Fact 2.13.27. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, $C \in \mathbb{F}^{m\times n}$, and $D \in \mathbb{F}^{m\times m}$, and assume that A and D are nonsingular. Then,

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}.$$

Fact 2.13.28. Let $A \in \mathbb{F}^{n\times n}, B \in \mathbb{F}^{m\times m}$, and $C \in \mathbb{F}^{m\times n}$. Then,

$$\det \begin{bmatrix} 0 & A \\ B & C \end{bmatrix} = \det \begin{bmatrix} C & B \\ A & 0 \end{bmatrix} = (-1)^{nm}(\det A)(\det B).$$

If, in addition, A and B are nonsingular, then

$$\begin{bmatrix} 0 & A \\ B & C \end{bmatrix}^{-1} = \begin{bmatrix} -B^{-1}CA^{-1} & B^{-1} \\ A^{-1} & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} C & B \\ A & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & A^{-1} \\ B^{-1} & -B^{-1}CA^{-1} \end{bmatrix}.$$

Fact 2.13.29. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, and $C \in \mathbb{F}^{m\times m}$, and assume that C is nonsingular. Then,

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix} = \begin{bmatrix} A - BC^{-1}B^{\mathrm{T}} & B \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{-1}B^{\mathrm{T}} & I \end{bmatrix}.$$

If, in addition, $A - BC^{-1}B^{\mathrm{T}}$ is nonsingular, then $\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}$ is nonsingular and

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}^{-1} = \begin{bmatrix} \big(A - BC^{-1}B^{\mathrm{T}}\big)^{-1} & -\big(A - BC^{-1}B^{\mathrm{T}}\big)^{-1}BC^{-1} \\ -C^{-1}B^{\mathrm{T}}\big(A - BC^{-1}B^{\mathrm{T}}\big)^{-1} & C^{-1}B^{\mathrm{T}}\big(A - BC^{-1}B^{\mathrm{T}}\big)^{-1}BC^{-1} + C^{-1} \end{bmatrix}.$$

Fact 2.13.30. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\det \begin{bmatrix} I & A \\ B & I \end{bmatrix} = \det(I - AB) = \det(I - BA).$$

If $\det(I - BA) \neq 0$, then

$$\begin{bmatrix} I & A \\ B & I \end{bmatrix}^{-1} = \begin{bmatrix} I + A(I - BA)^{-1}B & -A(I - BA)^{-1} \\ -(I - BA)^{-1}B & (I - BA)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} (I - AB)^{-1} & -(I - AB)^{-1}A \\ -B(I - AB)^{-1} & I + B(I - AB)^{-1}A \end{bmatrix}.$$

Fact 2.13.31. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix} = \frac{1}{2}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}\begin{bmatrix} A + B & 0 \\ 0 & A - B \end{bmatrix}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}.$$

Therefore,

$$\operatorname{rank} \begin{bmatrix} A & B \\ B & A \end{bmatrix} = \operatorname{rank}(A + B) + \operatorname{rank}(A - B).$$

Now, assume that $n = m$. Then,

$$\det \begin{bmatrix} A & B \\ B & A \end{bmatrix} = \det[(A + B)(A - B)] = \det\big(A^2 - B^2 - [A, B]\big).$$

If, in addition, $A + B$ and $A - B$ are nonsingular, then

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}^{-1} = \begin{bmatrix} (A + B)^{-1} + (A - B)^{-1} & (A + B)^{-1} - (A - B)^{-1} \\ (A + B)^{-1} - (A - B)^{-1} & (A + B)^{-1} + (A - B)^{-1} \end{bmatrix}.$$

Fact 2.13.32. Let $\mathcal{A} \triangleq \begin{bmatrix} A & B \\ 0_{m\times m} & C \end{bmatrix}$, where $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times n}$, and $C \in \mathbb{F}^{m\times n}$, and assume that CA is nonsingular. Furthermore, define $P \triangleq A(CA)^{-1}C$ and $P_\perp \triangleq I - P$. then $\mathcal{A}$ is nonsingular if and only if $P + P_\perp BP_\perp$

is nonsingular. In this case,

$$\mathcal{A}^{-1} = \begin{bmatrix} (CA)^{-1}(C - CBD) & -(CA)^{-1}CB(A - DBA)(CA)^{-1} \\ D & (A - DBA)(CA)^{-1} \end{bmatrix},$$

where $D \triangleq (P + P_\perp BP_\perp)^{-1}P_\perp$. (Proof: See [263].)

Fact 2.13.33. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times(n-m)}$, and assume that $\begin{bmatrix} A & B \end{bmatrix}$ is nonsingular and $A^*B = 0$. Then,

$$\begin{bmatrix} A & B \end{bmatrix}^{-1} = \begin{bmatrix} (A^*A)^{-1}A^* \\ (B^*B)^{-1}B^* \end{bmatrix}.$$

(Remark: See Fact 6.4.14.) (Problem: Find an expression for $\begin{bmatrix} A & B \end{bmatrix}^{-1}$ without assuming $A^*B = 0$.)

Fact 2.13.34. Let $M \triangleq \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be nonsingular, where $A \in \mathbb{F}^{n\times n}$ and $D \in \mathbb{F}^{m\times m}$, and let $\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \triangleq M^{-1}$, where $A' \in \mathbb{F}^{n\times n}$ and $D' \in \mathbb{F}^{m\times m}$. Then,

$$\det D' = \frac{\det A}{\det M}$$

and

$$\det A' = \frac{\det D}{\det M}.$$

Consequently, A is nonsingular if and only if D' is nonsingular, and D is nonsingular if and only if A' is nonsingular. (Proof: Use $M\begin{bmatrix} I & B' \\ 0 & D' \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & I \end{bmatrix}$. See [506].) (Remark: This identity is a special case of *Jacobi's identity.* See [287, p. 21].) (Remark: See Fact 3.6.7.)

Fact 2.13.35. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, and $C \in \mathbb{F}^{m\times l}$. Then,

$$\begin{bmatrix} I_n & A & B \\ 0 & I_m & C \\ 0 & 0 & I_l \end{bmatrix}^{-1} = \begin{bmatrix} I_n & -A & AC - B \\ 0 & I_m & -C \\ 0 & 0 & I_l \end{bmatrix}.$$

Fact 2.13.36. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, and define $A_0 \triangleq I_n$. Furthermore, for all $k = 1, \ldots, n$, let

$$\alpha_k = \tfrac{1}{k}\operatorname{tr} AA_{k-1},$$

and, for all $k = 1, \ldots, n-1$, let

$$A_k = AA_{k-1} - \alpha_k I.$$

Then,

$$A^{-1} = \tfrac{1}{\alpha_n}A_{n-1}.$$

(Remark: This result is due to Frame. See [74, p. 99].)

Fact 2.13.37. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and define $\{B_i\}_{i=1}^{\infty}$ by

$$B_{i+1} \triangleq 2B_i - B_iAB_i,$$

where $B_0 \in \mathbb{F}^{n\times n}$ satisfies $\operatorname{sprad}(I - B_0A) < 1$. Then,

$$B_i \to A^{-1}$$

as $i \to \infty$. (Proof: See [64, p. 167].) (Remark: This sequence is a Newton-Raphson algorithm.) (Remark: See Fact 6.3.18 for the case in which A is singular or not square.)

Fact 2.13.38. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, $A + A^{-*}$ is nonsingular. (Proof: Note that $AA^* + I$ is positive definite.)

Fact 2.13.39. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, $X = A^{-1}$ is the unique matrix satisfying

$$\operatorname{rank}\begin{bmatrix} A & I \\ I & X \end{bmatrix} = \operatorname{rank} A.$$

(Remark: See Fact 6.3.13 and Fact 6.5.5.) (Proof: See [203].)

2.14 Facts on Commutators

Fact 2.14.1. Let $A, B \in \mathbb{F}^{2\times 2}$. Then,

$$[A,B]^2 = \tfrac{1}{2}\operatorname{tr}[A,B]^2 I_2.$$

(Remark: See [211, 212].)

Fact 2.14.2. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $[A,B] = 0$. Then, $\left[A^k, B^l\right] = 0$ for all $k, l \in \mathbb{N}$.

Fact 2.14.3. Let $A, B, C \in \mathbb{F}^{n\times n}$. Then, the following identities hold:

i) $[A,A] = 0$.

ii) $[A,B] = [-A,-B] = -[B,A]$.

iii) $[A,B+C] = [A,B] + [A,C]$.

iv) $[\alpha A,B] = [A,\alpha B] = \alpha[A,B]$ for all $\alpha \in \mathbb{F}$.

v) $[A,[B,C]] + [B,[A,C]] + [C,[A,B]] = 0$.

vi) $[A,B]^{\mathrm{T}} = \left[B^{\mathrm{T}},A^{\mathrm{T}}\right] = -\left[A^{\mathrm{T}},B^{\mathrm{T}}\right]$.

vii) $\operatorname{tr}[A,B] = 0$.

viii) $\operatorname{tr} A^k[A,B] = \operatorname{tr} B^k[A,B] = 0$ for all $k \in \mathbb{P}$.

ix) $[[A,B],B-A] = [[B,A],A-B]$.

x) $[A,[A,B]] = -[A,[B,A]]$.

(Remark: *v)* is the *Jacobi identity*.)

Fact 2.14.4. Let $A,B \in \mathbb{F}^{n\times n}$. Then, for all $X \in \mathbb{F}^{n\times n}$,

$$\mathrm{ad}_{[A,B]} = [\mathrm{ad}_A, \mathrm{ad}_B],$$

that is,

$$\mathrm{ad}_{[A,B]}(X) = \mathrm{ad}_A[\mathrm{ad}_B(X)] - \mathrm{ad}_B[\mathrm{ad}_A(X)]$$

or

$$[[A,B],X] = [A,[B,X]] - [B,[A,X]].$$

Fact 2.14.5. Let $A \in \mathbb{F}^{n\times n}$ and, for all $X \in \mathbb{F}^{n\times n}$, define

$$\mathrm{ad}_A^k(X) \triangleq \begin{cases} \mathrm{ad}_A(X), & k=1,\\ \mathrm{ad}_A^{k-1}[\mathrm{ad}_A(X)], & k \ge 2. \end{cases}$$

Then, for all $X \in \mathbb{F}^{n\times n}$ and for all $k \ge 1$,

$$\mathrm{ad}_A^2(X) = [A,[A,X]] - [[A,X],A]$$

and

$$\mathrm{ad}_A^k(X) = \sum_{i=0}^{k} (-1)^{k-i}\binom{k}{i} A^i X A^{k-i}.$$

(Remark: The proof of Proposition 11.4.8 is based on $g\left(e^{t\,\mathrm{ad}_A}e^{t\,\mathrm{ad}_B}\right)$, where $g(z) \triangleq (\log z)/(z-1)$. See [496, p. 35].) (Remark: See Fact 11.11.4.) (Proof: For the last identity, see [466, pp. 176, 207].)

Fact 2.14.6. Let $A,B \in \mathbb{F}^{n\times n}$, and assume that $[A,B] = A$. Then, A is singular. (Proof: If A is nonsingular, then $\mathrm{tr}\,B = \mathrm{tr}\,ABA^{-1} = \mathrm{tr}\,B + n$.)

Fact 2.14.7. Let $A,B \in \mathbb{R}^{n\times n}$ be such that $AB = BA$. Then, there exists $C \in \mathbb{R}^{n\times n}$ such that $A^2 + B^2 = C^2$. (Proof: See [180].) (Remark: The result applies to real matrices only.)

2.15 Facts on Complex Matrices

Fact 2.15.1. Let $a,b \in \mathbb{R}$. Then, $\left[\begin{smallmatrix} a & b\\ -b & a\end{smallmatrix}\right]$ is a representation of the complex number $a + \jmath b$ that preserves addition, multiplication and inversion of complex numbers. In particular, if $a^2+b^2 \ne 0$, then

$$\begin{bmatrix} a & b\\ -b & a\end{bmatrix}^{-1} = \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{-b}{a^2+b^2}\\ \frac{b}{a^2+b^2} & \frac{a}{a^2+b^2}\end{bmatrix}$$

and

$$(a + \jmath b)^{-1} = \frac{a}{a^2+b^2} - \jmath\frac{b}{a^2+b^2}.$$

(Remark: $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ is a *rotation-dilation*. See Fact 3.11.1.)

Fact 2.15.2. Let $\nu, \omega \in \mathbb{R}$. Then,

$$\begin{bmatrix} \nu & \omega \\ -\omega & \nu \end{bmatrix} = \tfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ \jmath & -\jmath \end{bmatrix}\begin{bmatrix} \nu + \jmath\omega & 0 \\ 0 & \nu - \jmath\omega \end{bmatrix}\tfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ \jmath & -\jmath \end{bmatrix}^*$$

and

$$\begin{bmatrix} \nu & \omega \\ -\omega & \nu \end{bmatrix}^{-1} = \frac{1}{\nu^2+\omega^2}\begin{bmatrix} \nu & -\omega \\ \omega & \nu \end{bmatrix}.$$

(Remark: See Fact 2.15.1.)

Fact 2.15.3. Let $A, B \in \mathbb{R}^{n\times m}$. Then,

$$\begin{aligned}\begin{bmatrix} A & B \\ -B & A \end{bmatrix} &= \tfrac{1}{2}\begin{bmatrix} I & I \\ \jmath I & -\jmath I \end{bmatrix}\begin{bmatrix} A+\jmath B & 0 \\ 0 & A-\jmath B \end{bmatrix}\begin{bmatrix} I & -\jmath I \\ I & \jmath I \end{bmatrix}\\ &= \tfrac{1}{2}\begin{bmatrix} I & \jmath I \\ -\jmath I & -I \end{bmatrix}\begin{bmatrix} A-\jmath B & 0 \\ 0 & A+\jmath B \end{bmatrix}\begin{bmatrix} I & \jmath I \\ -\jmath I & -I \end{bmatrix}\\ &= \begin{bmatrix} I & 0 \\ \jmath I & I \end{bmatrix}\begin{bmatrix} A+\jmath B & B \\ 0 & A-\jmath B \end{bmatrix}\begin{bmatrix} I & 0 \\ -\jmath I & I \end{bmatrix}\end{aligned}$$

and

$$\operatorname{rank}(A+\jmath B) = \operatorname{rank}(A-\jmath B) = \tfrac{1}{2}\operatorname{rank}\begin{bmatrix} A & B \\ -B & A \end{bmatrix}.$$

Now, assume that $n = m$. Then,

$$\begin{aligned}\det\begin{bmatrix} A & B \\ -B & A \end{bmatrix} &= \det(A+\jmath B)\det(A-\jmath B)\\ &= |\det(A+\jmath B)|^2\\ &= \det\big[A^2+B^2+\jmath(AB-BA)\big]\\ &\ge 0\end{aligned}$$

and

$$\operatorname{mspec}\left(\begin{bmatrix} A & B \\ -B & A \end{bmatrix}\right) = \operatorname{mspec}(A+\jmath B)\cup\operatorname{mspec}(A-\jmath B).$$

If A is nonsingular, then

$$\det\begin{bmatrix} A & B \\ -B & A \end{bmatrix} = \det\big(A^2+ABA^{-1}B\big).$$

If $AB = BA$, then

$$\det \begin{bmatrix} A & B \\ -B & A \end{bmatrix} = \det\left(A^2 + B^2\right).$$

(Proof: If A is nonsingular, then use

$$\begin{bmatrix} A & B \\ -B & A \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ -A^{-1}B & I \end{bmatrix}$$

and

$$\det \begin{bmatrix} I & A^{-1}B \\ -A^{-1}B & I \end{bmatrix} = \det\left[I + \left(A^{-1}B\right)^2\right].$$

(Remark: See Fact 4.10.18 and [37, 551].)

Fact 2.15.4. Let $A, B \in \mathbb{R}^{n\times m}$ and $C, D \in \mathbb{R}^{m\times l}$. Then, $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} C & D \\ -D & C \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} A+C & B+D \\ -(B+D) & A+C \end{smallmatrix}\right]$ are representations of the complex matrices $A + \jmath B$, $C + \jmath D$, and their sum that preserve addition.

Fact 2.15.5. Let $A, B \in \mathbb{R}^{n\times m}$ and $C, D \in \mathbb{R}^{m\times l}$. Then, $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} C & D \\ -D & C \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} AC-BD & AD+BC \\ -(AD+BC) & AC-BD \end{smallmatrix}\right]$ are representations of the complex matrices $A + \jmath B$, $C + \jmath D$, and their product that preserve multiplication.

Fact 2.15.6. Let $A, B \in \mathbb{R}^{n\times n}$. Then, $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$ is a representation of the complex matrix $A + \jmath B$ that preserves addition, multiplication, and inversion of complex matrices. In particular, $A + \jmath B$ is nonsingular if and only if $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$ is nonsingular. Furthermore, if A is nonsingular, then $A + \jmath B$ is nonsingular if and only if $A + BA^{-1}B$ is nonsingular. In this case,

$$\begin{bmatrix} A & B \\ -B & A \end{bmatrix}^{-1} = \begin{bmatrix} \left(A + BA^{-1}B\right)^{-1} & -A^{-1}B\left(A + BA^{-1}B\right)^{-1} \\ A^{-1}B\left(A + BA^{-1}B\right)^{-1} & \left(A + BA^{-1}B\right)^{-1} \end{bmatrix}$$

and

$$(A + \jmath B)^{-1} = \left(A + BA^{-1}B\right)^{-1} - \jmath A^{-1}B\left(A + BA^{-1}B\right)^{-1}.$$

Finally, assume that B is nonsingular. Then, $A + \jmath B$ is nonsingular if and only if $B + AB^{-1}A$ is nonsingular. In this case,

$$\begin{bmatrix} A & B \\ -B & A \end{bmatrix}^{-1} = \begin{bmatrix} B^{-1}A\left(B + AB^{-1}A\right)^{-1} & -\left(B + AB^{-1}A\right)^{-1} \\ \left(B + AB^{-1}A\right)^{-1} & B^{-1}A\left(B + AB^{-1}A\right)^{-1} \end{bmatrix}$$

and

$$(A + \jmath B)^{-1} = B^{-1}A\left(B + AB^{-1}A\right)^{-1} - \jmath\left(B + AB^{-1}A\right)^{-1}.$$

(Problem: Consider the case in which A and B are singular.)

Fact 2.15.7. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\det\big(I + A\overline{A}\big) \ge 0.$$

(Proof: See [181].)

Fact 2.15.8. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\det \begin{bmatrix} A & B \\ -\overline{B} & \overline{A} \end{bmatrix} \ge 0.$$

If, in addition, A is nonsingular, then

$$\det \begin{bmatrix} A & B \\ -\overline{B} & \overline{A} \end{bmatrix} = |\det A| \det\Big(I + \overline{A^{-1}B}A^{-1}B\Big).$$

(Proof: See [628].)

Fact 2.15.9. Let $A, B \in \mathbb{R}^{n\times n}$, and define $C \in \mathbb{R}^{2n\times 2n}$ by $C \triangleq \begin{bmatrix} C_{11} & C_{12} & \cdots \\ C_{21} & \cdots & \\ \vdots & & \end{bmatrix}$, where $C_{ij} \triangleq \begin{bmatrix} A_{(i,j)} & B_{(i,j)} \\ -B_{(i,j)} & A_{(i,j)} \end{bmatrix}$ for all $i, j = 1, \ldots, n$. Then,

$$\det C = |\det(A + \jmath B)|^2.$$

(Proof: Note that

$$C = A \otimes I_2 + B \otimes J_2 = P_{2,n}(I_2 \otimes A + J_2 \otimes B)P_{2,n} = P_{2,n}\begin{bmatrix} A & B \\ -B & A \end{bmatrix}P_{2,n}.$$

See [109].)

2.16 Facts on Geometry

Fact 2.16.1. The points $x, y, z \in \mathbb{R}^2$ lie on one line if and only if

$$\det \begin{bmatrix} x & y & z \\ 1 & 1 & 1 \end{bmatrix} = 0.$$

The points $x, y, z \in \mathbb{R}^3$ lie on one line if and only if

$$\det \begin{bmatrix} x & y & z \end{bmatrix} = 0.$$

Fact 2.16.2. Let $\mathcal{S} \subset \mathbb{R}^2$ denote the triangle with vertices $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right], \left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right] \in \mathbb{R}^2$. Then,

$$\operatorname{area}(\mathcal{S}) = \tfrac{1}{2}\left|\det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}\right|.$$

Fact 2.16.3. Let $\mathcal{S} \subset \mathbb{R}^2$ denote the polygon with vertices $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right], \ldots,$

$\left[\begin{smallmatrix} x_n \\ y_n \end{smallmatrix}\right] \in \mathbb{R}^2$ arranged in counterclockwise order. Then,

$$\operatorname{area}(\mathcal{S}) = \tfrac{1}{2}\det\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} + \tfrac{1}{2}\det\begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} + \cdots$$

$$+ \tfrac{1}{2}\det\begin{bmatrix} x_{n-1} & x_n \\ y_{n-1} & y_n \end{bmatrix} + \tfrac{1}{2}\det\begin{bmatrix} x_n & x_1 \\ y_n & y_1 \end{bmatrix}.$$

(Remark: The polygon need not be convex, where "counterclockwise" is determined with respect to the inside of the polygon. See [529].)

Fact 2.16.4. Let $\mathcal{S} \subset \mathbb{R}^3$ denote the triangle with vertices $x, y, z \in \mathbb{R}^3$. Then,

$$\operatorname{area}(\mathcal{S}) = \tfrac{1}{2}\sqrt{[(y-x)\times(z-x)]^{\mathrm{T}}[(y-x)\times(z-x)]}.$$

Fact 2.16.5. Let $\mathcal{S} \subset \mathbb{R}^3$ denote the tetrahedron with vertices $x, y, z, w \in \mathbb{R}^3$. Then,

$$\operatorname{volume}(\mathcal{S}) = \tfrac{1}{6}\left|(x-w)^{\mathrm{T}}[(y-w)\times(z-w)]\right|.$$

Fact 2.16.6. Let $\mathcal{S} \subset \mathbb{R}^3$ denote the parallelepiped with vertices $x, y, z, y+z-x, w, w+y-x, w+z-x, w+z+y-2x \in \mathbb{R}^3$. Then,

$$\operatorname{volume}(\mathcal{S}) = \left|(w-x)^{\mathrm{T}}[(y-x)\times(z-x)]\right|.$$

Fact 2.16.7. Let $A \in \mathbb{R}^{n\times m}$, assume that $\operatorname{rank} A = m$, and let $\mathcal{S} \subset \mathbb{R}^n$ denote the parallelepiped in $\mathbb{R}^n$ generated by the columns of A. Then,

$$\operatorname{volume}(\mathcal{S}) = \left[\det\left(A^{\mathrm{T}}A\right)\right]^{1/2}.$$

If, in addition, $m = n$, then

$$\operatorname{volume}(\mathcal{S}) = |\det A|.$$

Fact 2.16.8. Let $\mathcal{S} \subset \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. Then,

$$\operatorname{volume}(A\mathcal{S}) = |\det A| \operatorname{volume}(\mathcal{S}).$$

(Remark: See [416, p. 468].)

2.17 Notes

The theory of determinants is discussed in [430, 560, 574]. The empty matrix is discussed in [435] and [484]. Convexity is the subject of [80, 103, 185,357,485,565,591]. Convex optimization theory is the subject of [79]. Our development of rank properties is based on [398]. Theorem 2.6.3 is based on [440]. The term "subdeterminant" is used in [456] and is equivalent to *minor*. The notation A^{A} for adjugate is used in [523]. Numerous papers on basic topics in matrix theory and linear algebra are collected in [129,130]. A geometric interpretation of $\mathcal{N}(A)$, $\mathcal{R}(A)$, $\mathcal{N}(A^{\mathrm{T}})$, and $\mathcal{R}(A^{\mathrm{T}})$ is given in [531]. Some reflections on matrix theory are given in [536, 549].

Chapter Three

Matrix Classes and Transformations

This chapter presents definitions of various types of matrices as well as transformations needed to analyze matrices.

3.1 Matrix Classes

In this section we categorize various types of matrices based upon their algebraic and structural properties.

The following definition introduces various types of square matrices.

Definition 3.1.1. For $A \in \mathbb{F}^{n\times n}$ define the following types of matrices:

i) A is *group invertible* if $\mathcal{R}(A) = \mathcal{R}(A^2)$.

ii) A is *range Hermitian* if $\mathcal{R}(A) = \mathcal{R}(A^*)$.

iii) A is *range symmetric* if $\mathcal{R}(A) = \mathcal{R}(A^{\mathrm{T}})$.

iv) A is *Hermitian* if $A = A^*$.

v) A is *symmetric* if $A = A^{\mathrm{T}}$.

vi) A is *skew Hermitian* if $A = -A^*$.

vii) A is *skew symmetric* if $A = -A^{\mathrm{T}}$.

viii) A is *normal* if $AA^* = A^*A$.

ix) A is *nonnegative semidefinite* ($A \geq 0$) if A is Hermitian and $x^*Ax \geq 0$ for all $x \in \mathbb{F}^n$.

x) A is *nonpositive semidefinite* ($A \leq 0$) if $-A$ is nonnegative semidefinite.

xi) A is *positive definite* ($A > 0$) if A is Hermitian and $x^*Ax > 0$ for all $x \in \mathbb{F}^n$ such that $x \neq 0$.

xii) A is *negative definite* ($A < 0$) if $-A$ is positive definite.

xiii) A is *semidissipative* if $A + A^*$ is nonpositive semidefinite.

xiv) A is *dissipative* if $A + A^*$ is negative definite.

xv) A is *unitary* if $A^*A = I$.

xvi) A is *orthogonal* if $A^{\mathrm{T}}A = I$.

xvii) A is a *projector* if A is Hermitian and idempotent.

xviii) A is a *reflector* if A is Hermitian and unitary.

xix) A is an *elementary projector* if there exists nonzero $x \in \mathbb{F}^n$ such that $A = I - (x^*x)^{-1}xx^*$.

xx) A is an *elementary reflector* if there exists nonzero $x \in \mathbb{F}^n$ such that $A = I - 2(x^*x)^{-1}xx^*$.

xxi) A is an *elementary matrix* if there exist $x, y \in \mathbb{F}^n$ such that $A = I - xy^{\mathrm{T}}$ and $x^{\mathrm{T}}y \neq 1$.

xxii) A is *involutory* if $A^2 = I$.

xxiii) A is *skew involutory* if $A^2 = -I$.

xxiv) A is *idempotent* if $A^2 = A$.

xxv) A is *tripotent* if $A^3 = A$.

xxvi) A is *nilpotent* if there exists $k \in \mathbb{P}$ such that $A^k = 0$.

xxvii) A is *reverse Hermitian* if $A = A^{\hat{*}}$.

xxviii) A is *reverse symmetric* if $A = A^{\hat{\mathrm{T}}}$.

xxix) A is a *permutation matrix* if every row of A and every column of A possesses one 1 and zeros otherwise.

Let $A \in \mathbb{F}^{n \times n}$ be Hermitian. Then, the function $f\colon\ \mathbb{F}^n \mapsto \mathbb{R}$ defined by

$$f(x) \triangleq x^*Ax \tag{3.1.1}$$

is a *quadratic form.*

The $n \times n$ *standard nilpotent matrix*, which has ones on the superdiagonal and zeros elsewhere, is denoted by N_n or just N. We define $N_1 \triangleq 0$ and $N_0 \triangleq 0_{0\times 0}$.

The following definition considers matrices that are not necessarily square.

Definition 3.1.2. For $A \in \mathbb{F}^{n \times m}$ define the following types of matrices:

i) A is *semicontractive* if $I_n - AA^*$ is nonnegative semidefinite.

ii) A is *contractive* if $I_n - AA^*$ is positive definite.

iii) A is *left inner* if $A^*A = I_m$.

iv) A is *right inner* if $AA^* = I_n$.

v) A is *centrohermitian* if $A = \hat{I}_n \overline{A} \hat{I}_m$.

vi) A is *centrosymmetric* if $A = \hat{I}_n A \hat{I}_m$.

vii) A is an *outer product* if there exist $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$ such that $A = xy^{\mathrm{T}}$.

The following definition introduces various types of structured matrices.

Definition 3.1.3. For $A \in \mathbb{F}^{n\times m}$ with $l \triangleq \min\{n, m\}$ define the following types of matrices:

i) A is *diagonal* if $A_{(i,j)} = 0$ for all $i \neq j$. If $n = m$, then

$$A = \operatorname{diag}\big(A_{(1,1)}, \ldots, A_{(n,n)}\big).$$

ii) A is *tridiagonal* if $A_{(i,j)} = 0$ for all $|i - j| > 1$.

iii) A is *reverse diagonal* if $A_{(i,j)} = 0$ for all $i + j \neq l + 1$. If $n = m$, then

$$A = \operatorname{revdiag}\big(A_{(1,n)}, \ldots, A_{(n,1)}\big).$$

iv) A is (*upper triangular, strictly upper triangular*) if $A_{(i,j)} = 0$ for all $(i \geq j, i > j)$.

v) A is (*lower triangular, strictly lower triangular*) if $A_{(i,j)} = 0$ for all $(i \leq j, i < j)$.

vi) A is (*upper Hessenberg, lower Hessenberg*) if $A_{(i,j)} = 0$ for all $(i > j + 1, i < j + 1)$.

vii) A is *Toeplitz* if $A_{(i,j)} = A_{(k,l)}$ for all $k - i = l - j$, that is,

$$A = \begin{bmatrix} a & b & c & \cdots \\ d & a & b & \ddots \\ e & d & a & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}.$$

viii) A is *Hankel* if $A_{(i,j)} = A_{(k,l)}$ for all $i+j = k+l$, that is,

$$A = \begin{bmatrix} a & b & c & \cdots \\ b & c & d & \iddots \\ c & d & e & \iddots \\ \vdots & \iddots & \iddots & \iddots \end{bmatrix}.$$

ix) A is *block diagonal* if

$$A = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{bmatrix} = \mathrm{diag}(A_1, \ldots, A_n),$$

where $A_i \in \mathbb{F}^{n_i \times n_i}$ for all $i = 1, \ldots, k$.

x) A is *upper block triangular* if

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ 0 & A_{22} & \cdots & A_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{kk} \end{bmatrix},$$

where $A_{ij} \in \mathbb{F}^{n_i \times n_j}$ for all $i, j = 1, \ldots, k$.

xi) A is *lower block triangular* if

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{bmatrix},$$

where $A_{ij} \in \mathbb{F}^{n_i \times n_j}$ for all $i, j = 1, \ldots, k$.

xii) A is *block Toeplitz* if $A_{(i,j)} = A_{(k,l)}$ for all $k - i = l - j$, that is,

$$A = \begin{bmatrix} A_1 & A_2 & A_3 & \cdots \\ A_4 & A_1 & A_2 & \ddots \\ A_5 & A_4 & A_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix},$$

where $A_i \in \mathbb{F}^{n_i \times m_i}$.

xiii) A is *block Hankel* if $A_{(i,j)} = A_{(k,l)}$ for all $i+j=k+l$, that is,

$$A = \begin{bmatrix} A_1 & A_2 & A_3 & \cdots \\ A_2 & A_3 & A_4 & \iddots \\ A_3 & A_4 & A_5 & \iddots \\ \vdots & \iddots & \iddots & \iddots \end{bmatrix},$$

where $A_i \in \mathbb{F}^{n_i \times m_i}$.

Define the matrix $J_n \in \mathbb{R}^{2n\times 2n}$ (or just J) by

$$J_{2n} \triangleq \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}. \tag{3.1.2}$$

In particular,

$$J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{3.1.3}$$

The following definition introduces various types of real matrices.

Definition 3.1.4. For $A \in \mathbb{R}^{n\times m}$ define the following types of matrices:

i) A is *nonnegative* ($A \geq\geq 0$) if $A_{(i,j)} \geq 0$ for all $i = 1,\ldots,n$ and $j = 1,\ldots,m$.

ii) A is *positive* ($A >> 0$) if $A_{(i,j)} > 0$ for all $i = 1,\ldots,n$ and $j = 1,\ldots,m$.

For $A \in \mathbb{R}^{2n\times 2n}$ define the following types of real matrices:

iii) A is *Hamiltonian* if $J^{-1}A^{\mathrm{T}}J = -A$.

iv) A is *symplectic* if A is nonsingular and $J^{-1}A^{\mathrm{T}}J = A^{-1}$.

Proposition 3.1.5. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) If A is Hermitian or skew Hermitian, then A is normal.

ii) If A is nonsingular or normal, then A is range Hermitian.

iii) If A is range Hermitian, idempotent, or tripotent, then A is group invertible.

iv) If A is a reflector, then A is tripotent.

v) If A is a permutation matrix, then A is orthogonal.

Proof. *i)* is immediate. To prove *ii)* note that if A is nonsingular, then

$\mathcal{R}(A) = \mathcal{R}(A^*) = \mathbb{F}^n$, and thus A is range Hermitian. If A is normal, then it follows from Theorem 2.4.3 that $\mathcal{R}(A) = \mathcal{R}(AA^*) = \mathcal{R}(A^*A) = \mathcal{R}(A^*)$, which proves that A is range Hermitian. To prove *iii*) note that if A is range Hermitian, then $\mathcal{R}(A) = \mathcal{R}(AA^*) = A\mathcal{R}(A^*) = A\mathcal{R}(A) = \mathcal{R}(A^2)$, while, if A is idempotent, then $\mathcal{R}(A) = \mathcal{R}(A^2)$. If A is tripotent, then $\mathcal{R}(A) = \mathcal{R}(A^3) = A^2\mathcal{R}(A) \subseteq \mathcal{R}(A^2) = A\mathcal{R}(A) \subseteq \mathcal{R}(A)$. Hence, $\mathcal{R}(A) = \mathcal{R}(A^2)$. □

3.2 Matrix Transformations

A variety of transformations can be employed for analyzing matrices.

Definition 3.2.1. Let $A, B \in \mathbb{F}^{n\times m}$. Then, the following terminology is defined:

i) A and B are *left equivalent* if there exists a nonsingular matrix $S_1 \in \mathbb{F}^{n\times n}$ such that $A = S_1B$.

ii) A and B are *right equivalent* if there exists a nonsingular matrix $S_2 \in \mathbb{F}^{m\times m}$ such that $A = BS_2$.

iii) A and B are *biequivalent* if there exist nonsingular matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ such that $A = S_1BS_2$.

iv) A and B are *unitarily left equivalent* if there exists a unitary matrix $S_1 \in \mathbb{F}^{n\times n}$ such that $A = S_1B$.

v) A and B are *unitarily right equivalent* if there exists a unitary matrix $S_2 \in \mathbb{F}^{m\times m}$ such that $A = BS_2$.

vi) A and B are *unitarily biequivalent* if there exist unitary matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ such that $A = S_1BS_2$.

Definition 3.2.2. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following terminology is defined:

i) A and B are *similar* if there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SBS^{-1}$.

ii) A and B are *congruent* if there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SBS^*$.

iii) A and B are *T-congruent* if there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SBS^{\mathrm{T}}$.

iv) A and B are *unitarily similar* if there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SBS^* = SBS^{-1}$.

The following results summarize some matrix properties that are preserved under left equivalence, right equivalence, biequivalence, similarity, congruence, and unitary similarity.

Proposition 3.2.3. Let $A, B \in \mathbb{F}^{n\times n}$. If A and B are similar, then the following statements hold:

i) A and B are biequivalent.

ii) $\operatorname{tr} A = \operatorname{tr} B$.

iii) $\det A = \det B$.

iv) A^k and B^k are similar for all $k \in \mathbb{P}$.

v) A^{k*} and B^{k*} are similar for all $k \in \mathbb{P}$.

vi) A is nonsingular if and only if B is; in this case, A^{-k} and B^{-k} are similar for all $k \in \mathbb{P}$.

vii) A is (group invertible, involutory, skew involutory, idempotent, tripotent, nilpotent) if and only if B is.

If A and B are congruent, then the following statements hold:

viii) A and B are biequivalent.

ix) A^* and B^* are congruent.

x) A is nonsingular if and only if B is; in this case, A^{-1} and B^{-1} are congruent.

xi) A is (range Hermitian, group invertible, Hermitian, skew Hermitian, nonnegative semidefinite, positive definite) if and only if B is.

If A and B are unitarily similar, then the following statements hold:

xii) A and B are similar.

xiii) A and B are congruent.

xiv) A is (range Hermitian, group invertible, normal, Hermitian, skew Hermitian, nonnegative semidefinite, positive definite, orthogonal, involutory, skew involutory, idempotent, tripotent, nilpotent) if and only if B is.

Definition 3.2.4. Let $\mathcal{S} \subseteq \mathbb{F}^{n\times n}$. Then, $\mathcal{S}$ is a *Lie algebra* if the following conditions are satisfied:

i) $\mathcal{S}$ is a subspace.

ii) If $A, B \in \mathcal{S}$, then $[A, B] \in \mathcal{S}$.

Proposition 3.2.5. The following sets are Lie algebras:

i) $\mathrm{gl}_{\mathbb{F}}(n) \triangleq \mathbb{F}^{n\times n}$.

ii) $\mathrm{pl}_{\mathbb{C}}(n) \triangleq \{A \in \mathbb{C}^{n\times n}\colon\ \mathrm{tr}\,A \in \mathbb{R}\}$.

iii) $\mathrm{sl}_{\mathbb{F}}(n) \triangleq \{A \in \mathbb{F}^{n\times n}\colon\ \mathrm{tr}\,A = 0\}$.

iv) $\mathrm{u}(n) \triangleq \{A \in \mathbb{C}^{n\times n}\colon\ A \text{ is skew Hermitian}\}$.

v) $\mathrm{su}(n) \triangleq \{A \in \mathbb{C}^{n\times n}\colon\ A \text{ is skew Hermitian and } \mathrm{tr}\,A = 0\}$.

vi) $\mathrm{so}(n) \triangleq \{A \in \mathbb{R}^{n\times n}\colon\ A \text{ is skew symmetric}\}$.

vii) $\mathrm{sp}(n) \triangleq \{A \in \mathbb{R}^{2n\times 2n}\colon\ A \text{ is Hamiltonian}\}$.

viii) $\mathrm{aff}_{\mathbb{F}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix}\colon\ A \in \mathrm{gl}_{\mathbb{F}}(n),\ b \in \mathbb{F}^n \right\}$.

ix) $\mathrm{se}_{\mathbb{C}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix}\colon\ A \in \mathrm{su}(n),\ b \in \mathbb{C}^n \right\}$.

x) $\mathrm{se}_{\mathbb{R}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix}\colon\ A \in \mathrm{so}(n),\ b \in \mathbb{R}^n \right\}$.

xi) $\mathrm{trans}_{\mathbb{F}}(n) \triangleq \left\{ \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}\colon\ b \in \mathbb{F}^n \right\}$.

Definition 3.2.6. Let $\mathcal{S} \subset \mathbb{F}^{n\times n}$. Then, $\mathcal{S}$ is a *group* if the following conditions are satisfied:

i) If $A \in \mathcal{S}$, then A is nonsingular.

ii) If $A \in \mathcal{S}$, then $A^{-1} \in \mathcal{S}$.

iii) If $A, B \in \mathcal{S}$, then $AB \in \mathcal{S}$.

Note that if $\mathcal{S} \subset \mathbb{F}^{n\times n}$ is a group, then $I_n \in \mathcal{S}$.

The following result lists several classical groups that are of importance in physics and engineering. In particular, $\mathrm{O}(1,3)$ is the *Lorentz group*, see, for example, [505, p. 126] or [496, p. 16].

Proposition 3.2.7. The following sets are groups:

i) $\mathrm{GL}_{\mathbb{F}}(n) \triangleq \{A \in \mathbb{F}^{n\times n}\colon\ \det A \neq 0\}$.

ii) $\mathrm{PL}_{\mathbb{F}}(n) \triangleq \{A \in \mathbb{F}^{n\times n}\colon\ \det A > 0\}$.

iii) $\mathrm{SL}_{\mathbb{F}}(n) \triangleq \{A \in \mathbb{F}^{n\times n}\colon\ \det A = 1\}$.

iv) $\mathrm{U}(n) \triangleq \{A \in \mathbb{C}^{n\times n}\colon\ A \text{ is unitary}\}$.

v) $\mathrm{O}(n) \triangleq \{A \in \mathbb{R}^{n\times n}\colon\ A \text{ is orthogonal}\}$.

vi) $\mathrm{U}(n,m) \triangleq \{A \in \mathbb{C}^{(n+m)\times(n+m)}\colon A^*\mathrm{diag}(I_n,-I_m)A = \mathrm{diag}(I_n,-I_m)\}$.

vii) $\mathrm{O}(n,m) \triangleq \{A \in \mathbb{R}^{(n+m)\times(n+m)}\colon A^{\mathrm{T}}\mathrm{diag}(I_n,-I_m)A = \mathrm{diag}(I_n,-I_m)\}$.

viii) $\mathrm{SU}(n) \triangleq \{A \in \mathrm{U}(n)\colon\ \det A = 1\}$.

ix) $\mathrm{SO}(n) \triangleq \{A \in \mathrm{O}(n)\colon \det A = 1\}$.

x) $\mathrm{Sp}(n) \triangleq \{A \in \mathbb{R}^{2n\times 2n}\colon\ A \text{ is symplectic}\}$.

xi) $\mathrm{Aff}_{\mathbb{F}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}\colon\ A \in \mathrm{GL}_{\mathbb{F}}(n),\ b \in \mathbb{F}^n \right\}$.

xii) $\mathrm{SE}_{\mathbb{C}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}\colon\ A \in \mathrm{SU}(n),\ b \in \mathbb{C}^n \right\}$.

xiii) $\mathrm{SE}_{\mathbb{R}}(n) \triangleq \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}\colon\ A \in \mathrm{SO}(n),\ b \in \mathbb{R}^n \right\}$.

xiv) $\mathrm{Trans}_{\mathbb{F}}(n) \triangleq \left\{ \begin{bmatrix} I & b \\ 0 & 1 \end{bmatrix}\colon\ b \in \mathbb{F}^n \right\}$.

The following result shows that groups can be used to define equivalence relations on $\mathbb{F}^{n\times m}$.

Proposition 3.2.8. Let $\mathcal{S}_1 \subset \mathbb{R}^{n\times n}$ and $\mathcal{S}_2 \subset \mathbb{R}^{m\times m}$ be groups. Then, the relation $\mathcal{R}$ defined on $\mathbb{F}^{n\times m}$ by

$$(A,B) \in \mathcal{R} \iff \text{there exist } S_1 \in \mathcal{S}_1 \text{ and } S_2 \in \mathcal{S}_2 \text{ such that } A = S_1BS_2$$

is an equivalence relation.

3.3 Facts on Range-Hermitian and Group-Invertible Matrices

Fact 3.3.1. Let $A \in \mathbb{F}^{n\times n}$. Then, A is range Hermitian if and only if $\mathcal{N}(A) = \mathcal{N}(A^*)$.

Fact 3.3.2. Let $A, B \in \mathbb{F}^{n\times n}$ be range Hermitian. Then,

$$\mathrm{rank}\, AB = \mathrm{rank}\, BA.$$

(Proof: See [52].)

Fact 3.3.3. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is group invertible.

ii) A^* is group invertible.

iii) $\mathcal{N}(A) = \mathcal{N}(A^2)$.

iv) $\mathcal{N}(A) \cap \mathcal{R}(A) = \{0\}$.

v) $\mathcal{N}(A) + \mathcal{R}(A) = \mathbb{F}^n$.

vi) A and A^2 are left equivalent.

vii) A and A^2 are right equivalent.

viii) $\operatorname{rank} A = \operatorname{rank} A^2$.

ix) $\operatorname{def} A = \operatorname{def} A^2$.

Fact 3.3.4. Let $A \in \mathbb{F}^{n\times n}$. If A is range Hermitian, then A is group invertible.

Fact 3.3.5. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A is dissipative and B is range Hermitian. Then, $\operatorname{ind} B = \operatorname{ind} AB$. (Proof: See [87].)

3.4 Facts on Hermitian and Skew-Hermitian Matrices

Fact 3.4.1. Let $A \in \mathbb{F}^{n\times m}$. Then, $AA^{\mathrm{T}} \in \mathbb{F}^{n\times n}$ and $A^{\mathrm{T}}A \in \mathbb{F}^{m\times m}$ are symmetric.

Fact 3.4.2. Let $A \in \mathbb{F}^{n\times n}$, let $k \in \mathbb{P}$, and assume that A is Hermitian. Then, $\mathcal{R}(A) = \mathcal{R}(A^k)$ and $\mathcal{N}(A) = \mathcal{N}(A^k)$.

Fact 3.4.3. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) $x^{\mathrm{T}}Ax = 0$ for all $x \in \mathbb{R}^n$ if and only if A is skew symmetric.

ii) A is symmetric and $x^{\mathrm{T}}Ax = 0$ for all $x \in \mathbb{R}^n$ if and only if $A = 0$.

Fact 3.4.4. Let $A \in \mathbb{C}^{n\times n}$. Then, the following statements hold:

i) x^*Ax is real for all $x \in \mathbb{C}^n$ if and only if A is Hermitian.

ii) x^*Ax is imaginary for all $x \in \mathbb{C}^n$ if and only if A is skew Hermitian.

iii) $x^*Ax = 0$ for all $x \in \mathbb{C}^n$ if and only if $A = 0$.

Fact 3.4.5. Let $A \in \mathbb{C}^{n\times n}$. Then, the following statements hold:

i) A is skew Hermitian if and only if $\jmath A$ is Hermitian.

ii) A is Hermitian if and only if $\jmath A$ is skew Hermitian.

iii) A is Hermitian if and only if $\operatorname{Re} A$ is symmetric and $\operatorname{Im} A$ is skew symmetric.

iv) A is skew Hermitian if and only if $\operatorname{Re} A$ is skew symmetric and $\operatorname{Im} A$ is symmetric.

v) A is nonnegative semidefinite if and only if $\operatorname{Re} A$ is nonnegative semidefinite.

vi) A is positive definite if and only if $\operatorname{Re} A$ is positive definite.

Fact 3.4.6. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) If A is (Hermitian, nonnegative semidefinite, positive definite), then so is A^{A}.

ii) If A is skew Hermitian and n is odd, then A^{A} is Hermitian.

iii) If A is skew Hermitian and n is even, then A^{A} is skew Hermitian.

iv) If A is normal, then so is A^{A}.

v) If A is diagonal, then so is A^{A}, and, for all $i = 1, \ldots, n$,

$$\left(A^{\mathrm{A}}\right)_{(i,i)} = \prod_{\substack{j=1\\ j\neq i}}^{n} A_{(j,j)}.$$

(Proof: Use Fact 2.13.9.) (Remark: See Fact 5.11.2.)

Fact 3.4.7. Let $A \in \mathbb{F}^{n\times n}$, assume that n is even, let $x \in \mathbb{F}^n$, and let $\alpha \in \mathbb{F}$. Then,

$$\det(A + \alpha x x^*) = \det A.$$

(Proof: Use Fact 2.13.2 and Fact 3.4.6.)

Fact 3.4.8. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is Hermitian.

ii) $A^2 = A^*A$.

iii) $\operatorname{tr} A^2 = \operatorname{tr} A^*A$.

(Proof: Use the Schur decomposition Theorem 5.4.1. See [347].) (Problem: If $AA^*A = A^*A^2$, then does it follow that A is normal?)

Fact 3.4.9. Let $A \in \mathbb{R}^{n\times n}$ be skew symmetric, and let $\alpha > 0$. Then, $-A^2$ is nonnegative semidefinite, $\det A \geq 0$, and $\det(\alpha I + A) > 0$. If, in addition, n is odd, then $\det A = 0$.

Fact 3.4.10. Let $A \in \mathbb{F}^{n\times n}$ be skew Hermitian. If n is even, then $\det A \geq 0$. If n is odd, then $\det A$ is imaginary. (Proof: The first statement follows from Proposition 5.5.25.)

Fact 3.4.11. Let $x, y \in \mathbb{F}^n$ and define

$$A \triangleq \begin{bmatrix} x & y \end{bmatrix}.$$

Then,

$$xy^* - yx^* = AJ_2A^*.$$

Furthermore, $xy^* - yx^*$ is skew Hermitian and has rank 0 or 2.

Fact 3.4.12. Let $x, y \in \mathbb{F}^n$. Then, the following statements hold:

i) xy^{T} is idempotent if and only if either $xy^{\mathrm{T}} = 0$ or $x^{\mathrm{T}}y = 1$.

ii) xy^{T} is Hermitian if and only if there exists $\alpha \in \mathbb{R}$ such that either $y = \alpha\overline{x}$ or $x = \alpha\overline{y}$.

Fact 3.4.13. Let $x, y \in \mathbb{F}^n$, and define $A \triangleq I - xy^{\mathrm{T}}$. Then, the following statements hold:

i) $\det A = 1 - x^{\mathrm{T}}y$.

ii) A is nonsingular if and only if $x^{\mathrm{T}}y \neq 1$.

iii) A is nonsingular if and only if A is elementary.

iv) $\operatorname{rank} A = n - 1$ if and only if $x^{\mathrm{T}}y = 1$.

v) A is Hermitian if and only if there exists $\alpha \in \mathbb{R}$ such that either $y = \alpha\overline{x}$ or $x = \alpha\overline{y}$.

vi) A is nonnegative semidefinite if and only if A is Hermitian and $x^{\mathrm{T}}y \leq 1$.

vii) A is positive definite if and only if A is Hermitian and $x^{\mathrm{T}}y < 1$.

viii) A is idempotent if and only if either $xy^{\mathrm{T}} = 0$ or $x^{\mathrm{T}}y = 1$.

ix) A is orthogonal if and only if either $x = 0$ or $y = \frac{1}{2}y^{\mathrm{T}}yx$.

x) A is involutory if and only if $x^{\mathrm{T}}y = 2$.

xi) A is a projector if and only if either $y = 0$ or $x = x^*xy$.

xii) A is a reflector if and only if either $y = 0$ or $2x = x^*xy$.

xiii) A is an elementary projector if and only if $x \neq 0$ and $y = (x^*x)^{-1}x$.

xiv) A is an elementary reflector if and only if $x \neq 0$ and $y = 2(x^*x)^{-1}x$.

(Remark: See Fact 3.5.9.)

Fact 3.4.14. Let $x, y \in \mathbb{F}^{n\times n}$ satisfy $x^{\mathrm{T}}y \neq 1$. Then, $I - xy^{\mathrm{T}}$ is nonsingular and

$$\left(I - xy^{\mathrm{T}}\right)^{-1} = I - \frac{1}{x^{\mathrm{T}}y - 1}xy^{\mathrm{T}}.$$

(Remark: The inverse of an elementary matrix is an elementary matrix.)

Fact 3.4.15. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, $\det A$ is real.

Fact 3.4.16. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then,

$$(\operatorname{tr} A)^2 \le (\operatorname{rank} A) \operatorname{tr} A^2.$$

Furthermore, equality holds if and only if there exists $\alpha \in \mathbb{R}$ such that $A^2 = \alpha A$. (Remark: See Fact 5.9.27.)

Fact 3.4.17. Let $A \in \mathbb{R}^{n\times n}$, and assume that A is skew symmetric. Then, $\operatorname{tr} A = 0$. If, in addition, $B \in \mathbb{R}^{n\times n}$ is symmetric, then $\operatorname{tr} AB = 0$.

Fact 3.4.18. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is skew Hermitian. Then, $\operatorname{Re}\operatorname{tr} A = 0$. If, in addition, $B \in \mathbb{F}^{n\times n}$ is Hermitian, then $\operatorname{Re}\operatorname{tr} AB = 0$.

Fact 3.4.19. Let $A \in \mathbb{F}^{n\times m}$. Then, A^*A is nonnegative semidefinite. Furthermore, A^*A is positive definite if and only if A is left invertible. In this case, A^{L} defined by

$$A^{\mathrm{L}} \triangleq (A^*A)^{-1}A^*$$

is a left inverse of A. (Remark: See Fact 2.13.23, Fact 3.4.20, and Fact 3.5.3.)

Fact 3.4.20. Let $A \in \mathbb{F}^{n\times m}$. Then, AA^* is nonnegative semidefinite. Furthermore, AA^* is positive definite if and only if A is right invertible. In this case, A^{R} defined by

$$A^{\mathrm{R}} \triangleq A^*(AA^*)^{-1}$$

is a right inverse of A. (Remark: See Fact 2.13.23, Fact 3.5.3, and Fact 3.4.19.)

Fact 3.4.21. Let $A \in \mathbb{F}^{n\times m}$. Then, A^*A, AA^*, $A + A^*$, and $\left[\begin{smallmatrix} 0 & A^* \\ A & 0 \end{smallmatrix}\right]$ are Hermitian, and $\left[\begin{smallmatrix} 0 & A^* \\ -A & 0 \end{smallmatrix}\right]$ and $A - A^*$ are skew Hermitian.

Fact 3.4.22. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist a unique Hermitian matrix $B \in \mathbb{F}^{n\times n}$ and a unique skew-Hermitian matrix $C \in \mathbb{F}^{n\times n}$ such that $A = B + C$. Specifically, if $A = \hat{B} + \jmath\hat{C}$, where $\hat{B}, \hat{C} \in \mathbb{R}^{n\times n}$, then $\hat{B}$ and $\hat{C}$ are given by

$$B = \tfrac{1}{2}(A + A^*) = \tfrac{1}{2}\left(\hat{B} + \hat{B}^{\mathrm{T}}\right) + \jmath\tfrac{1}{2}\left(\hat{C} - \hat{C}^{\mathrm{T}}\right)$$

and

$$C = \tfrac{1}{2}(A - A^*) = \tfrac{1}{2}\Big(\hat{B} - \hat{B}^{\mathrm{T}}\Big) + \jmath\tfrac{1}{2}\Big(\hat{C} + \hat{C}^{\mathrm{T}}\Big).$$

Furthermore, A is normal if and only if $BC = CB$. (Remark: See Fact 11.10.7.)

Fact 3.4.23. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist unique Hermitian matrices $B, C \in \mathbb{C}^{n\times n}$ such that $A = B + \jmath C$. Specifically, if $A = \hat{B} + \jmath\hat{C}$, where $\hat{B}, \hat{C} \in \mathbb{R}^{n\times n}$, then $\hat{B}$ and $\hat{C}$ are given by

$$B = \tfrac{1}{2}(A + A^*) = \tfrac{1}{2}\Big(\hat{B} + \hat{B}^{\mathrm{T}}\Big) + \jmath\tfrac{1}{2}\Big(\hat{C} - \hat{C}^{\mathrm{T}}\Big)$$

and

$$C = \tfrac{1}{2\jmath}(A - A^*) = \tfrac{1}{2}\Big(\hat{C} + \hat{C}^{\mathrm{T}}\Big) - \jmath\tfrac{1}{2}\Big(\hat{B} - \hat{B}^{\mathrm{T}}\Big).$$

Furthermore, A is normal if and only if $BC = CB$. (Remark: This result is the *Cartesian decomposition.*)

Fact 3.4.24. Let $x, y, z, w \in \mathbb{R}^3$, and define

$$C(x) \triangleq \begin{bmatrix} 0 & -x_{(3)} & x_{(2)} \\ x_{(3)} & 0 & -x_{(1)} \\ -x_{(2)} & x_{(1)} & 0 \end{bmatrix}.$$

Then, the following statements hold:

i) $x \times y = C(x)y$.

ii) $x \times x = C(x)x = 0$.

iii) $x \times y = -(y \times x) = C(x)y = -C(y)x$.

iv) If $x \times y \neq 0$, then $\mathcal{N}\big[(x \times y)^{\mathrm{T}}\big] = \mathcal{R}\big(\begin{bmatrix} x & y \end{bmatrix}\big)$.

v) $C(x \times y) = C[C(x)y] = [C(x), C(y)] = yx^{\mathrm{T}} - xy^{\mathrm{T}}$.

vi) $C^2(x) = xx^{\mathrm{T}} - \big(x^{\mathrm{T}}x\big)I$.

vii) If $x^{\mathrm{T}}x = 1$, then $C^3(x) = -C(x)$.

viii) If $x^{\mathrm{T}}x = 1$, then $C[(x \times y) \times x] = \big(I - xx^{\mathrm{T}}\big)y$.

ix) $\det\begin{bmatrix} x & y & z \end{bmatrix} = (x \times y)^{\mathrm{T}}z = x^{\mathrm{T}}(y \times z)$.

x) $(x \times y)^{\mathrm{T}}(x \times y) = \det\begin{bmatrix} x & y & x \times y \end{bmatrix}$.

xi) $(x \times y) \times z = \big(x^{\mathrm{T}}z\big)y - \big(y^{\mathrm{T}}z\big)x$.

xii) $x \times (y \times z) = \big(x^{\mathrm{T}}z\big)y - \big(x^{\mathrm{T}}y\big)z$.

xiii) $(x \times y)^{\mathrm{T}}(x \times y) = x^{\mathrm{T}}xy^{\mathrm{T}}y - \big(x^{\mathrm{T}}y\big)^2$.

xiv) $\sqrt{(x \times y)^{\mathrm{T}}(x \times y)} = \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}\,\sin\theta$, where θ is the angle between x

and y.

xv) $(x\times y)^{\rm T}(z\times w) = x^{\rm T}zy^{\rm T}w - x^{\rm T}wy^{\rm T}z = \det\begin{bmatrix} x^{\rm T}z & x^{\rm T}w \\ y^{\rm T}z & y^{\rm T}w\end{bmatrix}$.

xvi) $(x\times y)\times(z\times w) = x^{\rm T}(y\times w)z - x^{\rm T}(y\times z)w = x^{\rm T}(z\times w)y - y^{\rm T}(z\times w)x$.

xvii) $x\times[y\times(z\times w)] = (y^{\rm T}w)(x\times z) - (y^{\rm T}z)(x\times w)$.

xviii) $x\times[y\times(y\times x)] = y\times[x\times(y\times x)] = (y^{\rm T}x)(x\times y)$.

xix) If $A\in\mathbb{R}^{3\times3}$, then $A^{\rm T}(Ax\times Ay) = (\det A)(x\times y)$.

xx) If $A\in\mathbb{R}^{3\times3}$ is orthogonal and $\det A = 1$, then $A(x\times y) = Ax\times Ay$.

(Proof: Using *ix)*, $e_i^{\rm T}A^{\rm T}(Ax\times Ay) = \det\begin{bmatrix} Ax & Ay & Ae_i\end{bmatrix} = (\det A)e_i^{\rm T}(x\times y)$ for all $i = 1,2,3$, which proves *xvii)*.) (Remark: See [177, 447, 508, 539].)

Fact 3.4.25. Let $A, B\in\mathbb{R}^3$ be skew symmetric. Then,

$$\operatorname{tr} AB^3 = \tfrac{1}{2}(\operatorname{tr} AB)\big(\operatorname{tr} B^2\big)$$

and

$$\operatorname{tr} A^3B^3 = \tfrac{1}{4}\big(\operatorname{tr} A^2\big)(\operatorname{tr} AB)\big(\operatorname{tr} B^2\big) + \tfrac{1}{3}\big(\operatorname{tr} A^3\big)\big(\operatorname{tr} B^3\big).$$

(Proof: See [37].)

Fact 3.4.26. Let $A, B\in\mathbb{F}^{n\times n}$. If either A and B are Hermitian or A and B are skew Hermitian, then $[A,B]$ is skew Hermitian. Furthermore, if A is Hermitian and B is skew Hermitian, or vice versa, then $[A,B]$ is Hermitian.

Fact 3.4.27. Let $A\in\mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) $\operatorname{tr} A = 0$

ii) There exist $B, C\in\mathbb{F}^{n\times n}$ such that A is Hermitian, $\operatorname{tr} B = 0$, and $A = [B,C]$.

(Proof: See [221] and Fact 5.7.18. If all of the diagonal entries of A are zero, then let $B \triangleq \operatorname{diag}(1,\ldots,n)$, $C_{(i,i)} \triangleq 0$, and, for $i\ne j$, $C_{(i,j)} \triangleq A_{(i,j)}/(i-j)$. See [626, p. 110]. See also [466, p. 172].)

Fact 3.4.28. Let $A\in\mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is Hermitian and $\operatorname{tr} A = 0$.

ii) There exists a nonsingular matrix $B\in\mathbb{F}^{n\times n}$ such that $A = [B, B^*]$.

iii) There exist a Hermitian matrix $B\in\mathbb{F}^{n\times n}$ and a skew-Hermitian matrix $C\in\mathbb{F}^{n\times n}$ such that $A = [B,C]$

iv) There exist a skew-Hermitian matrix $B \in \mathbb{F}^{n\times n}$ and a Hermitian matrix $C \in \mathbb{F}^{n\times n}$ such that $A = [B, C]$

(Proof: See [542] and [221].)

Fact 3.4.29. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is skew Hermitian and $\operatorname{tr} A = 0$.

ii) There exists a nonsingular matrix $B \in \mathbb{F}^{n\times n}$ such that $A = [\jmath B, B^*]$.

iii) If $A \in \mathbb{C}^{n\times n}$ is skew Hermitian, then there exist Hermitian matrices $B, C \in \mathbb{F}^{n\times n}$ such that $A = [B, C]$.

(Proof: See [221] or use Fact 3.4.28.)

Fact 3.4.30. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is skew symmetric. Then, there exist symmetric matrices $B, C \in \mathbb{F}^{n\times n}$ such that $A = [B, C]$. (Proof: Use Fact 5.13.22. See [466, pp. 83, 89].) (Remark: All matrices can be complex.)

Fact 3.4.31. Let $A \in \mathbb{F}^{n\times n}$, and assume that $[A, [A, A^*]] = 0$. Then, A is normal. (Remark: See [626, p. 32].)

Fact 3.4.32. Let $A \in \mathbb{F}^{n\times n}$ and $k \in \mathbb{P}$. If A is (normal, Hermitian, unitary, involutory, nonnegative semidefinite, positive definite, idempotent, nilpotent), then so is A^k. If A is (skew Hermitian, skew involutory), then so is A^{2k+1}. If A is Hermitian, then A^{2k} is nonnegative semidefinite. If A is tripotent, then so is A^{3k}.

Fact 3.4.33. Let $x, y \in \mathbb{F}^n$, and assume that $x \neq 0$. Then, there exists a Hermitian matrix $A \in \mathbb{F}^{n\times n}$ such that $y = Ax$ if and only if x^*y is real. One such matrix is

$$A = (x^*x)^{-1}[yx^* + xy^* - x^*yI]\,.$$

(Remark: See Fact 2.11.12.)

Fact 3.4.34. Let $x, y \in \mathbb{F}^n$, and assume that $x \neq 0$. Then, there exists a positive-definite matrix $A \in \mathbb{F}^{n\times n}$ such that $y = Ax$ if and only if x^*y is real and positive. One such matrix is

$$A = I + (x^*y)^{-1}yy^* - (x^*x)^{-1}xx^*.$$

(Proof: To show that A is positive definite, note that the elementary projector $I - (x^*x)^{-1}xx^*$ is nonnegative semidefinite and $\operatorname{rank}[I - (x^*x)^{-1}xx^*] = n - 1$. Since $(x^*y)^{-1}yy^*$ is nonnegative semidefinite, it follows that $\mathcal{N}(A) \subseteq \mathcal{N}[I - (x^*x)^{-1}xx^*]$. Next, since $x^*y > 0$, it follows that $y^*x \neq 0$ and $y \neq 0$, and thus $x \notin \mathcal{N}(A)$. Consequently, $\mathcal{N}(A) \subset \mathcal{N}[I - (x^*x)^{-1}xx^*]$ (note proper inclusion), and thus $\operatorname{def} A < 1$. Hence, A is nonsingular.)

Fact 3.4.35. Let $x, y \in \mathbb{F}^n$. Then, there exists a skew-Hermitian matrix $A \in \mathbb{F}^{n\times n}$ such that $y = Ax$ if and only if either $y = 0$ or $x \neq 0$ and $x^*y = 0$. If $x \neq 0$ and $x^*y = 0$, then one such matrix is

$$A = (x^*x)^{-1}(yx^* - xy^*).$$

(Proof: See [376].)

Fact 3.4.36. Let $A \in \mathbb{R}^{n\times n}$ be positive definite. Then,

$$\{x \in \mathbb{R}^n\colon\ x^{\mathrm{T}}Ax \le 1\}$$

is an ellipsoid.

Fact 3.4.37. Let $x, y, z \in \mathbb{F}^n$ satisfy $x^*x = y^*y = z^*z = 1$. Then,

$$\sqrt{1 - |x^*y|^2} \le \sqrt{1 - |x^*z|^2} + \sqrt{1 - |y^*z|^2}.$$

Furthermore, if $A, B \in \mathbb{F}^{n\times n}$ are unitary, then

$$\sqrt{1 - \left|\tfrac{1}{n}\operatorname{tr} AB\right|^2} \le \sqrt{1 - \left|\tfrac{1}{n}\operatorname{tr} A\right|^2} + \sqrt{1 - \left|\tfrac{1}{n}\operatorname{tr} B\right|^2}.$$

(Proof: See [580].)

3.5 Facts on Projectors and Idempotent Matrices

Fact 3.5.1. Let $A \in \mathbb{F}^{n\times n}$ be a projector, and let $x \in \mathbb{F}^n$. Then, $x \in \mathcal{R}(A)$ if and only if $x = Ax$.

Fact 3.5.2. Let $A, B \in \mathbb{F}^{n\times n}$ be projectors, and assume that $\mathcal{R}(A) = \mathcal{R}(B)$. Then, $A = B$.

Fact 3.5.3. Let $A \in \mathbb{F}^{n\times m}$. If $\operatorname{rank} A = m$, then $B \triangleq A(A^*A)^{-1}A^*$ is a projector and $\operatorname{rank} B = m$. If $\operatorname{rank} A = n$, then $B \triangleq A^*(AA^*)^{-1}A$ is a projector and $\operatorname{rank} B = n$. (Remark: See Fact 2.13.23, Fact 3.4.19, and Fact 3.4.20.)

Fact 3.5.4. Let $A \in \mathbb{F}^{n\times n}$. Then, A is a projector if and only if $A = A^*A$.

Fact 3.5.5. Let $A \in \mathbb{F}^{n\times m}$, and assume that A is a projector. Then, A is nonnegative semidefinite.

Fact 3.5.6. Let $x \in \mathbb{F}^n$ be nonzero and define the elementary projector

$A \triangleq I - (x^*x)^{-1}xx^*$. Then, the following statements hold:

i) $\operatorname{rank} A = n - 1$.

ii) $\mathcal{N}(A) = \operatorname{span}\{x\}$.

iii) $\mathcal{R}(A) = \{x\}^\perp$.

iv) $2A - I$ is the elementary reflector $I - 2(x^*x)^{-1}xx^*$.

(Remark: If $y \in \mathbb{F}^n$, then Ay is the *projection* of y on $\{x\}^\perp$.

Fact 3.5.7. Let $A \in \mathbb{F}^{n\times n}$. Then, A is an elementary reflector if and only if A is a reflector and $\operatorname{tr} A = n - 2$. Furthermore, A is an elementary projector if and only if A is a projector and $\operatorname{tr} A = n - 1$. (Proof: See Proposition 5.5.25.)

Fact 3.5.8. Let $n > 1$, and let $\mathcal{S} \subset \mathbb{F}^n$ be a hyperplane. Then, there exists a unique elementary projector $A \in \mathbb{F}^{n\times n}$ such that $\mathcal{R}(A) = \mathcal{S}$ and $\mathcal{N}(A) = \mathcal{S}^\perp$. Furthermore, if $x \in \mathbb{F}^n$ is nonzero and $\mathcal{S} \triangleq \{x\}^\perp$, then $A = I - (x^*x)^{-1}xx^*$. (Remark: See Proposition 5.5.4.)

Fact 3.5.9. Let $A \in \mathbb{F}^{n\times n}$. Then, A is a projector and $\operatorname{rank} A = n - 1$ if and only if there exists nonzero $x \in \mathcal{N}(A)$ such that

$$A = I - (x^*x)^{-1}xx^*.$$

In this case, it follows that, for all $y \in \mathbb{F}^n$,

$$y^*y - y^*Ay = \frac{(y^*x)^2}{x^*x}.$$

Furthermore, for $y \in \mathbb{F}^n$, the following statements are equivalent:

i) $y^*Ay = y^*y$.

ii) $y^*x = 0$.

iii) $Ay = y$.

(Remark: See Fact 3.4.13.)

Fact 3.5.10. Let $A \in \mathbb{F}^{n\times n}$ be a projector, and let $x \in \mathbb{F}^n$. Then,

$$x^*Ax \le x^*x.$$

Furthermore, the following statements are equivalent:

i) $x^*Ax = x^*x$.

ii) $Ax = x$.

iii) $x \in \mathcal{R}(A)$.

Fact 3.5.11. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is idempotent. Then, A is a projector if and only if, for all $x \in \mathbb{F}^n$, $x^*Ax \le x^*x$. (Proof: See [466, p. 105].)

Fact 3.5.12. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\mathcal{N}(A) \subseteq \mathcal{R}(I - A)$$

and

$$\mathcal{R}(A) \subseteq \mathcal{N}(I - A).$$

Furthermore, the following statements are equivalent:

i) A is idempotent.

ii) $\mathcal{N}(A) = \mathcal{R}(I - A)$.

iii) $\mathcal{R}(A) = \mathcal{N}(I - A)$.

(Proof: See [269, p. 146].)

Fact 3.5.13. Let $A \in \mathbb{F}^{n\times n}$. Then, A is idempotent and $\operatorname{rank} A = 1$ if and only if there exist $x, y \in \mathbb{F}^n$ such that $y^{\mathrm{T}}x = 1$ and $A = xy^{\mathrm{T}}$.

Fact 3.5.14. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is idempotent. Then, A^{T}, $\overline{A}$, and A^* are idempotent.

Fact 3.5.15. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be complementary subspaces, and let $A \in \mathbb{F}^{n\times n}$ be the idempotent matrix associated with $\mathcal{S}_1, \mathcal{S}_2$. Then, A^{T} is the idempotent matrix associated with $\mathcal{S}_2^{\perp}, \mathcal{S}_1^{\perp}$. (Remark: See Fact 2.9.11.)

Fact 3.5.16. Let $A \in \mathbb{F}^{n\times n}$. Then, A is idempotent if and only if $\operatorname{rank} A + \operatorname{rank}(I - A) = n$.

Fact 3.5.17. Let $A, B \in \mathbb{R}^{n\times n}$ be idempotent and define $A_\perp \triangleq I - A$ and $B_\perp \triangleq I - B$. Then, the following identities hold:

i) $(A - B)^2 + (A_\perp - B)^2 = I$.

ii) $[A, B] = [B, A_\perp] = [B_\perp, A] = [A_\perp, B_\perp]$.

iii) $A - B = AB_\perp - A_\perp B$.

iv) $AB_\perp + BA_\perp = AB_\perp A + A_\perp BA_\perp$.

v) $A[A, B] = [A, B]A_\perp$.

vi) $B[A, B] = [A, B]B_\perp$.

(Proof: See [439].)

Fact 3.5.18. Let $A \in \mathbb{F}^{n\times n}$ and $\alpha \in \mathbb{F}$, where $\alpha \neq 0$. Then, the matrices

$$\begin{bmatrix} A & A^* \\ A^* & A \end{bmatrix}, \quad \begin{bmatrix} A & \alpha^{-1}A \\ \alpha(I-A) & I-A \end{bmatrix}, \quad \begin{bmatrix} A & \alpha^{-1}A \\ -\alpha A & -A \end{bmatrix}$$

are, respectively, normal, idempotent, and nilpotent.

Fact 3.5.19. Let $A, B \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) Assume that $A^3 = -A$. Then, $B \triangleq I + A + A^2$ satisfies $B^4 = I$, $B^{-1} = I - A + A^2$, $B^3 - B^2 + B - I = 0$, and $A = \frac{1}{2}(B - B^3)$. Furthermore, $I + A^2$ is idempotent.

ii) Assume that $B^4 = I$. Then, $A \triangleq \frac{1}{2}(B - B^{-1})$ satisfies $B^3 = -B$. Furthermore, $\frac{1}{4}(I + B + B^2 + B^3)$ is idempotent.

iii) Assume that $B^3 - B^2 + B - I = 0$. Then, $A \triangleq \frac{1}{2}(B - B^3)$ satisfies $A^3 = -A$ and $B = I + A + A^2$.

(Remark: The geometrical interpretation of these results is discussed in [197].)

Fact 3.5.20. Let $A \in \mathbb{F}^{n\times m}$. If $A^{\mathrm{L}} \in \mathbb{F}^{m\times n}$ is a left inverse of A, then AA^{L} is idempotent and $\operatorname{rank} A^{\mathrm{L}} = \operatorname{rank} A$. Furthermore, if $A^{\mathrm{R}} \in \mathbb{F}^{m\times n}$ is a right inverse of A, then $A^{\mathrm{R}}A$ is idempotent and $\operatorname{rank} A^{\mathrm{R}} = \operatorname{rank} A$.

Fact 3.5.21. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$, and assume that AB is nonsingular. Then, $B(AB)^{-1}A$ is idempotent.

Fact 3.5.22. Let $A, B \in \mathbb{F}^{n\times n}$ be idempotent. Then, $A + B$ is idempotent if and only if $AB = BA = 0$. (Proof: $AB + BA = 0$ implies $AB + ABA = ABA + BA = 0$, which implies that $AB - BA = 0$ and hence $AB = 0$. See [262, p. 250].)

Fact 3.5.23. If $A, B \in \mathbb{F}^{n\times n}$ are idempotent and $AB = 0$, then $A + B - BA$ is idempotent and $C \triangleq A - B$ is tripotent. Conversely, if $C \in \mathbb{F}^{n\times n}$ is tripotent, then $A \triangleq \frac{1}{2}(C^2 + C)$ and $B \triangleq \frac{1}{2}(C^2 - C)$ are idempotent and satisfy $C = A - B$ and $AB = BA = 0$. (Proof: See [407, p. 114].)

Fact 3.5.24. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and idempotent. Then, $A = I_n$.

Fact 3.5.25. Let $A \in \mathbb{F}^{n\times n}$ be idempotent. Then, so is $A_\perp \triangleq I - A$, and, furthermore, $AA_\perp = A_\perp A = 0$.

Fact 3.5.26. Let $A \in \mathbb{F}^{n\times n}$ be idempotent. Then,

$$\det(I + A) = 2^{\operatorname{tr} A}$$

and

$$(I+A)^{-1} = I - \tfrac{1}{2}A.$$

Fact 3.5.27. If $A \in \mathbb{F}^{n\times n}$ is idempotent, then $B \triangleq 2A - I$ is involutory, while if $B \in \mathbb{F}^{n\times n}$ is involutory, then $A \triangleq \frac{1}{2}(B+I)$ is idempotent. Furthermore, if $A \in \mathbb{F}^{n\times n}$ is a projector, then $B \triangleq 2A - I$ is a reflector, while if $B \in \mathbb{F}^{n\times n}$ is a reflector, then $A \triangleq \frac{1}{2}(B+I)$ is a projector.

Fact 3.5.28. Let $A \in \mathbb{F}^{n\times n}$, and assume that A satisfies two out of the three properties (Hermitian, idempotent, $A + A^* = 2AA^*$). Then, A satisfies the remaining property. (Proof: If A is idempotent and $2AA^* = A + A^*$, then $(2A-I)^{-1} = 2A - I = (2A^* - I)^{-1}$. Hence, A is Hermitian.) (Remark: These matrices are the projectors.) (Remark: The condition $A + A^* = 2AA^*$ is considered in Fact 3.5.29.) (Remark: See Fact 3.7.1 and Fact 3.7.5.)

Fact 3.5.29. If $B \in \mathbb{F}^{n\times n}$ is unitary and skew Hermitian, then $A \triangleq \frac{1}{2}(B+I)$ satisfies

$$A + A^* = 2AA^*.$$

Conversely, if $A \in \mathbb{F}^{n\times n}$ satisfies this equation, then $B \triangleq 2A - I$ is unitary. (Remark: See Fact 3.5.28.) (Remark: This equation has normal solutions such that $B \triangleq 2A - I$ is not skew Hermitian, for example, $A = 1/3 + \jmath\sqrt{2}/3$.) (Problem: Characterize all normal and nonnormal solutions.)

3.6 Facts on Unitary Matrices

Fact 3.6.1. Let $A \in \mathbb{F}^{n\times n}$ be unitary. Then, the following statements hold:

i) $U = U^{-*}$.

ii) $U^{\mathrm{T}} = \overline{U}^{-1} = \overline{U}^{*}$.

iii) $\overline{U} = U^{-\mathrm{T}} = \overline{U}^{-*}$.

iv) $U^* = U^{-1}$.

Fact 3.6.2. Let $A \in \mathbb{F}^{n\times n}$ be unitary. Then,

$$-n \le \operatorname{Re}\operatorname{tr} A \le n,$$
$$-n \le \operatorname{Im}\operatorname{tr} A \le n,$$

and

$$|\operatorname{tr} A| \le n.$$

Fact 3.6.3. Let $x, y \in \mathbb{F}^n$, and let $A \in \mathbb{F}^{n\times n}$ be unitary. Then, $x^*y = 0$

if and only if $(Ax)^*Ay = 0$.

Fact 3.6.4. Let $A \in \mathbb{F}^{n\times m}$. If A is (left inner, right inner), then A is (left invertible, right invertible) and A^* is a (left inverse, right inverse).

Fact 3.6.5. Let $A \in \mathbb{R}^{n\times n}$ be a permutation matrix. Then, A is orthogonal.

Fact 3.6.6. Let $A \in \mathbb{C}^{n\times n}$ be unitary. Then, $|\det A| = 1$.

Fact 3.6.7. Let $M \triangleq \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] \in \mathbb{F}^{(n+m)\times(n+m)}$ be unitary. Then,

$$\det A = (\det M)\overline{\det D}.$$

(Proof: Let $\left[\begin{smallmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{smallmatrix}\right] \triangleq A^{-1}$ and take the determinant of $A\left[\begin{smallmatrix} I & \hat{B} \\ 0 & \hat{D} \end{smallmatrix}\right] = \left[\begin{smallmatrix} A & 0 \\ C & I \end{smallmatrix}\right]$. See [3] or [506].) (Remark: See Fact 2.13.34.)

Fact 3.6.8. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is Hermitian, skew Hermitian, or unitary. Then, A is normal.

Fact 3.6.9. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is block diagonal. Then, A is (normal, Hermitian, unitary) if and only if every diagonally located block has the same property.

Fact 3.6.10. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, A is normal if and only if $A^{-1}A^*$ is unitary.

Fact 3.6.11. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and assume that A is (normal, Hermitian, skew Hermitian, unitary). Then, so is A^{-1}.

Fact 3.6.12. Let $A, B \in \mathbb{R}^{n\times n}$. Then, $A + \jmath B$ is (Hermitian, skew Hermitian, unitary) if and only if $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$ is (symmetric, skew symmetric, orthogonal).

Fact 3.6.13. Let $A \in \mathbb{F}^{n\times n}$ be semicontractive. Then, $B \in \mathbb{F}^{2n\times 2n}$ defined by

$$B \triangleq \begin{bmatrix} A & (I - AA^*)^{1/2} \\ (I - A^*A)^{1/2} & -A^* \end{bmatrix}$$

is unitary. (Remark: See [216, p. 180].)

Fact 3.6.14. Let $\theta \in \mathbb{R}$, and define the orthogonal matrix

$$A(\theta) \triangleq \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Now, let $\theta_1, \theta_2 \in \mathbb{R}$. Then,

$$A(\theta_1)A(\theta_2) = A(\theta_1 + \theta_2).$$

Consequently,

$$\begin{aligned}\cos(\theta_1 + \theta_2) &= (\cos\theta_1)\cos\theta_2 - (\sin\theta_1)\sin\theta_2,\\ \sin(\theta_1 + \theta_2) &= (\cos\theta_1)\sin\theta_2 + (\sin\theta_1)\cos\theta_2.\end{aligned}$$

Furthermore,

$$\mathrm{SO}(2) = \{A(\theta)\colon\ \theta \in \mathbb{R}\}.$$

(Remark: See Proposition 3.2.7 and Fact 11.9.3.)

Fact 3.6.15. Let $x, y, z \in \mathbb{R}^2$. If x is rotated according to the right hand rule through an angle $\theta \in \mathbb{R}$ about y, then the resulting vector $\hat{x} \in \mathbb{R}^2$ is given by

$$\hat{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} x + \begin{bmatrix} y_{(1)}(1-\cos\theta) + y_{(2)}\sin\theta \\ y_{(2)}(1-\cos\theta) + y_{(1)}\sin\theta \end{bmatrix}.$$

If x is reflected across the line passing through 0 and z and parallel to the line passing through 0 and y, then the resulting vector $\hat{x} \in \mathbb{R}^2$ is given by

$$\hat{x} = \begin{bmatrix} y_{(1)}^2 - y_{(2)}^2 & 2y_{(1)}y_{(2)} \\ 2y_{(1)}y_{(2)} & y_{(2)}^2 - y_{(1)}^2 \end{bmatrix} x + \begin{bmatrix} -z_{(1)}\left(y_{(1)}^2 - y_{(2)}^2 - 1\right) - 2z_{(2)}y_{(1)}y_{(2)} \\ -z_{(2)}\left(y_{(1)}^2 - y_{(2)}^2 - 1\right) - 2z_{(1)}y_{(1)}y_{(2)} \end{bmatrix}.$$

(Remark: These *affine planar transformations* are used in computer graphics. See [210, 464].)

Fact 3.6.16. Let $x, y \in \mathbb{R}^3$, and assume that $y^\mathrm{T}y = 1$. If x is rotated according to the right hand rule through an angle $\theta \in \mathbb{R}$ about the line passing through 0 and y, then the resulting vector $\hat{x} \in \mathbb{R}^3$ is given by

$$\hat{x} = x + (\sin\theta)(y \times x) + (1 - \cos\theta)[y \times (y \times x)].$$

(Proof: See [10].)

Fact 3.6.17. Let $x, y \in \mathbb{R}^n$. Then, there exists an orthogonal matrix $A \in \mathbb{R}^{n\times n}$ such that $y = Ax$ if and only if $x^\mathrm{T}x = y^\mathrm{T}y$. (Remark: One such matrix is given by a product of n plane rotations given by Fact 5.13.13. Another is given by the product of elementary reflectors given by Fact 5.13.12. See Fact 11.9.9 and Fact 3.7.3.) (Problem: Extend this result to $\mathbb{C}$.)

Fact 3.6.18. Let $A \in \mathbb{F}^{n\times n}$ be unitary, and let $x \in \mathbb{F}^n$ be such that $x^*x = 1$ and $Ax = -x$. Then, the following statements hold:

i) $\det(A + I) = 0$.

ii) $A + 2xx^*$ is unitary.

iii) $A = (A + 2xx^*)(I_n - 2xx^*) = (I_n - 2xx^*)(A + 2xx^*)$.

iv) $\det(A + 2xx^*) = -\det A$.

Fact 3.6.19. Let $A \in \mathbb{R}^{3\times3}$. Then, A is an orthogonal matrix if and only if there exist real numbers a, b, c, d, not all zero, such that

$$A = \frac{\pm1}{\alpha}\begin{bmatrix} a^2+b^2-c^2-d^2 & 2(bc+da) & 2(bd-ca) \\ 2(bc-da) & a^2-b^2+c^2-d^2 & 2(cd+ba) \\ 2(bd+ca) & 2(cd-ba) & a^2-b^2-c^2+d^2 \end{bmatrix},$$

where $\alpha \triangleq a^2 + b^2 + c^2 + d^2$. (Remark: This result is due to Rodrigues.)

Fact 3.6.20. Let $A \in \mathbb{R}^{n\times n}$, and assume that A is orthogonal. Then, either $\det A = 1$ or $\det A = -1$.

Fact 3.6.21. Let $A \in \mathbb{F}^{n\times n}$ and assume that A is involutory. Then, either $\det A = 1$ or $\det A = -1$.

Fact 3.6.22. Let $A \in \mathbb{F}^{n\times n}$ be unitary. Then, $\frac{1}{\sqrt{2}}\begin{bmatrix} A & -A \\ A & A \end{bmatrix}$ is also unitary.

Fact 3.6.23. If $A \in \mathbb{F}^{n\times n}$ is Hermitian, then $I + \jmath A$ is nonsingular and $B \triangleq (A - \jmath I)(A + \jmath I)^{-1}$ is unitary and $B - I$ is nonsingular. Conversely, if $B \in \mathbb{F}^{n\times n}$ is unitary and $B - I$ is nonsingular, then $A \triangleq \jmath(I + B)(I - B)^{-1}$ is Hermitian. (Proof: See [216, pp. 168, 169].) (Remark: $(A - \jmath I)(A + \jmath I)^{-1}$ is the *Cayley transform* of A. See Fact 3.6.24, Fact 3.6.25, Fact 3.9.8, and Fact 8.7.18, and Fact 11.15.9.) (Remark: The linear fractional transformation $f(s) \triangleq (s - \jmath)(s + \jmath)$ maps the upper half plane of $\mathbb{C}$ onto the unit disk in $\mathbb{C}$, and the real line onto the unit circle in $\mathbb{C}$.)

Fact 3.6.24. If $A \in \mathbb{F}^{n\times n}$ is skew Hermitian, then $I + A$ is nonsingular, $B \triangleq (I - A)(I + A)^{-1} = (I + A)^{-1}(I - A)$ is unitary, and $|\det B| = 1$. Conversely, if $B \in \mathbb{F}^{n\times n}$ is unitary and $I + B$ is nonsingular, then $A \triangleq (I + B)^{-1}(I - B)$ is skew Hermitian. Furthermore, if B is unitary, then there exist $\lambda \in \mathbb{C}$ and a skew-Hermitian matrix $A \in \mathbb{F}^{n\times n}$ such that $|\lambda| = 1$ and $B \triangleq (I - A)(I + A)^{-1}$. (Proof: See [289, p. 440] and [216, p. 184].)

Fact 3.6.25. If $A \in \mathbb{R}^{n\times n}$ is skew symmetric, then $I + A$ is nonsingular, $B \triangleq (I - A)(I + A)^{-1} = (I + A)^{-1}(I - A)$ is orthogonal, and $I + B$ is nonsingular. Equivalently, if $A \in \mathbb{R}^{n\times n}$ is skew symmetric, then there exists an orthogonal matrix $B \in \mathbb{R}^{n\times n}$ such that $I + B$ is nonsingular and $A = (I + B)^{-1}(I - B)$. Conversely, if $B \in \mathbb{R}^{n\times n}$ is orthogonal and $I + B$ is nonsingular, then $\det B = 1$ and $A \triangleq (I + B)^{-1}(I - B)$ is skew symmetric. Equivalently, if $B \in \mathbb{R}^{n\times n}$ is orthogonal and $I + B$ is nonsingular, then there exists a skew-symmetric matrix $A \in \mathbb{R}^{n\times n}$ such that $B = (I - A)(I + A)^{-1}$.

Fact 3.6.26. Let $A \in \mathbb{R}^{n\times n}$ be orthogonal. Then, there exist a skew-symmetric matrix $B \in \mathbb{R}^{n\times n}$ and a diagonal matrix $C \in \mathbb{R}^{n\times n}$, each of whose diagonal entries is either 1 or -1, such that

$$A = C(I - B)(I + B)^{-1}.$$

(Proof: See [466, p. 101].) (Remark: This result is due to Hsu.)

3.7 Facts on Reflectors

Fact 3.7.1. Let $A \in \mathbb{F}^{n\times n}$, and assume that A satisfies two out of the three properties (Hermitian, unitary, involutory). Then, A also satisfies the remaining property. (Remark: These matrices are the reflectors.) (Remark: See Fact 3.5.28 and Fact 3.7.5.)

Fact 3.7.2. Let $x \in \mathbb{F}^n$ be nonzero and define the elementary reflector $A \triangleq I - 2(x^*x)^{-1}xx^*$. Then, the following statements hold:

i) $\det A = -1$.

ii) If $y \in \mathbb{F}^n$, then Ay is the reflection of y across $\{x\}^\perp$.

iii) $Ax = -x$.

iv) $\frac{1}{2}(A + I)$ is the elementary projector $I - (x^*x)^{-1}xx^*$.

Fact 3.7.3. Let $x, y \in \mathbb{F}^n$. Then, there exists a unique elementary reflector $A \in \mathbb{F}^{n\times n}$ such that $y = Ax$ if and only if x^*y is real and $x^*x = y^*y$. If $x \neq y$, then A is given by

$$A = I - 2[(x - y)^*(x - y)]^{-1}(x - y)(x - y)^*.$$

(Remark: This result is the *reflection theorem*. See [229, pp. 16–18] and [484, p. 357]. See Fact 3.6.17 and Fact 11.9.9.)

Fact 3.7.4. Let $n > 1$, and let $\mathcal{S} \subset \mathbb{F}^n$ be a hyperplane. Then, there exists a unique elementary reflector $A \in \mathbb{F}^{n\times n}$ such that, for all $y = y_1 + y_2 \in \mathbb{F}^n$, where $y_1 \in \mathcal{S}$ and $y_2 = \mathcal{S}^\perp$, it follows that $Ay = y_1 - y_2$. Furthermore, if $\mathcal{S} = \{x\}^\perp$, then $A = I - 2(x^*x)^{-1}xx^*$.

Fact 3.7.5. Let $A \in \mathbb{F}^{n\times n}$, and assume that A satisfies two out of the three properties (skew Hermitian, unitary, skew involutory). Then, A also satisfies the remaining property. In particular, J_n satisfies all three properties. In addition, A^2 is a reflector. (Problem: Does every reflector have a skew-Hermitian, unitary square root?) (Remark: See Fact 3.5.28 and Fact 3.7.1.)

Fact 3.7.6. Let $A \in \mathbb{F}^{n\times n}$. Then, A is a reflector if and only if $A = AA^* + A^* - I$. (Proof: This condition is equivalent to $A = \frac{1}{2}(A+I)(A^*+I) - I$.)

3.8 Facts on Nilpotent Matrices

Fact 3.8.1. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and B are upper triangular. Then,

$$[A,B]^{n-1} = 0.$$

Hence, $[A,B]$ is nilpotent. (Remark: See [211, 212].)

Fact 3.8.2. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $[A,[A,B]] = 0$. Then, $[A,B]$ is nilpotent. (Remark: This result is due to Jacobson. See [207] or [287, p. 98].)

Fact 3.8.3. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $\left[A, B^2\right] = B$. Then, B is nilpotent. (Proof: See [493].)

Fact 3.8.4. Let $A \in \mathbb{R}^{n\times n}$. Then, $\operatorname{rank} A^k$ is a nonincreasing function of $k \in \mathbb{P}$. Furthermore, if there exists $k \in \{1,\ldots,n\}$ such that $\operatorname{rank} A^{k+1} = \operatorname{rank} A^k$, then $\operatorname{rank} A^l = \operatorname{rank} A^k$ for all $l \geq k$. Finally, if A is nilpotent and $A^l \neq 0$, then $\operatorname{rank} A^{k+1} < \operatorname{rank} A^k$ for all $k = 1,\ldots,l$.

Fact 3.8.5. Let $n \in \mathbb{P}$ and $k \in \{0,\ldots,n\}$. Then, $\operatorname{rank} N_n^k = n - k$.

Fact 3.8.6. Let $A \in \mathbb{F}^{n\times n}$. Then, A is nilpotent and $\operatorname{rank} A = 1$ if and only if there exist nonzero $x, y \in \mathbb{F}^n$ such that $y^{\mathrm{T}}x = 0$ and $A = xy^{\mathrm{T}}$.

Fact 3.8.7. Let $A \in \mathbb{R}^{n\times n}$ be nilpotent and assume that $A^k = 0$, where $k \in \mathbb{P}$. Then,

$$\det(I - A) = 1$$

and

$$(I - A)^{-1} = \sum_{i=0}^{k-1} A^i.$$

Fact 3.8.8. Let $\lambda \in \mathbb{F}$ and $n, k \in \mathbb{P}$. Then,

$$(\lambda I_n + N_n)^k = \begin{cases} \lambda^k I_n + \binom{k}{1}\lambda^{k-1}N_n + \cdots + \binom{k}{k}N_n^k, & k < n-1, \\ \lambda^k I_n + \binom{k}{1}\lambda^{k-1}N_n + \cdots + \binom{k}{n-1}\lambda^{k-n+1}N_n^{n-1}, & k \geq n-1, \end{cases}$$

that is, for $k \geq n-1$,

$$\begin{bmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}^k = \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{n-2}\lambda^{k-n+1} & \binom{k}{n-1}\lambda^{k-n+1} \\ 0 & \lambda^k & \ddots & \binom{k}{n-3}\lambda^{k-n+2} & \binom{k}{n-2}\lambda^{k-n+2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ 0 & 0 & \cdots & 0 & \lambda^k \end{bmatrix}.$$

Fact 3.8.9. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A is nilpotent and $AB = BA$. Then, $\det(A+B) = \det B$. (Proof: Use Fact 5.8.6.)

Fact 3.8.10. Let $A, B \in \mathbb{R}^{n\times n}$ be nilpotent and assume that $AB = BA$. Then, $A+B$ is nilpotent. (Proof: If $A^k = B^l = 0$, then $(A+B)^{k+l} = 0$.)

Fact 3.8.11. Let $A \in \mathbb{F}^{n\times n}$. Then, A is nilpotent if and only if, for all $k = 1, \ldots, n$, $\operatorname{tr} A^k = 0$. (Proof: See [466, p. 103].)

3.9 Facts on Hamiltonian and Symplectic Matrices

Fact 3.9.1. J_n is skew symmetric, skew involutory, and Hamiltonian, I_n is symplectic, and $\hat{I}_n$ is a symmetric permutation matrix.

Fact 3.9.2. Let $A \in \mathbb{R}^{2n\times 2n}$ be symplectic. Then, $\det A = 1$. Furthermore, $A \in \mathbb{R}^{2\times 2}$ is symplectic if and only if $\det A = 1$, that is, $\mathrm{SL}_{\mathbb{R}}(2) = \mathrm{Sp}(1)$. (Proof: See [45, p. 27] or [505, p. 128].)

Fact 3.9.3. Let $A \in \mathbb{R}^{2n\times 2n}$. If A is Hamiltonian and nonsingular, then A^{-1} is Hamiltonian. Now let $B \in \mathbb{R}^{2n\times 2n}$. If A and B are Hamiltonian, the $A+B$ is Hamiltonian.

Fact 3.9.4. Let $A \in \mathbb{R}^{2n\times 2n}$. Then, A is Hamiltonian if and only if $JA = (JA)^{\mathrm{T}}$. Furthermore, A is symplectic if and only if $A^{\mathrm{T}}JA = J$.

Fact 3.9.5. Let $A \in \mathbb{R}^{2n\times 2n}$ be Hamiltonian, and let $S \in \mathbb{R}^{2n\times 2n}$ be symplectic. Then, SAS^{-1} is Hamiltonian.

Fact 3.9.6. Let $\mathcal{A} \in \mathbb{R}^{2n\times 2n}$. Then, $\mathcal{A}$ is skew symmetric and Hamiltonian if and only if there exist a skew-symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a symmetric matrix $B \in \mathbb{R}^{n\times n}$ such that $\mathcal{A} = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}$.

Fact 3.9.7. Let $A \in \mathbb{R}^{2n\times 2n}$ be skew symmetric. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{2n\times 2n}$ such that $S^{\mathrm{T}}AS = J_n$. (Proof: See [45, p. 231].)

Fact 3.9.8. If $A \in \mathbb{R}^{2n\times 2n}$ is Hamiltonian and $A+I$ is nonsingular, then $B \triangleq (A-I)(A+I)^{-1}$ is symplectic and $I-B$ is nonsingular. Conversely, if $B \in \mathbb{R}^{2n\times 2n}$ is symplectic and $I-B$ is nonsingular, then $A = (I+B)(I-B)^{-1}$ is Hamiltonian. (Remark: See Fact 3.6.23, Fact 3.6.24, and Fact 3.6.25.)

3.10 Facts on Groups

Fact 3.10.1. The following subsets of $\mathbb{R}$ are groups:

i) $\{x \in \mathbb{R}\colon\ x \neq 0\}$.

ii) $\{x \in \mathbb{R}\colon\ x > 0\}$.

iii) $\{x \in \mathbb{R}\colon\ x \neq 0$ and x is rational$\}$.

iv) $\{x \in \mathbb{R}\colon\ x > 0$ and x is rational$\}$.

v) $\{-1, 1\}$.

vi) $\{1\}$.

Fact 3.10.2. The following subsets of $\mathbb{F}^{n\times n}$ are Lie algebras:

i) $\mathrm{ut}(n) \triangleq \{A \in \mathrm{gl}_{\mathbb{F}}(n)\colon\ A$ is upper triangular$\}$.

ii) $\mathrm{sut}(n) \triangleq \{A \in \mathrm{gl}_{\mathbb{F}}(n)\colon\ A$ is strictly upper triangular$\}$.

iii) $\{0_{n\times n}\}$.

Fact 3.10.3. The following subsets of $\mathbb{F}^{n\times n}$ are groups:

i) $\mathrm{UT}(n) \triangleq \{A \in \mathrm{GL}_{\mathbb{F}}(n)\colon\ A$ is upper triangular$\}$.

ii) $\mathrm{UT}_+(n) \triangleq \{A \in \mathrm{UT}(n)\colon\ A_{(i,i)} > 0$ for all $i = 1, \ldots, n\}$.

iii) $\mathrm{UT}_{\pm 1}(n) \triangleq \{A \in \mathrm{UT}(n)\colon\ A_{(i,i)} = \pm 1$ for all $i = 1, \ldots, n\}$.

iv) $\mathrm{SUT}(n) \triangleq \{A \in \mathrm{UT}(n)\colon\ A_{(i,i)} = 1$ for all $i = 1, \ldots, n\}$.

v) $\{I_n\}$.

(Remark: The matrices in $\mathrm{UT}_1(n)$ are *unipotent*. See Fact 5.13.6.)

Fact 3.10.4. Let $\mathcal{S} \subset \mathbb{F}^{n\times n}$, and assume that $\mathcal{S}$ is a group. Then, $\left\{A^{\mathrm{T}}\colon\ A \in \mathcal{S}\right\}$ and $\left\{\overline{A}\colon\ A \in \mathcal{S}\right\}$ are groups.

3.11 Facts on Quaternions

Fact 3.11.1. Define $Q_0, Q_2, Q_3 \in \mathbb{C}^{2\times 2}$ by

$$Q_0 \triangleq I_2,\ Q_1 \triangleq \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},\ Q_2 \triangleq \begin{bmatrix} \jmath & 0 \\ 0 & -\jmath \end{bmatrix},\ Q_3 \triangleq \begin{bmatrix} 0 & -\jmath \\ -\jmath & 0 \end{bmatrix}.$$

Then, the following statements hold:

i) $Q_0^* = Q_0$ and $Q_i^* = -Q_i$ for all $i = 1, 2, 3$.

ii) $Q_0^2 = Q_0$ and $Q_i^2 = -Q_0$ for all $i = 1, 2, 3$.

iii) $Q_iQ_j = -Q_jQ_i$ for all $1 \le i < j \le 3$.

iv) $Q_1Q_2 = Q_3$, $Q_2Q_3 = Q_1$, and $Q_3Q_1 = Q_2$.

v) $\{\pm Q_0, \pm Q_1, \pm Q_2, \pm Q_3\}$ is a group.

For $\beta \triangleq \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^4$ define

$$Q(\beta) \triangleq \sum_{i=0}^{3} \beta_i Q_i.$$

Then,

$$Q(\beta)Q^*(\beta) = \beta^{\mathrm{T}}\beta I_2$$

and

$$\det Q(\beta) = \beta^{\mathrm{T}}\beta.$$

Hence, if $\beta^{\mathrm{T}}\beta = 1$, then $Q(\beta)$ is unitary. Furthermore, the complex matrices Q_0, Q_1, Q_2, Q_3, and $Q(\beta)$ have the real representations

$$\mathcal{Q}_0 = I_4, \qquad \mathcal{Q}_1 = \begin{bmatrix} J_2 & 0 \\ 0 & J_2 \end{bmatrix},$$

$$\mathcal{Q}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathcal{Q}_3 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathcal{Q}(\beta) = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & -\beta_3 \\ -\beta_1 & \beta_0 & -\beta_3 & -\beta_2 \\ -\beta_2 & \beta_3 & \beta_0 & \beta_1 \\ \beta_3 & \beta_2 & -\beta_1 & \beta_0 \end{bmatrix}.$$

Hence,

$$\mathcal{Q}(\beta)\mathcal{Q}^{\mathrm{T}}(\beta) = \beta^{\mathrm{T}}\beta I_4$$

and

$$\det \mathcal{Q}(\beta) = \left(\beta^{\mathrm{T}}\beta\right)^2.$$

(Remark: Q_0, Q_1, Q_2, Q_3 represent the *quaternions* $1, \imath, \jmath, k$. See Fact 3.11.3. The quaternion group *v)* is isomorphic to SU(2).) (Remark: Matrices with quaternion entries and 4×4 matrix representations are considered in [38, 109, 248, 627]. For applications of quaternions, see [11, 250, 344].) (Remark: $\mathcal{Q}(\beta)$ has the form $\left[\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right]$, where A and $\hat{I}B$ are rotation-dilations. See Fact 2.15.1.)

Fact 3.11.2. Let $A \in \mathbb{C}^{2\times 2}$. Then, A is unitary if and only if there exist $\theta \in \mathbb{R}$ and $\beta \in \mathbb{R}^4$ such that $A = e^{\jmath\theta}Q(\beta)$, where $Q(\beta)$ is defined in Fact 3.11.1. (Proof: See [484, p. 228].)

Fact 3.11.3. Let $A_0, A_1, A_2, A_3 \in \mathbb{R}^{n\times n}$, let $\imath, \jmath, k$ satisfy

$$\begin{aligned} \imath^2 = \jmath^2 = k^2 &= -1, \\ \imath\jmath = k &= -\jmath\imath, \\ \jmath k = \imath &= -k\jmath, \\ k\imath = \jmath &= -\imath k, \end{aligned}$$

and let $A \triangleq A_0 + \imath A_1 + \jmath A_2 + kA_3$. Then,

$$\begin{bmatrix} A_0 & -A_1 & -A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{bmatrix} = U\mathrm{diag}(A, A, A, A)U,$$

where

$$U \triangleq \tfrac{1}{2}\begin{bmatrix} I & \imath I & \jmath I & kI \\ -\imath I & I & kI & -\jmath I \\ -\jmath I & -kI & I & \imath I \\ -kI & \jmath I & -\imath I & I \end{bmatrix}.$$

(Proof: See [551].) (Remark: k is not an integer here. $\imath, \jmath, k$ are the unit quaternions. This identity uses a similarity transformation to construct a real representation of quaternions. See Fact 2.12.14.)

3.12 Facts on Miscellaneous Types of Matrices

Fact 3.12.1. Let $A \in \mathbb{F}^{n\times m}$. Then, A is centrosymmetric if and only if $A^{\mathrm{T}} = A^{\hat{\mathrm{T}}}$. Furthermore, A is centrohermitian if and only if $A^* = A^{\hat{*}}$.

Fact 3.12.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. If A and B are both (centrohermitian, centrosymmetric), then so is AB.

Fact 3.12.3. Let $A \in \mathbb{F}^{n\times m}$. Then, A is (semicontractive, contractive) if and only if A^* is.

Fact 3.12.4. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is dissipative. Then,

A is nonsingular. (Proof: Suppose that A is singular, and let $x \in \mathcal{N}(A)$. Then, $x^*(A+A^*)x = 0$.) (Remark: If $A + A^*$ is nonsingular, then A is not necessarily nonsingular. Let $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$.)

Fact 3.12.5. Let $A \in \mathbb{R}^{n\times n}$ be tridiagonal with positive diagonal entries, and assume that, for all $i = 2, \ldots, n$,

$$A_{(i,i-1)}A_{(i-1,i)} < \tfrac{1}{4}(\cos \tfrac{\pi}{n+1})^{-2} A_{(i,i)}A_{(i-1,i-1)}.$$

Then, $\det A > 0$. (Proof: See [312].)

Fact 3.12.6. Let $A \in \mathbb{F}^{n\times n}$ be Toeplitz. Then, A is reverse symmetric.

Fact 3.12.7. Let $A \in \mathbb{F}^{n\times n}$. Then, A is Toeplitz if and only if there exist $a_0, \ldots, a_n \in \mathbb{F}$ and $b_1, \ldots, b_n \in \mathbb{F}$ such that

$$A = \sum_{i=1}^{n} b_i N_n^{i\mathrm{T}} + \sum_{i=0}^{n} a_i N_n^i.$$

Fact 3.12.8. Let $A \in \mathbb{F}^{n\times n}$, let $k \in \mathbb{P}$, and assume that A is (lower triangular, strictly lower triangular, upper triangular, strictly upper triangular). Then, so is A^k. If, in addition, A is Toeplitz, then so is A^k. (Remark: See Fact 11.10.1.)

Fact 3.12.9. Let $A \in \mathbb{F}^{n\times m}$. Then, the following statements hold:

i) If A is Toeplitz, then $\hat{I}A$ and $A\hat{I}$ are Hankel.

ii) If A is Hankel, then $\hat{I}A$ and $A\hat{I}$ are Toeplitz.

iii) A is Toeplitz if and only if $\hat{I}A\hat{I}$ is Toeplitz.

iv) A is Hankel if and only if $\hat{I}A\hat{I}$ is Hankel.

Fact 3.12.10. Let $A \in \mathbb{F}^{n\times n}$, assume that A is Hankel, and consider the following conditions:

i) A is Hermitian.

ii) A is real.

iii) A is symmetric.

Then, *i)* $\Longrightarrow$ *ii)* $\Longrightarrow$ *iii)*.

Fact 3.12.11. Let $A \in \mathbb{F}^{n\times n}$ be a partitioned matrix, each of whose blocks is a $k \times k$ (circulant, Hankel, Toeplitz) matrix. Then, A is similar to a block-(circulant, Hankel, Toeplitz) matrix. (Proof: See [60].)

Fact 3.12.12. For all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by $A_{(i,j)} \triangleq$

$1/(i+j-1)$. Then, A is Hankel and

$$\det A = \frac{[1!2!\cdots(n-1)!]^4}{1!2!\cdots(2n-1)!}.$$

Furthermore, for all $i,j = 1,\ldots,n$, A^{-1} has integer entries given by

$$\left(A^{-1}\right)_{(i,j)} = (-1)^{i+j}(i+j-1)\binom{n+i-1}{n-j}\binom{n+j-1}{n-1}\binom{i+j-2}{i-1}^2.$$

Finally, for large n,

$$\det A \approx 2^{-2n^2}.$$

(Remark: A is the *Hilbert matrix*, which is a Cauchy matrix. See [280, pp. 513], Fact 1.4.8, Fact 3.12.13, and Fact 8.7.29.)

Fact 3.12.13. Let $a_1,\ldots,a_n,b_1,\ldots,b_n \in \mathbb{R}$, assume that $a_i + b_j \neq 0$ for all $i,j = 1,\ldots,n$, and, for all $i,j = 1,\ldots,n$, define $A \in \mathbb{R}^{n\times n}$ by $A_{(i,j)} \triangleq 1/(a_i + b_j)$. Then, A is Hankel and

$$\det A = \frac{\prod_{1\le i<j\le n}(a_j - a_i)(b_j - b_i)}{\prod_{1\le i,j\le n}(a_i + b_j)}.$$

Now, assume that $a_1,\ldots,a_n$ are distinct and $b_1,\ldots,b_n$ are distinct. Then, A is nonsingular and

$$\left(A^{-1}\right)_{(i,j)} = \frac{\prod_{1\le k\le n}(a_j + b_k)(a_k + b_i)}{(a_j + b_i)\prod_{\substack{1\le k\le n\\ k\ne j}}(a_j - a_k)\prod_{\substack{1\le k\le n\\ k\ne i}}(b_i - b_k)}.$$

Furthermore,

$$1_{1\times n}A^{-1}1_{n\times 1} = \sum_{i=1}^n (a_i + b_i).$$

(Remark: A is a *Cauchy matrix.* See [280, p. 515], Fact 8.7.23, and Fact 1.4.8.)

Fact 3.12.14. Let $A \in \mathbb{R}^{n\times n}$ be tripotent. Then,

$$\operatorname{rank} A = \operatorname{rank} A^2 = \operatorname{tr} A^2.$$

Fact 3.12.15. Let $A \in \mathbb{F}^{n\times n}$. Then, A is nonsingular and tripotent if and only if A is involutory.

Fact 3.12.16. Let $A \in \mathbb{F}^{n\times n}$. Then, A is involutory if and only if $(A+I)(A-I)=0$.

Fact 3.12.17. $A \in \mathbb{R}^{n\times n}$, and assume that A is skew involutory. Then, n is even.

Fact 3.12.18. Let $x, y \in \mathbb{R}^n$, and assume that $x_{(1)} \ge \cdots \ge x_{(n)}$ and $y_{(1)} \ge \cdots \ge y_{(n)}$. Then, there exists a doubly stochastic matrix $A \in \mathbb{R}^{n\times n}$ such that $y = Ax$ if and only if y strongly majorizes x. (Remark: The matrix A is *doubly stochastic* if it is nonnegative, $1_{1\times n}A = 1_{1\times n}$, and $A1_{n\times 1} = 1_{n\times 1}$. This result is the *Hardy-Littlewood-Polya theorem*. See [93, p. 33], [287, p. 197], and [400, p. 22].)

3.13 Notes

In the literature on generalized inverses, range Hermitian matrices are traditionally called *EP matrices*. Elementary reflectors are traditionally called *Householder matrices* or *Householder reflections*.

Left equivalence, right equivalence, and biequivalence are treated in [484]. Each of the groups defined in Proposition 3.2.7 is actually a *Lie group*. Elementary treatments of Lie algebras and Lie groups are given in [36,45,157,196,227,299,455], while an advanced treatment appears in [571]. Some additional groups of structured matrices are given in [386].

Applications of the matrix inversion lemma are discussed in [256]. The terminology "idempotent" and "projector" is not standardized in the literature. Some writers use "projector" or "oblique projector" for idempotent, and "orthogonal projector" for projector. Centrosymmetric and centrohermitian matrices are discussed in [359,590]. Several characterizations of normal and almost normal matrices are given in [186,188,246]. Symplectic and Hamiltonian matrices are discussed in [354].

Chapter Four

Matrix Polynomials and Rational Transfer Functions

In this chapter we consider matrices whose entries are polynomials or rational functions. The decomposition of polynomial matrices in terms of the Smith form provides the foundation for developing canonical forms in Chapter 4. In this chapter we also present some basic properties of eigenvalues and eigenvectors as well as the minimal and characteristic polynomials of a square matrix. Finally, we consider the extension of the Smith form to the Smith-McMillan form for rational transfer functions.

4.1 Polynomials

A function $p\colon\ \mathbb{C} \mapsto \mathbb{C}$ of the form

$$p(s) = \beta_k s^k + \beta_{k-1}s^{k-1} + \cdots + \beta_1 s + \beta_0, \tag{4.1.1}$$

where $k \in \mathbb{N}$ and $\beta_0, \ldots, \beta_k \in \mathbb{F}$, is a *polynomial.* The set of polynomials is denoted by $\mathbb{F}[s]$. If the leading coefficient $\beta_k \in \mathbb{F}$ is nonzero, then the *degree* of p, denoted by $\deg p$, is k. If, in addition, $\beta_k = 1$, then p is *monic.* If $k = 0$, then p is *constant.* The degree of a nonzero constant polynomial is zero, while the degree of the zero polynomial is defined to be $-\infty$.

Let p_1 and p_2 be polynomials. Then,

$$\deg p_1 p_2 = \deg p_1 + \deg p_2. \tag{4.1.2}$$

If $p_1 = 0$ or $p_2 = 0$, then $\deg p_1 p_2 = \deg p_1 + \deg p_2 = -\infty$. If p_2 is a nonzero constant, then $\deg p_2 = 0$ and thus $\deg p_1 p_2 = \deg p_1$. Furthermore,

$$\deg(p_1 + p_2) \le \max\{\deg p_1, \deg p_2\}. \tag{4.1.3}$$

Therefore, $\deg(p_1 + p_2) = \max\{\deg p_1, \deg p_2\}$ if and only if either $\deg p_1 \ne \deg p_2$ or $p_1 = p_2 = 0$ or $\deg p_1 = \deg p_2 \ne -\infty$ and $\frac{\mathrm{d}^k}{\mathrm{d}s^k}[p_1(s) + p_2(s)] \ne 0$, where $k = \deg p_1 = \deg p_2$.

Let $p \in \mathbb{F}[s]$ be a polynomial of degree $k \geq 1$. Then, it follows from the *fundamental theorem of algebra* that p has k possibly repeated complex roots $\lambda_1, \ldots, \lambda_k$ so that p can be factored as

$$p(s) = \beta \prod_{i=1}^{k} (s - \lambda_i), \tag{4.1.4}$$

where $\beta \in \mathbb{F}$. The multiplicity of a root $\lambda \in \mathbb{C}$ of p is denoted by $\mathrm{m}_p(\lambda)$. If λ is not a root of p, then $\mathrm{m}_p(\lambda) = 0$. The multiset consisting of the roots of p including multiplicity is $\mathrm{mroots}(p) = \{\lambda_1, \ldots \lambda_k\}_{\mathrm{m}}$, while the set of roots of p ignoring multiplicity is $\mathrm{roots}(p) = \{\hat{\lambda}_1, \ldots \hat{\lambda}_l\}$, where $\sum_{i=1}^{l} \mathrm{m}_p(\hat{\lambda}_i) = k$. If $\mathbb{F} = \mathbb{R}$, then the multiplicity of a non-real root λ_i is equal to the multiplicity of its complex conjugate $\overline{\lambda_i}$. Hence, $\mathrm{mroots}(p)$ is *self conjugate*, that is, $\mathrm{mroots}(p) = \overline{\mathrm{mroots}(p)}$.

Let $p \in \mathbb{F}[s]$. If $p(-s) = p(s)$ for all $s \in \mathbb{C}$, then p is *even*, while, if $p(-s) = -p(s)$ for all $s \in \mathbb{C}$, then p is *odd.* If p is either odd or even, then $\mathrm{mroots}(p) = -\,\mathrm{mroots}(p)$. If $p \in \mathbb{R}[s]$ and there exists $q \in \mathbb{R}[s]$ such that $p(s) = q(s)q(-s)$ for all $s \in \mathbb{C}$, then p has a *spectral factorization.* If p has a spectral factorization, then p is even.

Proposition 4.1.1. Let $p \in \mathbb{R}[s]$. Then, the following statements are equivalent:

i) p has a spectral factorization.

ii) p is even and every imaginary root of p has even multiplicity.

iii) p is even and $p(\jmath\omega) \geq 0$ for all $\omega \in \mathbb{R}$.

Proof. The equivalence of *i)* and *ii)* is immediate. To prove *i)* $\Longrightarrow$ *iii)* note that, for all $\omega \in \mathbb{R}$,

$$p(\jmath\omega) = q(\jmath\omega)q(-\jmath\omega) = |q(\jmath\omega)|^2 \geq 0.$$

Conversely, to prove *iii)* $\Longrightarrow$ *i)* write $p = p_1p_2$, where all of the roots of p_1 are imaginary and none of the roots of p_2 are imaginary. Now, let z be a root of p_2. Then, $-z$, $\overline{z}$, and $-\overline{z}$ are also roots of p_2 with the same multiplicity as z. Hence, there exists a polynomial $p_{20} \in \mathbb{R}[s]$ such that $p_2(s) = p_{20}(s)p_{20}(-s)$ for all $s \in \mathbb{C}$.

Next, write $p_1(s) = \prod_{i=1}^{k} \big(s^2 + \omega_i^2\big)^{m_i}$, where $0 \leq \omega_1 < \cdots < \omega_k$ and $m_i \triangleq \mathrm{m}_{p_i}(\jmath\omega_i)$. Let ω_{i_0} denote the smallest element of the set $\{\omega_1, \ldots, \omega_k\}$ such that m_i is odd. Then, it follows that $p_1(\jmath\omega) = \prod_{i=1}^{k} \big(\omega_i^2 - \omega^2\big)^{m_i} < 0$ for all $\omega \in (\omega_{i_0}, \omega_{i_0+1})$, where $\omega_{k+1} \triangleq \infty$. However, note that $p_1(\jmath\omega) = p(\jmath\omega)/p_2(\jmath\omega) = p(\jmath\omega)/|p_{20}(\jmath\omega)|^2 \geq 0$ for all $\omega \in \mathbb{R}$, which is a contradiction. Therefore, m_i is even for all $i = 1, \ldots, k$, and thus $p_1(s) = p_{10}(s)p_{10}(-s)$

for all $s \in \mathbb{C}$, where $p_{10}(s) \triangleq \prod_{i=1}^{r}\left(s^2+\omega_i^2\right)^{m_i/2}$. Consequently, $p(s) = p_{10}(s)p_{20}(s)p_{10}(-s)p_{20}(-s)$ for all $s \in \mathbb{C}$. □

The following division algorithm is essential to the study of polynomials.

Lemma 4.1.2. Let $p_1, p_2 \in \mathbb{F}[s]$, and assume that p_2 is not the zero polynomial. Then, there exist unique polynomials $q, r \in \mathbb{F}[s]$ such that $\deg r < \deg p_2$ and

$$p_1 = qp_2 + r. \tag{4.1.5}$$

Proof. First note that if $\deg p_1 < \deg p_2$, then $q = 0$ and $r = p_1$. Hence, assume that $\deg p_1 = n \ge m = \deg p_2$ and write $p_1(s) = \beta_n s^n + \cdots + \beta_0$ and $p_2(s) = \gamma_m s^m + \cdots + \gamma_0$. If $n = 1$, then (4.1.5) is satisfied with $q(s) = \beta_1/\gamma_1$ and $r(s) = \beta_0 - \beta_1\gamma_0/\gamma_1$. Now, suppose that $n = 2$. Then, $\hat{p}_1(s) = p_1(s) - (\beta_2/\gamma_m)s^{2-m}p_2(s)$ has degree 1. Applying (4.1.5) with p_1 replaced by $\hat{p}_1$, it follows that there exist $q_1, r_1 \in \mathbb{F}[s]$ such that $\hat{p}_1 = q_1p_2 + r_1$ and such that $\deg r_1 < \deg p_2$. It thus follows that $p_1(s) = q_1(s)p_2(s) + r_1(s) + (\beta_2/\gamma_m)s^{2-m}p_2(s) = q(s)p_2(s) + r(s)$, where $q(s) = q_1(s) + (\beta_2/\gamma_m)s^{n-m}$ and $r = r_1$, which verifies (4.1.5). Similar arguments apply to successively larger values of n.

To prove uniqueness, suppose there exist polynomials $\hat{q}$ and $\hat{r}$ such that $\deg \hat{r} < \deg p_2$ and $p_1 = \hat{q}p_2 + \hat{r}$. Then, it follows that $(\hat{q} - q)p_2 = r - \hat{r}$. Next, note that $\deg(r - \hat{r}) < \deg p_2$. If $\hat{q} \ne q$, then $\deg p_2 \le \deg[(\hat{q} - q)p_2]$ so that $\deg(r - \hat{r}) < \deg[(\hat{q} - q)p_2]$, which is a contradiction. Thus, $\hat{q} = q$, and, hence, $r = \hat{r}$. □

In Lemma 4.1.2, q is the *quotient* of p_1 and p_2, while r is the *remainder*. If $\deg p_1 < \deg p_2$, then (4.1.5) is satisfied with $q = 0$ and $r = p_1$ so that $\deg r < \deg p_2$. Furthermore, if p_2 is a nonzero constant so that $\deg p_2 = 0$, then Lemma 4.1.2 implies that $q = p_1/p_2$ and $r = 0$, in which case $-\infty = \deg r < \deg p_2 = 0$. Finally, if $p_2(s) = s - \alpha$, where $\alpha \in \mathbb{F}$, then r is constant and thus $r(s) = p_1(\alpha)$. In general, if $r = 0$, then p_2 *divides* p_1, or, equivalently, p_1 is a *multiple* of p_2.

If a polynomial $p_3 \in \mathbb{F}[s]$ divides two polynomials $p_1, p_2 \in \mathbb{F}[s]$, then p_3 is a *common divisor* of p_1 and p_2. Given polynomials $p_1, p_2 \in \mathbb{F}[s]$, there exists a unique monic polynomial $p_3 \in \mathbb{F}[s]$, the *greatest common divisor* of p_1 and p_2, such that p_3 is a common divisor of p_1 and p_2 and such that every common divisor of p_1 and p_2 divides p_3. In addition, there exist polynomials $q_1, q_2 \in \mathbb{F}[s]$ such that the greatest common divisor p_3 of p_1 and p_2 is given by $p_3 = q_1p_1 + q_2p_2$. See [456, p. 113], for proofs of these results. Finally, p_1 and p_2 are *coprime* if their greatest common divisor is $p_3 = 1$, while a polynomial $p \in \mathbb{F}[s]$ is *irreducible* if there do not exist nonconstant polynomials $p_1, p_2 \in \mathbb{F}[s]$ such that $p = p_1 p_2$. For example, if $\mathbb{F} = \mathbb{R}$, then $p(s) = s^2 + s + 1$ is irreducible.

If a polynomial $p_3 \in \mathbb{F}[s]$ is a multiple of two polynomials $p_1, p_2 \in \mathbb{F}[s]$, then p_3 is a *common multiple* of p_1 and p_2. Given nonzero polynomials p_1 and p_2, there exists (see [456, p. 113]) a unique monic polynomial $p_3 \in \mathbb{F}[s]$, called the *least common multiple* of p_1 and p_2, that is a common multiple of p_1 and p_2 and that divides every common multiple of p_1 and p_2.

The polynomial $p \in \mathbb{F}[s]$ given by (4.1.1) can be evaluated with a square matrix argument $A \in \mathbb{F}^{n\times n}$ by defining

$$p(A) \triangleq \beta_k A^k + \beta_{k-1}A^{k-1} + \cdots + \beta_1 A + \beta_0 I. \tag{4.1.6}$$

4.2 Matrix Polynomials

The set $\mathbb{F}^{n\times m}[s]$ of *matrix polynomials* consists of matrix functions $P\colon\ \mathbb{C} \mapsto \mathbb{C}^{n\times m}$ all of whose entries are elements of $\mathbb{F}[s]$. A matrix polynomial $P \in \mathbb{F}^{n\times m}[s]$ can thus be written as

$$P(s) = s^k B_k + s^{k-1}B_{k-1} + \cdots + sB_1 + B_0, \tag{4.2.1}$$

where $B_0, \ldots, B_k \in \mathbb{F}^{n\times m}$. If B_k is nonzero, then the *degree* of P, denoted by $\deg P$, is k, while if $P = 0$, then $\deg P = -\infty$. If $n = m$ and B_k is nonsingular, then P is *regular*, while if $B_k = I$, then P is *monic*.

The following result, which generalizes Lemma 4.1.2, provides a division algorithm for matrix polynomials.

Lemma 4.2.1. Let $P_1, P_2 \in \mathbb{F}^{n\times n}[s]$, where P_2 is regular. Then, there exist unique matrix polynomials $Q, R, \hat{Q}, \hat{R} \in \mathbb{F}^{n\times n}[s]$ such that $\deg R < \deg P_2$, $\deg \hat{R} < \deg P_2$,

$$P_1 = QP_2 + R \tag{4.2.2}$$

and

$$P_1 = P_2\hat{Q} + \hat{R}. \tag{4.2.3}$$

Proof. See [456, pp. 134–135] or [230, p. 90]. □

If $R = 0$, then P_2 *right divides* P_1, while if $\hat{R} = 0$, then P_2 *left divides* P_1.

Let the matrix polynomial $P \in \mathbb{F}^{n\times m}[s]$ be given by (4.2.1). Then, P can be evaluated with a square matrix argument in two different ways, either from the right or from the left. For $A \in \mathbb{C}^{m\times m}$ define

$$P_{\mathrm{R}}(A) \triangleq B_kA^k + B_{k-1}A^{k-1} + \cdots + B_1A + B_0, \tag{4.2.4}$$

while, for $A \in \mathbb{C}^{n\times n}$, define

$$P_{\mathrm{L}}(A) \triangleq A^kB_k + A^{k-1}B_{k-1} + \cdots + AB_1 + B_0. \tag{4.2.5}$$

If $n = m$, then $P_{\mathrm{R}}(A)$ and $P_{\mathrm{L}}(A)$ can be evaluated for all $A \in \mathbb{F}^{n\times n}$, but are generally different.

The following result is useful.

Lemma 4.2.2. Let $Q, \hat{Q} \in \mathbb{F}^{n\times n}[s]$ and $A \in \mathbb{F}^{n\times n}$. Furthermore, define $P, \hat{P} \in \mathbb{F}^{n\times n}[s]$ by $P(s) \triangleq Q(s)(sI - A)$ and $\hat{P}(s) \triangleq (sI - A)\hat{Q}(s)$. Then, $P_{\mathrm{R}}(A) = 0$ and $\hat{P}_{\mathrm{L}}(A) = 0$.

Let $p \in \mathbb{F}[s]$ be given by (4.1.1) and define $P(s) \triangleq p(s)I_n = s^k\beta_kI_n + s^{k-1}\beta_{k-1}I_n + \cdots + s\beta_1I_n + \beta_0I_n \in \mathbb{F}^{n\times n}[s]$. For $A \in \mathbb{C}^{n\times n}$ it follows that $p(A) = P(A) = P_{\mathrm{R}}(A) = P_{\mathrm{L}}(A)$.

The following result specializes Lemma 4.2.1 to the case of matrix polynomial divisors of degree 1.

Corollary 4.2.3. Let $P \in \mathbb{F}^{n\times n}[s]$ and $A \in \mathbb{F}^{n\times n}$. Then, there exist unique matrix polynomials $Q, \hat{Q} \in \mathbb{F}^{n\times n}[s]$ and unique matrices $R, \hat{R} \in \mathbb{F}^{n\times n}$ such that

$$P(s) = Q(s)(sI - A) + R, \tag{4.2.6}$$

and

$$P(s) = (sI - A)\hat{Q}(s) + \hat{R}. \tag{4.2.7}$$

Furthermore, $R = P_{\mathrm{R}}(A)$ and $\hat{R} = P_{\mathrm{L}}(A)$.

Proof. In Lemma 4.2.1 set $P_1 = P$ and $P_2(s) = sI - A$. Since $\deg P_2 = 1$, it follows that $\deg R = \deg \hat{R} = 0$ and thus R and $\hat{R}$ are constant. Finally, the last statement follows from Lemma 4.2.2. □

Definition 4.2.4. Let $P \in \mathbb{F}^{n\times m}[s]$. Then, the *rank* of P is the nonnegative integer

$$\operatorname{rank} P \triangleq \max_{s\in\mathbb{C}} \operatorname{rank} P(s). \tag{4.2.8}$$

Let $P \in \mathbb{F}^{n\times n}[s]$. Then, $P(s) \in \mathbb{C}^{n\times n}$ for all $s \in \mathbb{C}$. Furthermore, $\det P$ is a polynomial in s, that is, $\det P \in \mathbb{F}[s]$.

Definition 4.2.5. Let $P \in \mathbb{F}^{n\times n}[s]$. Then, P is *nonsingular* if $\det P$ is not the zero polynomial; otherwise, P is *singular*.

Proposition 4.2.6. Let $P \in \mathbb{F}^{n\times n}[s]$, and assume that P is regular. Then, P is nonsingular.

Let $P \in \mathbb{F}^{n\times n}[s]$. If P is nonsingular, then the *inverse* P^{-1} of P can be constructed according to (2.7.21). In general, the entries of P^{-1} are rational functions of s (see Definition 4.7.1). For example, if $P(s) = \left[\begin{smallmatrix} s+2 & s+1 \\ s-2 & s-1 \end{smallmatrix}\right]$, then $P^{-1}(s) = \frac{1}{2}\left[\begin{smallmatrix} 1 & -\frac{s+1}{s-1} \\ -\frac{s-2}{s-1} & \frac{s+1}{s-1} \end{smallmatrix}\right]$. In certain cases P^{-1} is also a matrix polynomial. For example, if $P(s) = \left[\begin{smallmatrix} s & 1 \\ s^2+s-1 & s+1 \end{smallmatrix}\right]$, then $P^{-1}(s) = \left[\begin{smallmatrix} s+1 & -1 \\ -s^2-s+1 & s \end{smallmatrix}\right]$.

The following result is an extension of Proposition 2.7.7 from constant to matrix polynomials.

Proposition 4.2.7. Let $P \in \mathbb{F}^{n\times m}[s]$. Then, $\operatorname{rank} P$ is the order of the largest nonsingular matrix polynomial that is a submatrix of P.

Proof. For all $s \in \mathbb{C}$ it follows from Proposition 2.7.7 that $\operatorname{rank} P(s)$ is the order of the largest nonsingular submatrix of $P(s)$. Now, let $s_0 \in \mathbb{C}$ be such that $\operatorname{rank} P(s_0) = \operatorname{rank} P$. Then, $P(s_0)$ has a nonsingular submatrix of maximal order $\operatorname{rank} P$. Therefore, P has a nonsingular submatrix polynomial of maximal order $\operatorname{rank} P$. □

A matrix polynomial can be transformed by performing elementary row and column operations of the following types:

i) Multiply a row or a column by a nonzero constant.

ii) Interchange two rows or two columns.

iii) Add a polynomial multiple of one (row, column) to another (row, column).

These operations correspond to left multiplication or right multiplication by the elementary matrices

$$I_n + (\alpha - 1)E_{i,i} = \begin{bmatrix} I_{i-1} & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & I_{n-i} \end{bmatrix}, \tag{4.2.9}$$

where $\alpha \in \mathbb{F}$ is nonzero,

$$I_n + E_{i,j} + E_{j,i} - E_{i,i} - E_{j,j} = \begin{bmatrix} I_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{n-j} \end{bmatrix}, \tag{4.2.10}$$

where $i \neq j$, and as well as the *elementary matrix polynomial*

$$I_n + pE_{i,j} = \begin{bmatrix} I_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & p & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_{n-j} \end{bmatrix}, \tag{4.2.11}$$

where $i \neq j$ and $p \in \mathbb{F}[s]$. The matrices shown in (4.2.10) and (4.2.11) illustrate the case $i < j$. Applying these operations sequentially corresponds to forming products of elementary matrices and elementary matrix polynomials. Note that the elementary matrix polynomial $I + pE_{i,j}$ is nonsingular and that $(I+pE_{i,j})^{-1} = I - pE_{i,j}$ so that the inverse of an elementary matrix polynomial is an elementary matrix polynomial.

4.3 The Smith Decomposition and Similarity Invariants

Definition 4.3.1. Let $P \in \mathbb{F}^{n\times n}[s]$. Then, P is *unimodular* if P is the product of elementary matrices and elementary matrix polynomials.

The following result provides a canonical form, known as the *Smith form*, for matrix polynomials under unimodular transformation.

Theorem 4.3.2. Let $P \in \mathbb{F}^{n\times m}[s]$, and let $r \triangleq \operatorname{rank} P$. Then, there exist unimodular matrices $S_1 \in \mathbb{F}^{n\times n}[s]$ and $S_2 \in \mathbb{F}^{m\times m}[s]$ and monic polynomials $p_1, \ldots, p_r \in \mathbb{F}[s]$ such that p_i divides p_{i+1} for all $i = 1, \ldots, r-1$ and such that

$$P = S_1 \begin{bmatrix} p_1 & & & \\ & \ddots & & \\ & & p_r & \\ & & & 0_{(n-r)\times(m-r)} \end{bmatrix} S_2. \tag{4.3.1}$$

Furthermore, for all $i = 1, \ldots, r$, p_i is uniquely determined by

$$\Delta_i = p_1 \cdots p_i, \tag{4.3.2}$$

where Δ_i is the greatest common divisor of all $i \times i$ subdeterminants of P.

Proof. The result is obtained by sequentially applying elementary row and column operations to P. For details, see [321, pp. 390–392] or [456, pp. 125–128]. □

Corollary 4.3.3. Let $P \in \mathbb{R}^{n\times n}[s]$ be unimodular. Then, the Smith form of P is the identity.

Definition 4.3.4. The monic polynomials $p_1, \ldots, p_r \in \mathbb{F}[s]$ of the Smith form of $P \in \mathbb{F}^{n\times n}[s]$ are the *invariant polynomials* of P.

Proposition 4.3.5. Let $P \in \mathbb{F}^{n\times n}[s]$. Then, P is unimodular if and only if $\det P$ is a nonzero constant.

Proof. Necessity is immediate since every elementary matrix and every elementary matrix polynomial has a constant nonzero determinant. To prove sufficiency, note that, since $\det P$ is a nonzero constant, it follows from Theorem 4.3.2 that every invariant polynomial of P is also a nonzero constant. Consequently, P is a product of elementary matrices and elementary matrix polynomials and thus is unimodular. $\square$

Proposition 4.3.6. Let $P \in \mathbb{F}^{n\times n}[s]$. Then, the following statements are equivalent:

i) P is unimodular.

ii) P is nonsingular, and P^{-1} is a matrix polynomial.

iii) P is nonsingular, and P^{-1} is unimodular.

Proof. To prove *i)* $\Longrightarrow$ *ii)* suppose that P is unimodular. Then, it follows from Proposition 4.3.5 that $\det P$ is a nonzero constant. Therefore, P is nonsingular. Furthermore, since P^{A} is a matrix polynomial, it follows that $P^{-1} = (\det P)^{-1}P^{\mathrm{A}}$ is a matrix polynomial. To prove *ii)* $\Longrightarrow$ *iii)* suppose that P is nonsingular and P^{-1} is a matrix polynomial so that $\det P^{-1}$ is a polynomial. Since $\det P$ is a nonzero constant and $\det P^{-1} = 1/\det P$, it follows that $\det P^{-1}$ is also a nonzero constant. Thus, Proposition 4.3.5 implies that P^{-1} is unimodular. Finally, to prove *iii)* $\Longrightarrow$ *i)*, suppose that P is nonsingular and P^{-1} is unimodular. Then, since $\det P^{-1}$ is a nonzero constant, it follows that $\det P = 1/\det P^{-1}$ is a nonzero constant. Proposition 4.3.5 thus implies that P is unimodular. $\square$

Proposition 4.3.7. Let $A_1, B_1, A_2, B_2 \in \mathbb{F}^{n\times n}$, where A_2 is nonsingular, and define the matrix polynomials $P_1, P_2 \in \mathbb{F}^{n\times n}[s]$ by $P_1(s) \triangleq sA_1 + B_1$ and $P_2(s) \triangleq sA_2 + B_2$. Then, P_1 and P_2 have the same invariant polynomials if and only if there exist nonsingular matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that $P_2 = S_1P_1S_2$.

Proof. The sufficiency result is immediate. To prove necessity, note that it follows from Theorem 4.3.2 that there exist unimodular matrices $T_1, T_2 \in \mathbb{F}^{n\times n}[s]$ such that $P_2 = T_2P_1T_1$. Now, since P_2 is regular, it follows from Lemma 4.2.1 that there exist matrix polynomials $Q, \hat{Q} \in \mathbb{F}^{n\times n}[s]$ and constant matrices $R, \hat{R} \in \mathbb{F}^{n\times n}$ such that $T_1 = QP_2 + R$ and $T_2 = P_2\hat{Q} + \hat{R}$.

Next, we have

$$\begin{aligned}
P_2 &= T_2P_1T_1\\
&= (P_2\hat{Q} + \hat{R})P_1T_1\\
&= \hat{R}P_1T_1 + P_2\hat{Q}T_2^{-1}P_2\\
&= \hat{R}P_1(QP_2 + R) + P_2\hat{Q}T_2^{-1}P_2\\
&= \hat{R}P_1R + (T_2 - P_2\hat{Q})P_1QP_2 + P_2\hat{Q}T_2^{-1}P_2\\
&= \hat{R}P_1R + T_2P_1QP_2 + P_2\Big(-\hat{Q}P_1Q + \hat{Q}T_2^{-1}\Big)P_2\\
&= \hat{R}P_1R + P_2\Big(T_1^{-1}Q - \hat{Q}P_1Q + \hat{Q}T_2^{-1}\Big)P_2.
\end{aligned}$$

Since P_2 is regular and has degree 1, it follows that if $T_1^{-1}Q-\hat{Q}P_1Q+\hat{Q}T_2^{-1}$ is not zero, then $\deg P_2\Big(T_1^{-1}Q - \hat{Q}P_1Q + \hat{Q}T_2^{-1}\Big)P_2 \ge 2$. However, since P_2 and $\hat{R}P_1R$ have degree less than two, it follows that $T_1^{-1}Q - \hat{Q}P_1Q + \hat{Q}T_2^{-1} = 0$. Hence, $P_2 = \hat{R}P_1R$.

Next, to show that $\hat{R}$ and R are nonsingular, note that, for all $s \in \mathbb{C}$,

$$P_2(s) = \hat{R}P_1(s)R = s\hat{R}A_1R + \hat{R}B_1R,$$

which implies that $A_2 = S_1A_1S_2$, where $S_1 = \hat{R}$ and $S_2 = R$. Since A_2 is nonsingular, it follows that S_1 and S_2 are nonsingular. □

Definition 4.3.8. Let $A \in \mathbb{F}^{n\times n}$. Then, the invariant polynomials of $sI - A$ are the *similarity invariants* of A.

The following result provides necessary and sufficient conditions for two matrices to be similar.

Theorem 4.3.9. Let $A, B \in \mathbb{F}^{n\times n}$. Then, A and B are similar if and only if they have the same similarity invariants.

Proof. To prove necessity, assume that A and B are similar. Then, the matrices $sI - A$ and $sI - B$ have the same Smith form and thus the same similarity invariants. To prove sufficiency, it follows from Proposition 4.3.7 that there exist nonsingular matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that $sI - A = S_1(sI - B)S_2$. Thus, $S_1 = S_2^{-1}$, and, hence, $A = S_1BS_1^{-1}$. □

4.4 Eigenvalues

Let $A \in \mathbb{F}^{n\times n}$. Then, the matrix polynomial $sI - A \in \mathbb{F}^{n\times n}[s]$ is monic and has degree 1.

Definition 4.4.1. Let $A \in \mathbb{F}^{n\times n}$. Then, the *characteristic polynomial* of A is the polynomial $\chi_A \in \mathbb{F}[s]$ given by

$$\chi_A(s) \triangleq \det(sI - A). \tag{4.4.1}$$

Proposition 4.4.2. Let $A \in \mathbb{F}^{n\times n}$. Then, χ_A is monic and $\deg \chi_A = n$.

Let $A \in \mathbb{F}^{n\times n}$ and write the characteristic polynomial of A as

$$\chi_A(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0, \tag{4.4.2}$$

where $\beta_0, \ldots, \beta_{n-1} \in \mathbb{F}$. The *eigenvalues* of A are the n possibly repeated roots $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ of χ_A, that is, the solutions of the *characteristic equation*

$$\chi_A(s) = 0. \tag{4.4.3}$$

It is often convenient to denote the eigenvalues of A by $\lambda_1(A), \ldots, \lambda_n(A)$ or just $\lambda_1, \ldots, \lambda_n$. This notation may be ambiguous, however, since it does not uniquely specify which eigenvalue is denoted by λ_i. If, however, every eigenvalue of A is real, then we employ the notational convention

$$\lambda_1 \geq \cdots \geq \lambda_n, \tag{4.4.4}$$

and we define

$$\lambda_{\max}(A) \triangleq \lambda_1, \quad \lambda_{\min}(A) \triangleq \lambda_n. \tag{4.4.5}$$

Definition 4.4.3. Let $A \in \mathbb{F}^{n\times n}$. The *algebraic multiplicity* of an eigenvalue λ of A, denoted by $\mathrm{am}_A(\lambda)$, is the algebraic multiplicity of λ as a root of χ_A, that is,

$$\mathrm{am}_A(\lambda) \triangleq \mathrm{m}_{\chi_A}(\lambda). \tag{4.4.6}$$

The multiset consisting of the eigenvalues of A including their algebraic multiplicity, denoted by $\mathrm{mspec}(A)$, is the *multispectrum* of A, that is,

$$\mathrm{mspec}(A) \triangleq \mathrm{mroots}(\chi_A). \tag{4.4.7}$$

Ignoring algebraic multiplicity, $\mathrm{spec}(A)$ denotes the *spectrum* of A, that is,

$$\mathrm{spec}(A) \triangleq \mathrm{roots}(\chi_A). \tag{4.4.8}$$

If $\lambda \notin \mathrm{spec}(A)$, then $\lambda \notin \mathrm{roots}(\chi_A)$, and thus $\mathrm{am}_A(\lambda) = \mathrm{m}_{\chi_A}(\lambda) = 0$.

Let $A \in \mathbb{F}^{n\times n}$ and $\mathrm{mroots}(\chi_A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$\chi_A(s) = \prod_{i=1}^{n}(s - \lambda_i). \tag{4.4.9}$$

If $\mathbb{F} = \mathbb{R}$, then $\chi_A(s)$ has real coefficients, and thus the eigenvalues of A occur in complex conjugate pairs, that is, $\overline{\mathrm{mroots}(\chi_A)} = \mathrm{mroots}(\chi_A)$. Now, let $\mathrm{spec}(A) = \{\lambda_1, \ldots, \lambda_r\}$, and, for all $i = 1, \ldots, r$, let n_i denote the algebraic multiplicity of λ_i. Then,

$$\chi_A(s) = \prod_{i=1}^{r} (s - \lambda_i)^{n_i}. \tag{4.4.10}$$

The following result gives some basic properties of the spectrum of a matrix.

Proposition 4.4.4. Let $A, B \in \mathbb{F}^{n \times n}$. Then, the following statements hold:

i) $\chi_{A^{\mathrm{T}}} = \chi_A$.

ii) $\chi_{-A} = (-1)^n \chi_A$.

iii) $\mathrm{mspec}\big(A^{\mathrm{T}}\big) = \mathrm{mspec}(A)$.

iv) $\mathrm{mspec}\big(\overline{A}\big) = \overline{\mathrm{mspec}(A)}$.

v) $\mathrm{mspec}(A^*) = \overline{\mathrm{mspec}(A)}$.

vi) $0 \in \mathrm{spec}(A)$ if and only if $\det A = 0$.

vii) If either $k \in \mathbb{N}$ or A is nonsingular and $k \in \mathbb{Z}$, then

$$\mathrm{mspec}\Big(A^k\Big) = \Big\{\lambda^k\colon\ \lambda \in \mathrm{mspec}(A)\Big\}_{\mathrm{m}}. \tag{4.4.11}$$

viii) If $\alpha \in \mathbb{F}$, then $\mathrm{mspec}(\alpha I + A) = \alpha + \mathrm{mspec}(A)$.

ix) If $\alpha \in \mathbb{F}$, then $\mathrm{mspec}(\alpha A) = \alpha \mathrm{mspec}(A)$.

x) If $A = A^*$, then $\mathrm{spec}(A) \subset \mathbb{R}$.

xi) If A and B are similar, then $\chi_A = \chi_B$ and $\mathrm{mspec}(A) = \mathrm{mspec}(B)$.

Proof. To prove *i)* note that $\det\big(sI - A^{\mathrm{T}}\big) = \det\big[(sI - A)^{\mathrm{T}}\big] = \det(sI - A)$. To prove *ii)* note that $\chi_{-A} = \det(sI + A) = (-1)^n \det(-sI - A) = (-1)^n \chi_A(-s)$. Next, *iii)* follows from *i)*. Next, *iv)* follows from $\det(sI - \overline{A}) = \det(\overline{\overline{s}I - A}) = \overline{\det(\overline{s}I - A)}$, while *v)* follows from *iii)* and *iv)*. Next, *vi)* follows from the fact that $\chi_A(0) = (-1)^n \det A$. To prove *vii)* note that, if $\lambda \in \mathrm{spec}(A)$ and $x \in \mathbb{C}^n$ is an eigenvector of A associated with λ, then $A^2x = A(Ax) = A(\lambda x) = \lambda Ax = \lambda^2 x$. Similarly, if A is nonsingular, then $Ax = \lambda x$ implies that $A^{-1}x = \lambda^{-1}x$, and thus $A^{-2}x = \lambda^{-2}x$. Next, if $\lambda \in \mathrm{spec}(A)$ and $\alpha \in \mathbb{F}$, then $\det[(\alpha + \lambda)I - (\alpha I + A)] = \det(\lambda I - A) = 0$, which implies that $\alpha + \lambda \in \mathrm{spec}(\alpha I + A)$ and thus proves *viii)*. If $\lambda \in \mathrm{spec}(A)$ and $\alpha \in \mathbb{F}$, then $\det(\alpha\lambda I - \alpha A) = \alpha^n \det(\lambda I - A) = 0$, which implies that $\alpha\lambda \in \mathrm{spec}(\alpha A)$, which proves *ix*). To prove *x*), assume $A = A^*$, let $\lambda \in \operatorname{spec}(A)$, and let $x \in \mathbb{C}^n$ be an eigenvector of A associated with λ. Then, $\lambda = x^*Ax/x^*x$, which is real. Finally, the proof of *xi*) is immediate. □

The following result characterizes the coefficients of χ_A in terms of the eigenvalues of A.

Proposition 4.4.5. Let $A \in \mathbb{F}^{n\times n}$, let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$, and, for all $i = 1, \ldots, n$, let γ_i denote the sum of all $i\times i$ principal subdeterminants of A. Then, for all $i = 1, \ldots, n-1$,

$$\gamma_i = \sum \lambda_{j_1} \cdots \lambda_{j_i}, \tag{4.4.12}$$

where the summation in (4.4.12) is taken over all multisubsets of $\operatorname{mspec}(A)$ having i elements. Furthermore, for all $i = 0, \ldots, n-1$, the coefficient β_i of s^i in (4.4.2) is given by

$$\beta_i = (-1)^{n-i}\gamma_{n-i}. \tag{4.4.13}$$

In particular,

$$\beta_{n-1} = -\operatorname{tr} A = -\sum_{i=1}^{n} \lambda_i, \tag{4.4.14}$$

$$\beta_{n-2} = \tfrac{1}{2}\big[(\operatorname{tr} A)^2 - \operatorname{tr} A^2\big] = \sum \lambda_{j_1}\lambda_{j_2}, \tag{4.4.15}$$

$$\beta_1 = (-1)^{n-1}\operatorname{tr} A^{\mathrm{A}} = (-1)^{n-1}\sum \lambda_{j_1} \cdots \lambda_{j_{n-1}}, \tag{4.4.16}$$

$$\beta_0 = (-1)^n \det A = (-1)^n \prod_{i=1}^{n} \lambda_i. \tag{4.4.17}$$

Proof. The expression for γ_i given by (4.4.12) follows from the factored form of $\chi_A(s)$ given by (4.4.9), while the expression for β_i given by (4.4.13) follows by examining the cofactor expansion (2.7.15) of $\det(sI - A)$. For details, see [416, p. 495]. Equation (4.4.14) follows from (4.4.13) and the fact that the $(n-1)\times(n-1)$ principal subdeterminants of A are the diagonal entries $A_{(i,i)}$. Using

$$\sum_{i=1}^{n} \lambda_i^2 = \left(\sum_{i=1}^{n} \lambda_i\right)^2 - 2\sum \lambda_{j_1}\lambda_{j_2}$$

and (4.4.14) yields (4.4.15). Next, if A is nonsingular, then $\chi_{A^{-1}}(s) = (-s)^n\big(\det A^{-1}\big)\chi_A(1/s)$. Using (4.4.2) with s replaced by $1/s$ and (4.4.14), it follows that $\operatorname{tr} A^{-1} = (-1)^{n-1}\big(\det A^{-1}\big)\beta_1$, and, hence, (4.4.16) is satisfied. Using continuity for the case in which A is singular yields (4.4.16) for arbitrary A. Finally, $\beta_0 = \chi_A(0) = \det(0I - A) = (-1)^n \det A$, which verifies

(4.4.17). □

From the definition the adjugate of a matrix it follows that $(sI-A)^{\mathrm{A}} \in \mathbb{F}^{n\times n}[s]$ is a monic matrix polynomial of degree $n-1$ of the form

$$(sI-A)^{\mathrm{A}} = s^{n-1}I + s^{n-2}B_{n-2} + \cdots + sB_1 + B_0, \tag{4.4.18}$$

where $B_0, B_1, \ldots, B_{n-2} \in \mathbb{F}^{n\times n}$. Since $(sI-A)^{\mathrm{A}}$ is regular it follows from Proposition 4.2.6 that $(sI-A)^{\mathrm{A}}$ is a nonsingular polynomial matrix.

The next result is the *Cayley-Hamilton theorem*, which shows that every matrix is a "root" of its characteristic polynomial.

Theorem 4.4.6. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\chi_A(A) = 0. \tag{4.4.19}$$

Proof. Define $P, Q \in \mathbb{F}^{n\times n}[s]$ by $P(s) \triangleq \chi_A(s)I$ and $Q(s) \triangleq (sI-A)^{\mathrm{A}}$. Then, (4.7.2) implies that $P(s) = Q(s)(sI-A)$. It thus follows from Lemma 4.2.2 that $P_{\mathrm{R}}(A) = 0$. Furthermore, $\chi_A(A) = P(A) = P_{\mathrm{R}}(A)$. Hence, $\chi_A(A) = 0$. □

In the notation of (4.4.10), it thus follows from Theorem 4.4.6 that

$$\prod_{i=1}^{r}(\lambda_i I - A)^{n_i} = 0. \tag{4.4.20}$$

Lemma 4.4.7. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}s}\chi_A(s) = \operatorname{tr}\big[(sI-A)^{\mathrm{A}}\big] = \sum_{i=1}^{n}\det\big(sI - A_{[i,i]}\big). \tag{4.4.21}$$

Proof. It follows from (4.4.16) that $\frac{\mathrm{d}}{\mathrm{d}s}\chi_A(s)\big|_{s=0} = \beta_1 = (-1)^{n-1}\operatorname{tr} A^{\mathrm{A}}$. Hence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\chi_A(s) &= \frac{\mathrm{d}}{\mathrm{d}z}\det[(s+z)I - A]\bigg|_{z=0} = \frac{\mathrm{d}}{\mathrm{d}z}\det[zI-(-sI+A)]\bigg|_{z=0}\\ &= (-1)^{n-1}\operatorname{tr}\big[(-sI+A)^{\mathrm{A}}\big] = \operatorname{tr}\big[(sI-A)^{\mathrm{A}}\big].\end{aligned}$$ □

The following result, known as *Leverrier's algorithm*, provides a recursive formula for the coefficients $\beta_0, \ldots, \beta_{n-1}$ of χ_A and $B_0, \ldots, B_{n-2}$ of $(sI-A)^{\mathrm{A}}$.

Proposition 4.4.8. Let $A \in \mathbb{F}^{n\times n}$, let χ_A be given by (4.4.2), and let $(sI-A)^{\mathrm{A}}$ be given by (4.4.18). Then, $\beta_{n-1}, \ldots, \beta_0$ and $B_{n-2}, \ldots, B_0$ are given by

$$\beta_k = \frac{1}{k-n} \operatorname{tr} AB_k, \quad k = n-1, \ldots, 0, \tag{4.4.22}$$

$$B_{k-1} = AB_k + \beta_k I, \quad k = n-1, \ldots, 1, \tag{4.4.23}$$

where $B_{n-1} = I$.

Proof. Since $(sI - A)(sI - A)^{\mathrm{A}} = \chi_A(s)I$, it follows that

$$s^n I + s^{n-1}(B_{n-2} - A) + s^{n-2}(B_{n-3} - AB_{n-2}) + \cdots + s(B_0 - AB_1) - AB_0 = (s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0)I.$$

Equating coefficients of powers of s yields (4.4.23) along with $-AB_0 = \beta_0 I$. Taking the trace of this last identity yields $\beta_0 = -\frac{1}{n} \operatorname{tr} AB_0$, which confirms (4.4.22) for $k = 0$. Next, using (4.4.21) and (4.4.18), it follows that

$$\frac{\mathrm{d}}{\mathrm{d}s}\chi_A(s) = \sum_{k=1}^{n} k\beta_k s^{k-1} = \sum_{k=1}^{n} (\operatorname{tr} B_{k-1}) s^{k-1},$$

where $B_{n-1} \triangleq I_n$ and $\beta_n \triangleq 1$. Equating powers of s, it follows that $k\beta_k = \operatorname{tr} B_{k-1}$ for all $k = 1, \ldots, n$. Now, (4.4.23) implies that $k\beta_k = \operatorname{tr}(AB_k + \beta_k I)$ for all $k = 1, \ldots, n-1$, which implies (4.4.22). □

Proposition 4.4.9. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$, and assume that $m \le n$. Then,

$$\chi_{AB}(s) = s^{n-m}\chi_{BA}(s). \tag{4.4.24}$$

Consequently,

$$\operatorname{mspec}(AB) = \operatorname{mspec}(BA) \cup \{0, \ldots, 0\}_{\mathrm{m}}, \tag{4.4.25}$$

where the multiset $\{0, \ldots, 0\}_{\mathrm{m}}$ contains $n - m$ zeros.

Proof. First note that

$$\begin{bmatrix} 0_{m\times m} & 0_{m\times n} \\ A & AB \end{bmatrix} = \begin{bmatrix} I_m & -B \\ 0_{n\times m} & I_n \end{bmatrix} \begin{bmatrix} BA & 0_{m\times n} \\ A & 0_{n\times n} \end{bmatrix} \begin{bmatrix} I_m & B \\ 0_{n\times m} & I_n \end{bmatrix},$$

which shows that $\left[\begin{smallmatrix} 0_{m\times m} & 0_{m\times n} \\ A & AB \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} BA & 0_{m\times n} \\ A & 0_{n\times n} \end{smallmatrix}\right]$ are similar. It thus follows from *xi*) of Proposition 4.4.4 that $s^m\chi_{AB}(s) = s^n\chi_{BA}(s)$, which implies (4.4.24). Finally, (4.4.25) follows immediately from (4.4.24). □

If $n = m$, then Proposition 4.4.9 specializes to the following result.

Corollary 4.4.10. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\chi_{AB} = \chi_{BA}. \tag{4.4.26}$$

Consequently,

$$\operatorname{mspec}(AB) = \operatorname{mspec}(BA). \tag{4.4.27}$$

4.5 Eigenvectors

Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \mathbb{C}$ be an eigenvalue of A. Then, $\chi_A(\lambda) = \det(\lambda I - A) = 0$, and thus $\lambda I - A \in \mathbb{C}^{n\times n}$ is singular. Furthermore, $\mathcal{N}(\lambda I - A)$ is a nontrivial subspace of $\mathbb{C}^n$, that is, $\operatorname{def}(\lambda I - A) > 0$. If $x \in \mathcal{N}(\lambda I - A)$, that is, $Ax = \lambda x$, and $x \neq 0$, then x is an *eigenvector of A associated with* λ. Note that if A and λ are real, then there exists a real eigenvector associated with λ.

Definition 4.5.1. The *geometric multiplicity* of $\lambda \in \operatorname{spec}(A)$, denoted by $\operatorname{gm}_A(\lambda)$, is the number of linearly independent eigenvectors associated with λ, that is,

$$\operatorname{gm}_A(\lambda) \triangleq \operatorname{def}(\lambda I - A). \tag{4.5.1}$$

By convention, if $\lambda \notin \operatorname{spec}(A)$, then $\operatorname{gm}_A(\lambda) \triangleq 0$.

The spectral properties of normal matrices deserve special attention.

Lemma 4.5.2. Let $A \in \mathbb{F}^{n\times n}$ be normal, let $\lambda \in \operatorname{spec}(A)$, and let $x \in \mathbb{C}^n$ be an eigenvector of A associated with λ. Then, x is an eigenvector of A^* associated with $\overline{\lambda} \in \operatorname{spec}(A^*)$.

Proof. Since $\lambda \in \operatorname{spec}(A)$, *iii*) of Proposition 4.4.4 implies that $\overline{\lambda} \in \operatorname{spec}(A^*)$. Next, note that, since $Ax = \lambda x$, $x^*A^* = \overline{\lambda}x^*$, and $AA^* = A^*A$, it follows that

$$\begin{aligned}(A^*x - \overline{\lambda}x)^*(A^*x - \overline{\lambda}x) &= x^*AA^*x - \overline{\lambda}x^*Ax - \lambda x^*A^*x + \lambda\overline{\lambda}x^*x \\ &= x^*A^*Ax - \lambda\overline{\lambda}x^*x - \lambda\overline{\lambda}x^*x + \lambda\overline{\lambda}x^*x \\ &= \lambda\overline{\lambda}x^*x - \lambda\overline{\lambda}x^*x = 0.\end{aligned}$$

Hence, $A^*x = \overline{\lambda}x$. □

Proposition 4.5.3. Let $A \in \mathbb{F}^{n\times n}$. Then, eigenvectors associated with distinct eigenvalues of A are linearly independent. If, in addition, A is normal, then these eigenvectors are mutually orthogonal.

Proof. Let $\lambda_1, \lambda_2 \in \operatorname{spec}(A)$ be distinct with associated eigenvectors $x_1, x_2 \in \mathbb{C}^n$. Suppose that x_1 and x_2 are linearly dependent, that is, $x_1 = \alpha x_2$, where $\alpha \in \mathbb{C}$ and $\alpha \neq 0$. Then, $Ax_1 = \lambda_1 x_1 = \lambda_1\alpha x_2$, but also $Ax_1 = A\alpha x_2 = \alpha\lambda_2 x_2$. Hence, $\alpha(\lambda_1 - \lambda_2)x_2 = 0$, which contradicts $\alpha \neq 0$. Since pairwise linearly independence does not imply the linear independence of larger sets, next, let $\lambda_1, \lambda_2, \lambda_3 \in \operatorname{spec}(A)$ be distinct with associated eigenvectors $x_1, x_2, x_3 \in \mathbb{C}^n$. Suppose that x_1, x_2, x_3 are linearly dependent. In this case, there exist $a_1, a_2, a_3 \in \mathbb{C}$, not all zero, such that

$a_1x_1 + a_2x_2 + a_3x_3 = 0$. If $a_1 = 0$, then $a_2x_2 + a_3x_3 = 0$. But $\lambda_2 \neq \lambda_3$ implies that x_2 and x_3 are linearly independent, which in turn implies that $a_2 = 0$ and $a_3 = 0$. Since a_1, a_2, a_3 are not all zero, it follows that $a_1 \neq 0$. Therefore, $x_1 = \alpha x_2 + \beta x_3$, where $\alpha \triangleq -a_2/a_1$ and $\beta \triangleq -a_3/a_1$ are not both zero. Thus, $Ax_1 = A(\alpha x_2 + \beta x_3) = \alpha A x_2 + \beta A x_3 = \alpha\lambda_2 x_2 + \beta\lambda_3 x_3$. But, $Ax_1 = \lambda_1 x_1 = \lambda_1(\alpha x_2 + \beta x_3) = \alpha\lambda_1 x_2 + \beta\lambda_1 x_3$. Subtracting these relations yields $0 = \alpha(\lambda_1 - \lambda_2)x_2 + \beta(\lambda_1 - \lambda_3)x_3$. Since x_2 and x_3 are linearly independent, it follows that $\alpha(\lambda_1 - \lambda_2) = 0$ and $\beta(\lambda_1 - \lambda_3) = 0$. Since α and β are not both zero, it follows that $\lambda_1 = \lambda_2$ or $\lambda_1 = \lambda_3$, which contradicts the assumption that $\lambda_1, \lambda_2, \lambda_3$ are distinct. The same arguments apply to sets of four or more eigenvectors.

Now, suppose that A is normal and let $\lambda_1, \lambda_2 \in \operatorname{spec}(A)$ be distinct eigenvalues with associated eigenvectors $x_1, x_2 \in \mathbb{C}^n$. Then, by Lemma 4.5.2, $Ax_1 = \lambda_1 x_1$ implies that $A^*x_1 = \overline{\lambda}_1 x_1$. Consequently, $x_1^*A = \lambda_1 x_1^*$, which implies that $x_1^*Ax_2 = \lambda_1 x_1^* x_2$. Furthermore, $x_1^*Ax_2 = \lambda_2 x_1^* x_2$. It thus follows that $0 = (\lambda_1 - \lambda_2)x_1^*x_2$. Hence, $\lambda_1 \neq \lambda_2$ implies that $x_1^*x_2 = 0$. □

If $A \in \mathbb{R}^{n\times n}$ is symmetric, then Lemma 4.5.2 is not needed and the proof of Proposition 4.5.3 is simpler. In this case, it follows from x) of Proposition 4.4.4 that $\lambda_1, \lambda_2 \in \operatorname{spec}(A)$ are real and thus associated eigenvectors $x_1 \in \mathcal{N}(\lambda_1 I - A)$ and $x_2 \in \mathcal{N}(\lambda_1 I - A)$ can be chosen to be real. Hence, $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ imply that $x_2^{\mathrm{T}}Ax_1 = \lambda_1 x_2^{\mathrm{T}} x_1$ and $x_1^{\mathrm{T}}Ax_2 = \lambda_2 x_1^{\mathrm{T}} x_2$. Since $x_1^{\mathrm{T}}Ax_2 = x_2^{\mathrm{T}}A^{\mathrm{T}}x_1 = x_2^{\mathrm{T}}Ax_1$ and $x_1^{\mathrm{T}}x_2 = x_2^{\mathrm{T}}x_1$, it follows that $(\lambda_1 - \lambda_2)x_1^{\mathrm{T}}x_2 = 0$. Since $\lambda_1 \neq \lambda_2$, it follows that $x_1^{\mathrm{T}}x_2 = 0$.

We define the *spectral abscissa* of $A \in \mathbb{F}^{n\times n}$ by

$$\operatorname{spabs}(A) \triangleq \max\{\operatorname{Re}\lambda\colon\ \lambda \in \operatorname{spec}(A)\} \tag{4.5.2}$$

and the *spectral radius* of $A \in \mathbb{F}^{n\times n}$ by

$$\operatorname{sprad}(A) \triangleq \max\{|\lambda|\colon\ \lambda \in \operatorname{spec}(A)\}. \tag{4.5.3}$$

Let $A \in \mathbb{F}^{n\times n}$. Then, $\nu_-(A), \nu_0(A)$, and $\nu_+(A)$ denote the number of eigenvalues of A counting algebraic multiplicity having, respectively, negative, zero, and positive real part. Define the *inertia* of A by

$$\operatorname{In}(A) \triangleq \begin{bmatrix} \nu_-(A) \\ \nu_0(A) \\ \nu_+(A) \end{bmatrix}. \tag{4.5.4}$$

Note that $\operatorname{spabs}(A) < 0$ if and only if $\nu_-(A) = n$.

4.6 Minimal Polynomial

As we showed in Theorem 4.4.6, every square matrix $A \in \mathbb{F}^{n\times n}$ is a root of its characteristic polynomial. However, there may be polynomials of degree less than n having A as a root. In fact, the following result shows that there exists a unique monic polynomial that has A as a root and that divides all polynomials that have A as a root.

Theorem 4.6.1. Let $A \in \mathbb{F}^{n\times n}$. Then, there exists a unique monic polynomial $\mu_A \in \mathbb{F}[s]$ of minimal degree such that $\mu_A(A) = 0$. Furthermore, $\deg \mu_A \le n$, and μ_A divides every polynomial $p \in \mathbb{F}[s]$ satisfying $p(A) = 0$.

Proof. Since $\chi_A(A) = 0$ and $\deg \chi_A = n$, it follows that there exists a minimal positive integer $n_0 \le n$ such that there exists a monic polynomial $p_0 \in \mathbb{F}[s]$ satisfying $p_0(A) = 0$ and $\deg p_0 = n_0$. Let $p \in \mathbb{F}[s]$ satisfy $p(A) = 0$. Then, by Lemma 4.1.2, there exist $q, r \in \mathbb{F}[s]$ such that $p = qp_0 + r$ and $\deg r < \deg p_0$. However, $p(A) = p_0(A) = 0$ implies that $r(A) = 0$. If $r \ne 0$, then r can be normalized to obtain a monic polynomial of degree less than n_0, which contradicts the definition n_0. Hence, $r = 0$, which implies that p_0 divides p. This proves existence.

Now, suppose there exist two monic polynomials $p_0, \hat{p}_0 \in \mathbb{F}[s]$ of degree n_0 and such that $p_0(A) = \hat{p}_0(A) = 0$. By the previous argument, p_0 divides $\hat{p}_0$, and vice versa. Therefore, p_0 is a constant multiple of $\hat{p}_0$. Since p_0 and $\hat{p}_0$ are both monic, it follows that $p_0 = \hat{p}_0$. This proves uniqueness. Denote this polynomial by μ_A. □

The monic polynomial μ_A of least order having A as a root is the *minimal polynomial* of A.

The following result relates the characteristic and minimal polynomials of $A \in \mathbb{F}^{n\times n}$ to the similarity invariants of A. Note that $\operatorname{rank}(sI - A) = n$, so that A has n similarity invariants $p_1, \ldots, p_n \in \mathbb{F}[s]$. In this case, (4.3.1) becomes

$$sI - A = S_1(s)\begin{bmatrix} p_1(s) & & \\ & \ddots & \\ & & p_n(s) \end{bmatrix} S_2(s), \tag{4.6.1}$$

where $S_1, S_2 \in \mathbb{F}^{n\times n}[s]$ are unimodular and p_i divides p_{i+i} for all $i = 1, \ldots, n-1$.

Proposition 4.6.2. Let $A \in \mathbb{F}^{n\times n}$, and let $p_1, \ldots, p_n \in \mathbb{F}[s]$ be the similarity invariants of A, where p_i divides p_{i+1} for all $i = 1, \ldots, n-1$.

Then,

$$\chi_A = \prod_{i=1}^{n} p_i \tag{4.6.2}$$

and

$$\mu_A = p_n. \tag{4.6.3}$$

Proof. Using Theorem 4.3.2 and (4.6.1) it follows that

$$\chi_A(s) = \det(sI - A) = [\det S_1(s)]\,[\det S_2(s)] \prod_{i=1}^{n} p_i(s).$$

Since S_1 and S_2 are unimodular and χ_A and $p_1, \ldots, p_n$ are monic, it follows that $[\det S_1(s)][\det S_2(s)] = 1$, which proves (4.6.2).

To prove (4.6.3), first note that it follows from Theorem 4.3.2 that $\chi_A = \Delta_{n-1} p_n$, where $\Delta_{n-1} \in \mathbb{F}[s]$ is the greatest common divisor of all $(n-1) \times (n-1)$ subdeterminants of $sI - A$. Since the $(n-1) \times (n-1)$ subdeterminants of $sI - A$ are the entries of $\pm(sI - A)^{\mathrm{A}}$, it follows that Δ_{n-1} divides every entry of $(sI - A)^{\mathrm{A}}$. Hence, there exists a polynomial matrix $P \in \mathbb{F}^{n \times n}[s]$ such that $(sI - A)^{\mathrm{A}} = \Delta_{n-1}(s)P(s)$. Furthermore, since $(sI - A)^{\mathrm{A}}(sI - A) = \chi_A(s)I$, it follows that $\Delta_{n-1}(s)P(s)(sI - A) = \chi_A(s)I = \Delta_{n-1}(s)p_n(s)I$, and thus $P(s)(sI - A) = p_n(s)I$. Lemma 4.2.2 now implies that $p_n(A) = 0$.

Since $p_n(A) = 0$, it follows from Theorem 4.6.1 that μ_A divides p_n. Hence, let $q \in \mathbb{F}[s]$ be the monic polynomial satisfying $p_n = q\mu_A$. Furthermore, since $\mu_A(A) = 0$, it follows from Corollary 4.2.3 that there exists a polynomial matrix $Q \in \mathbb{F}^{n \times n}[s]$ such that $\mu_A(s)I = Q(s)(sI - A)$. Thus, $P(s)(sI - A) = p_n(s)I = q(s)\mu_A(s)I = q(s)Q(s)(sI - A)$, which implies that $P = qQ$. Thus, q divides every entry of P. However, since P was obtained by dividing $(sI - A)^{\mathrm{A}}$ by the greatest common divisor of all of its entries, it follows that the greatest common divisor of the entries of P is 1. Hence, $q = 1$, which implies that $p_n = \mu_A$, which proves (4.6.3). □

Proposition 4.6.2 shows that μ_A divides χ_A, which is also a consequence of Theorem 4.4.6 and Theorem 4.6.1. Proposition 4.6.2 also shows that $\mu_A = \chi_A$ if and only if $p_1 = \cdots = p_{n-1} = 1$, that is, if and only if $p_n = \chi_A$ is the only nonconstant similarity invariant of A. Note that, in general, it follows from (4.6.2) that $\sum_{i=1}^{n} \deg p_i = n$.

Finally, note that the similarity invariants of the $n \times n$ identity matrix I_n are given by $p_i(s) = s - 1$ for all $i = 1, \ldots, n$. Thus, $\chi_{I_n}(s) = (s-1)^n$ and $\mu_{I_n}(s) = s - 1$.

Proposition 4.6.3. Let $A \in \mathbb{F}^{n \times n}$, and assume that A and B are similar. Then,

$$\mu_A = \mu_B. \tag{4.6.4}$$

4.7 Rational Transfer Functions and the Smith-McMillan Decomposition

We now turn our attention to rational functions.

Definition 4.7.1. The set $\mathbb{F}(s)$ of *rational functions* consists of functions $g\colon \ \mathbb{C}\backslash\mathcal{S} \mapsto \mathbb{C}$, where $g(s) = p(s)/q(s)$, $p, q \in \mathbb{F}[s]$ are coprime, $q \neq 0$, and $\mathcal{S} \triangleq \operatorname{roots}(q)$. The rational function g is *strictly proper, proper, exactly proper, improper*, respectively, if $\deg p < \deg q$, $\deg p \le \deg q$, $\deg p = \deg q$, $\deg p > \deg q$. The *relative degree* of g, denoted by $\operatorname{reldeg} g$, is $\deg q - \deg p$. Finally, the roots of p are the *zeros* of g, while the roots of the denominator q are the *poles* of g.

Definition 4.7.2. The set $\mathbb{F}^{n\times m}(s)$ of *rational transfer functions* consists of matrices whose entries are elements of $\mathbb{F}(s)$. The rational transfer function $G \in \mathbb{F}^{n\times m}(s)$ is *strictly proper* if every entry of G is strictly proper, *proper* if every entry of G is proper, *exactly proper* if every entry of G is proper and at least one entry of G is exactly proper, and *improper* if at least one entry of G is improper. The *relative degree* of $G \in \mathbb{F}^{n\times m}(s)$, denoted by $\operatorname{reldeg} G$, is defined by

$$\operatorname{reldeg} G \triangleq \min_{\substack{i=1,\ldots,n\\ j=1,\ldots,m}} \operatorname{reldeg} G_{(i,j)}. \tag{4.7.1}$$

By writing $(sI - A)^{-1}$ as

$$(sI - A)^{-1} = \frac{1}{\chi_A(s)}(sI - A)^{\mathrm{A}}, \tag{4.7.2}$$

it follows from (4.4.18) that $(sI - A)^{-1}$ is a strictly proper rational transfer function. In fact, for all $i = 1, \ldots, n$,

$$\operatorname{reldeg} \left[(sI - A)^{-1}\right]_{(i,i)} = n - 1, \tag{4.7.3}$$

and thus

$$\operatorname{reldeg}\, (sI - A)^{-1} = n - 1. \tag{4.7.4}$$

The following result provides a canonical form, known as the *Smith-McMillan form*, for rational transfer functions under unimodular transformation. The following definition is an extension of Definition 4.2.4 for matrix polynomials.

Definition 4.7.3. Let $G \in \mathbb{F}^{n\times m}(s)$, and let $\mathcal{S}$ be as defined in Definition 4.7.2. Then, the rank of G is the nonnegative integer

$$\operatorname{rank} G \triangleq \max_{s\in\mathbb{C}\backslash\mathcal{S}} \operatorname{rank} G(s). \tag{4.7.5}$$

Theorem 4.7.4. Let $G \in \mathbb{F}^{n\times m}(s)$ and let $r \triangleq \operatorname{rank} G$. Then, there exist unimodular matrices $S_1 \in \mathbb{F}^{n\times n}[s]$ and $S_2 \in \mathbb{F}^{m\times m}[s]$ and monic polynomials $p_1,\ldots,p_r,q_1,\ldots,q_r \in \mathbb{F}[s]$ such that p_i and q_i are coprime for all $i = 1,\ldots,r$, p_i divides p_{i+1} for all $i = 1,\ldots,r-1$, q_{i+1} divides q_i for all $i = 1,\ldots,r-1$, and

$$G = S_1 \begin{bmatrix} p_1/q_1 & & & \\ & \ddots & & \\ & & p_r/q_r & \\ & & & 0_{(n-r)\times(m-r)} \end{bmatrix} S_2. \tag{4.7.6}$$

Proof. Let n_{ij}/d_{ij} denote the (i,j) entry of G, where $n_{ij}, d_{ij} \in \mathbb{F}[s]$ are coprime, and let $d \in \mathbb{F}[s]$ denote the least common multiple of d_{ij} for all $i = 1,\ldots,n$, and $j = 1,\ldots,m$. From Theorem 4.3.2 it follows that the polynomial matrix dG has a Smith form $\operatorname{diag}(\hat{p}_1,\ldots,\hat{p}_r, 0,\ldots,0)$, where $\hat{p}_1,\ldots,\hat{p}_r \in \mathbb{F}[s]$ and $\hat{p}_i$ divides $\hat{p}_{i+1}$ for all $i = 1,\ldots,r-1$. Now, divide this Smith form by d and express every rational function $\hat{p}_i/d$ in coprime form p_i/q_i so that p_i divides p_{i+1} for all $i = 1,\ldots,r-1$ and q_{i+1} divides q_i for all $i = 1,\ldots,r-1$. □

Let $g_1,\ldots,g_r \in \mathbb{F}^n(s)$. Then, $g_1,\ldots,g_r$ are *linearly independent* if $\alpha_1,\ldots,\alpha_r \in \mathbb{F}[s]$ and $\sum_{n=1}^r \alpha_i g_i = 0$ imply that $\alpha_1 = \cdots = \alpha_r = 0$. It can be seen that this definition is unchanged if $\alpha_1,\ldots,\alpha_r \in \mathbb{F}(s)$.

Proposition 4.7.5. Let $G \in \mathbb{F}^{n\times m}(s)$. Then, $\operatorname{rank} G$ is equal to the number of linearly independent columns of G.

As a special case, Proposition 4.7.5 applies to polynomial matrices $G \in \mathbb{F}^{n\times m}[s]$.

Definition 4.7.6. Let $G \in \mathbb{F}^{n\times m}(s)$, let $r \triangleq \operatorname{rank} G$, and let $p_1,\ldots,p_r$, $q_1,\ldots,q_r \in \mathbb{F}[s]$ be given by Theorem 4.7.4. Then, the *McMillan degree* of G is $\sum_{i=1}^r \deg q_i$. Furthermore, the *poles* of G are the roots of q_1, the *transmission zeros* of G are the roots of p_r, and the *blocking zeros* of G are the roots of p_1.

4.8 Facts on Polynomials

Fact 4.8.1. Let $p \in \mathbb{R}[s]$ be monic and define $q(s) \triangleq s^n p(1/s)$, where $n \triangleq \deg p$. If $0 \notin \operatorname{roots}(p)$, then $\deg(q) = n$ and

$$\operatorname{mroots}(q) = \{1/\lambda\colon\ \lambda \in \operatorname{mroots}(p)\}_{\mathrm{m}}.$$

If $0 \in \operatorname{roots}(p)$ with multiplicity r, then $\deg(q) = n - r$ and

$$\operatorname{mroots}(q) = \{1/\lambda\colon\ \lambda \neq 0 \text{ and } \lambda \in \operatorname{mroots}(p)\}_{\mathrm{m}}.$$

(Remark: See Fact 11.13.3 and Fact 11.13.4.)

Fact 4.8.2. Let $p \in \mathbb{F}^n$ be given by

$$p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0,$$

let $\beta_n \triangleq 1$, let $\operatorname{mroots}(p) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$, and define $\mu_1, \ldots, \mu_n$ by

$$\mu_i \triangleq \lambda_1^i + \cdots + \lambda_n^i.$$

Then, for all $k = 1, \ldots, n$,

$$k\beta_{n-k} + \mu_1\beta_{n-k+1} + \mu_2\beta_{n-k+2} + \cdots \mu_k\beta_n = 0.$$

That is,

$$\begin{bmatrix} n & \mu_1 & \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_n \\ 0 & n-1 & \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n-1} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 2 & \mu_1 & \mu_2 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \mu_1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n-1} \\ \beta_n \end{bmatrix} = 0.$$

Consequently, $\beta_1, \ldots, \beta_{n-1}$ are uniquely determined by $\mu_1, \ldots, \mu_n$. In particular,

$$\beta_{n-1} = -\mu_1$$

and

$$\beta_{n-2} = \tfrac{1}{2}\big[\mu_1^2 - \mu_2\big].$$

(Proof: See [287, p. 44] and [419, p. 9].) (Remark: These equations are *Newton's identities.*)

Fact 4.8.3. Let $p, q \in \mathbb{F}[s]$ be monic. Then, p and q are coprime if and only if their least common multiple is pq.

Fact 4.8.4. Let $p, q \in \mathbb{F}[s]$, where $p(s) = a_n s^n + \cdots + a_1 s + a_0$, $q(s) = b_m s^m + \cdots + b_1 s + b_0$, $\deg p = n$, and $\deg q = m$. Furthermore, define the

Toeplitz matrices $[p]^{(m)} \in \mathbb{F}^{m\times(n+m)}$ and $[q]^{(n)} \in \mathbb{F}^{n\times(n+m)}$ by

$$[p]^{(m)} \triangleq \begin{bmatrix} a_n & a_{n-1} & \cdots & a_1 & a_0 & 0 & 0 & \cdots & 0 \\ 0 & a_n & a_{n-1} & \cdots & a_1 & a_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}$$

and

$$[q]^{(n)} \triangleq \begin{bmatrix} b_m & b_{m-1} & \cdots & b_1 & b_0 & 0 & 0 & \cdots & 0 \\ 0 & b_m & b_{m-1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}.$$

Then, p and q are coprime if and only if

$$\det \begin{bmatrix} [p]^{(m)} \\ [q]^{(n)} \end{bmatrix} \neq 0.$$

(Proof: See [202, p. 162] or [466, pp. 187–191].) (Remark: $\left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]$ is the *Sylvester matrix*, and $\det \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right]$ is the *resultant* of p and q.) (Remark: The form $\left[\begin{smallmatrix} [p]^{(m)} \\ [q]^{(n)} \end{smallmatrix}\right]$ appears in [466, pp. 187–191]. The result is given in [202, p. 162] in terms of $\left[\begin{smallmatrix} \hat{I}[p]^{(m)} \\ \hat{I}[q]^{(n)} \end{smallmatrix}\right]\hat{I}$ and in [633, p. 85] in terms of $\left[\begin{smallmatrix} [p]^{(m)} \\ \hat{I}[q]^{(n)} \end{smallmatrix}\right]$.)

Fact 4.8.5. Let $p_1, \ldots, p_n \in \mathbb{F}[s]$, and let $d \in \mathbb{F}[s]$ be the greatest common divisor of $p_1, \ldots, p_n$. Then, there exist $q_1, \ldots, q_n \in \mathbb{F}[s]$ such that

$$d = \sum_{i=1}^{n} q_i p_i.$$

In addition, $p_1, \ldots, p_n$ are coprime if and only if there exist $q_1, \ldots, q_n \in \mathbb{F}[s]$ such that

$$1 = \sum_{i=1}^{n} q_i p_i.$$

(Proof: See [216, p. 16].) (Remark: The polynomial d is given by the *Bezout equation.*)

Fact 4.8.6. Let $p, q \in \mathbb{F}[s]$, where $p(s) = a_n s^n + \cdots + a_1 s + a_0$ and $q(s) = b_n s^n + \cdots + b_1 s + b_0$, and define $[p]^{(n)}, [q]^{(n)} \in \mathbb{F}^{n\times 2n}$ as in Fact 4.8.4. Furthermore, define

$$R(p,q) \triangleq \begin{bmatrix} [p]^{(m)} \\ [q]^{(n)} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix},$$

where $A_1, A_2, B_1, B_2 \in \mathbb{F}^{n\times n}$, and define $\hat{p}(s) \triangleq s^n p(-s)$ and $\hat{q}(s) \triangleq s^n q(-s)$.

Then,

$$\begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix} = \begin{bmatrix} \hat{p}\big(N^{\mathrm{T}}\big) & p(N) \\ \hat{q}\big(N^{\mathrm{T}}\big) & q(N) \end{bmatrix},$$

$$A_1B_1 = B_1A_1,$$

$$A_2B_2 = B_2A_2,$$

$$A_1B_2 + A_2B_1 = B_1A_2 + B_2A_1.$$

Therefore,

$$\begin{bmatrix} I & 0 \\ -B_1 & A_1 \end{bmatrix}\begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ 0 & A_1B_2 - B_1A_2 \end{bmatrix},$$

$$\begin{bmatrix} -B_2 & A_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix} = \begin{bmatrix} A_2B_1 - B_2A_1 & 0 \\ B_1 & B_2 \end{bmatrix},$$

and

$$\det R(p,q) = \det(A_1B_2 - B_1A_2) = \det(B_2A_1 - A_2B_1).$$

Now, define $B(p,q) \in \mathbb{F}^{n \times n}$ by

$$B(p,q) \triangleq (A_1B_2 - B_1A_2)\hat{I}.$$

Then, the following statements hold:

i) For all $s, \hat{s} \in \mathbb{C}$,

$$p(s)q(\hat{s}) - q(s)p(\hat{s}) = (s - \hat{s})\begin{bmatrix} 1 \\ s \\ \vdots \\ s^{n-1} \end{bmatrix}^{\mathrm{T}} B(p,q)\begin{bmatrix} 1 \\ \hat{s} \\ \vdots \\ \hat{s}^{n-1} \end{bmatrix}$$

ii) $B(p,q) = (B_2A_1 - A_2B_1)\hat{I} = \hat{I}\big(A_1^{\mathrm{T}}B_2^{\mathrm{T}} - B_1^{\mathrm{T}}A_2^{\mathrm{T}}\big) = \hat{I}\big(B_1^{\mathrm{T}}A_2^{\mathrm{T}} - A_1^{\mathrm{T}}B_2^{\mathrm{T}}\big)$.

iii) $\begin{bmatrix} 0 & B(p,q) \\ -B(p,q) & \end{bmatrix} = QR^{\mathrm{T}}(p,q)QR(p,q)Q$, where $Q \triangleq \begin{bmatrix} 0 & \hat{I} \\ -\hat{I} & 0 \end{bmatrix}$.

iv) $|\det B(p,q)| = |\det R(p,q)| = |\det q[C(p)]|$.

v) $B(p,q)$ and $\hat{B}(p,q)$ are symmetric.

vi) $B(p,q)$ is a linear function of (p,q).

vii) $B(p,q) = -B(q,p)$.

Now, assume that $\deg q \le \deg p = n$ and p is monic. Then, the following statements hold:

viii) $\operatorname{def} B(p,q)$ is equal to the degree of the greatest common divisor of p and q.

ix) p and q are coprime if and only if $B(p,q)$ is nonsingular.

x) If $B(p,q)$ is nonsingular, then $[B(p,q)]^{-1}$ is Hankel. In fact,

$$[B(p,q)]^{-1} = H(a/p),$$

where $a, b \in \mathbb{F}[s]$ satisfy the Bezout equation $aq + bp = 1$.

xi) If $q = q_1 q_2$, where $q_1, q_2 \in \mathbb{F}[s]$, then

$$B(p,q) = B(p,q_1)q_2[C(p)] = q_1\big[C^{\mathrm{T}}(p)\big]B(p,q_2).$$

xii) $B(p,q) = B(p,q)C(p) = C^{\mathrm{T}}(p)B(p,q)$.

xiii) $B(p,q) = B(p,1)q[C(p)] = q[C^{\mathrm{T}}(p)]B(p,1)$, where $B(p,1)$ is the Hankel matrix

$$B(p,1) = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & 1 \\ a_2 & a_3 & \iddots & 1 & 0 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ a_{n-1} & 1 & \iddots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

In particular, for $n = 3$ and $q(s) = s$, it follows that

$$\begin{bmatrix} -a_0 & 0 & 0 \\ 0 & a_2 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & 1 \\ a_2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}.$$

xiv) $$\begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix} = \begin{bmatrix} 0 & I \\ A_2^{-1}\hat{I} & B_2A_2^{-1} \end{bmatrix} \begin{bmatrix} B(p,q) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ A_1 & A_2 \end{bmatrix}.$$

xv) If p has distinct roots $\lambda_1, \ldots, \lambda_n$, then

$$V^{\mathrm{T}}(\lambda_1, \ldots, \lambda_n)B(p,q)V(\lambda_1, \ldots, \lambda_n) = \operatorname{diag}[q(\lambda_1)p'(\lambda_1), \ldots, q(\lambda_n)p'(\lambda_n)].$$

(Proof: See [202, pp. 164–167], [273], and [216, pp. 200–207]. To prove *ii)*, note that A_1, A_2, B_1, B_2 are square and Toeplitz, and thus reverse symmetric, that is, $A_1 = A_1^{\hat{\mathrm{T}}}$. See Fact 3.12.6.) (Remark: $B(p,q)$ is a *Bezout matrix.* See [65, 298], [466, p. 189], [566], and Fact 5.13.22.) (Remark: *xiii)* is the *Barnett factorization.* See [59, 566]. The definition of $B(p,q)$ and *ii)* are the *Gohberg-Semencul formulas.* See [216, p. 206].) (Remark: It follows from continuity that the determinant expressions are valid if A_1 or B_2 is singular. See Fact 2.12.16.) (Remark: The inverse of a Hankel matrix is a Bezout matrix. See [202, p. 174].)

Fact 4.8.7. Let $p, q \in \mathbb{F}[s]$, assume that q is monic, and $\deg p < \deg q = n$. Furthermore, define $g \in \mathbb{F}(s)$ by

$$g(s) \triangleq \frac{p(s)}{q(s)} = \sum_{i=1}^{\infty} \frac{g_i}{s^i}.$$

Finally, define the Hankel matrix

$$H(g) \triangleq \begin{bmatrix} g_1 & g_2 & \cdots & g_{n-1} & g_n \\ g_2 & g_3 & \iddots & g_n & g_{n+1} \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ g_{n-1} & g_n & \iddots & g_{2n-3} & g_{2n-2} \\ g_n & g_{n+1} & \cdots & g_{2n-2} & g_{2n-1} \end{bmatrix}.$$

Then, the following statements hold:

i) p and q are coprime if and only if $H(g)$ is nonsingular.

ii) If p and q are coprime, then $[H(g)]^{-1} = B(q,a)$, where $a,b \in \mathbb{F}[s]$ satisfy the Bezout equation $ap + bq = 1$.

iii) $B(q,p) = B(q,1)H(g)B(q,1)$.

iv) $B(q,p)$ and $H(g)$ are congruent.

v) $\operatorname{In} B(q,p) = \operatorname{In} H(g)$.

vi) $\det H(g) = \det B(q,p)$.

(Proof: See [216, pp. 215–221].)

Fact 4.8.8. Let $p \in \mathbb{R}[s]$, and define $g \in \mathbb{F}(s)$ by $g \triangleq q'/q$. Then, the following statements hold:

i) The number of distinct roots of q is $\operatorname{rank} B(q,q')$.

ii) q has n distinct roots if and only if $B(q,q')$ is nonsingular.

iii) The number of distinct real roots of q is $\operatorname{sig} B(q,q')$.

iv) q has n distinct, real roots if and only if $B(q,q')$ is positive definite.

v) The number of distinct complex roots of q is $2\nu_-[B(q,q')]$.

vi) q has n distinct, complex roots if and only if n is even and $\nu_-[B(q,q')] = n/2$.

vii) q has n real roots if and only if $B(q,q')$ is nonnegative semidefinite.

(Proof: See [216, p. 252].) (Remark: $q'(s) \triangleq (\mathrm{d}/\mathrm{d}s)q(s)$.)

Fact 4.8.9. Let $q \in \mathbb{F}[s]$, where $q(s) = \sum_{i=0}^{n} b_i s^i$, and define

$$\operatorname{coeff}(q) \triangleq \begin{bmatrix} b_n \\ \vdots \\ b_0 \end{bmatrix}.$$

Now, let $p \in \mathbb{F}[s]$, where $p(s) = \sum_{i=0}^n a_i s^i$. Then,

$$\operatorname{coeff}(pq) = A\operatorname{coeff}(q),$$

where $A \in \mathbb{F}^{2n\times(n+1)}$ is the Toeplitz matrix

$$A = \begin{bmatrix} a_n & 0 & 0 & \cdots & 0 \\ a_{n-1} & a_n & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ a_0 & a_1 & \cdots & \cdots & a_n \\ 0 & a_0 & \ddots & \cdots & a_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_0 & a_1 \end{bmatrix}.$$

In particular, if $n = 3$, then

$$A = \begin{bmatrix} a_2 & 0 & 0 \\ a_1 & a_2 & 0 \\ a_0 & a_1 & a_2 \\ 0 & a_0 & a_1 \end{bmatrix}.$$

Fact 4.8.10. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ be distinct and, for all $i = 1, \ldots, n$, define

$$p_i(s) \triangleq \prod_{\substack{j=1 \\ j\neq i}}^n \frac{s - \lambda_i}{\lambda_i - \lambda_j}.$$

Then, for all $i = 1, \ldots, n$,

$$p_i(\lambda_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

(Remark: This identity is the *Lagrange interpolation formula.*)

Fact 4.8.11. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\det(I + A) \neq 0$. Then, there exists a polynomial p of degree less than or equal to $n-1$ such that $(I + A)^{-1} = p(A)$.

Fact 4.8.12. indexPfaffian!skew-symmetric matrix!Fact 4.8.12Let $A \in \mathbb{R}^{n\times n}$ be skew symmetric and let the components of $x_A \in \mathbb{R}^{n(n-1)/2}$ be the entries $A_{(i,j)}$ for all $i > j$. Then, there exists a polynomial function $p\colon\ \mathbb{R}^{n(n-1)/2} \mapsto \mathbb{R}$ such that, for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^{n(n-1)/2}$,

$$p(\alpha x) = \alpha^{n/2} p(x)$$

and

$$\det A = p^2(x_A).$$

In particular,

$$\det\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix} = a^2$$

and

$$\det\begin{bmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{bmatrix} = (af - be + cd)^2.$$

(Proof: See [356, p. 224] and [466, pp. 125–127].) (Remark: The polynomial p is the *Pfaffian*, and this result is *Pfaff's theorem.*)

Fact 4.8.13. Let $G \in \mathbb{F}^{n\times m}(s)$, and let $G_{(i,j)} = n_{ij}/d_{ij}$, where $n_{ij} \in \mathbb{F}[s]$ and $d_{ij} \in \mathbb{F}[s]$ are coprime for all $i = 1,\dots,n$ and $j = 1,\dots,m$. Then, q_1 given by the Smith-McMillan form is the least common multiple of $d_{11}, d_{12}, \dots, d_{nm}$.

Fact 4.8.14. Let $G \in \mathbb{F}^{n\times m}(s)$, assume that $\operatorname{rank} G = m$, and let $\lambda \in \mathbb{C}$, where λ is not a pole of G. Then, λ is a transmission zero of G if and only if there exists $u \in \mathbb{C}^m$ such that $G(\lambda)u = 0$. Furthermore, if G is square, then λ is a transmission zero of G if and only if $\det G(\lambda) = 0$.

4.9 Facts on the Characteristic and Minimal Polynomials

Fact 4.9.1. Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathbb{R}^{2\times 2}$. Then, the following identities hold:

i) $$\operatorname{mspec}(A) = \left\{\tfrac{1}{2}\Big[a + d \pm \sqrt{(a-d)^2 + 4bc}\Big]\right\}_{\mathrm{m}} = \left\{\tfrac{1}{2}\Big[\operatorname{tr} A \pm \sqrt{(\operatorname{tr} A)^2 - 4\det A}\Big]\right\}_{\mathrm{m}}.$$

ii) $\chi_A(s) = s^2 - (\operatorname{tr} A)s + \det A.$

iii) $\det A = \frac{1}{2}\big[(\operatorname{tr} A)^2 - \operatorname{tr} A^2\big].$

iv) $(sI - A)^{\mathrm{A}} = sI + A - (\operatorname{tr} A)I.$

v) $A^{-1} = (\det A)^{-1}[(\operatorname{tr} A)I - A].$

vi) $A^{\mathrm{A}} = (\operatorname{tr} A)I - A.$

vii) $\operatorname{tr} A^{-1} = \operatorname{tr} A/\det A.$

Fact 4.9.2. Let $A, B \in \mathbb{F}^{2\times 2}$. Then,

$$AB + BA - (\operatorname{tr} A)B - (\operatorname{tr} B)A + [(\operatorname{tr} A)(\operatorname{tr} B) - \operatorname{tr} AB]I = 0.$$

Furthermore,

$$\det(A+B) - \det A - \det B = (\operatorname{tr} A)(\operatorname{tr} B) - \operatorname{tr}(AB).$$

(Proof: Apply the Cayley-Hamilton theorem to $A + xB$, differentiate with respect to x, and set $x = 0$. For the second identity, evaluate the Cayley-Hamilton theorem with $A + B$. See [211, 212, 364, 483] or [505, p. 37].)

Fact 4.9.3. Let $A \in \mathbb{R}^{3\times 3}$. Then, the following identities hold:

i) $\chi_A(s) = s^3 - (\operatorname{tr} A)s^2 + \big(\operatorname{tr} A^{\mathrm{A}}\big)s - \det A.$

ii) $\operatorname{tr} A^{\mathrm{A}} = \frac{1}{2}\big[(\operatorname{tr} A)^2 - \operatorname{tr} A^2\big].$

iii) $\det A = \frac{1}{3}\operatorname{tr} A^3 - \frac{1}{2}(\operatorname{tr} A)\operatorname{tr} A^2 + \frac{1}{6}(\operatorname{tr} A)^3.$

iv) $(sI-A)^{\mathrm{A}} = s^2 I + s[A - (\operatorname{tr} A)I] + A^2 - (\operatorname{tr} A)A + \frac{1}{2}\big[(\operatorname{tr} A)^2 - \operatorname{tr} A^2\big]I.$

Fact 4.9.4. Let $A, B, C \in \mathbb{F}^{3\times 3}$. Then,

$$\begin{aligned}&\sum \big[A'B'C' - \big(\operatorname{tr} A'\big)B'C' + \big(\operatorname{tr} A'\big)\big(\operatorname{tr} B'\big)C' - \big(\operatorname{tr} A'B'\big)C'\big]\\ &\quad - [(\operatorname{tr} A)(\operatorname{tr} B)\operatorname{tr} C - (\operatorname{tr} A)\operatorname{tr} BC - (\operatorname{tr} B)\operatorname{tr} CA - (\operatorname{tr} C)\operatorname{tr} AB + \operatorname{tr} ABC\\ &\quad + \operatorname{tr} CBA]I = 0,\end{aligned}$$

where the sum is taken over all six permutations A', B', C' of A, B, C. (Remark: This identity is the *polarized Cayley-Hamilton theorem*. See [37, 364, 483].)

Fact 4.9.5. Let $A \in \mathbb{F}^{n\times n}$, and let $\chi_A(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_0$. Then,

$$A^{\mathrm{A}} = (-1)^{n-1}\big(A^{n-1} + \beta_{n-1}A^{n-2} + \cdots + \beta_1 I\big).$$

Furthermore,

$$\operatorname{tr} A^{\mathrm{A}} = (-1)^{n-1}\chi_A'(0) = (-1)^{n-1}\beta_1.$$

(Proof: Use $A^{-1}\chi_A(A) = 0$. The second identity follows from (4.4.16) or Lemma 4.4.7.)

Fact 4.9.6. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, and let $\chi_A(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_0$. Then,

$$\begin{aligned}\chi_{A^{-1}}(s) &= \frac{1}{\det A}(-s)^n\chi_A(1/s)\\ &= s^n + (\beta_1/\beta_0)s^{n-1} + \cdots + (\beta_{n-1}/\beta_0)s + 1/\beta_0.\end{aligned}$$

(Remark: See Fact 5.12.2.)

Fact 4.9.7. Let $A \in \mathbb{F}^{n\times n}$, and assume that either A and $-A$ are similar or A^{T} and $-A$ are similar. Then,

$$\chi_A(s) = (-1)^n \chi_A(-s).$$

Furthermore, if n is even, then χ_A is even, whereas, if n is odd, then χ_A is odd.

Fact 4.9.8. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $s \in \mathbb{C}$,

$$(sI - A)^{\mathrm{A}} = \chi_A(s)(sI - A)^{-1} = \sum_{i=0}^{n-1} \chi_A^{[i]}(s)A^i,$$

where

$$\chi_A(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$$

and, for all $i = 0, \ldots, n-1$, the polynomial $\chi_A^{[i]}$ is defined by

$$\chi_A^{[i]}(s) \triangleq s^{n-i} + \beta_{n-1}s^{n-1-i} + \cdots + \beta_{i+1}.$$

Note that

$$\chi_A^{[n-1]}(s) = s + \beta_{n-1}, \quad \chi_A^{[n]}(s) = 1,$$

and that, for all $i = 0, \ldots, n-1$ and with $\chi_A^{[0]} \triangleq \chi_A$, the polynomials $\chi_A^{[i]}$ satisfy the recursion

$$s\chi_A^{[i+1]}(s) = \chi_A^{[i]}(s) - \beta_i.$$

(Proof: See [615, p. 31].)

Fact 4.9.9. Let $A \in \mathbb{R}^{n\times n}$ be skew symmetric. If n is even, then χ_A is even, whereas, if n is odd, then χ_A is odd.

Fact 4.9.10. Let $A \in \mathbb{F}^{n\times n}$. Then, $\chi_{\mathcal{A}}$ is even for all of the matrices $\mathcal{A}$ given by $\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$, $\begin{bmatrix} A & 0 \\ 0 & -A \end{bmatrix}$, and $\begin{bmatrix} A & 0 \\ 0 & -A^* \end{bmatrix}$.

Fact 4.9.11. Let $A, B \in \mathbb{F}^{n\times n}$, and define $\mathcal{A} \triangleq \begin{bmatrix} 0 & A \\ B & 0 \end{bmatrix}$. Then, $\chi_{\mathcal{A}}(s) = \chi_{AB}(s^2) = \chi_{BA}(s^2)$. Consequently, $\chi_{\mathcal{A}}$ is even. (Proof: Use Fact 2.12.16 and Proposition 4.4.9.)

Fact 4.9.12. Let $x, y, z, w \in \mathbb{F}^n$, and define $A \triangleq xy^{\mathrm{T}}$ and $B \triangleq xy^{\mathrm{T}} + zw^{\mathrm{T}}$. Then,

$$\chi_A(s) = s^{n-1}\big(s - x^{\mathrm{T}}y\big)$$

and

$$\chi_B(s) = s^{n-2}\big[s^2 - \big(x^{\mathrm{T}}y + z^{\mathrm{T}}w\big)s + x^{\mathrm{T}}yz^{\mathrm{T}}w - y^{\mathrm{T}}zx^{\mathrm{T}}w\big].$$

(Remark: See Fact 5.9.8.)

Fact 4.9.13. Let $x,y,z,w \in \mathbb{F}^{n-1}$, and define $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} 1 & x^{\mathrm{T}} \\ y & zw^{\mathrm{T}} \end{bmatrix}.$$

Then,

$$\chi_A(s) = s^{n-3}\big[s^3 - \big(1+w^{\mathrm{T}}z\big)s^2 + \big(w^{\mathrm{T}}z - x^{\mathrm{T}}y\big)s + w^{\mathrm{T}}zx^{\mathrm{T}}y - x^{\mathrm{T}}zw^{\mathrm{T}}y\big].$$

(Proof: See [176].)

Fact 4.9.14. Let $A \in \mathbb{R}^{2n\times 2n}$ be Hamiltonian. Then, χ_A is even.

Fact 4.9.15. Let $A,B,C \in \mathbb{R}^{n\times n}$ and define $\mathcal{A} \triangleq \left[\begin{smallmatrix} A & B \\ C & -A^{\mathrm{T}} \end{smallmatrix}\right]$. If B and C are symmetric, then $\mathcal{A}$ is Hamiltonian. If B and C are skew symmetric, then $\chi_{\mathcal{A}}$ is even, but $\mathcal{A}$ is not necessarily Hamiltonian. (Proof: For the second result replace J_n by $\left[\begin{smallmatrix} 0 & I_n \\ I_n & 0 \end{smallmatrix}\right]$.)

Fact 4.9.16. Let $A \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{n\times m}$, and define $\mathcal{A} \in \mathbb{R}^{2n\times 2n}$ by

$$\mathcal{A} \triangleq \begin{bmatrix} A & BB^{\mathrm{T}} \\ R & -A^{\mathrm{T}} \end{bmatrix}.$$

Then,

$$\chi_{\mathcal{A}}(s) = (-1)^n \chi_A(s)\chi_A(-s)\det\Big[I + B^{\mathrm{T}}\big(-sI - A^{\mathrm{T}}\big)^{-1}R(sI-A)^{-1}B\Big].$$

(Remark: If R is symmetric, then $\mathcal{A}$ is Hamiltonian, and it can be seen directly that $\chi_{\mathcal{A}}$ is even.) If, in addition, R is nonnegative semidefinite, then $(-1)^n\chi_{\mathcal{A}}$ has a spectral factorization. (Proof: Using (2.8.10) and (2.8.14) it follows that, for all $s \notin \pm\operatorname{spec}(A)$,

$$\begin{aligned}\chi_{\mathcal{A}}(s) &= \det(sI-A)\det\big[sI + A^{\mathrm{T}} - R(sI-A)^{-1}BB^{\mathrm{T}}\big] \\ &= (-1)^n\chi_A(s)\chi_A(-s)\det\Big[I - B^{\mathrm{T}}\big(sI + A^{\mathrm{T}}\big)^{-1}R(sI-A)^{-1}B\Big].\end{aligned}$$

To prove the second statement, note that, for $\omega \in \mathbb{R}$ such that $\jmath\omega \notin \operatorname{spec}(A)$, it follows that

$$\chi_{\mathcal{A}}(\jmath\omega) = (-1)^n\chi_A(\jmath\omega)\overline{\chi_A(\jmath\omega)}\det\big[I + B^{\mathrm{T}}(\jmath\omega I - A)^{-*}R(\jmath\omega I - A)^{-1}B\big]$$

and thus $(-1)^n\chi_{\mathcal{A}}(\jmath\omega) \ge 0$. By continuity, this inequality holds for all $\omega \in \mathbb{R}$. Now, Proposition 4.1.1 implies that $(-1)^n\chi_{\mathcal{A}}$ has a spectral factorization.) (Remark: Not all Hamiltonian matrices have this property. Consider $\left[\begin{smallmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \end{smallmatrix}\right]$, which has spectrum $\{\jmath, -\jmath, \sqrt{3}\jmath, -\sqrt{3}\jmath\}$.)

4.10 Facts on the Spectrum

Fact 4.10.1. Let $A \in \mathbb{F}^{n\times n}$, let $p \in \mathbb{F}[s]$, and define $B \triangleq p(A)$. Then, B is nonsingular if and only if $\operatorname{spec}(A) \cap \operatorname{roots}(p) = \varnothing$.

Fact 4.10.2. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. If $\operatorname{tr} A^k = \operatorname{tr} B^k$ for all $k \in \{1,\ldots,\max\{m,n\}\}$, then A and B have the same nonzero eigenvalues with the same algebraic multiplicity. Now, assume that $n = m$. Then, $\operatorname{tr} A^k = \operatorname{tr} B^k$ for all $k \in \{1,\ldots,n\}$ if and only if $\operatorname{mspec}(A) = \operatorname{mspec}(B)$. (Proof: Use *Newton's identities.* See Fact 4.8.2.) (Remark: This result yields Proposition 4.4.9 since $\operatorname{tr}(AB)^k = \operatorname{tr}(BA)^k$ for all $k \in \mathbb{P}$ and for all matrices A and B that are not square.) (Remark: Setting $B = 0_{n\times n}$ yields necessity in Fact 2.11.16.)

Fact 4.10.3. Let $A \in \mathbb{F}^{n\times n}$ and let $\operatorname{mspec}(A) = \{\lambda_1,\ldots,\lambda_n\}_{\mathrm{m}}$. Then,

$$\operatorname{mspec}\big(A^{\mathrm{A}}\big) = \begin{cases} \left\{\dfrac{\det A}{\lambda_1},\ldots,\dfrac{\det A}{\lambda_n}\right\}_{\mathrm{m}}, & \operatorname{rank} A = n,\\ \left\{\displaystyle\sum_{i=1}^{n} \det A_{[i,i]}, 0,\ldots,0\right\}_{\mathrm{m}}, & \operatorname{rank} A = n-1,\\ \{0,\ldots,0\}_{\mathrm{m}}, & \operatorname{rank} A < n-1. \end{cases}$$

(Remark: See Fact 2.13.7 and Fact 5.9.19.)

Fact 4.10.4. Let $a,b,c,d,\omega \in \mathbb{R}$, and define the skew-symmetric matrix $A \in \mathbb{R}^{4\times 4}$ by

$$A \triangleq \begin{bmatrix} 0 & \omega & a & b\\ -\omega & 0 & c & d\\ -a & -c & 0 & \omega\\ -b & -d & -\omega & 0 \end{bmatrix}.$$

Then,

$$\det A = \left[\omega^2 - (ad - bc)\right]^2.$$

Furthermore, A has a repeated eigenvalue if and only if either *i)* A is singular or *ii)* $a = -d$ and $b = c$. In case *i)*, A has the repeated eigenvalue 0, while in case *ii)*, A has the repeated eigenvalues $\jmath\sqrt{\omega^2+a^2+b^2}$ and $-\jmath\sqrt{\omega^2+a^2+b^2}$.

Fact 4.10.5. Let $A \in \mathbb{F}^{n\times n}$, and let $p \in \mathbb{F}[s]$. Then, μ_A divides p if and only if $\operatorname{spec}(A) \subseteq \operatorname{roots}(p)$ and, for all $\lambda \in \operatorname{spec}(A)$, $\operatorname{ind}_A(\lambda) \le \mathrm{m}_p(\lambda)$.

Fact 4.10.6. Let $A \in \mathbb{F}^{n\times n}$, let $\operatorname{mspec}(A) = \{\lambda_1,\ldots,\lambda_n\}_{\mathrm{m}}$, and let $p \in \mathbb{F}[s]$. Then,

$$\operatorname{mspec}[p(A)] = \{p(\lambda_1),\ldots,p(\lambda_n)\}_{\mathrm{m}}.$$

Furthermore, $\mathrm{roots}(p) \cap \mathrm{spec}(A) = \varnothing$ if and only if $p(A)$ is nonsingular. Finally, μ_A divides p if and only if $p(A) = 0$.

Fact 4.10.7. Let $A_1 \in \mathbb{F}^{n\times n}$, $A_{12} \in \mathbb{F}^{n\times m}$, and $A_2 \in \mathbb{F}^{m\times m}$, and define $A \in \mathbb{F}^{(n+m)\times(n+m)}$ by

$$A \triangleq \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}.$$

Then,

$$\chi_A = \chi_{A_1}\chi_{A_2}.$$

Now, write

$$A^k = \begin{bmatrix} A_1^k & B_k \\ 0 & A_2^k \end{bmatrix},$$

where $B_k \in \mathbb{F}^{n\times m}$ for all $k \in \mathbb{N}$. Then,

$$\chi_{A_1}(A) = \begin{bmatrix} 0 & \hat{B}_1 \\ 0 & \chi_{A_1}(A_2) \end{bmatrix}$$

and

$$\chi_{A_2}(A) = \begin{bmatrix} \chi_{A_2}(A_1) & \hat{B}_2 \\ 0 & 0 \end{bmatrix},$$

where $\hat{B}_1, \hat{B}_2 \in \mathbb{F}^{n\times m}$. Therefore,

$$\mathcal{R}[\chi_{A_2}(A)] \subseteq \mathcal{R}\left(\begin{bmatrix} I_n \\ 0 \end{bmatrix}\right) \subseteq \mathcal{N}[\chi_{A_1}(A)]$$

and

$$\chi_{A_2}(A_1)\hat{B}_1 + \hat{B}_2\chi_{A_1}(A_2) = 0.$$

Hence, $\chi_A(A) = \chi_{A_1}(A)\chi_{A_2}(A) = \chi_{A_2}(A)\chi_{A_1}(A) = 0$.

Fact 4.10.8. Let $A_1 \in \mathbb{F}^{n\times n}$, $A_{12} \in \mathbb{F}^{n\times m}$, and $A_2 \in \mathbb{F}^{m\times m}$, assume that $\mathrm{spec}(A_1) \cap \mathrm{spec}(A_2) = \varnothing$, and define $A \in \mathbb{F}^{(n+m)\times(n+m)}$ by

$$A \triangleq \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}.$$

Furthermore, let $\mu_1, \mu_2 \in \mathbb{F}[s]$ be such that

$$\mu_A = \mu_1\mu_2,$$
$$\mathrm{roots}(\mu_1) = \mathrm{spec}(A_1),$$
$$\mathrm{roots}(\mu_2) = \mathrm{spec}(A_2).$$

Now, write

$$A^k = \begin{bmatrix} A_1^k & B_k \\ 0 & A_2^k \end{bmatrix},$$

where $B_k \in \mathbb{F}^{n\times m}$ for all $k \in \mathbb{N}$. Then,

$$\mu_1(A) = \begin{bmatrix} 0 & \hat{B}_1 \\ 0 & \mu_1(A_2) \end{bmatrix}$$

and

$$\mu_2(A) = \begin{bmatrix} \mu_2(A_1) & \hat{B}_2 \\ 0 & 0 \end{bmatrix},$$

where $\hat{B}_1, \hat{B}_2 \in \mathbb{F}^{n\times m}$. Therefore,

$$\mathcal{R}[\mu_2(A)] \subseteq \mathcal{R}\left(\begin{bmatrix} I_n \\ 0 \end{bmatrix}\right) \subseteq \mathcal{N}[\mu_1(A)]$$

and

$$\mu_2(A_1)\hat{B}_1 + \hat{B}_2\mu_1(A_2) = 0.$$

Hence, $\mu_A(A) = \mu_1(A)\mu_2(A) = \mu_2(A)\mu_1(A) = 0.$

Fact 4.10.9. Let $A_1, A_2, A_3, A_4, B_1, B_2 \in \mathbb{F}^{n\times n}$, and define $A \in \mathbb{F}^{4n\times 4n}$ by

$$A \triangleq \begin{bmatrix} A_1 & B_1 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & A_3 & 0 \\ 0 & 0 & B_2 & A_4 \end{bmatrix}.$$

Then,

$$\operatorname{mspec}(A) = \bigcup_{i=1}^{4} \operatorname{mspec}(A_i).$$

Fact 4.10.10. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$, and assume that $m < n$. Then,

$$\operatorname{mspec}(I_n + AB) = \operatorname{mspec}(I_m + BA) \cup \{1, \ldots, 1\}_{\rm m}.$$

Fact 4.10.11. Let $a, b \in \mathbb{F}$, and define the Toeplitz matrix $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix}.$$

Then,

$$\operatorname{mspec}(A) = \{a + (n-1)b, a-b, \ldots, a-b\}_{\rm m}$$

and

$$A^2 + a_1A + a_0I = 0,$$

where $a_1 \triangleq -2a+(2-n)b$ and $a_0 \triangleq a^2+(n-2)ab+(1-n)b^2$. Furthermore, if A is nonsingular, then

$$A^{-1} = \frac{1}{a-b}I_n + \frac{b}{(b-a)[a+b(n-1)]}1_{n\times n}.$$

(Remark: See Fact 2.12.24.)

Fact 4.10.12. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{spec}(A) \subset \bigcup_{i=1}^{n} \left\{ \lambda \in \mathbb{C}\colon\ |\lambda - A_{(i,i)}| \le \sum_{j=1, j\ne i}^{n} |A_{(i,j)}| \right\}.$$

(Remark: This result is the *Gershgorin circle theorem.* See [115] for a proof and related results.)

Fact 4.10.13. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{spec}(A) \subset \bigcup_{\substack{i,j=1\\ i\ne j}}^{n} \left\{ \lambda \in \mathbb{C}\colon |\lambda - A_{(i,i)}||\lambda - A_{(j,j)}| \le \sum_{\substack{k=1\\ k\ne i}}^{n} |A_{(i,k)}| \sum_{\substack{k=1\\ k\ne j}}^{n} |A_{(j,k)}| \right\}.$$

(Remark: The inclusion region is the *ovals of Cassini.* The result is due to Brauer. See [287, p. 380].)

Fact 4.10.14. Let $A \in \mathbb{F}^{n\times n}$, and assume that, for all $i = 1, \ldots, n$,

$$\sum_{j=1, j\ne i}^{n} |A_{(i,j)}| < |A_{(i,i)}|.$$

Then, A is nonsingular. (Proof: Apply the Gershgorin circle theorem.) (Remark: This result is the *diagonal dominance theorem* and A is *diagonally dominant.* See [500] for a history of this result.) (Remark: For related results, see [189, 428, 470].) (Problem: Determine a lower bound for $|\det A|$ in terms of the difference between these quantities.)

Fact 4.10.15. Let $A \in \mathbb{F}^{n\times n}$, and, for $j = 1, \ldots, n$, define $b_j \triangleq \sum_{i=1}^{n} |A_{(i,j)}|$. Then,

$$\sum_{j=1}^{n} |A_{(j,j)}|/b_j \le \operatorname{rank} A.$$

(Proof: See [466, p. 67].) (Remark: See Fact 4.10.14.)

Fact 4.10.16. Let $A_1, \ldots, A_r \in \mathbb{F}^{n\times n}$ be normal and let $A \in \operatorname{co}\{A_1,$

$\ldots, A_r\}$. Then,

$$\operatorname{spec}(A) \subseteq \operatorname{co} \bigcup_{i=1,\ldots,r} \operatorname{spec}(A_i).$$

(Proof: See [584].)

Fact 4.10.17. Let $A \in \mathbb{F}^{n\times n}$, and define the *numerical range* of A by

$$\Theta(A) \triangleq \{x^*Ax\colon\ x \in \mathbb{C}^n \text{ and } x^*x = 1\}.$$

Then, $\Theta(A)$ is a closed, convex subset of $\mathbb{C}$. Furthermore,

$$\operatorname{co}\operatorname{spec}(A) \subseteq \Theta(A) \subseteq \operatorname{co}\{\nu_1 + \jmath\mu_1, \nu_1 + \jmath\mu_n, \nu_n + \jmath\mu_1, \nu_n + \jmath\mu_n\},$$

where

$$\nu_1 = \lambda_{\max}\big(\tfrac{1}{2}(A + A^*)\big), \qquad \nu_n = \lambda_{\min}\big(\tfrac{1}{2}(A + A^*)\big),$$
$$\mu_1 = \lambda_{\max}\Big(\tfrac{1}{2\jmath}(A - A^*)\Big), \qquad \mu_n = \lambda_{\min}\Big(\tfrac{1}{2\jmath}(A - A^*)\Big).$$

If, in addition, A is normal, then

$$\Theta(A) = \operatorname{co}\operatorname{spec}(A).$$

Conversely, if $n \le 4$ and $\Theta(A) = \operatorname{co}\operatorname{spec}(A)$, then A is normal. (Proof: See [252] or [289, pp. 11, 52].) (Remark: $\Theta(A)$ is called the *field of values* in [289, p. 5].)

Fact 4.10.18. Let $A, B \in \mathbb{R}^{n\times n}$. Then,

$$\operatorname{mspec}\left(\begin{bmatrix} A & B \\ -B & A \end{bmatrix}\right) = \operatorname{mspec}(A + \jmath B) \cup \operatorname{mspec}(A - \jmath B).$$

(Remark: See Fact 2.15.3.)

Fact 4.10.19. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and assume that $\operatorname{sprad}(I - A) < 1$. Then,

$$A^{-1} = \sum_{k=0}^{\infty} (I - A)^k.$$

4.11 Facts on Nonnegative Matrices

Fact 4.11.1. Let $A \in \mathbb{R}^{n\times n}$, where $n > 1$, and assume that A is nonnegative. Then, the following statements hold:

i) $\operatorname{sprad}(A)$ is an eigenvalue of A.

ii) There exists a nonnegative vector $x \in \mathbb{R}^n$ such that $Ax = \operatorname{sprad}(A)x$.

Furthermore, the following statements are equivalent:

iii) $(I + A)^{n-1}$ is positive.

iv) There do not exist $k>0$ and a permutation matrix $S\in\mathbb{R}^{n\times n}$ such that

$$SAS^{\mathrm{T}}=\begin{bmatrix} B & C\\ 0_{k\times(n-k)} & D\end{bmatrix}.$$

v) No eigenvector of A has a zero component.

vi) A has exactly one nonnegative eigenvector whose components sum to 1, and this eigenvector is positive.

A is *irreducible* if *iii)*-*vi)* are satisfied. If A is irreducible, then the following statements hold:

vii) $\operatorname{sprad}(A)>0$.

viii) $\operatorname{sprad}(A)$ is a simple eigenvalue of A.

ix) There exists a positive vector $x\in\mathbb{R}^n$ such that $Ax=\operatorname{sprad}(A)x$.

x) A has exactly one positive eigenvector whose components sum to 1.

xi) Assume that $\{\lambda_1,\ldots,\lambda_k\}_{\mathrm{m}}=\{\lambda\in\operatorname{mspec}(A)\colon\ |\lambda|=\operatorname{sprad}(A)\}_{\mathrm{m}}$. Then, $\lambda_1,\ldots,\lambda_k$ are distinct, and

$$\{\lambda_1,\ldots,\lambda_k\}=\{e^{2\pi\jmath i/k}\operatorname{sprad}(A)\colon i=1,\ldots,k\}.$$

Furthermore,

$$\operatorname{mspec}(A)=e^{2\pi\jmath/k}\operatorname{mspec}(A).$$

xii) If at least one diagonal entry of A is positive, then $\operatorname{sprad}(A)$ is the only eigenvalue of A whose absolute value is $\operatorname{sprad}(A)$.

In addition, the following statements are equivalent:

xiii) There exists $k>0$ such that A^k is positive.

xiv) A is irreducible and $|\lambda|<\operatorname{sprad}(A)$ for all $\lambda\in\operatorname{spec}(A)\backslash\{\operatorname{sprad}(A)\}$.

xv) A^{n^2-2n+2} is positive.

A is *primitive* if *xiii)*-*xiv)* are satisfied. (Example: $\left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right]$ is irreducible but not primitive.) Finally, assume that A is irreducible and let $x\in\mathbb{R}^n$ be positive and satisfy $Ax=\operatorname{sprad}(A)x$. Then, for all positive $x_0\in\mathbb{R}^n$, there exists a positive real number γ such that

$$\lim_{k\to\infty}\Big(A^kx_0-\gamma[\operatorname{sprad}(A)]^kx\Big)=0.$$

(Remark: For an arbitrary positive initial condition, the state of the difference equation $x_{k+1}=Ax_k$ approaches a distribution that is identical to the distribution of the eigenvector associated with the positive eigenvalue of maximum absolute value. In demography, this eigenvector is interpreted as the *stable age distribution*. See [329, pp. 47, 63].) (Proof: See [7, pp. 45–49], [81, pp. 26–28, 32, 55], [287, pp. 507–511], and [202].) (Remark:

This result is the *Perron-Frobenius theorem*.) (Remark: See Fact 11.14.18.) (Remark: Statement *xv*) is due to Wielandt. See [466, p. 157].)

Fact 4.11.2. Let $A \triangleq \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Then, $\chi_A(s) = s^2 - s - 1$ and $\operatorname{spec}(A) = \{\alpha, \beta\}$, where $\alpha \triangleq \frac{1}{2}(1+\sqrt{5})$ and $\beta \triangleq \frac{1}{2}(1-\sqrt{5})$ satisfy

$$\alpha - 1 = 1/\alpha, \qquad \beta - 1 = 1/\beta.$$

Furthermore, $\left[\begin{smallmatrix} \alpha \\ 1 \end{smallmatrix}\right]$ is an eigenvector of A associated with α. Now, for $k \ge 0$, consider the difference equation

$$x_{k+1} = Ax_k.$$

Then, for all $k \ge 0$,

$$x_k = A^k x_0$$

and

$$x_{k+2(1)} = x_{k+1(1)} + x_{k(1)}.$$

Furthermore, if x_0 is positive, then

$$\lim_{k\to\infty} \frac{x_{k(1)}}{x_{k(2)}} = \alpha.$$

In particular, if $x_0 \triangleq \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, then, for all $k \ge 0$,

$$x_k = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix},$$

where $F_1 \triangleq F_2 \triangleq 1$ and, for all $k \ge 1$, F_k satisfies

$$F_{k+2} = F_{k+1} + F_k.$$

Furthermore,

$$A^k = \begin{bmatrix} F_{k+1} & F_k \\ F_k & F_{k-1} \end{bmatrix}.$$

On the other hand, if $x_0 \triangleq \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$, then, for all $k \ge 0$,

$$x_k = \begin{bmatrix} L_{k+2} \\ L_{k+1} \end{bmatrix},$$

where $L_1 \triangleq 1$, $L_2 \triangleq 3$, and, for all $k \ge 1$, L_k satisfies

$$L_{k+2} = L_{k+1} + L_k.$$

Furthermore,

$$\lim_{k\to\infty} \frac{F_{k+1}}{F_k} = \frac{L_{k+1}}{L_k} = \alpha.$$

(Proof: Use the last statement of Fact 4.11.1.) (Remark: F_k is the kth *Fibonacci number*, L_k is the kth *Lucas number*, and α is the *golden mean*. See [339, pp. 6–8, 239–241, 362, 363].)

Fact 4.11.3. Consider the nonnegative companion matrix $A \in \mathbb{R}^{n\times n}$ defined by

$$A \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1/n & 1/n & 1/n & \cdots & 1/n & 1/n \end{bmatrix}.$$

Then, A is irreducible, 1 is a simple eigenvalue of A with associated eigenvector $1_{n\times 1}$, and $|\lambda| < 1$ for all $\lambda \in \operatorname{spec}(A)\backslash\{1\}$. Furthermore, if $x \in \mathbb{R}^n$, then

$$\lim_{k\to\infty} A^k x = \left[\frac{2}{n(n+1)}\sum_{i=1}^{n} i x_{(i-1)}\right] 1_{n\times 1}.$$

(Proof: See [261, pp. 82, 83, 263–266].) (Remark: The result also follows from Fact 4.11.1.)

Fact 4.11.4. Let $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^m$. Then, the following statements are equivalent:

i) If $x \in \mathbb{R}^m$ and $Ax \geq\geq 0$, then $b^{\mathrm{T}}x \geq 0$.

ii) There exists $y \in \mathbb{R}^n$ such that $y \geq\geq 0$ and $A^{\mathrm{T}}y = b$.

Equivalently, exactly one of the following two statements is satisfied:

i) There exists $x \in \mathbb{R}^m$ such that $Ax \geq\geq 0$ and $b^{\mathrm{T}}x < 0$.

ii) There exists $y \in \mathbb{R}^n$ such that $y \geq\geq 0$ and $A^{\mathrm{T}}y = b$.

(Proof: See [68, p. 47].) (Remark: This result is *Farkas' theorem.*)

Fact 4.11.5. Let $A \in \mathbb{R}^{n\times m}$. Then, the following statements are equivalent:

i) There exists $x \in \mathbb{R}^m$ such that $Ax >> 0$.

ii) If $y \in \mathbb{R}^n$ is nonzero and $y \geq\geq 0$, then $A^{\mathrm{T}}y \neq 0$.

Equivalently, exactly one of the following two statements is satisfied:

i) There exists $x \in \mathbb{R}^m$ such that $Ax >> 0$.

ii) There exists nonzero $y \in \mathbb{R}^n$ such that $y \geq\geq 0$ and $A^{\mathrm{T}}y = 0$.

(Proof: See [68, p. 47].) (Remark: This result is *Gordan's theorem.*)

Fact 4.11.6. Let $A \in \mathbb{C}^{n\times n}$, and define $|A| \in \mathbb{R}^{n\times n}$ by $|A|_{(i,j)} \triangleq |A_{(i,j)}|$

for all $i,j = 1,\ldots,n$. Then,

$$\operatorname{sprad}(A) \le \operatorname{sprad}(|A|).$$

(Proof: See [416, p. 619].)

Fact 4.11.7. Let $A, B \in \mathbb{R}^{n\times n}$, where $0 \le\le A \le\le B$. Then,

$$\operatorname{sprad}(A) \le \operatorname{sprad}(B).$$

If, in addition, $B \ne A$ and $A + B$ is irreducible, then

$$\operatorname{sprad}(A) < \operatorname{sprad}(B).$$

(Proof: See [74, p. 27].)

Fact 4.11.8. Let $A \in \mathbb{R}^{n\times n}$, assume that $A >> 0$, and let $\lambda \in \operatorname{spec}(A)\backslash\{\operatorname{sprad}(A)\}$. Then,

$$|\lambda| \le \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} \operatorname{sprad}(A),$$

where

$$A_{\max} \triangleq \max\left\{A_{(i,j)}\colon\ i,j = 1,\ldots,n\right\}$$

and

$$A_{\min} \triangleq \min\left\{A_{(i,j)}\colon\ i,j = 1,\ldots,n\right\}.$$

(Remark: This result is *Hopf's theorem.*)

Fact 4.11.9. Let $A \in \mathbb{R}^{n\times n}$, assume that A is nonnegative and primitive, and let $x, y \in \mathbb{R}^n$, where $x > 0$ and $y > 0$ satisfy $Ax = \operatorname{sprad}(A)x$ and $A^{\mathrm{T}}y = \operatorname{sprad}(A)y$. Then,

$$\lim_{i\to\infty} \left[\frac{1}{\operatorname{sprad}(A)} A\right]^i = xy^{\mathrm{T}}.$$

(Proof: See [287, p. 516].)

4.12 Notes

Much of the development in this chapter is based upon [456]. Additional discussions of the Smith and Smith-McMillan forms are given in [321] and [632]. The proofs of Lemma 4.4.7 and Leverrier's algorithm Proposition 4.4.8 are based on [484, p. 432, 433], where it is called the *Souriau-Frame algorithm.* Alternative proofs of Leverrier's algorithm are given in [63, 296]. The proof of Theorem 4.6.1 is based on [287]. Polynomial-based approaches to linear algebra are given in [120, 216], while polynomial matrices and rational transfer functions are studied in [230, 572].

Chapter Five

Matrix Decompositions

In this chapter we present several matrix decompositions, namely, the Smith, multi-companion, hypercompanion, Jordan, Schur, and singular value decompositions.

5.1 Smith Form

Our first decomposition involves rectangular matrices subject to a biequivalence transformation. This result is the specialization of the Smith decomposition given by Theorem 4.3.2 to constant matrices.

Theorem 5.1.1. Let $A \in \mathbb{F}^{n\times m}$ and $r \triangleq \operatorname{rank} A$. Then, there exist nonsingular matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ such that

$$A = S_1 \begin{bmatrix} I_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} S_2. \tag{5.1.1}$$

Corollary 5.1.2. Let $A, B \in \mathbb{F}^{n\times m}$. Then, A and B are biequivalent if and only if A and B have the same Smith form.

Proposition 5.1.3. Let $A, B \in \mathbb{F}^{n\times m}$. Then, the following statements hold:

i) A and B are left equivalent if and only if $\mathcal{N}(A) = \mathcal{N}(B)$.

ii) A and B are right equivalent if and only $\mathcal{R}(A) = \mathcal{R}(B)$.

iii) A and B are biequivalent if and only if $\operatorname{rank} A = \operatorname{rank} B$.

Proof. The proof of necessity is immediate in *i)*-*iii)*. Sufficiency in *iii)* follows from Corollary 5.1.2. For sufficiency in *i)* and *ii)*, see [484, pp. 179–181]. □

5.2 Multi-Companion Form

For the monic polynomial $p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0 \in \mathbb{F}[s]$ of degree $n \geq 1$, the *companion matrix* $C(p) \in \mathbb{F}^{n\times n}$ associated with p is defined to be

$$C(p) \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 & \cdots & -\beta_{n-2} & -\beta_{n-1} \end{bmatrix}. \tag{5.2.1}$$

If $n = 1$, then $p(s) = s + \beta_0$ and $C(p) = -\beta_0$. Furthermore, if $n = 0$ and $p = 1$, then we define $C(p) \triangleq 0_{0\times 0}$. Note that if $n \geq 1$, then $\operatorname{tr} C(p) = -\beta_{n-1}$ and $\det C(p) = (-1)^n\beta_0 = (-1)^n p(0)$.

It is easy to see that the characteristic polynomial of the companion matrix $C(p)$ is p. For example, let $n = 3$ so that

$$C(p) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 \end{bmatrix} \tag{5.2.2}$$

and thus

$$sI - C(p) = \begin{bmatrix} s & -1 & 0 \\ 0 & s & -1 \\ \beta_0 & \beta_1 & s+\beta_2 \end{bmatrix}. \tag{5.2.3}$$

Adding s times the second column and s^2 times the third column to the first column leaves the determinant of $sI - C(p)$ unchanged and yields

$$\begin{bmatrix} 0 & -1 & 0 \\ 0 & s & -1 \\ p(s) & \beta_1 & s+\beta_2 \end{bmatrix}, \tag{5.2.4}$$

Hence, $\chi_{C(p)} = p$. If $n = 0$ and $p = 1$, then we define $\chi_{C(p)} \triangleq \chi_{0_{0\times 0}} = 1$. The following result shows that companion matrices have the same characteristic and minimal polynomials.

Proposition 5.2.1. Let $p \in \mathbb{F}[s]$ be a monic polynomial having degree

n. Then, there exist unimodular matrices $S_1, S_2 \in \mathbb{F}^{n\times n}[s]$ such that

$$sI - C(p) = S_1(s)\begin{bmatrix} I_{n-1} & 0_{(n-1)\times 1} \\ 0_{1\times(n-1)} & p(s)\end{bmatrix} S_2(s). \tag{5.2.5}$$

Furthermore,

$$\chi_{C(p)}(s) = \mu_{C(p)}(s) = p(s). \tag{5.2.6}$$

Proof. Since $\chi_{C(p)} = p$, it follows that $\operatorname{rank}[sI - C(p)] = n$. Next, since $\det\big([sI - C(p)]_{[n,1]}\big) = (-1)^{n-1}$, it follows that $\Delta_{n-1} = 1$, where Δ_{n-1} is the greatest common divisor (which is monic by definition) of all $(n-1)\times(n-1)$ subdeterminants of $sI - C(p)$. Furthermore, since Δ_{i-1} divides Δ_i for all $i = 2, \ldots, n-1$, it follows that $\Delta_1 = \cdots = \Delta_{n-2} = 1$. Consequently, $p_1 = \cdots = p_{n-1} = 1$. Since, by Proposition 4.6.2, $\chi_{C(p)} = \prod_{i=1}^n p_i = p_n$ and $\mu_{C(p)} = p_n$, it follows that $\chi_{C(p)} = \mu_{C(p)} = p$. $\square$

Next, we consider block-diagonal matrices all of whose diagonally located blocks are companion matrices.

Lemma 5.2.2. Let $p_1, \ldots, p_n \in \mathbb{F}[s]$ be monic polynomials such that p_i divides p_{i+1} for all $i = 1, \ldots, n-1$ and $n = \sum_{i=1}^n \deg p_i$. Furthermore, define $C \triangleq \operatorname{diag}[C(p_1), \ldots, C(p_n)] \in \mathbb{F}^{n\times n}$. Then, there exist unimodular matrices $S_1, S_2 \in \mathbb{F}^{n\times n}[s]$ such that

$$sI - C = S_1(s)\begin{bmatrix} p_1(s) & & 0 \\ & \ddots & \\ 0 & & p_n(s)\end{bmatrix} S_2(s). \tag{5.2.7}$$

Proof. Letting $k_i = \deg p_i$, Proposition 5.2.1 implies that the Smith form of $sI_{k_i} - C(p_i)$ is $0_{0\times 0}$ if $k_i = 0$ and $\operatorname{diag}(I_{k_i - 1}, p_i)$ if $k_i \ge 1$. By combining these Smith forms it follows that there exist unimodular matrices $S_1, S_2 \in \mathbb{F}^{n\times n}[s]$ such that

$$\begin{aligned} sI - C &= \begin{bmatrix} sI_{k_1} - C(p_1) & & \\ & \ddots & \\ & & sI_{k_n} - C(p_n)\end{bmatrix} \\ &= S_1(s)\begin{bmatrix} p_1(s) & & 0 \\ & \ddots & \\ 0 & & p_n(s)\end{bmatrix} S_2(s). \end{aligned}$$

Since p_i divides p_{i+1} for all $i = 1, \ldots, n-1$, it follows that this diagonal matrix is the Smith form of $sI - C$. $\square$

The following result uses Lemma 5.2.2 to construct a canonical form, known as the *multi-companion form*, for square matrices under a similarity transformation.

Theorem 5.2.3. Let $A \in \mathbb{F}^{n\times n}$, and let $p_1, \ldots, p_n \in \mathbb{F}[s]$ denote the similarity invariants of A, where p_i divides p_{i+1} for all $i = 1, \ldots, n-1$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = S\begin{bmatrix} C(p_1) & & \\ & \ddots & \\ & & C(p_n) \end{bmatrix} S^{-1}. \tag{5.2.8}$$

Proof. Lemma 5.2.2 implies that the $n \times n$ matrix $sI - C$, where $C \triangleq \operatorname{diag}[C(p_1), \ldots, C(p_n)]$, has the Smith form $\operatorname{diag}(p_1, \ldots, p_n)$. Now, since $sI - A$ has the same similarity invariants as C, it follows from Theorem 4.3.9 that A and C are similar. □

Corollary 5.2.4. Let $A \in \mathbb{F}^{n\times n}$. Then, $\mu_A = \chi_A$ if and only if A is similar to $C(\chi_A)$.

Proof. Suppose that $\mu_A = \chi_A$. Then, it follows from Proposition 4.6.2 that $p_i = 1$ for all $i = 1, \ldots, n-1$ and $p_n = \chi_A$ is the only nonconstant similarity invariant of A. Thus, $C(p_i) = 0_{0\times 0}$ for all $i = 1, \ldots, n-1$, and it follows from Theorem 5.2.3 that A is similar to $C(\chi_A)$. The converse can be verified directly. □

Corollary 5.2.5. Let $A \in \mathbb{F}^{n\times n}$ be a companion matrix. Then, $\mu_A = \chi_A$.

Proof. The result is an immediate consequence of Corollary 5.2.5. Alternatively, if p is monic with degree $n-1$, then $[p(A)]_{(1,n)} = 1$. □

Note that if $A = I_n$, then the similarity invariants of A are $p_i(s) = s-1$ for all $i = 1, \ldots, n$. Thus, $C(p_i) = 1$ for all $i = 1, \ldots, n$, as expected.

Corollary 5.2.6. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A and B are similar.

ii) A and B have the same similarity invariants.

iii) A and B have the same multi-companion form.

The multi-companion form given by Theorem 5.2.3 provides a canonical form for A in terms of a block-diagonal matrix of companion matrices. As will be seen, however, the multi-companion form is only one such decomposition. The goal of the remainder of this section is to obtain an additional canonical form by applying a similarity transformation to the multi-companion form.

To begin, note that if A_i is similar to B_i for all $i = 1, \ldots, r$, then $\operatorname{diag}(A_1, \ldots, A_r)$ is similar to $\operatorname{diag}(B_1, \ldots, B_r)$. Therefore, it follows from Corollary 5.2.6 that, if $sI - A_i$ and $sI - B_i$ have the same Smith form for all $i = 1, \ldots, r$, then $sI - \operatorname{diag}(A_1, \ldots, A_r)$ and $sI - \operatorname{diag}(B_1, \ldots, B_r)$ have the same Smith form. The following lemma is needed.

Lemma 5.2.7. Let $A = \operatorname{diag}(A_1, A_2)$, where $A_i \in \mathbb{F}^{n_i \times n_i}$ for $i = 1, 2$. Then, μ_A is the least common multiple of μ_{A_1} and μ_{A_2}. In particular, if μ_{A_1} and μ_{A_2} are coprime, then $\mu_A = \mu_{A_1}\mu_{A_2}$.

Proof. Since $\mu_A(A) = 0$, it follows that $\mu_A(A_1) = 0$ and $\mu_A(A_2) = 0$. Therefore, Theorem 4.1.5 implies that μ_{A_1} and μ_{A_2} both divide μ_A. Consequently, the least common multiple q of μ_{A_1} and μ_{A_2} also divides μ_A. Since $q(A_1) = 0$ and $q(A_2) = 0$, it follows that $q(A) = 0$. Therefore, μ_A divides q. Hence, $q = \mu_A$. If, in addition, μ_{A_1} and μ_{A_2} are coprime, then $\mu_A = \mu_{A_1}\mu_{A_2}$. □

Proposition 5.2.8. Let $p \in \mathbb{F}[s]$ be a monic polynomial of positive degree n, and let $p = p_1 \cdots p_r$, where $p_1, \ldots, p_r \in \mathbb{F}[s]$ are monic and pairwise coprime polynomials. Then, the matrices $C(p)$ and $\operatorname{diag}[C(p_1), \ldots, C(p_r)]$ are similar.

Proof. Let $\hat{p}_2 = p_2 \cdots p_r$ and $\hat{C} \triangleq \operatorname{diag}[C(p_1), C(\hat{p}_2)]$. Since p_1 and $\hat{q}_2$ are coprime, it follows from Lemma 5.2.7 that $\mu_{\hat{C}} = \mu_{C(p_1)}\mu_{C(\hat{p}_2)}$. Furthermore, $\chi_{\hat{C}} = \chi_{C(p_1)}\chi_{C(\hat{p}_2)} = \mu_{\hat{C}}$. Hence, Corollary 5.2.4 implies that $\hat{C}$ is similar to $C(\chi_{\hat{C}})$. However, $\chi_{\hat{C}} = p_1 \cdots p_r = p$, so that $\hat{C}$ is similar to $C(p)$. If $r > 2$, then the same argument can be used to decompose $C(\hat{p}_2)$ to show that $C(p)$ is similar to $\operatorname{diag}[C(p_1), \ldots, C(p_r)]$. □

Proposition 5.2.8 can be used to decompose every companion block of a multi-companion form into smaller companion matrices. This procedure can be carried out for every companion block whose characteristic polynomial has coprime factors. For example, suppose that $A \in \mathbb{R}^{10 \times 10}$ has the similarity invariants $p_i(s) = 1$ for all $i = 1, \ldots, 7$, $p_8(s) = (s+1)^2$, $p_9(s) = (s+1)^2(s+2)$, and $p_{10}(s) = (s+1)^2(s+2)(s^2+3)$, so that, by Theorem 5.2.3 the multi-companion form of A is $\operatorname{diag}[C(p_8), C(p_9), C(p_{10})]$, where $C(p_8) \in \mathbb{R}^{2 \times 2}$, $C(p_9) \in \mathbb{R}^{3 \times 3}$, and $C(p_{10}) \in \mathbb{R}^{5 \times 5}$. According to Proposition 5.2.8, the companion matrices $C(p_9)$ and $C(p_{10})$ can be further decomposed. For example, $C(p_9)$ is similar to $\operatorname{diag}[C(p_{9,1}), C(p_{9,2})]$, where $p_{9,1}(s) = (s+1)^2$ and $p_{9,2}(s) = s+2$ are coprime. Furthermore,

$C(p_{10})$ is similar to four different diagonal matrices, three of which have two companion blocks while the fourth has three companion blocks. Since $p_8(s) = (s+1)^2$ does not have nonconstant coprime factors, however, it follows that the companion matrix $C(p_8)$ cannot be decomposed into smaller companion matrices.

The largest number of companion blocks achievable by similarity transformation is obtained by factoring every similarity invariant into *elementary divisors*, which are powers of irreducible polynomials that are nonconstant, monic, and pairwise coprime. In the above example, this factorization is given by $p_9(s) = p_{9,1}(s)p_{9,2}(s)$, where $p_{9,1}(s) = (s+1)^2$ and $p_{9,2}(s) = s+2$, and by $p_{10} = p_{10,1}p_{10,2}p_{10,3}$, where $p_{10,1}(s) = (s+1)^2$, $p_{10,2}(s) = s+2$, and $p_{10,3}(s) = s^2+3$. The elementary divisors of A are thus $(s+1)^2$, $(s+1)^2$, $s+2$, $(s+1)^2$, $s+2$, and s^2+3, which yields six companion blocks. Viewing $A \in \mathbb{C}^{n\times n}$ we can further factor $p_{10,3}(s) = (s+\jmath\sqrt{3})(s-\jmath\sqrt{3})$, which yields a total of seven companion blocks. From Proposition 5.2.8 and Theorem 5.2.3 we obtain the *elementary multi-companion form*, which provides another canonical form for A.

Theorem 5.2.9. Let $A \in \mathbb{F}^{n\times n}$, and let $q_1^{l_1}, \ldots, q_h^{l_h} \in \mathbb{F}[s]$ be the elementary divisors of A, where $l_1, \ldots, l_h \in \mathbb{P}$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = S\begin{bmatrix} C\Big(q_1^{l_1}\Big) & & \\ & \ddots & \\ & & C\Big(q_h^{l_h}\Big) \end{bmatrix}S^{-1}. \tag{5.2.9}$$

5.3 Hypercompanion Form and Jordan Form

In this section we present an alternative form of the companion blocks of the elementary multi-companion form (5.2.9). To do this we define the *hypercompanion matrix* $\mathcal{H}_l(q)$ associated with the elementary divisor $q^l \in \mathbb{F}[s]$, where $l \in \mathbb{P}$, as follows. For $q(s) = s - \lambda \in \mathbb{C}[s]$, define the $l \times l$ Toeplitz hypercompanion matrix

$$\mathcal{H}_l(q) \triangleq \lambda I_l + N_l = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}, \tag{5.3.1}$$

while, for $q(s) = s^2 - \beta_1 s - \beta_0 \in \mathbb{R}[s]$, define the $2l \times 2l$ real, tridiagonal hypercompanion matrix

$$\mathcal{H}_l(q) \triangleq \begin{bmatrix} 0 & 1 & & & & & & \\ \beta_0 & \beta_1 & 1 & & & 0 & & \\ & 0 & 0 & 1 & & & & \\ & & \beta_0 & \beta_1 & 1 & & & \\ & & & \ddots & \ddots & \ddots & & \\ & 0 & & & \ddots & 0 & 1 \\ & & & & & \beta_0 & \beta_1 \end{bmatrix}. \tag{5.3.2}$$

The following result shows that the hypercompanion matrix $\mathcal{H}_l(q)$ is similar to the companion matrix $C(q^l)$ associated with the elementary divisor q^l of $\mathcal{H}_l(q)$.

Lemma 5.3.1. Let $l \in \mathbb{P}$, and let $q(s) = s - \lambda \in \mathbb{C}[s]$ or $q(s) = s^2 - \beta_1 s - \beta_0 \in \mathbb{R}[s]$. Then, q^l is the only elementary divisor of $\mathcal{H}_l(q)$, and $\mathcal{H}_l(q)$ is similar to $C\big(q^l\big)$.

Proof. Let k denote the order of $\mathcal{H}_l(q)$. Then, $\chi_{\mathcal{H}_l(q)} = q^l$ and $\det\big([sI - C_l(q)]_{[k,1]}\big) = (-1)^{k-1}$. Hence, as in the proof of Proposition 5.2.1, it follows that $\chi_{\mathcal{H}_l(q)} = \mu_{\mathcal{H}_l(q)}$. Corollary 5.2.4 now implies that $\mathcal{H}_l(q)$ is similar to $C\big(q^l\big)$. □

Proposition 5.2.8 and Lemma 5.3.1 yield the following canonical form, which is known as the *hypercompanion form.*

Theorem 5.3.2. Let $A \in \mathbb{F}^{n\times n}$, and let $q_1^{l_1}, \ldots, q_h^{l_h} \in \mathbb{F}[s]$ be the elementary divisors of A, where $l_1, \ldots, l_h \in \mathbb{P}$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = S \begin{bmatrix} \mathcal{H}_{l_1}(q_1) & & \\ & \ddots & \\ & & \mathcal{H}_{l_h}(q_h) \end{bmatrix} S^{-1}. \tag{5.3.3}$$

Next, consider Theorem 5.3.3 with $\mathbb{F} = \mathbb{C}$. In this case, every elementary divisor $q_i^{l_i}$ is of the form $(s - \lambda_i)^{l_i}$, where $\lambda_i \in \mathbb{C}$. Furthermore, $S \in \mathbb{C}^{n\times n}$, and the hypercompanion form (5.3.4) is a block-diagonal matrix all of whose diagonally located blocks are of the form (5.3.1). The hypercompanion form (5.3.4) with every diagonally located block of the form (5.3.1) is the *Jordan form* given by the following result.

Theorem 5.3.3. Let $A \in \mathbb{F}^{n\times n}$, and let $q_1^{l_1}, \ldots, q_h^{l_h} \in \mathbb{C}[s]$ be the elementary divisors of A, where $l_1, \ldots, l_h \in \mathbb{P}$ and $q_1, \ldots, q_h \in \mathbb{C}[s]$ are linear. Then, there exists a nonsingular matrix $S \in \mathbb{C}^{n\times n}$ such that

$$A = S \begin{bmatrix} \mathcal{H}_{l_1}(q_1) & & \\ & \ddots & \\ & & \mathcal{H}_{l_h}(q_h) \end{bmatrix} S^{-1}. \tag{5.3.4}$$

Corollary 5.3.4. Let $p \in \mathbb{F}[s]$, let $\lambda_1, \ldots, \lambda_r$ denote the distinct roots of p, and, for $i = 1, \ldots, r$, let $l_i \triangleq \mathrm{m}_p(\lambda_i)$ and $p_i(s) \triangleq s - \lambda_i$. Then, $C(p)$ is similar to $\mathrm{diag}[\mathcal{H}_{l_1}(p_1), \ldots, \mathcal{H}_{l_r}(p_r)]$.

To illustrate the structure of the Jordan form, let $l_i = 3$ and $q_i(s) = s - \lambda_i$, where $\lambda_i \in \mathbb{C}$. Then, $\mathcal{H}_{l_i}(q_i)$ is the 3×3 matrix

$$\mathcal{H}_{l_i}(q_i) = \lambda_i I_3 + N_3 = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix} \tag{5.3.5}$$

so that $\mathrm{mspec}[\mathcal{H}_{l_i}(q_i)] = \{\lambda_i, \lambda_i, \lambda_i\}_{\mathrm{m}}$. If $\mathcal{H}_{l_i}(q_i)$ is the only diagonally located block of the Jordan form associated with the eigenvalue λ_i, then the algebraic multiplicity of λ_i is equal to 3 while its geometric multiplicity is equal to 1.

Now, consider Theorem 5.3.3 with $\mathbb{F} = \mathbb{R}$. In this case, every elementary divisor $q_i^{l_i}$ is either of the form $(s-\lambda_i)^{l_i}$ or of the form $(s^2 - \beta_{1i}s - \beta_{0i})^{l_i}$, where $\beta_{0i}, \beta_{1i} \in \mathbb{R}$. Furthermore, $S \in \mathbb{R}^{n\times n}$ and the hypercompanion form (5.3.4) is a block-diagonal matrix whose diagonally located blocks are real matrices of the form (5.3.1) or (5.3.2). In this case, (5.3.4) is the *real hypercompanion form.*

Applying an additional real similarity transformation to each diagonally located block of the real hypercompanion form yields the *real Jordan form.* To do this, define the *real Jordan matrix* $\mathcal{J}_l(q)$ for $l \in \mathbb{P}$ as follows. For $q(s) = s - \lambda \in \mathbb{F}[s]$ define $\mathcal{J}_l(q) \triangleq \mathcal{H}_l(q)$, while if $q(s) = s^2 - \beta_1 s - \beta_0 \in \mathbb{F}[s]$ is irreducible with a nonreal root $\lambda = \nu + \jmath\omega$, then define the $2l \times 2l$ upper-

Hessenberg matrix

$$\mathcal{J}_l(q) \triangleq \begin{bmatrix} \nu & \omega & 1 & 0 & & & & \\ -\omega & \nu & 0 & 1 & \ddots & & 0 & \\ & & \nu & \omega & 1 & \ddots & & \\ & & -\omega & \nu & 0 & \ddots & \ddots & \\ & & & & \ddots & \ddots & 1 & 0 \\ & & & & & \ddots & 0 & 1 \\ & 0 & & & & & \nu & \omega \\ & & & & & & -\omega & \nu \end{bmatrix}. \tag{5.3.6}$$

Theorem 5.3.5. Let $A \in \mathbb{R}^{n\times n}$, and let $q_1^{l_1}, \ldots, q_h^{l_h} \in \mathbb{R}[s]$, where $l_1, \ldots, l_h \in \mathbb{P}$ are the elementary divisors of A. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S \begin{bmatrix} \mathcal{J}_{l_1}(q_1) & & 0 \\ & \ddots & \\ 0 & & \mathcal{J}_{l_h}(q_h) \end{bmatrix} S^{-1}. \tag{5.3.7}$$

Proof. It need only be shown that $\mathcal{J}_l(q)$ and $\mathcal{H}_l(q)$ are similar in the case that $q(s) = s^2 - \beta_1 s - \beta_0$ is an irreducible quadratic. Let $\lambda = \nu + \jmath\omega$ denote a root of q so that $\beta_1 = 2\nu$ and $\beta_0 = -(\nu^2 + \omega^2)$. Then,

$$\mathcal{H}_1(q) = \begin{bmatrix} 0 & 1 \\ \beta_0 & \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \nu & \omega \end{bmatrix} \begin{bmatrix} \nu & \omega \\ -\omega & \nu \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\nu/\omega & 1/\omega \end{bmatrix} = S\mathcal{J}_1(q)S^{-1}.$$

The transformation matrix $S = \left[\begin{smallmatrix} 1 & 0 \\ \nu & \omega \end{smallmatrix}\right]$ is not unique; an alternative choice is $S = \left[\begin{smallmatrix} \omega & \nu \\ 0 & \nu^2+\omega^2 \end{smallmatrix}\right]$. Similarly,

$$\mathcal{H}_2(q) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \beta_0 & \beta_1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \beta_0 & \beta_1 \end{bmatrix} = S \begin{bmatrix} \nu & \omega & 1 & 0 \\ -\omega & \nu & 0 & 1 \\ 0 & 0 & \nu & \omega \\ 0 & 0 & -\omega & \nu \end{bmatrix} S^{-1} = S\mathcal{J}_2(q)S^{-1},$$

where

$$S \triangleq \begin{bmatrix} \omega & \nu & \omega & \nu \\ 0 & \nu^2+\omega^2 & \omega & \nu^2+\omega^2+\nu \\ 0 & 0 & -2\omega\nu & 2\omega^2 \\ 0 & 0 & -2\omega(\nu^2+\omega^2) & 0 \end{bmatrix}. \qquad \square$$

Finally, we relate the real Jordan form (5.3.7) to the Jordan form (5.3.4) by showing that every diagonally located block of the form (5.3.6) is similar to a pair of Jordan blocks of the form (5.3.1). For example, if

$q(s) = s^2 - 2\nu s + \nu^2 + \omega^2$ with roots $\lambda = \nu + \jmath\omega$ and $\overline{\lambda} = \nu - \jmath\omega$, then

$$\mathcal{H}_1(q) = \begin{bmatrix} \nu & \omega \\ -\omega & \nu \end{bmatrix} = \tfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ \jmath & -\jmath \end{bmatrix}\begin{bmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{bmatrix}\tfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\jmath \\ 1 & \jmath \end{bmatrix}, \tag{5.3.8}$$

while

$$\mathcal{H}_2(q) = \begin{bmatrix} \nu & \omega & 1 & 0 \\ -\omega & \nu & 0 & 1 \\ 0 & 0 & \nu & \omega \\ 0 & 0 & -\omega & \nu \end{bmatrix} = S\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \overline{\lambda} & 1 \\ 0 & 0 & 0 & \overline{\lambda} \end{bmatrix}S^{-1}, \tag{5.3.9}$$

where

$$S = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 & 0 \\ \jmath & 0 & -\jmath & 0 \\ 0 & 1 & 0 & 1 \\ 0 & \jmath & 0 & -\jmath \end{bmatrix} \tag{5.3.10}$$

and

$$S^{-1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\jmath & 0 & 0 \\ 0 & 0 & 1 & -\jmath \\ 1 & \jmath & 0 & 0 \\ 0 & 0 & 1 & \jmath \end{bmatrix}. \tag{5.3.11}$$

Example 5.3.6. Let $A, B \in \mathbb{R}^{4\times4}$ and $C \in \mathbb{C}^{4\times4}$ be given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -16 & 0 & -8 & 0 \end{bmatrix}, \tag{5.3.12}$$

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -4 & 0 \end{bmatrix}, \tag{5.3.13}$$

and

$$C = \begin{bmatrix} 2\jmath & 1 & 0 & 0 \\ 0 & 2\jmath & 0 & 0 \\ 0 & 0 & -2\jmath & 1 \\ 0 & 0 & 0 & -2\jmath \end{bmatrix}. \tag{5.3.14}$$

Then, A is in companion form, B is in real hypercompanion form, and C is in Jordan form. Furthermore, A, B, and C are similar.

Example 5.3.7. Let $A, B \in \mathbb{R}^{6\times 6}$ and $C \in \mathbb{C}^{6\times 6}$ be given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -27 & 54 & -63 & 44 & -21 & 6 \end{bmatrix} \tag{5.3.15}$$

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -3 & 2 \end{bmatrix}, \tag{5.3.16}$$

and

$$C = \begin{bmatrix} 1+\jmath\sqrt{2} & 1 & 0 & 0 & 0 & 0 \\ 0 & 1+\jmath\sqrt{2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1+\jmath\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\jmath\sqrt{2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1-\jmath\sqrt{2} & 1 \\ 0 & 0 & 0 & 0 & 0 & 1-\jmath\sqrt{2} \end{bmatrix}. \tag{5.3.17}$$

Then, A is in companion form, B is in real hypercompanion form, and C is in Jordan form. Furthermore, A, B, and C are similar.

The next result shows that every matrix is similar to its transpose by means of a symmetric similarity transformation. This result is due to Frobenius.

Corollary 5.3.8. Let $A \in \mathbb{F}^{n\times n}$. Then, there exists a symmetric nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SA^{\mathrm{T}}S^{-1}$.

Proof. It follows from Theorem 5.3.3 that there exists a nonsingular matrix $\hat{S} \in \mathbb{C}^{n\times n}$ such that $A = \hat{S}B\hat{S}^{-1}$, where $B = \operatorname{diag}(B_1, \ldots, B_r)$ is the Jordan form of A and $B_i \in \mathbb{C}^{n_i\times n_i}$ for all $i = 1, \ldots, r$. Now, define the symmetric nonsingular matrix $S \triangleq \hat{S}\tilde{I}\hat{S}^{\mathrm{T}}$, where $\tilde{I} \triangleq \operatorname{diag}\left(\hat{I}_{n_1}, \ldots, \hat{I}_{n_r}\right)$ is symmetric and involutory. Furthermore, note that $\hat{I}_{n_i}B_i\hat{I}_{n_i} = B_i^{\mathrm{T}}$ for all $i = 1, \ldots, r$ so that $\tilde{I}B\tilde{I} = B^{\mathrm{T}}$ and thus $\tilde{I}B^{\mathrm{T}}\tilde{I} = B$. Hence, it follows that

$$\begin{aligned} SA^{\mathrm{T}}S^{-1} &= S\hat{S}^{-\mathrm{T}}B^{\mathrm{T}}\hat{S}^{\mathrm{T}}S^{-1} = \hat{S}\tilde{I}\hat{S}^{\mathrm{T}}\hat{S}^{-\mathrm{T}}B^{\mathrm{T}}\hat{S}^{\mathrm{T}}\hat{S}^{-\mathrm{T}}\tilde{I}\hat{S}^{-1} \\ &= \hat{S}\tilde{I}B^{\mathrm{T}}\tilde{I}\hat{S}^{-1} = \hat{S}B\hat{S}^{-1} = A. \end{aligned}$$

If A is real, then a similar argument based on the real Jordan form shows that S can be chosen to be real. □

Corollary 5.3.9. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist symmetric matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that S_2 is nonsingular and $A = S_1S_2$.

Proof. From Corollary 5.3.8 it follows that there exists a symmetric, nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SA^{\mathrm{T}}S^{-1}$. Now, let $S_1 = SA^{\mathrm{T}}$ and $S_2 = S^{-1}$. Note that S_2 is symmetric and nonsingular. Furthermore, $S_1^{\mathrm{T}} = AS = SA^{\mathrm{T}} = S_1$, which shows that S_1 is symmetric. □

Note that Corollary 5.3.9 follows from Corollary 5.3.8. If $A = S_1S_2$, where S_1, S_2 are symmetric and S_2 is nonsingular, then $A = S_2^{-1}S_2S_1S_2 = S_2^{-1}A^{\mathrm{T}}S_2$.

5.4 Schur Form

Next, we consider a decomposition involving a unitary transformation and an upper triangular matrix called the *Schur form*.

Theorem 5.4.1. Let $A \in \mathbb{C}^{n\times n}$. Then, there exists a unitary matrix $S \in \mathbb{C}^{n\times n}$ and an upper triangular matrix $B \in \mathbb{C}^{n\times n}$ such that

$$A = SBS^*. \tag{5.4.1}$$

Proof. Let $\lambda_1 \in \mathbb{C}$ be an eigenvalue of A with associated eigenvector $x \in \mathbb{C}^n$ chosen such that $x^*x = 1$. Furthermore, let $S_1 \triangleq \begin{bmatrix} x & \hat{S}_1 \end{bmatrix} \in \mathbb{C}^{n\times n}$ be unitary, where $\hat{S}_1 \in \mathbb{C}^{n\times(n-1)}$ satisfies $\hat{S}_1^*S_1 = I_{n-1}$ and $x^*\hat{S}_1 = 0_{1\times(n-1)}$. Then, $S_1e_1 = x$ and

$$\mathrm{col}_1(S_1^{-1}AS_1) = S_1^{-1}Ax = \lambda_1S_1^{-1}x = \lambda_1e_1.$$

Consequently,

$$A = S_1\begin{bmatrix} \lambda_1 & C_1 \\ 0_{(n-1)\times 1} & A_1 \end{bmatrix}S_1^{-1},$$

where $C_1 \in \mathbb{C}^{1\times(n-1)}$ and $A_1 \in \mathbb{C}^{(n-1)\times(n-1)}$. Next, let $S_{20} \in \mathbb{C}^{(n-1)\times(n-1)}$ be a unitary matrix such that

$$A_1 = S_{20}\begin{bmatrix} \lambda_2 & C_2 \\ 0_{(n-2)\times 1} & A_2 \end{bmatrix}S_{20}^{-1},$$

where $C_2 \in \mathbb{C}^{1\times(n-2)}$ and $A_2 \in \mathbb{C}^{(n-2)\times(n-2)}$. Hence,

$$A = S_1S_2\begin{bmatrix} \lambda_1 & C_{11} & C_{12} \\ 0 & \lambda_2 & C_2 \\ 0 & 0 & A_2 \end{bmatrix}S_2^{-1}S_1,$$

where $C_1 = \begin{bmatrix} C_{11} & C_{12} \end{bmatrix}$, $C_{11} \in \mathbb{C}$, and $S_2 \triangleq \left[\begin{smallmatrix} 1 & 0 \\ 0 & S_{20} \end{smallmatrix}\right]$ is unitary. Proceeding in a similar manner yields (5.4.1) with $S \triangleq S_1S_2\cdots S_{n-1}$, where $S_1,\dots,S_{n-1} \in \mathbb{C}^{n\times n}$ are unitary. □

It can be seen that the diagonal entries of B are the eigenvalues of A.

As with the real Jordan form, there exists a *real Schur form*.

Corollary 5.4.2. Let $A \in \mathbb{R}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1,\dots,\lambda_r\}_{\mathrm{m}} \cup \{\nu_1 + \jmath\omega_1, \nu_1 - \jmath\omega_1, \dots, \nu_l + \jmath\omega_l, \nu_l - \jmath\omega_l\}_{\mathrm{m}}$, where $\lambda_1,\dots,\lambda_r \in \mathbb{R}$ and, for all $i = 1,\dots,l$, $\nu_i, \omega_i \in \mathbb{R}$ and $\omega_i \neq 0$. Then, there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = SBS^{\mathrm{T}}, \tag{5.4.2}$$

where B is upper block triangular and the diagonally located blocks $B_1,\dots,B_r \in \mathbb{R}$ and $\hat{B}_1,\dots,\hat{B}_l \in \mathbb{R}^{2\times 2}$ of B are $B_i \triangleq [\lambda_i]$ for all $i = 1,\dots,r$ and $\hat{B}_i \triangleq \left[\begin{smallmatrix} \nu_i & \omega_i \\ \omega_i & -\nu_i \end{smallmatrix}\right]$ for all $i = 1,\dots,l$.

Corollary 5.4.3. Let $A \in \mathbb{R}^{n\times n}$, and assume that A has real spectrum. Then, there exist an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ and an upper triangular matrix $B \in \mathbb{R}^{n\times n}$ such that

$$A = SBS^{\mathrm{T}}. \tag{5.4.3}$$

The Schur decomposition reveals the structure of range-Hermitian matrices and thus, as a special case, normal matrices.

Corollary 5.4.4. Let $A \in \mathbb{F}^{n\times n}$. Then, A is range Hermitian if and only if there exist a unitary matrix $S \in \mathbb{F}^{n\times n}$ and a nonsingular matrix $B \in \mathbb{F}^{r\times r}$, where $r \triangleq \operatorname{rank} A$, such that

$$A = S\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}S^*. \tag{5.4.4}$$

In addition, A is normal if and only if there exist a unitary matrix $S \in \mathbb{C}^{n\times n}$ and a diagonal matrix $B \in \mathbb{C}^{n\times n}$ such that

$$A = SBS^*. \tag{5.4.5}$$

Proof. Suppose that A is range Hermitian and let $A = SBS^*$, where B is the real Schur form of A and $S \in \mathbb{F}^{n\times n}$ is unitary. Assume A is singular and choose S such that $B_{(j,j)} = B_{(j+1,j+1)} = \cdots = B_{(n,n)} = 0$ and such that all other diagonal entries of B are nonzero. Thus, $\mathrm{row}_n(B) = 0$, which implies that $e_n \notin \mathcal{R}(B)$. Since A is range Hermitian, it follows that $\mathcal{R}(B) = \mathcal{R}(B^*)$ so that $e_n \notin \mathcal{R}(B^*)$. Thus, $\mathrm{col}_n(B) = \mathrm{row}_n(B^*) = 0$. If, in addition, $B_{(n-1,n-1)} = 0$, then $\mathrm{col}_{n-1}(B) = 0$. Repeating this argument shows that B has the form $\left[\begin{smallmatrix} \hat{B} & 0 \\ 0 & 0 \end{smallmatrix}\right]$, where $\hat{B}$ is nonsingular.

Now, suppose that A is normal and let $A = SBS^*$, where $B \in \mathbb{C}^{n\times n}$ is upper triangular and $S \in \mathbb{C}^{n\times n}$ is unitary. Since A is normal, it follows that $AA^* = A^*A$, which implies that $BB^* = B^*B$. Since B is upper triangular, it follows that $(B^*B)_{(1,1)} = B_{(1,1)}\overline{B}_{(1,1)}$, whereas $(BB^*)_{(1,1)} = \mathrm{row}_1(B)[\mathrm{row}_1(B)]^* = \sum_{i=1}^n B_{(1,i)}\overline{B}_{(1,i)}$. Since $(B^*B)_{(1,1)} = (BB^*)_{(1,1)}$, it follows that $B_{(1,i)} = 0$ for all $i = 2,\ldots,n$. Continuing in a similar fashion row by row, it follows that B is diagonal. □

Corollary 5.4.5. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, there exist a unitary matrix $S \in \mathbb{F}^{n\times n}$ and a diagonal matrix $B \in \mathbb{R}^{n\times n}$ such that

$$A = SBS^*. \tag{5.4.6}$$

If, in addition, A is (nonnegative semidefinite, positive definite), then the diagonal entries of B are (nonnegative, positive).

Proof. It follows from Corollary 5.4.4 that there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ and a diagonal matrix $B \in \mathbb{F}^{n\times n}$ such that $A = SBS^*$. If A is nonnegative semidefinite, then $x^*Ax \geq 0$ for all $x \in \mathbb{F}^n$. Choosing $x = Se_i$ it follows that $B_{(i,i)} = e_i^{\mathrm{T}}Be_i = e_i^{\mathrm{T}}S^*ASe_i \geq 0$ for all $i = 1,\ldots,n$. If A is positive definite, then $B_{(i,i)} > 0$ for all $i = 1,\ldots,n$. □

Proposition 5.4.6. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = S\begin{bmatrix} -I_{\nu_-(A)} & 0 & 0 \\ 0 & 0_{\nu_0(A)\times\nu_0(A)} & 0 \\ 0 & 0 & I_{\nu_+(A)} \end{bmatrix}S^*. \tag{5.4.7}$$

Furthermore,

$$\mathrm{rank}\, A = \nu_+(A) + \nu_-(A). \tag{5.4.8}$$

Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, the quantity

$$\mathrm{sig}(A) \triangleq \nu_+(A) - \nu_-(A) \tag{5.4.9}$$

is the *signature* of A.

Proof. Since A is Hermitian, it follows from Corollary 5.4.5 that there exist a unitary matrix $\hat{S} \in \mathbb{F}^{n\times n}$ and a diagonal matrix $B \in \mathbb{R}^{n\times n}$ such that $A = \hat{S}B\hat{S}^*$. Choose S to order the diagonal entries of B such that $B = \operatorname{diag}(B_1, 0, -B_2)$, where the diagonal matrices B_1, B_2 are both positive definite. Now, define $\hat{B} \triangleq \operatorname{diag}(B_1, I, B_2)$. Then, $B = \hat{B}^{1/2}D\hat{B}^{1/2}$, where $D = \operatorname{diag}\big(I_{\nu_-(A)}, 0_{\nu_0(A)\times\nu_0(A)}, -I_{\nu_+(A)}\big)$. Consequently, $A = \hat{S}\hat{B}^{1/2}D\hat{B}^{1/2}\hat{S}^*$. □

Corollary 5.4.7. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, A and B are congruent if and only if $\operatorname{In}(A) = \operatorname{In}(B)$.

In Proposition 4.5.3 it was shown that eigenvectors associated with a collection of distinct eigenvalues of a normal matrix are mutually orthogonal. Thus, a normal matrix will have at least as many mutually orthogonal eigenvectors as it has distinct eigenvalues. The next result, which is an immediate consequence of Corollary 5.4.4, shows that every $n \times n$ normal matrix actually has n mutually orthogonal eigenvectors. In fact, the converse is also true.

Corollary 5.4.8. Let $A \in \mathbb{C}^{n\times n}$. Then, A is normal if and only if A has n mutually orthogonal eigenvectors.

There is also a *real normal form*, which is analogous to the real Schur form.

Corollary 5.4.9. Let $A \in \mathbb{R}^{n\times n}$ be range symmetric. Then, there exist an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ and a nonsingular matrix $B \in \mathbb{R}^{r\times r}$, where $r \triangleq \operatorname{rank} A$, such that

$$A = S\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}S^{\mathrm{T}}. \tag{5.4.10}$$

In addition, assume that A is normal and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_r\}_{\mathrm{m}} \cup \{\nu_1 + \jmath\omega_1, \nu_1 - \jmath\omega_1, \ldots, \nu_l + \jmath\omega_l, \nu_l - \jmath\omega_l\}_{\mathrm{m}}$, where $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ and, for all $i = 1, \ldots, l$, $\nu_i, \omega_i \in \mathbb{R}$ and $\omega_i \neq 0$. Then, there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = SBS^{\mathrm{T}}, \tag{5.4.11}$$

where $B \triangleq \operatorname{diag}(B_1, \ldots, B_r, \hat{B}_1, \ldots, \hat{B}_l)$, $B_i \triangleq [\lambda_i]$ for all $i = 1, \ldots, r$, and $\hat{B}_i \triangleq \left[\begin{smallmatrix} \nu_i & \omega_i \\ -\omega_i & \nu_i \end{smallmatrix}\right]$ for all $i = 1, \ldots, l$.

5.5 Eigenstructure Properties

Definition 5.5.1. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \mathbb{C}$. Then, the *index of* λ *with respect to* A, denoted by $\operatorname{ind}_A(\lambda)$, is the smallest nonnegative integer k such that

$$\mathcal{R}\Big[(\lambda I - A)^k\Big] = \mathcal{R}\Big[(\lambda I - A)^{k+1}\Big]. \tag{5.5.1}$$

Furthermore, the *index of* A, denoted by $\operatorname{ind} A$, is the smallest nonnegative integer k such that

$$\mathcal{R}\left(A^k\right) = \mathcal{R}\left(A^{k+1}\right), \tag{5.5.2}$$

that is, $\operatorname{ind} A = \operatorname{ind}_A(0)$.

Note that $\lambda \notin \operatorname{spec}(A)$ if and only if $\operatorname{ind}_A(\lambda) = 0$. Hence, $0 \notin \operatorname{spec}(A)$ if and only if $\operatorname{ind} A = \operatorname{ind}_A(0) = 0$. Hence, A is nonsingular if and only if $\operatorname{ind} A = 0$.

Proposition 5.5.2. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \mathbb{C}$. Then, $\operatorname{ind}_A(\lambda)$ is the smallest nonnegative integer k such that

$$\operatorname{rank}\left[(\lambda I - A)^k\right] = \operatorname{rank}\left[(\lambda I - A)^{k+1}\right]. \tag{5.5.3}$$

Furthermore, $\operatorname{ind} A$ is the smallest nonnegative integer k such that

$$\operatorname{rank}\left(A^k\right) = \operatorname{rank}\left(A^{k+1}\right). \tag{5.5.4}$$

Proof. Corollary 2.4.2 implies that $\mathcal{R}\big[(\lambda I - A)^k\big] \subseteq \mathcal{R}\big[(\lambda I - A)^{k+1}\big]$. Now, Lemma 2.3.4 implies that $\mathcal{R}\big[(\lambda I - A)^k\big] = \mathcal{R}\big[(\lambda I - A)^{k+1}\big]$ if and only if $\operatorname{rank}\big[(\lambda I - A)^k\big] = \operatorname{rank}\big[(\lambda I - A)^{k+1}\big]$. □

Proposition 5.5.3. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \operatorname{spec}(A)$. Then, the following statements hold:

i) $\operatorname{ind}_A(\lambda)$ is the order of the largest Jordan block of A associated with λ.

ii) $\operatorname{gm}_A(\lambda)$ is equal to the number of Jordan blocks of A associated with λ.

iii) $\operatorname{ind}_A(\lambda) \le \operatorname{am}_A(\lambda)$.

iv) $\operatorname{gm}_A(\lambda) \le \operatorname{am}_A(\lambda)$.

v) $\operatorname{ind}_A(\lambda) + \operatorname{gm}_A(\lambda) \le \operatorname{am}_A(\lambda) + 1$.

vi) $\operatorname{rank} A = n - \operatorname{gm}_A(0)$.

Proposition 5.5.4. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be a subspace. Then, there exists a unique projector $A \in \mathbb{F}^{n\times n}$ such that $\mathcal{S} = \mathcal{R}(A)$. Furthermore, $x \in \mathcal{S}$ if and only if $x = Ax$.

Proof. See [416, p. 386]. □

For a subspace $\mathcal{S} \subseteq \mathbb{F}^n$, the matrix $A \in \mathbb{F}^{n\times n}$ given by Proposition 5.5.4 is the *projector onto* $\mathcal{S}$.

Let $A \in \mathbb{F}^{n\times n}$ be an idempotent matrix. Then, the *complementary idempotent matrix* defined by

$$A_\perp \triangleq I - A \tag{5.5.5}$$

is also idempotent. If A is a projector, then $A_\perp$ is the *complementary projector*.

Proposition 5.5.5. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be a subspace and let $A \in \mathbb{F}^{n\times n}$ be the projector onto $\mathcal{S}$. Then, $A_\perp$ is the projector onto $\mathcal{S}^\perp$. Furthermore,

$$\mathcal{R}(A)^\perp = \mathcal{N}(A) = \mathcal{R}(A_\perp). \tag{5.5.6}$$

Proposition 5.5.6. Let $A \in \mathbb{F}^{n\times n}$, and let k be a positive integer. Then, $\operatorname{ind} A \le k$ if and only if $\mathcal{R}(A^k)$ and $\mathcal{N}(A^k)$ are complementary subspaces.

Corollary 5.5.7. Let $A \in \mathbb{F}^{n\times n}$. Then, A is group invertible if and only if $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are complementary subspaces.

Proposition 5.5.8. Let $A \in \mathbb{F}^{n\times n}$, and let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$ be complementary subspaces. Then, there exists a unique idempotent matrix $A \in \mathbb{F}^{n\times n}$ such that $\mathcal{R}(A) = \mathcal{S}_1$ and $\mathcal{N}(A) = \mathcal{S}_2$. Furthermore, $\mathcal{R}(A_\perp) = \mathcal{S}_2$ and $\mathcal{N}(A_\perp) = \mathcal{S}_1$.

Proof. See [82, p. 118] or [416, p. 386]. □

For complementary subspaces $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{F}^n$, the unique idempotent matrix $A \in \mathbb{F}^{n\times n}$ given by Proposition 5.5.8 is the *idempotent matrix onto* $\mathcal{S}_1 = \mathcal{R}(A)$ *along* $\mathcal{S}_2 = \mathcal{N}(A)$.

Proposition 5.5.9. Let $A \in \mathbb{F}^{n\times n}$, and let $r \triangleq \operatorname{rank} A$. Then, A is group invertible if and only if there exist $B \in \mathbb{F}^{n\times r}$ and $C \in \mathbb{F}^{r\times n}$ such that $\operatorname{rank} B = \operatorname{rank} C = r$. Furthermore, the idempotent matrix $P \triangleq B(CB)^{-1}C$ is the idempotent matrix onto $\mathcal{R}(A)$ along $\mathcal{N}(A)$.

Proof. See [416, p. 634]. □

An alternative expression for the idempotent matrix onto $\mathcal{R}(A)$ along $\mathcal{N}(A)$ is given by Proposition 6.2.2.

Definition 5.5.10. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \operatorname{spec}(A)$. Then, the following terminology is defined:

i) λ is *simple* if $\operatorname{am}_A(\lambda) = 1$.

ii) A is *simple* if every eigenvalue of A is simple.

iii) λ is *cyclic* if $\mathrm{gm}_A(\lambda) = 1$.

iv) A is *cyclic* if every eigenvalue of A is cyclic.

v) λ is *derogatory* if $\mathrm{gm}_A(\lambda) > 1$.

vi) A is *derogatory* if A has at least one derogatory eigenvalue.

vii) λ is *semisimple* if $\mathrm{gm}_A(\lambda) = \mathrm{am}_A(\lambda)$.

viii) A is *semisimple* if every eigenvalue of A is semisimple.

ix) λ is *defective* if $\mathrm{gm}_A(\lambda) < \mathrm{am}_A(\lambda)$.

x) A is *defective* if A has at least one defective eigenvalue.

xi) A is *diagonalizable over* $\mathbb{C}$ if A is *semisimple.*

xii) $A \in \mathbb{R}^{n\times n}$ is *diagonalizable over* $\mathbb{R}$ if A is *semisimple* and every eigenvalue of A is real.

Proposition 5.5.11. Let $A \in \mathbb{F}^{n\times n}$ and $\lambda \in \mathrm{spec}(A)$. Then, λ is simple if and only if λ is cyclic and semisimple.

Proposition 5.5.12. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \mathrm{spec}(A)$. Then,

$$\mathrm{def}\Big[(\lambda I - A)^{\mathrm{ind}_A(\lambda)}\Big] = \mathrm{am}_A(\lambda). \tag{5.5.7}$$

Theorem 5.3.3 yields the following result, which shows that the subspaces $\mathcal{N}\big[(\lambda I - A)^k\big]$, where $\lambda \in \mathrm{spec}(A)$ and $k = \mathrm{ind}_A(\lambda)$, provide a decomposition of $\mathbb{F}^n$.

Proposition 5.5.13. Let $A \in \mathbb{F}^{n\times n}$, let $\mathrm{spec}(A) = \{\lambda_1, \ldots, \lambda_r\}$, and, for all $i = 1, \ldots, r$, let $k_i \triangleq \mathrm{ind}_A(\lambda_i)$. Then, the following statements hold:

i) $\mathcal{N}\big[(\lambda_i I - A)^{k_i}\big] \cap \mathcal{N}\big[(\lambda_j I - A)^{k_j}\big] = \{0\}$ for all $i, j = 1, \ldots, r$ such that $i \neq j$.

ii) $\sum_{i=1}^r \mathcal{N}\big[(\lambda_i I - A)^{k_i}\big] = \mathbb{F}^n$.

Proposition 5.5.14. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \mathrm{spec}(A)$. Then, the following statements are equivalent:

i) λ is semisimple.

ii) $\mathrm{def}(\lambda I - A) = \mathrm{def}\big[(\lambda I - A)^2\big]$.

iii) $\mathcal{N}(\lambda I - A) = \mathcal{N}\big[(\lambda I - A)^2\big]$.

iv) $\mathrm{ind}_A(\lambda) = 1$.

Proof. To prove that *i)* implies *ii)*, suppose that λ is semisimple so that $\mathrm{gm}_A(\lambda) = \mathrm{am}_A(\lambda)$ and thus $\mathrm{def}(\lambda I - A) = \mathrm{am}_A(\lambda)$. Then, it follows from Proposition 5.5.12 that $\mathrm{def}\big[(\lambda I - A)^k\big] = \mathrm{am}_A(\lambda)$, where $k \triangleq \mathrm{ind}_A(\lambda)$. Therefore, it follows from Corollary 2.5.6 that $\mathrm{am}_A(\lambda) = \mathrm{def}(\lambda I - A) \le \mathrm{def}\big[(\lambda I - A)^2\big] \le \mathrm{def}\big[(\lambda I - A)^k\big] = \mathrm{am}_A(\lambda)$, which implies that $\mathrm{def}(\lambda I - A) = \mathrm{def}\big[(\lambda I - A)^2\big]$.

To prove that *ii*) implies *iii*), note that it follows from Corollary 2.5.6 that $\mathcal{N}(\lambda I - A) \subseteq \mathcal{N}\big[(\lambda I - A)^2\big]$. Since, by *ii*), these subspaces have equal dimension, it follows from Lemma 2.3.4 that these subspaces are equal. Conversely, *iii*) implies *ii*).

Finally, *iv*) is equivalent to the fact that every Jordan block of A associated with λ has order 1, which is equivalent to the fact that the geometric multiplicity of λ is equal to the algebraic multiplicity of λ, that is, that λ is semisimple. □

Corollary 5.5.15. Let $A \in \mathbb{F}^{n\times n}$. Then, A is group invertible if and only if $\mathrm{ind}\,A \le 1$.

Proposition 5.5.16. Suppose $A, B \in \mathbb{F}^{n\times n}$ are similar. Then, the following statements hold:

i) $\mathrm{mspec}(A) = \mathrm{mspec}(B)$.

ii) For all $\lambda \in \mathrm{spec}(A)$, $\mathrm{gm}_A(\lambda) = \mathrm{gm}_B(\lambda)$.

Proposition 5.5.17. Let $A \in \mathbb{F}^{n\times n}$. Then, A is semisimple if and only if A is similar to a normal matrix.

The following result is an extension of Corollary 5.3.9.

Proposition 5.5.18. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is diagonalizable over $\mathbb{R}$.

ii) There exists a positive-definite matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SA^*S^{-1}$.

iii) There exist a Hermitian matrix $S_1 \in \mathbb{F}^{n\times n}$ and a positive-definite matrix $S_2 \in \mathbb{F}^{n\times n}$ such that $A = S_1S_2$.

Proof. To prove that *i)* implies *ii)*, let nonsingular $\hat{S} \in \mathbb{F}^{n\times n}$ be such that $A = \hat{S}B\hat{S}^{-1}$, where $B \in \mathbb{R}^{n\times n}$ is diagonal. Then, $B = \hat{S}^{-1}A\hat{S} = \hat{S}^*A^*\hat{S}^{-*}$. Hence, $A = \hat{S}B\hat{S}^{-1} = \hat{S}\Big(\hat{S}^*A^*\hat{S}^{-*}\Big)\hat{S}^{-1} = \Big(\hat{S}\hat{S}^*\Big)A^*\Big(\hat{S}\hat{S}^*\Big)^{-1} = SA^*S^{-1}$, where $S \triangleq \hat{S}\hat{S}^*$ is positive definite. To show that *ii)* implies *iii)*, note that $A = SA^*S^{-1} = S_1S_2$, where $S_1 \triangleq SA^*$ and $S_2 = S^{-1}$. Since $S_1^* = (SA^*)^* = AS^* = AS = SA^* = S_1$, it follows that S_1 is Hermitian. Furthermore, since S is positive definite, it follows that S^{-1}, and hence S_2, is also positive definite. Finally, to prove that *iii)* implies *i)*, note that $A = S_1S_2 = S_2^{-1/2}\Big(S_2^{1/2}S_1S_2^{1/2}\Big)S_2^{1/2}$. Since $S_2^{1/2}S_1S_2^{1/2}$ is Hermitian, it follows from Corollary 5.4.5 that $S_2^{1/2}S_1S_2^{1/2}$ is diagonalizable over $\mathbb{R}$. Consequently, A is diagonalizable over $\mathbb{R}$. □

If a matrix is block triangular, then the following result shows that its eigenvalues and their algebraic multiplicity are determined by the diagonally located blocks. If, in addition, the matrix is block diagonal, then the geometric multiplicities of its eigenvalues are determined by the diagonally located blocks.

Proposition 5.5.19. Let $A \in \mathbb{F}^{n\times n}$ be either upper block triangular or lower block triangular with diagonally located blocks $A_{11}, \ldots, A_{rr}$, where $A_{ii} \in \mathbb{F}^{n_i\times n_i}$ for all $i = 1, \ldots, r$. Then,

$$\mathrm{am}_A(\lambda) = \sum_{i=1}^r \mathrm{am}_{A_{ii}}(\lambda). \tag{5.5.8}$$

Hence,

$$\mathrm{mspec}(A) = \bigcup_{i=1}^r \mathrm{mspec}(A_{ii}). \tag{5.5.9}$$

Now, assume that A is block diagonal. Then,

$$\mathrm{gm}_A(\lambda) = \sum_{i=1}^r \mathrm{gm}_{A_{ii}}(\lambda). \tag{5.5.10}$$

Proposition 5.5.20. Let $A \in \mathbb{F}^{n \times n}$, let $\operatorname{spec}(A) = \{\lambda_1, \ldots, \lambda_r\}$, and let $k_i \triangleq \operatorname{ind}_A(\lambda_i)$ for all $i = 1, \ldots, r$. Then,

$$\mu_A(s) = \prod_{i=1}^{r} (s - \lambda_i)^{k_i} \tag{5.5.11}$$

and

$$\deg \mu_A = \sum_{i=1}^{r} k_i. \tag{5.5.12}$$

Furthermore, the following statements are equivalent:

i) $\mu_A = \chi_A$.

ii) A is cyclic.

iii) For all $\lambda \in \operatorname{spec}(A)$, the Jordan form of A contains exactly one block associated with λ.

Proof. Let $A = SBS^{-1}$, where $B = \operatorname{diag}(B_1, \ldots, B_{n_\mathrm{h}})$ denotes the Jordan form of A given by (5.3.4). Let $\lambda_i \in \operatorname{spec}(A)$, and let B_j be a Jordan block associated with λ_i. Then, the order of B_j is less than or equal to k_i. Consequently, $(B_j - \lambda_i I)^{k_i} = 0$.

Next, let $p(s)$ denote the right-hand side of (5.5.11). Thus,

$$\begin{aligned} p(A) &= \prod_{i=1}^{r} (A - \lambda_i I)^{k_i} = S \left[\prod_{i=1}^{r} (B - \lambda_i I)^{k_i} \right] S^{-1} \\ &= S \operatorname{diag}\left(\prod_{i=1}^{r} (B_1 - \lambda_i I)^{k_i}, \ldots, \prod_{i=1}^{r} (B_{n_\mathrm{h}} - \lambda_i I)^{k_i} \right) S^{-1} = 0. \end{aligned}$$

Therefore, it follows from Theorem 4.6.1 that μ_A divides p. Furthermore, note that if k_i is replaced by $\hat{k}_i < k_i$, then $p(A) \neq 0$. Hence, p is the minimal polynomial of A. The equivalence of *i)* and *ii)* is now immediate, while the equivalence of *ii)* and *iii)* follows from Theorem 5.3.5. □

Example 5.5.21. The matrix $\left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$ is normal but is neither symmetric nor skew symmetric, while the matrix $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ is normal but is neither symmetric nor semisimple with real eigenvalues.

Example 5.5.22. The matrices $\left[\begin{smallmatrix} 1 & 0 \\ 2 & -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right]$ are diagonalizable over $\mathbb{R}$ but not normal, while the matrix $\left[\begin{smallmatrix} -1 & 1 \\ -2 & 1 \end{smallmatrix}\right]$ is diagonalizable but is neither normal nor diagonalizable over $\mathbb{R}$.

Example 5.5.23. The product of the Hermitian matrices $\left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 & 1 \\ 1 & -2 \end{smallmatrix}\right]$ has has no real eigenvalues.

Example 5.5.24. The matrices $\left[\begin{smallmatrix}1 & 0\\ 0 & 2\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0 & 1\\ -2 & 3\end{smallmatrix}\right]$ are similar, whereas $\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0 & 1\\ -1 & 2\end{smallmatrix}\right]$ have the same spectrum but are not similar.

Proposition 5.5.25. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) A is singular if and only if $0 \in \operatorname{spec}(A)$.

ii) A is group invertible if and only if either A is nonsingular or $0 \in \operatorname{spec}(A)$ is semisimple.

iii) A is Hermitian if and only if A is normal and $\operatorname{spec}(A) \subset \mathbb{R}$.

iv) A is skew Hermitian if and only if A is normal and $\operatorname{spec}(A) \subset \jmath\mathbb{R}$.

v) A is nonnegative semidefinite if and only if A is normal and $\operatorname{spec}(A) \subset [0,\infty)$.

vi) A is positive definite if and only if A is normal and $\operatorname{spec}(A) \subset (0,\infty)$.

vii) A is unitary if and only if A is normal and $\operatorname{spec}(A) \subset \{\lambda \in \mathbb{C}\colon\ |\lambda| = 1\}$.

viii) A is involutory if and only if A is semisimple and $\operatorname{spec}(A) \subseteq \{-1, 1\}$.

ix) A is skew involutory if and only if A is semisimple and $\operatorname{spec}(A) \subseteq \{-\jmath, \jmath\}$.

x) A is idempotent if and only if A is semisimple and $\operatorname{spec}(A) \subseteq \{0, 1\}$.

xi) A is tripotent if and only if A is semisimple and $\operatorname{spec}(A) \subseteq \{-1, 0, 1\}$.

xii) A is nilpotent if and only if $\operatorname{spec}(A) = \{0\}$.

xiii) A is a projector if and only if A is normal and $\operatorname{spec}(A) = \{0, 1\}$.

xiv) A is a reflector if and only if A is normal and $\operatorname{spec}(A) = \{-1, 1\}$.

xv) A is an elementary projector if and only if A is normal and $\operatorname{mspec}(A) = \{0, 1, \ldots, 1\}_{\mathrm{m}}$.

xvi) A is an elementary reflector if and only if A is normal and $\operatorname{mspec}(A) = \{-1, 1, \ldots, 1\}_{\mathrm{m}}$.

xvii) A is an elementary matrix if and only if A is normal and $\operatorname{mspec}(A) = \{\alpha, 1, \ldots, 1\}_{\mathrm{m}}$, where $\alpha \neq 0$.

If, furthermore, $A \in \mathbb{R}^{2n\times 2n}$, then the following statements hold:

xviii) If A is Hamiltonian, then $\operatorname{mspec}(A) = -\operatorname{mspec}(A)$.

ix) If A is symplectic, then $\operatorname{mspec}(A) = \{1/\lambda\colon\ \lambda \in \operatorname{mspec}(A)\}_{\mathrm{m}}$.

5.6 Singular Value Decomposition

The third matrix decomposition that we consider is the *singular value decomposition*. Unlike the Jordan and Schur decompositions, the singular value decomposition applies to matrices that are not necessarily square. Let $A \in \mathbb{F}^{n\times m}$, where $A \neq 0$, and consider the nonnegative-semidefinite matrices $AA^* \in \mathbb{F}^{n\times n}$ and $A^*A \in \mathbb{F}^{m\times m}$. It follows from Proposition 4.4.9 that AA^* and A^*A have the same nonzero eigenvalues with the same algebraic multiplicities. Since AA^* and A^*A are nonnegative semidefinite, it follows that they have the same *positive* eigenvalues with the same algebraic multiplicities. Furthermore, since AA^* is Hermitian, it follows that the number of positive eigenvalues of AA^* (or A^*A) counting algebraic multiplicity is equal to the rank of AA^* (or A^*A). Since $\operatorname{rank} A = \operatorname{rank} AA^* = \operatorname{rank} A^*A$, it thus follows that AA^* and A^*A both have r positive eigenvalues, where $r \triangleq \operatorname{rank} A$.

Definition 5.6.1. Let $A \in \mathbb{F}^{n\times m}$. Then, the *singular values* of A are the $\min\{n,m\}$ nonnegative numbers $\sigma_1(A),\ldots,\sigma_{\min\{n,m\}}(A)$, where, for all $i = 1,\ldots,\min\{n,m\}$,

$$\sigma_i(A) \triangleq \begin{cases} [\lambda_i(AA^*)]^{1/2}, & n \le m, \\ [\lambda_i(A^*A)]^{1/2}, & m \le n. \end{cases} \tag{5.6.1}$$

Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\sigma_1(A) \ge \cdots \ge \sigma_{\min\{n,m\}}(A) \ge 0. \tag{5.6.2}$$

If $A \neq 0$, then

$$\sigma_1(A) \ge \cdots \ge \sigma_r(A) > \sigma_{r+1}(A) = \cdots = \sigma_{\min\{n,m\}}(A) = 0, \tag{5.6.3}$$

where $r \triangleq \operatorname{rank} A$. For convenience, define

$$\sigma_{\max}(A) \triangleq \sigma_1(A), \tag{5.6.4}$$

and, if $n = m$,

$$\sigma_{\min}(A) \triangleq \sigma_n(A). \tag{5.6.5}$$

Note that

$$\sigma_{\max}(0_{n\times n}) = \sigma_{\min}(0_{n\times n}) = 0, \tag{5.6.6}$$

and, for all $i = 1,\ldots,\min\{n,m\}$,

$$\sigma_i(A) = \sigma_i(A^*) = \sigma_i(\overline{A}) = \sigma_i\big(A^{\mathrm{T}}\big). \tag{5.6.7}$$

Proposition 5.6.2. Let $A \in \mathbb{F}^{n\times m}$, where $A \neq 0$. Then, the following statements are equivalent:

i) $\operatorname{rank} A = n$.

ii) $\sigma_n(A) > 0$.

The following statements are also equivalent:

iii) $\operatorname{rank} A = m$.

iv) $\sigma_m(A) > 0$.

Now, assume that $n = m$. Then, the following statements are also equivalent:

v) A is nonsingular.

vi) $\sigma_{\min}(A) > 0$.

We now state the singular value decomposition.

Theorem 5.6.3. Let $A \in \mathbb{F}^{n\times m}$ where $A \neq 0$, let $r \triangleq \operatorname{rank} A$, and define $B \triangleq \operatorname{diag}[\sigma_1(A), \ldots, \sigma_r(A)]$. Then, there exist unitary matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ such that

$$A = S_1 \begin{bmatrix} B & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} S_2. \tag{5.6.8}$$

Proof. For convenience, assume $r < \min\{n, m\}$, since otherwise the zero matrices become empty matrices. By Corollary 5.4.5 there exists a unitary matrix $U \in \mathbb{F}^{n\times n}$ such that

$$AA^* = U \begin{bmatrix} B^2 & 0 \\ 0 & 0 \end{bmatrix} U^*.$$

Partition $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$, where $U_1 \in \mathbb{F}^{n\times r}$ and $U_2 \in \mathbb{F}^{n\times(n-r)}$. Since $U^*U = I_n$, it follows that $U_1^*U_1 = I_r$ and $U_1^*U = \begin{bmatrix} I_r & 0_{r\times(n-r)} \end{bmatrix}$. Now, define $V_1 \triangleq A^*U_1B^{-1} \in \mathbb{F}^{m\times r}$ and note that

$$V_1^*V_1 = B^{-1}U_1^*AA^*U_1B^{-1} = B^{-1}U_1^*U \begin{bmatrix} B^2 & 0 \\ 0 & 0 \end{bmatrix} U^*U_1B^{-1} = I_r.$$

Next, note that, since $U_2^*U = \begin{bmatrix} 0_{(n-r)\times r} & I_{n-r} \end{bmatrix}$, it follows that

$$U_2^*AA^* = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} B^2 & 0 \\ 0 & 0 \end{bmatrix} U^* = 0.$$

However, since $\mathcal{R}(A) = \mathcal{R}(AA^*)$, it follows that $U_2^*A = 0$. Finally, let $V_2 \in \mathbb{F}^{m\times(m-r)}$ be such that $V \triangleq \begin{bmatrix} V_1 & V_2 \end{bmatrix} \in \mathbb{F}^{m\times m}$ is unitary. Hence, we have

$$\begin{aligned} U \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} V^* &= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = U_1BV_1^* = U_1BB^{-1}U_1^*A \\ &= U_1U_1^*A = (U_1U_1^* + U_2U_2^*)A = UU^*A = A, \end{aligned}$$

which yields (5.6.8) with $S_1 = U$ and $S_2 = V^*$. □

An immediate corollary of the singular value decomposition is the *polar decomposition.*

Corollary 5.6.4. Let $A \in \mathbb{F}^{n\times n}$. Then, there exists a nonnegative-semidefinite matrix $M \in \mathbb{F}^{n\times n}$ and a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = MS. \tag{5.6.9}$$

Proof. It follows from the singular value decomposition that there exist unitary matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ and a diagonal positive-definite matrix $B \in \mathbb{F}^{r\times r}$, where $r \triangleq \operatorname{rank} A$, such that $A = S_1\left[\begin{smallmatrix} B & 0 \\ 0 & 0 \end{smallmatrix}\right]S_2$. Hence,

$$A = S_1\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}S_1^*S_1S_2 = MS,$$

where $M \triangleq S_1\left[\begin{smallmatrix} B & 0 \\ 0 & 0 \end{smallmatrix}\right]S_1^*$ is nonnegative semidefinite and $S \triangleq S_1S_2$ is unitary. □

Proposition 5.6.5. Let $A \in \mathbb{F}^{n\times m}$, let $r \triangleq \operatorname{rank} A$, and define the Hermitian matrix $\mathcal{A} \triangleq \left[\begin{smallmatrix} 0 & A \\ A^* & 0 \end{smallmatrix}\right] \in \mathbb{F}^{(n+m)\times(n+m)}$. Then, $\operatorname{rank} \mathcal{A} = 2r$, and the $2r$ nonzero eigenvalues of $\mathcal{A}$ are the r positive singular values of A and their negatives.

Proof. Since $\chi_{\mathcal{A}}(s) = s^2I - A^*A$, it follows that

$$\operatorname{mspec}(\mathcal{A})\backslash\{0,\ldots,0\}_{\mathrm{m}} = \{\sigma_1(A), -\sigma_1(A), \ldots, \sigma_r(A), -\sigma_r(A)\}_{\mathrm{m}}. \qquad \square$$

5.7 Facts on Matrix Transformations Involving One Matrix

Fact 5.7.1. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{spec}(A) = \{1\}$. Then, A^k is similar to A for all $k \in \mathbb{P}$.

Fact 5.7.2. Let $A \in \mathbb{F}^{n\times n}$ be normal. Then, the Schur form of A is equal to the Jordan form of A.

Fact 5.7.3. Let $A \in \mathbb{R}^{n\times n}$. Then, there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that $-1 \notin \operatorname{spec}(S)$ and SAS^{T} is diagonal. (Proof: See [466, p. 101].) (Remark: This result is due to Hsu.)

Fact 5.7.4. Let $A \in \mathbb{F}^{n\times n}$, and assume there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $S^{-1}AS$ is upper triangular. Then, for all $r = 1,\ldots,n$, $\mathcal{R}\big(S\left[\begin{smallmatrix} I_r \\ 0 \end{smallmatrix}\right]\big)$ is an invariant subspace of A. (Remark: Analogous results hold for lower triangular matrices and for block-triangular matrices.)

Fact 5.7.5. Let $A \in \mathbb{F}^{n\times n}$. Then, $\left[\begin{smallmatrix} A & 0 \\ 0 & -A \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right]$ are unitarily similar. (Proof: Use the unitary transformation $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix} I & -I \\ I & I \end{smallmatrix}\right]$.)

Fact 5.7.6. Let $A \in \mathbb{F}^{n\times n}$. Then, there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that S^*AS has equal diagonal entries. (Remark: The diagonal entries are equal to $(1/n)\operatorname{tr} A$.) (Proof: See [206] or [466, p. 78]. This result is due to Parker. See [221].)

Fact 5.7.7. Let $A \in \mathbb{R}^{n\times n}$, and assume that A is not of the form aI, where $a \in \mathbb{R}$. Then, A is similar to a matrix with diagonal entries $0, \ldots, 0, \operatorname{tr} A$. (Proof: See [466, p. 77].) (Remark: This result is due to Gibson.)

Fact 5.7.8. Let $A \in \mathbb{R}^{n\times n}$, and assume that A is not zero. Then, A is similar to a matrix all of whose diagonal entries are nonzero. (Proof: See [466, p. 79].) (Remark: This result is due to Marcus and Purves.)

Fact 5.7.9. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian, let $S \in \mathbb{F}^{m\times n}$, and assume that $\operatorname{rank} S = n$. Then, $\nu_+(SAS^{\mathrm{T}}) = \nu_+(A)$ and $\nu_-(SAS^{\mathrm{T}}) = \nu_-(A)$. (Proof: See [216, p. 194].)

Fact 5.7.10. Let $A \in \mathbb{F}^{n\times n}$ be symmetric. Then, there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = SBS^{\mathrm{T}},$$

where

$$B \triangleq \operatorname{diag}[\sigma_1(A), \ldots, \sigma_n(A)].$$

(Proof: See [287, p. 207].) (Remark: A is symmetric, complex, and T-congruent to B.)

Fact 5.7.11. Let $A \in \mathbb{F}^{n\times n}$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ and a skew-Hermitian matrix $B \in \mathbb{F}^{n\times n}$ such that

$$A = S\left(\begin{bmatrix} I_{\nu_+(A+A^*)} & 0 & 0 \\ 0 & 0_{\nu_0(A+A^*)\times\nu_0(A+A^*)} & 0 \\ 0 & 0 & -I_{\nu_-(A+A^*)} \end{bmatrix} + B\right)S^*.$$

(Proof: Write $A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*)$ and apply Proposition 5.4.6 to $\frac{1}{2}(A + A^*)$.)

Fact 5.7.12. Let $A \in \mathbb{F}^{n\times n}$, and let $r \triangleq \operatorname{rank} A$. Then, A is group invertible if and only if there exist a nonsingular matrix $B \in \mathbb{F}^{r\times r}$ and a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} S^{-1}.$$

Fact 5.7.13. Let $A \in \mathbb{F}^{n\times n}$ be normal. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A^{\mathrm{T}} = SAS^{-1}$$

and such that $S = S^{\mathrm{T}}$ and $S^{-1} = \overline{S}$. (Remark: If $\mathbb{F} = \mathbb{R}$, then S is a reflector.) (Proof: For $\mathbb{F} = \mathbb{C}$, let $A = UBU^*$, where U is unitary and B is diagonal. Then, $A^{\mathrm{T}} = SA\overline{S}$, where $S \triangleq \overline{U}U^{-1}$. For $\mathbb{F} = \mathbb{R}$, use the real normal form and let $S \triangleq U\tilde{I}U^{\mathrm{T}}$, where U is orthogonal and $\tilde{I} \triangleq \operatorname{diag}(\hat{I}, \ldots, \hat{I})$.)

Fact 5.7.14. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, there exists an involutory matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A^{\mathrm{T}} = SAS^{\mathrm{T}}.$$

(Remark: A^{T}, not A^*.) (Proof: See [240].)

Fact 5.7.15. Let $n \in \mathbb{P}$. Then,

$$\hat{I}_n = \begin{cases} S\begin{bmatrix} -I_{n/2} & 0 \\ 0 & -I_{n/2} \end{bmatrix} S^{\mathrm{T}}, & n \text{ even}, \\[2ex] S\begin{bmatrix} -I_{n/2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_{n/2} \end{bmatrix} S^{\mathrm{T}}, & n \text{ odd}, \end{cases}$$

where

$$S \triangleq \begin{cases} \frac{1}{\sqrt{2}}\begin{bmatrix} I_{n/2} & -\hat{I}_{n/2} \\ \hat{I}_{n/2} & I_{n/2} \end{bmatrix}, & n \text{ even}, \\[2ex] \frac{1}{\sqrt{2}}\begin{bmatrix} I_{n/2} & 0 & -\hat{I}_{n/2} \\ 0 & \sqrt{2} & 0 \\ \hat{I}_{n/2} & 0 & I_{n/2} \end{bmatrix}, & n \text{ odd}. \end{cases}$$

Therefore,

$$\operatorname{mspec}\left(\hat{I}_n\right) = \begin{cases} \{-1, 1, \ldots, -1, 1\}_{\mathrm{m}}, & n \text{ even}, \\ \{1, -1, 1, \ldots, -1, 1\}_{\mathrm{m}}, & n \text{ odd}. \end{cases}$$

(Remark: See [590].)

Fact 5.7.16. Let $A \in \mathbb{F}^{n\times n}$ be unitary and let $m \le n/2$. Then, there exist unitary matrices $U, V \in \mathbb{F}^{n\times n}$ such that

$$A = U\begin{bmatrix} \Gamma & -\Sigma & 0 \\ \Sigma & \Gamma & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix} V,$$

where $\Gamma, \Sigma \in \mathbb{R}^{m\times m}$ are diagonal and nonnegative semidefinite and satisfy

$$\Gamma^2 + \Sigma^2 = I_m.$$

(Proof: See [525, p. 37].) (Remark: This result is the *CS decomposition.*)

Fact 5.7.17. Let $A \in \mathbb{C}^{n\times n}$. Then, there exists $B \in \mathbb{R}^{n\times n}$ such that $A\overline{A}$ and B^2 are similar. (Proof: See [180].)

Fact 5.7.18. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) $\operatorname{tr} A = 0$.

ii) There exist $B, C \in \mathbb{F}^{n\times n}$ such that $A = [B, C]$.

iii) A is unitarily similar to a matrix whose diagonal entries are zero.

(Remark: This result is *Shoda's theorem.* See [4, 220, 325, 333] or [258, p. 146].)

5.8 Facts on Matrix Transformations Involving Two or More Matrices

Fact 5.8.1. Let $A, B \in \mathbb{F}^{n\times m}$. Then, A and B are in the same equivalence class of $\mathbb{F}^{n\times m}$ induced by equivalence if and only if A and B are equivalent to $\left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Now, let $n = m$. Then, A and B are in the same equivalence class of $\mathbb{F}^{n\times n}$ induced by (similarity, unitary similarity) if and only if A and B have the same (Jordan, Schur) form.

Fact 5.8.2. Left equivalence, right equivalence, biequivalence, unitary left equivalence, unitary right equivalence, and unitary biequivalence are equivalence relations on $\mathbb{F}^{n\times m}$. Similarity, congruence, and unitary similarity are equivalence relations on $\mathbb{F}^{n\times n}$.

Fact 5.8.3. Let $A, B \in \mathbb{F}^{n\times n}$ be normal and assume that A and B are similar. Then, A and B are unitarily similar. (Proof: Since A and B are similar, it follows that $\operatorname{mspec}(A) = \operatorname{mspec}(B)$. Since A and B are normal, it follows that they are unitarily similar to the same diagonal matrix.)

Fact 5.8.4. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that either A or B is nonsingular. Then, AB and BA are similar. (Proof: If A is nonsingular, then $AB = A(BA)A^{-1}$.)

Fact 5.8.5. Let $A, B \in \mathbb{R}^{n\times n}$ be projectors. Then, AB and BA are unitarily similar. (Remark: This result is due to Dixmier. See [474].)

Fact 5.8.6. Let $\mathcal{S} \subset \mathbb{F}^{n\times n}$, and assume that $AB = BA$ for all $A, B \in \mathcal{S}$. Then, there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that, for all $A \in \mathcal{S}$, SAS^* is upper triangular. (Proof: See [287, p. 81] and [473].) (Remark: See Fact 8.11.5.)

Fact 5.8.7. Let $\mathcal{S} \subset \mathbb{F}^{n\times n}$, and assume that every matrix $A \in \mathcal{S}$ is normal. Then, $AB = BA$ for all $A, B \in \mathcal{S}$ if and only if there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that, for all $A \in \mathcal{S}$, SAS^* is diagonal. (Remark: See Fact 8.11.2 and [287, pp. 103, 172].)

Fact 5.8.8. Let $\mathcal{S} \subset \mathbb{F}^{n\times n}$, and assume that every matrix $A \in \mathcal{S}$ is diagonalizable over $\mathbb{F}$. Then, $AB = BA$ for all $A, B \in \mathcal{S}$ if and only if there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that, for all $A \in \mathcal{S}$, SAS^{-1} is diagonal. (Proof: See [287, p. 52].)

Fact 5.8.9. Let $A, B \in \mathbb{F}^{n\times m}$. Then, the following statements hold:

i) The matrices A and B are unitarily left equivalent if and only if $A^*A = B^*B$.

ii) The matrices A and B are unitarily right equivalent if and only if $AA^* = BB^*$.

iii) The matrices A and B are unitarily biequivalent if and only if A and B have the same singular values with the same multiplicity.

(Proof: See [293] and [484, pp. 372, 373].) (Remark: In [293] A and B need not be the same size.) (Remark: The singular value decomposition provides a canonical form under unitary biequivalence in analogy with the Smith form under biequivalence.) (Remark: Note that $AA^* = BB^*$ implies $\mathcal{R}(A) = \mathcal{R}(B)$, which implies that right equivalence, which is an alternative proof of the immediate fact that unitary right equivalence implies right equivalence.)

Fact 5.8.10. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $A^*A = B^*B$ if and only if there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that $A = SB$.

ii) $A^*A \le B^*B$ if and only if there exists $S \in \mathbb{F}^{n\times n}$ such that $A = SB$ and $S^*S \le I$.

iii) $A^*B+B^*A=0$ if and only if there exists a unitary matrix $S\in\mathbb{F}^{n\times n}$ such that $(I-S)A=(I+S)B$.

iv) $A^*B+B^*A\geq 0$ if and only if there exists $S\in\mathbb{F}^{n\times n}$ such that $(I-S)A=(I+S)B$ and $S^*S\leq I$.

(Proof: See [476].) (Remark: Statements *iii*) and *iv*) follow from *i*) and *ii*) by replacing A and B with $A-B$ and $A+B$, respectively.)

Fact 5.8.11. Let $A\in\mathbb{F}^{n\times n}$, $B\in\mathbb{F}^{m\times m}$, and $C\in\mathbb{F}^{n\times m}$. Then, there exists $X\in\mathbb{F}^{n\times m}$ satisfying

$$AX+XB+C=0$$

if and only if the matrices

$$\begin{bmatrix} A & 0 \\ 0 & -B \end{bmatrix}, \qquad \begin{bmatrix} A & C \\ 0 & -B \end{bmatrix}$$

are similar. (Proof: See [353, pp. 422–424] or [466, pp. 194–195]. For necessity, the similarity transformation is given by $\left[\begin{smallmatrix} I & X \\ 0 & I \end{smallmatrix}\right]$.) (Remark: $AX+XB+C=0$ is *Sylvester's equation.* See Proposition 7.2.4 and Proposition 11.7.3.) (Remark: This result is due to Roth.)

Fact 5.8.12. Let $A\in\mathbb{F}^{n\times n}$, $B\in\mathbb{F}^{m\times m}$, and $C\in\mathbb{F}^{n\times m}$. Then, there exist $X,Y\in\mathbb{F}^{n\times m}$ satisfying

$$AX+YB+C=0$$

if and only if

$$\operatorname{rank}\begin{bmatrix} A & 0 \\ 0 & -B \end{bmatrix}=\operatorname{rank}\begin{bmatrix} A & C \\ 0 & -B \end{bmatrix}.$$

(Proof: See [466, pp. 194–195].) (Remark: $AX+YB+C=0$ is a generalization of Sylvester's equation. See Fact 5.8.11.) (Remark: This result is due to Roth.)

5.9 Facts on Eigenvalues and Singular Values Involving One Matrix

Fact 5.9.1. Let $A\in\mathbb{F}^{n\times n}$, let $\alpha\in\mathbb{F}$, and assume that $A^2=\alpha A$. Then, $\operatorname{spec}(A)\subseteq\{0,\alpha\}$.

Fact 5.9.2. Let $A\in\mathbb{F}^{n\times n}$ be Hermitian and let $\alpha\in\mathbb{R}$. Then, $A^2=\alpha A$ if and only if $\operatorname{spec}(A)\subseteq\{0,\alpha\}$. (Remark: See Fact 3.4.16.)

Fact 5.9.3. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then,

$$\operatorname{spabs}(A) = \lambda_{\max}(A),$$

$$\operatorname{sprad}(A) = \sigma_{\max}(A) = \max\{|\lambda_{\min}(A)|, \lambda_{\max}(A)\},$$

and

$$\operatorname{spabs}(A) = \lambda_{\max}(A).$$

If, in addition, A is nonnegative semidefinite, then

$$\operatorname{sprad}(A) = \sigma_{\max}(A) = \operatorname{spabs}(A) = \lambda_{\max}(A).$$

Fact 5.9.4. Let $A \in \mathbb{F}^{n\times n}$ be skew Hermitian. Then, the eigenvalues of A are imaginary. (Proof: Let $\lambda \in \operatorname{spec}(A)$. Since $0 \le AA^* = -A^2$, it follows that $-\lambda^2 \ge 0$ and thus $\lambda^2 \le 0$.)

Fact 5.9.5. Let $A \in \mathbb{F}^{n\times n}$, assume that every eigenvalue of A is real, and assume that exactly r eigenvalues of A, including algebraic multiplicity, are nonzero. Then,

$$(\operatorname{tr} A)^2 \le r \operatorname{tr} A^2.$$

Furthermore, equality holds if and only if the nonzero eigenvalues of A are equal. (Remark: For arbitrary $A \in \mathbb{F}^{n\times n}$ with r nonzero eigenvalues, it is not generally true that $|\operatorname{tr} A|^2 \le r|\operatorname{tr} A^2|$. For example, consider $\operatorname{mspec}(A) = \{1, 1, \jmath, -\jmath\}_{\mathrm{m}}$.)

Fact 5.9.6. Let $A \in \mathbb{R}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$\sum_{i=1}^{n} (\operatorname{Re} \lambda_i)(\operatorname{Im} \lambda_i) = 0$$

and

$$\operatorname{tr} A^2 = \sum_{i=1}^{n} (\operatorname{Re} \lambda_i)^2 - \sum_{i=1}^{n} (\operatorname{Im} \lambda_i)^2.$$

Fact 5.9.7. Let $a_1, \ldots, a_n > 0$, and define the symmetric matrix $A \in \mathbb{R}^{n\times n}$ by $A_{(i,j)} \triangleq a_i + a_j$ for all $i, j = 1, \ldots, n$. Then,

$$\operatorname{rank} A = 2,$$

$$\operatorname{spec}(A) = \left\{ \left(\sum_{i=1}^{n} a_i\right) + \sqrt{\sum_{i=1}^{n} a_i^2},\ \left(\sum_{i=1}^{n} a_i\right) - \sqrt{\sum_{i=1}^{n} a_i^2},\ 0 \right\},$$

and

$$\lambda_{\min}(A) < 0 < \operatorname{tr} A = 2\sum_{i=1}^{n} a_i < \lambda_{\max}(A).$$

(Proof: $A = a1_{1\times n} + 1_{n\times 1}a^{\mathrm{T}}$, where $a \triangleq \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^{\mathrm{T}}$.) (Remark: See Fact 8.7.25.)

Fact 5.9.8. Let $x, y \in \mathbb{R}^n$. Then,

$$\operatorname{mspec}\big(xy^{\mathrm{T}} + yx^{\mathrm{T}}\big) = \Big\{x^{\mathrm{T}}y + \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}, x^{\mathrm{T}}y - \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}, 0, \ldots, 0\Big\}_{\mathrm{m}},$$

$$\operatorname{sprad}\big(xy^{\mathrm{T}} + yx^{\mathrm{T}}\big) = \begin{cases} x^{\mathrm{T}}y + \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}, & x^{\mathrm{T}}y \ge 0, \\ \Big|x^{\mathrm{T}}y - \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}\Big|, & x^{\mathrm{T}}y \le 0, \end{cases}$$

and

$$\operatorname{sprad}\big(xy^{\mathrm{T}} + yx^{\mathrm{T}}\big) = x^{\mathrm{T}}y + \sqrt{x^{\mathrm{T}}xy^{\mathrm{T}}y}.$$

(Problem: Extend this result to $\mathbb{C}$ and $xy^{\mathrm{T}} + zw^{\mathrm{T}}$. See Fact 4.9.12.)

Fact 5.9.9. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$\operatorname{mspec}\big(A^{-1}\big) = \big\{\lambda_1^{-1}, \ldots, \lambda_n^{-1}\big\}_{\mathrm{m}},$$

$$\operatorname{mspec}\big[(I + A)^{-1}\big] = \big\{(1+\lambda_1)^{-1}, \ldots, (1+\lambda_n)^{-1}\big\}_{\mathrm{m}},$$

$$\operatorname{mspec}\big[(I + A)^{2}\big] = \big\{(1+\lambda_1)^{2}, \ldots, (1+\lambda_n)^{2}\big\}_{\mathrm{m}},$$

$$\operatorname{mspec}\big[A(I + A)^{-1}\big] = \big\{\lambda_1(1+\lambda_1)^{-1}, \ldots, \lambda_n(1+\lambda_n)^{-1}\big\}_{\mathrm{m}}.$$

Fact 5.9.10. Let $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$. Then,

$$\sigma_{\max}(xy^*) = \sqrt{x^*xy^*y}.$$

If, in addition, $m = n$, then

$$\operatorname{mspec}(xy^*) = \{x^*y, 0, \ldots, 0\}_{\mathrm{m}},$$

$$\operatorname{mspec}(I + xy^*) = \{1 + x^*y, 1, \ldots, 1\}_{\mathrm{m}},$$

$$\operatorname{sprad}(xy^*) = |x^*y|,$$

$$\operatorname{spabs}(xy^*) = \max\{0, \operatorname{Re} x^*y\}.$$

Fact 5.9.11. Let $A \in \mathbb{F}^{n\times n}$ and $\operatorname{rank} A = 1$. Then,

$$\sigma_{\max}(A) = \sigma_{\min}(A) = (\operatorname{tr} AA^*)^{1/2}.$$

Fact 5.9.12. Let $x, y \in \mathbb{F}^n$, and assume that $x^*y \ne 0$. Then,

$$\sigma_{\max}\Big[(x^*y)^{-1}xy^*\Big] \ge 1.$$

Fact 5.9.13. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \cdots, \lambda_n\}_{\mathrm{m}}$, where $\lambda_1, \ldots, \lambda_n$ are ordered such that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. Then, for all $k = 1, \ldots, n$,

$$\prod_{i=1}^{k} |\lambda_i| \le \prod_{i=1}^{k} \sigma_i(A)$$

with equality for $k = n$, that is,

$$|\det A| = \prod_{i=1}^{n} |\lambda_i| = \prod_{i=1}^{n} \sigma_i(A).$$

Hence, for all $k = 1, \ldots, n$,

$$\prod_{i=k}^{n} \sigma_i(A) \le \prod_{i=k}^{n} |\lambda_i|.$$

(Proof: See [93, p. 43], [289, p. 171], or [625, p. 19].) (Remark: This result is due to Weyl.) (Remark: See Fact 8.14.16 and Fact 9.11.16.)

Fact 5.9.14. Let $\beta_0, \ldots, \beta_{n-1} \in \mathbb{R}$, define $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 & \cdots & -\beta_{n-2} & -\beta_{n-1} \end{bmatrix},$$

and define $\alpha \triangleq 1 + \sum_{i=1}^{n-1} \beta_i^2$. Then,

$$\sigma_1(A) = \sqrt{\tfrac{1}{2}\Big(\alpha + \sqrt{\alpha^2 + 4\beta_0^2}\Big)},$$

$$\sigma_2(A) = \cdots = \sigma_{n-1}(A) = 1,$$

$$\sigma_n(A) = \sqrt{\tfrac{1}{2}\Big(\alpha - \sqrt{\alpha^2 + 4\beta_0^2}\Big)}.$$

(Proof: See [326, 334] or [280, p. 523].)

Fact 5.9.15. Let $\beta \in \mathbb{C}$. Then,

$$\sigma_{\max}\left(\begin{bmatrix} 1 & 2\beta \\ 0 & 1 \end{bmatrix}\right) = |\beta| + \sqrt{1 + |\beta|^2}$$

and

$$\sigma_{\min}\left(\begin{bmatrix} 1 & 2\beta \\ 0 & 1 \end{bmatrix}\right) = \sqrt{1+|\beta|^2} - |\beta|.$$

(Proof: See [370].) (Remark: Inequalities involving the singular values of block-triangular matrices are given in [370].)

Fact 5.9.16. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\sigma_{\max}\left(\begin{bmatrix} I & 2A \\ 0 & I \end{bmatrix}\right) = \sigma_{\max}(A) + \sqrt{1+\sigma_{\max}^2(A)}.$$

(Proof: See [280, p. 116].)

Fact 5.9.17. Let $A \in \mathbb{F}^{n\times m}$, and let $r = \operatorname{rank} A$. Then, for all $i = 1,\ldots,r$,

$$\sigma_i(AA^*) = \sigma_i(A^*A) = \sigma_i^2(A).$$

In particular,

$$\sigma_{\max}(AA^*) = \sigma_{\max}^2(A),$$

and, if $n = m$, then

$$\sigma_{\min}(AA^*) = \sigma_{\min}^2(A).$$

Furthermore, for all $i = 1,\ldots,r$,

$$\sigma_i(AA^*A) = \sigma_i^3(A).$$

Fact 5.9.18. Let $A \in \mathbb{F}^{n\times n}$. Then, $\sigma_{\max}(A) \le 1$ if and only if $A^*A \le I$.

Fact 5.9.19. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $i = 1,\ldots,n$,

$$\sigma_i\big(A^{\mathrm{A}}\big) = \prod_{\substack{j=1\\ j\ne n+1-i}}^{n} \sigma_j(A).$$

(Proof: See Fact 4.10.3 and [466, p. 149].)

Fact 5.9.20. Let $A \in \mathbb{F}^{n\times n}$. Then, $\sigma_1(A) = \sigma_n(A)$ if and only if there exist $\lambda \in \mathbb{F}$ and a unitary matrix $B \in \mathbb{F}^{n\times n}$ such that $A = \lambda B$. (Proof: See [466, pp. 149, 165].)

Fact 5.9.21. Let $A \in \mathbb{R}^{n\times n}$, and let $\lambda \in \operatorname{spec}(A)$. Then, the following inequalities hold:

i) $\sigma_{\min}(A) \le |\lambda| \le \sigma_{\max}(A)$.

ii) $\lambda_{\min}\big[\frac{1}{2}\big(A+A^{\mathrm{T}}\big)\big] \le \operatorname{Re}\lambda \le \lambda_{\max}\big[\frac{1}{2}\big(A+A^{\mathrm{T}}\big)\big]$.

iii) $\lambda_{\min}\Big[\frac{1}{2j}\big(A-A^{\mathrm{T}}\big)\Big] \le \operatorname{Im}\lambda \le \lambda_{\max}\Big[\frac{1}{2j}\big(A-A^{\mathrm{T}}\big)\Big]$.

(Remark: *i*) is *Browne's theorem*, *ii*) is *Bendixson's theorem*, and *iii*) is *Hirsch's theorem*. See [395, pp. 140–144]. See Fact 9.10.6.)

Fact 5.9.22. Let $A \in \mathbb{R}^{n\times n}$, where $n \geq 2$, be the tridiagonal matrix

$$A \triangleq \begin{bmatrix} b_1 & c_1 & 0 & \cdots & 0 & 0 \\ a_1 & b_2 & c_2 & \cdots & 0 & 0 \\ 0 & a_2 & b_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & b_{n-1} & c_{n-1} \\ 0 & 0 & 0 & \cdots & a_{n-1} & b_n \end{bmatrix},$$

and assume that $a_ic_i > 0$ for all $i = 1,\ldots,n-1$. Then, A is simple and every eigenvalue of A is real. (Proof: SAS^{-1} is symmetric, where $S \triangleq \operatorname{diag}(d_1,\ldots,d_n)$, $d_1 \triangleq 1$, and $d_{i+1} \triangleq (c_i/a_i)^{1/2}d_i$ for all $i = 1,\ldots,n-1$. For a proof of the fact that A is simple, see [202, p. 198].)

Fact 5.9.23. Let $A \in \mathbb{R}^{n\times n}$ be the tridiagonal matrix

$$A \triangleq \begin{bmatrix} 0 & 1 & 0 & & & & \\ n-1 & 0 & 2 & & & 0 & \\ 0 & n-2 & 0 & \ddots & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & 0 & n-2 & 0 \\ & 0 & & \ddots & 2 & 0 & n-1 \\ & & & & 0 & 1 & 0 \end{bmatrix}.$$

Then,

$$\chi_A(s) = \prod_{i=1}^{n}[s-(n+1-2i)].$$

Hence,

$$\operatorname{spec}(A) = \begin{cases} \{n-1, -(n-1), \ldots, 1, -1\}, & n \text{ even}, \\ \{n-1, -(n-1), \ldots, 2, -2, 0\}, & n \text{ odd}. \end{cases}$$

(Proof: See [537].)

Fact 5.9.24. Let $A \in \mathbb{R}^{n\times n}$, where $n \ge 1$, be the tridiagonal matrix

$$A \triangleq \begin{bmatrix} b & c & 0 & \cdots & 0 & 0 \\ a & b & c & \cdots & 0 & 0 \\ 0 & a & b & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & b & c \\ 0 & 0 & 0 & \cdots & a & b \end{bmatrix},$$

and assume that $ab > 0$. Then,

$$\operatorname{spec}(A) = \{b + \sqrt{ac}\cos[i\pi/(n+1)]\colon\ i = 1, \ldots, n\}.$$

(Remark: See [280, p. 522].)

Fact 5.9.25. Let $a_1, \ldots, a_n \in \mathbb{R}^n$ be linearly independent and, for all $i = 1, \ldots, n$, define

$$A_i \triangleq I - \left(a_i^{\mathrm{T}} a_i\right)^{-1} a_i a_i^{\mathrm{T}}.$$

Then,

$$\sigma_{\max}(A_n A_{n-1} \cdots A_1) < 1.$$

Fact 5.9.26. Let $A \in \mathbb{R}^{n\times n}$, and assume that A has real eigenvalues. Then,

$$\begin{aligned} \lambda_{\min}(A) &\le \tfrac{1}{n}\operatorname{tr} A - \sqrt{\tfrac{1}{n^2-n}\left[\operatorname{tr} A^2 - \tfrac{1}{n}(\operatorname{tr} A)^2\right]} \\ &\le \tfrac{1}{n}\operatorname{tr} A + \sqrt{\tfrac{1}{n^2-n}\left[\operatorname{tr} A^2 - \tfrac{1}{n}(\operatorname{tr} A)^2\right]} \\ &\le \lambda_{\max}(A) \\ &\le \tfrac{1}{n}\operatorname{tr} A + \sqrt{\tfrac{n-1}{n}\left[\operatorname{tr} A^2 - \tfrac{1}{n}(\operatorname{tr} A)^2\right]}. \end{aligned}$$

Furthermore, for all $i = 1, \ldots, n$,

$$\left|\lambda_i(A) - \tfrac{1}{n}\operatorname{tr} A\right| \le \sqrt{\tfrac{n-1}{n}\left[\operatorname{tr} A^2 - \tfrac{1}{n}(\operatorname{tr} A)^2\right]}.$$

(Proof: See [610].)

Fact 5.9.27. Let $A \in \mathbb{R}^{n\times n}$, and assume that $r \triangleq \operatorname{rank} A \ge 2$. If $r \operatorname{tr} A^2 \le (\operatorname{tr} A)^2$, then

$$\operatorname{sprad}(A) \ge \sqrt{\frac{(\operatorname{tr} A)^2 - \operatorname{tr} A^2}{r(r-1)}}.$$

If $(\operatorname{tr} A)^2 \le r \operatorname{tr} A^2$, then

$$\operatorname{sprad}(A) \ge \frac{|\operatorname{tr} A|}{r} + \sqrt{\frac{r \operatorname{tr} A^2 - (\operatorname{tr} A)^2}{r^2(r-1)}}.$$

If $\operatorname{rank} A = 2$, then equality holds in both cases. Finally, if A is skew symmetric, then

$$\operatorname{sprad}(A) \ge \sqrt{\frac{3}{r(r-1)}}\,\|A\|_{\mathrm{F}}.$$

(Proof: See [295].)

Fact 5.9.28. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{spabs}(A) \le \tfrac{1}{2}\lambda_{\max}(A + A^*).$$

Furthermore, equality holds if and only if A is normal. (Proof: See *xii*) and *xiv*) of Fact 9.10.8.)

5.10 Facts on Eigenvalues and Singular Values Involving Two or More Matrices

Fact 5.10.1. Let $A, B \in \mathbb{F}^{n\times n}$ be normal. Then,

$$\min \operatorname{Re} \sum_{i=1}^{n} \lambda_i(A)\lambda_{\sigma(i)}(B) \le \operatorname{Re} \operatorname{tr} AB \le \max \operatorname{Re} \sum_{i=1}^{n} \lambda_i(A)\lambda_{\sigma(i)}(B),$$

where "max" and "min" are taken over all permutations σ of the eigenvalues of B. If, in addition, A and B are Hermitian, then

$$\sum_{i=1}^{n} \lambda_i(A)\lambda_i(B) \le \operatorname{tr} AB \le \sum_{i=1}^{n} \lambda_i(A)\lambda_i(B).$$

(Proof: See [392].) (Remark: See Proposition 8.4.13 and Fact 8.12.14.)

Fact 5.10.2. Let $A, B \in \mathbb{R}^{n\times n}$, assume that B is symmetric, and define $C \triangleq \frac{1}{2}(A + A^{\mathrm{T}})$. Then,

$$\lambda_{\min}(C)\operatorname{tr} B - \lambda_{\min}(B)[n\lambda_{\min}(C) - \operatorname{tr} A]$$
$$\le \operatorname{tr} AB \le \lambda_{\max}(C)\operatorname{tr} B - \lambda_{\max}(B)[n\lambda_{\max}(C) - \operatorname{tr} A].$$

(Proof: See [195].) (Remark: See Fact 5.10.1, Proposition 8.4.13, and Fact 8.12.14. Extensions are given in [451].)

Fact 5.10.3. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$|\operatorname{tr} AB| \le \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B).$$

(Proof: See [466, p. 148].) (Remark: This result is due to Mirsky.)

Fact 5.10.4. Let $A, B \in \mathbb{R}^{n\times n}$, and assume that $AB = BA$. Then,

$$\operatorname{sprad}(AB) \le \operatorname{sprad}(A)\operatorname{sprad}(B),$$

$$\operatorname{sprad}(A+B) \le \operatorname{sprad}(A) + \operatorname{sprad}(B).$$

(Remark: If $AB \ne BA$, then both of these inequalities may be violated. Consider $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right]$.)

Fact 5.10.5. Let $M \in \mathbb{R}^{r\times r}$ be positive definite, let $C, K \in \mathbb{R}^{r\times r}$ be nonnegative semidefinite, and consider the equation

$$M\ddot{q} + C\dot{q} + Kq = 0.$$

Then, $x(t) \triangleq \left[\begin{smallmatrix} q(t) \\ \dot{q}(t) \end{smallmatrix}\right]$, satisfies $\dot{x}(t) = Ax(t)$, where A is the $2r \times 2r$ matrix

$$A \triangleq \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix}.$$

Furthermore,

$$\det A = \frac{\det K}{\det M}$$

and

$$\operatorname{rank} A = r + \operatorname{rank} K.$$

Hence, A is nonsingular if and only if K is positive definite. In this case,

$$A^{-1} = \begin{bmatrix} -K^{-1}C & -K^{-1}M \\ I & 0 \end{bmatrix}.$$

Finally, let $\lambda \in \mathbb{C}$. Then, $\lambda \in \operatorname{spec}(A)$ if and only if $\det(\lambda^2 M + \lambda C + K) = 0$. (Remark: M, C, K are mass, damping, and stiffness matrices. See [85].)

Fact 5.10.6. Let $M, C, K \in \mathbb{R}^{r\times r}$, and assume that M is positive definite and C and K are nonnegative semidefinite. Furthermore, let $\lambda \in \mathbb{C}$ satisfy $\det(\lambda^2 M + \lambda C + K) = 0$. Then, $\operatorname{Re}\lambda \le 0$. Furthermore, if C and K are positive definite, then $\operatorname{Re}\lambda < 0$.

Fact 5.10.7. Let $A, B \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite. Then, every eigenvalue λ of $\left[\begin{smallmatrix} 0 & B \\ -A & 0 \end{smallmatrix}\right]$ satisfies $\operatorname{Re}\lambda = 0$. (Proof: Square this matrix.) (Problem: What happens if A and B have different dimensions?) In addition, let $C \in \mathbb{R}^{n\times n}$ be (nonnegative semidefinite, positive definite).

Then, every eigenvalue of $\begin{bmatrix} 0 & A \\ -B & -C \end{bmatrix}$ satisfies $(\operatorname{Re}\lambda \le 0, \operatorname{Re}\lambda < 0)$. (Problem: Consider also $\begin{bmatrix} -C & A \\ -B & -C \end{bmatrix}$ and $\begin{bmatrix} -C & A \\ -A & -C \end{bmatrix}$.)

5.11 Facts on Matrix Eigenstructure

Fact 5.11.1. Let $A \in \mathbb{F}^{n\times n}$. Then, $\mathcal{R}(A) = \mathcal{R}(A^2)$ if and only if $\operatorname{ind} A \le 1$.

Fact 5.11.2. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is diagonalizable. Then, are A^{A}, A^*, $\overline{A}$, and A^{T} are diagonalizable. If, in addition, A is nonsingular, then A^{-1} is diagonalizable. (Proof: See Fact 2.13.9 and Fact 3.4.6.)

Fact 5.11.3. Let $A \in \mathbb{F}^{n\times n}$ be diagonalizable over $\mathbb{F}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, and let $B \triangleq \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. If, for all $i = 1, \ldots, n$, $x_i \in \mathbb{F}^n$ is an eigenvector of A associated with λ_i, then $A = SBS^{-1}$, where $S \triangleq \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$. Conversely, if $S \in \mathbb{F}^{n\times n}$ is nonsingular and $A = SBS^{-1}$, then, for all $i = 1, \ldots, n$, $\operatorname{col}_i(S)$ is an associated eigenvector.

Fact 5.11.4. Let $A \in \mathbb{F}^{n\times n}$, let $S \in \mathbb{F}^{n\times n}$, assume that S is nonsingular, let $\lambda \in \mathbb{C}$, and assume that $\operatorname{row}_1(S^{-1}AS) = \lambda e_1^{\mathrm{T}}$. Then, $\lambda \in \operatorname{spec}(A)$, and $\operatorname{col}_1(S)$ is an associated eigenvector.

Fact 5.11.5. Let $A \in \mathbb{F}^{n\times n}$. Then, A is cyclic if and only if there exists $x \in \mathbb{F}^n$ such that $\begin{bmatrix} x & Ax & \cdots & A^{n-1}x \end{bmatrix}$ is nonsingular.

Fact 5.11.6. Let $A \in \mathbb{R}^{n\times n}$. Then, A is cyclic and diagonalizable over $\mathbb{R}$ if and only if A is simple.

Fact 5.11.7. Let $A = \operatorname{revdiag}(a_1, \ldots, a_n) \in \mathbb{R}^{n\times n}$. Then, A is semisimple if and only if, for all $i = 1, \ldots, n$, a_i and a_{n+1-i} are either both zero or both nonzero. (Proof: See [258, p. 116], [328], or [466, pp. 68, 86].)

Fact 5.11.8. Let $A \in \mathbb{F}^{n\times n}$. The A has at least m real eigenvalues and m associated linearly independent eigenvectors if and only if there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $AS = SA^*$. (Proof: See [466, pp. 68, 86].) (Remark: See Proposition 5.5.18.) (Remark: This result is due to Drazin and Haynsworth.)

Fact 5.11.9. Let $A \in \mathbb{F}^{n\times n}$ be normal and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then, there exist $x_1, \ldots, x_n \in \mathbb{C}^n$ such that $x_i^* x_j = \delta_{ij}$ for all $i, j = 1, \ldots, n$

and

$$A = \sum_{i=1}^{n} \lambda_i x_i x_i^*.$$

Fact 5.11.10. Let $A \in \mathbb{F}^{n\times n}$, assume that A is normal, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then, the singular values of A are $|\lambda_1|, \ldots, |\lambda_n|$.

Fact 5.11.11. Let $A \in \mathbb{F}^{n\times n}$ be idempotent. Then, A is diagonalizable over $\mathbb{R}$, $\operatorname{spec}(A) \subset \{0, 1\}$, and $\operatorname{tr} A = \operatorname{rank} A$.

Fact 5.11.12. Let $A \in \mathbb{F}^{n\times n}$ be either involutory or skew involutory. Then, A is semisimple.

Fact 5.11.13. Let $A \in \mathbb{R}^{n\times n}$ be involutory. Then, A is diagonalizable over $\mathbb{R}$.

Fact 5.11.14. Let $A \in \mathbb{F}^{n\times n}$ be semisimple and assume that $A^3 = A^2$. Then, A is idempotent.

Fact 5.11.15. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{spec}(A) = \{0, \lambda_1, \ldots, \lambda_r\}$. Then, A is group invertible if and only if $\operatorname{rank} A = \sum_{i=1}^{r} \operatorname{am}_A(\lambda_i)$.

Fact 5.11.16. Let $A \in \mathbb{F}^{n\times n}$. Then, every matrix $B \in \mathbb{F}^{n\times n}$ satisfying $AB = BA$ is a polynomial in A if and only if A is cyclic.

Fact 5.11.17. Let $A, B \in \mathbb{C}^{n\times n}$, and assume that $AB = BA$. Then, there exists a nonzero vector $x \in \mathbb{C}^n$ that is an eigenvector of both A and B. (Proof: See [287, p. 51].)

Fact 5.11.18. Let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) If A and B are Hermitian, then AB is Hermitian if and only if $AB = BA$.

ii) If A is normal and $AB = BA$, then $A^*B = BA^*$.

iii) If B is Hermitian and $AB = BA$, then $A^*B = BA^*$.

iv) If A and B are normal and $AB = BA$, then AB is normal.

v) If A, B, and AB are normal, then BA is normal.

vi) If A and B are normal and either A or B has the property that distinct eigenvalues have unequal absolute values, then AB is normal if and only if $AB = BA$.

vii) If A and B are normal, either A or B is nonnegative semidefinite, and AB is normal, then AB is normal if and only if $AB = BA$.

(Proof: See [154, 597], [259, p. 157], [262, p. 157], and [466, p. 102].)

Fact 5.11.19. Let $A, B, C \in \mathbb{F}^{n\times n}$, and assume that A and B are normal and $AC = CB$. Then, $A^*C = CB^*$. (Proof: Consider $\left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & C \\ 0 & 0 \end{smallmatrix}\right]$ in *ii*) of Fact 5.11.18. See [259, p. 104] or [262, p. 321].) (Remark: This result is the *Putnam-Fuglede theorem.*)

Fact 5.11.20. Let $A, B \in \mathbb{R}^{n\times n}$ be skew symmetric. Then, there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S\begin{bmatrix} 0_{(n-l)\times(n-l)} & A_{12} \\ -A_{12}^{\mathrm{T}} & A_{22} \end{bmatrix} S^{\mathrm{T}}$$

and

$$B = S\begin{bmatrix} B_{11} & B_{12} \\ -B_{12}^{\mathrm{T}} & 0_{l\times l} \end{bmatrix} S^{\mathrm{T}},$$

where $l \triangleq \lfloor n/2 \rfloor$. Consequently,

$$\operatorname{mspec}(AB) = \operatorname{mspec}\big(-A_{12}B_{12}^{\mathrm{T}}\big) \cup \operatorname{mspec}\big(-A_{12}^{\mathrm{T}}B_{12}\big),$$

and thus every nonzero eigenvalue of AB has even algebraic multiplicity. (Proof: See [13].)

Fact 5.11.21. Let $A, B \in \mathbb{R}^{n\times n}$ be skew symmetric. If n is even, then there exists a monic polynomial p of degree $n/2$ such that $\chi_{AB}(s) = p^2(s)$ and $p(AB) = 0$. If n is odd, then there exists a monic polynomial $p(s)$ of degree $(n-1)/2$ such that $\chi_{AB}(s) = sp^2(s)$ and $ABp(AB) = 0$. Consequently, if n is (even, odd), then χ_{AB} is (even, odd) and (every, every nonzero) eigenvalue of AB has even algebraic multiplicity and geometric multiplicity of at least 2. (Proof: See [183, 241].)

Fact 5.11.22. Let $A, B \in \mathbb{F}^{n\times n}$ be projectors. Then, $\operatorname{spec}(AB) \subset [0, 1]$ and $\operatorname{spec}(A - B) \subset [-1, 1]$. (Proof: See [19] or [466, p. 147].) (Remark: The first result is due to Afriat.)

Fact 5.11.23. Let $q(t)$ denote the displacement of a mass $m > 0$ connected to a spring $k \ge 0$ and dashpot $c \ge 0$ and subject to a force $f(t)$. Then, $q(t)$ satisfies

$$m\ddot{q}(t) + c\dot{q}(t) + kq(t) = f(t)$$

or

$$\ddot{q}(t) + \frac{c}{m}\dot{q}(t) + \frac{k}{m}q(t) = \frac{1}{m}f(t).$$

Now, define the *natural frequency* $\omega_{\mathrm{n}} \triangleq \sqrt{k/m}$ and, if $k > 0$, the *damping* *ratio* $\zeta \triangleq c/2\sqrt{km}$ to obtain

$$\ddot{q}(t) + 2\zeta\omega_{\rm n}\dot{q}(t) + \omega_{\rm n}^2 q(t) = \frac{1}{m}f(t).$$

If $k = 0$, then set $\omega_{\rm n} = 0$, and $\zeta\omega_{\rm n} = c/2m$. Next, define $x_1(t) \triangleq q(t)$ and $x_2(t) \triangleq \dot{q}(t)$ so that this equation can be written as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_{\rm n}^2 & -2\zeta\omega_{\rm n} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} f(t).$$

The eigenvalues of the companion matrix $A_{\rm c} \triangleq \left[\begin{smallmatrix} 0 & 1 \\ -\omega_{\rm n}^2 & -2\zeta\omega_{\rm n} \end{smallmatrix}\right]$ are given by

$$\operatorname{mspec}(A_{\rm c}) = \begin{cases} \{-\zeta\omega_{\rm n} - \jmath\omega_{\rm d}, -\zeta\omega_n + \jmath\omega_{\rm d}\}_{\rm m}, & 0 \le \zeta \le 1, \\ \left\{(-\zeta - \sqrt{\zeta^2 - 1})\omega_{\rm n}, (-\zeta + \sqrt{\zeta^2 - 1})\omega_{\rm n}\right\}, & \zeta > 1, \end{cases}$$

where $\omega_{\rm d} \triangleq \omega_{\rm n}\sqrt{1 - \zeta^2}$ is the *damped natural frequency.* The matrix $A_{\rm c}$ has repeated eigenvalues in exactly two cases, namely,

$$\operatorname{mspec}(A_{\rm c}) = \begin{cases} \{0, 0\}_{\rm m}, & \omega_{\rm n} = 0, \\ \{-\omega_{\rm n}, -\omega_{\rm n}\}_{\rm m}, & \zeta = 1. \end{cases}$$

In both of these cases the matrix $A_{\rm c}$ is defective. In the case $\omega_{\rm n} = 0$, the matrix $A_{\rm c}$ is also in Jordan form, while in the case $\zeta = 1$, it follows that $A_{\rm c} = SA_{\rm J}S^{-1}$, where $S \triangleq \left[\begin{smallmatrix} -1 & 0 \\ \omega_{\rm n} & -1 \end{smallmatrix}\right]$ and $A_{\rm J}$ is the Jordan form matrix $A_{\rm J} \triangleq \left[\begin{smallmatrix} -\omega_{\rm n} & 1 \\ 0 & -\omega_{\rm n} \end{smallmatrix}\right]$. If $A_{\rm c}$ is not defective, that is, if $\omega_{\rm n} \ne 0$ and $\zeta \ne 1$, then the Jordan form $A_{\rm J}$ of $A_{\rm c}$ is given by

$$A_{\rm J} \triangleq \begin{cases} \begin{bmatrix} -\zeta\omega_{\rm n} + \jmath\omega_{\rm d} & 0 \\ 0 & -\zeta\omega_{\rm n} - \jmath\omega_{\rm d} \end{bmatrix}, & 0 \le \zeta < 1,\ \omega_{\rm n} \ne 0, \\ \begin{bmatrix} \left(-\zeta - \sqrt{\zeta^2 - 1}\right)\omega_{\rm n} & 0 \\ 0 & \left(-\zeta + \sqrt{\zeta^2 - 1}\right)\omega_{\rm n} \end{bmatrix}, & \zeta > 1,\ \omega_{\rm n} \ne 0. \end{cases}$$

In the case $0 \le \zeta < 1$ and $\omega_{\rm n} \ne 0$, define the real normal form

$$A_{\rm n} \triangleq \begin{bmatrix} -\zeta\omega_{\rm n} & \omega_{\rm d} \\ -\omega_{\rm d} & -\zeta\omega_{\rm n} \end{bmatrix}.$$

The matrices $A_{\rm c}, A_{\rm J}$, and $A_{\rm n}$ are related by the similarity transformations

$$A_{\rm c} = S_1 A_{\rm J} S_1^{-1} = S_2 A_{\rm n} S_2^{-1}, \quad A_{\rm J} = S_3 A_{\rm n} S_3^{-1},$$

where

$$S_1 \triangleq \begin{bmatrix} 1 & 1 \\ -\zeta\omega_{\rm n} + \jmath\omega_{\rm d} & -\zeta\omega_{\rm n} - \jmath\omega_{\rm d} \end{bmatrix}, \quad S_1^{-1} = \frac{\jmath}{2\omega_{\rm d}}\begin{bmatrix} -\zeta\omega_{\rm n} - \jmath\omega_{\rm d} & -1 \\ \zeta\omega_{\rm n} - \jmath\omega_{\rm d} & 1 \end{bmatrix},$$

$$S_2 \triangleq \frac{1}{\omega_{\rm d}}\begin{bmatrix} 1 & 0 \\ -\zeta\omega_{\rm n} & \omega_{\rm d} \end{bmatrix}, \qquad S_2^{-1} = \begin{bmatrix} \omega_{\rm d} & 0 \\ \zeta\omega_{\rm n} & 1 \end{bmatrix},$$

$$S_3 \triangleq \frac{1}{2\omega_{\rm d}}\begin{bmatrix} 1 & -\jmath \\ 1 & \jmath \end{bmatrix}, \qquad S_3^{-1} = \omega_{\rm d}\begin{bmatrix} 1 & 1 \\ \jmath & -\jmath \end{bmatrix}.$$

In the case $\zeta > 1$ and $\omega_{\rm n} \neq 0$, the matrices $A_{\rm c}$ and $A_{\rm J}$ are related by

$$A_{\rm c} = S_4 A_{\rm J} S_4^{-1},$$

where

$$S_4 \triangleq \begin{bmatrix} 1 & 1 \\ -\zeta\omega_{\rm n} + \jmath\omega_{\rm d} & -\zeta\omega_{\rm n} - \jmath\omega_{\rm d} \end{bmatrix}, \quad S_4^{-1} = \frac{\jmath}{2\omega_{\rm d}}\begin{bmatrix} -\zeta\omega_{\rm n} - \jmath\omega_{\rm d} & -1 \\ \zeta\omega_{\rm n} - \jmath\omega_{\rm d} & 1 \end{bmatrix}.$$

Finally, define the energy coordinates matrix

$$A_{\rm e} \triangleq \begin{bmatrix} 0 & \omega_{\rm n} \\ -\omega_{\rm n} & -2\zeta\omega_{\rm n} \end{bmatrix}.$$

Then, $A_{\rm e} = S_5 A_{\rm c} S_5^{-1}$, where

$$S_5 \triangleq \sqrt{\tfrac{m}{2}}\begin{bmatrix} 1/\omega_{\rm n} & 0 \\ 0 & 1 \end{bmatrix}.$$

5.12 Facts on Companion, Vandermonde, and Circulant Matrices

Fact 5.12.1. Let $p \in \mathbb{F}[s]$, where $p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_0$, and define $C_{\rm b}(p), C_{\rm r}(p), C_{\rm t}(p), C_{\rm l}(p) \in \mathbb{F}^{n\times n}$ by

$$C_{\rm b}(p) \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 & \cdots & -\beta_{n-2} & -\beta_{n-1} \end{bmatrix},$$

$$C_{\mathrm{r}}(p) \triangleq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -\beta_0 \\ 1 & 0 & 0 & \cdots & 0 & -\beta_1 \\ 0 & 1 & 0 & \cdots & 0 & -\beta_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & -\beta_{n-2} \\ 0 & 0 & 0 & \cdots & 1 & -\beta_{n-1} \end{bmatrix},$$

$$C_{\mathrm{t}}(p) \triangleq \begin{bmatrix} -\beta_{n-1} & -\beta_{n-2} & \cdots & -\beta_2 & -\beta_1 & -\beta_0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \ddots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

$$C_{\mathrm{l}}(p) \triangleq \begin{bmatrix} -\beta_{n-1} & 1 & \cdots & 0 & 0 & 0 \\ -\beta_{n-2} & 0 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -\beta_2 & 0 & \cdots & 0 & 1 & 0 \\ -\beta_1 & 0 & \cdots & 0 & 0 & 1 \\ -\beta_0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}.$$

Then,

$$C_{\mathrm{r}}(p) = C_{\mathrm{b}}^{\mathrm{T}}(p), \quad C_{\mathrm{l}}(p) = C_{\mathrm{t}}^{\mathrm{T}}(p),$$
$$C_{\mathrm{t}}(p) = \hat{I}C_{\mathrm{b}}(p)\hat{I}, \quad C_{\mathrm{l}}(p) = \hat{I}C_{\mathrm{r}}(p)\hat{I},$$
$$C_{\mathrm{l}}(p) = C_{\mathrm{b}}^{\hat{\mathrm{T}}}(p), \quad C_{\mathrm{t}}(p) = C_{\mathrm{r}}^{\hat{\mathrm{T}}}(p),$$

and

$$\chi_{C_{\mathrm{b}}(p)} = \chi_{C_{\mathrm{r}}(p)} = \chi_{C_{\mathrm{t}}(p)} = \chi_{C_{\mathrm{l}}(p)} = p.$$

Furthermore,

$$C_{\mathrm{r}}(p) = SC_{\mathrm{b}}(p)S^{-1}$$

and

$$C_{\mathrm{t}}(p) = \hat{S}C_{\mathrm{l}}(p)\hat{S}^{-1}$$

where $S, \hat{S} \in \mathbb{F}^{n\times n}$ are the Hankel matrices

$$S \triangleq \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{n-1} & 1 \\ \beta_2 & \beta_3 & .\cdot^{\cdot} & 1 & 0 \\ \vdots & .\cdot^{\cdot} & .\cdot^{\cdot} & .\cdot^{\cdot} & \vdots \\ \beta_{n-1} & 1 & .\cdot^{\cdot} & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

and

$$\hat{S} \triangleq \hat{I}S\hat{I} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & .\cdot^{\cdot} & 1 & \beta_{n-1} \\ \vdots & .\cdot^{\cdot} & .\cdot^{\cdot} & .\cdot^{\cdot} & \vdots \\ 0 & 1 & .\cdot^{\cdot} & \beta_3 & \beta_2 \\ 1 & \beta_{n-1} & \cdots & \beta_2 & \beta_1 \end{bmatrix}.$$

(Remark: $(C_{\rm b}(p), C_{\rm r}(p), C_{\rm t}(p), C_{\rm l}(p))$ are the (*bottom, right, top, left*) companion matrices. See [64, p. 282] and [321, p. 659].) (Remark: $S = B(p,1)$, where $B(p,1)$ is a Bezout matrix. See Fact 4.8.6.)

Fact 5.12.2. Let $p \in \mathbb{F}[s]$, where $p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_0$, assume that $\beta_0 \neq 0$, and let

$$C_{\rm b}(p) \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 & \cdots & -\beta_{n-2} & -\beta_{n-1} \end{bmatrix}.$$

Then,

$$C_{\rm b}^{-1}(p) = C_{\rm t}(\hat{p}) = \begin{bmatrix} -\beta_1/\beta_0 & \cdots & -\beta_{n-2}/\beta_0 & -\beta_{n-1}/\beta_0 & -1/\beta_0 \\ 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

where $\hat{p}(s) \triangleq \beta_0^{-1}s^n p(1/s)$. (Remark: See Fact 4.9.6.)

Fact 5.12.3. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, and define the *Vandermonde matrix*

$V(\lambda_1,\dots,\lambda_n)\in\mathbb{F}^{n\times n}$ by

$$V(\lambda_1,\dots,\lambda_n)\triangleq\begin{bmatrix}1 & 1 & \cdots & 1\\ \lambda_1 & \lambda_2 & \cdots & \lambda_n\\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2\\ \lambda_1^3 & \lambda_2^3 & \cdots & \lambda_n^3\\ \vdots & \vdots & \cdots & \vdots\\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1}\end{bmatrix}.$$

Then,

$$\det V(\lambda_1,\dots,\lambda_n)=\prod_{i>j}(\lambda_i-\lambda_j).$$

Thus, $V(\lambda_1,\dots,\lambda_n)$ is nonsingular if and only if $\lambda_1,\dots,\lambda_n$ are distinct. (Remark: This result yields Proposition 4.5.3. Let $x_1,\dots,x_k$ be eigenvectors of $V(\lambda_1,\dots,\lambda_n)$ associated with distinct eigenvalues $\lambda_1,\dots,\lambda_k$ of $V(\lambda_1,\dots,\lambda_n)$. Assume $\alpha_1x_1+\cdots+\alpha_kx_k=0$ so that $V^i(\lambda_1,\dots,\lambda_n)(\alpha_1x_1+\cdots+\alpha_kx_k)=\alpha_1\lambda_1^ix_i+\cdots+\alpha_k\lambda_k^ix_k=0$ for all $i=0,1,\dots,k-1$. Let $X\triangleq\begin{bmatrix}x_1 & \cdots & x_k\end{bmatrix}\in\mathbb{F}^{n\times k}$ and $D\triangleq\operatorname{diag}(\alpha_1,\dots,\alpha_k)$. Then, $XDV^{\mathrm{T}}(\lambda_1,\dots,\lambda_k)=0$, which implies that $XD=0$. Hence, $\alpha_ix_i=0$ for all $i=1,\dots,k$, and thus $\alpha_1=\alpha_2=\cdots=\alpha_k=0$.)

Fact 5.12.4. Let $\lambda_1,\dots,\lambda_n\in\mathbb{F}$ and, for $i=1,\dots,n$, define

$$p_i(s)\triangleq\prod_{\substack{j=1\\ j\neq i}}^{n}(s-\lambda_j).$$

Furthermore, define $A\in\mathbb{F}^{n\times n}$ by

$$A\triangleq\begin{bmatrix}p_1(0) & \frac{1}{1!}p_1'(0) & \cdots & \frac{1}{(n-1)!}p_1^{(n-1)}(0)\\ \vdots & \cdots & \cdots & \vdots\\ p_n(0) & \frac{1}{1!}p_n'(0) & \cdots & \frac{1}{(n-1)!}p_n^{(n-1)}(0)\end{bmatrix}.$$

Then,

$$\operatorname{diag}[p_1(s),\dots,p_n(s)]=AV(s,\dots,s).$$

(Proof: See [202, p. 159].)

Fact 5.12.5. Let $p\in\mathbb{F}[s]$, where $p(s)=s^n+\beta_{n-1}s^{n-1}+\cdots+\beta_1s+\beta_0$, and assume that p has distinct roots $\lambda_1,\dots,\lambda_n\in\mathbb{C}$. Then,

$$C(p)=V^{-1}(\lambda_1,\dots,\lambda_n)\operatorname{diag}(\lambda_1,\dots,\lambda_n)V(\lambda_1,\dots,\lambda_n).$$

Fact 5.12.6. Let $A\in\mathbb{F}^{n\times n}$. Then, A is cyclic if and only if A is similar to a companion matrix. (Proof: The result follows from Corollary 5.3.4. Alternatively, let $\operatorname{spec}(A) = \{\lambda_1, \ldots, \lambda_r\}$ and $A = SBS^{-1}$, where $S \in \mathbb{C}^{n \times n}$ is nonsingular and $B = \operatorname{diag}(B_1, \ldots, B_r)$ is the Jordan form of A, where, for all $i = 1, \ldots, r$, $B_i \in \mathbb{C}^{n_i \times n_i}$ and $\lambda_i, \ldots, \lambda_i$ are the diagonal entries of B_i. Now, define $R \in \mathbb{C}^{n \times n}$ by $R \triangleq \begin{bmatrix} R_1 & \cdots & R_r \end{bmatrix} \in \mathbb{C}^{n \times n}$, where, for all $i = 1, \ldots, r$, $R_i \in \mathbb{C}^{n \times n_i}$ is the matrix

$$R_i \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \therefore & \vdots \\ \lambda_i^{n-2} & \binom{n-2}{1}\lambda_i^{n-3} & \cdots & \binom{n-2}{n_i-1}\lambda_i^{n-n_i-1} \\ \lambda_i^{n-1} & \binom{n-1}{1}\lambda_i^{n-2} & \cdots & \binom{n-1}{n_i-1}\lambda_i^{n-n_i} \end{bmatrix}.$$

Then, since $\lambda_1, \ldots, \lambda_r$ are distinct, it follows that R is nonsingular. Furthermore, $C = RBR^{-1}$ is in companion form and thus $A = SR^{-1}CRS$. If $n_i = 1$ for all $i = 1, \ldots, r$, then R is a Vandermonde matrix. See Fact 5.12.3 and Fact 5.12.5.)

Fact 5.12.7. Let $a_0, \ldots, a_{n-1} \in \mathbb{F}$, and define $\operatorname{circ}(a_0, \ldots, a_{n-1}) \in \mathbb{F}^{n \times n}$ by

$$\operatorname{circ}(a_0, \ldots, a_{n-1}) \triangleq \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-3} & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \ddots & a_{n-4} & a_{n-3} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a_2 & a_3 & a_4 & \ddots & a_0 & a_1 \\ a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_0 \end{bmatrix}.$$

A matrix of this form is *circulant*. Furthermore, define the *primary circulant*

$$P \triangleq \operatorname{circ}(0, 1, 0, \ldots, 0) \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Finally, define $p(s) \triangleq a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$. Then, the following statements hold:

i) $\operatorname{circ}(a_0, \ldots, a_{n-1}) = p(P)$.

ii) If $A, B \in \mathbb{F}^{n \times n}$ are circulant, then A and B commute and AB is circulant.

iii) If A is circulant, then A^* is circulant.

iv) If A is circulant and $k \geq 0$, then A^k is circulant.

v) If A is nonsingular and circulant, then A^{-1} is circulant.

vi) $A \in \mathbb{F}^{n\times n}$ is circulant if and only if $A = PAP^{\mathrm{T}}$.

vii) P is an orthogonal matrix, and $P^n = I_n$.

viii) $P = C(p)$, where $p \in \mathbb{F}[s]$ is defined by $p(s) \triangleq s^n - 1$.

ix) If $A \in \mathbb{F}^{n\times n}$ is circulant, then A is reverse symmetric, Toeplitz, and normal.

x) $A \in \mathbb{F}^{n\times n}$ is normal if and only if A is unitarily similar to a normal matrix.

Next, let $\theta \triangleq e^{2\pi \jmath/n}$, and define the *Fourier matrix* $S \in \mathbb{C}^{n\times n}$ by

$$S \triangleq n^{-1/2}V\big(1, \theta, \ldots, \theta^{n-1}\big) = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \theta & \theta^2 & \cdots & \theta^{n-1} \\ 1 & \theta^2 & \theta^4 & \cdots & \theta^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \theta^{n-1} & \theta^{n-2} & \cdots & \theta \end{bmatrix}.$$

Then, the following statements hold:

i) S is symmetric and unitary.

ii) $S^4 = I_n$.

iii) $\operatorname{spec}(S) = \{1, -1, \jmath, -\jmath\}$.

iv) $\operatorname{Re} S$ and $\operatorname{Im} S$ are symmetric, commute, and satisfy $(\operatorname{Re} S)^2 + (\operatorname{Im} S)^2 = I_n$.

v) $SPS^{-1} = \operatorname{diag}\big(1, \theta, \ldots, \theta^{n-1}\big)$.

vi) $S\operatorname{circ}(a_0, \ldots, a_{n-1})S^{-1} = \operatorname{diag}\big[p(1), p(\theta), \ldots, p\big(\theta^{n-1}\big)\big]$.

vii) $\operatorname{mspec}[\operatorname{circ}(a_0, \ldots, a_{n-1})] = \big\{p(1), p(\theta), p\big(\theta^2\big), \ldots, p\big(\theta^{n-1}\big)\big\}_{\mathrm{m}}$.

viii) $\operatorname{spec}(P) = \big\{1, \theta, \theta^2, \ldots, \theta^{n-1}\big\}$.

(Proof: See [7, pp. 81–98], [163, p. 81], and [629, pp. 106–110].) (Remark: Circulant matrices play an important role in digital signal processing, specifically, in the efficient implementation of the *fast Fourier transform.* See [415, pp. 356–380] and [569, pp. 206, 207].) (Remark: If a real Toeplitz matrix is normal, then it must be either symmetric, skew-symmetric, circulant, or skew circulant. See [34] and the references therein.)

5.13 Facts on Matrix Factorizations

Fact 5.13.1. Let $A \in \mathbb{F}^{n\times n}$. Then, A is normal if and only if there exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that $A^* = AS$. (Proof: See [466, pp. 102, 113].)

Fact 5.13.2. Let $A \in \mathbb{F}^{m\times m}$ and $B \in \mathbb{F}^{n\times n}$. Then, there exist $C \in \mathbb{F}^{m\times n}$ and $D \in \mathbb{F}^{n\times m}$ such that $A = CD$ and $B = DC$ if and only if the following statements hold:

i) The Jordan blocks associated with nonzero eigenvalues are identical in A and B.

ii) Let $n_1 \geq n_2 \geq \cdots \geq n_r$ denote the sizes of the Jordan blocks of A associated with $0 \in \operatorname{spec}(A)$, and let $m_1 \geq m_2 \geq \cdots \geq m_r$ denote the sizes of the Jordan blocks of B associated with $0 \in \operatorname{spec}(B)$, where $n_i = 0$ or $m_i = 0$ as needed. Then, $|n_i - m_i| \leq 1$ for all $i = 1, \ldots, r$.

(Proof: See [315].) (Remark: See Fact 5.13.3.)

Fact 5.13.3. Let $A, B \in \mathbb{F}^{n\times n}$ be nonsingular. Then, A and B are similar if and only if there exist nonsingular matrices $C, D \in \mathbb{F}^{n\times n}$ such that $A = CD$ and $B = DC$. (Proof: Sufficiency follows from Fact 5.8.4. Necessity is a special case of Fact 5.13.2.)

Fact 5.13.4. Let $A, B \in \mathbb{F}^{n\times n}$ be nonsingular. Then, $\det A = \det B$ if and only if there exist nonsingular matrices $C, D, E \in \mathbb{R}^{n\times n}$ such that $A = CDE$ and $B = EDC$. (Remark: This result is due to Shoda and Taussky-Todd. See [110].)

Fact 5.13.5. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist $B, C \in \mathbb{F}^{n\times n}$ such that B is unitary, C is upper triangular, and $A = BC$. If, in addition, A is nonsingular, then there exist unique $B, C \in \mathbb{F}^{n\times n}$ such that B is unitary, C is upper triangular with positive diagonal entries, and $A = BC$. (Proof: See [287, p. 112] or [484, p. 362].) (Remark: This result is the *QR decomposition*. The orthogonal matrix B is constructed as a product of elementary reflectors.)

Fact 5.13.6. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = m$. Then, there a unique matrix $B \in \mathbb{F}^{n\times m}$ and a matrix $C \in \mathbb{F}^{m\times m}$ such that $B^*B = I_m$, C is upper triangular with positive diagonal entries, and $A = BC$. (Proof: See [287, p. 15] or [484, p. 206].) (Remark: $C \in \mathrm{UT}_+(n)$. See Fact 3.10.3.) (Remark: This result is *Gram-Schmidt orthonormalization*.)

Fact 5.13.7. Let $A \in \mathbb{F}^{n\times n}$, let $r \triangleq \operatorname{rank} A$, and assume that the first r leading principal subdeterminants of A are nonzero. Then, there exist

$B, C \in \mathbb{F}^{n\times n}$ such that B is lower triangular, C is upper triangular, and $A = BC$. Either B or C can be chosen to be nonsingular. Furthermore, both B and C are nonsingular if and only if A is nonsingular. (Proof: See [287, p. 160].) (Remark: This result is the *LU decomposition.*)

Fact 5.13.8. Let $A \in \mathbb{F}^{n\times n}$, and let $r \triangleq \operatorname{rank} A$. Then, A is range Hermitian if and only if there exist a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ and a nonsingular matrix $B \in \mathbb{F}^{r\times r}$ such that

$$A = S\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}S^*.$$

(Remark: S need not be unitary for sufficiency. See Corollary 5.4.4.) (Proof: Use the QR decomposition Fact 5.13.5 to let $S \triangleq \hat{S}R$, where $\hat{S}$ is unitary and R is upper triangular.)

Fact 5.13.9. Let $A \in \mathbb{F}^{n\times n}$. Then, A is nonsingular if and only if A is the product of elementary matrices. (Problem: How many factors are needed?)

Fact 5.13.10. Let $A \in \mathbb{F}^{n\times n}$ be a projector, and let $r \triangleq \operatorname{rank} A$. Then, there exist nonzero $x_1, \ldots, x_{n-r} \in \mathbb{F}^n$ such that $x_i^*x_j = 0$ for all $i \neq j$ and such that

$$A = \prod_{i=1}^{n-r}\left[I - (x_i^*x_i)^{-1}x_ix_i^*\right].$$

(Remark: Every projector is the product of mutually orthogonal elementary projectors.) (Proof: A is unitarily similar to $\operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)$, which can be written as the product of elementary projectors.)

Fact 5.13.11. Let $A \in \mathbb{F}^{n\times n}$. Then, A is a reflector if and only if there exist $m \leq n$ nonzero vectors $x_1, \ldots, x_m \in \mathbb{F}^n$ such that $x_i^*x_j = 0$ for all $i \neq j$ and such that

$$A = \prod_{i=1}^{m}\left[I - 2(x_i^*x_i)^{-1}x_ix_i^*\right].$$

In this case, m is the algebraic multiplicity of $-1 \in \operatorname{spec}(A)$. (Remark: Every reflector is the product of mutually orthogonal elementary reflectors.) (Proof: A is unitarily similar to $\operatorname{diag}(\pm1, \ldots, \pm1)$, which can be written as the product of elementary reflectors.)

Fact 5.13.12. Let $A \in \mathbb{F}^{n\times n}$. Then, A is unitary if and only if there exist nonzero vectors $x_1, \ldots, x_m \in \mathbb{F}^n$ such that

$$A = \prod_{i=1}^{m}\left[I - 2(x_i^*x_i)^{-1}x_ix_i^*\right].$$

(Remark: Every unitary matrix is the product of elementary reflectors. This factorization is a result of Cartan and Dieudonne. See [45, p. 24] and [498, 564]. The minimal number of factors is unsettled; see Fact 3.7.3. See Fact 3.6.17.)

Fact 5.13.13. Let $A \in \mathbb{R}^{n\times n}$, where $n \geq 2$. Then, A is orthogonal if and only if there exist $\theta_1, \ldots, \theta_n \in \mathbb{R}$ and $j_1, \ldots, j_n, k_1, \ldots, k_n \in \{1, \ldots, n\}$ such that

$$A = \operatorname{sign}(\det A)\prod_{i=1}^{n} P(\theta_i, j_i, k_i),$$

where

$$P(\theta, j, k) \triangleq I_n + [(\cos\theta) - 1](E_{j,j} + E_{k,k}) + (\sin\theta)(E_{j,k} - E_{k,j}).$$

(Remark: $P(\theta, j, k)$ is a *plane* or *Givens rotation.* See Fact 3.6.17.) (Problem: Generalize this result to $\mathbb{C}^{n\times n}$.)

Fact 5.13.14. Let $A \in \mathbb{F}^{n\times n}$. Then, $A^{2*}A = A^*A^2$ if and only if there exist a projector $B \in \mathbb{F}^{n\times n}$ and a Hermitian matrix $C \in \mathbb{F}^{n\times n}$ such that $A = BC$. (Proof: See [474].)

Fact 5.13.15. Let $A \in \mathbb{R}^{n\times n}$. Then, $|\det A| = 1$ if and only if A is the product of $n+2$ or fewer involutory matrices that have exactly one negative eigenvalue. In addition, the following statements hold:

i) If $n = 2$, then 3 or fewer factors are needed.

ii) If $A \neq \alpha I$ for all $\alpha \in \mathbb{R}$ and $\det A = (-1)^n$, then n or fewer factors are needed.

iii) If $\det A = (-1)^{n+1}$, then $n + 1$ or fewer factors are needed.

(Proof: See [133,472].) (Remark: The minimal number of factors for unitary A is given in [182].)

Fact 5.13.16. Let $A \in \mathbb{F}^{n\times n}$, and define $r_0 \triangleq n$ and $r_k \triangleq \operatorname{rank} A^k$ for all $k = 1, 2, \ldots$. Then, there exists $B \in \mathbb{C}^{n\times n}$ such that $A = B^2$ if and only if the sequence $\{r_k - r_{k+1}\}_{k=0}^{\infty}$ does not contain two successive occurrences of the same odd integer and, if $r_0 - r_1$ is odd, then $r_0 + r_2 \geq 1 + 2r_1$. Now, assume that $A \in \mathbb{R}^{n\times n}$. Then, there exists $B \in \mathbb{R}^{n\times n}$ such that $A = B^2$ if and only if the above condition holds and, for every negative eigenvalue λ of A and for every positive integer k, the Jordan form of A has an even number of $k \times k$ blocks associated with λ. (Proof: See [289, p. 472].) (Remark: See Fact 11.14.31.) (Remark: For all $l \geq 2$, $A \triangleq N_l$ does not have a complex square root.) (Remark: Uniqueness is discussed in [314]. mth roots are considered in [468].)

Fact 5.13.17. Let $A \in \mathbb{C}^{n\times n}$ be group invertible. Then, there exists $B \in \mathbb{C}^{n\times n}$ such that $A = B^2$.

Fact 5.13.18. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and define $\{P_k\}_{k=0}^{\infty} \subset \mathbb{F}^{n\times n}$ and $\{Q_k\}_{k=0}^{\infty} \subset \mathbb{F}^{n\times n}$ by

$$P_0 \triangleq A, \qquad Q_0 \triangleq I,$$

and, for $k \in \mathbb{P}$,

$$P_{k+1} \triangleq \tfrac{1}{2}\big(P_k + Q_k^{-1}\big),$$

$$Q_{k+1} \triangleq \tfrac{1}{2}\big(Q_k + P_k^{-1}\big).$$

Then,

$$B \triangleq \lim_{k\to\infty} P_k$$

exists and satisfies $B^2 = A$. Furthermore,

$$\lim_{k\to\infty} Q_k = A^{-1}.$$

(Proof: See [170, 277].) (Remark: This sequence is a modified Newton-Raphson algorithm based on the *matrix sign function*. See [327].) (Remark: See Fact 8.7.20.)

Fact 5.13.19. Let $A \in \mathbb{C}^{n\times n}$ be nonsingular. Then, there exist a semisimple matrix $S_1 \in \mathbb{C}^{n\times n}$ and a nilpotent matrix $S_2 \in \mathbb{C}^{n\times n}$ such that $S_1S_2 = S_2S_1$ and $A = S_1(I+S_2)$. (Proof: The result follows from the Jordan decomposition.)

Fact 5.13.20. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and let $r \triangleq \operatorname{rank} A$. Then, there exists $B \in \mathbb{F}^{n\times r}$ such that $A = BB^*$.

Fact 5.13.21. Let $A \in \mathbb{F}^{n\times n}$, and let $k \in \mathbb{P}$. Then, there exists a unique matrix $B \in \mathbb{F}^{n\times n}$ such that

$$A = B(B^*B)^k.$$

(Proof: See [461].)

Fact 5.13.22. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist symmetric matrices $B, C \in \mathbb{F}^{n\times n}$, one of which is singular, such that $A = BC$. (Proof: See [466, p. 82].) (Remark: Note that

$$\begin{bmatrix} \beta_1 & \beta_2 & 1 \\ \beta_2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 \end{bmatrix} = \begin{bmatrix} -\beta_0 & 0 & 0 \\ 0 & \beta_2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

and use Theorem 5.2.3.) (Remark: This result is due to Frobenius. The identity is a *Bezout matrix factorization*; see Fact 4.8.6. See [104, 105, 260].)

(Remark: Symmetric, not Hermitian.)

Fact 5.13.23. Let $A \in \mathbb{C}^{n\times n}$. Then, $\det A$ is real if and only if A is the product of four Hermitian matrices. Furthermore, four is the smallest number of factors in general. (Proof: See [618].)

Fact 5.13.24. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is the product of two nonnegative-semidefinite matrices if and only if A is similar to a nonnegative-semidefinite matrix.

ii) If A is nilpotent, then A is the product of three nonnegative-semidefinite matrices.

iii) If A is singular, then A is the product of four nonnegative-semidefinite matrices.

iv) $\det A > 0$ and $A \neq \alpha I$ for all $\alpha \leq 0$ if and only if A is the product of four positive-definite matrices.

v) $\det A > 0$ if and only if A is the product of five positive-definite matrices.

(Proof: [48, 260, 617, 618].) (Remark: See [618] for factorizations of complex matrices and operators.) (Example:

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 7 & 10 \end{bmatrix}\begin{bmatrix} 13/2 & -5 \\ -5 & 4 \end{bmatrix}\begin{bmatrix} 8 & 5 \\ 5 & 13/4 \end{bmatrix}\begin{bmatrix} 25/8 & -11/2 \\ -11/2 & 10 \end{bmatrix}.)$$

Fact 5.13.25. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) $A = BC$, where $B \in \mathbf{S}^n$ and $C \in \mathbf{N}^n$, if and only if A^2 is diagonalizable over $\mathbb{R}$ and $\operatorname{spec}(A) \subset [0,\infty)$.

ii) $A = BC$, where $B \in \mathbf{S}^n$ and $C \in \mathbf{P}^n$, if and only if A is diagonalizable over $\mathbb{R}$.

iii) $A = BC$, where $B, C \in \mathbf{N}^n$, if and only if $A = DE$, where $D \in \mathbf{N}^n$ and $E \in \mathbf{P}^n$.

iv) $A = BC$, where $B \in \mathbf{N}^n$ and $C \in \mathbf{P}^n$, if and only if A is diagonalizable over $\mathbb{R}$ and $\operatorname{spec}(A) \subset [0,\infty)$.

v) $A = BC$, where $B, C \in \mathbf{P}^n$, if and only if A is diagonalizable over $\mathbb{R}$ and $\operatorname{spec}(A) \subset [0,\infty)$.

(Proof: See [286, 614, 617].)

Fact 5.13.26. Let $A \in \mathbb{R}^{n\times n}$ be singular and assume that A is not a 2×2 nilpotent matrix. Then, there exist nilpotent matrices $B, C \in \mathbb{R}^{n\times n}$ such that $A = BC$ and $\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} A$. (Proof: See [616].)

Fact 5.13.27. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular. Then, A is similar to A^{-1} if and only if A is the product of two involutory matrices. If, in addition, A is orthogonal, then A is the product of two reflectors. (Proof: See [53, 179, 612, 613] or [466, p. 108].) (Problem: Construct these reflectors for $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$.)

Fact 5.13.28. Let $A \in \mathbb{R}^{n\times n}$. Then, $|\det A| = 1$ if and only if A is the product of four or fewer involutory matrices. (Proof: [54, 253, 517].)

Fact 5.13.29. Let $A \in \mathbb{R}^{n\times n}$. Then, A is the identity or singular if and only if A is the product of n or fewer idempotent matrices. Furthermore, $\operatorname{rank}(A - I) \le k\operatorname{def}(A)$, where $k \in \mathbb{N}$, if and only if A is the product of k idempotent matrices. (Proof: See [55].) (Problem: Explicitly construct the two factors when $\operatorname{rank} A = 1$ and A is not idempotent. Example: $\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$.)

Fact 5.13.30. Let $A \in \mathbb{R}^{n\times n}$, where $n \ge 2$. Then, A is the product of two commutators. (Proof: See [618].)

Fact 5.13.31. Let $A \in \mathbb{R}^{n\times n}$, and assume that $\det A = 1$. Then, there exist nonsingular matrices $B, C \in \mathbb{R}^{n\times n}$ such that $A = BCB^{-1}C^{-1}$. (Proof: See [507].) (Remark: The product is a *multiplicative commutator*. This result is due to Shoda.)

Fact 5.13.32. Let $A \in \mathbb{R}^{n\times n}$ be orthogonal and assume that $\det A = 1$. Then, there exist reflectors $B, C \in \mathbb{R}^{n\times n}$ such that $A = BCB^{-1}C^{-1}$. (Proof: See [544].)

Fact 5.13.33. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, there exists an involutory matrix $B \in \mathbb{F}^{n\times n}$ and a symmetric matrix $C \in \mathbb{F}^{n\times n}$ such that $A = BC$. (Proof: See [240].)

Fact 5.13.34. Let $A \in \mathbb{F}^{n\times n}$, and assume that n is even. Then, the following statements are equivalent:

i) A is the product of two skew-symmetric matrices.

ii) Every elementary divisor of A has even algebraic multiplicity.

iii) There exists $B \in \mathbb{F}^{n/2\times n/2}$ such that A is similar to $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$.

(Remark: In *i)* the factors are skew symmetric even when A is complex.) (Proof: See [241, 618].)

Fact 5.13.35. Let $A \in \mathbb{R}^{n\times n}$ be skew symmetric. If $n = 4, 8, 12\ldots,$ then A is the product of five or fewer skew-symmetric matrices. If $n = 6, 10, 14, \ldots,$ then A is the product of seven or fewer skew-symmetric matrices. (Proof: See [348].)

Fact 5.13.36. Let $A \in \mathbb{F}^{n\times n}$. Then, there exist a symmetric matrix $B \in \mathbb{F}^{n\times n}$ and a skew-symmetric matrix $C \in \mathbb{F}^{n\times n}$ such that $A = BC$ if and only if A is similar to $-A$. (Proof: See [487].)

Fact 5.13.37. Let $A \in \mathbb{F}^{n\times m}$, and let $r \triangleq \operatorname{rank} A$. Then, there exist $B \in \mathbb{F}^{n\times r}$ and $C \in \mathbb{R}^{r\times m}$ such that $A = BC$. Furthermore, $\operatorname{rank} B = \operatorname{rank} C = r$.

Fact 5.13.38. Let $A \in \mathbb{F}^{n\times m}$, where $n \le m$. Then, there exist $M \in \mathbb{F}^{n\times n}$ and $S \in \mathbb{F}^{n\times m}$ such that M is nonnegative semidefinite, S satisfies $SS^* = I_n$, and $A = MS$. Furthermore, M is given uniquely by $M = (AA^*)^{1/2}$. If, in addition, $\operatorname{rank} A = n$, then S is given uniquely by $S = (AA^*)^{-1/2}A$.

Fact 5.13.39. Let $A \in \mathbb{F}^{n\times m}$, where $m \le n$. Then, there exist $M \in \mathbb{F}^{m\times m}$ and $S \in \mathbb{F}^{n\times m}$ such that M is nonnegative semidefinite, S satisfies $S^*S = I_m$, and $A = SM$. Furthermore, M is given uniquely by $M = (A^*A)^{1/2}$. If, in addition, $\operatorname{rank} A = m$, then S is given uniquely by $S = A(A^*A)^{-1/2}$.

Fact 5.13.40. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular. Then, these exist unique matrices $M, S \in \mathbb{F}^{n\times n}$ such that $A = MS$, M is nonnegative semidefinite, and S is unitary. Furthermore, S is given uniquely by $S = (AA^*)^{-1/2}A$. In addition, A is nonsingular if and only if M is unique. In this case, M is given by $M = (AA^*)^{1/2}$.

Fact 5.13.41. Let $M_1, M_2 \in \mathbb{F}^{n\times n}$ be positive definite, let $S_1, S_2 \in \mathbb{F}^{n\times n}$ be unitary, and assume that $M_1S_1 = S_2M_2$. Then, $S_1 = S_2$. (Proof: Let $A = M_1S_1 = S_2M_2$. Then, $S_1 = \left(S_2M_2^2S_2^*\right)^{-1/2}S_2M_2 = S_2$.)

Fact 5.13.42. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular and let $M, S \in \mathbb{F}^{n\times n}$ be such that $A = MS$, M is nonnegative semidefinite, and S is unitary. Then, A is normal if and only if $MS = SM$. (Proof: See [287, p. 414].)

5.14 Notes

It is sometimes useful to define block-companion form matrices in which the scalars are replaced by matrix blocks [231]. The companion form illustrates but one connection between matrices and polynomials. Additional connections are given by the *comrade form*, *Leslie form*, *Schwarz form*, *Routh form*, *confederate form*, and *congenial form*. See [61, 64] and Fact 11.14.23 and Fact 11.14.24 for the Schwarz and Routh forms.

The multi-companion form and the elementary multi-companion form are generally know as *rational canonical forms*, while the multi-companion form is traditionally called the *Frobenius canonical form* [66]. The derivation of the Jordan form by means of the elementary multi-companion form and the hypercompanion form follows [456]. Corollary 5.3.8, Corollary 5.3.9, and Proposition 5.5.18 are given in [104, 105, 534, 535, 538]. Corollary 5.3.9 is due to Frobenius. Canonical forms for congruence transformations are given in [360, 548].

Chapter Six

Generalized Inverses

Generalized inverses provide a useful extension of the matrix inverse to singular matrices and to rectangular matrices that are neither left nor right invertible.

6.1 Moore-Penrose Generalized Inverse

Let $A \in \mathbb{F}^{n\times m}$. If A is nonzero, then, by the singular value decomposition Theorem 5.6.3, there exist orthogonal matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ such that

$$A = S_1 \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} S_2, \tag{6.1.1}$$

where $B \triangleq \operatorname{diag}[\sigma_1(A), \ldots, \sigma_r(A)]$, $r \triangleq \operatorname{rank} A$, and $\sigma_1(A) \ge \sigma_2(A) \ge \cdots \ge \sigma_r(A) > 0$ are the positive singular values of A. In (6.1.1), some of the bordering zero matrices may be empty. Then, the *(Moore-Penrose) generalized inverse* A^+ of A is the $m \times n$ matrix

$$A^+ \triangleq S_2^* \begin{bmatrix} B^{-1} & 0 \\ 0 & 0 \end{bmatrix} S_1^*. \tag{6.1.2}$$

If $A = 0_{n\times m}$, then $A^+ \triangleq 0_{m\times n}$, while if $m = n$ and $\det A \ne 0$, then $A^+ = A^{-1}$. In general, it is helpful to remember that A^+ and A^* are the same size. It is easy to verify that A^+ satisfies

$$AA^+A = A, \tag{6.1.3}$$

$$A^+AA^+ = A^+, \tag{6.1.4}$$

$$(AA^+)^* = AA^+, \tag{6.1.5}$$

$$(A^+A)^* = A^+A. \tag{6.1.6}$$

Hence, for all $A \in \mathbb{F}^{n\times m}$ there exists a matrix $X \in \mathbb{F}^{m\times n}$ satisfying the four conditions

$$AXA = A, \tag{6.1.7}$$
$$XAX = X, \tag{6.1.8}$$
$$(AX)^* = AX, \tag{6.1.9}$$
$$(XA)^* = XA. \tag{6.1.10}$$

We now show that X is uniquely defined by (6.1.7)-(6.1.10).

Theorem 6.1.1. Let $A \in \mathbb{F}^{n\times m}$. Then, $X = A^+$ is the unique matrix $X \in \mathbb{F}^{m\times n}$ satisfying (6.1.7)-(6.1.10).

Proof. Suppose there exists $X \in \mathbb{F}^{m\times n}$ satisfying (6.1.7)-(6.1.10). Then,

$$\begin{aligned} X &= XAX = X(AX)^* = XX^*A^* = XX^*(AA^+A)^* = XX^*A^*A^{+*}A^* \\ &= X(AX)^*(AA^+)^* = XAXAA^+ = XAA^+ = (XA)^*A^+ = A^*X^*A^+ \\ &= (AA^+A)^*X^*A^+ = A^*A^{+*}A^*X^*A^+ = (A^+A)^*(XA)^*A^+ \\ &= A^+AXAA^+ = A^+AA^+ = A^+. \end{aligned}$$ □

Given $A \in \mathbb{F}^{n\times m}$, $X \in \mathbb{F}^{m\times n}$ is a *(1)-inverse* of A if (6.1.7) holds, a *(1,2)-inverse* of A if (6.1.7) and (6.1.8) hold, etc.

Proposition 6.1.2. Let $A \in \mathbb{F}^{n\times m}$, and assume that A is right invertible. Then, $X \in \mathbb{F}^{m\times n}$ is a right inverse of A if and only if X is a (1)-inverse of A. Furthermore, every right inverse (or, equivalently, every (1)-inverse) of A is also a (2,3)-inverse of A.

Proof. Suppose that $AX = I_n$, that is, $X \in \mathbb{F}^{m\times n}$ is a right inverse of A. Then, $AXA = A$, which implies that X is a (1)-inverse of A. Conversely, let X be a (1)-inverse of A, that is, $AXA = A$. Then, letting $\hat{X} \in \mathbb{F}^{m\times n}$ denote a right inverse of A, it follows that $AX = AXA\hat{X} = A\hat{X} = I_n$. Hence, X is a right inverse of A. Finally, if X is a right inverse of A, then it is also a (2,3)-inverse of A. □

Proposition 6.1.3. Let $A \in \mathbb{F}^{n\times m}$, and assume that A is left invertible. Then, $X \in \mathbb{F}^{m\times n}$ is a left inverse of A if and only if X is a (1)-inverse of A. Furthermore, every left inverse (or, equivalently, every (1)-inverse) of A is also a (2,4)-inverse of A.

It can now be seen that A^+ is a particular (right, left) inverse when A is (right, left) invertible.

Corollary 6.1.4. Let $A \in \mathbb{F}^{n\times m}$. If A is right invertible, then A^+ is a right inverse of A. Furthermore, if A is left invertible, then A^+ is a left inverse of A.

The following result provides an explicit expression for A^+ when A is right or left invertible. It is helpful to note that A is (right, left) invertible if and only if $(AA^*,\ A^*A)$ is positive definite.

Proposition 6.1.5. Let $A \in \mathbb{F}^{n\times m}$. If A is right invertible, then

$$A^+ = A^*(AA^*)^{-1}. \tag{6.1.11}$$

If A is left invertible, then

$$A^+ = (A^*A)^{-1}A^*. \tag{6.1.12}$$

Proof. The result follows by verifying (6.1.7)-(6.1.10) with $X = A^+$. □

Proposition 6.1.6. Let $A \in \mathbb{F}^{n\times m}$. Then, the following statements hold:

i) $A = 0$ if and only if $A^+ = 0$.

ii) $(A^+)^+ = A$.

iii) $\overline{A}^+ = \overline{A^+}$.

iv) $\left(A^{\mathrm{T}}\right)^+ = (A^+)^{\mathrm{T}} = A^{+\mathrm{T}}$.

v) $(A^*)^+ = (A^+)^* \triangleq A^{+*}$.

vi) $\mathcal{R}(A) = \mathcal{R}(AA^+) = \mathcal{R}(AA^*) = \mathcal{N}(I - AA^+)$.

vii) $\mathcal{R}(A^*) = \mathcal{R}(A^*A) = \mathcal{R}(A^+) = \mathcal{R}(A^+A)$.

viii) $\mathcal{N}(A) = \mathcal{N}(A^+A) = \mathcal{N}(A^*A) = \mathcal{R}(I - A^+A)$.

ix) $\mathcal{N}(A^*) = \mathcal{N}(A^+) = \mathcal{N}(AA^+) = \mathcal{R}(I - AA^+)$.

x) AA^+ is the projector onto $\mathcal{R}(A)$.

xi) A^+A is the projector onto $\mathcal{R}(A^*)$.

xii) $I - A^+A$ is the projector onto $\mathcal{N}(A)$.

xiii) $I - AA^+$ is the projector onto $\mathcal{N}(A^*)$.

xiv) $x \in \mathcal{R}(A)$ if and only if $x = AA^+x$.

xv) $\operatorname{rank} A = \operatorname{rank} A^+ = \operatorname{rank} AA^+ = \operatorname{rank} A^+A = \operatorname{tr} AA^+ = \operatorname{tr} A^+A$.

xvi) $(A^*A)^+ = A^+A^{+*}$.

xvii) $(AA^*)^+ = A^{+*}A^+$.

xviii) $AA^+ = A(A^*A)^+A^*$.

xix) $A^+A = A^*(AA^*)^+A.$

xx) $A = AA^*A^{*+} = A^{*+}A^*A.$

xxi) $A^* = A^*AA^+ = A^+AA^*.$

xxii) $A^+ = A^*(AA^*)^+ = (A^*A)^+A^*.$

xxiii) $A^{+*} = (AA^*)^+A = A(A^*A)^+.$

xxiv) $A = A(A^*A)^+A^*A = AA^*A(A^*A)^+.$

xxv) $A = AA^*(AA^*)^+A = (AA^*)^+AA^*A.$

xxvi) If $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ are unitary, then $(S_1AS_2)^+ = S_2^*A^+S_1^*$.

xxvii) If A is (normal, Hermitian, nonnegative semidefinite, positive definite), then so is A^+.

xxviii) A is range Hermitian if and only if $AA^+ = A^+A$.

Theorem 2.6.3 showed that the equation $Ax = b$, where $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$, has a solution $x \in \mathbb{F}^m$ if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$. In particular, $Ax = b$ has a unique solution $x \in \mathbb{F}^m$ if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} = m$, while $Ax = b$ has infinitely many solutions if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix} < m$. We are now prepared to characterize these nonunique solutions.

Proposition 6.1.7. Let $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$. Then, the following statements are equivalent:

i) There exists $x \in \mathbb{F}^m$ satisfying $Ax = b$.

ii) $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & b \end{bmatrix}$.

iii) $b \in \mathcal{R}(A)$.

iv) $AA^+b = b$.

Now, assume that *i)*-*iv)* are satisfied. Then, the following statements hold:

v) If $x \in \mathbb{F}^m$ satisfies $Ax = b$, then

$$x = A^+b + (I - A^+A)x. \tag{6.1.13}$$

vi) For all $y \in \mathbb{F}^m$, $x \in \mathbb{F}^m$ given by

$$x = A^+b + (I - A^+A)y \tag{6.1.14}$$

satisfies $Ax = b$.

vii) Let $x \in \mathbb{F}^m$ be given by (6.1.14), where $y \in \mathbb{F}^m$. Then, $y = 0$ minimizes x^*x.

viii) Assume $\operatorname{rank} A = m$. Then, there exists a unique $x \in \mathbb{F}^m$ satisfying

$Ax = b$ given by $x = A^+b$. If, in addition, $A^{\mathrm{L}} \in \mathbb{F}^{m\times m}$ is a left inverse of A, then $A^{\mathrm{L}}b = A^+b$.

ix) Assume $\operatorname{rank} A = n$, and let $A^{\mathrm{R}} \in \mathbb{F}^{m\times n}$ be a right inverse of A. Then, $x = A^{\mathrm{R}}b$ satisfies $Ax = b$.

Proof. The equivalence of *i)*-*iii)* is immediate. To prove the equivalence of *iv)*, note that if there exists $x \in \mathbb{F}^n$ satisfying $Ax = b$, then $b = Ax = AA^+Ax = AA^+b$. Conversely, if $b = AA^+b$, then $x = A^+b$ satisfies $Ax = b$.

Now, suppose that *i)*-*iv)* are satisfied. To prove *v)* let $x \in \mathbb{F}^m$ satisfy $Ax = b$ so that $A^+Ax = A^+b$. Hence, $x = x + A^+b - A^+Ax = A^+b + (I - A^+A)x$. To prove *vi)* let $y \in \mathbb{F}^m$, and let $x \in \mathbb{F}^m$ be given by (6.1.14). Then, $Ax = AA^+b = b$. To prove *vii)* let $y \in \mathbb{F}^m$, and let $x \in \mathbb{F}^n$ be given by (6.1.14). Then, $x^*x = b^*A^{+*}A^+b + y^*(I - A^+A)y$. Therefore, x^*x is minimized by $y = 0$. To prove *viii)* suppose that $\operatorname{rank} A = m$. Then, A is left invertible, and it follows from Corollary 6.1.4 that A^+ is a left inverse of A. Hence, it follows from (6.1.13) that $x = A^+b$ is the unique solution to $Ax = b$. In addition, $x = A^{\mathrm{L}}b$. To prove *ix)* let $x = A^{\mathrm{R}}b$ and note that $AA^{\mathrm{R}}b = b$. □

Definition 6.1.8. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{k\times m}$, and $D \in \mathbb{F}^{k\times l}$, and define $\mathcal{A} \triangleq \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{F}^{(n+k)\times(m+l)}$. Then, the *Schur complement* $D|\mathcal{A}$ of D with respect to $\mathcal{A}$ is defined by

$$D|\mathcal{A} \triangleq A - BD^+C. \tag{6.1.15}$$

Likewise, the *Schur complement* $A|\mathcal{A}$ of A with respect to $\mathcal{A}$ is defined by

$$A|\mathcal{A} \triangleq D - CA^+B. \tag{6.1.16}$$

6.2 Drazin Generalized Inverse

We now introduce a different type of generalized inverse, which applies only to square matrices but which is more useful in certain applications. Let $A \in \mathbb{F}^{n\times n}$. Then, A has a decomposition

$$A = S\begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}S^{-1}, \tag{6.2.1}$$

where $S \in \mathbb{F}^{n\times n}$ is nonsingular, $J_1 \in \mathbb{F}^{m\times m}$ is nonsingular, and $J_2 \in \mathbb{F}^{(n-m)\times(n-m)}$ is nilpotent. Then, the *Drazin generalized inverse* A^{D} of A is the matrix

$$A^{\mathrm{D}} \triangleq S\begin{bmatrix} J_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}S^{-1}. \tag{6.2.2}$$

Let $A \in \mathbb{F}^{n\times n}$. Then, it follows from Definition 5.5.1 that $\operatorname{ind} A = \operatorname{ind}_A(0)$. If A is nonsingular, then $\operatorname{ind} A = 0$, whereas $\operatorname{ind} A = 1$ if and only if A is singular and the zero eigenvalue of A is semisimple. In particular, $\operatorname{ind} 0_{n\times n} = 1$. Note that $\operatorname{ind} A$ is the size of the largest Jordan block of A associated with the zero eigenvalue of A.

It can be seen that A^{D} satisfies

$$A^{\mathrm{D}}AA^{\mathrm{D}} = A^{\mathrm{D}}, \tag{6.2.3}$$

$$AA^{\mathrm{D}} = A^{\mathrm{D}}A, \tag{6.2.4}$$

$$A^{k+1}A^{\mathrm{D}} = A^k, \tag{6.2.5}$$

where $k = \operatorname{ind} A$. Hence, for all $A \in \mathbb{F}^{n\times n}$ such that $\operatorname{ind} A = k$ there exists a matrix $X \in \mathbb{F}^{n\times n}$ satisfying the three conditions

$$XAX = X, \tag{6.2.6}$$

$$AX = XA, \tag{6.2.7}$$

$$A^{k+1}X = A^k. \tag{6.2.8}$$

We now show that X is uniquely defined by (6.2.6)-(6.2.8).

Theorem 6.2.1. Let $A \in \mathbb{F}^{n\times n}$, and let $k \triangleq \operatorname{ind} A$. Then, $X = A^{\mathrm{D}}$ is the unique matrix $X \in \mathbb{F}^{n\times n}$ satisfying (6.2.6)-(6.2.8).

Proof. Let $X \in \mathbb{F}^{n\times n}$ satisfy (6.2.6)-(6.2.8). If $k = 0$, then it follows from (6.2.8) that $X = A^{-1}$. Hence, let $A = S\begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}S^{-1}$, where $k = \operatorname{ind} A \ge 1$, $S \in \mathbb{F}^{n\times n}$ is nonsingular, $J_1 \in \mathbb{F}^{m\times m}$ is nonsingular, and $J_2 \in \mathbb{F}^{(n-m)\times(n-m)}$ is nilpotent. Now, let $\hat{X} \triangleq S^{-1}XS = \begin{bmatrix} \hat{X}_1 & \hat{X}_{12} \\ \hat{X}_{21} & \hat{X}_2 \end{bmatrix}$ be partitioned conformably with $S^{-1}AS = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}$. Since, by (6.2.7), $\hat{A}\hat{X} = \hat{X}\hat{A}$, it follows that $J_1\hat{X}_1 = \hat{X}_1J_1$, $J_1\hat{X}_{12} = \hat{X}_{12}J_2$, $J_2\hat{X}_{21} = \hat{X}_{21}J_1$, and $J_2\hat{X}_2 = \hat{X}_2J_2$. Since $J_2^k = 0$, it follows that $J_1\hat{X}_{12}J_2^{k-1} = 0$, and thus $\hat{X}_{12}J_2^{k-1} = 0$. By repeating this argument, it follows that $J_1\hat{X}_{12}J_2 = 0$, and thus $\hat{X}_{12}J_2 = 0$, which implies that $J_1\hat{X}_{12} = 0$ and thus $\hat{X}_{12} = 0$. Similarly, $\hat{X}_{21} = 0$, so that $\hat{X} = \begin{bmatrix} \hat{X}_1 & 0 \\ 0 & \hat{X}_2 \end{bmatrix}$. Now, (6.2.8) implies that $J_1^{k+1}\hat{X}_1 = J_1^k$ and hence $\hat{X}_1 = J_1^{-1}$. Next, (6.2.6) implies that $\hat{X}_2J_2\hat{X}_2 = \hat{X}_2$, which, together with $J_2\hat{X}_2 = \hat{X}_2J_2$, yields $\hat{X}_2^2J_2 = \hat{X}_2$. Consequently, $0 = \hat{X}_2^2J_2^k = \hat{X}_2J_2^{k-1}$ and thus, by repeating this argument, $\hat{X}_2 = 0$. Therefore, $A^{\mathrm{D}} = S\begin{bmatrix} J_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}S^{-1} = S\begin{bmatrix} \hat{X}_1 & 0 \\ 0 & 0 \end{bmatrix}S^{-1} = S\hat{X}S^{-1} = X$. □

Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{ind} A \le 1$ so that A is group invertible. In this case, the Drazin inverse A^{D} is denoted by $A^{\#}$, which is the *group* *generalized inverse* of A. Therefore, $A^{\#}$ satisfies

$$A^{\#}AA^{\#} = A^{\#}, \tag{6.2.9}$$

$$AA^{\#} = A^{\#}A, \tag{6.2.10}$$

$$AA^{\#}A = A, \tag{6.2.11}$$

while $A^{\#}$ is the unique matrix $X \in \mathbb{F}^{n\times n}$ satisfying

$$XAX = X, \tag{6.2.12}$$

$$AX = XA, \tag{6.2.13}$$

$$AXA = A. \tag{6.2.14}$$

Proposition 6.2.2. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is group invertible. Then, the following statements hold:

i) $A = 0$ if and only if $A^{\#} = 0$.

ii) $(A^{\#})^{\#} = A$.

iii) If A is idempotent, then $A^{\#} = A$.

iv) $AA^{\#}$ and $A^{\#}A$ are idempotent.

v) $\left(A^{\mathrm{T}}\right)^{\#} = (A^{\#})^{\mathrm{T}}$.

vi) $\operatorname{rank} A = \operatorname{rank} A^{\#} = \operatorname{rank} AA^{\#} = \operatorname{rank} A^{\#}A$.

vii) $\mathcal{R}(A) = \mathcal{R}(AA^{\#}) = \mathcal{N}(I - AA^{\#}) = \mathcal{R}(AA^{+}) = \mathcal{N}(I - AA^{+})$.

viii) $\mathcal{N}(A) = \mathcal{N}(AA^{\#}) = \mathcal{R}(I - AA^{\#}) = \mathcal{N}(A^{+}A) = \mathcal{R}(I - A^{+}A)$.

ix) $AA^{\#}$ is the idempotent matrix onto $\mathcal{R}(A)$ along $\mathcal{N}(A)$.

An alternative expression for the idempotent matrix onto $\mathcal{R}(A)$ along $\mathcal{N}(A)$ is given by Proposition 5.5.9.

6.3 Facts on the Moore-Penrose Generalized Inverse Involving One Matrix

Fact 6.3.1. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = 1$. Then,

$$A^{+} = (\operatorname{tr} AA^{*})^{-1}A^{*}.$$

Consequently, if $x \in \mathbb{F}^{n}$ and $y \in \mathbb{F}^{n}$ are nonzero, then

$$(xy^{*})^{+} = (x^{*}xy^{*}y)^{-1}yx^{*}.$$

Fact 6.3.2. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = m$. Then,

$$(AA^{*})^{+} = A(A^{*}A)^{-2}A^{*}.$$

Fact 6.3.3. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$A^+ = \lim_{\alpha\downarrow 0} A^*(AA^* + \alpha I)^{-1} = \lim_{\alpha\downarrow 0} (A^*A + \alpha I)^{-1}A^*.$$

Fact 6.3.4. Let $A \in \mathbb{F}^{n\times m}$, let $\chi_{AA^*}(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$, and let $n-k$ denote the smallest integer in $\{0,\ldots,n-1\}$ such that $\beta_k \neq 0$. Then,

$$A^+ = -\beta_{n-k}^{-1}A^*\Big[(AA^*)^{k-1} + \beta_{n-1}(AA^*)^{k-2} + \cdots + \beta_{n-k+1}I\Big].$$

(Proof: See [168].)

Fact 6.3.5. Let $A \in \mathbb{F}^{n\times n}$ and assume that A is Hermitian. Then, $\operatorname{In} A = \operatorname{In} A^+$.

Fact 6.3.6. Let $A \in \mathbb{F}^{n\times n}$ be a projector. Then, $A^+ = A$.

Fact 6.3.7. Let $A \in \mathbb{F}^{n\times n}$. Then, $A^+ = A$ if and only if A is tripotent and A^2 is Hermitian.

Fact 6.3.8. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is idempotent. Then,

$$A^+A + (I-A)(I-A)^+ = I.$$

(Proof: $\mathcal{N}(A) = \mathcal{R}(I - A^+A) = \mathcal{R}(I-A) = \mathcal{R}[(I-A)(I-A^+)]$.)

Fact 6.3.9. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is idempotent. Then,

$$A^*A^+A = A^+A$$

and

$$AA^+A^* = AA^+.$$

(Proof: Note that A^*A^+A is a projector and $\mathcal{R}(A^*A^+A) = \mathcal{R}(A^*) = \mathcal{R}(A^+A)$.)

Fact 6.3.10. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is idempotent. Then, $A + A^* - I$ is nonsingular, and

$$(A + A^* - I)^{-1} = AA^+ + A^+A - I.$$

(Proof: Use Fact 6.3.9.) (Remark: See [416, p. 457] for a geometric interpretation of this identity.)

Fact 6.3.11. Let $A \in \mathbb{F}^{n\times n}$, and let $r \triangleq \operatorname{rank} A$. Then, $A^+ = A^*$ if and only if $\sigma_1(A) = \sigma_r(A) = 1$.

Fact 6.3.12. Let $A \in \mathbb{F}^{n\times m}$ where $A \neq 0$, and let $r \triangleq \operatorname{rank} A$. Then, for all $i = 1,\ldots,r$, the singular values of A^+ are given by

$$\sigma_i(A^+) = \sigma_{r+1-i}^{-1}(A).$$

In particular,

$$\sigma_r(A) = 1/\sigma_{\max}(A^+).$$

If, in addition, $A \in \mathbb{F}^{n\times n}$ and A is nonsingular, then

$$\sigma_{\min}(A) = 1/\sigma_{\max}\big(A^{-1}\big).$$

Fact 6.3.13. Let $A \in \mathbb{F}^{n\times m}$. Then, $X = A^+$ is the unique matrix satisfying

$$\operatorname{rank}\begin{bmatrix} A & AA^+ \\ A^+A & X \end{bmatrix} = \operatorname{rank} A.$$

(Remark: See Fact 2.13.39 and Fact 6.5.5.) (Proof: See [203].)

Fact 6.3.14. Let $A \in \mathbb{F}^{n\times n}$ be centrohermitian. Then, A^+ is centrohermitian. (Proof: See [359].)

Fact 6.3.15. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) $A^2 = AA^*A$.

ii) A is the product of two projectors.

iii) $A = A(A^+)^2A$.

(Remark: This result is due to Crimmins. See [474].)

Fact 6.3.16. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$A^+ = 4(I + A^+A)^+A^+(I + AA^+)^+.$$

(Proof: Use Fact 6.4.20 with $B = A$.)

Fact 6.3.17. Let $A \in \mathbb{F}^{n\times n}$ be unitary. Then,

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} A^i = I - (A - I)(A - I)^+.$$

(Remark: $I-(A-I)(A-I)^+$ is the projector onto $\{x\colon\ Ax = x\} = \mathcal{N}(A-I)$.) (Remark: This result is the *ergodic theorem.*) (Proof: Use Fact 11.15.12 and Fact 11.15.14 and note that $(A - I)^* = (A - I)^+$. See [258, p. 185].)

Fact 6.3.18. Let $A \in \mathbb{F}^{n\times m}$, and define $\{B_i\}_{i=1}^\infty$ by

$$B_{i+1} \triangleq 2B_i - B_iAB_i,$$

where $B_0 \triangleq \alpha A^*$ and $\alpha \in (0, 2/\sigma_{\max}^2(A))$. Then,

$$\lim_{i\to\infty} B_i = A^+.$$

(Proof: See [64, p. 259] or [124, p. 250]. This result is due to Ben-Israel.) (Remark: This sequence is a Newton-Raphson algorithm.) (Remark: B_0 satisfies $\operatorname{sprad}(I - B_0A) < 1$.) (Remark: For the case in which A is square and nonsingular, see Fact 2.13.37.) (Problem: Does convergence hold for all $B_0 \in \mathbb{F}^{n\times n}$ satisfying $\operatorname{sprad}(I - B_0A) < 1$?)

6.4 Facts on the Moore-Penrose Generalized Inverse Involving Two or More Matrices

Fact 6.4.1. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, $AB = 0$ if and only if $B^+A^+ = 0$.

Fact 6.4.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then, $A^+B = 0$ if and only if $A^*B = 0$.

Fact 6.4.3. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$(AB)^+ = B_1^+A_1^+,$$

where $B_1 \triangleq A^+AB$ and $A_1 \triangleq AB_1B_1^+$. That is,

$$(AB)^+ = (A^+AB)^+\big[AB(A^+AB)^+\big]^+.$$

(Proof: See [6, p. 55].) (Remark: This result is due to Cline.)

Fact 6.4.4. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$(AB)^+ = B^+A^+$$

if and only if $\mathcal{R}(BB^*A^*) \subseteq \mathcal{R}(A^*)$ and $\mathcal{R}(A^*AB) \subseteq \mathcal{R}(B)$. (Proof: See [6, p. 53].) (Remark: This result is due to Greville.)

Fact 6.4.5. Let $A \in \mathbb{F}^{n\times r}$ and $B \in \mathbb{F}^{r\times m}$, and assume that $\operatorname{rank} A = \operatorname{rank} B = r$. Then,

$$(AB)^+ = B^+A^+ = B^*(BB^*)^{-1}(A^*A)^{-1}A^*.$$

Fact 6.4.6. Let $A, B \in \mathbb{F}^{n\times n}$ be range Hermitian. If $(AB)^+ = A^+B^+$, then AB is range Hermitian. (Proof: See [268].) (Remark: See Fact 8.9.10.)

Fact 6.4.7. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, and assume that $\operatorname{rank} B = m$. Then,

$$AB(AB)^+ = AA^+.$$

Fact 6.4.8. Let $A \in \mathbb{F}^{n\times m}$, let $B \in \mathbb{F}^{m\times n}$ satisfy $BAA^* = A^*$, and let

$C \in \mathbb{F}^{m\times n}$ satisfy $A^*AC = A^*$. Then,

$$A^+ = BAC.$$

(Proof: See [6, p. 36].) (Remark: This result is due to Decell.)

Fact 6.4.9. Let $A \in \mathbb{F}^{n\times m}$. Then, there exists $B \in \mathbb{F}^{m\times m}$ satisfying $BAB = B$ if and only if there exist projectors $C \in \mathbb{F}^{n\times n}$ and $D \in \mathbb{F}^{m\times m}$ such that $B = (CAD)^+$. (Proof: See [245].)

Fact 6.4.10. Let $A \in \mathbb{F}^{n\times n}$. Then, A is idempotent if and only if there exist projectors $B, C \in \mathbb{F}^{n\times n}$ such that $A = (BC)^+$. (Proof: Let $A = I$ in Fact 6.4.9.) (Remark: See [247].)

Fact 6.4.11. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{k\times m}$, $D \in \mathbb{F}^{k\times l}$. Then,

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A & B \end{bmatrix} &= \operatorname{rank} A + \operatorname{rank}\big(B - AA^+B\big)\\ &= \operatorname{rank} B + \operatorname{rank}\big(A - BB^+A\big),\end{aligned}$$

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A \\ C \end{bmatrix} &= \operatorname{rank} A + \operatorname{rank}\big(C - CA^+A\big)\\ &= \operatorname{rank} C + \operatorname{rank}\big(A - AC^+C\big),\end{aligned}$$

$$\operatorname{rank}\begin{bmatrix} A & B \\ C & 0 \end{bmatrix} = \operatorname{rank} B + \operatorname{rank} C + \operatorname{rank}\big[\big(I_n - BB^+\big)A\big(I_m - C^+C\big)\big].$$

Now, define $\mathcal{A} \triangleq \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$. Then,

$$\begin{aligned}\operatorname{rank}\mathcal{A} = {}& \operatorname{rank} A + \operatorname{rank} X + \operatorname{rank} Y\\ &+ \operatorname{rank}\big[\big(I_k - YY^+\big)(D|\mathcal{A})\big(I_p - X^+X\big)\big],\end{aligned}$$

where $X \triangleq B - AA^+B$ and $Y \triangleq C - CA^+A$. Consequently,

$$\operatorname{rank} A + \operatorname{rank}(D|\mathcal{A}) \le \operatorname{rank}\mathcal{A}.$$

Furthermore, if $AA^+B = B$ and $CA^+A = C$, then

$$\operatorname{rank} A + \operatorname{rank}(D|\mathcal{A}) = \operatorname{rank}\mathcal{A}.$$

Finally, if $n = m$ and A is nonsingular, then

$$\operatorname{rank} A + \operatorname{rank}\big(D - CA^{-1}B\big) \le \operatorname{rank}\mathcal{A}.$$

(Proof: See [128, 398].) (Remark: With certain restrictions the generalized inverses can be replaced by (1)-inverses.) (Remark: See Proposition 2.8.3.)

Fact 6.4.12. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} 0 & A \\ B & I \end{bmatrix} &= \operatorname{rank} A + \operatorname{rank}\begin{bmatrix} B & I - A^+A \end{bmatrix} \\ &= \operatorname{rank}\begin{bmatrix} A \\ I - BB^+ \end{bmatrix} + \operatorname{rank} B \\ &= \operatorname{rank} A + \operatorname{rank} B + \operatorname{rank}\big[\big(I - BB^+\big)\big(I - A^+A\big)\big] \\ &= n + \operatorname{rank} AB.\end{aligned}$$

Hence, the following statements hold:

i) $\operatorname{rank} AB = \operatorname{rank} A + \operatorname{rank} B - n$ if and only if $(I - BB^+)(I - A^+A) = 0$.

ii) $\operatorname{rank} AB = \operatorname{rank} A$ if and only if $\begin{bmatrix} B & I - A^+A \end{bmatrix}$ is right invertible.

iii) $\operatorname{rank} AB = \operatorname{rank} B$ if and only if $\left[\begin{smallmatrix} A \\ I - BB^+ \end{smallmatrix}\right]$ is left invertible.

(Proof: See [398].) (Remark: The generalized inverses can be replaced by arbitrary (1)-inverses.)

Fact 6.4.13. Let $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$. Then,

$$\begin{bmatrix} A & b \end{bmatrix}^+ = \begin{bmatrix} A^+[I - bc] \\ c \end{bmatrix},$$

where

$$c \triangleq \begin{cases} (b - AA^+b)^+, & b \ne AA^+b, \\ \dfrac{b^*(AA^*)^+}{1 + b^*(AA^*)^+b}, & b = AA^+b. \end{cases}$$

(Proof: See [6, p. 44], [202, p. 270], or [505, p. 148].) (Remark: This result is due to Greville.)

Fact 6.4.14. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\begin{bmatrix} A & B \end{bmatrix}^+ = \begin{bmatrix} A^+ - A^+B(C^+ + D) \\ C^+ + D \end{bmatrix},$$

where

$$C \triangleq (I - AA^+)B$$

and

$$D \triangleq (I - C^+C)[I + (I - C^+C)B^*(AA^*)^+B(I - C^+C)]^{-1}B^*(AA^*)^+(I - BC^+).$$

Furthermore,

$$\begin{bmatrix} A & B \end{bmatrix}^+ = \begin{cases} \begin{bmatrix} A^*(AA^* + BB^*)^{-1} \\ B^*(AA^* + BB^*)^{-1} \end{bmatrix}, & \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = n, \\ \begin{bmatrix} A^*A & A^*B \\ B^*A & B^*B \end{bmatrix}^{-1} \begin{bmatrix} A^* \\ B^* \end{bmatrix}, & \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = m + l, \\ \begin{bmatrix} A^*(AA^*)^{-1}(I - BE) \\ E \end{bmatrix}, & \operatorname{rank} A = n, \end{cases}$$

where

$$E \triangleq \left[I + B^*(AA^*)^{-1}B\right]^{-1} B^*(AA^*)^{-1}.$$

(Proof: See [147] or [387, p. 14].) (Remark: If $\begin{bmatrix} A & B \end{bmatrix}$ is square and nonsingular and $A^*B = 0$, then the second expression yields Fact 2.13.33.)

Fact 6.4.15. Let $A \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite, let $B \in \mathbb{F}^{n \times m}$, and define

$$\mathcal{A} \triangleq \begin{bmatrix} A & B \\ B^* & 0 \end{bmatrix}.$$

Then,

$$\mathcal{A}^+ = \begin{bmatrix} C^+ - C^+BD^+B^*C^+ & C^+BD^+ \\ (C^+BD^+)^* & DD^+ - D^+ \end{bmatrix},$$

where

$$C \triangleq A + BB^*, \qquad D \triangleq B^+C^+C.$$

(Proof: See [388, p. 58].) (Remark: Representations for the generalized inverse of a partitioned matrix are given in [47, 57, 76, 121, 124, 266, 301, 414, 415, 417, 418, 478, 489, 550, 593].)

Fact 6.4.16. Let $A \in \mathbb{F}^{n \times n}$ be Hermitian, let $b \in \mathbb{F}^n$, and define $S \triangleq I - A^+A$. Then,

$$(A + bb^*)^+ = \begin{cases} \left[I - (b^*Sb)^{-1}Sbb^*\right]A^+\left[I - (b^*Sb)^{-1}bb^*S\right] + (b^*Sb)^{-2}Sbb^*S, & Sb \neq 0, \\ A^+ - (1 + b^*A^+b)A^+bb^*A^+, & 1 + b^*A^+b \neq 0, \\ \left[I - (b^*A^{2+}b)^{-1}A^+bb^*A^+\right]A^+\left[I - (b^*A^{2+}b)^{-1}A^+bb^*A^+\right], & b^*A^+b = 0. \end{cases}$$

(Proof: See [421].) (Remark: Expressions for $(A + BB^*)^+$, where $B \in \mathbb{F}^{n \times l}$, are given in [421].)

Fact 6.4.17. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, let $C \in \mathbb{F}^{m\times m}$ be positive definite, and let $B \in \mathbb{F}^{n\times m}$. Then,

$$(A+BCB^*)^+ = A^+ - A^+B\big(C^{-1} + B^*A^+B\big)^{-1}B^*A^+$$

if and only if

$$AA^+B = B.$$

(Proof: See [442].) (Remark: $AA^+B = B$ is equivalent to $\mathcal{R}(B) \subseteq \mathcal{R}(A)$.)

Fact 6.4.18. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $A^*B = 0$ and $BA^* = 0$. Then,

$$(A+B)^+ = A^+ + B^+.$$

(Proof: Use Fact 2.10.6 and Fact 6.4.19. See [148].) (Remark: This result is due to Penrose.)

Fact 6.4.19. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank}(A+B) = \operatorname{rank} A + \operatorname{rank} B$. Then,

$$(A+B)^+ = (I - C^+B)A^+(I - BC^+) + C^+,$$

where $C \triangleq (I - AA^+)B(I - A^+A)$. (Proof: See [148].)

Fact 6.4.20. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$(A+B)^+ = (I + A^+B)^+(A^+ + A^+BA^+)(I + BA^+)^+$$

if and only if $AA^+B = B = BA^+A$. Furthermore, if $n = m$ and A is nonsingular, then

$$(A+B)^+ = \big(I + A^{-1}B\big)^+\big(A^{-1} + A^{-1}BA^{-1}\big)\big(I + BA^{-1}\big)^+.$$

(Proof: See [148].) (Remark: If A and $A+B$ are nonsingular, then the last statement yields $(A+B)^{-1} = (A+B)^{-1}(A+B)(A+B)^{-1}$ for which the assumption that A is nonsingular is superfluous.)

Fact 6.4.21. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{l\times k}$, and $C \in \mathbb{F}^{n\times k}$. Then, there exists $X \in \mathbb{F}^{m\times l}$ satisfying $AXB = C$ if and only if $AA^+CB^+B = C$. Furthermore, X satisfies $AXB = C$ if and only if there exists $Y \in \mathbb{F}^{m\times l}$ such that

$$X = A^+CB^+ + Y - A^+AYBB^+.$$

Finally, if $Y = 0$, then $\operatorname{tr} X^*X$ is minimized. (Proof: Use Proposition 6.1.7. See [388, p. 37] and, for Hermitian solutions, see [330].)

Fact 6.4.22. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = m$. Then, $A^{\mathrm{L}} \in \mathbb{F}^{m\times n}$ is a left inverse of A if and only if there exists $B \in \mathbb{F}^{m\times n}$ such that

$$A^{\mathrm{L}} = A^+ + B(I - AA^+).$$

(Proof: Use Fact 6.4.16 with $A = C = I_m$.)

Fact 6.4.23. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = n$. Then, $A^{\mathrm{R}} \in \mathbb{F}^{m\times n}$ is a right inverse of A if and only if there exists $B \in \mathbb{F}^{m\times n}$ such that

$$A^{\mathrm{R}} = A^+ + (I - A^+A)B.$$

(Proof: Use Fact 6.4.21 with $B = C = I_n$.)

Fact 6.4.24. Let $A \in \mathbb{F}^{n\times n}$, $x, y \in \mathbb{F}^n$, and $a \in \mathbb{F}$, and assume that $x \in \mathcal{R}(A)$. Then,

$$\begin{bmatrix} A & x \\ y^{\mathrm{T}} & a \end{bmatrix} = \begin{bmatrix} I & 0 \\ y^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} A & 0 \\ y^{\mathrm{T}} - y^{\mathrm{T}}A & a - y^{\mathrm{T}}A^+x \end{bmatrix} \begin{bmatrix} I & A^+x \\ 0 & 1 \end{bmatrix}.$$

(Remark: See Fact 2.12.4 and Fact 2.12.13 and note that $x = AA^+x$.) (Problem: Obtain a factorization for the case $x \notin \mathcal{R}(A)$.)

Fact 6.4.25. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\det \begin{bmatrix} A^*A & B^*A \\ B^*A & B^*B \end{bmatrix} = \det(A^*A)\det[B^*(I - AA^+)B]$$
$$= \det(B^*B)\det[A^*(I - BB^+)A].$$

Fact 6.4.26. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, $C \in \mathbb{F}^{m\times n}$, and $D \in \mathbb{F}^{m\times m}$, assume that either $\operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank} A$ or $\operatorname{rank} \begin{bmatrix} A \\ C \end{bmatrix} = \operatorname{rank} A$, and let A^- be a (1)-inverse of A. Then,

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = (\det A)\det(D - CA^-B).$$

(Proof: See [64, p. 266].)

Fact 6.4.27. Let $A, B \in \mathbb{F}^{n\times n}$ be projectors. Then,

$$\lim_{k\to\infty} A(BA)^k = 2A(A + B)^+B.$$

Furthermore, $2A(A + B)^+B$ is the projector onto $\mathcal{R}(A) \cap \mathcal{R}(B)$. (Proof: See [20].) (Remark: See Fact 6.4.28 and Fact 8.9.9.)

Fact 6.4.28. Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times l}$. Then,

$$\mathcal{R}(A) \cap \mathcal{R}(B) = \mathcal{R}[AA^+(AA^+ + BB^+)^+BB^+].$$

(Remark: See Theorem 2.3.1, and Fact 8.9.9.)

Fact 6.4.29. Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times l}$. Then, $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ if and only if $BB^+A = A$. (Proof: See [6, p. 35].)

Fact 6.4.30. Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times l}$. Then,

$$\operatorname{rank} AA^+(AA^+ + BB^+)^+BB^+ = \operatorname{rank} A + \operatorname{rank} B - \operatorname{rank}\begin{bmatrix} A & B \end{bmatrix}.$$

(Proof: Use Fact 6.4.28, Fact 2.10.26, and Fact 2.10.22.)

Fact 6.4.31. Let $A \in \mathbb{F}^{n\times m}$ and $b \in \mathbb{F}^n$, and define $f(x) \triangleq (Ax - b)^*(Ax - b)$, where $x \in \mathbb{F}^m$. Then, x minimizes f if and only if there exists $y \in \mathbb{F}^m$ such that

$$x = A^+b + (I - A^+A)y.$$

In this case,

$$f(x) = b^*(I - AA^+)b.$$

Finally, f has a unique minimizer if and only if A is left invertible. (Remark: The minimization of f is the *least squares problem*. See [6, 100].)

Fact 6.4.32. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, and define

$$f(X) \triangleq \operatorname{tr}[(AX - B)^*(AX - B)],$$

where $X \in \mathbb{F}^{m\times l}$. Then, $X = A^+B$ minimizes f. (Problem: Determine all minimizers.) (Problem: Consider $f(X) = \operatorname{tr}[(AX - B)^*C(AX - B)]$, where $C \in \mathbb{F}^{n\times n}$ is positive definite.)

Fact 6.4.33. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times m}$, and define

$$f(X) \triangleq \operatorname{tr}[(XA - B)^*(XA - B)],$$

where $X \in \mathbb{F}^{l\times n}$. Then, $X = BA^+$ minimizes f.

Fact 6.4.34. Let $A, B \in \mathbb{F}^{n\times m}$, and define

$$f(X) \triangleq \operatorname{tr}[(AX - B)^*(AX - B)],$$

where $X \in \mathbb{F}^{m\times m}$ is unitary. Then, $X = S_1S_2$ minimizes f, where $S_1\left[\begin{smallmatrix} \hat{B} & 0 \\ 0 & 0 \end{smallmatrix}\right]S_2$ is the singular value decomposition of A^*B. (Proof: See [64, p. 224].)

Fact 6.4.35. Let $A \triangleq \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right] \in \mathbb{F}^{(n+m)\times(n+m)}$, $B \in \mathbb{F}^{(n+m)\times l}$, $C \in \mathbb{F}^{l\times(n+m)}$, $D \in \mathbb{F}^{l\times l}$, and $\mathcal{A} \triangleq \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$, and assume that A and A_{11} are nonsingular. Then,

$$A|\mathcal{A} = (A_{11}|A)|(A_{11}|\mathcal{A}).$$

(Proof: See [466, pp. 18, 19].) (Remark: This result is due to Haynsworth.) (Problem: Is the result true if either A or A_{11} is singular?)

6.5 Facts on the Drazin and Group Generalized Inverses

Fact 6.5.1. Let $A \in \mathbb{F}^{n\times n}$. Then, AA^{D} is idempotent.

Fact 6.5.2. Let $A \in \mathbb{F}^{n\times n}$. Then, $A = A^{\mathrm{D}}$ if and only if A is tripotent.

Fact 6.5.3. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$(A^*)^{\mathrm{D}} = \left(A^{\mathrm{D}}\right)^*.$$

Fact 6.5.4. Let $A \in \mathbb{F}^{n\times n}$, and let $r \in \mathbb{P}$. Then,

$$\left(A^{\mathrm{D}}\right)^r = (A^r)^{\mathrm{D}}.$$

Fact 6.5.5. Let $A \in \mathbb{F}^{n\times n}$. Then, $X = A^{\mathrm{D}}$ is the unique matrix satisfying

$$\operatorname{rank}\begin{bmatrix} A & AA^{\mathrm{D}} \\ A^{\mathrm{D}}A & X \end{bmatrix} = \operatorname{rank} A.$$

(Remark: See Fact 2.13.39 and Fact 6.3.13.) (Proof: See [631].)

Fact 6.5.6. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $AB = BA$. Then,

$$(AB)^{\mathrm{D}} = B^{\mathrm{D}}A^{\mathrm{D}},$$
$$A^{\mathrm{D}}B = BA^{\mathrm{D}},$$
$$AB^{\mathrm{D}} = B^{\mathrm{D}}A.$$

Fact 6.5.7. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{ind} A = \operatorname{rank} A = 1$. Then,

$$A^{\#} = \left(\operatorname{tr} A^2\right)^{-1}A.$$

Consequently, if $x, y \in \mathbb{F}^n$ satisfy $x^*y \neq 0$, then

$$(xy^*)^{\#} = (x^*y)^{-2}xy^*.$$

In particular, $1_{n\times n}^{\#} = n^{-2}1_{n\times n}$.

Fact 6.5.8. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is range Hermitian.

ii) $A^+A = AA^+$.

iii) $A^+ = A^{\mathrm{D}}$.

iv) $\operatorname{ind} A \le 1$ and $A^+ = A^{\#}$.

v) $\operatorname{ind} A \le 1$ and $(A^+)^2 = \left(A^2\right)^+$.

vi) There exists a nonsingular matrix $B \in \mathbb{F}^{n\times n}$ such that $A = A^*B$.

(Proof: To prove *i*) $\Longrightarrow$ *vi*) use Corollary 5.4.4 and $B = S\begin{bmatrix} B_0^{-*}B_0 & 0 \\ 0 & I \end{bmatrix}S^*$.)

Fact 6.5.9. Let $A \in \mathbb{F}^{n\times n}$. Then, A is group invertible if and only if $\lim_{\alpha\to 0}(A+\alpha I)^{-1}A$ exists. In this case,

$$\lim_{\alpha\to 0}(A+\alpha I)^{-1}A = AA^{\#}.$$

6.6 Notes

The proof of the uniqueness of A^+ is given in [388]. Most of the results given in this chapter can be found in [124]. Reverse order laws for the generalized inverse of a product are discussed in [592]. Additional books on generalized inverses include [78,106,477]. Generalized inverses are widely used in least squares methods; see [102, 124, 355]. Applications to singular differential equations are considered in [123]. Historical remarks are given in [77].

Chapter Seven

Kronecker and Schur Algebra

In this chapter we introduce Kronecker matrix algebra, which is useful for analyzing linear matrix equations.

7.1 Kronecker Product

For $A \in \mathbb{F}^{n \times m}$ define the *vec* operator as

$$\operatorname{vec} A \triangleq \begin{bmatrix} \operatorname{col}_1(A) \\ \vdots \\ \operatorname{col}_m(A) \end{bmatrix} \in \mathbb{F}^{nm}, \tag{7.1.1}$$

which is the column vector of size $nm \times 1$ obtained by stacking the columns of A. We recover A from $\operatorname{vec} A$ by writing

$$A = \operatorname{vec}^{-1}(\operatorname{vec} A) \tag{7.1.2}$$

Proposition 7.1.1. Let $A \in \mathbb{F}^{n \times m}$ and $B \in \mathbb{F}^{m \times n}$. Then,

$$\operatorname{tr} AB = \left(\operatorname{vec} A^{\mathrm{T}}\right)^{\mathrm{T}} \operatorname{vec} B = \left(\operatorname{vec} B^{\mathrm{T}}\right)^{\mathrm{T}} \operatorname{vec} A. \tag{7.1.3}$$

Proof. Note that

$$\begin{aligned} \operatorname{tr} AB &= \sum_{i=1}^{n} e_i^{\mathrm{T}} A B e_i = \sum_{i=1}^{n} \operatorname{row}_i(A) \operatorname{col}_i(B) \\ &= \sum_{i=1}^{n} \left[\operatorname{col}_i\left(A^{\mathrm{T}}\right)\right]^{\mathrm{T}} \operatorname{col}_i(B) \\ &= \begin{bmatrix} \operatorname{col}_1^{\mathrm{T}}\left(A^{\mathrm{T}}\right) & \cdots & \operatorname{col}_n^{\mathrm{T}}\left(A^{\mathrm{T}}\right) \end{bmatrix} \begin{bmatrix} \operatorname{col}_1(B) \\ \vdots \\ \operatorname{col}_n(B) \end{bmatrix} \\ &= \left(\operatorname{vec} A^{\mathrm{T}}\right)^{\mathrm{T}} \operatorname{vec} B. \end{aligned}$$

□

Next, we introduce the Kronecker product.

Definition 7.1.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then, the *Kronecker product* $A \otimes B \in \mathbb{F}^{nl\times mk}$ of A is the partitioned matrix

$$A \otimes B \triangleq \begin{bmatrix} A_{(1,1)}B & A_{(1,2)}B & \cdots & A_{(1,m)}B \\ \vdots & \vdots & \ldots & \vdots \\ A_{(n,1)}B & A_{(n,2)}B & \cdots & A_{(n,m)}B \end{bmatrix}. \tag{7.1.4}$$

Unlike matrix multiplication, the Kronecker product $A \otimes B$ does not entail a restriction on either the size of A or the size of B.

The following results are immediate consequences of the definition of the Kronecker product.

Proposition 7.1.3. Let $\alpha \in \mathbb{F}$, $A \in \mathbb{F}^{n\times m}$, and $B \in \mathbb{F}^{l\times k}$. Then,

$$A \otimes (\alpha B) = (\alpha A) \otimes B = \alpha(A \otimes B), \tag{7.1.5}$$

$$\overline{A \otimes B} = \overline{A} \otimes \overline{B}, \tag{7.1.6}$$

$$(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}, \tag{7.1.7}$$

$$(A \otimes B)^* = A^* \otimes B^*. \tag{7.1.8}$$

Proposition 7.1.4. Let $A, B \in \mathbb{F}^{n\times m}$ and $C \in \mathbb{F}^{l\times k}$. Then,

$$(A + B) \otimes C = A \otimes C + B \otimes C \tag{7.1.9}$$

and

$$C \otimes (A + B) = C \otimes A + C \otimes B. \tag{7.1.10}$$

Proposition 7.1.5. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{l\times k}$, and $C \in \mathbb{F}^{j\times i}$. Then,

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C. \tag{7.1.11}$$

Hence, we write $A \otimes B \otimes C$ for $A \otimes (B \otimes C)$ and $(A \otimes B) \otimes C$.

The next result illustrates an important form of compatibility between matrix multiplication and the Kronecker product.

Proposition 7.1.6. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{l\times k}$, $C \in \mathbb{F}^{m\times j}$, and $D \in \mathbb{F}^{k\times i}$, and assume that $mj = lk$. Then,

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \tag{7.1.12}$$

Proof. Note that the ij block of $(A\otimes B)(C\otimes D)$ is given by

$$\begin{aligned}[(A\otimes B)(C\otimes D)]_{ij} &= \begin{bmatrix} A_{(i,1)}B & \cdots & A_{(i,m)}B \end{bmatrix}\begin{bmatrix} C_{(1,j)}D \\ \vdots \\ C_{(m,j)}D \end{bmatrix}\\ &= \sum_{k=1}^{m} A_{(i,k)}C_{(k,j)}BD = (AC)_{(i,j)}BD \\ &= (AC\otimes BD)_{ij}. \end{aligned}$$ □

Next, we consider the inverse of a Kronecker product.

Proposition 7.1.7. Suppose $A\in\mathbb{F}^{n\times n}$ and $B\in\mathbb{F}^{m\times m}$ are nonsingular. Then,

$$(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}. \tag{7.1.13}$$

Proof. Note that

$$(A\otimes B)\left(A^{-1}\otimes B^{-1}\right) = AA^{-1}\otimes BB^{-1} = I_n\otimes I_m = I_{nm}.$$ □

Proposition 7.1.8. Let $x\in\mathbb{F}^n$ and $y\in\mathbb{F}^m$. Then,

$$xy^{\mathrm{T}} = x\otimes y^{\mathrm{T}} = y^{\mathrm{T}}\otimes x \tag{7.1.14}$$

and

$$\operatorname{vec} xy^{\mathrm{T}} = y\otimes x \tag{7.1.15}$$

The following result concerns the vec of the product of three matrices.

Proposition 7.1.9. Let $A\in\mathbb{F}^{n\times m}$, $B\in\mathbb{F}^{m\times l}$, and $C\in\mathbb{F}^{l\times k}$. Then,

$$\operatorname{vec}(ABC) = \left(C^{\mathrm{T}}\otimes A\right)\operatorname{vec} B. \tag{7.1.16}$$

Proof. Using (7.1.12) and (7.1.15), it follows that

$$\begin{aligned}\operatorname{vec} ABC &= \operatorname{vec}\sum_{i=1}^{l} A\mathrm{col}_i(B)e_i^{\mathrm{T}}C = \sum_{i=1}^{l}\operatorname{vec}\left[A\mathrm{col}_i(B)\left(C^{\mathrm{T}}e_i\right)^{\mathrm{T}}\right]\\ &= \sum_{i=1}^{l}\left[C^{\mathrm{T}}e_i\right]\otimes\left[A\mathrm{col}_i(B)\right] = \left(C^{\mathrm{T}}\otimes A\right)\sum_{i=1}^{l} e_i\otimes\mathrm{col}_i(B)\\ &= \left(C^{\mathrm{T}}\otimes A\right)\sum_{i=1}^{l}\operatorname{vec}\left[\mathrm{col}_i(B)e_i^{\mathrm{T}}\right] = \left(C^{\mathrm{T}}\otimes A\right)\operatorname{vec} B.\end{aligned}$$ □

The following result concerns eigenvalues and eigenvectors of the Kronecker product of two matrices.

Proposition 7.1.10. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\operatorname{mspec}(A\otimes B) = \{\lambda\mu\colon\ \lambda \in \operatorname{mspec}(A), \mu \in \operatorname{mspec}(B)\}_{\mathrm{m}}. \tag{7.1.17}$$

If, in addition, $x \in \mathbb{C}^n$ is an eigenvector of A associated with $\lambda \in \operatorname{spec}(A)$ and $y \in \mathbb{C}^n$ is an eigenvector of B associated with $\mu \in \operatorname{spec}(B)$, then $x\otimes y$ is an eigenvector of $A\otimes B$ associated with $\lambda\mu$.

Proof. Using (7.1.12), we have

$$(A\otimes B)(x\otimes y) = (Ax)\otimes(By) = (\lambda x)\otimes(\mu y) = \lambda\mu(x\otimes y). \qquad \square$$

Proposition 7.1.10 shows that $\operatorname{mspec}(A\otimes B) = \operatorname{mspec}(B\otimes A)$. Consequently, it follows that $\det(A\otimes B) = \det(B\otimes A)$ and $\operatorname{tr}(A\otimes B) = \operatorname{tr}(B\otimes A)$. The following results are generalizations of these identities.

Proposition 7.1.11. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\det(A\otimes B) = \det(B\otimes A) = (\det A)^m(\det B)^n. \tag{7.1.18}$$

Proof. Let $\operatorname{mspec}(A) = \{\lambda_1,\ldots,\lambda_n\}_{\mathrm{m}}$ and $\operatorname{mspec}(B) = \{\mu_1,\ldots,\mu_m\}_{\mathrm{m}}$. Then, Proposition 7.1.10 implies that

$$\begin{aligned}\det(A\otimes B) &= \prod_{i,j=1}^{n,m}\lambda_i\mu_j = \left(\lambda_1^m\prod_{j=1}^{m}\mu_j\right)\cdots\left(\lambda_n^m\prod_{j=1}^{m}\mu_j\right)\\ &= (\lambda_1\cdots\lambda_n)^m(\mu_1\cdots\mu_m)^n = (\det A)^m(\det B)^n. \qquad \square\end{aligned}$$

Proposition 7.1.12. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\operatorname{tr}(A\otimes B) = \operatorname{tr}(B\otimes A) = (\operatorname{tr} A)(\operatorname{tr} B). \tag{7.1.19}$$

Proof. Note that

$$\begin{aligned}\operatorname{tr}(A\otimes B) &= \operatorname{tr}(A_{(1,1)}B) + \cdots + \operatorname{tr}(A_{(n,n)}B)\\ &= [A_{(1,1)} + \cdots + A_{(n,n)}]\operatorname{tr} B = (\operatorname{tr} A)(\operatorname{tr} B). \qquad \square\end{aligned}$$

Next, define the *Kronecker permutation matrix* $P_{n,m} \in \mathbb{F}^{nm\times nm}$ by

$$P_{n,m} \triangleq \sum_{i,j=1}^{n,m} E_{i,j,n\times m}\otimes E_{j,i,m\times n}. \tag{7.1.20}$$

Proposition 7.1.13. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{vec} A^{\mathrm{T}} = P_{n,m}\operatorname{vec} A. \tag{7.1.21}$$

7.2 Kronecker Sum and Linear Matrix Equations

Next, we define the Kronecker sum of two square matrices.

Definition 7.2.1. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then, the *Kronecker sum* $A \oplus B \in \mathbb{F}^{nm\times nm}$ of A and B is

$$A \oplus B \triangleq A \otimes I_m + I_n \otimes B. \tag{7.2.1}$$

Proposition 7.2.2. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and $C \in \mathbb{F}^{l\times l}$. Then,

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C. \tag{7.2.2}$$

Hence, we write $A \oplus B \oplus C$ for $A \oplus (B \oplus C)$ and $(A \oplus B) \oplus C$.

In Proposition 7.1.10 it was shown that if $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$, then $\lambda\mu \in \operatorname{spec}(A \otimes B)$. Next, we present an analogous result involving Kronecker sums.

Proposition 7.2.3. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\operatorname{mspec}(A \oplus B) = \{\lambda + \mu\colon\ \lambda \in \operatorname{mspec}(A),\ \mu \in \operatorname{mspec}(B)\}_{\mathrm{m}}. \tag{7.2.3}$$

Now, let $x \in \mathbb{C}^n$ be an eigenvector of A associated with $\lambda \in \operatorname{spec}(A)$, and let $y \in \mathbb{C}^m$ be an eigenvector of B associated with $\mu \in \operatorname{spec}(B)$. Then, $x \otimes y$ is an eigenvector of $A \oplus B$ associated with $\lambda + \mu$.

Proof. Note that

$$\begin{aligned}(A \oplus B)(x \otimes y) &= (A \otimes I_m)(x \otimes y) + (I_n \otimes B)(x \otimes y)\\ &= (Ax \otimes y) + (x \otimes By) = (\lambda x \otimes y) + (x \otimes \mu y)\\ &= \lambda(x \otimes y) + \mu(x \otimes y) = (\lambda + \mu)(x \otimes y).\end{aligned}$$

□

The next result concerns the existence and uniqueness of solutions to *Sylvester's equation.* See Fact 5.8.11 and Proposition 11.7.3.

Proposition 7.2.4. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and $C \in \mathbb{F}^{n\times m}$. Then, $X \in \mathbb{F}^{n\times m}$ satisfies

$$AX + XB + C = 0 \tag{7.2.4}$$

if and only if X satisfies

$$\big(B^{\mathrm{T}} \oplus A\big)\operatorname{vec} X + \operatorname{vec} C = 0. \tag{7.2.5}$$

Consequently, $B^{\mathrm{T}} \oplus A$ is nonsingular if and only if there exists a unique matrix $X \in \mathbb{F}^{n\times m}$ satisfying (7.2.4). In this case, X is given by

$$X = -\operatorname{vec}^{-1}\Big[\big(B^{\mathrm{T}} \oplus A\big)^{-1} \operatorname{vec} C\Big]. \tag{7.2.6}$$

Furthermore, $B^{\mathrm{T}} \oplus A$ is singular and $\operatorname{rank} B^{\mathrm{T}} \oplus A = \operatorname{rank}\left[\begin{array}{cc} B^{\mathrm{T}} \oplus A & \operatorname{vec} C \end{array}\right]$ if and only if there exist infinitely many matrices $X \in \mathbb{F}^{n\times m}$ satisfying (7.4.15). Then, the set of solutions of (7.2.4) is given by $X + \mathcal{N}\big(B^{\mathrm{T}} \oplus A\big)$.

Proof. Note that (7.2.4) is equivalent to

$$\begin{aligned} 0 &= \operatorname{vec}(AXI + IXB) + \operatorname{vec} C = (I \otimes A) \operatorname{vec} X + (B^{*} \otimes I) \operatorname{vec} X + \operatorname{vec} C \\ &= (B^{*} \otimes I + I \otimes A) \operatorname{vec} X + \operatorname{vec} C = (B^{*} \oplus A) \operatorname{vec} X + \operatorname{vec} C, \end{aligned}$$

which yields (7.2.5). The remaining results follow from Corollary 2.6.5. □

7.3 Schur Product

An alternative form of vector and matrix multiplication is given by the *Schur product*. If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times m}$, then $A \circ B \in \mathbb{F}^{n\times m}$ is defined by

$$(A \circ B)_{(i,j)} \triangleq A_{(i,j)} B_{(i,j)}, \tag{7.3.1}$$

that is, $A \circ B$ is formed by means of entry-by-entry multiplication. For matrices $A, B, C \in \mathbb{F}^{n\times m}$, the commutative, associative, and distributive identities

$$A \circ B = B \circ A, \tag{7.3.2}$$

$$A \circ (B \circ C) = (A \circ B) \circ C, \tag{7.3.3}$$

$$A \circ (B + C) = A \circ B + A \circ C \tag{7.3.4}$$

are valid. For a real scalar $\alpha \ge 0$ and $A \in \mathbb{F}^{n\times m}$, the *Schur power* $A^{\{\alpha\}}$ is defined by

$$\left(A^{\{\alpha\}}\right)_{(i,j)} \triangleq \left(A_{(i,j)}\right)^{\alpha}. \tag{7.3.5}$$

Thus, $A^{\{2\}} = A \circ A$. Note that $A^{\{0\}} = 1_{n\times m}$, while $\alpha < 0$ is allowed if A has no zero entries. Finally, for all $A \in \mathbb{F}^{n\times m}$,

$$A \circ 1_{n\times m} = 1_{n\times m} \circ A = A. \tag{7.3.6}$$

Proposition 7.3.1. Let $A, \in \mathbb{F}^{n\times n}$. Then, $A \circ B$ is a submatrix of $A \otimes B$ consisting of rows $\operatorname{row}_1(A\otimes B), \operatorname{row}_{n+2}(A\otimes B), \operatorname{row}_{2n+3}(A\otimes B), \ldots, \operatorname{row}_{n^2}(A\otimes B)$ and columns $\operatorname{col}_1(A\otimes B), \operatorname{col}_{m+2}(A\otimes B), \operatorname{col}_{2m+3}(A\otimes B), \ldots, \operatorname{col}_{m^2}(A\otimes B)$. If, in addition, $n = m$, then $A \circ B$ is a principal submatrix of $A \otimes B$.

Proof. See [394] or [289, p. 304]. □

7.4 Facts on the Kronecker Product

Fact 7.4.1. Let $x, y \in \mathbb{F}^n$. Then,

$$x \otimes y = (x \otimes I_n)y = (I_n \otimes y)x.$$

Fact 7.4.2. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$ be (diagonal, upper triangular, lower triangular). Then, so is $A \otimes B$.

Fact 7.4.3. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and $l \in \mathbb{P}$. Then,

$$(A \otimes B)^l = A^l \otimes B^l.$$

Fact 7.4.4. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{vec} A = (I_m \otimes A) \operatorname{vec} I_m = \left(A^{\mathrm{T}} \otimes I_n\right) \operatorname{vec} I_n.$$

Fact 7.4.5. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\operatorname{vec} AB = (I_l \otimes A) \operatorname{vec} B = \left(B^{\mathrm{T}} \otimes A\right) \operatorname{vec} I_m = \sum_{i=1}^{m} \operatorname{col}_i\left(B^{\mathrm{T}}\right) \otimes \operatorname{col}_i(A).$$

Fact 7.4.6. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $C \in \mathbb{F}^{l\times n}$. Then,

$$\operatorname{tr} ABC = (\operatorname{vec} A)^{\mathrm{T}} (B \otimes I) \operatorname{vec} C^{\mathrm{T}}.$$

Fact 7.4.7. Let $A, B, C \in \mathbb{F}^{n\times n}$, where C is symmetric. Then,

$$(\operatorname{vec} C)^{\mathrm{T}} (A \otimes B) \operatorname{vec} C = (\operatorname{vec} C)^{\mathrm{T}} (B \otimes A) \operatorname{vec} C.$$

Fact 7.4.8. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, $C \in \mathbb{F}^{l\times k}$, and $D \in \mathbb{F}^{k\times n}$. Then,

$$\operatorname{tr} ABCD = (\operatorname{vec} A)^{\mathrm{T}} \left(B \otimes D^{\mathrm{T}}\right) \operatorname{vec} C^{\mathrm{T}}.$$

Fact 7.4.9. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $k \in \mathbb{P}$. Then,

$$(AB)^{\otimes k} = A^{\otimes k} B^{\otimes k},$$

where $A^{\otimes k} \triangleq A \otimes A \otimes \cdots \otimes A$, with A appearing k times.

Fact 7.4.10. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$(A \oplus A)^2 = A^2 \oplus A^2 + 2A \otimes A.$$

Fact 7.4.11. Let $A, C \in \mathbb{F}^{n\times m}$ and $B, D \in \mathbb{F}^{l\times k}$, and assume that A is (left equivalent, right equivalent, biequivalent) to C and B is (left equivalent, right equivalent, biequivalent) to D. Then, $A \otimes B$ is (left equivalent, right equivalent, biequivalent) to $C \otimes D$.

Fact 7.4.12. Let $A, B, C, D \in \mathbb{F}^{n\times n}$, and assume that A is (similar, congruent, unitarily similar) to C and B is (similar, congruent, unitarily similar) to D. Then, $A\otimes B$ is (similar, congruent, unitarily similar) to $C\otimes D$.

Fact 7.4.13. Let $A_1, \ldots, A_r \in \mathbb{F}^{n\times n}$ be (Hermitian, nonnegative semidefinite, positive definite, range Hermitian, normal, semisimple, group invertible). Then, so is $A_1 \otimes \cdots \otimes A_r$.

Fact 7.4.14. Let $A_1, \ldots, A_l \in \mathbb{F}^{n\times n}$ be skew Hermitian. If l is (even, odd), then $A_1 \otimes \cdots \otimes A_l$ is (Hermitian, skew Hermitian).

Fact 7.4.15. Let $A_1, \ldots, A_l \in \mathbb{F}^{n\times n}$ be (Hermitian, nonnegative semidefinite, positive definite, skew Hermitian). Then, so is $A_1 \oplus \cdots \oplus A_l$.

Fact 7.4.16. Let $A_{i,j} \in \mathbb{F}^{n_i\times n_j}$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, l$. Then,

$$\begin{bmatrix} A_{11} & A_{22} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \cdot\vdots\cdot & \cdot\vdots\cdot \end{bmatrix} \otimes B = \begin{bmatrix} A_{11}\otimes B & A_{22}\otimes B & \cdots \\ A_{21}\otimes B & A_{22}\otimes B & \cdots \\ \vdots & \cdot\vdots\cdot & \cdot\vdots\cdot \end{bmatrix}.$$

Fact 7.4.17. Let $x \in \mathbb{F}^k$, and let $A_i \in \mathbb{F}^{n\times n_i}$ for all $i = 1, \ldots, l$. Then,

$$x \otimes \begin{bmatrix} A_1 & \cdots & A_l \end{bmatrix} = \begin{bmatrix} x\otimes A_1 & \cdots & x\otimes A_l \end{bmatrix}.$$

Fact 7.4.18. Let $A \in \mathbb{F}^{n\times n}$ be (range Hermitian, normal). Then, so is $A \oplus A$.

Fact 7.4.19. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then, the eigenvalues of $\sum_{i,j=1,1}^{k,l} \gamma_{ij} A^i \otimes B^j$ are of the form $\sum_{i,j=1,1}^{k,l} \gamma_{ij}\lambda^i\mu^j$, where $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$ and an associated eigenvector is given by $x \otimes y$, where $x \in \mathbb{F}^n$ is an eigenvector of A associated with $\lambda \in \operatorname{spec}(A)$ and $y \in \mathbb{F}^n$ is an eigenvector of B associated with $\mu \in \operatorname{spec}(B)$. (Remark: This result is due to Stephanos.) (Proof: Let $Ax = \lambda x$ and $By = \mu y$. Then, $\gamma_{ij}(A^i\otimes B^j)(x\otimes y) = \gamma_{ij}\lambda^i\mu^j(x \otimes y)$. See [217], [353, p. 411], or [384, p. 83].)

Fact 7.4.20. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then,

$$\operatorname{rank}(A \otimes B) = (\operatorname{rank} A)(\operatorname{rank} B).$$

(Proof: Use the singular value decomposition of $A\otimes B$.) (Remark: See Fact 8.15.9.)

Fact 7.4.21. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$, and assume that $nl = mk$ and $n \neq m$. Then, $A \otimes B$ is singular. (Proof: See [289, p. 250].)

Fact 7.4.22. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$. Then, the algebraic multiplicity of the zero eigenvalue of $A \otimes B$ is greater than or equal to $|n-m|\min\{n,m\}$. (Proof: See [289, p. 249].)

Fact 7.4.23. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and let $\gamma \in \operatorname{spec}(A \otimes B)$. Then,

$$\sum \operatorname{gm}_A(\lambda)\operatorname{gm}_B(\mu) \le \operatorname{gm}_{A\otimes B}(\gamma) \le \operatorname{am}_{A\otimes B}(\gamma) = \sum \operatorname{am}_A(\lambda)\operatorname{am}_B(\mu),$$

where both sums are taken over all $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$ such that $\lambda\mu = \gamma$.

Fact 7.4.24. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and let $\gamma \in \operatorname{spec}(A \otimes B)$. Then, $\operatorname{ind}_{A\otimes B}(\gamma) \le 1$ if and only if $\operatorname{ind}_A(\lambda) \le 1$ and $\operatorname{ind}_B(\mu) \le 1$ for all $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$ such that $\lambda\mu = \gamma$.

Fact 7.4.25. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\operatorname{ind} A \otimes B = \max\{\operatorname{ind} A, \operatorname{ind} B\}.$$

Fact 7.4.26. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and let $\gamma \in \operatorname{spec}(A \oplus B)$. Then,

$$\sum \operatorname{gm}_A(\lambda)\operatorname{gm}_B(\mu) \le \operatorname{gm}_{A\oplus B}(\gamma) \le \operatorname{am}_{A\oplus B}(\gamma) = \sum \operatorname{am}_A(\lambda)\operatorname{am}_B(\mu),$$

where both sums are taken over all $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$ such that $\lambda + \mu = \gamma$.

Fact 7.4.27. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and let $\gamma \in \operatorname{spec}(A \oplus B)$. Then, $\operatorname{ind}_{A\oplus B}(\gamma) \le 1$ if and only if $\operatorname{ind}_A(\lambda) \le 1$ and $\operatorname{ind}_B(\mu) \le 1$ for all $\lambda \in \operatorname{spec}(A)$ and $\mu \in \operatorname{spec}(B)$ such that $\lambda + \mu = \gamma$.

Fact 7.4.28. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$, where B is nonnegative semidefinite, and let $\operatorname{mspec}(B) = \{\lambda_1, \ldots, \lambda_m\}_{\mathrm{m}}$. Then,

$$\det(A \oplus B) = \prod_{i=1}^{n} \det(\lambda_i I + A).$$

(Proof: See [419, p. 40].) (Remark: Expressions for $\det(A\otimes B + C\otimes D)$ are given in [419].) (Problem: Weaken the assumption that B is nonnegative semidefinite.)

Fact 7.4.29. The Kronecker permutation matrix has the following properties:

i) $P_{n,m}$ is a permutation matrix.

ii) $P_{n,m}^{\mathrm{T}} = P_{m,n}$.

iii) $P_{n,m}$ is orthogonal.

iv) $P_{n,m}P_{m,n} = I_{nm}$.

v) $P_{1,m} = I_m$ and $P_{n,1} = I_n$.

vi) If $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$, then

$$P_{n,m}(y \otimes x) = x \otimes y.$$

vii) If $A \in \mathbb{F}^{n\times m}$, then

$$P_{n,l}(I_l \otimes A) = (A \otimes I_l)P_{m,l}.$$

viii) If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$, then

$$P_{n,l}(A \otimes B)P_{m,k} = B \otimes A$$

and

$$\mathrm{vec}(A \otimes B) = (I_m \otimes P_{k,n} \otimes I_l)[(\mathrm{vec}\,A) \otimes (\mathrm{vec}\,B)].$$

ix) If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$, then

$$\mathrm{tr}\,AB = \mathrm{tr}[P_{m,n}(A \otimes B)].$$

Fact 7.4.30. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and $C \in \mathbb{F}^{n\times m}$, and assume that $\det(B^{\mathrm{T}} \oplus A) \neq 0$. Then, $X \in \mathbb{F}^{n\times m}$ satisfies

$$A^2X + 2AXB + XB^2 + C = 0$$

if and only if

$$X = -\,\mathrm{vec}^{-1}\Big[\big(B^{\mathrm{T}} \oplus A\big)^{-2}\mathrm{vec}\,C\Big].$$

Fact 7.4.31. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then,

$$(A \otimes B)^+ = A^+ \otimes B^+.$$

Fact 7.4.32. Let $A \in \mathbb{F}^{n\times m}$, and let $k \in \mathbb{P}$ satisfy $1 \le k \le \min\{n, m\}$. Furthermore, define the kth *compound* $A^{(k)}$ to be the $\binom{n}{k} \times \binom{m}{k}$ matrix whose entries are $k \times k$ subdeterminants of A, ordered lexicographically. (Example: For $n = k = 3$, subsets of the rows and columns of A are chosen in the order $(1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), \ldots$.) Specifically, $\big(A^{(k)}\big)_{(i,j)}$ is the $k \times k$ subdeterminant of A corresponding to the ith selection of k rows of A and the jth selection of k columns of A. Then, the following statements hold:

i) $\big[A^{(k)}\big]^{\mathrm{T}} = \big[A^{\mathrm{T}}\big]^{(k)}$.

ii) $\det A^{(k)} = (\det A)^{\binom{n-1}{k-1}}$.

iii) If $n = m$ and A is nonsingular, then $\big[A^{(k)}\big]^{-1} = \big[A^{-1}\big]^{(k)}$.

iv) If $B \in \mathbb{F}^{m\times l}$, then $(AB)^{(k)} = A^{(k)}B^{(k)}$.

Now, assume that $n = m$, let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\rm m}$, and, for $i = 0, \ldots, k$, define $A^{(k,i)}$ by

$$(A+sI)^{(k)} = s^kA^{(k,0)} + s^{k-1}A^{(k,1)} + \cdots + sA^{(k,k-1)} + A^{(k,k)}.$$

Then,

$$\operatorname{mspec}\Big[A^{(2,1)}\Big] = \{\lambda_i + \lambda_j\colon\ i,j = 1,\ldots,n, i<j\}_{\rm m},$$

$$\operatorname{mspec}\Big(A^{(2)}\Big) = \{\lambda_i\lambda_j\colon\ i,j = 1,\ldots,n, i<j\}_{\rm m},$$

and

$$\operatorname{mspec}\bigg(\Big[A^{(2,1)}\Big]^2 - 4A^{(2)}\bigg) = \big\{(\lambda_i - \lambda_j)^2\colon\ i,j = 1,\ldots,n, i<j\big\}_{\rm m}.$$

(Proof: See [202, pp. 142–155] and [466, p. 124].) (Remark: $\big(A^{(2,1)}\big)^2 - 4A^{(2)}$ is the *discriminant* of A. The discriminant of A is singular if and only if A has a repeated eigenvalue.) (Remark: The compound operation is related to the *bialternate product* since $\operatorname{mspec}(2A\cdot I) = \operatorname{mspec}\big(A^{(2,1)}\big)$ and $\operatorname{mspec}(A\cdot A) = \operatorname{mspec}\big(A^{(2)}\big)$. See [217,239], [319, pp. 313–320], and [384, pp. 84, 85].) (Problem: Express $A\cdot B$ in terms of compounds.)

7.5 Facts on the Schur Product

Fact 7.5.1. Let $x, y, z \in \mathbb{F}^n$. Then,

$$x^{\rm T}(y\circ z) = z^{\rm T}(x\circ y) = y^{\rm T}(x\circ z).$$

Fact 7.5.2. Let $w, y \in \mathbb{F}^n$ and $x, z \in \mathbb{F}^m$. Then,

$$\big(wx^{\rm T}\big)\circ\big(yz^{\rm T}\big) = (w\circ y)(x\circ z)^{\rm T}.$$

Fact 7.5.3. Let $A \in \mathbb{F}^{n\times n}$ and $d \in \mathbb{F}^n$. Then,

$$\operatorname{diag}(d)A = A\circ d1_{1\times n}.$$

Fact 7.5.4. Let $A \in \mathbb{F}^{n\times m}$, $D_1 \in \mathbb{F}^{n\times n}$, and $D_2 \in \mathbb{F}^{m\times m}$, where D_1 and D_2 are diagonal. Then,

$$(D_1A)\circ(BD_2) = D_1(A\circ B)D_2.$$

Fact 7.5.5. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then,

$$\operatorname{rank}(A\circ B) \le \operatorname{rank}(A\otimes B) = (\operatorname{rank} A)(\operatorname{rank} B).$$

(Proof: Use Proposition 7.3.1.) (Remark: See Fact 8.15.9.)

Fact 7.5.6. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{tr}\left[(A\circ B)(A\circ B)^{\mathrm{T}}\right] = \operatorname{tr}\left[(A\circ A)(B\circ B)^{\mathrm{T}}\right].$$

Fact 7.5.7. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times n}$, $a \in \mathbb{F}^m$, and $b \in \mathbb{F}^n$. Then,

$$\operatorname{tr}\left[A\left(B\circ ab^{\mathrm{T}}\right)\right] = b^{\mathrm{T}}\left(A\circ B^{\mathrm{T}}\right)a.$$

Fact 7.5.8. Let $A, B \in \mathbb{F}^{n\times m}$ and $C \in \mathbb{F}^{m\times n}$. Then,

$$I\circ\left[A\left(B^{\mathrm{T}}\circ C\right)\right] = I\circ[(A\circ B)C] = I\circ\left[\left(A\circ C^{\mathrm{T}}\right)B^{\mathrm{T}}\right].$$

Hence,

$$\operatorname{tr}\left[A\left(B^{\mathrm{T}}\circ C\right)\right] = \operatorname{tr}[(A\circ B)C] = \operatorname{tr}\left[\left(A\circ C^{\mathrm{T}}\right)B^{\mathrm{T}}\right].$$

Fact 7.5.9. Let $x \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times m}$, and define $x^A \in \mathbb{R}^n$ by

$$x^A \triangleq \begin{bmatrix} x_{(1)}^{A_{(1,1)}}\cdots x_{(m)}^{A_{(1,m)}} \\ \vdots \\ x_{(1)}^{A_{(n,1)}}\cdots x_{(m)}^{A_{(n,m)}} \end{bmatrix},$$

where every entry is assumed to exist. Then, the following statements hold:

i) If $a \in \mathbb{R}$, then $a^x = \begin{bmatrix} a^{x_{(1)}} \\ \vdots \\ a^{x_{(m)}} \end{bmatrix}$.

ii) $x^{-A} = \left(x^A\right)^{\{-1\}}$.

iii) If $y \in \mathbb{R}^m$, then $(x\circ y)^A = x^A \circ y^A$.

iv) If $B \in \mathbb{R}^{n\times m}$, then $x^{A+B} = x^A \circ x^B$.

v) If $B \in \mathbb{R}^{l\times n}$, then $\left(x^A\right)^B = x^{BA}$.

vi) If $a \in \mathbb{R}$, then $(a^x)^A = a^{Ax}$.

vii) If $A^{\mathrm{L}} \in \mathbb{R}^{m\times n}$ is a left inverse of A and $y = x^A$, then $x = y^{A^{\mathrm{L}}}$.

viii) If $A \in \mathbb{R}^{n\times n}$ is nonsingular and $y = x^A$, then $x = y^{A^{-1}}$.

ix) Define $f(x) \triangleq x^A$. Then, $f'(x) = \operatorname{diag}\left(x^A\right)A\operatorname{diag}\left(x^{\{-1\}}\right)$.

(Remark: These operations arise in modeling chemical reaction kinetics. See [365].)

Fact 7.5.10. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular. Then,

$$\left(A\circ A^{-\mathrm{T}}\right)1_{n\times 1} = 1_{n\times 1}$$

and

$$1_{1\times n}\left(A\circ A^{-\mathrm{T}}\right) = 1_{1\times n}.$$

(Proof: See [316].)

Fact 7.5.11. Let $A \in \mathbb{R}^{n\times n}$, and assume that $A \geq\geq 0$. Then,

$$\operatorname{sprad}\Big[\big(A \circ A^{\mathrm{T}}\big)^{\{1/2\}}\Big] \leq \operatorname{sprad}(A) \leq \operatorname{sprad}\big[\tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\big].$$

(Proof: See [502].)

Fact 7.5.12. Let $A_1, \ldots, A_r \in \mathbb{R}^{n\times n}$ and $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$, where $A_i \geq\geq 0$ for all $i = 1, \ldots, r$, $\alpha_i > 0$ for all $i = 1, \ldots, r$, and $\sum_{i=1}^r \alpha_i \geq 1$. Then,

$$\operatorname{sprad}\Big(A_1^{\{\alpha_1\}} \circ \cdots \circ A_r^{\{\alpha_r\}}\Big) \leq \prod_{i=1}^{r} [\operatorname{sprad}(A_i)]^{\alpha_i}.$$

In particular, let $A \in \mathbb{R}^{n\times n}$ be such that $A \geq\geq 0$. Then, for all $\alpha \geq 1$,

$$\operatorname{sprad}\Big(A^{\{\alpha\}}\Big) \leq [\operatorname{sprad}(A)]^{\alpha}$$

and, for all $\alpha \leq 1$,

$$[\operatorname{sprad}(A)]^{\alpha} \leq \operatorname{sprad}\Big(A^{\{\alpha\}}\Big).$$

Furthermore,

$$\operatorname{sprad}\Big(A^{\{1/2\}} \circ A^{\mathrm{T}\{1/2\}}\Big) \leq \operatorname{sprad}(A)$$

and

$$[\operatorname{sprad}(A \circ A)]^{1/2} \leq \operatorname{sprad}(A).$$

If, in addition, $B \in \mathbb{R}^{n\times n}$ is such that $B \geq\geq 0$, then

$$\operatorname{sprad}(A \circ B) \leq [\operatorname{sprad}(A \circ A)\operatorname{sprad}(B \circ B)]^{1/2} \leq \operatorname{sprad}(A)\operatorname{sprad}(B)$$

and

$$\operatorname{sprad}\Big(A^{\{1/2\}} \circ B^{\{1/2\}}\Big) \leq \sqrt{\operatorname{sprad}(A)\operatorname{sprad}(B)}.$$

If, in addition, $A >> 0$ and $B >> 0$, then

$$\operatorname{sprad}(A \circ B) < \operatorname{sprad}(A)\operatorname{sprad}(B).$$

(Proof: See [187, 322].)

7.6 Notes

A history of the Kronecker product is given in [275]. Kronecker matrix algebra is discussed in [111,242,276,388,412,518,575]. Applications to signal processing are considered in [479].

The fact that the Schur product is a principal submatrix of the Kronecker product is noted in [394]. A variation of Kronecker matrix algebra for symmetric matrices can be developed in terms of the half-vectorization operator “vech” and associated elimination and duplication matrices [276, 387, 559].

Generalizations of the Schur and Kronecker products, known as the block-Kronecker, Khatri-Rao, and Tracy-Singh products, are discussed in [292, 303, 338, 377]. Another related operation is the *bialternate product*, which is a variation of the compound operation discussed in Fact 7.4.32. See [217, 239], [319, pp. 313–320], and [384, pp. 84, 85]. The Schur product is also called the Hadamard product.

Chapter Eight

Nonnegative-Semidefinite Matrices

In this chapter we focus on nonnegative-semidefinite and positive-definite matrices. These matrices arise in a variety of applications, such as covariance analysis in signal processing and controllability analysis in linear system theory, and they have many special properties.

8.1 Nonnegative-Semidefinite and Positive-Definite Orderings

Let $A \in \mathbb{F}^{n\times n}$ be a Hermitian matrix. As shown in Corollary 5.4.5, A is unitarily similar to a real diagonal matrix whose diagonal entries are the eigenvalues of A. We denote these eigenvalues by $\lambda_1, \ldots, \lambda_n$ or, for clarity, by $\lambda_1(A), \ldots, \lambda_n(A)$. As in Chapter 3, we employ the convention

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \tag{8.1.1}$$

and, for convenience, we define

$$\lambda_{\max}(A) \triangleq \lambda_1, \quad \lambda_{\min}(A) \triangleq \lambda_n. \tag{8.1.2}$$

Then, A is nonnegative semidefinite if and only if $\lambda_{\min}(A) \geq 0$, while A is positive definite if and only if $\lambda_{\min}(A) > 0$.

For convenience, let $\mathbf{H}^n, \mathbf{N}^n$, and $\mathbf{P}^n$ denote, respectively, the Hermitian, nonnegative-semidefinite, and positive-definite matrices in $\mathbb{F}^{n\times n}$. Hence, $\mathbf{P}^n \subset \mathbf{N}^n \subset \mathbf{H}^n$. If $A \in \mathbf{N}^n$, then we write $A \geq 0$, while if $A \in \mathbf{P}^n$, then we write $A > 0$. If $A, B \in \mathbf{H}^n$, then $A - B \in \mathbf{N}^n$ is possible even if neither A nor B is nonnegative semidefinite. In this case, we write $A \geq B$ or $B \leq A$. Similarly, $A - B \in \mathbf{P}^n$ is denoted by $A > B$ or $B < A$. This notation is consistent with the case $n = 1$, where $\mathbf{H}^1 = \mathbb{R}$, $\mathbf{N}^1 = [0, \infty)$, and $\mathbf{P}^1 = (0, \infty)$.

Note that, since $0 \in \mathbf{N}^n$, it follows that $\mathbf{N}^n$ is a pointed cone. Furthermore, if $A, -A \in \mathbf{N}^n$, then $x^*Ax = 0$ for all $x \in \mathbb{F}^n$, which implies that

$A = 0$. Hence, $\mathbf{N}^n$ is a one-sided cone. Finally, $\mathbf{N}^n$ and $\mathbf{P}^n$ are convex cones since, if $A, B \in \mathbf{N}^n$, then $\alpha A + \beta B \in \mathbf{N}^n$ for all $\alpha, \beta > 0$ and likewise for $\mathbf{P}^n$. The following result shows that the relation "$\leq$" is a partial ordering on $\mathbf{H}^n$.

Proposition 8.1.1. The relation "$\leq$" is reflexive, antisymmetric, and transitive on $\mathbf{H}^n$, that is, if $A, B, C \in \mathbf{H}^n$, then the following statements hold:

i) $A \leq A$.

ii) If $A \leq B$ and $B \leq A$, then $A = B$.

iii) If $A \leq B$ and $B \leq C$, then $A \leq C$.

Proof. Since $\mathbf{N}^n$ is a pointed, one-sided, and convex cone, it follows from Proposition 2.3.6 that the relation "$\leq$" is reflexive, antisymmetric, and transitive. □

Additional properties of "$\leq$" and "$<$" are given by the following result.

Proposition 8.1.2. Let $A, B, C, D \in \mathbf{H}^n$. Then, the following statements hold:

i) If $A \geq 0$, then $\alpha A \geq 0$ for all $\alpha \geq 0$, and $\alpha A \leq 0$ for all $\alpha \leq 0$.

ii) If $A > 0$, then $\alpha A > 0$ for all $\alpha > 0$, and $\alpha A < 0$ for all $\alpha < 0$.

iii) If $A \geq 0$ and $B \geq 0$, then $\alpha A + \beta B \geq 0$ for all $\alpha, \beta \geq 0$.

iv) If $A \geq 0$ and $B > 0$, then $A + B > 0$.

v) $A^2 \geq 0$.

vi) $A^2 > 0$ if and only if $\det A \neq 0$.

vii) If $A \leq B$ and $B < C$, then $A < C$.

viii) If $A < B$ and $B \leq C$, then $A < C$.

ix) If $A \leq B$ and $C \leq D$, then $A + C \leq B + D$.

x) If $A \leq B$ and $C < D$, then $A + C < B + D$.

Furthermore, let $S \in \mathbb{F}^{m \times n}$. Then, the following statements hold:

xi) If $A \leq B$, then $SAS^* \leq SBS^*$.

xii) If $A < B$ and $\operatorname{rank} S = m$, then $SAS^* < SBS^*$.

xiii) If $SAS^* \leq SBS^*$ and $\operatorname{rank} S = n$, then $A \leq B$.

xiv) If $SAS^* < SBS^*$ and $\operatorname{rank} S = n$, then $m = n$ and $A < B$.

Proof. Results $i) - xi)$ are immediate. To prove $xii)$ note that $A < B$ implies that $(B-A)^{1/2}$ is positive definite. Thus, $\operatorname{rank} S(A-B)^{1/2} = m$, which implies that $S(A-B)S^*$ is positive definite. To prove $xiii)$ note that, since $\operatorname{rank} S = n$, it follows that S has a left inverse $S^{\mathrm{L}} \in \mathbb{F}^{n\times m}$. Thus, $xi)$ implies that $A = S^{\mathrm{L}}SAS^*S^{\mathrm{L}*} \le S^{\mathrm{L}}SBS^*S^{\mathrm{L}*} = B$. To prove $xiv)$, note that, since $S(B-A)S^*$ is positive definite, it follows that $\operatorname{rank} S = m$. Hence, $m = n$ and S is nonsingular. Thus, $xii)$ implies that $A = S^{-1}SAS^*S^{-*} < S^{-1}SBS^*S^{-*} = B$. □

The following result is an immediate consequence of Corollary 5.4.7.

Corollary 8.1.3. Let $A, B \in \mathbf{H}^n$ and assume that A and B are congruent. Then, A is nonnegative semidefinite if and only if B is nonnegative semidefinite. Furthermore, A is positive definite if and only if B is positive definite.

Lemma 8.1.4. Let $A \in \mathbf{P}^n$. If $A \le I$, then $A^{-1} \ge I$. Furthermore, if $A < I$, then $A^{-1} > I$.

Proof. Since $A \le I$, it follows from $xi)$ of Proposition 8.1.2 that $I = A^{-1/2}AA^{-1/2} \le A^{-1/2}IA^{-1/2} = A^{-1}$. Similarly, $A < I$ implies that $I = A^{-1/2}AA^{-1/2} < A^{-1/2}IA^{-1/2} = A^{-1}$. □

Proposition 8.1.5. Let $A, B \in \mathbf{H}^n$ be both positive definite or both negative definite. If $A \le B$, then $B^{-1} \le A^{-1}$. If, in addition, $A < B$, then $B^{-1} < A^{-1}$.

8.2 Submatrices

We first consider some identities involving a partitioned nonnegative-semi-definite matrix.

Lemma 8.2.1. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^{\mathrm{T}} & A_{22} \end{bmatrix} \in \mathbf{N}^{n+m}$. Then,

$$A_{12} = A_{11}A_{11}^{+}A_{12}, \tag{8.2.1}$$

$$A_{12} = A_{12}A_{22}A_{22}^{+}. \tag{8.2.2}$$

Proof. Since $A \ge 0$, it follows from Corollary 5.4.5 that $A = BB^*$, where $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \in \mathbb{F}^{(n+m)\times r}$ and $r \triangleq \operatorname{rank} A$. Thus, $A_{11} = B_1B_1^*$, $A_{12} = B_1B_2^*$, and $A_{22} = B_2B_2^*$. Since A_{11} is Hermitian, it follows that A_{11}^{+} is also Hermitian. Next, defining $S \triangleq B_1 - B_1B_1^*(B_1B_1^*)^{+}B_1$, it follows that $SS^* = 0$ and thus $\operatorname{tr} SS^* = 0$. Hence, Lemma 2.2.3 implies that $S = 0$, and thus $B_1 = B_1B_1^*(B_1B_1^*)^{+}B_1$. Consequently, $B_1B_2^* = B_1B_1^*(B_1B_1^*)^{+}B_1B_2$, that is, $A_{12} = A_{11}A_{11}^{+}A_{12}$. The second result is analogous. □

Corollary 8.2.2. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbf{N}^{n+m}$. Then, the following statements hold:

i) $\mathcal{R}(A_{12}) \subseteq \mathcal{R}(A_{11})$.

ii) $\mathcal{R}(A_{12}^*) \subseteq \mathcal{R}(A_{22})$.

iii) $\operatorname{rank}\begin{bmatrix} A_{11} & A_{12} \end{bmatrix} = \operatorname{rank} A_{11}$.

iv) $\operatorname{rank}\begin{bmatrix} A_{12}^* & A_{22} \end{bmatrix} = \operatorname{rank} A_{22}$.

Proof. Results *i)* and *ii)* follow from (8.2.1) and (8.2.2), while *iii)* and *iv)* are consequences of *i)* and *ii)*. □

Next, if (8.2.1) holds, then the partitioned matrix $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$ can be factored as

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ A_{12}^*A_{11}^{+} & I \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & A_{11}|A \end{bmatrix} \begin{bmatrix} I & A_{11}^{+}A_{12} \\ 0 & I \end{bmatrix}, \tag{8.2.3}$$

while if (8.2.2) holds, then

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} = \begin{bmatrix} I & A_{12}A_{22}^{+} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{22}|A & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ A_{22}^{+}A_{12}^* & I \end{bmatrix}, \tag{8.2.4}$$

where

$$A_{11}|A = A_{22} - A_{12}^*A_{11}^{+}A_{12} \tag{8.2.5}$$

and

$$A_{22}|A = A_{11} - A_{12}A_{22}^{+}A_{12}^*. \tag{8.2.6}$$

Hence, it follows from Lemma 8.2.1 that, if A is nonnegative semidefinite, then (8.2.3) and (8.2.4) are valid, and, furthermore, the Schur complements $A_{11}|A$ and $A_{22}|A$ are both nonnegative semidefinite. Consequently, we have the following result.

Proposition 8.2.3. Let $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbf{H}^{n+m}$. Then, the following statements are equivalent:

i) $A \geq 0$.

ii) $A_{11} \geq 0, A_{12} = A_{11}A_{11}^{+}A_{12}$, and $A_{12}^*A_{11}^{+}A_{12} \leq A_{22}$.

iii) $A_{22} \geq 0$, $A_{12} = A_{12}A_{22}A_{22}^{+}$, and $A_{12}A_{22}^{+}A_{12}^* \leq A_{11}$.

The following statements are also equivalent:

iv) $A > 0$.

v) $A_{11} > 0$ and $A_{12}^*A_{11}^{-1}A_{12} < A_{22}$.

vi) $A_{22} > 0$ and $A_{12}A_{22}^{-1}A_{12}^* < A_{11}$.

The following result follows from (2.8.16) and (2.8.17).

Proposition 8.2.4. Let $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbf{P}^{n+m}$. Then,

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{11}|A)^{-1}A_{12}^*A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{11}|A)^{-1} \\ -(A_{11}|A)^{-1}A_{12}^*A_{11}^{-1} & (A_{11}|A)^{-1} \end{bmatrix} \tag{8.2.7}$$

and

$$A^{-1} = \begin{bmatrix} (A_{22}|A)^{-1} & -(A_{22}|A)^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{12}^*(A_{22}|A)^{-1} & A_{22}^{-1}A_{12}^*(A_{22}|A)^{-1}A_{12}A_{22}^{-1} + A_{22}^{-1} \end{bmatrix}, \tag{8.2.8}$$

where

$$A_{11}|A = A_{22} - A_{12}^*A_{11}^{-1}A_{12} \tag{8.2.9}$$

and

$$A_{22}|A = A_{11} - A_{12}A_{22}^{-1}A_{12}^*. \tag{8.2.10}$$

Now, let $A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix}$. Then,

$$B_{11}|A^{-1} = A_{22}^{-1} \tag{8.2.11}$$

and

$$B_{22}|A^{-1} = A_{11}^{-1}. \tag{8.2.12}$$

Lemma 8.2.5. Let $A \in \mathbb{F}^{n\times n}$, $b \in \mathbb{F}^n$, and $a \in \mathbb{R}$. Then, $B \triangleq \begin{bmatrix} A & b \\ b^* & a \end{bmatrix}$ is nonnegative semidefinite if and only if A is nonnegative semidefinite, $b = AA^+b$, and $b^*A^+b \le a$. Furthermore, B is positive definite if and only if A is positive definite and $b^*A^{-1}b < a$. In this case,

$$\det B = (\det A)\big(a - b^*A^{-1}b\big). \tag{8.2.13}$$

For the following result note that a matrix is a principal submatrix of itself and the determinant of a matrix is also a principal subdeterminant.

Proposition 8.2.6. Let $A \in \mathbf{H}^n$. Then, the following statements are equivalent:

i) A is nonnegative semidefinite.

ii) Every principal submatrix of A is nonnegative semidefinite.

iii) Every principal subdeterminant of A is nonnegative.

iv) For all $i = 1, \ldots, n$, the sum of all $i \times i$ principal subdeterminants of A is nonnegative.

Proof. To prove *i)* $\implies$ *ii)*, let $\hat{A} \in \mathbb{F}^{m\times m}$ be the principal submatrix of A obtained from A by retaining rows and columns $i_1, \ldots, i_m$. Then, $\hat{A} = S^{\mathrm{T}}AS$, where $S \triangleq \begin{bmatrix} e_{i_1} & \cdots & e_{i_m} \end{bmatrix} \in \mathbb{R}^{n\times m}$. Now, let $\hat{x} \in \mathbb{F}^m$. Since A is nonnegative semidefinite, it follows that $\hat{x}^*\hat{A}\hat{x} = \hat{x}^*S^{\mathrm{T}}AS\hat{x} \geq 0$, and thus $\hat{A}$ is nonnegative semidefinite.

Next, the implications *ii)* $\implies$ *iii)* $\implies$ *iv)* are immediate. To prove *iv)* $\implies$ *i)*, note that it follows from Proposition 4.4.5 that

$$\chi_A(s) = \sum_{i=0}^{n} \beta_i s^i = \sum_{i=0}^{n} (-1)^{n-i}\gamma_{n-i}s^i = (-1)^n \sum_{i=0}^{n} \gamma_{n-i}(-s)^i, \tag{8.2.14}$$

where, for all $i = 1, \ldots, n$, γ_i is the sum of all $i \times i$ principal subdeterminants of A, and $\beta_n = \gamma_0 = 1$. By assumption, $\gamma_i \geq 0$ for all $i = 1, \ldots, n$. Now, suppose that there exists $\lambda \in \operatorname{spec}(A)$ such that $\lambda < 0$. Then, $0 = (-1)^n\chi_A(\lambda) = \sum_{i=0}^{n} \gamma_{n-i}(-\lambda)^i > 0$, which is a contradiction. □

Proposition 8.2.7. Let $A \in \mathbf{H}^n$. Then, the following statements are equivalent:

i) A is positive definite.

ii) Every principal submatrix of A is positive definite.

iii) Every principal subdeterminant of A is positive.

iv) Every leading principal submatrix of A is positive definite.

v) Every leading principal subdeterminant of A is positive.

Proof. To prove *i)* $\implies$ *ii)*, let $\hat{A} \in \mathbb{F}^{m\times m}$ and S be as in the proof of Proposition 8.2.6 and let $\hat{x}$ be nonzero so that $S\hat{x}$ is nonzero. Since A is positive definite, it follows that $\hat{x}^*\hat{A}\hat{x} = \hat{x}^*S^{\mathrm{T}}AS\hat{x} > 0$ and hence $\hat{A}$ is positive definite.

Next, the implications *i)* $\implies$ *ii)* $\implies$ *iii)* $\implies$ *v)* and *ii)* $\implies$ *iv)* $\implies$ *v)* are immediate. To prove *v)* $\implies$ *i)*, suppose that the leading principal submatrix $A_i \in \mathbb{F}^{i\times i}$ has positive determinant for all $i = 1, \ldots, n$,. The result is true for $n = 1$. For $n \geq 2$, we show that if A_i is positive definite, then so is A_{i+1}. Writing $A_{i+1} = \begin{bmatrix} A_i & b_i \\ b_i^* & a_i \end{bmatrix}$, it follows from Lemma 8.2.5 that $\det A_{i+1} = (\det A_i)\big(a_i - b_i^*A_i^{-1}b_i\big) > 0$ and hence $a_i - b_i^*A_i^{-1}b_i = \det A_{i+1}/\det A_i > 0$. Lemma 8.2.5 now implies that A_{i+1} is positive definite. Using this argument for all $i = 2, \ldots, n$ implies that A is positive definite. □

The example $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ shows that every principal subdeterminant of A, rather than just the leading principal subdeterminants of A, must be checked to determine whether A is nonnegative semidefinite. A less obvious example is $A = \begin{bmatrix} 1&1&1\\ 1&1&1\\ 1&1&0 \end{bmatrix}$, whose eigenvalues are 0, $1+\sqrt{3}$, and $1-\sqrt{3}$. In this case, the principal subdeterminant $\det A_{[1,1]} = \det \begin{bmatrix} 1&1\\ 1&0 \end{bmatrix} < 0$.

Corollary 8.2.8. Let $A \in \mathbf{N}^n$. Then, every diagonally located square submatrix of A is nonnegative semidefinite. If, in addition, A is positive definite, then every diagonally located square submatrix of A is positive definite.

8.3 Simultaneous Diagonalization

This section considers the simultaneous diagonalization of a pair of matrices $A, B \in \mathbf{H}^n$. There are two types of simultaneous diagonalization. *Cogredient diagonalization* involves a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and SBS^* are both diagonal, whereas *contragredient diagonalization* involves finding a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and $S^{-*}BS^{-1}$ are both diagonal. Both types of simultaneous transformation involve only congruence transformations. We begin by assuming that one of the matrices is positive definite, in which case the results are quite simple to prove. Our first result involves cogredient diagonalization.

Theorem 8.3.1. Let $A, B \in \mathbf{H}^n$ and assume that A is positive definite. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $SAS^* = I$ and SBS^* is diagonal.

Proof. Setting $S_1 = A^{-1/2}$ it follows that $S_1AS_1^* = I$. Now, since $S_1BS_1^*$ is Hermitian, it follows from Corollary 5.4.5 that there exists a unitary matrix $S_2 \in \mathbb{F}^{n\times n}$ such that $SBS^* = S_2S_1BS_1^*S_2^*$ is diagonal, where $S = S_2S_1$. Finally, $SAS^* = S_2S_1AS_1^*S_2^* = S_2IS_2^* = I$. □

An analogous result holds for contragedient diagonalization.

Theorem 8.3.2. Let $A, B \in \mathbf{H}^n$, and assume that A is positive definite. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $SAS^* = I$ and $S^{-*}BS^{-1}$ is diagonal.

Proof. Setting $S_1 = A^{-1/2}$ it follows that $S_1AS_1^* = I$. Since $S_1^{-*}BS_1^{-1}$ is Hermitian, it follows that there exists a unitary matrix $S_2 \in \mathbb{F}^{n\times n}$ such that $S^{-*}BS^{-1} = S_2^{-*}S_1^{-*}BS_1^{-1}S_2^{-1} = S_2\big(S_1^{-*}BS_1^{-1}\big)S_2^*$ is diagonal, where $S = S_2S_1$. Finally, $SAS^* = S_2S_1AS_1^*S_2^* = S_2IS_2^* = I$. □

Corollary 8.3.3. Let $A, B \in \mathbf{P}^n$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and $S^{-*}BS^{-1}$ are equal and diagonal.

Proof. By Theorem 8.3.2 there exists a nonsingular matrix $S_1 \in \mathbb{F}^{n\times n}$ such that $S_1AS_1^* = I$ and $B_1 = S_1^{-*}BS_1^{-1}$ is diagonal. Defining $S \triangleq B_1^{1/4}S_1$ yields $SAS^* = S^{-*}BS^{-1} = B_1^{1/2}$. □

The transformation S of Corollary 8.3.3 is a *balancing transformation.*

Next, we weaken the requirement in Theorem 8.3.1 and Theorem 8.3.2 that A be positive definite by assuming only that A is nonnegative semidefinite. In this case, however, we assume that B is also nonnegative semidefinite.

Theorem 8.3.4. Let $A, B \in \mathbf{N}^n$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $SAS^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and SBS^* is diagonal.

Proof. Let the nonsingular matrix $S_1 \in \mathbb{F}^{n\times n}$ be such that $S_1AS_1^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and similarly partition $S_1BS_1^* = \left[\begin{smallmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{smallmatrix}\right]$, which is nonnegative semidefinite. Letting $S_2 \triangleq \left[\begin{smallmatrix} I & -B_{12}B_{22}^+ \\ 0 & I \end{smallmatrix}\right]$ it follows from Lemma 8.2.1 that

$$S_2S_1BS_1^*S_2^* = \begin{bmatrix} B_{11} - B_{12}B_{22}^+B_{12}^* & 0 \\ 0 & B_{22} \end{bmatrix}.$$

Next, let U_1 and U_2 be unitary matrices such that $U_1(B_{11} - B_{12}B_{22}^+B_{12}^*)U_1^*$ and $U_2B_{22}U_2^*$ are diagonal. Then, defining $S_3 \triangleq \left[\begin{smallmatrix} U_1 & 0 \\ 0 & U_2 \end{smallmatrix}\right]$ and $S \triangleq S_3S_2S_1$, it follows that $SAS^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $SBS^* = S_3S_2S_1BS_1^*S_2^*S_3^*$ is diagonal. □

Theorem 8.3.5. Let $A, B \in \mathbf{N}^n$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $SAS^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $S^{-*}BS^{-1}$ is diagonal.

Proof. Let $S_1 \in \mathbb{F}^{n\times n}$ be a nonsingular matrix such that $S_1AS_1^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and similarly partition $S_1^{-*}BS_1^{-1} = \left[\begin{smallmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{smallmatrix}\right]$, which is nonnegative semidefinite. Letting $S_2 \triangleq \left[\begin{smallmatrix} I & B_{11}^+B_{12} \\ 0 & I \end{smallmatrix}\right]$, it follows that

$$S_2^{-*}S_1^{-*}BS_1^{-1}S_2^{-1} = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} - B_{12}^*B_{11}^+B_{12} \end{bmatrix}.$$

Now, let U_1 and U_2 be unitary matrices such that $U_1B_{11}U_1^*$ and $U_2(B_{22} - B_{12}^*B_{11}^+B_{12})U_2^*$ are diagonal. Then, defining $S_3 \triangleq \left[\begin{smallmatrix} U_1 & 0 \\ 0 & U_2 \end{smallmatrix}\right]$ and $S \triangleq S_3S_2S_1$, it follows that $SAS^* = \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $S^{-*}BS^{-1} = S_3^{-*}S_2^{-*}S_1^{-*}BS_1^{-1}S_2^{-1}S_3^{-1}$ is diagonal. □

Corollary 8.3.6. Let $A, B \in \mathbf{N}^n$. Then, AB is semisimple, and every eigenvalue of AB is nonnegative. If, in addition, A and B are positive definite, then every eigenvalue of AB is positive.

Proof. It follows from Theorem 8.3.5 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $A_1 = SAS^*$ and $B_1 = S^{-*}BS^{-1}$ are diagonal with nonnegative diagonal entries. Hence, $AB = S^{-1}A_1B_1S$ is semisimple and has nonnegative eigenvalues. □

A more direct approach to showing that AB has nonnegative eigenvalues is to use Corollary 4.4.10 and note that $\lambda_i(AB) = \lambda_i\big(B^{1/2}AB^{1/2}\big) \geq 0$.

Corollary 8.3.7. Let $A, B \in \mathbf{N}^n$ and assume that $\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} AB$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $SAS^* = S^{-*}BS^{-1}$ and such that SAS^* is diagonal.

Proof. By Theorem 8.3.5 there exists a nonsingular matrix $S_1 \in \mathbb{F}^{n\times n}$ such that $S_1AS_1^* = \left[\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right]$, where $r \triangleq \operatorname{rank} A$, and such that $B_1 = S_1^{-*}BS_1^{-1}$ is diagonal. Hence, $AB = S_1^{-1}\left[\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right]B_1S_1$. Since $\operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} AB = r$, it follows that $B_1 = \left[\begin{smallmatrix} \hat{B}_1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, where $\hat{B}_1 \in \mathbb{F}^{r\times r}$ is positive diagonal. Hence, $S_1^{-*}BS_1^{-1} = \left[\begin{smallmatrix} \hat{B}_1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Now, define $S_2 \triangleq \left[\begin{smallmatrix} \hat{B}_1^{1/4} & 0 \\ 0 & I \end{smallmatrix}\right]$ and $S \triangleq S_2S_1$. Then, $SAS^* = S_2S_1AS_1^*S_2^* = \left[\begin{smallmatrix} \hat{B}_1^{1/2} & 0 \\ 0 & 0 \end{smallmatrix}\right] = S_2^{-*}S_1^{-*}BS_1^{-1}S_2^{-1} = S^{-*}BS^{-1}$. □

8.4 Eigenvalue Inequalities

Next, we turn our attention to inequalities involving eigenvalues. We begin with a series of lemmas.

Lemma 8.4.1. Let $A \in \mathbf{H}^n$ and let $\beta \in \mathbb{R}$. Then, the following statements hold:

i) $\beta I \leq A$ if and only if $\beta \leq \lambda_{\min}(A)$.

ii) $\beta I < A$ if and only if $\beta < \lambda_{\min}(A)$.

iii) $A \leq \beta I$ if and only if $\lambda_{\max}(A) \leq \beta$.

iv) $A < \beta I$ if and only if $\lambda_{\max}(A) < \beta$.

Proof. To prove *i)* assume that $\beta I \leq A$, and let $S \in \mathbb{F}^{n\times n}$ be a unitary matrix such that $B = SAS^*$ is diagonal. Then, $\beta I \leq B$, which yields $\beta \leq \lambda_{\min}(B) = \lambda_{\min}(A)$. Conversely, let $S \in \mathbb{F}^{n\times n}$ be a unitary matrix such that $B = SAS^*$ is diagonal. Since the diagonal entries of B are the eigenvalues of A, it follows that $\lambda_{\min}(A)I \leq B$, which implies that $\beta I \leq \lambda_{\min}(A)I \leq S^*BS = A$. Results *ii)*, *iii)* and *iv)* are proved in a similar manner. □

Corollary 8.4.2. Let $A \in \mathbf{H}^n$. Then,

$$\lambda_{\min}(A)I \le A \le \lambda_{\max}(A)I. \tag{8.4.1}$$

Proof. The result follows from *i*) and *ii*) of Lemma 8.4.1 with $\beta = \lambda_{\min}(A)$ and $\beta = \lambda_{\max}(A)$, respectively. □

Lemma 8.4.3. Let $A \in \mathbf{H}^n$. Then,

$$\lambda_{\min}(A) = \min_{x \in \mathbb{F}^n \setminus \{0\}} \frac{x^*Ax}{x^*x} \tag{8.4.2}$$

and

$$\lambda_{\max}(A) = \max_{x \in \mathbb{F}^n \setminus \{0\}} \frac{x^*Ax}{x^*x}. \tag{8.4.3}$$

Proof. It follows from (8.4.1) that $\lambda_{\min}(A) \le x^*Ax/x^*x$ for all nonzero $x \in \mathbb{F}^n$. Letting $x \in \mathbb{F}^n$ be an eigenvector of A associated with $\lambda_{\min}(A)$, it follows that this lower bound is attained. This proves (8.4.2). An analogous argument yields (8.4.3). □

The following result is the *Cauchy interlacing theorem*.

Lemma 8.4.4. Let $A \in \mathbf{H}^n$ and let A_0 be an $(n-1)\times(n-1)$ principal submatrix of A. Then, for all $i = 1, \ldots, n-1$,

$$\lambda_{i+1}(A) \le \lambda_i(A_0) \le \lambda_i(A). \tag{8.4.4}$$

Proof. Suppose that the chain of inequalities (8.4.4) does not hold. In particular, first suppose that the right-most inequality in (8.4.4) that is not true is $\lambda_i(A_0) \le \lambda_i(A)$, so that $\lambda_i(A) < \lambda_i(A_0)$. Choose δ such that $\lambda_i(A) < \delta < \lambda_i(A_0)$ and such that δ is not an eigenvalue of A_0. If $i = 1$, then $A - \delta I$ is negative definite, while if $i \ge 2$, then $\lambda_i(A) < \delta < \lambda_i(A_0) \le \lambda_{i-1}(A_0) \le \lambda_{i-1}(A)$, so that $A - \delta I$ has $i-1$ positive eigenvalues. Thus, $\nu_+(A - \delta I) = i - 1$. Furthermore, since $\delta < \lambda_i(A_0)$, it follows that $\nu_+(A_0 - \delta I) \ge i$.

Now, assume for convenience that the rows and columns of A are ordered so that A_0 is the $(n-1)\times(n-1)$ leading principal submatrix of A. Thus, $A = \begin{bmatrix} A_0 & \beta \\ \beta^* & \gamma \end{bmatrix}$, where $\beta \in \mathbb{F}^{n-1}$ and $\gamma \in \mathbb{F}$. Next, note the identity

$$\begin{aligned} &A - \delta I \\ &= \begin{bmatrix} I & 0 \\ \beta^*(A_0-\delta I)^{-1} & 1 \end{bmatrix} \begin{bmatrix} A_0 - \delta I & 0 \\ 0 & \gamma - \delta - \beta^*(A_0-\delta I)^{-1}\beta \end{bmatrix} \begin{bmatrix} I & (A_0-\delta I)^{-1}\beta \\ 0 & 1 \end{bmatrix}, \end{aligned} \tag{8.4.5}$$

where $A_0 - \delta I$ is nonsingular since δ was chosen to not be an eigenvalue of A_0. Since the right-hand side of this identity involves a congruence transformation and, since $\nu_+(A_0 - \delta I) \ge i$, it follows from Corollary 5.4.7 that $\nu_+(A - \delta I) \ge i$. However, this contradicts the fact that $\nu_+(A - \delta I) = i - 1$.

Finally, suppose that the right-most inequality in (8.4.4) that is not true is $\lambda_{i+1}(A) \le \lambda_i(A_0)$, so that $\lambda_i(A_0) < \lambda_{i+1}(A)$. Choose δ such that $\lambda_i(A_0) < \delta < \lambda_{i+1}(A)$ and such that δ is not an eigenvalue of A_0. Then, it follows that $\nu_+(A - \delta I) \ge i + 1$ and $\nu_+(A_0 - \delta I) = i - 1$. Using the congruence transformation as in the previous case, it follows that $\nu_+(A - \delta I) \le i$, which contradicts the fact that $\nu_+(A - \delta I) \ge i + 1$. □

The following result is the *inclusion principle*.

Theorem 8.4.5. Let $A \in \mathbf{H}^n$ and let $A_0 \in \mathbf{H}^k$ be a $k \times k$ principal submatrix of A. Then, for all $i = 1, \ldots, k$,

$$\lambda_{i+n-k}(A) \le \lambda_i(A_0) \le \lambda_i(A). \tag{8.4.6}$$

Proof. If $k = n - 1$, then the result is given by Lemma 8.4.4. Hence, let $k = n - 2$, and let A_1 denote an $(n-1) \times (n-1)$ principal submatrix of A such that the $(n-2) \times (n-2)$ principal submatrix A_0 of A is also a principal submatrix of A_1. Therefore, Lemma 8.4.4 implies that $\lambda_n(A) \le \lambda_{n-1}(A_1) \le \cdots \le \lambda_2(A_1) \le \lambda_2(A) \le \lambda_1(A_1) \le \lambda_1(A)$ and $\lambda_{n-1}(A_1) \le \lambda_{n-2}(A_0) \le \cdots \le \lambda_2(A_0) \le \lambda_2(A_1) \le \lambda_1(A_0) \le \lambda_1(A_1)$. Combining these inequalities yields $\lambda_{i+2}(A) \le \lambda_i(A_0) \le \lambda_i(A)$ for all $i = 1, \ldots, n-2$, while proceeding in a similar manner with $k < n - 2$ yields (8.4.6). □

Corollary 8.4.6. Let $A \in \mathbf{H}^n$ and let $A_0 \in \mathbf{H}^k$ be a $k \times k$ principal submatrix of A. Then,

$$\lambda_{\min}(A) \le \lambda_{\min}(A_0) \le \lambda_{\max}(A_0) \le \lambda_{\max}(A) \tag{8.4.7}$$

and

$$\lambda_{\min}(A_0) \le \lambda_k(A). \tag{8.4.8}$$

Corollary 8.4.7. Let $A \in \mathbf{H}^n$. Then,

$$\lambda_{\min}(A) \le \mathrm{d}_{\min}(A) \le \mathrm{d}_{\max}(A) \le \lambda_{\max}(A). \tag{8.4.9}$$

Lemma 8.4.8. Let $A, B \in \mathbf{H}^n$, and assume that $A \le B$ and $\mathrm{mspec}(A) = \mathrm{mspec}(B)$. Then, $A = B$.

Proof. Let $\alpha \ge 0$ be such that $0 < \hat{A} \le \hat{B}$, where $\hat{A} \triangleq A + \alpha I$ and $\hat{B} \triangleq B + \alpha I$. Note that $\mathrm{mspec}(\hat{A}) = \mathrm{mspec}(\hat{B})$ and thus $\det \hat{A} = \det \hat{B}$. Next, it follows that $I \le \hat{A}^{-1/2}\hat{B}\hat{A}^{-1/2}$. Hence, it follows from *i*) of Lemma 8.4.1 that $\lambda_{\min}\Big(\hat{A}^{-1/2}\hat{B}\hat{A}^{-1/2}\Big) \ge 1$. Furthermore, $\det\Big(\hat{A}^{-1/2}\hat{B}\hat{A}^{-1/2}\Big) = \det \hat{B}/\det \hat{A} = 1$, which implies that $\lambda_i(\hat{A}^{-1/2}\hat{B}\hat{A}^{-1/2}) = 1$ for all $i = 1, \ldots, n$. Hence,

$\hat{A}^{-1/2}\hat{B}\hat{A}^{-1/2} = I$ and thus $\hat{A} = \hat{B}$. Hence, $A = B$. □

The following result is the *monotonicity theorem* or *Weyl's inequality.*

Theorem 8.4.9. Let $A, B \in \mathbf{H}^n$, and assume that $A \le B$. Then, for all $i = 1, \ldots, n$,

$$\lambda_i(A) \le \lambda_i(B). \tag{8.4.10}$$

If $A \ne B$, then there exists $i \in \{1, \ldots, n\}$ such that

$$\lambda_i(A) < \lambda_i(B). \tag{8.4.11}$$

If $A < B$, then (8.4.11) holds for all $i = 1, \ldots, n$.

Proof. Since $A \le B$, it follows from Corollary 8.4.2 that $\lambda_{\min}(A)I \le A \le B \le \lambda_{\max}(B)I$. Hence, by *iii*) and *i*) of Lemma 8.4.1 it follows that $\lambda_{\min}(A) \le \lambda_{\min}(B)$ and $\lambda_{\max}(A) \le \lambda_{\max}(B)$. Next, let $S \in \mathbb{F}^{n \times n}$ be a unitary matrix such that $SAS^* = \operatorname{diag}[\lambda_1(A), \ldots, \lambda_n(A)]$. Furthermore, for $2 \le i \le n-1$, let $A_0 = \operatorname{diag}[\lambda_1(A), \ldots, \lambda_i(A)]$ and B_0 denote the $i \times i$ leading principal submatrices of SAS^* and SBS^*, respectively. Since $A \le B$, it follows that $A_0 \le B_0$, which implies that $\lambda_{\min}(A_0) \le \lambda_{\min}(B_0)$. It now follows from (8.4.8) that

$$\lambda_i(A) = \lambda_{\min}(A_0) \le \lambda_{\min}(B_0) \le \lambda_i(SBS^*) = \lambda_i(B),$$

which proves (8.4.10). If $A \ne B$, then it follows from Lemma 8.4.8 that $\operatorname{mspec}(A) \ne \operatorname{mspec}(B)$ and thus there exists $i \in \{1, \ldots, n\}$ such that (8.4.11) holds. If $A < B$, then $\lambda_{\min}(A_0) < \lambda_{\min}(B_0)$, which implies that (8.4.11) holds for all $i = 1, \ldots, n$. □

Corollary 8.4.10. Let $A, B \in \mathbf{H}^n$. Then, the following statements hold:

i) If $A \le B$, then $\operatorname{tr} A \le \operatorname{tr} B$.

ii) If $A \le B$ and $\operatorname{tr} A = \operatorname{tr} B$, then $A = B$.

iii) If $A < B$, then $\operatorname{tr} A < \operatorname{tr} B$.

iv) If $0 \le A \le B$, then $0 \le \det A \le \det B$.

v) If $0 \le A < B$, then $0 \le \det A < \det B$.

vi) If $0 < A \le B$ and $\det A = \det B$, then $A = B$.

Proof. Statements *i*), *iii*), *iv*), *v*) follow from Theorem 8.4.9. To prove *ii*) note that, since $A \le B$ and $\operatorname{tr} A = \operatorname{tr} B$, it follows from Theorem 8.4.9 that $\operatorname{mspec}(A) = \operatorname{mspec}(B)$. Now, Lemma 8.4.8 implies that $A = B$. A similar argument yields *vi*). □

The following result, which is a generalization of Theorem 8.4.9, is due to Weyl.

Theorem 8.4.11. Let $A, B \in \mathbf{H}^n$. If $i + j \geq n + 1$, then

$$\lambda_i(A) + \lambda_j(B) \leq \lambda_{i+j-n}(A + B). \tag{8.4.12}$$

If $i + j \leq n + 1$, then

$$\lambda_{i+j-1}(A + B) \leq \lambda_i(A) + \lambda_j(B). \tag{8.4.13}$$

In particular, for all $i = 1, \ldots, n$,

$$\lambda_i(A) + \lambda_{\min}(B) \leq \lambda_i(A + B) \leq \lambda_i(A) + \lambda_{\max}(B), \tag{8.4.14}$$

$$\lambda_{\min}(A) + \lambda_{\min}(B) \leq \lambda_{\min}(A + B) \leq \lambda_{\min}(A) + \lambda_{\max}(B), \tag{8.4.15}$$

$$\lambda_{\max}(A) + \lambda_{\min}(B) \leq \lambda_{\max}(A + B) \leq \lambda_{\max}(A) + \lambda_{\max}(B). \tag{8.4.16}$$

Proof. See [287, p. 182]. □

Lemma 8.4.12. Let $A, B, C \in \mathbf{H}^n$. If $A \leq B$ and C is nonnegative semidefinite, then

$$\operatorname{tr} AC \leq \operatorname{tr} BC. \tag{8.4.17}$$

If $A < B$ and C is positive definite, then

$$\operatorname{tr} AC < \operatorname{tr} BC. \tag{8.4.18}$$

Proof. Since $C^{1/2}AC^{1/2} \leq C^{1/2}BC^{1/2}$, it follows from *i*) of Corollary 8.4.10 that

$$\operatorname{tr} AC = \operatorname{tr} C^{1/2}AC^{1/2} \leq \operatorname{tr} C^{1/2}BC^{1/2} = \operatorname{tr} BC.$$

Result (8.4.18) follows from *ii*) of Corollary 8.4.10 in a similar fashion. □

Proposition 8.4.13. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that B is nonnegative semidefinite. Then,

$$\tfrac{1}{2}\lambda_{\min}(A + A^*)\operatorname{tr} B \leq \operatorname{tr} AB \leq \tfrac{1}{2}\lambda_{\max}(A + A^*)\operatorname{tr} B. \tag{8.4.19}$$

If, in addition, A is Hermitian, then

$$\lambda_{\min}(A)\operatorname{tr} B \leq \operatorname{tr} AB \leq \lambda_{\max}(A)\operatorname{tr} B. \tag{8.4.20}$$

Proof. It follows from Corollary 8.4.2 that $\frac{1}{2}\lambda_{\min}(A+A^*)I \leq \frac{1}{2}(A+A^*)$, while Lemma 8.4.12 implies that $\frac{1}{2}\lambda_{\min}(A+A^*)\operatorname{tr} B = \operatorname{tr} \frac{1}{2}\lambda_{\min}(A+A^*)IB \leq \operatorname{tr} \frac{1}{2}(A + A^*)B = \operatorname{tr} AB$, which proves the left-hand inequality of (8.4.19). Similarly, the right-hand inequality holds. □

Proposition 8.4.14. Let $A, B \in \mathbf{P}^n$, and assume that $\det B = 1$.

Then,

$$(\det A)^{1/n} \le \tfrac{1}{n}\operatorname{tr} AB. \tag{8.4.21}$$

Furthermore, equality holds if and only if $B = (\det A)^{1/n}A^{-1}$.

Proof. Using the arithmetic-mean-geometric-mean inequality given by Fact 1.4.9 it follows that

$$\begin{aligned}(\det A)^{1/n} = \left(\det B^{1/2}AB^{1/2}\right)^{1/n} &= \left[\prod_{i=1}^{n}\lambda_i\left(B^{1/2}AB^{1/2}\right)\right]^{1/n}\\ &\le \tfrac{1}{n}\sum_{i=1}^{n}\lambda_i\left(B^{1/2}AB^{1/2}\right) = \tfrac{1}{n}\operatorname{tr} AB.\end{aligned}$$

Equality holds if and only if there exists $\beta > 0$ such that $B^{1/2}AB^{1/2} = \beta I$. In this case, $\beta = (\det A)^{1/n}$ and $B = (\det A)^{1/n}A^{-1}$. □

The following corollary of Proposition 8.4.14 is *Minkowski's determinant theorem*.

Corollary 8.4.15. Let $A, B \in \mathbf{N}^n$. Then,

$$\det A + \det B \le \left[(\det A)^{1/n} + (\det B)^{1/n}\right]^n \le \det(A+B). \tag{8.4.22}$$

If $B = 0$ or $\det(A+B) = 0$, then both inequalities become identities. If there exists $\alpha \ge 0$ such that $B = \alpha A$, then the right-hand inequality becomes an identity. Conversely, if $A+B$ is positive definite and the right-hand inequality holds as an identity, then there exists $\alpha \ge 0$ such that either $B = \alpha A$ or $A = \alpha B$. Finally, if A is positive definite and both inequalities hold as identities, then $B = 0$.

Proof. The left-hand inequality is immediate. To prove the right-hand inequality, note that it follows from Proposition 8.4.14 that

$$\begin{aligned}(\det A)^{1/n} + (\det B)^{1/n} &\le \tfrac{1}{n}\operatorname{tr}\left[A[\det(A+B)]^{1/n}(A+B)^{-1}\right]\\ &\quad + \tfrac{1}{n}\operatorname{tr}\left[B[\det(A+B)]^{1/n}(A+B)^{-1}\right]\\ &= [\det(A+B)]^{1/n}.\end{aligned}$$

If $B = 0$ or $\det(A+B) = 0$, then both inequalities become identities, while if there exists $\alpha \ge 0$ such that $B = \alpha A$, then

$$\left[(\det A)^{1/n} + (\det B)^{1/n}\right]^n = (1+\alpha)^n \det A = \det[(1+\alpha)A].$$

Now, suppose that $A+B$ is positive definite and the right-hand inequality holds as an identity. Then, either A or B is positive definite. Hence, suppose that A is positive definite. Multiplying the identity $(\det A)^{1/n} + (\det B)^{1/n} =$

$[\det(A+B)]^{1/n}$ by $(\det A)^{-1/n}$ yields

$$1+\left(\det A^{-1/2}BA^{-1/2}\right)^{1/n}=\left[\det\left(I+A^{-1/2}BA^{-1/2}\right)\right]^{1/n}.$$

Letting $\lambda_1,\ldots,\lambda_n$ denote the eigenvalues of $A^{-1/2}BA^{-1/2}$ it follows that $1+(\lambda_1\cdots\lambda_n)^{1/n}=[(1+\lambda_1)\cdots(1+\lambda_n)]^{1/n}$. It now follows from Fact 1.4.12 that $\lambda_1=\cdots=\lambda_n$. Now, suppose that A is positive definite and both inequalities hold as identities. Then, it follows that $1+\det A^{-1/2}BA^{-1/2}=\det\left(1+A^{-1/2}BA^{-1/2}\right)$, which implies that $1+\lambda_1\cdots\lambda_n=(1+\lambda_1)\cdots(1+\lambda_n)$, where $\lambda_1,\ldots,\lambda_n$ are the eigenvalues of $A^{-1/2}BA^{-1/2}$. Consequently, $B=0$.

Finally, suppose that A is positive definite and both inequalities hold as identities. Since $\det A>0$, it follows from the left-hand identity that $\det B=0$. Hence, the right-hand identity implies that $\det A=\det(A+B)$. Since $A\le A+B$, it follows from v) of Corollary 8.4.10 that $B=0$. □

8.5 Matrix Inequalities

Lemma 8.5.1. Let $A,B\in\mathbf{H}^n$ and assume that $0\le A\le B$. Then, $\mathcal{R}(A)\subseteq\mathcal{R}(B)$.

Proof. Let $x\in\mathcal{N}(B)$. Then, $x^*Bx=0$ and thus $x^*Ax=0$, which implies $Ax=0$. Hence, $\mathcal{N}(B)\subseteq\mathcal{N}(A)$ and thus $\mathcal{N}(A)^\perp\subseteq\mathcal{N}(B)^\perp$. Since A and B are Hermitian, it follows from Theorem 2.4.3 that $\mathcal{R}(A)=\mathcal{N}(A)^\perp$ and $\mathcal{R}(B)=\mathcal{N}(B)^\perp$. Hence, $\mathcal{R}(A)\subseteq\mathcal{R}(B)$. □

The following result is the *Douglas-Fillmore-Williams lemma.*

Theorem 8.5.2. Let $A\in\mathbb{F}^{n\times m}$ and $B\in\mathbb{F}^{n\times l}$. Then, the following statements are equivalent:

i) There exists a matrix $C\in\mathbb{F}^{l\times m}$ such that $A=BC$.

ii) There exists $\alpha>0$ such that $AA^*\le\alpha BB^*$.

iii) $\mathcal{R}(A)\subseteq\mathcal{R}(B)$.

Proof. First we prove that *i*) implies *ii*). Since $A=BC$, it follows that $AA^*=BCC^*B^*$. Since $CC^*\le\lambda_{\max}(CC^*)I$, it follows that $AA^*\le\alpha BB^*$, where $\alpha\triangleq\lambda_{\max}(CC^*)$. To prove that *ii*) implies *iii*), first note that Lemma 8.5.1 implies that $\mathcal{R}(AA^*)\subseteq\mathcal{R}(\alpha BB^*)=\mathcal{R}(BB^*)$. Since, by Theorem 2.4.3, $\mathcal{R}(AA^*)=\mathcal{R}(A)$ and $\mathcal{R}(BB^*)=\mathcal{R}(B)$, it follows that $\mathcal{R}(A)\subseteq\mathcal{R}(B)$. Finally, to prove that *iii*) implies *i*), use Theorem 5.6.3 to write $B=S_1\left[\begin{smallmatrix}D&0\\0&0\end{smallmatrix}\right]S_2$, where $S_1\in\mathbb{F}^{n\times n}$ and $S_2\in\mathbb{F}^{l\times l}$ are unitary and $D\in\mathbb{R}^{r\times r}$ is diagonal with positive diagonal entries, where $r\triangleq\operatorname{rank}B$. Since

$\mathcal{R}(S_1^*A) \subseteq \mathcal{R}(S_1^*B)$ and $S_1^*B = \left[\begin{smallmatrix} D & 0 \\ 0 & 0 \end{smallmatrix}\right] S_2$, it follows that $S_1^*A = \left[\begin{smallmatrix} A_1 \\ 0 \end{smallmatrix}\right]$, where $A_1 \in \mathbb{F}^{r\times m}$. Consequently,

$$A = S_1\begin{bmatrix} A_1 \\ 0 \end{bmatrix} = S_1\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} S_2S_2^*\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} A_1 \\ 0 \end{bmatrix} = BC,$$

where $C \triangleq S_2^*\left[\begin{smallmatrix} D^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right]\left[\begin{smallmatrix} A_1 \\ 0 \end{smallmatrix}\right] \in \mathbb{F}^{l\times m}$. □

Proposition 8.5.3. Let $\{A_i\}_{i=1}^\infty \subset \mathbf{N}^n$ satisfy $0 \le A_i \le A_j$ for all $i \le j$, and assume that there exists $B \in \mathbf{N}^n$ satisfying $A_i \le B$ for all $i \in \mathbb{P}$. Then, $A \triangleq \lim_{i\to\infty} A_i$ exists and satisfies $0 \le A \le B$.

Proof. Let $k \in \{1,\ldots,n\}$. Then, the sequence $\{A_{i(k,k)}\}_{i=1}^\infty$ is nondecreasing and bounded from above. Hence, $A_{(k,k)} \triangleq \lim_{i\to\infty} A_{i(k,k)}$ exists. Now, let $k,l \in \{1,\ldots,n\}$, where $k \ne l$. Since $A_i \le A_j$ for all $i < j$, it follows that $(e_k+e_l)^*A_i(e_k+e_l) \le (e_k+e_l)^*A_j(e_k+e_l)$, which implies that $A_{i(k,l)} - A_{j(k,l)} \le \frac{1}{2}\big[A_{j(k,k)} - A_{i(k,k)} + A_{j(l,l)} - A_{i(l,l)}\big]$. Alternatively, replacing e_k+e_l by e_k-e_l yields $A_{j(k,l)} - A_{i(k,l)} \le \frac{1}{2}\big[A_{j(k,k)} - A_{i(k,k)} + A_{j(l,l)} - A_{i(l,l)}\big]$. Thus, $A_{i(k,l)} - A_{j(k,l)} \to 0$ as $i,j \to \infty$, which implies that $A_{(k,l)} \triangleq \lim_{i\to\infty} A_{i(k,l)}$ exists. Hence, $A \triangleq \lim_{i\to\infty} A_i$ exists. Since $A_i \le B$ for all $i = 1,2,\ldots$, it follows that $A \le B$. □

Let $A = SBS^* \in \mathbb{F}^{n\times n}$ be Hermitian, where $S \in \mathbb{F}^{n\times n}$ is unitary, $B \in \mathbb{R}^{n\times n}$ is diagonal, $\operatorname{spec}(A) \subset \mathcal{D}$, and $\mathcal{D} \subset \mathbb{R}$. Furthermore, let $f\colon\ \mathcal{D} \mapsto \mathbb{R}$. Then, we define $f(A) \in \mathbf{H}^n$ by

$$f(A) \triangleq Sf(B)S^*, \tag{8.5.1}$$

where $[f(B)]_{(i,i)} \triangleq f(B_{(i,i)})$. In particular, suppose that A is nonnegative semidefinite. Then, for all $r \ge 0$ (not necessarily an integer), $A^r = SB^rS^*$ is nonnegative semidefinite, where, for all $i = 1,\ldots,n$, $(B^r)_{(i,i)} = \big(B_{(i,i)}\big)^r$. Note that $A^0 \triangleq I$. In particular, $A^{1/2} = SB^{1/2}S^*$ is a nonnegative-semidefinite *square root* of A since $A^{1/2}A^{1/2} = SB^{1/2}S^*SB^{1/2}S^* = SBS^* = A$. Hence, if $C \in \mathbb{F}^{n\times m}$, then C^*C is nonnegative semidefinite, and we define

$$\langle C\rangle \triangleq \operatorname{tr}(C^*C)^{1/2}. \tag{8.5.2}$$

If A is positive definite, then A^r is positive definite for all $r \in \mathbb{R}$, and, if $r \ne 0$, then $(A^r)^{1/r} = A$. If, in addition, A is positive definite, then $\log A = S(\log B)S^* \in \mathbf{H}^n$, where $(\log B)_{(i,i)} = \log B_{(i,i)}$.

If $0 \le A \le B$, then it does not necessarily follow that $A^2 \le B^2$. Consider $A \triangleq \left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix} 2 & 0 \\ 0 & 8 \end{smallmatrix}\right]$. However, the following result, known as *Furuta's inequality*, is valid.

Proposition 8.5.4. Let $A,B \in \mathbf{N}^n$, and assume that $0 \le A \le B$.

Furthermore, let $p, q, r \in \mathbb{R}$ satisfy $p \geq 0$, $q \geq 1$, $r \geq 0$, and $p+2r \leq (1+2r)q$. Then,

$$A^{(p+2r)/q} \leq (A^r B^p A^r)^{1/q} \tag{8.5.3}$$

and

$$(B^r A^p B^r)^{1/q} \leq B^{(p+2r)/q}. \tag{8.5.4}$$

Proof. See [218]. □

Corollary 8.5.5. Let $A, B \in \mathbf{N}^n$, and assume that $0 \leq A \leq B$. Then,

$$A^2 \leq \left(AB^2A\right)^{1/2} \tag{8.5.5}$$

and

$$\left(BA^2B\right)^{1/2} \leq B^2. \tag{8.5.6}$$

Proof. In Proposition 8.5.4 set $r = 1$, $p = 2$, and $q = 2$. □

Corollary 8.5.6. Let $A, B, C \in \mathbf{N}^n$, and assume that $0 \leq A \leq C \leq B$. Then,

$$\left(CA^2C\right)^{1/2} \leq C^2 \leq \left(CB^2C\right)^{1/2}. \tag{8.5.7}$$

Proof. The result follows directly from Corollary 8.5.5. See also [583]. □

The following result provides representations for A^r, where $r \in [0, 1)$.

Proposition 8.5.7. Let $A \in \mathbf{P}^n$ and $r \in (0, 1)$. Then,

$$A^r = \left(\cos\frac{r\pi}{2}\right)I + \frac{\sin r\pi}{\pi}\int_0^\infty \left[\frac{x^{r+1}}{1+x^2}I - (A+xI)^{-1}x^r\right] \mathrm{d}x \tag{8.5.8}$$

and

$$A^r = \frac{\sin r\pi}{\pi}\int_0^\infty (A+xI)^{-1}Ax^{r-1}\,\mathrm{d}x. \tag{8.5.9}$$

Proof. Let $t \geq 0$. As shown in [90], [93, p. 143],

$$\int_0^\infty \left[\frac{x^{r+1}}{1+x^2} - \frac{x^r}{t+x}\right] \mathrm{d}x = \frac{\pi}{\sin r\pi}\left(t^r - \cos\frac{r\pi}{2}\right).$$

Solving for t^r and replacing t by A yields (8.5.8). Likewise, it follows from [633, p. 448, formula 589] that

$$\int_0^\infty \frac{tx^{r-1}}{t+x}\,\mathrm{d}x = \frac{t^r\pi}{\sin r\pi}.$$

Replacing t by A yields (8.5.9). $\square$

The following result is the *Lowner-Heinz* inequality.

Corollary 8.5.8. Let $A, B \in \mathbf{N}^n$, assume that $0 \le A \le B$, and let $r \in [0,1]$. Then, $A^r \le B^r$. If, in addition, $A < B$ and $r \in (0,1]$, then $A^r < B^r$.

Proof. Let $0 < A \le B$, and let $r \in (0,1)$. In Proposition 8.5.4, replace p, q, r with $r, 1, 0$. The first result now follows from (8.5.3). Alternatively, it follows from (8.5.8) of Proposition 8.5.7 that

$$B^r - A^r = \frac{\sin r\pi}{\pi} \int_0^\infty \big[(A + xI)^{-1} - (B + xI)^{-1}\big] x^r \, \mathrm{d}x.$$

Since $A \le B$, it follows from Proposition 8.1.5 that, for all $x \ge 0$, $(B + xI)^{-1} \le (A + xI)^{-1}$. Hence, $A^r \le B^r$. By continuity, the result holds for $A, B \in \mathbf{N}^n$ and $r \in [0,1]$. In the case $A < B$, it follows from Proposition 8.1.5 that, for all $x \ge 0$, $(B + xI)^{-1} < (A + xI)^{-1}$, so that $A^r < B^r$.

Alternatively, it follows from (8.5.9) of Proposition 8.5.7 that

$$B^r - A^r = \frac{\sin r\pi}{\pi} \int_0^\infty \big[(A + xI)^{-1}A - (B + xI)^{-1}B\big] x^{r-1} \, \mathrm{d}x.$$

Since $A \le B$, it follows that, for all $x \ge 0$, $(B + xI)^{-1}B \le (A + xI)^{-1}A$. Hence, $A^r \le B^r$. For yet another proof, see [625, p. 2]. $\square$

Many of the results already given involve functions that are nondecreasing or increasing on suitable sets of matrices.

Definition 8.5.9. Let $\mathcal{D} \subseteq \mathbf{H}^n$, and let $\phi\colon\ \mathcal{D} \mapsto \mathbf{H}^m$. The function ϕ is *nondecreasing* if $\phi(A) \le \phi(B)$ for all $A, B \in \mathcal{D}$ such that $A \le B$, it is *increasing* if it is nondecreasing and $\phi(A) < \phi(B)$ for all $A, B \in \mathcal{D}$ such that $A < B$, and it is *strongly increasing* if it is nondecreasing and $\phi(A) < \phi(B)$ for all $A, B \in \mathcal{D}$ such that $A \le B$ and $A \ne B$. The function ϕ is (*nonincreasing, decreasing, strongly decreasing*) if $-\phi$ is (nondecreasing, increasing, strongly increasing).

Proposition 8.5.10. The following functions are nondecreasing:

i) $\phi\colon\ \mathbf{H}^n \mapsto \mathbf{H}^n$ defined by $\phi(A) \triangleq BAB^*$, where $B \in \mathbb{F}^{m \times n}$.

ii) $\phi\colon\ \mathbf{H}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq \operatorname{tr} AB$, where $B \in \mathbf{N}^n$.

iii) $\phi\colon\ \mathbf{N}^{n+m} \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq A_{22}|A$, where $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$.

The following functions are increasing:

iv) ϕ: $\mathbf{H}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq \lambda_i(A)$, where $i \in \{1, \ldots, n\}$.

v) ϕ: $\mathbf{N}^n \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq A^r$, where $r \in [0, 1]$.

vi) ϕ: $\mathbf{N}^n \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq A^{1/2}$.

vii) ϕ: $\mathbf{P}^n \mapsto -\mathbf{P}^n$ defined by $\phi(A) \triangleq -A^{-r}$, where $r \in [0, 1]$.

viii) ϕ: $\mathbf{P}^n \mapsto -\mathbf{P}^n$ defined by $\phi(A) \triangleq -A^{-1}$.

ix) ϕ: $\mathbf{P}^n \mapsto -\mathbf{P}^n$ defined by $\phi(A) \triangleq -A^{-1/2}$.

x) ϕ: $-\mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq (-A)^{-r}$, where $r \in [0, 1]$.

xi) ϕ: $-\mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq -A^{-1}$.

xii) ϕ: $-\mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq -A^{-1/2}$.

xiii) ϕ: $\mathbf{H}^n \mapsto \mathbf{H}^m$ defined by $\phi(A) \triangleq BAB^*$, where $B \in \mathbb{F}^{m\times n}$ and $\operatorname{rank} B = m$.

xiv) ϕ: $\mathbf{P}^{n+m} \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq A_{22}|A$, where $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$.

xv) ϕ: $\mathbf{P}^{n+m} \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq -(A_{22}|A)^{-1}$, where $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$.

xvi) ϕ: $\mathbf{P}^n \mapsto \mathbf{H}^m$ defined by $\phi(A) \triangleq \log A$.

The following functions are strongly increasing:

xvii) ϕ: $\mathbf{H}^n \mapsto [0, \infty)$ defined by $\phi(A) \triangleq \operatorname{tr} BAB^*$, where $B \in \mathbb{F}^{m\times n}$ and $\operatorname{rank} B = m$.

xviii) ϕ: $\mathbf{H}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq \operatorname{tr} AB$, where $B \in \mathbf{P}^n$.

xix) ϕ: $\mathbf{N}^n \mapsto [0, \infty)$ defined by $\phi(A) \triangleq \det A$.

Proof. For the proof of *iii)*, see [369]. □

Finally, we consider convex functions defined with respect to matrix inequalities.

Definition 8.5.11. Let $\mathcal{D} \subseteq \mathbb{F}^{n\times m}$ be a convex set and let ϕ: $\mathcal{D} \mapsto \mathbf{H}^p$. The function ϕ is *convex* if

$$\phi[\alpha A_1 + (1-\alpha)A_2] \le \alpha\phi(A_1) + (1-\alpha)\phi(A_2) \tag{8.5.10}$$

for all $\alpha \in [0, 1]$ and $A_1, A_2 \in \mathcal{D}$. The function ϕ is *concave* if $-\phi$ is convex.

Lemma 8.5.12. Let $\mathcal{D} \subseteq \mathbb{F}^{n\times m}$ and $\mathcal{S} \subseteq \mathbf{H}^p$ be convex sets, and let ϕ_1: $\mathcal{D} \mapsto \mathcal{S}$ and ϕ_2: $\mathcal{S} \mapsto \mathbf{H}^q$. Then, the following statements hold:

i) If ϕ_1 is convex and ϕ_2 is nondecreasing and convex, then $\phi_2 \bullet \phi_1\colon\ \mathcal{D} \mapsto \mathbf{H}^q$ is convex.

ii) If ϕ_1 is concave and ϕ_2 is nonincreasing and convex, then $\phi_2 \bullet \phi_1\colon\ \mathcal{D} \mapsto \mathbf{H}^q$ is convex.

iii) If $\mathcal{S}$ is symmetric, $\phi_2(-A) = -\phi_2(A)$ for all $A \in \mathcal{S}$, ϕ_1 is concave, and ϕ_2 is nonincreasing and concave, then $\phi_2 \bullet \phi_1\colon\ \mathcal{D} \mapsto \mathbf{H}^q$ is convex.

iv) If $\mathcal{S}$ is symmetric, $\phi_2(-A) = -\phi_2(A)$ for all $A \in \mathcal{S}$, ϕ_1 is convex, and ϕ_2 is nondecreasing and concave, then $\phi_2 \bullet \phi_1\colon\ \mathcal{D} \mapsto \mathbf{H}^q$ is convex.

Proof. To prove *i)* and *ii)*, let $\alpha \in [0,1]$ and $A_1, A_2 \in \mathcal{D}$. In both cases it follows that

$$\begin{aligned}\phi_2(\phi_1[\alpha A_1 + (1-\alpha)A_2]) &\le \phi_2[\alpha\phi_1(A_1) + (1-\alpha)\phi_1(A_2)]\\ &\le \alpha\phi_2[\phi_1(A_1)] + (1-\alpha)\phi_2[\phi_1(A_2)].\end{aligned}$$

Statements *iii)* and *iv)* follow from *i)* and *ii)*, respectively. □

Proposition 8.5.13. The following functions are convex:

i) $\phi\colon\ \mathbf{N}^n \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq A^r$, where $r \in [1,2]$.

ii) $\phi\colon\ \mathbf{N}^n \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq A^2$.

iii) $\phi\colon\ \mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq A^{-r}$, where $r \in [0,1]$.

iv) $\phi\colon\ \mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq A^{-1}$.

v) $\phi\colon\ \mathbf{P}^n \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq A^{-1/2}$.

vi) $\phi\colon\ \mathbf{N}^n \mapsto -\mathbf{N}^n$ defined by $\phi(A) \triangleq -A^r$, where $r \in [0,1]$.

vii) $\phi\colon\ \mathbf{N}^n \mapsto -\mathbf{N}^n$ defined by $\phi(A) \triangleq -A^{1/2}$.

viii) $\phi\colon\ \mathbf{N}^n \mapsto \mathbf{H}^m$ defined by $\phi(A) \triangleq \gamma BAB^*$, where $\gamma \in \mathbb{R}$ and $B \in \mathbb{F}^{m\times n}$.

ix) $\phi\colon\ \mathbf{N}^n \mapsto \mathbf{N}^m$ defined by $\phi(A) \triangleq BA^rB^*$, where $B \in \mathbb{F}^{m\times n}$ and $r \in [1,2]$.

x) $\phi\colon\ \mathbf{P}^n \mapsto \mathbf{N}^m$ defined by $\phi(A) \triangleq BA^{-r}B^*$, where $B \in \mathbb{F}^{m\times n}$ and $r \in [0,1]$.

xi) $\phi\colon\ \mathbf{N}^n \mapsto -\mathbf{N}^m$ defined by $\phi(A) \triangleq -BA^rB^*$, where $B \in \mathbb{F}^{m\times n}$ and $r \in [0,1]$.

xii) $\phi\colon\ \mathbf{P}^n \mapsto -\mathbf{P}^m$ defined by $\phi(A) \triangleq -(BA^{-r}B^*)^{-p}$, where $B \in \mathbb{F}^{m\times n}$ has rank m and $r, p \in [0,1]$.

xiii) ϕ: $\mathbb{F}^{n\times m} \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq ABA^*$, where $B \in \mathbf{N}^m$.

xiv) ϕ: $\mathbf{P}^n \times \mathbb{F}^{m\times n} \mapsto \mathbf{N}^m$ defined by $\phi(A,B) \triangleq BA^{-1}B^*$.

xv) ϕ: $\mathbf{N}^{n+m} \mapsto \mathbf{N}^n$ defined by $\phi(A) \triangleq -A_{22}|A$, where $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$.

xvi) ϕ: $\mathbf{P}^{n+m} \mapsto \mathbf{P}^n$ defined by $\phi(A) \triangleq (A_{22}|A)^{-1}$, where $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$.

xvii) ϕ: $\mathbf{P}^n \mapsto (0,\infty)$ defined by $\phi(A) \triangleq \operatorname{tr} A^{-r}$, where $r > 0$.

xviii) ϕ: $\mathbf{P}^n \mapsto (-\infty,0)$ defined by $\phi(A) \triangleq -\left(\operatorname{tr} A^{-r}\right)^{-p}$, where $r,p \in [0,1]$.

xix) ϕ: $\mathbf{N}^n \times \mathbf{N}^n \mapsto (-\infty,0]$ defined by $\phi(A,B) \triangleq -\operatorname{tr}\left(A^r + B^r\right)^{1/r}$, where $r \in [0,1]$.

xx) ϕ: $\mathbf{N}^n \times \mathbf{N}^n \mapsto [0,\infty)$ defined by $\phi(A,B) \triangleq \operatorname{tr}\left(A^2 + B^2\right)^{1/2}$.

xxi) ϕ: $\mathbf{N}^n \times \mathbf{N}^m \mapsto \mathbb{R}$ defined by $\phi(A,B) \triangleq -\operatorname{tr} A^r X B^p X^*$, where $X \in \mathbb{F}^{n\times m}$, $r,p \ge 0$, and $r+p \le 1$.

xxii) ϕ: $\mathbf{N}^n \mapsto (-\infty,0)$ defined by $\phi(A) \triangleq -\operatorname{tr} A^r X A^p X^*$, where $X \in \mathbb{F}^{n\times n}$, $r,p \ge 0$, and $r+p \le 1$.

xxiii) ϕ: $\mathbf{P}^n \times \mathbf{P}^m \times \mathbb{F}^{m\times n} \mapsto \mathbb{R}$ defined by $\phi(A,B,X) \triangleq (\operatorname{tr} A^{-p} X B^{-r} X^*)^q$, where $r,p \ge 0$, $r+p \le 1$, and $q \ge (2-r-p)^{-1}$.

xxiv) ϕ: $\mathbf{P}^n \times \mathbb{F}^{n\times n} \mapsto [0,\infty)$ defined by $\phi(A,X) \triangleq \operatorname{tr} A^{-p} X A^{-r} X^*$, where $r,p \ge 0$ and $r+p \le 1$.

xxv) ϕ: $\mathbf{P}^n \times \mathbb{F}^{n\times n} \mapsto [0,\infty)$ defined by $\phi(A) \triangleq \operatorname{tr} A^{-p} X A^{-r} X^*$, where $r,p \in [0,1]$ and $X \in \mathbb{F}^{n\times n}$.

xxvi) ϕ: $\mathbf{P}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq \operatorname{tr}([A^r, X][A^{1-r}, X])$, where $X \in \mathbf{H}^n$.

xxvii) ϕ: $\mathbf{P}^n \mapsto \mathbf{H}^m$ defined by $\phi(A) \triangleq A \log A$.

xxviii) ϕ: $\mathbf{N}^n \backslash \{0\} \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq -\log \operatorname{tr} A^r$, where $r \in [0,1]$.

xxix) ϕ: $\mathbf{P}^n \times \mathbf{P}^n \mapsto (0,\infty)$ defined by $\phi(A,B) \triangleq \operatorname{tr}[A(\log A - \log B)]$.

xxx) ϕ: $\mathbf{N}^n \mapsto (-\infty,0]$ defined by $\phi(A) \triangleq -(\det A)^{1/n}$.

xxxi) ϕ: $\mathbf{P}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq -\log\det A$.

xxxii) ϕ: $\mathbf{P}^n \mapsto (0,\infty)$ defined by $\phi(A) \triangleq \det A^{-1}$.

xxxiii) ϕ: $\mathbf{N}^n \times \mathbf{N}^m \mapsto -\mathbf{N}^{nm}$ defined by $\phi(A,B) \triangleq -A^r \otimes B^{1-r}$, where $r \in [0,1]$.

xxxiv) ϕ: $\mathbf{N}^n \times \mathbf{N}^n \mapsto -\mathbf{N}^n$ defined by $\phi(A,B) \triangleq -A^r \circ B^{1-r}$, where $r \in [0,1]$.

xxxv) ϕ: $\mathbf{H}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq \sum_{i=1}^{k} \lambda_i(A)$, where $k \in \{1,\ldots,n\}$.

xxxvi) ϕ: $\mathbf{H}^n \mapsto \mathbb{R}$ defined by $\phi(A) \triangleq -\sum_{i=k}^{n} \lambda_i(A)$, where $k \in \{1,\ldots,n\}$.

Proof. Statements *i)* and *iii)* are proved in [23] and [93, p. 123].

Let $\alpha \in [0,1]$ for the remainder of the proof.

To prove *ii)* directly, let $A_1, A_2 \in \mathbf{H}^n$. Since

$$\alpha(1-\alpha) = \left(\alpha-\alpha^2\right)^{1/2}\left[(1-\alpha)-(1-\alpha)^2\right]^{1/2},$$

it follows that

$$\begin{aligned} 0 &\le \left[\left(\alpha-\alpha^2\right)^{1/2}A_1 - \left[(1-\alpha)-(1-\alpha)^2\right]^{1/2}A_2\right]^2 \\ &= \left(\alpha-\alpha^2\right)A_1^2 + \left[(1-\alpha)-(1-\alpha)^2\right]A_2^2 - \alpha(1-\alpha)(A_1A_2+A_2A_1). \end{aligned}$$

Hence,

$$[\alpha A_1 + (1-\alpha)A_2]^2 \le \alpha A_1^2 + (1-\alpha)A_2^2,$$

which shows that $\phi(A) = A^2$ is convex.

To prove *iv)* directly, let $A_1, A_2 \in \mathbf{P}^n$. Then, $\left[\begin{smallmatrix} A_1^{-1} & I \\ I & A_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} A_2^{-1} & I \\ I & A_2 \end{smallmatrix}\right]$ are nonnegative semidefinite, and thus

$$\begin{aligned} &\alpha\begin{bmatrix} A_1^{-1} & I \\ I & A_1 \end{bmatrix} + (1-\alpha)\begin{bmatrix} A_2^{-1} & I \\ I & A_2 \end{bmatrix} \\ &\qquad = \begin{bmatrix} \alpha A_1^{-1} + (1-\alpha)A_2^{-1} & I \\ I & \alpha A_1 + (1-\alpha)A_2 \end{bmatrix} \end{aligned}$$

is nonnegative semidefinite. It now follows from Proposition 8.2.3 that $[\alpha A_1 + (1-\alpha)A_2]^{-1} \le \alpha A_1^{-1} + (1-\alpha)A_2^{-1}$, which shows that $\phi(A) = A^{-1}$ is convex.

To prove *v)* directly, note that $\phi(A) = A^{-1/2} = \phi_2[\phi_1(A)]$, where $\phi_1(A) \triangleq A^{1/2}$ and $\phi_2(B) \triangleq B^{-1}$. It follows from *vii)* that ϕ_1 is concave, while it follows from *iv)* that ϕ_2 is convex. Furthermore, *viii)* of Proposition 8.5.10 implies that ϕ_2 is nonincreasing. It thus follows from *ii)* of Lemma 8.5.12 that $\phi(A) = A^{-1/2}$ is convex.

To prove *vi)*, let $A \in \mathbf{P}^n$ and note that $\phi(A) = -A^r = \phi_2[\phi_1(A)]$, where $\phi_1(A) \triangleq A^{-r}$ and $\phi_2(B) \triangleq -B^{-1}$. It follows from *iii)* that ϕ_1 is convex, while it follows from *iv)* that ϕ_2 is concave. Furthermore, *viii)* of Proposition 8.5.10 implies that ϕ_2 is nondecreasing. It thus follows from *iv)* of Lemma 8.5.12 that $\phi(A) = A^r$ is convex on $\mathbf{P}^n$. Continuity implies that $\phi(A) = A^r$ is convex on $\mathbf{N}^n$.

To prove *vii*) directly, let $A_1, A_2 \in \mathbf{N}^n$. Then,

$$0 \le \alpha(1-\alpha)\Big(A_1^{1/2} - A_2^{1/2}\Big)^2,$$

which is equivalent to

$$\Big[\alpha A_1^{1/2} + (1-\alpha)A_2^{1/2}\Big]^2 \le \alpha A_1 + (1-\alpha)A_2.$$

Using *vi*) of Proposition 8.5.10 yields

$$\alpha A_1^{1/2} + (1-\alpha)A_2^{1/2} \le [\alpha A_1 + (1-\alpha)A_2]^{1/2}.$$

Finally, multiplying by -1 shows that $\phi(A) = -A^{1/2}$ is convex.

The proof of *viii*) is immediate. Statements *ix*), *x*), *xi*) follow from *i*), *iii*), and *vi*), respectively.

To prove *xii*), note that $\phi(A) = -\left(BA^{-r}B^*\right)^{-p} = \phi_2[\phi_1(A)]$, where $\phi_1(A) = -BA^{-r}B^*$ and $\phi_2(C) = C^{-p}$. Statement *x*) implies that ϕ_1 is concave, while *iii*) implies that ϕ_2 is convex. Furthermore, *vii*) of Proposition 8.5.10 implies that ϕ_2 is nonincreasing. It thus follows from *ii*) of Lemma 8.5.12 that $\phi(A) = -\left(BA^{-r}B^*\right)^{-p}$ is convex.

To prove *xiii*), let $A_1, A_2 \in \mathbb{F}^{n\times m}$, and let $B \in \mathbf{N}^m$. Then,

$$\begin{aligned} 0 &\le \alpha(1-\alpha)(A_1 - A_2)B(A_1-A_2)^* \\ &= \alpha A_1BA_1^* + (1-\alpha)A_2BA_2^* - [\alpha A_1 + (1-\alpha)A_2]B[\alpha A_1 + (1-\alpha)A_2]^*. \end{aligned}$$

Thus,

$$[\alpha A_1 + (1-\alpha)A_2]B[\alpha A_1 + (1-\alpha)A_2]^* \le \alpha A_1BA_1^* + (1-\alpha)A_2BA_2^*,$$

which shows that $\phi(A) = ABA^*$ is convex.

To prove *xiv*), let $A_1, A_2 \in \mathbf{P}^n$ and $B_1, B_2 \in \mathbb{F}^{m\times n}$. Then, it follows from Proposition 8.2.3 that $\left[\begin{smallmatrix} B_1A_1^{-1}B_1^* & B_1 \\ B_1^* & A_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} B_2A_2^{-1}B_2^* & B_2 \\ B_2^* & A_2 \end{smallmatrix}\right]$ are nonnegative semidefinite and thus

$$\begin{aligned} &\alpha\begin{bmatrix} B_1A_1^{-1}B_1^* & B_1 \\ B_1^* & A_1 \end{bmatrix} + (1-\alpha)\begin{bmatrix} B_2A_2^{-1}B_2^* & B_2 \\ B_2^* & A_2 \end{bmatrix} \\ &\qquad = \begin{bmatrix} \alpha B_1A_1^{-1}B_1^* + (1-\alpha)B_2A_2^{-1}B_2^* & \alpha B_1 + (1-\alpha)B_2 \\ \alpha B_1^* + (1-\alpha)B_2^* & \alpha A_1 + (1-\alpha)A_2 \end{bmatrix} \end{aligned}$$

is nonnegative semidefinite. It thus follows from Proposition 8.2.3 that

$$[\alpha B_1 + (1-\alpha)B_2][\alpha A_1 + (1-\alpha)A_2]^{-1}[\alpha B_1 + (1-\alpha)B_2]^* \le \alpha B_1 A_1^{-1} B_1^* + (1-\alpha)B_2 A_2^{-1} B_2^*,$$

which shows that $\phi(A,B) = BA^{-1}B^*$ is convex.

To prove *xv*), let $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbf{P}^{n+m}$ and $B \triangleq \begin{bmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix} \in \mathbf{P}^{n+m}$. Then, it follows from *xiv*) with A_1, B_1, A_2, B_2 replaced by $A_{22}, A_{12}, B_{22}, B_{12}$, respectively, that

$$[\alpha A_{12} + (1-\alpha)B_{12}][\alpha A_{22} + (1-\alpha)B_{22}]^{-1}[\alpha A_{12} + (1-\alpha)B_{12}]^* \le \alpha A_{12}A_{22}^{-1}A_{12}^* + (1-\alpha)B_{12}B_{22}^{-1}B_{12}^*.$$

Hence,

$$\begin{aligned} -[\alpha A_{22} + (1-\alpha)B_{22}]|[\alpha A + (1-\alpha)B] &= [\alpha A_{12} + (1-\alpha)B_{12}][\alpha A_{22} + (1-\alpha)B_{22}]^{-1}[\alpha A_{12} + (1-\alpha)B_{12}]^* \\ &\quad - [\alpha A_{11} + (1-\alpha)B_{11}] \\ &\le \alpha\big(A_{12}A_{22}^{-1}A_{12}^* - A_{11}\big) + (1-\alpha)(B_{12}B_{22}^{-1}B_{12}^* - B_{11}) \\ &= \alpha(-A_{22}|A) + (1-\alpha)(-B_{22}|B), \end{aligned}$$

which shows that $\phi(A) \triangleq -A_{22}|A$ is convex. By continuity, the result holds for $A \in \mathbf{N}^{n+m}$.

To prove *xvi*), note that $\phi(A) = (A_{22}|A)^{-1} = \phi_2[\phi_1(A)]$, where $\phi_1(A) = A_{22}|A$ and $\phi_2(B) = B^{-1}$. It follows from *xv*) that ϕ_1 is concave, while it follows from *iv*) that ϕ_2 is convex. Furthermore, *viii*) of Proposition 8.5.10 implies that ϕ_2 is nonincreasing. It thus follows from Lemma 8.5.12 that $\phi(A) \triangleq (A_{22}|A)^{-1}$ is convex.

Result *xvii*) is given in by Theorem 9 of [372].

To prove *xviii*), note that $\phi(A) = -(\operatorname{tr} A^{-r})^{-p} = \phi_2[\phi_1(A)]$, where $\phi_1(A) = \operatorname{tr} A^{-r}$ and $\phi_2(B) = -B^{-p}$. Statement *iii*) implies that ϕ_1 is convex and that ϕ_2 is concave. Furthermore, *vii*) of Proposition 8.5.10 implies that ϕ_2 is nondecreasing. It thus follows from *iv*) of Lemma 8.5.12 that $\phi(A) = -(\operatorname{tr} A^{-r})^{-p}$ is convex.

Results *xix*) and *xx*) are proved in [126].

Results *xxi*)–*xxv*) are given by Corollary 1.1, Theorem 1, Corollary 2.1, Theorem 2, and Theorem 8, respectively, of [126]. A proof of *xxi*) in the case $p = 1 - r$ is given in [93, p. 273].

Result *xxvi*) is proved in [126] and [93, p. 274].

Result *xxvii*) is given in [93, p. 123].

To prove *xviii*), note that $\phi(A) = -\log \operatorname{tr} A^r = \phi_2[\phi_1(A)]$, where $\phi_1(A) = \operatorname{tr} A^r$ and $\phi_2(x) = -\log x$. Statement *vi*) implies that ϕ_1 is concave. Furthermore, ϕ_2 is convex and nonincreasing. It thus follows from *ii*) of Lemma 8.5.12 that $\phi(A) = -\log \operatorname{tr} A^r$ is convex.

Result *xxix*) is given in [93, p. 275].

To prove *xxx*), let $A_1, A_2 \in \mathbf{N}^n$. From Corollary 8.4.15 it follows that $(\det A_1)^{1/n} + (\det A_2)^{1/n} \le [\det(A_1 + A_2)]^{1/n}$. Replacing A_1 and A_2 by αA_1 and $(1-\alpha)A_2$, respectively, and multiplying by -1 shows that $\phi(A) = -(\det A)^{1/n}$ is convex.

To prove *xxxi*), note that $\phi(A) = -n\log\big[(\det A)^{1/n}\big] = \phi_2[\phi_1(A)]$, where $\phi_1(A) = (\det A)^{1/n}$ and $\phi_2(x) = -n\log x$. It follows from *xix*) that ϕ_1 is concave. Since ϕ_2 is nonincreasing and convex, it follows from *ii*) of Lemma 8.5.12 that $\phi(A) = -\log\det A$ is convex.

To prove *xxxii*), note that $\phi(A) = \det A^{-1} = \phi_2[\phi_1(A)]$, where $\phi_1(A) = \log\det A^{-1}$ and $\phi_2(x) = e^x$. It follows from *xx*) that ϕ_1 is convex. Since ϕ_2 is nondecreasing and convex, it follows from *i*) of Lemma 8.5.12 that $\phi(A) = \det A^{-1}$ is convex.

Next, *xxxiii*) is given in [93, p. 273] and [625, p. 9]. Statement *xxxiv*) is given in [625, p. 9].

Finally, *xxxv*) is given in [400, p. 478]. Statement *xxxvi*) follows immediately from *xxxv*). □

The following result is a corollary of *xv*) of Proposition 8.5.13 for the case$\alpha = 1/2$. Versions of this result appear in [128, 272, 369] and [466, p. 152].

Corollary 8.5.14. Let $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{n+m}$ and $B \triangleq \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix} \in \mathbb{F}^{n+m}$, and assume that A and B are nonnegative semidefinite. Then,

$$A_{11}|A + B_{11}|B \le (A_{11} + B_{11})|(A + B).$$

The following corollary of *xxxv*) gives a strong majorization condition for the eigenvalues of a pair of Hermitian matrices.

Corollary 8.5.15. Let $A, B \in \mathbf{H}^n$. Then, for all $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} \lambda_i(A+B) \le \sum_{i=1}^{k} [\lambda_i(A) + \lambda_i(B)] \tag{8.5.11}$$

with equality for $k = n$.

Proof. See [93, p. 69], [289, p. 201], or [400, p. 478]. □

8.6 Facts on Range and Rank

Fact 8.6.1. Let $A, B \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite. Then, there exists $\alpha > 0$ such that $A \le \alpha B$ if and only if $\mathcal{R}(A) \subseteq \mathcal{R}(B)$. In this case, $\operatorname{rank} A \le \operatorname{rank} B$. (Proof: Use Theorem 8.5.2 and Corollary 8.5.8.)

Fact 8.6.2. Let $A, B \in \mathbb{F}^{n \times n}$, and assume that A is nonnegative semidefinite and B is either nonnegative semidefinite or skew Hermitian. Then, the following identities hold:

i) $\mathcal{N}(A+B) = \mathcal{N}(A) \cap \mathcal{N}(B)$.

ii) $\mathcal{R}(A+B) = \mathcal{R}(A) + \mathcal{R}(B)$.

(Proof: Use $[(\mathcal{N}(A) \cap \mathcal{N}(B)]^\perp = \mathcal{R}(A) + \mathcal{R}(B)$.)

Fact 8.6.3. Let $A \in \mathbb{F}^{n \times n}$, and assume that $A + A^* \ge 0$. Then, the following identities hold:

i) $\mathcal{N}(A) = \mathcal{N}(A + A^*) \cap \mathcal{N}(A - A^*)$.

ii) $\mathcal{R}(A) = \mathcal{R}(A + A^*) + \mathcal{R}(A - A^*)$.

iii) $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A + A^* & A - A^* \end{bmatrix}$.

Fact 8.6.4. Let $A, B \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite. Then,

$$\operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank}(A+B)$$

and

$$\operatorname{rank} \begin{bmatrix} A & B \\ 0 & A \end{bmatrix} = \operatorname{rank} A + \operatorname{rank}(A+B).$$

(Proof: Using Fact 8.6.2,

$$\begin{aligned}\mathcal{R}\left(\begin{bmatrix} A & B \end{bmatrix}\right) &= \mathcal{R}\left(\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}\right) = \mathcal{R}\big(A^2 + B^2\big) = \mathcal{R}\big(A^2\big) + \mathcal{R}\big(B^2\big) \\ &= \mathcal{R}(A) + \mathcal{R}(B) = \mathcal{R}(A+B).\end{aligned}$$

Alternatively, it follows from Fact 6.4.11 that

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A & B \end{bmatrix} &= \operatorname{rank}\begin{bmatrix} A+B & B \end{bmatrix}\\ &= \operatorname{rank}(A+B) + \operatorname{rank}[B-(A+B)(A+B)^+B].\end{aligned}$$

Next, note that

$$\begin{aligned}\operatorname{rank}[B-(A+B)(A+B)^+B] &= \operatorname{rank}\Big(B^{1/2}\big[I-(A+B)(A+B)^+\big]B^{1/2}\Big)\\ &\le \operatorname{rank}\Big(B^{1/2}\big[I-BB^+\big]B^{1/2}\Big) = 0.\end{aligned}$$

For the second result use Theorem 8.3.4 to simultaneously diagonalize A and B.)

8.7 Facts on Identities and Inequalities Involving One Matrix

Fact 8.7.1. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and assume that there exists $i \in \{1,\ldots,n\}$ such that $A_{(i,i)} = 0$. Then, $\operatorname{row}_i(A) = 0$ and $\operatorname{col}_i(A) = 0$.

Fact 8.7.2. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, $A_{(i,i)} \ge 0$ for all $i = 1,\ldots,n$, and $|A_{(i,j)}|^2 \le A_{(i,i)}A_{(j,j)}$ for all $i,j = 1,\ldots,n$.

Fact 8.7.3. Let $A \in \mathbb{F}^{n\times n}$. Then, $A \ge 0$ if and only if $A \ge -A$.

Fact 8.7.4. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, $A^2 \ge 0$.

Fact 8.7.5. Let $A \in \mathbb{F}^{n\times n}$ be skew Hermitian. Then, $A^2 \le 0$.

Fact 8.7.6. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$(A+A^*)^2 \ge 0$$

and

$$(A-A^*)^2 \le 0.$$

Fact 8.7.7. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$A^2 + A^{2*} \le AA^* + A^*A.$$

Equality holds if and only if $A = A^*$.

Fact 8.7.8. Let $A \in \mathbb{F}^{n\times n}$, and let $\alpha > 0$. Then,

$$A + A^* \le \alpha I + \alpha^{-1}AA^*.$$

Equality holds if and only if $A = \alpha I$.

Fact 8.7.9. Let $A \in \mathbb{F}^{n \times n}$ be positive definite. Then,

$$2I \le A + A^{-1}.$$

Equality holds if and only if $A = I$.

Fact 8.7.10. Let $A \in \mathbb{F}^{n \times n}$ be Hermitian. Then, $A^2 \le A$ if and only if $0 \le A \le I$.

Fact 8.7.11. Let $A \in \mathbb{F}^{n \times n}$ be Hermitian. Then, $\alpha I + A \ge 0$ if and only if $\alpha \ge -\lambda_{\min}(A)$. Furthermore,

$$A^2 + A + \tfrac{1}{4}I \ge 0.$$

Fact 8.7.12. Let $A \in \mathbb{F}^{n \times m}$. Then, $AA^* \le I_n$ if and only if $A^*A \le I_m$.

Fact 8.7.13. Let $A \in \mathbb{F}^{n \times n}$, and assume that either $AA^* \le A^*A$ or $A^*A \le AA^*$. Then, A is normal. (Proof: Use the Schur decomposition.)

Fact 8.7.14. Let $A \in \mathbb{F}^{n \times n}$ be a projector. Then,

$$0 \le A \le I.$$

Fact 8.7.15. Let $A \in \mathbb{F}^{n \times m}$. Then,

$$(AA^*)^{1/2}A = A(A^*A)^{1/2}.$$

Fact 8.7.16. Let $A \in \mathbb{F}^{n \times m}$, and assume that A^*A is nonsingular. Then,

$$(AA^*)^{1/2} = A(A^*A)^{-1/2}A^*.$$

Fact 8.7.17. Let $A \in \mathbb{F}^{n \times n}$ be nonsingular. Then, $(AA^*)^{-1/2}A$ is unitary.

Fact 8.7.18. Let $A \in \mathbb{F}^{n \times n}$. Then, A is positive definite if and only if $I + A$ is nonsingular and the matrices $I - B$ and $I + B$ are positive definite, where $B \triangleq (I + A)^{-1}(I - A)$. (Proof: See [191].) (Remark: For additional results on the Cayley transform, see Fact 3.6.23, Fact 3.6.24, Fact 3.6.25, Fact 3.9.8, and Fact 11.15.9.)

Fact 8.7.19. Let $A \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite and let $k \in \mathbb{P}$. Then, there exists a unique nonnegative-semidefinite matrix $B \in \mathbb{F}^{n \times n}$ such that $B^k = A$. (Proof: See [287, p. 405].) (Problem: Find a direct proof of uniqueness for $k = 2$ and extend to nonintegral powers.)

Fact 8.7.20. Let $A \in \mathbb{R}^{n \times n}$ be positive definite, assume that $A \le I$, and define $\{B_k\}_{k=0}^{\infty}$ by $B_0 \triangleq 0$ and

$$B_{k+1} \triangleq B_k + \tfrac{1}{2}(A - B_k^2).$$

Then,

$$\lim_{k\to\infty} B_k = A^{1/2}.$$

(Proof: See [74, p. 181].) (Remark: See Fact 5.13.18.)

Fact 8.7.21. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular and define $\{B_k\}_{k=0}^{\infty}$ by $B_0 \triangleq A$ and

$$B_{k+1} \triangleq \tfrac{1}{2}\Big(B_k + B_k^{-\mathrm{T}}\Big).$$

Then,

$$\lim_{k\to\infty} B_k = \big(AA^{\mathrm{T}}\big)^{-1/2}A.$$

(Remark: The limit is unitary. See Fact 8.7.17. See [64, p. 224].)

Fact 8.7.22. Let $0 \le \alpha_1 \le \cdots \le \alpha_n$, and define $A \in \mathbb{R}^{n\times n}$ by $A_{(i,j)} \triangleq \min\{\alpha_i, \alpha_j\}$ for all $i, j = 1, \ldots, n$. Then, A is nonnegative semidefinite. (Problem: Determine $\operatorname{rank} A$. When is A positive definite?) (Remark: When $\alpha_i = i$ for all $i = 1, \ldots, n$, the matrix A is a covariance matrix arising in the theory of Brownian motion.)

Fact 8.7.23. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ be such that $\operatorname{Re}\lambda_i < 0$ for all $i = 1, \ldots, n$, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{C}^{n\times n}$ by

$$A_{(i,j)} \triangleq \frac{-1}{\lambda_i + \overline{\lambda_j}}.$$

Then, A is nonnegative semidefinite. (Proof: Note that $A = 2B \circ (1_{n\times n} - C)^{\{-1\}}$, where $B_{(i,j)} = \frac{1}{(\lambda_i - 1)(\overline{\lambda_j} - 1)}$ and $C_{(i,j)} = \frac{(\lambda_i + 1)(\overline{\lambda_j} + 1)}{(\lambda_i - 1)(\overline{\lambda_j} - 1)}$. Then, note that B is nonnegative semidefinite and that $(1_{n\times n} - C)^{\{-1\}} = 1_{n\times n} + C + C^{\{2\}} + C^{\{3\}} + \cdots$. Alternatively, A satisfies a Lyapunov equation with coefficient $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. See [289, p. 348].) (Remark: A is a Cauchy matrix. See Fact 3.12.13 and Fact 8.7.29.)

Fact 8.7.24. Let $a_1, \ldots, a_n \ge 0$ and $p \in \mathbb{R}$, assume that either $a_1, \ldots, a_n$ are positive or p is positive, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by

$$A_{(i,j)} \triangleq (a_i a_j)^p.$$

Then, A is nonnegative semidefinite. (Proof: $A = a^{\{p\}}a^{\{p\}\mathrm{T}}$, where $a \triangleq \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^{\mathrm{T}}$.)

Fact 8.7.25. Let $a_1, \ldots, a_n > 0$, let $\alpha > 0$, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by

$$A_{(i,j)} \triangleq \frac{1}{(a_i + a_j)^\alpha}.$$

Then, A is nonnegative semidefinite. (Proof: See [462].) (Remark: See Fact 5.9.7.)

Fact 8.7.26. Let $a_1, \ldots, a_n > 0$, let $r \in [-1, 1]$, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by

$$A_{(i,j)} \triangleq \frac{a_i^r + a_j^r}{a_i + a_j}.$$

Then, A is nonnegative semidefinite. (Proof: See [625, p. 74].)

Fact 8.7.27. Let $a_1, \ldots, a_n > 0$, let $q > 0$, let $p \in [-q, q]$, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by

$$A_{(i,j)} \triangleq \frac{a_i^p + a_j^p}{a_i^q + a_j^q}.$$

Then, A is nonnegative semidefinite. (Proof: In Fact 8.7.26, replace a_i by $1/a_i$, and let $r = p/q$. See [405] for the case $q \ge p \ge 0$.) (Remark: The case $q = 1$ and $p = 0$ yields a Cauchy matrix. In the case $n = 2$, $A \ge 0$ yields Fact 1.4.6.) (Problem: When is A positive definite?)

Fact 8.7.28. Let $a_1, \ldots, a_n > 0$, let $p \in [-1, 1]$ and $q \in (-2, 2]$, and, for all $i, j = 1, \ldots, n$, define $A \in \mathbb{R}^{n\times n}$ by

$$A_{(i,j)} \triangleq \frac{a_i^p + a_j^p}{a_i^2 + qa_ia_j + a_j^2}.$$

Then, A is nonnegative semidefinite. (Proof: See [624] or [625, p. 76].)

Fact 8.7.29. Let $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ be positive and, for all $i, j = 1, \ldots, n$, define the Cauchy matrix $A \in \mathbb{R}^{n\times n}$ by $A_{(i,j)} \triangleq 1/(a_i + b_j)$. Then, A is nonnegative semidefinite. If, in addition, $a_1 < \cdots < a_n$ are distinct and $b_1 < \cdots < b_n$ are distinct, then A is positive definite. In particular, the Hilbert matrix is positive definite. (Remark: See Fact 3.12.12 and Fact 3.12.13.) (Problem: Extend this result to complex entries and generalize Fact 8.7.23.)

Fact 8.7.30. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian, assume that $A_{(i,i)} > 0$ for all $i = 1, \ldots, n$, and assume that, for all $i, j = 1, \ldots, n$,

$$|A_{(i,j)}| < \tfrac{1}{n-1}\sqrt{A_{(i,i)}A_{(j,j)}}.$$

Then, A is positive definite. (Proof: Note that

$$x^*Ax = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\begin{bmatrix} x_{(i)} \\ x_{(j)} \end{bmatrix}^*\begin{bmatrix} \frac{1}{n-1}A_{(i,i)} & A_{(i,j)} \\ \overline{A_{(i,j)}} & \frac{1}{n-1}A_{(j,j)} \end{bmatrix}\begin{bmatrix} x_{(i)} \\ x_{(j)} \end{bmatrix}.)$$

(Remark: This result is due to Roup.)

Fact 8.7.31. Let $\alpha_0, \ldots, \alpha_n > 0$, and define the tridiagonal matrix $A \in \mathbb{R}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} \alpha_0+\alpha_1 & -\alpha_1 & 0 & 0 & \cdots & 0 \\ -\alpha_1 & \alpha_1+\alpha_2 & -\alpha_2 & 0 & \cdots & 0 \\ 0 & -\alpha_2 & \alpha_2+\alpha_3 & -\alpha_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha_{n-1}+\alpha_n \end{bmatrix}.$$

Then, A is positive definite. (Proof: For $k = 2, \ldots, n$, the $k \times k$ leading principal subdeterminant of A is given by $\left[\sum_{i=0}^{k} \alpha_i^{-1}\right]\alpha_0\alpha_1\cdots\alpha_k$. See [66, p. 115].) (Remark: A a stiffness matrix arising in structural analysis.)

Fact 8.7.32. Let $x_1, \ldots, x_n \in \mathbb{F}^n$, and define $A \in \mathbb{F}^{n\times n}$ by $A_{(i,j)} \triangleq x_i^* x_j$ for all $i, j = 1, \ldots, n$, and $B \triangleq \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$. Then, $A = B^*B$. Consequently, A is nonnegative semidefinite and $\operatorname{rank} A = \operatorname{rank} B$. Conversely, let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, there exist $x_1, \ldots, x_n \in \mathbb{F}^n$ such that $A = B^*B$, where $B = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$. (Proof: The converse is an immediate consequence of Corollary 5.4.5.) (Remark: A is the *Gram matrix* of $x_1, \ldots, x_n$.)

Fact 8.7.33. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, there exists $B \in \mathbb{F}^{n\times n}$ such that B is upper triangular, B has nonnegative diagonal entries, and $A = BB^*$. If, in addition, A is positive definite, then B is unique and has positive diagonal entries. (Remark: This result is the *Cholesky decomposition.*)

Fact 8.7.34. Let $x \in \mathbb{F}^n$. Then,

$$xx^* \le x^*xI.$$

Fact 8.7.35. Let $A \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = m$. Then,

$$0 \le A(A^*A)^{-1}A^* \le I.$$

Fact 8.7.36. Let $A \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$A^{-1} \le \frac{\alpha+\beta}{\alpha\beta}I - \frac{1}{\alpha\beta}A \le \frac{(\alpha+\beta)^2}{4\alpha\beta}A^{-1},$$

where $\alpha \triangleq \lambda_{\max}(A)$ and $\beta \triangleq \lambda_{\min}(A)$. (Proof: See [401].)

Fact 8.7.37. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{n\times n}$ be Hermitian, assume that A_{22} is nonsingular, and let $S \triangleq \begin{bmatrix} I & -A_{12}A_{22}^{-1} \end{bmatrix}$. Then,

$$A_{11} - A_{12}A_{22}^{-1}A_{12}^* = SAS^*.$$

If, in addition, A is (nonnegative semidefinite, positive definite), then so is $A_{11} - A_{12}A_{22}^{-1}A_{12}^*$.

Fact 8.7.38. Let $A \in \mathbb{F}^{n\times m}$, and define

$$\mathcal{A} \triangleq \begin{bmatrix} (AA^*)^{1/2} & A \\ A^* & (A^*A)^{1/2} \end{bmatrix}.$$

Then, $\mathcal{A}$ is nonnegative semidefinite.

Fact 8.7.39. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, $\begin{bmatrix} A & A \\ A & A \end{bmatrix}$ and $\begin{bmatrix} A & -A \\ -A & A \end{bmatrix}$ are nonnegative semidefinite. Furthermore, if $\begin{bmatrix} \alpha & \beta \\ \overline{\beta} & \gamma \end{bmatrix} \in \mathbb{F}^{2\times 2}$ is nonnegative semidefinite, then so is $\begin{bmatrix} \alpha A & \overline{\beta} A \\ \beta A & \gamma A \end{bmatrix}$. Finally, if A and $\begin{bmatrix} \alpha & \beta \\ \overline{\beta} & \gamma \end{bmatrix}$ are positive definite, then $\begin{bmatrix} \alpha A & \overline{\beta} A \\ \beta A & \gamma A \end{bmatrix}$ is positive definite. (Proof: Use Fact 7.4.13.)

Fact 8.7.40. Let $A_{11}, A_{12}, A_{22} \in \mathbb{F}^{n\times n}$, assume that $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{2n\times 2n}$ is nonnegative semidefinite, and assume that $\begin{bmatrix} \alpha & \beta \\ \overline{\beta} & \gamma \end{bmatrix} \in \mathbb{F}^{2\times 2}$ is nonnegative semidefinite. Then, $\begin{bmatrix} \alpha A_{11} & \beta A_{12} \\ \overline{\beta} A_{12}^* & \gamma A_{22} \end{bmatrix}$ is nonnegative semidefinite. If, in addition, $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$ is positive definite and $\alpha, \beta > 0$, then $\begin{bmatrix} \alpha A_{11} & \beta A_{12} \\ \overline{\beta} A_{12}^* & \gamma A_{22} \end{bmatrix}$ is positive definite. (Proof: Note that $\begin{bmatrix} \alpha A_{11} & \beta A_{12} \\ \overline{\beta} A_{12}^* & \gamma A_{22} \end{bmatrix} = \left(\begin{bmatrix} \alpha & \beta \\ \overline{\beta} & \gamma \end{bmatrix} \otimes 1_{n\times n} \right) \circ \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$ and use Fact 8.15.6 and Fact 7.4.13.) (Problem: Extend this result to nonsquare A_{12}.)

Fact 8.7.41. Let $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{2n\times 2n}$ be nonnegative semidefinite, where $A_{11}, A_{22} \in \mathbb{F}^{n\times n}$. Then,

$$-A_{11} - A_{22} \le A_{12} + A_{12}^* \le A_{11} + A_{22}.$$

If, in addition, $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}$ is positive definite, then

$$-A_{11} - A_{22} < A_{12} + A_{12}^* < A_{11} + A_{22}.$$

(Proof: Consider $S\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}S^{\mathrm{T}}$, where $S \triangleq \begin{bmatrix} I & I \end{bmatrix}$ and $S \triangleq \begin{bmatrix} I & -I \end{bmatrix}$.)

Fact 8.7.42. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, $-A \le B \le A$ if and only if $\left[\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right]$ is nonnegative semidefinite. Furthermore, $-A < B < A$ if and only if $\left[\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right]$ is positive definite.

Fact 8.7.43. Let $A \in \mathbb{R}^{n\times n}$ be positive definite, let $\mathcal{S} \subseteq \{1, \ldots, n\}$, and let $A_{[\mathcal{S}]}$ denote the principal submatrix of A obtained by deleting $\mathrm{row}_i(A)$ and $\mathrm{col}_i(A)$ for all $i \in \mathcal{S}$. Then,

$$\left(A_{[\mathcal{S}]}\right)^{-1} \le \left(A^{-1}\right)_{[\mathcal{S}]}.$$

(Proof: See [287, p. 474].) (Remark: Generalizations of this result are given in [143].)

Fact 8.7.44. Let $A \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$n + \log\det A \le n(\det A)^{1/n} \le \operatorname{tr} A \le \left(n \operatorname{tr} A^2\right)^{1/2},$$

with equality if and only if $A = I$.

Fact 8.7.45. Let $A \triangleq \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \cdot\vdots\cdot & \vdots \\ A_{1k} & \cdots & A_{kk} \end{bmatrix}$, where $A_{ij} \in \mathbb{F}^{n_i\times n_j}$ for all $i, j = 1, \ldots, k$, and assume that A is positive definite. Furthermore, define $\hat{A} \triangleq \begin{bmatrix} \hat{A}_{11} & \cdots & \hat{A}_{1k} \\ \vdots & \cdot\vdots\cdot & \vdots \\ \hat{A}_{1k} & \cdots & \hat{A}_{kk} \end{bmatrix}$, where $\hat{A}_{ij} = 1_{1\times n_i} A_{ij} 1_{n_j\times 1}$ is the sum of the entries of A_{ij} for all $i, j = 1, \ldots, k$. Then, $\hat{A}$ is positive definite. (Proof: $\hat{A} = BAB^{\mathrm{T}}$, where the entries of $B \in \mathbb{R}^{n\times n}$ are zeros and ones. See [22].)

8.8 Facts on Identities and Inequalities Involving Two or More Matrices

Fact 8.8.1. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$\left(A^{-1} + B^{-1}\right)^{-1} = A(A+B)^{-1}B.$$

Fact 8.8.2. Let $A \in \mathbb{F}^{n\times n}$ be positive semidefinite, let $A \in \mathbb{F}^{n\times n}$ be Hermitian, and assume that $A + B$ is nonsingular. Then,

$$(A+B)^{-1} + (A+B)^{-1}B(A+B)^{-1} \le A^{-1}.$$

If, in addition, B is nonsingular, the inequality is strict. (Proof: The inequality is equivalent to $BA^{-1}B \ge 0$. See [443].)

Fact 8.8.3. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times m}$, and assume that B is nonnegative semidefinite. Then, $ABA^* = 0$ if and only if $AB = 0$.

Fact 8.8.4. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, AB is nonnegative semidefinite if and only if AB is normal.

Fact 8.8.5. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and assume that either *i*) A and B are nonnegative semidefinite or *ii*) either A or B is positive definite. Then, AB is group invertible. (Proof: Use Theorem 8.3.2 and Theorem 8.3.5.)

Fact 8.8.6. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and assume that A and $AB+BA$ are positive definite. Then, B is positive definite. (Proof: See [356, p. 120] or [599]. Alternatively, the result follows from Corollary 11.7.4.)

Fact 8.8.7. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and assume that $A \le B$. Then, $A_{(i,i)} \le B_{(i,i)}$ for all $i = 1, \ldots, n$.

Fact 8.8.8. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, $B \le A$ if and only if $BA^{-1}B \le B$.

Fact 8.8.9. Let $A, B, C, D \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and assume that $0 < D \le C$ and $BCB \le ADA$. Then, $B \le A$. (Proof: See [40, 134].)

Fact 8.8.10. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that $0 \le A \le B$. Then,

$$\left(A + \tfrac{1}{4}A^2\right)^{1/2} \le \left(B + \tfrac{1}{4}B^2\right)^{1/2}.$$

(Proof: See [425].)

Fact 8.8.11. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and let $B \in \mathbb{F}^{l\times n}$. Then, BAB^* is positive definite if and only if $B\left(A + A^2\right)B^*$ is positive definite. (Proof: Diagonalize A using a unitary transformation and note that $BA^{1/2}$ and $B\left(A + A^2\right)^{1/2}$ have the same rank.)

Fact 8.8.12. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$, and assume that $\operatorname{rank} B = l$. Then,

$$0 \le A^*B(B^*B)^{-1}B^*A \le A^*A.$$

If, in particular, $m = l = 1$, then

$$|A^*B|^2 \le A^*AB^*B.$$

(Remark: This result is the Cauchy-Schwarz inequality. See Fact 8.13.13.)

Fact 8.8.13. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{m\times n}$, where $\operatorname{rank} B = m$. Then,

$$0 \le B^*(BAB^*)^{-1}B \le A^{-1}$$

and $A^{-1} - B^*(BAB^*)^{-1}B$ is nonnegative semidefinite and has rank $n-m$. (Proof: $I - A^{1/2}B^*(BAB^*)^{-1}BA^{1/2}$ is a projector.)

Fact 8.8.14. Let $A_1, \ldots, A_k \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and let $p, q \in \mathbb{R}$ satisfy $1 \le p \le q$. Then,

$$\left(\frac{1}{k}\sum_{i=1}^{k} A_i^p\right)^{1/p} \le \left(\frac{1}{k}\sum_{i=1}^{k} A_i^q\right)^{1/q}.$$

(Proof: See [90].)

Fact 8.8.15. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, there exists a Hermitian matrix $C \in \mathbb{F}^{n\times n}$ that is a least upper bound for A and B in the sense that $A \le C$, $B \le C$, and, if $D \in \mathbb{F}^{n\times n}$ is a Hermitian matrix satisfying $A \le D$ and $B \le D$, then $C \le D$. (Proof: First consider the case in which A and B are both nonnegative semidefinite.) (Problem: Generalize to three or more matrices.)

Fact 8.8.16. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p, q \in \mathbb{R}$ satisfy $p \ge q \ge 0$. Then,

$$\left[\tfrac{1}{2}(A^q + B^q)\right]^{1/q} \le \left[\tfrac{1}{2}(A^p + B^p)\right]^{1/p}.$$

Furthermore,

$$\mu(A,B) \triangleq \lim_{r\to\infty} \left[\tfrac{1}{2}(A^r + B^r)\right]^{1/r}$$

exists and satisfies

$$A \le \mu(A,B), \quad B \le \mu(A,B).$$

(Proof: See [75].) (Problem: If $A \le C$ and $B \le C$, then does it follow that $\mu(A,B) \le C$? See [27, 323].)

Fact 8.8.17. Let $\left[\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right] \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, where C is positive definite, and let D be positive definite. Then, $\left[\begin{smallmatrix} A+D & B \\ B^* & C \end{smallmatrix}\right]$ is positive definite.

Fact 8.8.18. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $C, D \in \mathbb{F}^{n\times n}$ are positive definite. Then,

$$(A+B)(C+D)^{-1}(A+B)^* \le AC^{-1}A^* + BD^{-1}B^*.$$

(Proof: Form the Schur complement of $A+B$ with respect to the nonnegative-semidefinite matrices $\left[\begin{smallmatrix} AC^{-1}A^* & A \\ A^* & C \end{smallmatrix}\right] + \left[\begin{smallmatrix} BD^{-1}B^* & B \\ B^* & D \end{smallmatrix}\right]$. See [272, 373] or [466,

p. 151].) (Remark: Replacing A, B, C, D by $\alpha B_1, (1-\alpha)B_2, \alpha A_1, (1-\alpha)A_2$ yields *xiv*) of Proposition 8.5.13.)

Fact 8.8.19. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite, let $C \in \mathbb{F}^{n\times n}$ satisfy $B = C^*C$, and let $\alpha \in [0,1]$. Then,

$$C^*\big(C^{-*}AC^{-1}\big)^{\alpha}C \leq \alpha A + (1-\alpha)B.$$

If, in addition, $\alpha \in (0,1)$, then equality holds if and only if $A = B$. (Proof: See [413].)

Fact 8.8.20. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$\begin{aligned} A^{1/2}\Big(A^{-1/2}BA^{-1/2}\Big)^{1/2}A^{1/2} &= A\big(A^{-1}B\big)^{1/2} \\ &= (A+B)\big[(A+B)^{-1}A(A+B)^{-1}B\big]^{1/2}, \end{aligned}$$

where $\big(A^{-1}B\big)^{1/2}$ has positive eigenvalues and satisfies $\big[(A^{-1}B)^{1/2}\big]^2 = A^{-1}B$. Denote the above quantity by $A\#B$. Then,

$$A\#B = B\#A,$$

$$2\big(A^{-1}+B^{-1}\big)^{-1} \leq A\#B \leq \tfrac{1}{2}(A+B),$$

$$(A\#B)B^{-1}(A\#B) = A^{-1},$$

$$\begin{bmatrix} A & A\#B \\ A\#B & B \end{bmatrix} \geq 0.$$

Furthermore, if $X \in \mathbf{H}^n$ and $\left[\begin{smallmatrix} A & X \\ X & B \end{smallmatrix}\right]$ is nonnegative semidefinite, then $X \leq A\#B$. Finally, if $\alpha \in [0,1]$, then

$$\big[\alpha A^{-1} + (1-\alpha)B^{-1}\big]^{-1} \leq A^{1/2}\Big(A^{-1/2}BA^{-1/2}\Big)^{1-\alpha}A^{1/2} \leq \alpha A + (1-\alpha)B,$$

or, equivalently,

$$[\alpha A + (1-\alpha)B]^{-1} \leq A^{-1/2}\Big(A^{-1/2}BA^{-1/2}\Big)^{\alpha-1}A^{-1/2} \leq \alpha A^{-1} + (1-\alpha)B^{-1}.$$

Hence,

$$\operatorname{tr}\,[\alpha A + (1-\alpha)B]^{-1} \leq \operatorname{tr}\left[A^{-1}\Big(A^{-1/2}BA^{-1/2}\Big)^{\alpha-1}\right] \leq \operatorname{tr}\big[\alpha A^{-1} + (1-\alpha)B^{-1}\big].$$

(Proof: See [553].) (Remark: These inequalities improve *iv*) of Proposition 8.5.13. Alternative means and their differences are considered in [8]. $A\#B$ is the *geometric mean* of A and B. A related mean is defined in [205].) (Problem: Does $\left[\begin{smallmatrix} A & X \\ X & B \end{smallmatrix}\right] > 0$ imply that $-(A\#B) < X < A\#B$?) (Remark: A geometric mean for an arbitrary number of positive-definite matrices is given in [28].)

Fact 8.8.21. Let $\{x_i\}_{i=1}^\infty \subset \mathbb{R}^n$ be such that $\sum_{i=1}^\infty x_i$ exists, and let $\{A_i\}_{i=1}^\infty \subset \mathbf{N}^n$ be such that $A_i \le A_{i+1}$ for all $i \in \mathbb{P}$ and $\lim_{i\to\infty} \operatorname{tr} A_i = \infty$. Then,

$$\lim_{k\to\infty} (\operatorname{tr} A_k)^{-1} \sum_{i=1}^k A_i x_i = 0.$$

If, in addition A_i is positive definite for all $i \in \mathbb{P}$ and $\{\lambda_{\max}(A_i)/\lambda_{\min}(A_i)\}_{i=1}^\infty$ is bounded, then

$$\lim_{k\to\infty} A_k^{-1} \sum_{i=1}^k A_i x_i = 0.$$

(Proof: See [16].) (Remark: These identities are matrix versions of the *Kronecker lemma.*)

8.9 Facts on Generalized Inverses

Fact 8.9.1. Let $A \in \mathbb{F}^{m\times m}$ be nonnegative semidefinite. Then, the following statements hold:

i) $A^+ = A^{\mathrm{D}} = A^\# \ge 0.$

ii) $\operatorname{rank} A = \operatorname{rank} A^+.$

iii) $\left(A^{1/2}\right)^+ = (A^+)^{1/2}.$

iv) $A^{1/2} = A(A^+)^{1/2} = (A^+)^{1/2}A.$

v) $AA^+ = A^{1/2}\left(A^{1/2}\right)^+.$

Fact 8.9.2. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$A = (A+B)(A+B)^+A.$$

Fact 8.9.3. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, $A \le B$ if and only if $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ and $\operatorname{sprad}(B^+A) \le 1$. (Proof: See [520].)

Fact 8.9.4. Let $A, B \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite, and assume that $A \le B$. Then, the following statements are equivalent:

i) $B^+ \le A^+.$

ii) $\mathcal{R}(A) = \mathcal{R}(B).$

iii) $\operatorname{rank} A = \operatorname{rank} B.$

Furthermore, the following statements are equivalent:

iv) $A^+ \le B^+.$

v) $A^2 = AB$.

(Proof: See [267, 420].)

Fact 8.9.5. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that $A \le B$. Then,

$$0 \le AA^+ \le BB^+.$$

If, in addition, $\operatorname{rank} A = \operatorname{rank} B$, then

$$AA^+ = BB^+.$$

Fact 8.9.6. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that $A \le B$. Then,

$$0 \le AB^+A \le A \le A + B\big[\big(I - AA^+\big)B\big(I - AA^+\big)\big]^+B \le B.$$

(Proof: See [267].)

Fact 8.9.7. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\operatorname{spec}\big[(A+B)^+A\big] \subset [0,1].$$

(Proof: Let C be positive definite and satisfy $B \le C$. Then, $(A+C)^{-1/2}C(A+C)^{-1/2} \le I$. The result now follows from Fact 8.9.8.)

Fact 8.9.8. Let $A, B, C \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that $B \le C$. Then, for all $i = 1, \ldots, n$,

$$\lambda_i\big[(A+B)^+B\big] \le \lambda_i\big[(A+C)^+C\big].$$

Consequently,

$$\operatorname{tr}\big[(A+B)^+B\big] \le \operatorname{tr}\big[(A+C)^+C\big].$$

(Proof: See [579].) (Remark: See Fact 8.9.7.)

Fact 8.9.9. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and define

$$A\!:\!B \triangleq A(A+B)^+B.$$

Then,

$$A\!:\!B = B - B(A+B)^+A = A - A(A+B)^+B = B\!:\!A,$$
$$\mathcal{R}(A\!:\!B) = \mathcal{R}(A) \cap \mathcal{R}(B),$$

for all $\alpha, \beta > 0$,

$$\big(\alpha^{-1}A\big)\!:\!\big(\beta^{-1}B\big) \le \alpha A + \beta B,$$

$A\!:\!B \ge X$ for all nonnegative-semidefinite matrices $X \in \mathbb{F}^{n\times n}$ such that

$$\begin{bmatrix} A+B & A \\ A & A-X \end{bmatrix} \ge 0,$$

and ϕ: $\mathbf{N}^n \times \mathbf{N}^n \mapsto -\mathbf{N}^n$ defined by $\phi(A,B) \triangleq -A{:}B$ is convex. If A and B are projectors, then

$$A{:}B = (A^+ + B^+)^+$$

and $2(A{:}B)$ is the projector onto $\mathcal{R}(A) \cap \mathcal{R}(B)$. If $A+B$ is positive definite, then

$$A{:}B = A(A+B)^{-1}B.$$

If A and B are positive definite, then

$$A{:}B = \left(A^{-1} + B^{-1}\right)^{-1}.$$

Let $C, D \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$(A{:}B){:}C = A{:}(B{:}C)$$

and

$$A{:}C + B{:}D \le (A+B){:}(C+D).$$

(Proof: See [17, 18, 21, 340], [477, p. 189], and [625, p. 9].) (Remark: $A{:}B$ is the *parallel sum* of A and B.) (Remark: See Fact 6.4.27 and Fact 6.4.28.)

Fact 8.9.10. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. If $(AB)^+ = B^+A^+$, then AB is range Hermitian. Furthermore, the following statements are equivalent:

i) AB is range Hermitian.

ii) $(AB)^\# = B^+A^+$.

iii) $(AB)^+ = B^+A^+$.

(Proof: See [408].) (Remark: See Fact 6.4.6.)

Fact 8.9.11. Let $A \in \mathbb{F}^{n\times n}$ and $C \in \mathbb{F}^{m\times m}$ be nonnegative semidefinite, let $B \in \mathbb{F}^{n\times m}$, and define $X \triangleq A^{+1/2}BC^{+1/2}$. Then, the following statements are equivalent:

i) $\left[\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right]$ is nonnegative semidefinite.

ii) $AA^+B = B$ and $X^*X \le I_m$.

iii) $BC^+C = B$ and $X^*X \le I_m$.

iv) $B = A^{1/2}XC^{1/2}$ and $X^*X \le I_m$.

(Remark: This result provides an explicit expression for X given in [625, p. 15].)

8.10 Facts on Identities and Inequalities Involving Quadratic Forms

Fact 8.10.1. Let $x, y \in \mathbb{F}^n$. Then, $xx^* \le yy^*$ if and only if there exists $\alpha \in \mathbb{F}$ such that $|\alpha| \in [0,1]$ and $x = \alpha y$.

Fact 8.10.2. Let $x, y \in \mathbb{F}^n$. Then, $xy^* + yx^* \ge 0$ if and only x and y are linearly dependent. (Proof: Evaluate the product of the nonzero eigenvalues of $xy^* + yx^*$ and use the Cauchy-Schwarz inequality $|x^*y|^2 \le x^*xy^*y$.)

Fact 8.10.3. Let $A \in \mathbb{F}^{n\times n}$ be positive definite, and let $x, y \in \mathbb{F}^n$. Then,

$$2\mathrm{Re}\, x^*y \le x^*Ax + y^*A^{-1}y.$$

(Proof: $\big(A^{1/2}x - A^{-1/2}y\big)^*\big(A^{1/2}x - A^{-1/2}y\big) \ge 0.$)

Fact 8.10.4. Let $A \in \mathbb{F}^{n\times n}$ be positive definite, and let $x, y \in \mathbb{F}^n$. Then,

$$|x^*y|^2 \le (x^*Ax)\big(y^*A^{-1}y\big).$$

(Proof: Use Fact 8.8.12 with A replaced by $A^{1/2}x$ and B replaced by $A^{-1/2}y$.)

Fact 8.10.5. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $x \in \mathbb{F}^n$. Then,

$$(x^*x)^2 \le (x^*Ax)\big(x^*A^{-1}x\big) \le \frac{(\alpha+\beta)^2}{4\alpha\beta}(x^*x)^2,$$

where $\alpha \triangleq \lambda_{\min}(A)$ and $\beta \triangleq \lambda_{\max}(A)$. (Remark: The second inequality is the *Kantorovich inequality.* See Fact 1.4.14 and [9]. See also [378].)

Fact 8.10.6. Let $A \in \mathbb{F}^{n\times n}$ be positive definite, let $y \in \mathbb{F}^n$, let $\alpha > 0$, and define $f\colon\ \mathbb{F}^n \mapsto \mathbb{R}$ by $f(x) \triangleq |x^*y|^2$. Then,

$$x_0 = \sqrt{\frac{\alpha}{y^*A^{-1}y}}A^{-1}y$$

minimizes $f(x)$ subject to $x^*Ax \le \alpha$. Furthermore, $f(x_0) = \alpha y^*A^{-1}y$. (Proof: See [14].)

Fact 8.10.7. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and let $x \in \mathbb{F}^n$. Then,

$$\big(x^*A^2x\big)^2 \le (x^*Ax)\big(x^*A^3x\big)$$

and

$$(x^*Ax)^2 \le (x^*x)\big(x^*A^2x\big).$$

Fact 8.10.8. Let $A, B \in \mathbb{R}^n$, and assume that A is Hermitian is B is positive definite. Then,

$$\lambda_{\max}\big(AB^{-1}\big) = \max\{\lambda \in \mathbb{R}\colon\ \det(A - \lambda B) = 0\} = \min_{x\in\mathbb{F}^n\setminus\{0\}} \frac{x^*Ax}{x^*Bx}.$$

(Proof: Use Lemma 8.4.3.)

Fact 8.10.9. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A is positive definite and B is nonnegative semidefinite. Then,

$$4(x^*x)(x^*Bx) < (x^*Ax)^2$$

for all nonzero $x \in \mathbb{R}^n$ if and only if there exists $\alpha > 0$ such that

$$\alpha I + \alpha^{-1}B < A.$$

In this case, $4B < A^2$ and hence $2B^{1/2} < A$. (Proof: Sufficiency follows from $\alpha x^*x + \alpha^{-1}x^*Bx < x^*Ax$. Necessity follows from Fact 8.10.10. The last result follows from $(A - 2\alpha I)^2 \geq 0$ or $2B^{1/2} \leq \alpha I + \alpha^{-1}B$.)

Fact 8.10.10. Let $A, B, C \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that

$$4(x^*Cx)(x^*Bx) < (x^*Ax)^2$$

for all nonzero $x \in \mathbb{R}^n$. Then, there exists $\alpha > 0$ such that

$$\alpha C + \alpha^{-1}B < A.$$

(Proof: See [457].)

Fact 8.10.11. Let $A, B \in \mathbb{F}^{n\times n}$, where A is Hermitian and B is nonnegative semidefinite. Then, $x^*Ax < 0$ for all $x \in \mathbb{F}^n$ such that $Bx = 0$ and $x \neq 0$ if and only if there exists $\alpha > 0$ such that $A < \alpha B$. (Proof: Suppose that for every $\alpha > 0$ there exists $x \neq 0$ such that $x^*Ax \geq \alpha x^*Bx$. Now, $Bx = 0$ implies that $x^*Ax \geq 0$.)

Fact 8.10.12. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and linearly independent. Then, the following statements are equivalent:

i) There exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha A + \beta B$ is positive definite.

ii) Either $x^*Ax \geq 0$ for all $x \in \{y \in \mathbb{F}^n\colon\ y^*By = 0\}$ or $x^*Ax \leq 0$ for all $x \in \{y \in \mathbb{F}^n\colon\ y^*By = 0\}$.

Now, assume that $\mathbb{F} = \mathbb{R}$ and $n \geq 3$. Then, the following statement is equivalent to *i)* and *ii)*:

iii) $\{x \in \mathbb{R}^n\colon\ x^{\mathrm{T}}Ax = x^{\mathrm{T}}Bx = 0\} = \{0\}$.

(Remark: The equivalence of *i)* and *ii)* is *Finsler's lemma.* A history of this result is given in [563].)

Fact 8.10.13. Let $A \in \mathbb{R}^{n\times n}$ be positive definite. Then,

$$\int_{\mathbb{R}^n} e^{-x^{\mathrm{T}}Ax}\,\mathrm{d}x = \frac{\pi^{n/2}}{\sqrt{\det A}}.$$

Fact 8.10.14. Let $A, B \in \mathbb{R}^{n\times n}$ be positive definite and, for $k = 0, 1, 2, 3$, define

$$\mathcal{I}_k \triangleq \frac{1}{(2\pi)^{n/2}\sqrt{\det A}} \int_{\mathbb{R}^n} \left(x^{\mathrm{T}}Bx\right)^k e^{-\frac{1}{2}x^{\mathrm{T}}A^{-1}x}\,\mathrm{d}x.$$

Then,

$$\mathcal{I}_0 = 1,$$
$$\mathcal{I}_1 = \operatorname{tr} AB,$$
$$\mathcal{I}_2 = (\operatorname{tr} AB)^2 + 2\operatorname{tr}(AB)^2,$$
$$\mathcal{I}_3 = (\operatorname{tr} AB)^3 + 6(\operatorname{tr} AB)\left[\operatorname{tr}(AB)^2\right] + 8\operatorname{tr}(AB)^3.$$

(Proof: See [419, p. 80].) (Remark: These identities are *Lancaster's formulas.*)

Fact 8.10.15. Let $A \in \mathbb{R}^{n\times n}$ be positive definite, let $B \in \mathbb{R}^{n\times n}$, let $a, b \in \mathbb{R}^n$, and let $\alpha, \beta \in \mathbb{R}$. Then,

$$\int_{\mathbb{R}^n} \left(x^{\mathrm{T}}Bx + b^{\mathrm{T}}x + \beta\right) e^{-(x^{\mathrm{T}}Ax + a^{\mathrm{T}}x + \alpha)}\,\mathrm{d}x$$
$$= \frac{\pi^{n/2}}{2\sqrt{\det A}}\left[2\beta + \operatorname{tr}\left(A^{-1}B\right) - b^{\mathrm{T}}A^{-1}a + \tfrac{1}{2}a^{\mathrm{T}}A^{-1}BA^{-1}a\right] e^{\frac{1}{4}a^{\mathrm{T}}A^{-1}a-\alpha}.$$

(Proof: See [269, p. 322].)

Fact 8.10.16. Let $A \in \mathbb{R}^{n\times n}$ be symmetric, let $b \in \mathbb{R}^n$ and $a \in \mathbb{R}$, and define $f\colon\ \mathbb{R}^n \mapsto \mathbb{R}$ by $f(x) \triangleq x^{\mathrm{T}}Ax + b^{\mathrm{T}}x + a$. Then, f is convex if and only if A is nonnegative semidefinite, while f is strictly convex if and only if A is positive definite. (Remark: *Strictly convex* means that $f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2)$ for all $\alpha \in (0,1)$ and for all $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$.) Furthermore, f has a minimizer if and only if $b \in \mathcal{R}(A)$. The point $x_0 \in \mathbb{R}^n$ is a minimizer of f if and only if x_0 satisfies $2x_0^{\mathrm{T}}A + b^{\mathrm{T}} = 0$. The minimum of f is given by $f(x_0) = c - x_0^{\mathrm{T}}Ax_0$. Furthermore, if A is positive definite, then $x_0 = -\frac{1}{2}A^{-1}b$ is the unique minimizer of f, and the minimum of f is given by $f(x_0) = c - \frac{1}{4}b^{\mathrm{T}}A^{-1}b$.

8.11 Facts on Matrix Transformations

Fact 8.11.1. Let $A \in \mathbb{F}^{n\times n}$. Then, AA^* and A^*A are unitarily similar.

Fact 8.11.2. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, and assume that A is nonsingular. Then, the following statements are equivalent:

i) There exists a unitary matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and SBS^* are diagonal.

ii) $AB = BA$.

iii) $A^{-1}B$ is Hermitian.

(Proof: See [287, p. 229].) (Remark: The equivalence of *i)* and *ii)* is given by Fact 5.8.7.)

Fact 8.11.3. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, and assume that A is nonsingular. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and SBS^* are diagonal if and only if $A^{-1}B$ is diagonalizable over $\mathbb{R}$. (Proof: See [287, p. 229] or [466, p. 95].)

Fact 8.11.4. Let $A, B \in \mathbb{F}^{n\times n}$ be symmetric, and assume that A is nonsingular. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^{T} and SBS^{T} are diagonal if and only if $A^{-1}B$ is diagonalizable. (Proof: See [287, p. 229] and [563].) (Remark: A and B are complex symmetric.)

Fact 8.11.5. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $\{x \in \mathbb{F}^n\colon\ x^*Ax = x^*Bx = 0\} = \{0\}$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and SBS^* are upper triangular. (Proof: See [466, p. 96].) (Remark: See Fact 8.11.6 and Fact 5.8.6.)

Fact 8.11.6. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, and assume that $\{x \in \mathbb{F}^n\colon\ x^*Ax = x^*Bx = 0\} = \{0\}$. Then, there exists a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that SAS^* and SBS^* are diagonal. (Proof: The result follows from Fact 8.11.6. See [389] or [466, p. 96].)

Fact 8.11.7. Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric and nonsingular, and assume there exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha A + \beta B$ is positive definite. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that SAS^{T} and SBS^{T} are diagonal. (Remark: This result is due to Weierstrass. See [563].) (Remark: Suppose that B is positive definite. Then, by necessity of Fact 8.11.3, it follows that $A^{-1}B$ is diagonalizable over $\mathbb{R}$. This proves *iii)* $\Longrightarrow$ *i)* of Proposition 5.5.18.)

Fact 8.11.8. Let $A \in \mathbb{F}^{n\times n}$. Then, A is diagonalizable over $\mathbb{F}$ with (nonnegative, positive) eigenvalues if and only if there exist (nonnegative-semidefinite, positive-definite) matrices $B, C \in \mathbb{F}^{n\times n}$ such that $A = BC$. (Proof: To prove sufficiency, use Theorem 8.3.5 and note that $A = S^{-1} \cdot (SBS^*)\left(S^{-*}CS^{-1}\right)S$.)

8.12 Facts on the Trace

Fact 8.12.1. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and B are both Hermitian or both skew Hermitian. Then, $\operatorname{tr} AB$ is real.

Fact 8.12.2. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, and assume that $-A \le B \le A$. Then,

$$\operatorname{tr} B^2 \le \operatorname{tr} A^2.$$

(Proof: $0 \le \operatorname{tr}[(A-B)(A+B)] = \operatorname{tr} A^2 - \operatorname{tr} B^2$. See [555].)

Fact 8.12.3. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, $AB = 0$ if and only if $\operatorname{tr} AB = 0$.

Fact 8.12.4. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p, q \ge 1$ satisfy $1/p + 1/q = 1$. Then,

$$\operatorname{tr} AB \le (\operatorname{tr} A^p)^{1/p}(\operatorname{tr} B^q)^{1/q}.$$

Furthermore, equality holds if and only if A^{p-1} and B are linearly dependent. (Remark: This result is a matrix version of Holder's inequality.)

Fact 8.12.5. Let $A, B \in \mathbb{F}^{n\times n}$, and let $k \in \mathbb{N}$. Then,

$$|\operatorname{tr}\,(AB)^{2k}| \le \operatorname{tr}\,(A^*ABB^*)^k \le \operatorname{tr}(A^*A)^k(BB^*)^k.$$

(Proof: See [622].)

Fact 8.12.6. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, and let $k \in \mathbb{P}$. Then,

$$|\operatorname{tr}\,(AB)^{2k}| \le \operatorname{tr}\left(A^2B^2\right)^2 \le \begin{cases} \operatorname{tr} A^{2k}B^{2k} \\ \left(\operatorname{tr} A^2B^2\right)^k \end{cases}.$$

(Proof: See [622].)

Fact 8.12.7. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\operatorname{tr} AB \le \left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{1/2}\right]^2 \le (\operatorname{tr} A)(\operatorname{tr} B) \le \tfrac{1}{4}(\operatorname{tr} A + \operatorname{tr} B)^2,$$

(Remark: Note that

$$\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{1/2} = \sum_{i=1}^n \lambda_i^{1/2}(AB)$$

and

$$\operatorname{tr} AB = \operatorname{tr} A^{1/2}BA^{1/2} = \operatorname{tr}\left[\left(A^{1/2}BA^{1/2}\right)^{1/2}\left(A^{1/2}BA^{1/2}\right)^{1/2}\right].$$

The second inequality follows from Proposition 9.3.6 with $p = q = 2$, $r = 1$, and A and B replaced by $A^{1/2}$ and $B^{1/2}$.)

Fact 8.12.8. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p \ge 0$ and $r \ge 1$. Then,

$$\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{pr} \le \operatorname{tr}\left(A^{r/2}B^rA^{r/2}\right)^p.$$

In particular,

$$\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{2p} \le \operatorname{tr}\left(AB^2A\right)^p$$

and

$$\operatorname{tr} AB \le \operatorname{tr}\,(AB^2A)^{1/2}.$$

(Proof: Use Fact 8.14.6 and Fact 8.14.7.) (Remark: This inequality is due to Araki. See [33] and [93, p. 258].) (Problem: Compare the upper bounds

$$\operatorname{tr} AB \le \left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{1/2}\right]^2$$

and

$$\operatorname{tr} AB \le \operatorname{tr}\left(AB^2A\right)^{1/2}.)$$

Fact 8.12.9. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $k, m \in \mathbb{P}$, where $m \ge k$. Then,

$$\operatorname{tr}\left(A^kB^k\right)^m \le \operatorname{tr}\,(A^mB^m)^k.$$

In particular,

$$\operatorname{tr}\,(AB)^m \le \operatorname{tr} A^mB^m.$$

If, in addition, m is even, then

$$\operatorname{tr}\,(AB)^m \le \operatorname{tr}\left(A^2B^2\right)^{m/2} \le \operatorname{tr} A^mB^m.$$

(Proof: Use Fact 8.14.6 and Fact 8.14.7.) (Remark: The result $\operatorname{tr}\,(AB)^m \le \operatorname{tr} A^mB^m$ is the *Lieb-Thirring inequality.* See [93, p. 279]. The inequality $\operatorname{tr}\,(AB)^m \le \operatorname{tr}\left(A^2B^2\right)^{m/2}$ follows from Fact 8.12.8. See [622].) (Problem: Compare the upper bounds

$$\operatorname{tr} AB \le \left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{1/2}\right]^2$$

and

$$\operatorname{tr} AB \le \operatorname{tr}\left(AB^2A\right)^{1/2}.)$$

Fact 8.12.10. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p \ge r \ge 0$. Then,

$$\left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{p}\right]^{1/p} \le \left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{r}\right]^{1/r}.$$

In particular,

$$\left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{2}\right]^{1/2} \le \operatorname{tr} AB \le \begin{cases} \operatorname{tr}\ \left(AB^2A\right)^{1/2} \\ \left[\operatorname{tr}\left(A^{1/2}BA^{1/2}\right)^{1/2}\right]^{2} \end{cases}.$$

(Proof: The result follows from the power sum inequality Fact 1.4.13. See [159].)

Fact 8.12.11. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, assume that $A \le B$, and let $p, q \ge 0$. Then,

$$\operatorname{tr} A^pB^q \le \operatorname{tr} B^{p+q}.$$

If, in addition, A and B are positive definite, then this inequality holds for all $p, q \in \mathbb{R}$ satisfying $q \ge -1$ and $p + q \ge 0$. (Proof: See [107].)

Fact 8.12.12. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $\alpha \in [0, 1]$. Then,

$$\operatorname{tr} A^\alpha B^{1-\alpha} \le (\operatorname{tr} A)^\alpha(\operatorname{tr} B)^{1-\alpha} \le \operatorname{tr}[\alpha A + (1-\alpha)B].$$

Furthermore, the first inequality is an equality if and only if A and B are linearly dependent, while the second inequality is an equality if and only if $A = B$. (Remark: See Fact 1.4.2 and Fact 8.12.13.)

Fact 8.12.13. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite, and let $\alpha \in [0, 1]$. Then,

$$\operatorname{tr} A^{-\alpha}B^{\alpha-1} \le \left(\operatorname{tr} A^{-1}\right)^\alpha\left(\operatorname{tr} B^{-1}\right)^{1-\alpha} \le \operatorname{tr}\left[\alpha A^{-1} + (1-\alpha)B^{-1}\right]$$

and

$$\operatorname{tr}\left[\alpha A + (1-\alpha)B\right]^{-1} \le \left(\operatorname{tr} A^{-1}\right)^\alpha\left(\operatorname{tr} B^{-1}\right)^{1-\alpha} \le \operatorname{tr}\left[\alpha A^{-1} + (1-\alpha)B^{-1}\right].$$

(Remark: The lower inequalities refine the convexity of $\phi(A) = \operatorname{tr} A^{-1}$. See Fact 1.4.2 and Fact 8.12.12.) (Problem: Compare this result to Fact 8.8.20.)

Fact 8.12.14. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that B is nonnegative semidefinite. Then,

$$|\operatorname{tr} AB| \le \sigma_{\max}(A)\operatorname{tr} B.$$

(Proof: Use Proposition 8.4.13 and $\sigma_{\max}(A + A^*) \le 2\sigma_{\max}(A)$.) (Remark: See Fact 5.10.1.)

Fact 8.12.15. Let $A, B \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite, and let $p \geq 1$. Then,

$$[\operatorname{tr}(A^p + B^p)]^{1/p} \leq [\operatorname{tr}(A+B)^p]^{1/p} \leq (\operatorname{tr} A^p)^{1/p} + (\operatorname{tr} B^p)^{1/p}.$$

(Proof: See [107].) (Remark: The first inequality is the *McCarthy inequality.* The second inequality is a special case of the triangle inequality for the norm $\|\cdot\|_{\sigma p}$ and a matrix version of *Minkowski's inequality.*)

Fact 8.12.16. Let $A, B \in \mathbb{F}^{n \times n}$, assume that B is nonnegative semidefinite, and assume that $A^*A \leq B$. Then,

$$\operatorname{tr} A \leq \operatorname{tr} B^{1/2}.$$

(Proof: $\sum_{i=1}^n |\lambda_i| \leq \sum_{i=1}^n \sigma_i(A) = \operatorname{tr}(A^*A)^{1/2} \leq \operatorname{tr} B^{1/2}$. See [71].)

Fact 8.12.17. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be Hermitian. Then, A is nonnegative semidefinite if and only if

$$\operatorname{tr} BA_{12}^* \leq \operatorname{tr}\left(A_{11}^{1/2} B A_{22} B^* A_{11}^{1/2}\right)^{1/2}$$

for all $B \in \mathbb{F}^{n \times m}$. (Proof: See [71].)

Fact 8.12.18. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be nonnegative semidefinite. Then,

$$\operatorname{tr} A_{12}^* A_{12} \leq (\operatorname{tr} A_{11})(\operatorname{tr} A_{22}).$$

(Proof: See [454].)

Fact 8.12.19. Let $A, B \in \mathbb{F}^{n \times n}$ be positive definite. Then,

$$\operatorname{tr}(A - B) \leq \operatorname{tr}[A(\log A - \log B)]$$

and

$$(\log \operatorname{tr} A - \log \operatorname{tr} B)\operatorname{tr} A \leq \operatorname{tr}[A(\log A - \log B)].$$

(Proof: See [93, p. 281] and [69].) (Remark: The second inequality is equivalent to the thermodynamic inequality. See Fact 11.11.22.) (Remark: $\operatorname{tr}[A(\log A - \log B)]$ is the *relative entropy of Umegaki.*)

8.13 Facts on the Determinant

Fact 8.13.1. Let $A \in \mathbb{F}^{n \times n}$ be such that $A + A^*$ is positive definite. Then,

$$\det \tfrac{1}{2}(A + A^*) \leq |\det A|.$$

Furthermore, equality holds if and only if A is Hermitian. (Remark: This result is the *Ostrowski-Taussky inequality.*)

Fact 8.13.2. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A is positive definite and B is Hermitian. Then,

$$\det A \le |\det(A + \jmath B)|.$$

Furthermore, equality holds if and only if $B = 0$. (Proof: See [466, pp. 146, 163].)

Fact 8.13.3. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are positive definite, and assume that $B \le A$. Then,

$$\det A + n \det B \le \det(A + B).$$

(Proof: See [466, pp. 154, 166].)

Fact 8.13.4. Let $A \in \mathbb{F}^{n\times n}$ be such that $\frac{1}{2\jmath}(A - A^*)$ is positive definite. Then,

$$B \triangleq \left[\tfrac{1}{2}(A + A^*)\right]^{1/2} A^{-1} A^* \left[\tfrac{1}{2}(A + A^*)\right]^{-1/2}$$

is unitary. (Proof: See [194].) (Remark: A is *strictly dissipative* if $\frac{1}{2\jmath}(A - A^*)$ is positive definite. A is strictly dissipative if and only if $-\jmath A$ is dissipative. See [192, 193].) (Remark: $A^{-1}A^*$ is similar to a unitary matrix. See Fact 3.6.10.)

Fact 8.13.5. Let $A \in \mathbb{R}^{n\times n}$ be such that $A + A^{\mathrm{T}}$ is positive definite. Then,

$$\left[\det \tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]\left[\tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]^{-1} \le (\det A)\left[\tfrac{1}{2}\big(A^{-1} + A^{-\mathrm{T}}\big)\right].$$

Furthermore,

$$\left[\det \tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]\left[\tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]^{-1} < (\det A)\left[\tfrac{1}{2}\big(A^{-1} + A^{-\mathrm{T}}\big)\right]$$

if and only if $\operatorname{rank}\big(A - A^{\mathrm{T}}\big) \ge 4$. Finally, if $n \ge 4$ and $A - A^{\mathrm{T}}$ is nonsingular, then

$$(\det A)\left[\tfrac{1}{2}\big(A^{-1} + A^{-\mathrm{T}}\big)\right] < \left[\det A - \det \tfrac{1}{2}\big(A - A^{\mathrm{T}}\big)\right]\left[\tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]^{-1}.$$

(Proof: See [193, 310].) (Remark: This result does not hold for complex matrices.) (Problem: If $A + A^{\mathrm{T}}$ is nonnegative semidefinite, does it follow that $\left[\tfrac{1}{2}\big(A + A^{\mathrm{T}}\big)\right]^{\mathrm{A}} \le \tfrac{1}{2}\big(A^{\mathrm{A}} + A^{\mathrm{AT}}\big)$?)

Fact 8.13.6. Let $A, B \in \mathbb{F}^{n\times n}$, assume that B is Hermitian, and assume that $A^*BA < A + A^*$. Then, $\det A \ne 0$.

Fact 8.13.7. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite and let $\alpha \in [0, 1]$.

Then,

$$(\det A)^{\alpha}(\det B)^{1-\alpha} \le \det[\alpha A + (1-\alpha)B].$$

Furthermore, equality holds if and only if $A = B$. (Remark: This result is due to Bergstrom.)

Fact 8.13.8. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are nonnegative semidefinite, assume that $0 \le A \le B$, and let $\alpha \in [0,1]$. Then,

$$\det[\alpha A + (1-\alpha)B] \le \alpha \det A + (1-\alpha)\det B.$$

(Proof: See [588].)

Fact 8.13.9. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and B are positive definite. Then,

$$\frac{\det A}{\det A_{[1,1]}} + \frac{\det B}{\det B_{[1,1]}} \le \frac{\det(A+B)}{\det\big(A_{[1,1]} + B_{[1,1]}\big)}.$$

(Proof: See [466, p. 145].)

Fact 8.13.10. Let $A_1, \ldots, A_k \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$. Then,

$$\det\left(\sum_{i=1}^{k} \lambda_i A_i\right) \le \det\left(\sum_{i=1}^{k} |\lambda_i| A_i\right).$$

(Proof: See [466, p. 144].)

Fact 8.13.11. Let $A, B, C \in \mathbb{R}^{n\times n}$, let $D \triangleq A + \jmath B$, and assume that $CB + B^{\mathrm{T}}C^{\mathrm{T}} < D + D^*$. Then, $\det A \ne 0$.

Fact 8.13.12. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are nonnegative semidefinite, and let $m \in \mathbb{P}$. Then,

$$n^{1/m}(\det AB)^{1/n} \le (\operatorname{tr} A^m B^m)^{1/m}.$$

(Proof: See [159].) (Remark: Assuming $\det B = 1$ and setting $m = 1$ yields Proposition 8.4.14.)

Fact 8.13.13. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$|\det AB^*|^2 \le (\det AA^*)(\det BB^*).$$

(Proof: Apply Fact 8.13.23 to $\left[\begin{smallmatrix} AA^* & AB^* \\ BA^* & BB^* \end{smallmatrix}\right]$.) (Remark: See Fact 8.8.12.)

Fact 8.13.14. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{m\times n}$, where $\operatorname{rank} B = m$. Then,

$$(\det BB^*)^2 \le (\det BAB^*)\det BA^{-1}B^*.$$

(Proof: Use Fact 8.8.13.)

Fact 8.13.15. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{m\times n}$, and $D \in \mathbb{F}^{l\times n}$. Then,

$$|\det(AC+BD)|^2 \le \det(AA^*+BB^*)\det(C^*C+D^*D).$$

(Proof: Use $SS^* \ge 0$, where $S \triangleq \begin{bmatrix} A & B \\ C^* & D^* \end{bmatrix}$.)

Fact 8.13.16. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$|\det(I+AB^*)|^2 \le \det(I+AA^*)\det(I+BB^*).$$

(Proof: Specialize Fact 8.13.15.)

Fact 8.13.17. Let $A, B \in \mathbb{F}^{n\times n}$, assume that $A+A^* > 0$ and $B+B^* \ge 0$, and let $\alpha > 0$. Then,

$$\operatorname{mspec}(\alpha I + AB) \cap (-\infty, 0] = \varnothing.$$

Hence,

$$\det(\alpha I + AB) > 0.$$

(Proof: See [254].) (Remark: Equivalently, $-A$ is dissipative and $-B$ is semidissipative.) (Problem: Find a positive lower bound for $\det(\alpha I + AB)$ in terms of α, A, and B.)

Fact 8.13.18. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\begin{bmatrix} I+A^*A & (A+B)^* \\ A+B & I+BB^* \end{bmatrix} = \begin{bmatrix} I & A^* \\ B & I \end{bmatrix} \begin{bmatrix} I & B^* \\ A & I \end{bmatrix} \ge 0$$

and

$$(A+B)^*(I+BB^*)^{-1}(A+B) \le I + A^*A.$$

If, in addition, $n = m$, then

$$|\det(A+B)|^2 \le \det(I+A^*A)\det(I+BB^*).$$

(Proof: See [630].)

Fact 8.13.19. Let $A, B \in \mathbb{F}^{n\times m}$. Then, there exist unitary matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that

$$I + \langle A+B\rangle \le S_1(I+\langle A\rangle)^{1/2}S_2(I+\langle B\rangle)S_2^*(I+\langle A\rangle)^{1/2}S_1^*.$$

Therefore,

$$\det(I+\langle A+B\rangle) \le \det(I+\langle A\rangle)\det(I+\langle B\rangle).$$

(Proof: See [24, 545].) (Remark: This result is due to Seiler and Simon.)

Fact 8.13.20. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $I - A^*A$ and $I - B^*B$ are positive definite. Then,

$$\begin{bmatrix} (I-A^*A)^{-1} & (I-B^*A)^{-1} \\ (I-A^*B)^{-1} & (I-B^*B)^{-1} \end{bmatrix} \geq 0,$$

$$I - B^*B \leq (I-B^*A)(I-A^*A)^{-1}(I-A^*B),$$

$$0 < \det(I-A^*A)\det(I-B^*B) \leq [\det(I-A^*B)]^2.$$

(Remark: These results are *Hua's inequalities*. See [24].)

Fact 8.13.21. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$|\det A| \leq \prod_{i=1}^{n}\left(\sum_{j=1}^{n}|A_{(i,j)}|^2\right)^{1/2}.$$

Furthermore, equality holds if and only if AA^* is diagonal. (Remark: Replace A with AA^* in Fact 8.14.5.)

Fact 8.13.22. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be positive definite. Then,

$$\begin{aligned}\det A &= (\det A_{11})\det\big(A_{22} - A_{12}^*A_{11}^{-1}A_{12}\big)\\ &\leq (\det A_{11})\det A_{22}\\ &\leq \prod_{i=1}^{n+m}A_{(i,i)}.\end{aligned}$$

If, in addition, $n = m$, then

$$0 < (\det A_{11})\det A_{22} - |\det A_{12}|^2 \leq \det A \leq (\det A_{11})\det A_{22}.$$

(Proof: Since $0 \leq A_{12}^*A_{11}^{-1}A_{12} < A_{22}$, it follows that $|\det A_{12}|^2/\det A_{11} < \det A_{22}$. Use Fact 8.13.23. Also, see [466, p. 142].) (Remark: $\det A \leq (\det A_{11})\det A_{22}$ is *Fischer's inequality*.)

Fact 8.13.23. Let $A = \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\ A_{1k}^{\mathrm{T}} & \cdots & A_{kk} \end{bmatrix}$ be nonnegative semidefinite, where $A_{ij} \in \mathbb{F}^{n\times n}$ for all $i, j = 1, \ldots, k$. Then,

$$\det\begin{bmatrix} \det A_{11} & \cdots & \det A_{1k} \\ \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\ \det A_{1k} & \cdots & \det A_{kk} \end{bmatrix} \leq \det A$$

and

$$\begin{bmatrix} \operatorname{tr} A_{11} & \cdots & \operatorname{tr} A_{1k} \\ \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\ \operatorname{tr} A_{1k} & \cdots & \operatorname{tr} A_{kk} \end{bmatrix} \geq 0.$$

(Remark: The matrix whose (i,j) entry is $\det A_{ij}$ is a *determinantal compression* of A. See [165, 166, 454, 543].)

8.14 Facts on Eigenvalues and Singular Values

Fact 8.14.1. Let $A \in \mathbb{F}^{n\times m}$. Then,

$$\operatorname{tr}\langle A\rangle = \sum_{i=1}^{\min\{n,m\}} \sigma_i(A).$$

Fact 8.14.2. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $i = 1,\ldots,n$,

$$\left|\lambda_i\left[\tfrac{1}{2}(A + A^*)\right]\right| \leq \sigma_i(A).$$

Hence,

$$|\operatorname{tr} A| \leq \operatorname{tr}\langle A\rangle.$$

(Proof: See [289, p. 151] or [516].)

Fact 8.14.3. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1,\cdots,\lambda_n\}_{\mathrm{m}}$, where $\lambda_1,\ldots,\lambda_n$ are ordered such that $|\lambda_1| \geq \cdots \geq |\lambda_n|$. If $r > 0$ or $r \in \mathbb{R}$ and A is nonsingular, then, for all $k = 1,\ldots,n$,

$$\sum_{i=1}^{k} |\lambda_i|^r \leq \sum_{i=1}^{k} \sigma_i^r(A).$$

In particular, for all $k = 1,\ldots,n$,

$$\sum_{i=1}^{k} |\lambda_i| \leq \sum_{i=1}^{k} \sigma_i(A).$$

Hence,

$$|\operatorname{tr} A| \leq \sum_{i=1}^{n} |\lambda_i| \leq \sum_{i=1}^{n} \sigma_i(A) = \operatorname{tr}\langle A\rangle.$$

Furthermore, for all $k = 1,\ldots,n$,

$$\sum_{i=1}^{k} |\lambda_i|^2 \leq \sum_{i=1}^{k} \sigma_i^2(A).$$

Hence,

$$|\mathrm{tr}\, A^2| \le \sum_{i=1}^n |\lambda_i|^2 \le \sum_{i=1}^n \sigma_i\big(A^2\big) = \mathrm{tr}\,\langle A^2\rangle \le \sum_{i=1}^n \sigma_i^2(A) = \mathrm{tr}\, A^*A.$$

(Proof: The result follows from Fact 8.16.5 and Fact 5.9.13. See [93, p. 42], [289, p. 176], or [625, p. 19]. See Fact 9.11.15 for the inequality $\mathrm{tr}\,\langle A^2\rangle = \mathrm{tr}\,\big(A^{2*}A^2\big)^{1/2} \le \mathrm{tr}\, A^*A$.) Finally,

$$\sum_{i=1}^n |\lambda_i|^2 = \mathrm{tr}\, A^*A$$

if and only if A is normal. (Proof: See [466, p. 146].) (Remark: $\sum_{i=1}^n |\lambda_i|^2 \le \mathrm{tr}\, A^*A$ is *Schur's inequality.* See Fact 9.10.2.) (Problem: Determine when equality holds for the remaining inequalities.)

Fact 8.14.4. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, for all $k = 1,\ldots,n$,

$$\sum_{i=1}^k d_i(A) \le \sum_{i=1}^k \lambda_i(A)$$

with equality for $k = n$, that is,

$$\mathrm{tr}\, A = \sum_{i=1}^n d_i(A) = \sum_{i=1}^n \lambda_i(A).$$

Hence, for all $k = 1,\ldots,n$,

$$\sum_{i=k}^n \lambda_i(A) \le \sum_{i=k}^n d_i(A).$$

(Proof: See [93, p. 35], [287, p. 193], or [625, p. 18].) (Remark: This result is *Schur's theorem.*)

Fact 8.14.5. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, for all $k = 1,\ldots,n$,

$$\prod_{i=k}^n \lambda_i(A) \le \prod_{i=k}^n d_i(A).$$

In particular,

$$\det A \le \prod_{i=1}^n A_{(i,i)}.$$

Now, assume that A is positive definite. Then, equality holds if and only if A is diagonal. (Proof: See [287, p. 200], [625, p. 18], and [287, p. 477].) (Remark: The case $k = n$ is *Hadamard's inequality.*)

Fact 8.14.6. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. If $p \ge 1$, then

$$\sum_{i=1}^{n} \lambda_i^p(A)\lambda_{n-i+1}^p(B) \le \operatorname{tr}\,(AB)^p \le \operatorname{tr} A^pB^p \le \sum_{i=1}^{n} \lambda_i^p(A)\lambda_i^p(B).$$

If $0 \le p \le 1$, then

$$\sum_{i=1}^{n} \lambda_i^p(A)\lambda_{n-i+1}^p(B) \le \operatorname{tr} A^pB^p \le \operatorname{tr}\,(AB)^p \le \sum_{i=1}^{n} \lambda_i^p(A)\lambda_i^p(B).$$

Now, suppose that A and B are positive definite. If $p \le -1$, then

$$\sum_{i=1}^{n} \lambda_i^p(A)\lambda_{n-i+1}^p(B) \le \operatorname{tr}\,(AB)^p \le \operatorname{tr} A^pB^p \le \sum_{i=1}^{n} \lambda_i^p(A)\lambda_i^p(B).$$

If $-1 \le p \le 0$, then

$$\sum_{i=1}^{n} \lambda_i^p(A)\lambda_{n-i+1}^p(B) \le \operatorname{tr} A^pB^p \le \operatorname{tr}\,(AB)^p \le \sum_{i=1}^{n} \lambda_i^p(A)\lambda_i^p(B).$$

(Proof: See [578]. See also [122, 358, 374, 581].) (Remark: See Fact 8.12.8. See Fact 8.12.5 for the indefinite case.)

Fact 8.14.7. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p \ge r \ge 0$. Then,

$$\begin{bmatrix} \lambda_1^{1/p}(A^pB^p) & \cdots & \lambda_n^{1/p}(A^pB^p) \end{bmatrix}$$

weakly log majorizes and thus weakly majorizes

$$\begin{bmatrix} \lambda_1^{1/r}(A^rB^r) & \cdots & \lambda_n^{1/r}(A^rB^r) \end{bmatrix}.$$

(Proof: See [93, p. 257] or [625, p. 20] and Fact 8.16.5.)

Fact 8.14.8. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\lambda_{\max}(A+B) \le \max\{\lambda_{\max}(A), \lambda_{\max}(B)\} + \lambda_{\max}\Big(A^{1/2}B^{1/2}\Big).$$

(Proof: See [335].)

Fact 8.14.9. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\lambda_{\max}(A+B) \le \tfrac{1}{2}\left[\lambda_{\max}(A) + \lambda_{\max}(B) + \sqrt{[\lambda_{\max}(A) - \lambda_{\max}(B)]^2 + 4\lambda_{\max}^2\big(A^{1/2}B^{1/2}\big)}\right].$$

(Proof: See [337].)

Fact 8.14.10. Let $f\colon\ \mathbb{R} \mapsto \mathbb{R}$ be convex, and let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, for all $\alpha \in [0,1]$,

$$\left[\ \alpha\lambda_1 f(A) + (1-\alpha)\lambda_1 f(B) \quad \cdots \quad \alpha\lambda_n f(A) + (1-\alpha)\lambda_n f(B)\ \right]$$

weakly majorizes

$$\left[\ \lambda_1 f(\alpha A + (1-\alpha)B) \quad \cdots \quad \lambda_n f(\alpha A + (1-\alpha)B)\ \right].$$

If, in addition, f is either nonincreasing or nondecreasing, then, for all $i = 1,\ldots,n$,

$$\lambda_i f(\alpha A + (1-\alpha)B) \le \alpha\lambda_i f(A) + (1-\alpha)\lambda_i f(B).$$

(Proof: See [42].)

Fact 8.14.11. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. If $r \in [0,1]$, then

$$\left[\ \lambda_1(A^r + B^r) \quad \cdots \quad \lambda_n(A^r + B^r)\ \right]$$

weakly majorizes

$$\left[\ \lambda_1[(A+B)^r] \quad \cdots \quad \lambda_n[(A+B)^r]\ \right],$$

and, for all $i = 1,\ldots,n$,

$$2^{1-r}\lambda_i[(A+B)^r] \le \lambda_i(A^r + B^r).$$

If $r \ge 1$, then

$$\left[\ \lambda_1[(A+B)^r] \quad \cdots \quad \lambda_n[(A+B)^r]\ \right]$$

weakly majorizes

$$\left[\ \lambda_1(A^r + B^r) \quad \cdots \quad \lambda_n(A^r + B^r)\ \right],$$

and, for all $i = 1,\ldots,n$,

$$\lambda_i(A^r + B^r) \le 2^{r-1}\lambda_i[(A+B)^r].$$

(Proof: The result follows from Fact 8.14.10. See [29, 41, 42].)

Fact 8.14.12. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian and let $S \in \mathbb{R}^{k\times n}$ satisfy $SS^* = I_k$. Then, for all $i = 1,\ldots,k$,

$$\lambda_{i+n-k}(A) \le \lambda_i(SAS^*) \le \lambda_i(A).$$

Consequently,

$$\sum_{i=1}^k \lambda_{i+n-k}(A) \le \operatorname{tr} SAS^* \le \sum_{i=1}^k \lambda_i(A)$$

and

$$\prod_{i=1}^k \lambda_{i+n-k}(A) \le \det SAS^* \le \prod_{i=1}^k \lambda_i(A).$$

(Proof: See [287, p. 190].) (Remark: This result is the *Poincare separation theorem.*)

Fact 8.14.13. Let $A \in \mathbb{F}^{n\times n}$ be Hermitian. Then, for all $k = 1,\ldots,n$,

$$\sum_{i=n+1-i}^{n} \lambda_i = \min\{\operatorname{tr} S^*AS\colon\ S \in \mathbb{F}^{n\times k} \text{ and } S^*S = I_k\}.$$

(Proof: See [289, p. 191].) (Remark: This result is the *minimum principle.*)

Fact 8.14.14. Let $A \in \mathbb{F}^{n\times n}$. Then, $\begin{bmatrix} I & A \\ A^* & I \end{bmatrix}$ is nonnegative semidefinite if and only if $\sigma_{\max}(A) \le 1$. Furthermore, $\begin{bmatrix} I & A \\ A^* & I \end{bmatrix}$ is positive definite if and only if $\sigma_{\max}(A) < 1$. (Proof: Note that

$$\begin{bmatrix} I & A \\ A^* & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ A^* & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I - A^*A \end{bmatrix} \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}.)$$

Fact 8.14.15. Let $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be nonnegative semidefinite. Then,

$$\sigma_{\max}^2(A_{12}) \le \sigma_{\max}(A_{11})\sigma_{\max}(A_{22}).$$

(Proof: Use $A_{22} \ge A_{12}^*A_{11}^+A_{12} \ge 0$, factor $A_{11}^+ = MM^*$, where M has full column rank, and recall that $\sigma_{\max}(SS^*) = \sigma_{\max}^2(S)$.) (Problem: Consider alternative norms.)

Fact 8.14.16. Let $A, B \in \mathbb{F}^{n\times m}$ be nonnegative semidefinite. Then, for all $k = 1,\ldots,n$,

$$\prod_{i=1}^{k} \lambda_i(AB) \le \prod_{i=1}^{k} \sigma_i(AB) \le \prod_{i=1}^{k} \lambda_i(A)\lambda_i(B)$$

with equality for $k = n$. Furthermore, for all $k = 1,\ldots,n$,

$$\prod_{i=k}^{n} \lambda_i(A)\lambda_i(B) \le \prod_{i=k}^{n} \sigma_i(AB) \le \prod_{i=k}^{n} \lambda_i(AB).$$

(Proof: Use Fact 5.9.13 and Fact 9.11.16.)

Fact 8.14.17. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite. If $q \ge 1$, then

$$\sigma_{\max}^q(AB) \le \sigma_{\max}(A^qB^q).$$

If $p \ge q > 0$, then

$$\sigma_{\max}^{1/q}(A^qB^q) \le \sigma_{\max}^{1/p}(A^pB^p).$$

(Proof: See [219].)

8.15 Facts on the Schur and Kronecker Products

Fact 8.15.1. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and assume that every entry of A is nonzero. Then, $A^{\{-1\}}$ is nonnegative semidefinite if and only if $\operatorname{rank} A = 1$. (Proof: See [363].)

Fact 8.15.2. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and let $k \in \mathbb{P}$. If $r \in [0,1]$, then

$$(A^r)^{\{k\}} \le \left(A^{\{k\}}\right)^r.$$

If $r \in [1,2]$, then

$$\left(A^{\{k\}}\right)^r \le (A^r)^{\{k\}}.$$

If A is positive definite and $r \in [0,1]$, then

$$\left(A^{\{k\}}\right)^{-r} \le \left(A^{-r}\right)^{\{k\}}.$$

(Proof: See [625, p. 8].)

Fact 8.15.3. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$(I \circ A)^2 \le \tfrac{1}{2}(I \circ A^2 + A \circ A) \le I \circ A^2.$$

Now, assume that A is positive definite. Then,

$$\left(A \circ A^{-1}\right)^{-1} \le I \le \left(A^{1/2} \circ A^{-1/2}\right)^2 \le \tfrac{1}{2}\left(I + A \circ A^{-1}\right) \le A \circ A^{-1},$$
$$(A \circ A)^{-1} \le A^{-1} \circ A^{-1},$$

and

$$1 \in \operatorname{spec}\left(A \circ A^{-1}\right).$$

Define $\Phi(A) \triangleq A \circ A^{-1}$ and, for all $k \in \mathbb{P}$, define

$$\Phi^{(k+1)}(A) \triangleq \Phi\left[\Phi^{(k)}(A)\right],$$

where $\Phi^{(1)}(A) \triangleq \Phi(A)$. Then, for all $k \in \mathbb{P}$,

$$\Phi^{(k)}(A) \ge I$$

and

$$\lim_{k\to\infty} \Phi^{(k)}(A) = I.$$

(Proof: See [201, 316, 577] and [287, p. 475].) (Remark: The convergence result also holds if A is an H-matrix [316]. $A \circ A^{-1}$ is the *relative gain array.*)

Fact 8.15.4. Let $A \triangleq \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ and $B \triangleq \begin{bmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix} \in \mathbf{P}^{(n+m)\times(n+m)}$, and assume that A and B are nonnegative semidefinite. Then,

$$(A_{11}|A) \circ (B_{11}|B) \leq (A_{11}|A) \circ B_{22} \leq (A_{11} \circ B_{11})|(A \circ B).$$

(Proof: See [369].)

Fact 8.15.5. Let $A \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite and assume that $I_n \circ A = I_n$. Then,

$$\det A \leq \lambda_{\min}(A \circ A).$$

(Proof: See [589].)

Fact 8.15.6. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and B are nonnegative semidefinite. Then, $A \circ B$ is nonnegative semidefinite. If, in addition, B is positive definite and all of the diagonal entries of A are positive, then $A \circ B$ is positive definite. (Proof: By Fact 7.4.13, $A \otimes B$ is nonnegative semidefinite, and the Schur product $A \circ B$ is a principal submatrix of the Kronecker product. If A is positive definite, use Fact 8.15.12 to obtain $\det(A \circ B) > 0$.) (Remark: The first result is *Schur's theorem*.)

Fact 8.15.7. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is positive definite. Then, there exist positive-definite matrices $B, C \in \mathbb{F}^{n\times n}$ such that $A = B \circ C$. (Remark: See [466, pp. 154, 166].) (Remark: This result is due to Djokovic.)

Fact 8.15.8. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\left(1_{1\times n} A^{-1} 1_{n\times 1}\right)^{-1} B \leq A \circ B.$$

(Proof: See [204].)

Fact 8.15.9. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\operatorname{rank} B \leq \operatorname{rank}(A \circ B) \leq \operatorname{rank}(A \otimes B) = (\operatorname{rank} A)(\operatorname{rank} B).$$

(Remark: See Fact 7.4.20, Fact 7.5.5, and Fact 8.15.8.) (Remark: The first inequality is due to Djokovic. See [466, pp. 154, 166].)

Fact 8.15.10. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. If $p \geq 1$, then

$$\operatorname{tr}\,(A \circ B)^p \leq \operatorname{tr} A^p \circ B^p.$$

If $0 \leq p \leq 1$, then

$$\operatorname{tr} A^p \circ B^p \leq \operatorname{tr}\,(A \circ B)^p.$$

Now, assume that A and B are positive definite. If $p \leq 0$, then

$$\operatorname{tr}\,(A \circ B)^p \leq \operatorname{tr} A^p \circ B^p.$$

(Proof: See [581].)

Fact 8.15.11. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then, for all $k = 1, \ldots, n$,

$$\prod_{i=k}^{n}\lambda_i(A)\lambda_i(B) \le \prod_{i=k}^{n}\sigma_i(AB) \le \prod_{i=k}^{n}\lambda_i(AB) \le \prod_{i=k}^{n}\lambda_i^2(A\#B) \le \prod_{i=k}^{n}\lambda_i(A\circ B).$$

Consequently,

$$\lambda_{\min}(AB)I \le A\circ B$$

and

$$\det AB = [\det(A\#B)]^2 \le \det(A\circ B).$$

(Proof: See [25, 201], [625, p. 21], and Fact 8.14.16.)

Fact 8.15.12. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$\det AB \le \left(\prod_{i=1}^{n} A_{(i,i)}\right)\det B \le \det(A\circ B).$$

If, in addition, A and B are positive definite, then the right-hand inequality is an equality if and only if B is diagonal. (Proof: See [397].) (Remark: The left-hand inequality follows from Hadamard's inequality Fact 8.14.5. The right-hand inequality is *Oppenheim's inequality.*) (Problem: Compare $\left(\prod_{i=1}^{n} A_{(i,i)}\right)\det B$ and $[\det(A\#B)]^2$.)

Fact 8.15.13. Let $A_1, A_2, B_1, B_2 \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and assume that $0 \le A_1 \le B_1$ and $0 \le A_2 \le B_2$. Then,

$$0 \le A_1 \otimes A_2 \le B_1 \otimes B_2$$

and

$$0 \le A_1 \circ A_2 \le B_1 \circ B_2.$$

(Proof: See [23].) (Problem: Under which conditions are these inequalities strict?)

Fact 8.15.14. Let $A_1, \ldots, A_k, B_1, \ldots, B_k \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$(A_1 + B_1) \otimes \cdots \otimes (A_k + B_k) \le A_1 \otimes \cdots \otimes A_k + B_1 \otimes \cdots \otimes B_k.$$

(Proof: See [412, p. 143].)

Fact 8.15.15. Let $A_1, A_2, B_1, B_2 \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, assume that $0 \le A_1 \le B_1$ and $0 \le A_2 \le B_2$, and let $\alpha \in [0, 1]$. Then,

$$[\alpha A_1 + (1-\alpha)B_1] \otimes [\alpha A_2 + (1-\alpha)B_2] \le \alpha(A_1 \otimes A_2) + (1-\alpha)(B_1 \otimes B_2).$$

(Proof: See [588].)

Fact 8.15.16. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, for all $i = 1,\ldots,n$,

$$\lambda_n(A)\lambda_n(B) \le \lambda_{i+n^2-n}(A\otimes B) \le \lambda_i(A\circ B) \le \lambda_i(A\otimes B) \le \lambda_1(A)\lambda_1(B).$$

(Proof: The result follows from Proposition 7.3.1 and Theorem 8.4.5. For A, B nonnegative semidefinite, the result is given in [394].)

Fact 8.15.17. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, assume that $0 \le A \le B$, and let $k \in \mathbb{P}$. Then,

$$A^{\{k\}} \le B^{\{k\}}.$$

(Proof: $0 \le (B - A)\circ(B + A)$ implies $A \circ A \le B \circ B$.)

Fact 8.15.18. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. If $r \in [0, 1]$, then

$$A^r \circ B^r \le (A \circ B)^r.$$

If $r \in [1, 2]$, then

$$(A \circ B)^r \le A^r \circ B^r.$$

If A and B are positive definite and $r \in [0, 1]$, then

$$(A \circ B)^{-r} \le A^{-r} \circ B^{-r}.$$

Therefore,

$$(A \circ B)^2 \le A^2 \circ B^2,$$
$$A \circ B \le \left(A^2 \circ B^2\right)^{1/2},$$
$$A^{1/2} \circ B^{1/2} \le (A \circ B)^{1/2}.$$

Furthermore,

$$A^2 \circ B^2 - \tfrac{1}{4}(\beta - \alpha)^2 I \le (A \circ B)^2 \le \tfrac{1}{2}\Big[A^2 \circ B^2 + (AB)^{\{2\}}\Big] \le A^2 \circ B^2$$

and

$$A \circ B \le \left(A^2 \circ B^2\right)^{1/2} \le \frac{\alpha + \beta}{2\sqrt{\alpha\beta}} A \circ B,$$

where $\alpha \triangleq \lambda_{\min}(A \otimes B)$ and $\beta \triangleq \lambda_{\max}(A \otimes B)$. Hence,

$$\begin{aligned} A \circ B - \tfrac{1}{4}\Big(\sqrt{\beta} - \sqrt{\alpha}\Big)^2 I &\le \Big(A^{1/2} \circ B^{1/2}\Big)^2 \\ &\le \tfrac{1}{2}\Big[A \circ B + \Big(A^{1/2}B^{1/2}\Big)^{\{2\}}\Big] \\ &\le A \circ B \\ &\le \big(A^2 \circ B^2\big)^{1/2} \\ &\le \frac{\alpha + \beta}{2\sqrt{\alpha\beta}} A \circ B. \end{aligned}$$

(Proof: See [23, 427, 577], [287, p. 475], and [625, p. 8].)

Fact 8.15.19. Let $A, B \in \mathbb{F}^{n \times n}$ be nonnegative semidefinite and let $p, q \in [1, \infty)$ be such that $p \le q$. Then,

$$(A^p \circ B^p)^{1/p} \le (A^q \circ B^q)^{1/q}.$$

(Proof: Since $p/q \le 1$, it follows from Fact 8.15.18 that $A^p \circ B^p = (A^q)^{p/q} \circ (A^q)^{p/q} \le (A^q \circ B^q)^{p/q}$. Then, use Corollary 8.5.8 with p replaced by $1/p$. See [625, p. 8].)

Fact 8.15.20. Let $A, B \in \mathbb{F}^{n \times n}$ be positive definite and let p, q be nonzero integers such that $p \le q$. Then,

$$(A^p \circ B^p)^{1/p} \le (A^q \circ B^q)^{1/q}.$$

In particular,

$$\big(A^{-1} \circ B^{-1}\big)^{-1} \le A \circ B,$$
$$(A \circ B)^{-1} \le A^{-1} \circ B^{-1},$$

and, for all $p \in \mathbb{P}$,

$$A \circ B \le (A^p \circ B^p)^{1/p},$$
$$A^{1/p} \circ B^{1/p} \le (A \circ B)^{1/p}.$$

Furthermore,

$$(A \circ B)^{-1} \le A^{-1} \circ B^{-1} \le \frac{(\alpha + \beta)^2}{4\alpha\beta}(A \circ B)^{-1},$$

where $\alpha \triangleq \lambda_{\min}(A \otimes B)$ and $\beta \triangleq \lambda_{\max}(A \otimes B)$. (Proof: See [427].) (Problem: Consider real numbers $p \le q \le -1$ to unify this result with Fact 8.15.19.)

Fact 8.15.21. Let $A, B \in \mathbb{F}^{n \times n}$ be positive definite. Then,

$$I \circ (\log A + \log B) \le \log(A \circ B).$$

(Proof: See [23, 625].) (Remark: See Fact 11.11.20.)

Fact 8.15.22. Let $A, B \in \mathbb{F}^{n\times n}$ be positive definite, and let $C, D \in \mathbb{F}^{m\times n}$. Then,

$$(C \circ D)(A \circ B)^{-1}(C \circ D)^* \le \big(CA^{-1}C^*\big)\circ\big(DB^{-1}D^*\big).$$

In particular,

$$(A \circ B)^{-1} \le A^{-1} \circ B^{-1}$$

and

$$(C \circ D)(C \circ D)^* \le (CC^*) \circ (DD^*).$$

(Proof: Form the Schur complement $A_{22\mathrm{c}}$ of the Schur product of the nonnegative-semidefinite matrices $\left[\begin{smallmatrix} A & C^* \\ C & CA^{-1}C^* \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} B & D^* \\ D & DB^{-1}D^* \end{smallmatrix}\right]$. See [396,582] or [625, p. 13].)

Fact 8.15.23. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p, q \in (1, \infty)$ satisfy $1/p + 1/q = 1$. Then,

$$(A \circ B) + (C \circ D) \le (A^p + C^p)^{1/p} \circ (B^q + D^q)^{1/q}.$$

(Proof: Use *xxiv*) of Proposition 8.5.13 with $r = 1/p$. See [625, p. 10].) (Remark: Note the relationship between the *conjugate parameters* p, q and the *barycentric coordinates* $\alpha, 1 - \alpha$. See Fact 1.4.16.)

Fact 8.15.24. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$(A \circ B)(A \circ B)^* \le \tfrac{1}{2}(AA^* \circ BB^* + AB^* \circ BA^*) \le AA^* \circ BB^*.$$

(Proof: See [291, 577].)

8.16 Facts on Majorization

Fact 8.16.1. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \ge \cdots \ge x_{(n)}$ and $y_{(1)} \ge \cdots \ge y_{(n)}$, assume that y strongly majorizes x, let $f\colon\ [\min\{x_{(n)}, y_{(n)}\}, y_{(1)}] \mapsto \mathbb{R}$, and assume that f is convex. Then, $\begin{bmatrix} f(y_{(1)}) & \cdots & f(y_{(n)}) \end{bmatrix}^{\mathrm{T}}$ weakly majorizes $\begin{bmatrix} f(x_{(1)}) & \cdots & f(x_{(n)}) \end{bmatrix}^{\mathrm{T}}$. (Proof: See [93, p. 42], [289, p. 173], or [400, p. 116].)

Fact 8.16.2. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \ge \cdots \ge x_{(n)} \ge 0$ and $y_{(1)} \ge \cdots \ge y_{(n)} \ge 0$, assume that y strongly log majorizes x, let $f\colon\ [0, \infty) \mapsto \mathbb{R}$, and assume that $g(z) \triangleq f(e^z)$ is convex. Then, $\begin{bmatrix} f(y_{(1)}) & \cdots & f(y_{(n)}) \end{bmatrix}^{\mathrm{T}}$ weakly majorizes $\begin{bmatrix} f(x_{(1)}) & \cdots & f(x_{(n)}) \end{bmatrix}^{\mathrm{T}}$. (Proof: Apply Fact 8.16.1.)

Fact 8.16.3. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \ge \cdots \ge x_{(n)}$ and $y_{(1)} \ge \cdots \ge y_{(n)}$, assume that y weakly majorizes x, let $f\colon\ [\min\{x_{(n)}, y_{(n)}\}, y_{(1)}] \mapsto \mathbb{R}$, and assume that f is convex and increasing. Then, $\begin{bmatrix} f(y_{(1)}) & \cdots & f(y_{(n)}) \end{bmatrix}^{\mathrm{T}}$

weakly majorizes $\left[\begin{array}{ccc} f(x_{(1)}) & \cdots & f(x_{(n)}) \end{array}\right]^{\mathrm{T}}$. (Proof: See [93, p. 42], [289, p. 173], or [400, p. 116].)

Fact 8.16.4. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \geq \cdots \geq x_{(n)} \geq 0$ and $y_{(1)} \geq \cdots \geq y_{(n)} \geq 0$, assume that y log majorizes x, let $f\colon [0,\infty) \mapsto \mathbb{R}$, and assume that $g(z) \triangleq f(e^z)$ is convex and increasing. Then, $\left[\begin{array}{ccc} f(y_{(1)}) & \cdots & f(y_{(n)}) \end{array}\right]^{\mathrm{T}}$ weakly majorizes $\left[\begin{array}{ccc} f(x_{(1)}) & \cdots & f(x_{(n)}) \end{array}\right]^{\mathrm{T}}$. (Proof: Use Fact 8.16.3.)

Fact 8.16.5. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \geq \cdots \geq x_{(n)} \geq 0$ and $y_{(1)} \geq \cdots \geq y_{(n)} \geq 0$, and assume that y weakly log majorizes x. Then, y weakly majorizes x. (Proof: Use Fact 8.16.3 with $f(t) = e^t$. See [625, p. 19].)

Fact 8.16.6. Let $x, y \in \mathbb{R}^n$, where $x_{(1)} \geq \cdots \geq x_{(n)} \geq 0$ and $y_{(1)} \geq \cdots \geq y_{(n)} \geq 0$, assume that y weakly majorizes x, let $p \in [1,\infty)$, and let $r > 0$. Then, for all $k = 1, \ldots, n$,

$$\left(\sum_{i=1}^{k} x_{(i)}^{p}\right)^{r} \leq \left(\sum_{i=1}^{k} y_{(i)}^{p}\right)^{r}.$$

(Proof: Use Fact 8.16.3. See [400, p. 96].) (Remark: $\phi(x) \triangleq \left(\sum_{i=1}^{k} x_{(i)}^{p}\right)^{1/p}$ is a *symmetric gauge function*.)

8.17 Notes

The ordering $A \leq B$ is traditionally called the *Loewner ordering*. Proposition 8.2.3 is given in [5] and [342] with extensions in [71]. The proof of Proposition 8.2.6 is based on [113, p. 120], as suggested in [533]. The proof given in [222, p. 307] is incomplete.

Theorem 8.3.4 is due to Newcomb [437].

Proposition 8.4.13 is given in [284, 429]. Special cases such as Fact 8.12.14 appear in numerous papers.

The proofs of Lemma 8.4.4 and Theorem 8.4.5 are based on [525]. Theorem 8.4.9 can also be obtained as a corollary of the *Fischer minimax theorem* given in [287, 400], which provides a geometric characterization of the eigenvalues of a symmetric matrix. Theorem 8.3.5 appears in [477, p. 121]. Theorem 8.5.2 is given in [21]. Additional inequalities appear in [422].

Functions that are nondecreasing on $\mathbf{P}^n$ are characterized by the theory of *monotone matrix functions* [93, 184]. See [425] for a summary of the principal results.

The literative on convex maps is extensive. Result *xiv*) of Proposition 8.5.13 is due to Lieb and Ruskai [373]. Result *xxi*) is the *Lieb concavity theorem* [372]. Result *xxxiii*) is due to Ando. Results *xxxv*) and *xxxvi*) are due to Fan. Some extensions to strict convexity are considered in [400]. See also [23, 411, 431].

Products of positive-definite matrices are studied in [48–51, 617]. Alternative orderings for nonnegative-semidefinite matrices are considered in [46, 267].

Essays on the legacy of Issai Schur appear in [318].

Chapter Nine

Norms

Norms are used to quantify vectors and norms, and they play a basic role in convergence analysis. This chapter introduces vector and matrix norms and their numerous properties.

9.1 Vector Norms

For $\alpha \in \mathbb{F}$, let $|\alpha|$ denote the absolute value of α. For $x \in \mathbb{F}^n$ and $A \in \mathbb{F}^{n\times m}$, every component of x and every entry of A can be replaced by its absolute value to obtain $|x| \in \mathbb{R}^n$ and $|A| \in \mathbb{R}^{n\times m}$ defined by

$$|x|_{(i)} \triangleq |x_{(i)}| \tag{9.1.1}$$

for all $i = 1,\dots,n$ and

$$|A|_{(i,j)} \triangleq |A_{(i,j)}| \tag{9.1.2}$$

for all $i = 1,\dots,n$ and $j = 1,\dots,m$. For many applications it is useful to have a scalar measure of the magnitude of x or A. *Norms* provide such measures.

Definition 9.1.1. A *norm* $\|\cdot\|$ on $\mathbb{F}^n$ is a function $\|\cdot\|\colon\ \mathbb{F}^n \mapsto \mathbb{R}$ that satisfies the following conditions:

i) $\|x\| \ge 0$ for all $x \in \mathbb{F}^n$.

ii) $\|x\| = 0$ if and only if $x = 0$.

iii) $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{F}$ and $x \in \mathbb{F}^n$.

iv) $\|x+y\| \le \|x\| + \|y\|$ for all $x, y \in \mathbb{F}^n$.

Condition *iv)* is the *triangle inequality.*

A norm $\|\cdot\|$ on $\mathbb{F}^n$ is *monotone* if $|x| \le\le |y|$ implies that $\|x\| \le \|y\|$ for all $x, y \in \mathbb{F}^n$, while $\|\cdot\|$ is *absolute* if $\||x|\| = \|x\|$ for all $x \in \mathbb{F}^n$.

Proposition 9.1.2. Let $\|\cdot\|$ be a norm on $\mathbb{F}^n$. Then, $\|\cdot\|$ is monotone if and only if $\|\cdot\|$ is absolute.

Proof. First, suppose that $\|\cdot\|$ is monotone. Let $x \in \mathbb{F}^n$, and define $y \triangleq |x|$. Then, $|y| = |x|$ and thus $|y| \leq\leq |x|$ and $|y| \leq\leq |x|$. Hence, $\|x\| \leq \|y\|$ and $\|y\| \leq \|x\|$, which implies that $\|x\| = \|y\|$. Thus, $\||x|\| = \|y\| = \|x\|$, which proves that $\|\cdot\|$ is absolute.

Conversely, suppose that $\|\cdot\|$ is absolute and, for convenience, let $n = 2$. Now, let $x, y \in \mathbb{F}^2$ be such that $|x| \leq\leq |y|$. Then, there exist $\alpha_1, \alpha_2 \in [0,1]$ and $\theta_1, \theta_2 \in \mathbb{R}$ such that $x_{(i)} = \alpha_i e^{\jmath\theta_i} y_{(i)}$ for $i = 1, 2$. Since $\|\cdot\|$ is absolute, it follows that

$$\begin{aligned}
\|x\| &= \left\| \begin{bmatrix} \alpha_1 e^{\jmath\theta_1} y_{(1)} \\ \alpha_2 e^{\jmath\theta_2} y_{(2)} \end{bmatrix} \right\| \\
&= \left\| \begin{bmatrix} \alpha_1 & |y_{(1)}| \\ \alpha_2 & |y_{(2)}| \end{bmatrix} \right\| \\
&= \left\| \tfrac{1}{2}(1-\alpha_1) \begin{bmatrix} -|y_{(1)}| \\ |\alpha_2||y_{(2)}| \end{bmatrix} + \tfrac{1}{2}(1-\alpha_1) \begin{bmatrix} |y_{(1)}| \\ \alpha_2 |y_{(2)}| \end{bmatrix} + \alpha_1 \begin{bmatrix} |y_{(1)}| \\ \alpha_2 |y_{(2)}| \end{bmatrix} \right\| \\
&\leq \left[\tfrac{1}{2}(1-\alpha_1) + \tfrac{1}{2}(1-\alpha_1) + \alpha_1\right] \left\| \begin{bmatrix} |y_{(1)}| \\ \alpha_2 |y_{(2)}| \end{bmatrix} \right\| \\
&= \left\| \begin{bmatrix} |y_{(1)}| \\ \alpha_2 |y_{(2)}| \end{bmatrix} \right\| \\
&= \left\| \tfrac{1}{2}(1-\alpha_2) \begin{bmatrix} |y_{(1)}| \\ -|y_{(2)}| \end{bmatrix} + \tfrac{1}{2}(1-\alpha_2) \begin{bmatrix} |y_{(1)}| \\ |y_{(2)}| \end{bmatrix} + \alpha_2 \begin{bmatrix} |y_{(1)}| \\ |y_{(2)}| \end{bmatrix} \right\| \\
&\leq \left\| \begin{bmatrix} |y_{(1)}| \\ |y_{(2)}| \end{bmatrix} \right\| \\
&= ||y||.
\end{aligned}$$

Thus, $\|\cdot\|$ is monotone. □

As we shall see, there are many different norms. A useful class of norms consists of the *Holder norms* defined by

$$\|x\|_p \triangleq \begin{cases} \left(\displaystyle\sum_{i=1}^n |x_{(i)}|^p \right)^{1/p}, & 1 \leq p < \infty, \\ \displaystyle\max_{i \in \{1,\ldots,n\}} |x_{(i)}|, & p = \infty. \end{cases} \tag{9.1.3}$$

These norms depend on *Minkowski's inequality* given by the following result.

Lemma 9.1.3. Let $p \in [1, \infty]$, and let $x, y \in \mathbb{F}^n$. Then,

$$\|x + y\|_p \le \|x\|_p + \|y\|_p. \tag{9.1.4}$$

If $p = 1$, then equality holds if and only if, for all $i = 1, \ldots, n$, there exists $\alpha_i \ge 0$ such that either $x_{(i)} = \alpha_i y_{(i)}$ or $y_{(i)} = \alpha_i x_{(i)}$. If $p \in (1, \infty)$, then equality holds if and only if there exists $\alpha \ge 0$ such that either $x = \alpha y$ or $y = \alpha x$.

Proof. See [70, 395] and Fact 1.4.17. □

Proposition 9.1.4. Let $p \in [1, \infty]$. Then, $\|\cdot\|_p$ is a norm on $\mathbb{F}^n$.

For $p = 1$,

$$\|x\|_1 = \sum_{i=1}^{n} |x_{(i)}| \tag{9.1.5}$$

is the *absolute sum norm*; for $p = 2$,

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_{(i)}|^2 \right)^{1/2} = \sqrt{x^* x} \tag{9.1.6}$$

is the *Euclidean norm*; and, for $p = \infty$,

$$\|x\|_\infty = \max_{i \in \{1, \ldots, n\}} |x_{(i)}| \tag{9.1.7}$$

is the *infinity norm*.

Proposition 9.1.5. Let $1 \le p \le q \le \infty$, and let $x \in \mathbb{F}^n$. Then,

$$\|x\|_\infty \le \|x\|_q \le \|x\|_p \le \|x\|_1. \tag{9.1.8}$$

Assume, in addition, that $1 \le p < q \le \infty$. Then, x has at least two nonzero components if and only if

$$\|x\|_\infty < \|x\|_q < \|x\|_p < \|x\|_1. \tag{9.1.9}$$

Proof. If either $p = q$ or $x = 0$ or x has exactly one nonzero component, then $\|x\|_q = \|x\|_p$. Hence, to prove both (9.1.8) and (9.1.9) it suffices to prove (9.1.9) in the case that $1 < p < q < \infty$ and x has at least two nonzero components. Thus, let $n \ge 2$, let $x \in \mathbb{F}^n$ have at least two nonzero components, and define $f\colon [1, \infty) \to [0, \infty)$ by $f(\beta) \triangleq \|x\|_\beta$. Hence,

$$f'(\beta) = \tfrac{1}{\beta} \|x\|_\beta^{1-\beta} \sum_{i=1}^{n} \gamma_i,$$

where, for all $i = 1, \ldots, n$,

$$\gamma_i \triangleq \begin{cases} |x_i|^\beta\big(\log|x_{(i)}| - \log\|x\|_\beta\big), & x_{(i)} \neq 0, \\ 0, & x_{(i)} = 0. \end{cases}$$

If $x_{(i)} \neq 0$, then $\log|x_{(i)}| < \log\|x\|_\beta$. It thus follows that $f'(\beta) < 0$, which implies that f is decreasing on $[1, \infty)$. Hence, (9.1.9) holds. □

The following result is *Holder's inequality.* For this result we interpret $1/\infty = 0$.

Proposition 9.1.6. Let $p, q \in [0, \infty]$ satisfy $1/p + 1/q = 1$, and let $x, y \in \mathbb{F}^n$. Then,

$$|x^*y| \le \|x\|_p\|y\|_q. \tag{9.1.10}$$

Furthermore, equality holds if and only if $|x^*y| = |x|^{\mathrm{T}}|y|$ and

$$\begin{cases} |x| \circ |y| = \|y\|_\infty|x|, & p = 1, \\ |x|^{\{p\}}\text{and}|y|^{\{q\}} \text{ are linearly dependent}, & 1 < p < \infty, \\ |x| \circ |y| = \|x\|_\infty|y|, & p = \infty. \end{cases} \tag{9.1.11}$$

Proof. See [117, p. 127], [287, pp. 534–536], and Fact 1.4.16. □

The case $p = q = 2$ is the *Cauchy-Schwarz inequality.*

Corollary 9.1.7. Let $x, y \in \mathbb{F}^n$. Then,

$$|x^*y| \le \|x\|_2\|y\|_2. \tag{9.1.12}$$

Furthermore, equality holds if and only if x and y are linearly dependent.

Proof. Assume $y \neq 0$, and define $M \triangleq \begin{bmatrix} \sqrt{y^*y}I & (y^*y)^{-1/2}y \end{bmatrix}$. Since $M^*M = \begin{bmatrix} y^*yI & y \\ y^* & 1 \end{bmatrix}$ is nonnegative semidefinite, it follows from *iii*) of Proposition 8.2.3 that $yy^* \le y^*yI$. Therefore, $x^*yy^*x \le x^*xy^*y$, which is equivalent to (9.1.12).

Now, suppose that x and y are linearly dependent. Then, there exists $\beta \in \mathbb{F}$ such that either $x = \beta y$ or $y = \beta x$. In both cases it follows that $|x^*y| = \|x\|_2\|y\|_2$. Conversely, define $f\colon\ \mathbb{F}^n \times \mathbb{F}^n \to [0, \infty)$ by $f(\mu, \nu) \triangleq \mu^*\mu\nu^*\nu - |\mu^*\nu|^2$. Now, suppose that $f(x, y) = 0$ so that (x, y) minimizes f. Then, it follows that $f_\mu(x, y) = 0$, which implies that $y^*yx = y^*xy$. Hence, x and y are linearly dependent. □

The norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{F}^n$ are *equivalent* if there exist $\alpha, \beta > 0$

such that

$$\alpha\|x\| \le \|x\|' \le \beta\|x\| \tag{9.1.13}$$

for all $x \in \mathbb{F}^n$. Note that these inequalities can be written as

$$\tfrac{1}{\beta}\|x\|' \le \|x\| \le \tfrac{1}{\alpha}\|x\|'. \tag{9.1.14}$$

Hence, the word "equivalent" is justified.

Theorem 9.1.8. Let $\|\cdot\|$ and $\|\cdot\|'$ be norms on $\mathbb{F}^n$. Then, $\|\cdot\|$ and $\|\cdot\|'$ are equivalent.

Proof. See [287, p. 272]. □

9.2 Matrix Norms

One way to define norms for matrices is by viewing a matrix $A \in \mathbb{F}^{n\times m}$ as a vector in $\mathbb{F}^{nm}$, for example, as $\operatorname{vec} A$.

Definition 9.2.1. A *norm* $\|\cdot\|$ on $\mathbb{F}^{n\times m}$ is a function $\|\cdot\|\colon\ \mathbb{F}^{n\times m} \mapsto \mathbb{R}$ that satisfies the following conditions:

i) $\|A\| \ge 0$ for all $A \in \mathbb{F}^{n\times m}$.

ii) $\|A\| = 0$ if and only if $A = 0$.

iii) $\|\alpha A\| = |\alpha|\,\|A\|$ for all $\alpha \in \mathbb{F}$.

iv) $\|A+B\| \le \|A\| + \|B\|$ for all $A, B \in \mathbb{F}^{n\times m}$.

If $\|\cdot\|$ is a norm on $\mathbb{F}^{nm}$, then $\|\cdot\|'$ defined by $\|A\|' \triangleq \|\operatorname{vec} A\|$ is a norm on $\mathbb{F}^{n\times m}$. For example, Holder norms can be defined for matrices by choosing $\|\cdot\| = \|\cdot\|_p$. Hence, for all $A \in \mathbb{F}^{n\times m}$ define

$$\|A\|_p \triangleq \begin{cases} \left(\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} |A_{(i,j)}|^p \right)^{1/p}, & 1 \le p < \infty, \\ \displaystyle\max_{\substack{i\in\{1,\ldots,n\}\\ j\in\{1,\ldots,m\}}} |A_{(i,j)}|, & p = \infty. \end{cases} \tag{9.2.1}$$

Note that the same symbol $\|\cdot\|_p$ is used to denote the Holder norm for both vectors and matrices. This notation is consistent since, if $A \in \mathbb{F}^{n\times 1}$, then $\|A\|_p$ coincides with the vector Holder norm. Furthermore, if $A \in \mathbb{F}^{n\times m}$ and $1 \le p \le \infty$, then

$$\|A\|_p = \|\operatorname{vec} A\|_p. \tag{9.2.2}$$

It follows from (9.1.8) that, if $A \in \mathbb{F}^{n\times m}$ and $1 \le p \le q$, then

$$\|A\|_\infty \le \|A\|_q \le \|A\|_p \le \|A\|_1. \tag{9.2.3}$$

If, in addition, $1 < p < q < \infty$ and A has at least two nonzero entries, then

$$\|A\|_\infty < \|A\|_q < \|A\|_p < \|A\|_1. \tag{9.2.4}$$

The Holder norms in the cases $p = 1, 2, \infty$ are the most commonly used. Let $A \in \mathbb{F}^{n\times m}$. For $p = 2$ we define the *Frobenius norm* $\|\cdot\|_{\mathrm{F}}$ by

$$\|A\|_{\mathrm{F}} \triangleq \|A\|_2. \tag{9.2.5}$$

Since $\|A\|_2 = \|\operatorname{vec} A\|_2$, it follows that

$$\|A\|_{\mathrm{F}} = \|A\|_2 = \|\operatorname{vec} A\|_2 = \|\operatorname{vec} A\|_{\mathrm{F}}. \tag{9.2.6}$$

It is easy to see that

$$\|A\|_{\mathrm{F}} = \sqrt{\operatorname{tr} A^*A}. \tag{9.2.7}$$

Let $\|\cdot\|$ be a norm on $\mathbb{F}^{n\times m}$. If $\|S_1AS_2\| = \|A\|$ for all $A \in \mathbb{F}^{n\times m}$ and for all unitary matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$, then $\|\cdot\|$ is *unitarily invariant*. Now, let $m = n$. If $\|A\| = \|A^*\|$ for all $A \in \mathbb{F}^{n\times n}$, then $\|\cdot\|$ is *self adjoint*. If $\|I_n\| = 1$, then $\|\cdot\|$ is *normalized*. Note that the Frobenius norm is not normalized since $\|I_n\|_{\mathrm{F}} = \sqrt{n}$. If $\|SAS^*\| = \|A\|$ for all $A \in \mathbb{F}^{n\times n}$ and for all unitary $S \in \mathbb{F}^{n\times n}$, then $\|\cdot\|$ is *weakly unitarily invariant*.

An important class of norms can be defined in terms of singular values. Let $\sigma_1(A) \ge \sigma_2(A) \ge \cdots$ denote the singular values of $A \in \mathbb{F}^{n\times m}$. The following result gives a weak majorization condition for singular values.

Proposition 9.2.2. Let $A, B \in \mathbb{F}^{n\times m}$. Then, for all $k = 1, \ldots, \min\{n, m\}$,

$$\sum_{i=1}^{k}[\sigma_i(A) - \sigma_i(B)] \le \sum_{i=1}^{k}\sigma_i(A+B) \le \sum_{i=1}^{k}[\sigma_i(A) + \sigma_i(B)]. \tag{9.2.8}$$

In particular,

$$\sigma_{\max}(A+B) \le \sigma_{\max}(A) + \sigma_{\max}(B) \tag{9.2.9}$$

and

$$\operatorname{tr}\langle A+B\rangle \le \operatorname{tr}\langle A\rangle + \operatorname{tr}\langle B\rangle. \tag{9.2.10}$$

Proof. Define $\mathcal{A}, \mathcal{B} \in \mathbf{H}^{n+m}$ by $\mathcal{A} \triangleq \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ and $\mathcal{B} \triangleq \begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix}$. Then, Corollary 8.5.15 implies that, for all $k = 1, \ldots, n+m$,

$$\sum_{i=1}^{k}\lambda_i(\mathcal{A}+\mathcal{B}) \le \sum_{i=1}^{k}[\lambda_i(\mathcal{A}) + \lambda_i(\mathcal{B})].$$

Now, consider $k \le \min\{n,m\}$. Then, it follows from Proposition 5.6.5 that, for all $i = 1,\ldots,k$, $\lambda_i(\mathcal{A}) = \sigma_i(A)$. Setting $k = 1$ yields (9.2.9), while setting $k = \min\{n,m\}$ and using Fact 8.14.1 yields (9.2.10). □

Proposition 9.2.3. Let $p \in [1,\infty]$, and let $A \in \mathbb{F}^{n\times m}$. Then, $\|\cdot\|_{\sigma p}$ defined by

$$\|A\|_{\sigma p} \triangleq \begin{cases} \left(\sum_{i=1}^{\min\{n,m\}} \sigma_i^p(A)\right)^{1/p}, & 1 \le p < \infty, \\ \sigma_{\max}(A), & p = \infty, \end{cases} \tag{9.2.11}$$

is a norm on $\mathbb{F}^{n\times m}$.

Proof. Let $p \in [1,\infty]$. Then, it follows from Proposition 9.2.2 and Minkowski's inequality Fact 1.4.17 that

$$\begin{aligned} \|A+B\|_{\sigma p} &= \left(\sum_{i=1}^{\min\{n,m\}} \sigma_i^p(A+B)\right)^{1/p} \\ &\le \left(\sum_{i=1}^{\min\{n,m\}} [\sigma_i(A)+\sigma_i(B)]^p\right)^{1/p} \\ &\le \left(\sum_{i=1}^{\min\{n,m\}} \sigma_i^p(A)\right)^{1/p} + \left(\sum_{i=1}^{\min\{n,m\}} \sigma_i^p(B)\right)^{1/p} \\ &= \|A\|_{\sigma p} + \|B\|_{\sigma p}. \end{aligned}$$

□

The norm $\|\cdot\|_{\sigma p}$ is a *Schatten norm*. Let $A \in \mathbb{F}^{n\times m}$. Then, for all $p \in [1,\infty)$,

$$\|A\|_{\sigma p} = (\operatorname{tr}\langle A\rangle^p)^{1/p}. \tag{9.2.12}$$

Important special cases are

$$\|A\|_{\sigma 1} = \sigma_1(A) + \cdots + \sigma_{\min\{n,m\}}(A) = \operatorname{tr}\langle A\rangle, \tag{9.2.13}$$

$$\|A\|_{\sigma 2} = \left[\sigma_1^2(A) + \cdots + \sigma_{\min\{n,m\}}^2(A)\right]^{1/2} = (\operatorname{tr} A^*A)^{1/2} = \|A\|_{\mathrm{F}}, \tag{9.2.14}$$

and

$$\|A\|_{\sigma\infty} = \sigma_1(A) = \sigma_{\max}(A), \tag{9.2.15}$$

which are the *trace norm*, Frobenius norm, and *spectral norm*, respectively.

By applying Proposition 9.1.5 to the vector $\begin{bmatrix} \sigma_1(A) & \cdots & \sigma_{\min\{n,m\}}(A) \end{bmatrix}^{\mathrm{T}}$, we obtain the following result.

Proposition 9.2.4. Let $p, q \in [1, \infty)$, where $p \le q$, and let $A \in \mathbb{F}^{n\times m}$. Then,

$$\|A\|_{\sigma\infty} \le \|A\|_{\sigma q} \le \|A\|_{\sigma p} \le \|A\|_{\sigma 1}. \tag{9.2.16}$$

Assume, in addition, that $1 < p < q < \infty$ and $\operatorname{rank} A \ge 2$. Then,

$$\|A\|_{\infty} < \|A\|_{q} < \|A\|_{p} < \|A\|_{1}. \tag{9.2.17}$$

The norms $\|\cdot\|_{\sigma p}$ are not very interesting when applied to vectors. Let $x \in \mathbb{F}^n = \mathbb{F}^{n\times 1}$. Then, $\sigma_{\max}(x) = (x^*x)^{1/2} = \|x\|_2$, and, since $\operatorname{rank} x \le 1$, it follows that, for all $p \in [1, \infty]$,

$$\|x\|_{\sigma p} = \|x\|_2. \tag{9.2.18}$$

Proposition 9.2.5. Let $A \in \mathbb{F}^{n\times m}$. If $p \in (0, 2]$, then

$$\|A\|_{\sigma p} \le \|A\|_p. \tag{9.2.19}$$

If $p \ge 2$, then

$$\|A\|_p \le \|A\|_{\sigma p}. \tag{9.2.20}$$

Proof. See [625, p. 50]. □

Proposition 9.2.6. Let $\|\cdot\|$ be a norm on $\mathbb{F}^{n\times n}$, and let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{sprad}(A) = \lim_{k\to\infty} \|A^k\|^{1/k}. \tag{9.2.21}$$

Proof. See [287, p. 322]. □

9.3 Compatible Norms

The norms $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ on $\mathbb{F}^{n\times l}$, $\mathbb{F}^{n\times m}$, and $\mathbb{F}^{m\times l}$, respectively, are *compatible* if, for all $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$,

$$\|AB\| \le \|A\|'\|B\|''. \tag{9.3.1}$$

For $l = 1$, the norms $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ on $\mathbb{F}^n$, $\mathbb{F}^{n\times m}$, and $\mathbb{F}^m$, respectively, are compatible if, for all $A \in \mathbb{F}^{n\times m}$ and $x \in \mathbb{F}^m$,

$$\|Ax\| \le \|A\|'\|x\|''. \tag{9.3.2}$$

Furthermore, the norm $\|\cdot\|$ on $\mathbb{F}^n$ is *compatible* with the norm $\|\cdot\|'$ on $\mathbb{F}^{n\times n}$ if, for all $A \in \mathbb{F}^{n\times n}$ and $x \in \mathbb{F}^n$,

$$\|Ax\| \le \|A\|'\|x\|. \tag{9.3.3}$$

Note that $\|I_n\|' \ge 1$. The norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ is *submultiplicative* if, for all $A, B \in \mathbb{F}^{n\times n}$,

$$\|AB\| \le \|A\|\|B\|. \tag{9.3.4}$$

Hence, the norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ is submultiplicative if and only if $\|\cdot\|$, $\|\cdot\|$, and $\|\cdot\|$ are compatible. In this case, $\|I_n\| \ge 1$.

Proposition 9.3.1. Let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$, and let $y \in \mathbb{F}^n$. Then, $\|x\|' \triangleq \|xy^*\|$ is a norm on $\mathbb{F}^n$, and $\|\cdot\|'$ is compatible with $\|\cdot\|$.

Proposition 9.3.2. Let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$, and let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{sprad}(A) \le \|A\|. \tag{9.3.5}$$

Proof. Use Proposition 9.3.1 to construct a norm $\|\cdot\|'$ on $\mathbb{F}^n$ that is compatible with $\|\cdot\|$. Furthermore, let $A \in \mathbb{F}^{n\times n}$, let $\lambda \in \operatorname{spec}(A)$, and let $x \in \mathbb{C}^n$ be an eigenvector of A associated with λ. Then, $Ax = \lambda x$ implies that $|\lambda|\|x\|' = \|Ax\|' \le \|A\|\|x\|'$, and thus $|\lambda| \le \|A\|$, which implies (9.3.5). □

Proposition 9.3.3. Let $A \in \mathbb{F}^{n\times n}$, and let $\varepsilon > 0$. Then, there exists a submultiplicative norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ such that

$$\operatorname{sprad}(A) \le \|A\| \le \operatorname{sprad}(A) + \varepsilon. \tag{9.3.6}$$

Proof. See [287, p. 297]. □

Corollary 9.3.4. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{sprad}(A) < 1$. Then, there exists a submultiplicative norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ such that $\|A\| < 1$.

We now identify some compatible norms. We begin with the Holder norms.

Proposition 9.3.5. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. If $p \in [1, 2]$, then

$$\|AB\|_p \le \|A\|_p\|B\|_p. \tag{9.3.7}$$

If $p \in [2, \infty]$ and q satisfies $1/p + 1/q = 1$, then

$$\|AB\|_p \le \|A\|_p\|B\|_q \tag{9.3.8}$$

and

$$\|AB\|_p \le \|A\|_q\|B\|_p. \tag{9.3.9}$$

Proof. First let $1 \le p \le 2$ so that $q \triangleq p/(p-1) \ge 2$. Using Holder's inequality (9.1.10) and (9.1.8) with $p \le q$ yields

$$\begin{aligned}
\|AB\|_p &= \left(\sum_{i,j=1}^{n,l} |\mathrm{row}_i(A)\mathrm{col}_j(B)|^p \right)^{1/p} \\
&\le \left(\sum_{i,j=1}^{n,l} \|\mathrm{row}_i(A)\|_p^p \|\mathrm{col}_j(B)\|_q^p \right)^{1/p} \\
&= \left(\sum_{i=1}^{n} \|\mathrm{row}_i(A)\|_p^p \right)^{1/p} \left(\sum_{j=1}^{l} \|\mathrm{col}_j(B)\|_q^p \right)^{1/p} \\
&\le \left(\sum_{i=1}^{n} \|\mathrm{row}_i(A)\|_p^p \right)^{1/p} \left(\sum_{j=1}^{l} \|\mathrm{col}_j(B)\|_p^p \right)^{1/p} \\
&= \|A\|_p \|B\|_p.
\end{aligned}$$

Next, let $2 \le p \le \infty$ so that $q \triangleq p/(p-1) \le 2$. Using Holder's inequality (9.1.10) and (9.1.8) with $q \le p$ yields

$$\begin{aligned}
\|AB\|_p &\le \left(\sum_{i=1}^{n} \|\mathrm{row}_i(A)\|_p^p \right)^{1/p} \left(\sum_{j=1}^{l} \|\mathrm{col}_j(B)\|_q^p \right)^{1/p} \\
&\le \left(\sum_{i=1}^{n} \|\mathrm{row}_i(A)\|_p^p \right)^{1/p} \left(\sum_{j=1}^{l} \|\mathrm{col}_j(B)\|_q^q \right)^{1/q} \\
&= \|A\|_p \|B\|_q.
\end{aligned}$$

Similarly, it can be shown that (9.3.9) holds. □

Proposition 9.3.6. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $p, q \in [1,\infty]$, and let $r \triangleq 1/(1/p + 1/q) \ge 1$. Then,

$$\|AB\|_{\sigma r} \le \|A\|_{\sigma p} \|B\|_{\sigma q}. \tag{9.3.10}$$

Proof. Using Proposition 9.6.3 and Holder's inequality with $1/(p/r)+$

$1/(q/r) = 1$, it follows that

$$\begin{aligned}
\|AB\|_{\sigma r} &= \left(\sum_{i=1}^{\min\{n,m,l\}} \sigma_i^r(AB)\right)^{1/r} \\
&\le \left(\sum_{i=1}^{\min\{n,m,l\}} \sigma_i^r(A)\sigma_i^r(B)\right)^{1/r} \\
&\le \left[\left(\sum_{i=1}^{\min\{n,m,l\}} \sigma_i^p(A)\right)^{r/p}\left(\sum_{i=1}^{\min\{n,m,l\}} \sigma_i^q(B)\right)^{r/q}\right]^{1/r} \\
&= \|A\|_{\sigma p}\|B\|_{\sigma q}.
\end{aligned}$$

□

Let $A, B \in \mathbb{F}^{n\times m}$. Using (9.2.16) and (9.3.10) it follows that

$$\|AB\|_{\sigma\infty} \le \|AB\|_{\sigma 2} \le \left\{\begin{array}{c} \|A\|_{\sigma\infty}\|B\|_{\sigma 2} \\ \|A\|_{\sigma 2}\|B\|_{\sigma\infty} \\ \|AB\|_{\sigma 1} \end{array}\right\} \le \|A\|_{\sigma 2}\|B\|_{\sigma 2} \tag{9.3.11}$$

or, equivalently,

$$\sigma_{\max}(AB) \le \|AB\|_{\mathrm{F}} \le \left\{\begin{array}{c} \sigma_{\max}(A)\|B\|_{\mathrm{F}} \\ \|A\|_{\mathrm{F}}\sigma_{\max}(B) \\ \mathrm{tr}\,\langle AB\rangle \end{array}\right\} \le \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}. \tag{9.3.12}$$

Also, for all $r \in [1, \infty]$,

$$\|AB\|_{\sigma r} \le \left\{\begin{array}{l} \|A\|_{\sigma r}\sigma_{\max}(B) \\ \sigma_{\max}(A)\|B\|_{\sigma r} \end{array}\right. . \tag{9.3.13}$$

In particular, setting $r = \infty$ yields

$$\sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B). \tag{9.3.14}$$

Note that the inequality $\|AB\|_{\mathrm{F}} \le \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}$ in (9.3.12) is equivalent to (9.3.7) with $p = 2$ as well as (9.3.8) and (9.3.9) with $p = q = 2$. Finally, it follows from the Cauchy-Schwarz inequality Corollary 9.1.7 that

$$|\mathrm{tr}\, A^*B| \le \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}. \tag{9.3.15}$$

9.4 Induced Norms

In this section we consider the case in which there exists nonzero $x \in \mathbb{F}^m$ such that (9.3.3) holds as an equality. This condition characterizes a special class of norms on $\mathbb{F}^{n\times n}$, namely, the *induced norms.*

Definition 9.4.1. Let $\|\cdot\|''$ and $\|\cdot\|$ be norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively. Then, $\|\cdot\|'\colon\ \mathbb{F}^{n\times m} \mapsto \mathbb{F}$ defined by

$$\|A\|' = \max_{x\in\mathbb{F}^m\backslash\{0\}} \frac{\|Ax\|}{\|x\|''} \tag{9.4.1}$$

is an *induced norm* on $\mathbb{F}^{n\times m}$. In this case, $\|\cdot\|'$ is *induced by* $\|\cdot\|''$ and $\|\cdot\|$. If $m=n$ and $\|\cdot\|''=\|\cdot\|$, then $\|\cdot\|'$ is *induced by* $\|\cdot\|$, and $\|\cdot\|'$ is an *equi-induced norm.*

The next result confirms that $\|\cdot\|'$ defined by (9.4.1) is indeed a norm.

Theorem 9.4.2. Every induced norm is a norm. Furthermore, every equi-induced norm is normalized.

Proof. See [287, p. 293].

Let $A \in \mathbb{F}^{n\times m}$. It can be seen that (9.4.1) is equivalent to

$$\|A\|' = \max_{x\in\{y\in\mathbb{F}^m\colon\ 0<\|y\|''\le 1\}} \frac{\|Ax\|}{\|x\|''} \tag{9.4.2}$$

as well as

$$\|A\|' = \max_{x\in\{y\in\mathbb{F}^m\colon\ \|y\|''=1\}} \|Ax\|. \tag{9.4.3}$$

Theorem 10.3.7 implies that the maximum in (9.4.3) exists. Since, for all $x \ne 0$,

$$\|A\|' = \max_{x\in\mathbb{F}^m\backslash\{0\}} \frac{\|Ax\|}{\|x\|''} \ge \frac{\|Ax\|}{\|x\|''} \tag{9.4.4}$$

it follows that, for all $x \in \mathbb{F}^m$,

$$\|Ax\| \le \|A\|'\|x\|'' \tag{9.4.5}$$

so that $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ are compatible. If $m=n$ and $\|\cdot\|''=\|\cdot\|$, then the norm $\|\cdot\|$ is compatible with the induced norm $\|\cdot\|'$. The next result shows that compatible norms can be obtained from induced norms.

Proposition 9.4.3. Let $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ be norms on $\mathbb{F}^l$, $\mathbb{F}^m$, and $\mathbb{F}^n$, respectively. Furthermore, let $\|\cdot\|'''$ be the norm on $\mathbb{F}^{m\times l}$ induced by $\|\cdot\|$ and $\|\cdot\|'$, let $\|\cdot\|''''$ be the norm on $\mathbb{F}^{n\times m}$ induced by $\|\cdot\|'$ and $\|\cdot\|''$, and let $\|\cdot\|'''''$ be the norm on $\mathbb{F}^{n\times l}$ induced by $\|\cdot\|$ and $\|\cdot\|''$. If $A \in \mathbb{F}^{n\times m}$

and $B \in \mathbb{F}^{m\times l}$, then
$$\|AB\|''''' \le \|A\|''''\|B\|'''. \tag{9.4.6}$$

Proof. Note that, for all $x \in \mathbb{F}^l$, $\|Bx\|' \le \|B\|'''\|x\|$, and, for all $y \in \mathbb{F}^m$, $\|Ay\|'' \le \|A\|''''\|y\|'$. Hence, for all $x \in \mathbb{F}^l$, $\|ABx\|'' \le \|A\|''''\|Bx\|' \le \|A\|''''\|B\|'''\|x\|$, which implies that
$$\|AB\|''''' = \max_{x\in\mathbb{F}^l\backslash\{0\}} \frac{\|ABx\|''}{\|x\|} \le \|A\|''''\|B\|'''.$$
□

Corollary 9.4.4. Every equi-induced norm is submultiplicative.

The following result is a consequence of Corollary 9.4.4 and Proposition 9.3.2.

Corollary 9.4.5. Let $\|\cdot\|$ be an equi-induced norm on $\mathbb{F}^{n\times n}$, and let $A \in \mathbb{F}^{n\times n}$. Then,
$$\operatorname{sprad}(A) \le \|A\|. \tag{9.4.7}$$

By assigning $\|\cdot\|_p$ to $\mathbb{F}^m$ and $\|\cdot\|_q$ to $\mathbb{F}^n$, the *Holder-induced norm* on $\mathbb{F}^{n\times m}$ is defined by
$$\|A\|_{q,p} \triangleq \max_{x\in\mathbb{F}^m\backslash\{0\}} \frac{\|Ax\|_q}{\|x\|_p}. \tag{9.4.8}$$

Proposition 9.4.6. Let $p, q, p', q' \in [1,\infty]$, where $p \le p'$ and $q \le q'$, and let $A \in \mathbb{F}^{n\times m}$. Then,
$$\|A\|_{q',p} \le \|A\|_{q,p} \le \|A\|_{q,p'}. \tag{9.4.9}$$

Proof. The result follows from Proposition 9.1.5. □

The following result gives explicit expressions for several Holder-induced norms.

Proposition 9.4.7. Let $A \in \mathbb{F}^{n\times m}$. Then,
$$\|A\|_{2,2} = \sigma_{\max}(A). \tag{9.4.10}$$
Now, let $p \in [1,\infty]$. Then,
$$\|A\|_{p,1} = \max_{i\in\{1,\ldots,m\}} \|\operatorname{col}_i(A)\|_p. \tag{9.4.11}$$
Finally, let $q \in [1,\infty]$ satisfy $1/p + 1/q = 1$. Then,
$$\|A\|_{\infty,p} = \max_{i\in\{1,\ldots,n\}} \|\operatorname{row}_i(A)\|_q. \tag{9.4.12}$$

Proof. Since A^*A is Hermitian, it follows from Corollary 8.4.2 that, for all $x \in \mathbb{F}^m$,

$$x^*A^*Ax \leq \lambda_{\max}(A^*A)x^*x,$$

which implies that, for all $x \in \mathbb{F}^m$, $\|Ax\|_2 \leq \sigma_{\max}(A)\|x\|_2$, and thus $\|A\|_{2,2} \leq \sigma_{\max}(A)$. Now, let $x \in \mathbb{F}^{n\times n}$ be an eigenvector associated with $\lambda_{\max}(A^*A)$ so that $\|Ax\|_2 = \sigma_{\max}(A)\|x\|_2$, which implies that $\sigma_{\max}(A) \leq \|A\|_{2,2}$. Hence, (9.4.10) holds.

Next, note that, for all $x \in \mathbb{F}^m$,

$$\|Ax\|_p = \left\|\sum_{i=1}^m x_{(i)}\mathrm{col}_i(A)\right\|_p \leq \sum_{i=1}^m |x_{(i)}|\,\|\mathrm{col}_i(A)\|_p \leq \max_{i\in\{1,\ldots,m\}}\|\mathrm{col}_i(A)\|_p\|x\|_1,$$

and hence $\|A\|_{p,1} \leq \max_{i\in\{1,\ldots,m\}}\|\mathrm{col}_i(A)\|_p$. Next, let $j \in \{1,\ldots,m\}$ be such that $\|\mathrm{col}_j(A)\|_p = \max_{i\in\{1,\ldots,m\}}\|\mathrm{col}_i(A)\|_p$. Now, since $\|e_j\|_1 = 1$, it follows that $\|Ae_j\|_p = \|\mathrm{col}_j(A)\|_p\|e_j\|_1$, which implies that

$$\max_{i\in\{1,\ldots,n\}}\|\mathrm{col}_i(A)\|_p = \|\mathrm{col}_j(A)\|_p \leq \|A\|_{p,1},$$

and hence (9.4.11) holds.

Next, for all $x \in \mathbb{F}^m$, it follows from Holder's inequality (9.1.10) that

$$\|Ax\|_\infty = \max_{i\in\{1,\ldots,n\}}|\mathrm{row}_i(A)x| \leq \max_{i\in\{1,\ldots,n\}}\|\mathrm{row}_i(A)\|_q\|x\|_p,$$

which implies that $\|A\|_{\infty,p} \leq \max_{i\in\{1,\ldots,n\}}\|\mathrm{row}_i(A)\|_q$. Next, let $j\in\{1,\ldots,n\}$ be such that $\|\mathrm{row}_j(A)\|_q = \max_{i\in\{1,\ldots,n\}}\|\mathrm{row}_i(A)\|_q$, and let nonzero $x \in \mathbb{F}^m$ be such that $|\mathrm{row}_j(A)x| = \|\mathrm{row}_j(A)\|_q\|x\|_p$. Hence,

$$\|Ax\|_\infty = \max_{i\in\{1,\ldots,n\}}|\mathrm{row}_i(A)x| \geq |\mathrm{row}_j(A)x| = \|\mathrm{row}_j(A)\|_q\|x\|_p,$$

which implies that

$$\max_{i\in\{1,\ldots,n\}}\|\mathrm{row}_i(A)\|_q = \|\mathrm{row}_j(A)\|_q \leq \|A\|_{\infty,p},$$

and thus (9.4.12) holds. □

Note that

$$\max_{i\in\{1,\ldots,m\}}\|\mathrm{col}_i(A)\|_2 = \mathrm{d}_{\max}^{1/2}(A^*A) \tag{9.4.13}$$

and

$$\max_{i\in\{1,\ldots,n\}}\|\mathrm{row}_i(A)\|_2 = \mathrm{d}_{\max}^{1/2}(AA^*). \tag{9.4.14}$$

Therefore, it follows from Proposition 9.4.7 that

$$\|A\|_{1,1} = \max_{i\in\{1,\ldots,m\}} \|\mathrm{col}_i(A)\|_1, \tag{9.4.15}$$

$$\|A\|_{2,1} = \mathrm{d}_{\max}^{1/2}(A^*A), \tag{9.4.16}$$

$$\|A\|_{\infty,1} = \|A\|_\infty = \max_{\substack{i\in\{1,\ldots,n\}\\ j\in\{1,\ldots,m\}}} |A_{(i,j)}|, \tag{9.4.17}$$

$$\|A\|_{\infty,2} = \mathrm{d}_{\max}^{1/2}(AA^*), \tag{9.4.18}$$

$$\|A\|_{\infty,\infty} = \max_{i\in\{1,\ldots,n\}} \|\mathrm{row}_i(A)\|_1. \tag{9.4.19}$$

For convenience, we define the *column norm*

$$\|A\|_{\mathrm{col}} \triangleq \|A\|_{1,1} \tag{9.4.20}$$

and the *row norm*

$$\|A\|_{\mathrm{row}} \triangleq \|A\|_{\infty,\infty}. \tag{9.4.21}$$

Proposition 9.4.8. Let $p,q \in [1,\infty]$ be such that $1/p + 1/q = 1$, and let $A \in \mathbb{F}^{n\times m}$. Then,

$$\|A\|_{q,p} \le \|A\|_q. \tag{9.4.22}$$

Proof. For $p = 1$ and $q = \infty$, (9.4.22) follows from (9.4.17). For $q < \infty$ and $x \in \mathbb{F}^n$, it follows from Holder's inequality (9.1.10) that

$$\begin{aligned}\|Ax\|_q &= \left(\sum_{i=i}^n |\mathrm{row}_i(A)x|^q\right)^{1/q} \le \left(\sum_{i=1}^n \|\mathrm{row}_i(A)\|_q^q \|x\|_p^q\right)^{1/q}\\ &= \left(\sum_{i=1}^n \sum_{j=1}^m |A_{(i,j)}|^q\right)^{1/q} \|x\|_p = \|A\|_q \|x\|_p,\end{aligned}$$

which implies (9.4.22). □

Next, we specialize Proposition 9.4.3 to the Holder-induced norms.

Corollary 9.4.9. Let $1 \le p,q,r \le \infty$, and let $A \in \mathbb{F}^{n\times m}$ and $A \in \mathbb{F}^{m\times l}$. Then,

$$\|AB\|_{r,p} \le \|A\|_{r,q}\|B\|_{q,p} \tag{9.4.23}$$

In particular,

$$\|AB\|_{\rm col} \le \|A\|_{\rm col}\|B\|_{\rm col}, \tag{9.4.24}$$

$$\sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B), \tag{9.4.25}$$

$$\|AB\|_{\rm row} \le \|A\|_{\rm row}\|B\|_{\rm row}, \tag{9.4.26}$$

$$\|AB\|_{\infty} \le \|A\|_{\infty}\|B\|_{\rm col}, \tag{9.4.27}$$

$$\|AB\|_{\infty} \le \|A\|_{\rm row}\|B\|_{\infty}, \tag{9.4.28}$$

$$\mathrm{d}_{\max}^{1/2}(B^*A^*AB) \le \mathrm{d}_{\max}^{1/2}(A^*A)\|B\|_{\rm col}, \tag{9.4.29}$$

$$\mathrm{d}_{\max}^{1/2}(B^*A^*AB) \le \sigma_{\max}(A)\mathrm{d}_{\max}^{1/2}(B^*B), \tag{9.4.30}$$

$$\mathrm{d}_{\max}^{1/2}(ABB^*A^*) \le \mathrm{d}_{\max}^{1/2}(AA^*)\sigma_{\max}(B), \tag{9.4.31}$$

$$\mathrm{d}_{\max}^{1/2}(ABB^*A^*) \le \|B\|_{\rm row}\mathrm{d}_{\max}^{1/2}(BB^*). \tag{9.4.32}$$

The following result is often useful.

Proposition 9.4.10. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\mathrm{sprad}(A) < 1$. Then, there exists a submultiplicative norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ such that $\|A\| < 1$. Furthermore, the series $\sum_{k=0}^{\infty} A^k$ converges absolutely, and

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k. \tag{9.4.33}$$

Finally,

$$\frac{1}{1+\|A\|} \le \left\|(I-A)^{-1}\right\| \le \frac{1}{1-\|A\|} + \|I\| - 1. \tag{9.4.34}$$

If, in addition, $\|\cdot\|$ is normalized, then

$$\frac{1}{1+\|A\|} \le \left\|(I-A)^{-1}\right\| \le \frac{1}{1-\|A\|}. \tag{9.4.35}$$

Proof. Corollary 9.3.4 implies that there exists a submultiplicative norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ such that $\|A\| < 1$. It thus follows that

$$\left\|\sum_{k=0}^{\infty} A^k\right\| \le \sum_{k=0}^{\infty} \|A^k\| \le \|I\| - 1 + \sum_{k=0}^{\infty} \|A\|^k = \frac{1}{1-\|A\|} + \|I\| - 1,$$

which proves that the series $\sum_{k=0}^{\infty} A^k$ converges absolutely.

Next, we show that $I - A$ is nonsingular. If $I - A$ is singular, then there exists a nonzero vector $x \in \mathbb{C}^n$ such that $Ax = x$. Hence, $1 \in \mathrm{spec}(A)$, which contradicts $\operatorname{sprad}(A) < 1$. Next, to verify (9.4.33), note that

$$(I-A)\sum_{k=0}^{\infty}A^k = \sum_{k=0}^{\infty}A^k - \sum_{k=1}^{\infty}A^k = I + \sum_{k=1}^{\infty}A^k - \sum_{k=1}^{\infty}A^k = I,$$

which implies (9.4.33) and thus the right-hand inequality in (9.4.34). Furthermore,

$$\begin{aligned} 1 &\le \|I\| \\ &= \left\|(I-A)(I-A)^{-1}\right\| \\ &\le \|I-A\|\left\|(I-A)^{-1}\right\| \\ &\le (1+\|A\|)\left\|(I-A)^{-1}\right\|, \end{aligned}$$

which yields the left-hand inequality in (9.4.34). □

9.5 Induced Lower Bound

We now consider a variation of the induced norm.

Definition 9.5.1. Let $\|\cdot\|$ and $\|\cdot\|'$ denote norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively. Then, for $A \in \mathbb{F}^{n\times m}$, $\ell\colon\ \mathbb{F}^{n\times m} \mapsto \mathbb{R}$ defined by

$$\ell(A) \triangleq \begin{cases} \min\limits_{y\in\mathcal{R}(A)\backslash\{0\}}\ \max\limits_{x\in\{z\in\mathbb{F}^m\colon\, Az=y\}} \frac{\|y\|'}{\|x\|}, & A \ne 0,\\[2ex] 0, & A = 0, \end{cases} \tag{9.5.1}$$

is the *lower bound induced by* $\|\cdot\|$ and $\|\cdot\|'$. Equivalently,

$$\ell(A) \triangleq \begin{cases} \min\limits_{y\in\mathcal{R}(A)\backslash\{0\}}\ \max\limits_{z\in\mathcal{N}(A)} \frac{\|Ax\|'}{\|x+z\|}, & A \ne 0,\\[2ex] 0, & A = 0. \end{cases} \tag{9.5.2}$$

Proposition 9.5.2. Let $\|\cdot\|$ and $\|\cdot\|'$ be norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively, let $\|\cdot\|''$ be the norm induced by $\|\cdot\|$ and $\|\cdot\|'$, let $\|\cdot\|'''$ be the norm induced by $\|\cdot\|'$ and $\|\cdot\|$, and let ℓ be the lower bound induced by $\|\cdot\|$ and $\|\cdot\|'$. Then, the following statements hold:

i) $\ell(A)$ exists for all $A \in \mathbb{F}^{n\times m}$, that is, the minimum in (9.5.1) is attained.

ii) If $A \in \mathbb{F}^{n\times m}$,then $\ell(A) = 0$ if and only if $A = 0$.

iii) For all $A \in \mathbb{F}^{n\times m}$ there exists $x \in \mathbb{F}^m$ such that

$$\ell(A)\|x\| = \|Ax\|'. \tag{9.5.3}$$

iv) For all $A \in \mathbb{F}^{n\times m}$,

$$\ell(A) \le \|A\|''. \tag{9.5.4}$$

v) If $A \ne 0$ and B is a (1)-inverse of A, then

$$1/\|B\|''' \le \ell(A) \le \|B\|'''. \tag{9.5.5}$$

vi) If $A, B \in \mathbb{F}^{n\times m}$ and either $\mathcal{R}(A) \subseteq \mathcal{R}(A+B)$ or $\mathcal{N}(A) \subseteq \mathcal{N}(A+B)$, then

$$\ell(A) - \|B\|''' \le \ell(A+B). \tag{9.5.6}$$

vii) If $A, B \in \mathbb{F}^{n\times m}$ and either $\mathcal{R}(A+B) \subseteq \mathcal{R}(A)$ or $\mathcal{N}(A+B) \subseteq \mathcal{N}(A)$, then

$$\ell(A+B) \le \ell(A) + \|B\|'''. \tag{9.5.7}$$

viii) If $n = m$ and $A \in \mathbb{F}^{n\times n}$ is nonsingular, then

$$\ell(A) = 1/\|A^{-1}\|'''. \tag{9.5.8}$$

Proof. See [243]. □

Proposition 9.5.3. Let $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ be norms on $\mathbb{F}^l$, $\mathbb{F}^m$, and $\mathbb{F}^n$, respectively, let $\|\cdot\|'''$ denote the norm on $\mathbb{F}^{m\times l}$ induced by $\|\cdot\|$ and $\|\cdot\|'$, let $\|\cdot\|''''$ denote the norm on $\mathbb{F}^{n\times m}$ induced by $\|\cdot\|'$ and $\|\cdot\|''$, and let $\|\cdot\|'''''$ denote the norm on $\mathbb{F}^{n\times l}$ induced by $\|\cdot\|$ and $\|\cdot\|''$. If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, then

$$\ell(A)\ell'(B) \le \ell''(AB). \tag{9.5.9}$$

In addition, the following statements hold:

i) If either $\operatorname{rank} B = \operatorname{rank} AB$ or $\operatorname{def} B = \operatorname{def} AB$, then

$$\ell''(AB) \le \|A\|''\ell(B). \tag{9.5.10}$$

ii) If $\operatorname{rank} A = \operatorname{rank} AB$, then

$$\ell''(AB) \le \ell(A)\|B\|''''. \tag{9.5.11}$$

iii) If $\operatorname{rank} B = m$, then

$$\|A\|''\ell(B) \le \|AB\|'''''. \tag{9.5.12}$$

iv) If $\operatorname{rank} A = m$, then

$$\ell(A)\|B\|'''' \le \|AB\|'''''. \tag{9.5.13}$$

Proof. See [243]. □

By assigning $\|\cdot\|_p$ to $\mathbb{F}^m$ and $\|\cdot\|_q$ to $\mathbb{F}^n$, the *Holder-induced lower bound* on $\mathbb{F}^{n\times m}$ is defined by

$$\ell_{q,p}(A) \triangleq \begin{cases} \min\limits_{y\in\mathcal{R}(A)\backslash\{0\}} \max\limits_{x\in\{z\in\mathbb{F}^m\colon Az=y\}} \frac{\|y\|'_q}{\|x\|_p}, & A \neq 0, \\ 0, & A = 0. \end{cases} \tag{9.5.14}$$

The following result shows that $\ell_{2,2}(A)$ is the smallest positive singular value of A.

Proposition 9.5.4. Let $A \in \mathbb{F}^{n\times m}$, assume that A is nonzero, and let $r \triangleq \operatorname{rank} A$. Then,

$$\ell_{2,2}(A) = \sigma_r(A). \tag{9.5.15}$$

Proof. The result follows from the singular value decomposition. □

Corollary 9.5.5. Let $A \in \mathbb{F}^{n\times m}$. If A is right invertible, then

$$\ell_{2,2}(A) = \sigma_n(A). \tag{9.5.16}$$

If A is left invertible, then

$$\ell_{2,2}(A) = \sigma_m(A). \tag{9.5.17}$$

Finally, if $n = m$ and A is nonsingular, then

$$\ell_{2,2}\big(A^{-1}\big) = \sigma_{\min}\big(A^{-1}\big) = \frac{1}{\sigma_{\max}(A)}. \tag{9.5.18}$$

Proof. Use Proposition 5.6.2 and Fact 6.3.12. □

In contrast to the submultiplicativity condition (9.4.5) satisfied by the induced norm, the induced lower bound satisfies a supermultiplicativity condition. The following result is analogous to Proposition 9.4.3.

Proposition 9.5.6. Let $\|\cdot\|$, $\|\cdot\|'$, and $\|\cdot\|''$ be norms on $\mathbb{F}^l$, $\mathbb{F}^m$, and $\mathbb{F}^n$, respectively. Let $\ell(\cdot)$ be the lower bound induced by $\|\cdot\|$ and $\|\cdot\|'$, let $\ell'(\cdot)$ be the lower bound induced by $\|\cdot\|'$ and $\|\cdot\|''$, let $\ell''(\cdot)$ be the lower bound induced by $\|\cdot\|$ and $\|\cdot\|''$, let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$, and assume that either A or B is right invertible. Then,

$$\ell'(A)\ell(B) \le \ell''(AB). \tag{9.5.19}$$

Furthermore, if $1 \le p, q, r \le \infty$, then

$$\ell_{r,q}(A)\ell_{q,p}(B) \le \ell_{r,p}(AB). \tag{9.5.20}$$

In particular,

$$\sigma_m(A)\sigma_l(B) \le \sigma_l(AB). \tag{9.5.21}$$

Proof. See [243] and [353, pp. 369, 370]. □

9.6 Singular Value Inequalities

Proposition 9.6.1. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, for all $i \in \{1,\ldots,\min\{n,m\}\}$ and $j \in \{1,\ldots,\min\{m,l\}\}$ such that $i+j \le \min\{n,l\}+1$,

$$\sigma_{i+j-1}(AB) \le \sigma_i(A)\sigma_j(B). \tag{9.6.1}$$

In particular, for all $j = 1,\ldots,\min\{n,m,l\}$,

$$\sigma_j(AB) \le \sigma_{\max}(A)\sigma_j(B). \tag{9.6.2}$$

and, for all $i = 1,\ldots,\min\{n,m,l\}$,

$$\sigma_i(AB) \le \sigma_i(A)\sigma_{\max}(B). \tag{9.6.3}$$

Proof. See [289, p. 178]. □

Proposition 9.6.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then, for all $k = 1,\ldots,\min\{n,m,l\}$,

$$\prod_{i=1}^{k}\sigma_i(AB) \le \prod_{i=1}^{k}\sigma_i(A)\sigma_i(B).$$

If, in addition, $n = m = l$, then

$$\prod_{i=1}^{n}\sigma_i(AB) = \prod_{i=1}^{n}\sigma_i(A)\sigma_i(B).$$

Proof. See [289, p. 172]. □

Proposition 9.6.3. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. If $r \ge 0$, then, for all $k = 1,\ldots,\min\{n,m,l\}$,

$$\sum_{i=1}^{k}\sigma_i^r(AB) \le \sum_{i=1}^{k}\sigma_i^r(A)\sigma_i^r(B). \tag{9.6.4}$$

In particular, for all $k = 1,\ldots,\min\{n,m,l\}$,

$$\sum_{i=1}^{k}\sigma_i(AB) \le \sum_{i=1}^{k}\sigma_i(A)\sigma_i(B). \tag{9.6.5}$$

If $r < 0$, $n = m = l$, and A and B are nonsingular, then

$$\sum_{i=1}^{n}\sigma_i^r(AB) \le \sum_{i=1}^{n}\sigma_i^r(A)\sigma_i^r(B). \tag{9.6.6}$$

Proof. The first statement follows from Proposition 9.6.2 and Fact 8.16.2. For the case $r < 0$, use Fact 8.16.4. See [289, p. 177] or [93, p.

94]. □

Proposition 9.6.4. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. If $m \le n$, then, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_m(A)\sigma_i(B) \le \sigma_i(AB). \tag{9.6.7}$$

If $m \le l$, then, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_i(A)\sigma_m(B) \le \sigma_i(AB). \tag{9.6.8}$$

Proof. Corollary 8.4.2 implies that $\sigma_m^2(A)I_m = \lambda_{\min}(A^*A)I_m \le A^*A$, which implies that $\sigma_m^2(A)B^*B \le B^*A^*AB$. Hence, it follows from the monotonicity theorem Theorem 8.4.9 that, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_m(A)\sigma_i(B) = \lambda_i\big[\sigma_m^2(A)B^*B\big]^{1/2} \le \lambda_i^{1/2}(B^*A^*AB) = \sigma_i(AB),$$

which proves the left-hand inequality in (9.6.7). Similarly, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_i(A)\sigma_m(B) = \lambda_i\big[\sigma_m^2(B)AA^*\big]^{1/2} \le \lambda_i^{1/2}(ABB^*A^*) = \sigma_i(AB). \qquad \square$$

Corollary 9.6.5. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\sigma_m(A)\sigma_{\min\{n,m,l\}}(B) \le \sigma_{\min\{n,m,l\}}(AB) \le \sigma_{\max}(A)\sigma_{\min\{n,m,l\}}(B), \tag{9.6.9}$$

$$\sigma_m(A)\sigma_{\max}(B) \le \sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B), \tag{9.6.10}$$

$$\sigma_{\min\{n,m,l\}}(A)\sigma_m(B) \le \sigma_{\min\{n,m,l\}}(AB) \le \sigma_{\min\{n,m,l\}}(A)\sigma_{\max}(B), \tag{9.6.11}$$

$$\sigma_{\max}(A)\sigma_m(B) \le \sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B). \tag{9.6.12}$$

Specializing Corollary 9.6.5 to the case in which A or B is square yields the following result.

Corollary 9.6.6. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{n\times l}$. Then, for all $i = 1, \ldots, \min\{n, l\}\}$,

$$\sigma_{\min}(A)\sigma_i(B) \le \sigma_i(AB) \le \sigma_{\max}(A)\sigma_i(B). \tag{9.6.13}$$

In particular,

$$\sigma_{\min}(A)\sigma_{\max}(B) \le \sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B). \tag{9.6.14}$$

If $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times m}$, then, for all $i = 1, \ldots, \min\{n, m\}\}$,

$$\sigma_i(A)\sigma_{\min}(B) \le \sigma_i(AB) \le \sigma_i(A)\sigma_{\max}(B). \tag{9.6.15}$$

In particular,

$$\sigma_{\max}(A)\sigma_{\min}(B) \le \sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B). \tag{9.6.16}$$

Corollary 9.6.7. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. If $m \le n$, then

$$\sigma_m(A)\|B\|_{\mathrm{F}} \le \|AB\|_{\mathrm{F}}. \tag{9.6.17}$$

If $m \le l$, then

$$\|A\|_{\mathrm{F}}\sigma_m(B) \le \|AB\|_{\mathrm{F}}. \tag{9.6.18}$$

Proposition 9.6.8. Let $A, B \in \mathbb{F}^{n\times m}$. Then, for all $i, j \in \{1, \ldots, \min\{n, m\}\}$ such that $i + j \le \min\{n, m\} + 1$,

$$\sigma_{i+j-1}(A + B) \le \sigma_i(A) + \sigma_j(B) \tag{9.6.19}$$

and

$$\sigma_{i+j-1}(A) - \sigma_j(B) \le \sigma_i(A + B). \tag{9.6.20}$$

Proof. See [289, p. 178]. □

Corollary 9.6.9. Let $A, B \in \mathbb{F}^{n\times m}$. Then,

$$\sigma_n(A) - \sigma_{\max}(B) \le \sigma_n(A + B) \le \sigma_n(A) + \sigma_{\max}(B). \tag{9.6.21}$$

Proof. The result follows from Proposition 9.6.8. Alternatively, it follows from Lemma 8.4.3 and the Cauchy-Schwarz inequality Corollary 9.1.7 that, for all $x \in \mathbb{F}^n$,

$$\begin{aligned}\lambda_{\min}[(A + B)(A + B)^*] &\le \frac{x^*(AA^* + BB^* + AB^* + BA^*)x}{x^*x}\\ &= \frac{x^*AA^*x}{\|x\|_2^2} + \frac{x^*BB^*x}{\|x\|_2^2} + \frac{2x^*AB^*x}{\|x\|_2^2}\\ &\le \frac{x^*AA^*x}{\|x\|_2^2} + \sigma_{\max}^2(B) + 2\frac{(x^*AA^*x)^{1/2}}{\|x\|_2^2}\sigma_{\max}(B).\end{aligned}$$

Minimizing with respect to x and using Lemma 8.4.3 yields

$$\begin{aligned}\sigma_n^2(A + B) &= \lambda_{\min}[(A + B)(A + B)^*]\\ &\le \lambda_{\min}(AA^*) + \sigma_{\max}^2(B) + 2\lambda_{\min}^{1/2}(AA^*)\sigma_{\max}(B)\\ &= [\sigma_n(A) + \sigma_{\max}(B)]^2,\end{aligned}$$

which proves the right-hand inequality of (9.6.21). Finally, the left-hand inequality follows from the right-hand inequality with A and B replaced by $A + B$ and $-B$, respectively. □

9.7 Facts on Vector Norms

Fact 9.7.1. Let $x, y \in \mathbb{F}^n$. Then, x and y are linearly dependent if and only if $|x|^{\{2\}}$ and $|y|^{\{2\}}$ are linearly dependent and $|x^*y| = |x|^{\mathrm{T}}|y|$. (Remark:

This equivalence clarifies the relationship between (9.1.11) with $p = 2$ and Corollary 9.1.7.)

Fact 9.7.2. Let $x, y \in \mathbb{F}^n$, and let $\|\cdot\|$ be a norm $\mathbb{F}^n$. Then,

$$|\|x\| - \|y\|| \le \|x+y\|$$

and

$$|\|x\| - \|y\|| \le \|x-y\|.$$

Fact 9.7.3. Let $x, y \in \mathbb{F}^n$, and let $\|\cdot\|$ be a norm on $\mathbb{F}^n$. Then, the following statements hold:

i) If there exists $\beta \ge 0$ such that either $x = \beta y$ or $y = \beta x$, then $\|x+y\| = \|x\| + \|y\|$.

ii) If $\|x+y\| = \|x\| + \|y\|$ and x and y are linearly dependent, then there exists $\beta \ge 0$ such that either $x = \beta y$ or $y = \beta x$.

iii) If $\|x+y\|_2 = \|x\|_2 + \|y\|_2$, then there exists $\beta \ge 0$ such that either $x = \beta y$ or $y = \beta x$.

(Proof: For *iii)* use *v)* of Fact 9.7.4.) (Problem: Consider *iii)* with alternative norms.) (Problem: If x and y are linearly independent, then does it follow that $\|x+y\| < \|x\| + \|y\|$?)

Fact 9.7.4. Let $x, y \in \mathbb{F}^n$. Then, the following statements hold:

i) $\frac{1}{2}\big(\|x+y\|_2^2 + \|x-y\|_2^2\big) = \|x\|_2^2 + \|y\|_2^2$.

ii) $\operatorname{Re} x^*y = \frac{1}{4}\big(\|x+y\|_2^2 - \|x-y\|_2^2\big) = \frac{1}{2}\big(\|x+y\|_2^2 - \|x\|_2^2 - \|y\|_2^2\big)$.

iii) $\|x-y\|_2 = \sqrt{\|x\|_2^2 + \|y\|_2^2 - 2\operatorname{Re} x^*y}$.

iv) $\|x+y\|_2\|x-y\|_2 \le \|x\|_2^2 + \|y\|_2^2$.

v) If $\|x+y\|_2 = \|x\|_2 + \|y\|_2$, then $\operatorname{Im} x^*y = 0$ and $\operatorname{Re} x^*y \ge 0$.

Furthermore, the following statements are equivalent:

vi) $\|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$.

vii) $\|x-y\|_2 = \|x+y\|_2$.

viii) $\operatorname{Re} x^*y = 0$.

(Remark: *i)* is the *parallelogram law*, which relates the diagonals and the sides of a parallelogram, *ii)* is the *polarization identity*, *iii)* is the *cosine law*, and the equivalence of *vi)* and *viii)* is the *Pythagorean theorem*.)

Fact 9.7.5. Let $x, y \in \mathbb{F}^n$ be nonzero. Then,

$$\|x\|_2 + \|y\|_2 \le \frac{2\|x-y\|}{\left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\|}.$$

(Proof: See [629, p. 28].) (Problem: Interpret this inequality geometrically.)

Fact 9.7.6. Let $x \in \mathbb{F}^n$, and let $p, q \in [1, \infty]$ satisfy $1/p + 1/q = 1$. Then,

$$\|x\|_2 \le \sqrt{\|x\|_p \|x\|_q}.$$

Fact 9.7.7. Let $x, y \in \mathbb{F}^n$, let $p \in (0, 1]$, and define $\|\cdot\|_p$ as in (9.1.3). Then,

$$\|x\|_p + \|y\|_p \le \|x + y\|_p.$$

(Remark: This result is a *reverse triangle inequality.*)

Fact 9.7.8. Let $y \in \mathbb{F}^n$, let $\|\cdot\|$ be a norm on $\mathbb{F}^n$, let $\|\cdot\|'$ be the norm on $\mathbb{F}^{n\times n}$ induced by $\|\cdot\|$, and define

$$\|y\|_{\mathrm{D}} \triangleq \max_{x\in\{z\in\mathbb{F}^n:\ \|z\|=1\}} |y^*x|.$$

Then, $\|\cdot\|_{\mathrm{D}}$ is a norm on $\mathbb{F}^n$. Furthermore,

$$\|y\| = \max_{x\in\{z\in\mathbb{F}^n:\ \|z\|_{\mathrm{D}}=1\}} |y^*x|.$$

Hence, for all $x \in \mathbb{F}^n$,

$$|x^*y| \le \|x\|\|y\|_{\mathrm{D}}.$$

In addition,

$$\|xy^*\|' = \|x\|\,\|y\|_{\mathrm{D}}.$$

Finally, let $p \in [1, \infty]$, and let $1/p + 1/q = 1$. Then,

$$\|\cdot\|_{p\mathrm{D}} = \|\cdot\|_q.$$

Hence, for all $x \in \mathbb{F}^n$,

$$|x^*y| \le \|x\|_p\|y\|_q$$

and

$$\|xy^*\|_{p,p} = \|x\|_p\|y\|_q.$$

(Proof: See [525, p. 57].) (Remark: $\|\cdot\|_{\mathrm{D}}$ is the *dual norm* of $\|\cdot\|$.)

Fact 9.7.9. Let $\|\cdot\|$ be a norm on $\mathbb{F}^n$, and let $\alpha > 0$. Then, $\{x \in \mathbb{F}^n\colon\ \|x\| \le \alpha\}$ is convex.

Fact 9.7.10. Let $x \in \mathbb{R}^n$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Then, $x^{\mathrm{T}}y > 0$ for all $y \in \mathbb{B}_{\|x\|}(x) = \{z \in \mathbb{R}^n\colon\ \|z - x\| < \|x\|\}$.

Fact 9.7.11. Let $x, y \in \mathbb{R}^n$ be nonzero, assume that $x^{\mathrm{T}}y = 0$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Then, $\|x\| \le \|x+y\|$. (Proof: If $\|x+y\| < \|x\|$, then $x+y \in \mathbb{B}_{\|x\|}(0)$, and thus $y \in \mathbb{B}_{\|x\|}(-x)$. By Fact 9.7.10, $x^{\mathrm{T}}y < 0$.) (Remark: See [98, 371] for related results concerning matrices.)

Fact 9.7.12. Let $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$. Then,

$$\sigma_{\max}(xy^*) = \|xy^*\|_{\mathrm{F}} = \|x\|_2\|y\|_2$$

and

$$\sigma_{\max}(xx^*) = \|xx^*\|_{\mathrm{F}} = \|x\|_2^2.$$

Fact 9.7.13. Let $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$. Then,

$$\|x \otimes y\|_2 = \left\|\mathrm{vec}\big(x \otimes y^{\mathrm{T}}\big)\right\|_2 = \left\|\mathrm{vec}\big(yx^{\mathrm{T}}\big)\right\|_2 = \left\|yx^{\mathrm{T}}\right\|_2 = \|x\|_2\|y\|_2.$$

Fact 9.7.14. Let $x \in \mathbb{F}^n$, and let $1 \le p, q \le \infty$. Then,

$$\|x\|_p = \|x\|_{p,q}.$$

Fact 9.7.15. Let $x \in \mathbb{F}^n$, and let $p, q \in [1, \infty)$, where $p \le q$. Then,

$$\|x\|_q \le \|x\|_p \le n^{1/p-1/q}\|x\|_q.$$

(Proof: See [279], [280, p. 107].) (Remark: See Fact 9.8.13.)

Fact 9.7.16. Let $A \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$\|x\|_A \triangleq (x^*Ax)^{1/2}$$

is a norm on $\mathbb{F}^n$.

Fact 9.7.17. Let $\|\cdot\|$ and $\|\cdot\|'$ be norms on $\mathbb{F}^n$ and let $\alpha, \beta > 0$. Then, $\alpha\|\cdot\| + \beta\|\cdot\|'$ is also a norm on $\mathbb{F}^n$. Furthermore, $\max\{\|\cdot\|, \|\cdot\|'\}$ is a norm on $\mathbb{F}^n$. (Remark: $\min\{\|\cdot\|, \|\cdot\|'\}$ is not generally a norm.)

Fact 9.7.18. Let $A \in \mathbb{F}^{n\times n}$, assume that A is nonsingular, and let $\|\cdot\|$ be a norm on $\mathbb{F}^n$. Then, $\|x\|' \triangleq \|Ax\|$ is a norm on $\mathbb{F}^n$.

Fact 9.7.19. Let $x \in \mathbb{F}^n$, and let $p \in [1, \infty]$. Then,

$$\|\overline{x}\|_p = \|x\|_p.$$

9.8 Facts on Matrix Norms Involving One Matrix

Fact 9.8.1. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\mathrm{sprad}(A) < 1$. Then, there exists a submultiplicative matrix norm $\|\cdot\|$ on $\mathbb{F}^{n\times n}$ such that $\|A\| < 1$.

Furthermore,

$$\lim_{k\to\infty} A^k = 0.$$

Fact 9.8.2. Let $A \in \mathbb{F}^{n\times n}$ be nonsingular, and let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$. Then,

$$\|A^{-1}\| \ge \|I_n\|/\|A\|.$$

Fact 9.8.3. Let $A \in \mathbb{F}^{n\times n}$, assume that A is nonzero and idempotent, and let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$. Then,

$$\|A\| \ge 1.$$

Fact 9.8.4. Let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then, $\|\cdot\|$ is self adjoint.

Fact 9.8.5. Let $A \in \mathbb{F}^{n\times m}$, let $\|\cdot\|$ be a norm on $\mathbb{F}^{n\times m}$, and define $\|A\|' \triangleq \|A^*\|$. Then, $\|\cdot\|'$ is a norm on $\mathbb{F}^{m\times n}$. If, in addition, $n = m$ and $\|\cdot\|$ is induced by $\|\cdot\|''$, then $\|\cdot\|'$ is induced by $\|\cdot\|''_{\mathrm{D}}$. (Proof: See [287, p. 309] and Fact 9.8.8.) (Remark: See Fact 9.7.8 for the definition of the dual norm. $\|\cdot\|'$ is the *adjoint norm* of $\|\cdot\|$.) (Problem: Generalize this result to matrices that are not square and norms that are not equi-induced.)

Fact 9.8.6. Let $1 \le p \le \infty$. Then, $\|\cdot\|_{\sigma p}$ is unitarily invariant.

Fact 9.8.7. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is nonnegative semidefinite. Then,

$$\|A\|_{1,\infty} = \max_{x\in\{z\in\mathbb{F}^n\colon\ \|z\|_\infty=1\}} x^*Ax.$$

(Remark: This result is due to Tao. See [490] and [280, p. 116].)

Fact 9.8.8. Let $A \in \mathbb{F}^{n\times m}$, and let $p, q \in [1,\infty]$ be such that $1/p+1/q = 1$. Then,

$$\|A^*\|_{p,p} = \|A\|_{q,q}.$$

In particular,

$$\|A^*\|_{\mathrm{col}} = \|A\|_{\mathrm{row}}.$$

(Proof: See Fact 9.8.5.)

Fact 9.8.9. Let $A \in \mathbb{F}^{n\times m}$, and let $p, q \in [1,\infty]$ be such that $1/p+1/q = 1$. Then,

$$\left\|\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}\right\|_{p,p} = \max\{\|A\|_{p,p}, \|A\|_{q,q}\}.$$

In particular,

$$\left\| \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \right\|_{\mathrm{col}} = \left\| \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \right\|_{\mathrm{row}} = \max\{\|A\|_{\mathrm{col}}, \|A\|_{\mathrm{row}}\}.$$

Fact 9.8.10. Let $A \in \mathbb{F}^{n\times m}$. Then, the following inequalities hold:

i) $\|A\|_{\mathrm{F}} \le \|A\|_1 \le \sqrt{mn}\|A\|_{\mathrm{F}}$.

ii) $\|A\|_\infty \le \|A\|_1 \le mn\|A\|_\infty$.

iii) $\|A\|_{\mathrm{col}} \le \|A\|_1 \le m\|A\|_{\mathrm{col}}$.

iv) $\|A\|_{\mathrm{row}} \le \|A\|_1 \le n\|A\|_{\mathrm{row}}$.

v) $\sigma_{\max}(A) \le \|A\|_1 \le \sqrt{mn\,\mathrm{rank}\,A}\,\sigma_{\max}(A)$.

vi) $\|A\|_\infty \le \|A\|_{\mathrm{F}} \le \sqrt{mn}\|A\|_\infty$.

vii) $\frac{1}{\sqrt{n}}\|A\|_{\mathrm{col}} \le \|A\|_{\mathrm{F}} \le \sqrt{m}\|A\|_{\mathrm{col}}$.

viii) $\frac{1}{\sqrt{m}}\|A\|_{\mathrm{row}} \le \|A\|_{\mathrm{F}} \le \sqrt{n}\|A\|_{\mathrm{row}}$.

ix) $\sigma_{\max}(A) \le \|A\|_{\mathrm{F}} \le \sqrt{\mathrm{rank}\,A}\,\sigma_{\max}(A)$.

x) $\frac{1}{n}\|A\|_{\mathrm{col}} \le \|A\|_\infty \le \|A\|_{\mathrm{col}}$.

xi) $\frac{1}{m}\|A\|_{\mathrm{row}} \le \|A\|_\infty \le \|A\|_{\mathrm{row}}$.

xii) $\frac{1}{\sqrt{mn}}\sigma_{\max}(A) \le \|A\|_\infty \le \sigma_{\max}(A)$.

xiii) $\frac{1}{m}\|A\|_{\mathrm{row}} \le \|A\|_{\mathrm{col}} \le n\|A\|_{\mathrm{row}}$.

xiv) $\frac{1}{\sqrt{m}}\sigma_{\max}(A) \le \|A\|_{\mathrm{col}} \le \sqrt{n}\sigma_{\max}(A)$.

xv) $\frac{1}{\sqrt{n}}\sigma_{\max}(A) \le \|A\|_{\mathrm{row}} \le \sqrt{m}\sigma_{\max}(A)$.

(Remark: See [280, p. 115] for matrices that attain these bounds.)

Fact 9.8.11. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\|A^{\mathrm{A}}\|_{\mathrm{F}} \le n^{(2-n)/2}\|A\|_{\mathrm{F}}^{n-1}.$$

(Proof: See [466, pp. 151, 165].)

Fact 9.8.12. Let $A \in \mathbb{F}^{n\times n}$, let $\|\cdot\|$ and $\|\cdot\|'$ be norms on $\mathbb{F}^n$, and define the induced norms

$$\|A\|'' \triangleq \max_{x\in\{y\in\mathbb{F}^m\colon\ \|y\|=1\}} \|Ax\|$$

and

$$\|A\|''' \triangleq \max_{x\in\{y\in\mathbb{F}^m\colon\ \|y\|'=1\}} \|Ax\|'.$$

Then,

$$\max_{A\in\{X\in\mathbb{F}^{n\times n}\colon X\neq 0\}} \frac{\|A\|''}{\|A\|'''} = \max_{A\in\{X\in\mathbb{F}^{n\times n}\colon X\neq 0\}} \frac{\|A\|'''}{\|A\|''}$$

$$= \max_{x\in\{y\in\mathbb{F}^{n}\colon y\neq 0\}} \frac{\|x\|}{\|x\|'} \max_{x\in\{y\in\mathbb{F}^{n}\colon y\neq 0\}} \frac{\|x\|'}{\|x\|}.$$

(Proof: See [287, p. 303].) (Remark: This symmetry property is evident in Fact 9.8.10.)

Fact 9.8.13. Let $A \in \mathbb{F}^{n\times n}$, and let $p, q \in [1,\infty]$. Then,

$$\|A\|_{p,p} \le \begin{cases} n^{1/p-1/q}\|A\|_{q,q}, & p \le q, \\ n^{1/q-1/p}\|A\|_{q,q}, & q \le p. \end{cases}$$

Consequently,

$$n^{1/p-1}\|A\|_{\rm col} \le \|A\|_{p,p} \le n^{1-1/p}\|A\|_{\rm col},$$

$$n^{-|1/p-1/2|}\sigma_{\max}(A) \le \|A\|_{p,p} \le n^{|1/p-1/2|}\sigma_{\max}(A),$$

$$n^{-1/p}\|A\|_{\rm col} \le \|A\|_{p,p} \le n^{1/p}\|A\|_{\rm row}.$$

(Proof: See [279] and [280, p. 112].) (Remark: See Fact 9.7.15.) (Problem: Extend these inequalities to matrices that are not square.)

Fact 9.8.14. Let $A \in \mathbb{F}^{n\times m}$, $p, q \in [1,\infty]$, and $\alpha \in [0,1]$, and let $r \triangleq pq/[(1-\alpha)p + \alpha q]$. Then,

$$\|A\|_{r,r} \le \|A\|_{p,p}^{\alpha}\|A\|_{q,q}^{1-\alpha}.$$

(Proof: See [279] or [280, p. 113].)

Fact 9.8.15. Let $A \in \mathbb{F}^{n\times m}$, and let $p \in [1,\infty]$. Then,

$$\|A\|_{p,p} \le \|A\|_{\rm col}^{1/p}\|A\|_{\rm row}^{1-1/p}.$$

In particular,

$$\sigma_{\max}(A) \le \sqrt{\|A\|_{\rm col}\|A\|_{\rm row}}.$$

(Proof: Set $\alpha = 1/p$, $p = 1$, and $q = \infty$ in Fact 9.8.14. See [280, p. 113]. To prove the special case $p = 2$ directly, note that $\lambda_{\max}(A^*A) \le \|A^*A\|_{\rm col} \le \|A^*\|_{\rm col}\|A\|_{\rm col} = \|A\|_{\rm row}\|A\|_{\rm col}$.)

Fact 9.8.16. Let $A \in \mathbb{F}^{n\times m}$, and let $p \in [1,2]$. Then,

$$\|A\|_{p,p} \le \|A\|_{\rm col}^{2/p-1}\sigma_{\max}^{2-2/p}(A).$$

(Proof: Let $\alpha = 2/p - 1$, $p = 1$, and $q = 2$ in Fact 9.8.14. See [280, p. 113].)

Fact 9.8.17. Let $A \in \mathbb{F}^{n\times n}$, and let $p \in [1,\infty]$. Then,

$$\|A\|_{p,p} \le \||A|\|_{p,p} \le n^{\min\{1/p,1-1/p\}}\|A\|_{p,p} \le \sqrt{n}\|A\|_{p,p}.$$

(Remark: See [280, p. 117].)

Fact 9.8.18. Let $A \in \mathbb{F}^{n\times m}$, and let $p,q \in [1,\infty]$. Then,

$$\|\overline{A}\|_{q,p} = \|A\|_{q,p}.$$

Fact 9.8.19. Let $A \in \mathbb{F}^{n\times m}$, and let $p,q \in [1,\infty]$. Then,

$$\|A^*\|_{q,p} = \|A\|_{p/(p-1),q/(q-1)}.$$

Fact 9.8.20. Let $A \in \mathbb{F}^{n\times m}$, and let $p,q \in [1,\infty]$. Then,

$$\|A\|_{q,p} \le \begin{cases} \|A\|_{p/(p-1)}, & 1/p+1/q \le 1, \\ \|A\|_q, & 1/p+1/q \ge 1. \end{cases}$$

Fact 9.8.21. Let $A \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|\langle A\rangle\| = \|A\|.$$

Fact 9.8.22. Let $A \in \mathbb{F}^{n\times n}$, let $S \in \mathbb{F}^{n\times n}$ be nonsingular, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|A\| \le \tfrac{1}{2}\|SAS^{-1} + S^*AS^{-*}\|.$$

(Proof: See [30, 107].)

Fact 9.8.23. Let $A \in \mathbb{F}^{n\times n}$, assume that A is nonnegative semidefinite, and let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$. Then,

$$\|A\|^{1/2} \le \left\|A^{1/2}\right\|.$$

In particular,

$$\sigma_{\max}^{1/2}(A) = \sigma_{\max}\left(A^{1/2}\right).$$

Fact 9.8.24. Let $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}$ be nonnegative semidefinite, let $\|\cdot\|$ and $\|\cdot\|'$ be unitarily invariant norms on $\mathbb{F}^{n\times n}$ and $\mathbb{F}^{m\times m}$, respectively, and let $p > 0$. Then,

$$\|\langle A_{12}\rangle^p\|'^2 \le \|A_{11}^p\|\|A_{22}^p\|'.$$

(Proof: See [291].)

Fact 9.8.25. Let $A \in \mathbb{F}^{n\times n}$, let $\|\cdot\|$ be a norm on $\mathbb{F}^n$, let $\|\cdot\|_{\mathrm{D}}$ denote the dual norm on $\mathbb{F}^n$, and let $\|\cdot\|'$ denote norm induced by $\|\cdot\|$ on $\mathbb{F}^{n\times n}$.

Then,

$$\|A\|' = \max_{\substack{x,y\in\mathbb{F}^n\\ x,y\neq 0}} \frac{\operatorname{Re} y^*Ax}{\|y\|_{\mathrm{D}}\|x\|}.$$

(Proof: See [280, p. 115].) (Remark: See Fact 9.7.8 for the definition of the dual norm.) (Problem: Generalize this result to obtain Fact 9.8.26 as a special case.)

Fact 9.8.26. Let $A\in\mathbb{F}^{n\times m}$, and let $p,q\in[1,\infty]$. Then,

$$\|A\|_{q,p} = \max_{\substack{x\in\mathbb{F}^m, y\in\mathbb{F}^n\\ x,y\neq 0}} \frac{|y^*Ax|}{\|y\|_{q/(q-1)}\|x\|_p}.$$

Fact 9.8.27. Let $A\in\mathbb{F}^{n\times m}$, and let $p,q\in[1,\infty]$ satisfy $1/p+1/q=1$. Then,

$$\|A\|_{p,p} = \max_{\substack{x\in\mathbb{F}^m, y\in\mathbb{F}^n\\ x,y\neq 0}} \frac{|y^*Ax|}{\|y\|_q\|x\|_p} = \max_{\substack{x\in\mathbb{F}^m, y\in\mathbb{F}^n\\ x,y\neq 0}} \frac{|y^*Ax|}{\|y\|_{p/(p-1)}\|x\|_p}.$$

(Remark: See Fact 9.11.2 for the case $p=2$.)

Fact 9.8.28. Let $A\in\mathbb{F}^{n\times n}$, and assume that A is positive definite. Then,

$$\min_{x\in\mathbb{F}^n\setminus\{0\}} \frac{x^*Ax}{\|Ax\|_2\|x\|_2} = \frac{2\sqrt{\alpha\beta}}{\alpha+\beta}$$

and

$$\min_{\alpha\geq 0}\sigma_{\max}(\alpha A - I) = \frac{\alpha-\beta}{\alpha+\beta}.$$

where $\alpha \triangleq \lambda_{\max}(A)$ and $\beta \triangleq \lambda_{\min}(A)$. (Proof: See [251].) (Remark: These quantities are *antieigenvalues*.)

Fact 9.8.29. Let $A\in\mathbb{F}^{n\times n}$, and define

$$\operatorname{nrad}(A) \triangleq \max\{|x^*Ax|:\ x\in\mathbb{C}^n \text{ and } x^*x\leq 1\}.$$

Then, the following statements hold:

i) $\operatorname{nrad}(A) = \max\{|z|:\ z\in\Theta(A)\}$.

ii) $\operatorname{sprad}(A) \leq \operatorname{nrad}(A) \leq \operatorname{nrad}(|A|) = \frac{1}{2}\operatorname{sprad}\big(|A|+|A|^{\mathrm{T}}\big)$.

iii) $\frac{1}{2}\sigma_{\max}(A) \leq \operatorname{nrad}(A) \leq \sigma_{\max}(A)$.

iv) If A is normal, then $\operatorname{sprad}(A) = \operatorname{nrad}(A) = \sigma_{\max}(A)$.

v) $\operatorname{nrad}\big(A^k\big) \leq [\operatorname{nrad}(A)]^k$ for all $k\in\mathbb{N}$.

vi) $\operatorname{nrad}(\cdot)$ is a weakly unitarily invariant norm on $\mathbb{F}^{n\times n}$.

vii) $\mathrm{nrad}(\cdot)$ is not a submultiplicative norm on $\mathbb{F}^{n\times n}$.

viii) $\|\cdot\| \triangleq \alpha\mathrm{nrad}(\cdot)$ is a submultiplicative norm on $\mathbb{F}^{n\times n}$ if and only if $\alpha \ge 4$.

ix) $\mathrm{nrad}(AB) \le \mathrm{nrad}(A)\mathrm{nrad}(B)$ for all $A, B \in \mathbb{F}^{n\times n}$ such that either A or B is normal.

x) $\mathrm{nrad}(A \circ B) \le \alpha\mathrm{nrad}(A)\mathrm{nrad}(B)$ for all $A, B \in \mathbb{F}^{n\times n}$ if and only if $\alpha \ge 2$.

xi) $\mathrm{nrad}(A \oplus B) = \max\{\mathrm{nrad}(A), \mathrm{nrad}(B)\}$ for all $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$.

(Proof: See [287, p. 331] and [289, pp. 43, 44].) (Remark: $\mathrm{nrad}(\cdot)$ is not submultiplicative. $\mathrm{nrad}(A)$ is the *numerical radius* of A. $\Theta(A)$ is the numerical range. See Fact 4.10.17.) (Remark: *vii)* is the *power inequality.*)

Fact 9.8.30. Let $A \in \mathbb{F}^{n\times m}$, let $\gamma > \sigma_{\max}(A)$, and define $\beta \triangleq \sigma_{\max}(A)/\gamma$. Then,

$$\|A\|_{\mathrm{F}} \le \sqrt{-[\gamma^2/(2\pi)]\log\det(I - \gamma^{-2}A^*A)} \le \beta^{-1}\sqrt{-\log(1-\beta^2)}\|A\|_{\mathrm{F}}.$$

(Proof: See [108].)

Fact 9.8.31. Let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then, $\|A\| = 1$ for all $A \in \mathbb{F}^{n\times n}$ such that $\mathrm{rank}\, A = 1$ if and only if $\|E_{1,1}\| = 1$. (Proof: $\|A\| = \|E_{1,1}\|\sigma_{\max}(A)$.) (Remark: These equivalent normalizations are used in [525, p. 74] and [93], respectively.)

Fact 9.8.32. Let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) $\sigma_{\max}(A) \le \|A\|$ for all $A \in \mathbb{F}^{n\times n}$.

ii) $\|\cdot\|$ is submultiplicative.

iii) $\|A^2\| \le \|A\|^2$ for all $A \in \mathbb{F}^{n\times n}$.

iv) $\|A^k\| \le \|A\|^k$ for all $k \in \mathbb{P}$ and $A \in \mathbb{F}^{n\times n}$.

v) $\|A \circ B\| \le \|A\|\|B\|$ for all $A, B \in \mathbb{F}^{n\times n}$.

vi) $\mathrm{sprad}(A) \le \|A\|$ for all $A \in \mathbb{C}^{n\times n}$.

vii) $\|Ax\|_2 \le \|A\|\|x\|_2$ for all $A \in \mathbb{C}^{n\times n}$ and $x \in \mathbb{C}^n$.

viii) $\|A\|_\infty \le \|A\|$ for all $A \in \mathbb{C}^{n\times n}$.

ix) $\|E_{1,1}\| \ge 1$.

x) $\sigma_{\max}(A) \le \|A\|$ for all $A \in \mathbb{C}^{n\times n}$ such that $\mathrm{rank}\, A = 1$.

(Proof: The equivalence of *i)* – *vii)* is given in [288] and [289, p. 211]. Since

$\|A\| = \|E_{1,1}\|\sigma_{\max}(A)$ for all $A \in \mathbb{F}^{n\times n}$ such that $\operatorname{rank} A = 1$, it follows that *vii*) and *viii*) are equivalent. To prove *ix*) $\Longrightarrow$ *x*) let $A \in \mathbb{C}^{n\times n}$ satisfy $\operatorname{rank} A = 1$. Then, $\|A\| = \sigma_{\max}(A)\|E_{1,1}\| \geq \sigma_{\max}(A)$. To show *x*) $\Longrightarrow$ *ii*), define $\|\cdot\|' \triangleq \|E_{1,1}\|^{-1}\|\cdot\|$. Since $\|E_{1,1}\|' = 1$, it follows from [93, p. 94] that $\|\cdot\|'$ is submultiplicative. Since $\|E_{1,1}\|^{-1} \leq 1$, it follows that $\|\cdot\|$ is also submultiplicative. Alternatively, $\|A\|' = \sigma_{\max}(A)$ for all $A \in \mathbb{F}^{n\times n}$ having rank 1. Then, Corollary 3.10 of [525, p. 80] implies that $\|\cdot\|'$, and thus $\|\cdot\|$ is submultiplicative.)

Fact 9.8.33. Let $\Phi\colon\ \mathbb{F}^n \mapsto \mathbb{F}$ satisfy the following conditions:

i) If $x \neq 0$, then $\Phi(x) > 0$.

ii) $\Phi(\alpha x) = |\alpha|\Phi(x)$ for all $\alpha \in \mathbb{R}$.

iii) $\Phi(x+y) \leq \Phi(x) + \Phi(y)$ for all $x, y \in \mathbb{F}^n$.

iv) If $A \in \mathbb{F}^{n\times n}$ is a permutation matrix, then $\Phi(Ax) = \Phi(x)$ for all $x \in \mathbb{F}^n$.

v) $\Phi(|x|) = \Phi(x)$ for all $x \in \mathbb{F}^n$.

Furthermore, for $A \in \mathbb{F}^{n\times m}$, define

$$\|A\| \triangleq \Phi(\sigma_1(A), \ldots, \sigma_n(A)).$$

Then, $\|\cdot\|$ is a unitarily invariant norm. Conversely, if $\|\cdot\|$ is a unitarily invariant norm on $\mathbb{F}^{n\times m}$, where $n \leq m$, then $\Phi\colon\ \mathbb{F}^n \mapsto \mathbb{F}$ defined by

$$\Phi(x) \triangleq \left\| \begin{bmatrix} x_{(1)} & & 0 & \\ & \ddots & & \\ & & x_{(n)} & \\ 0 & & & 0_{n\times(m-n)} \end{bmatrix} \right\|$$

satisfies *i*)-*v*). (Proof: See [525, pp. 75–76].) (Remark: Φ is a *symmetric gauge function*. This result is due to von Neumann. See Fact 8.16.6.)

Fact 9.8.34. Let $\|\cdot\|$ and $\|\cdot\|'$ denote norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively, and define $\hat{\ell}\colon\ \mathbb{F}^{n\times m} \mapsto \mathbb{R}$ by

$$\hat{\ell}(A) \triangleq \min_{x\in\mathbb{F}^m\backslash\{0\}} \frac{\|Ax\|'}{\|x\|},$$

or, equivalently,

$$\hat{\ell}(A) \triangleq \min_{x\in\{y\in\mathbb{F}^m:\ \|y\|=1\}} \|Ax\|'.$$

Then, for $A \in \mathbb{F}^{n\times n}$, the following statements hold:

i) $\hat{\ell}(A) \geq 0$.

ii) $\hat{\ell}(A) > 0$ if and only if $\operatorname{rank} A = m$

iii) $\hat{\ell}(A) = \ell(A)$ if and only if either $A = 0$ or $\operatorname{rank} A = m$.

(Proof: See [353, pp. 369, 370].) (Remark: $\hat{\ell}$ is a weaker version of ℓ.)

9.9 Facts on Matrix Norms Involving Two or More Matrices

Fact 9.9.1. $\|\cdot\|'_\infty \triangleq n\|\cdot\|_\infty$ is submultiplicative on $\mathbb{F}^{n\times n}$. (Remark: It is not generally true that $\|AB\|_\infty \le \|A\|_\infty\|B\|_\infty$. For example, let $A = B = \left[\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right]$.)

Fact 9.9.2. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$. Then,

$$\|AB\|_\infty \le m\|A\|_\infty\|B\|_\infty.$$

Furthermore, if $A = 1_{n\times m}$ and $B = 1_{m\times l}$, then $\|AB\|_\infty = m\|A\|_\infty\|B\|_\infty$.

Fact 9.9.3. Let $A, B \in \mathbb{F}^{n\times n}$ and let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$. Then, $\|AB\| \le \|A\|\|B\|$. Hence, if $\|A\| \le 1$ and $\|B\| \le 1$, then $\|AB\| \le 1$, and if either $\|A\| < 1$ or $\|B\| < 1$, then $\|AB\| < 1$. (Remark: $\operatorname{sprad}(A) < 1$ and $\operatorname{sprad}(B) < 1$ do not imply that $\operatorname{sprad}(AB) < 1$. Let $A = B^{\mathrm{T}} = \left[\begin{smallmatrix}0 & 2\\ 0 & 0\end{smallmatrix}\right]$.)

Fact 9.9.4. Let $\|\cdot\|$ be a norm on $\mathbb{F}^{m\times m}$, and let

$$\delta > \sup\left\{\frac{\|AB\|}{\|A\|\|B\|}:\ A, B \in \mathbb{F}^{m\times m}, A, B \ne 0\right\}.$$

Then, $\|\cdot\|' = \delta\|\cdot\|$ is a submultiplicative norm on $\mathbb{F}^{m\times m}$. (Proof: See [287, p. 323].)

Fact 9.9.5. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$. Then, $\|\cdot\|' \triangleq 2\|\cdot\|$ is submultiplicative and satisfies

$$\|[A, B]\|' \le \|A\|'\|B\|'.$$

Fact 9.9.6. Let $\|\cdot\|$ be a normalized, submultiplicative norm on $\mathbb{F}^{n\times n}$. Then, $\|\cdot\ \|$ is equi-induced if and only if $\|A\| \le \|A\|'$ for all $A \in \mathbb{F}^{n\times n}$ and for all normalized submultiplicative norms $\|\cdot\|'$ on $\mathbb{F}^{n\times n}$. (Proof: See [528].) (Remark: As shown in [138, 164], not every normalized submultiplicative norm on $\mathbb{F}^{n\times n}$ is equi-induced or induced.)

Fact 9.9.7. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are Hermitian, let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$, and let $k \in \mathbb{N}$. Then,

$$\|(A - B)^{2k+1}\| \le 2^{2k}\|A^{2k+1} - B^{2k+1}\|.$$

(Proof: See [93, p. 294].)

Fact 9.9.8. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are nonnegative semidefinite, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|(A-B)^2\| \le \|A^2 - B^2\|.$$

(Proof: See [336].)

Fact 9.9.9. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A and B are nonnegative semidefinite. Then,

$$\|AB - BA\|_{\mathrm{F}}^2 + \|(A-B)^2\|_{\mathrm{F}}^2 \le \|A^2 - B^2\|_{\mathrm{F}}^2.$$

(Proof: See [336].)

Fact 9.9.10. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|AB\| \le \sigma_{\max}(A)\|B\|$$

and

$$\|AB\| \le \|A\|\sigma_{\max}(B).$$

(Proof: See [336].)

Fact 9.9.11. Let $A, B \in \mathbb{F}^{n\times m}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. If $p > 0$, then

$$\|\langle B^*A\rangle^p\|^2 \le \|(A^*A)^p\|\,\|(B^*B)^p\|.$$

In particular,

$$\|(A^*BB^*A)^{1/4}\|^2 \le \|A\|\,\|B\|$$

and

$$\|A^*B\|^2 \le \|A^*A\|\,\|B^*B\|.$$

Furthermore,

$$\operatorname{tr}\langle B^*A\rangle \le \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}$$

and

$$\left[\operatorname{tr}(A^*BB^*A)^{1/4}\right]^2 \le (\operatorname{tr}\langle A\rangle)(\operatorname{tr}\langle B\rangle).$$

(Proof: See [291].) (Problem: Noting Fact 9.10.5, compare the lower bounds for $\|A\|_{\mathrm{F}}^2\|B\|_{\mathrm{F}}^2$ given by

$$|\operatorname{tr}(A^*B)^2| \le \operatorname{tr} AA^*BB^* \le \|A\|_{\mathrm{F}}^2\|B\|_{\mathrm{F}}^2$$

and

$$\left[\operatorname{tr}(A^*BB^*A)^{1/2}\right]^2 \le \|A\|_{\mathrm{F}}^2\|B\|_{\mathrm{F}}^2.)$$

Fact 9.9.12. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$(2\|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}})^{1/2} \le \left[\|A\|_{\mathrm{F}}^2 + \|B\|_{\mathrm{F}}^2\right]^{1/2} \le \|A+B\|_{\mathrm{F}} \le \sqrt{2}\left[\|A\|_{\mathrm{F}}^2 + \|B\|_{\mathrm{F}}^2\right]^{1/2}.$$

Fact 9.9.13. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|AB\| \le \tfrac{1}{4}\|(A+B)^2\|.$$

In particular,

$$\operatorname{tr} AB \le \operatorname{tr}\left(AB^2A\right)^{1/2} \le \tfrac{1}{4}\operatorname{tr}\,(A+B)^2,$$
$$\operatorname{tr}\,(AB)^2 \le \operatorname{tr} A^2B^2 \le \tfrac{1}{16}\operatorname{tr}\,(A+B)^4,$$
$$\sigma_{\max}(AB) \le \tfrac{1}{4}\sigma_{\max}\left[(A+B)^2\right].$$

(Proof: See [625, p. 77] or [97]. The inequalities $\operatorname{tr} AB \le \operatorname{tr}\left(AB^2A\right)^{1/2}$ and $\operatorname{tr}\,(AB)^2 \le \operatorname{tr} A^2B^2$ follow from Fact 8.12.8.) (Problem: Noting Fact 9.9.12, compare the lower bounds for $\|A+B\|_{\mathrm{F}}$ given by

$$(2\|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}})^{1/2} \le \left[\|A\|_{\mathrm{F}}^2 + \|B\|_{\mathrm{F}}^2\right]^{1/2} \le \|A+B\|_{\mathrm{F}}$$

and

$$2\|AB\|_{\mathrm{F}}^{1/2} \le \|(A+B)^2\|_{\mathrm{F}}^{1/2} \le \|A+B\|_{\mathrm{F}}.)$$

Fact 9.9.14. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times l}$, and $p, q, q', r \in [1,\infty]$, and assume that $1/q + 1/q' = 1$. Then,

$$\|AB\|_p \le \varepsilon_{pq}(n)\varepsilon_{pr}(l)\varepsilon_{q'r}(m)\|A\|_q\|B\|_r,$$

where

$$\varepsilon_{pq}(n) \triangleq \begin{cases} 1, & p \ge q, \\ n^{1/p-1/q}, & q \ge p. \end{cases}$$

Furthermore, there exist $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times l}$ such that equality holds. (Proof: See [233].) (Remark: Related results are given in [198, 233–235, 366, 552]

Fact 9.9.15. Let $A, B \in \mathbb{C}^{n\times m}$. Then, there exist unitary matrices $S_1, S_2 \in \mathbb{C}^{m\times m}$ such that

$$\langle A+B\rangle \le S_1\langle A\rangle S_1^* + S_2\langle B\rangle S_2^*.$$

(Remark: This result is a matrix version of the triangle inequality. See [24, 546].)

Fact 9.9.16. Let $A, X, B \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|A^*XB\| \le \tfrac{1}{2}\|AA^*X + XBB^*\|.$$

In particular,

$$\|A^*B\| \le \tfrac{1}{2}\|AA^* + BB^*\|.$$

(Proof: See [94, 96].) (Remark: See Fact 9.12.20.)

Fact 9.9.17. Let $A, B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite, and let $p \in [1,\infty]$. Then,

$$\|A - B\|_{\sigma 2p}^2 \le \|A^2 - B^2\|_{\sigma p}.$$

(Proof: See [332].)

Fact 9.9.18. Let $A, B \in \mathbb{F}^{n\times n}$. If $p \in (0,2]$, then

$$2^{p-1}(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p) \le \|A+B\|_{\sigma p}^p + \|A-B\|_{\sigma p}^p \le 2(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p).$$

If $p \in [2,\infty)$, then

$$2(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p) \le \|A+B\|_{\sigma p}^p + \|A-B\|_{\sigma p}^p \le 2^{p-1}(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p).$$

If $p \in (1,2]$ and $1/p + 1/q = 1$, then

$$\|A+B\|_{\sigma p}^q + \|A-B\|_{\sigma p}^q \le 2(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p)^{q/p}.$$

If $p \in [2,\infty)$ and $1/p + 1/q = 1$, then

$$2(\|A\|_{\sigma p}^p + \|B\|_{\sigma p}^p)^{q/p} \le \|A+B\|_{\sigma p}^q + \|A-B\|_{\sigma p}^q.$$

(Proof: See [283].) (Remark: These inequalities are versions of the *Clarkson inequalities.*) (Remark: See [283] for extensions to unitarily invariant norms.)

Fact 9.9.19. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\|\langle A\rangle - \langle B\rangle\|_{\mathrm F}^2 + \|\langle A^*\rangle - \langle B^*\rangle\|_{\mathrm F}^2 \le 2\|A-B\|_{\mathrm F}^2.$$

If, in addition, A and B are Hermitian, then

$$\|\langle A\rangle - \langle B\rangle\|_{\mathrm F} \le \|A-B\|_{\mathrm F}.$$

(Proof: See [24, 331].) (Remark: This inequality generalizes a result due to Araki and Yamagami.)

Fact 9.9.20. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\|\langle A\rangle - \langle B\rangle\|_{\mathrm F}^2 + \|\langle A^*\rangle - \langle B^*\rangle\|_{\mathrm F}^2 \le 2\|A-B\|_{\mathrm F}^2.$$

If, in addition, A and B are Hermitian, then

$$\|\langle A\rangle - \langle B\rangle\|_{\mathrm F} \le \|A-B\|_{\mathrm F}.$$

(Proof: See [24, 331].)

Fact 9.9.21. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\|\langle A\rangle - \langle B\rangle\| \le \sqrt{2\|A+B\|\,\|A-B\|}.$$

(Proof: See [24].) (Remark: This result is due to Kosaki and Bhatia.)

Fact 9.9.22. Let $A, B \in \mathbb{F}^{n\times n}$, and let $p \ge 1$. Then,

$$\|\langle A\rangle - \langle B\rangle\|_{\sigma p} \le \max\left\{2^{1/p-1/2}, 1\right\}\sqrt{\|A+B|_{\sigma p}\|A-B\|_{\sigma p}}.$$

(Proof: See [24].) (Remark: This result is due to Kittaneh, Kosaki, and Bhatia.)

Fact 9.9.23. Let $A \in \mathbb{F}^{n\times n}$, let $B \in \mathbb{F}^{n\times n}$, and assume that B is Hermitian. Then,

$$\sigma_{\max}\big[A - \tfrac{1}{2}(A+A^*)\big] \le \sigma_{\max}(A-B)$$

and

$$\|A - \tfrac{1}{2}(A+A^*)\|_{\mathrm{F}} \le \|A-B\|_{\mathrm{F}}.$$

(Proof: See [466, p. 150].)

Fact 9.9.24. Let $A, M, S, B \in \mathbb{F}^{n\times n}$, and assume that $A = MS$, M is nonnegative semidefinite, and S and B are unitary. Then,

$$\|A-S\|_{\mathrm{F}} \le \|A-B\|_{\mathrm{F}}.$$

(Proof: See [466, p. 150].) (Remark: $A = MS$ is the polar decomposition of A. See Corollary 5.6.4.)

Fact 9.9.25. Let $A \in \mathbb{F}^{n\times n}$, let $\|\cdot\|$ be a normalized submultiplicative norm on $\mathbb{F}^{n\times n}$, and assume that $\|I - A\| < 1$. Then, A is nonsingular.

Fact 9.9.26. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A is nonsingular, let $\|\cdot\|$ be a normalized submultiplicative norm on $\mathbb{F}^{n\times n}$, and assume that $\|A-B\| < 1/\big\|A^{-1}\big\|$. Then, B is nonsingular.

Fact 9.9.27. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and $A+B$ are nonsingular, and let $\|\cdot\|$ be a normalized submultiplicative norm on $\mathbb{F}^{n\times n}$. Then,

$$\big\|A^{-1} - (A+B)^{-1}\big\| \le \big\|A^{-1}\big\|\,\big\|(A+B)^{-1}\big\|\,\|B\|.$$

If, in addition, $\|A^{-1}B\| < 1$, then

$$\big\|A^{-1} + (A+B)^{-1}\big\| \le \frac{\|A^{-1}\|\,\|A^{-1}B\|}{1 - \|A^{-1}B\|}.$$

Furthermore, if $\|A^{-1}B\| < 1$ and $\|B\| < 1/\|A^{-1}\|$, then

$$\left\|A^{-1} - (A+B)^{-1}\right\| \le \frac{\|A^{-1}\|^2\|B\|}{1 - \|A^{-1}\|\,\|B\|}.$$

Fact 9.9.28. Let $A \in \mathbb{F}^{n\times n}$, assume that A is nonsingular, let $E \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a normalized norm on $\mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}(A+E)^{-1} &= A^{-1}\big(I + EA^{-1}\big)^{-1}\\ &= A^{-1} - A^{-1}EA^{-1} + O\big(\|E\|^2\big).\end{aligned}$$

Fact 9.9.29. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{l\times k}$. Then,

$$\|A\otimes B\|_{\mathrm{col}} = \|A\|_{\mathrm{col}}\|B\|_{\mathrm{col}},$$

$$\|A\otimes B\|_{\infty} = \|A\|_{\infty}\|B\|_{\infty},$$

$$\|A\otimes B\|_{\mathrm{row}} = \|A\|_{\mathrm{row}}\|B\|_{\mathrm{row}}.$$

Furthermore, if $p \in [1,\infty]$, then

$$\|A\otimes B\|_p = \|A\|_p\|B\|_p.$$

Fact 9.9.30. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times m}$. Then,

$$\|A\circ B\|^2 \le \|A^*A\|\,\|B^*B\|.$$

(Proof: See [290].)

Fact 9.9.31. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular, let $b \in \mathbb{R}^n$, and let $\hat{x} \in \mathbb{R}^n$. Then,

$$\frac{1}{\kappa(A)}\frac{\|A\hat{x}-b\|}{\|b\|} \le \frac{\|\hat{x}-A^{-1}b\|}{\|A^{-1}b\|} \le \kappa(A)\frac{\|A\hat{x}-b\|}{\|b\|},$$

where $\kappa(A) \triangleq \|A\|\,\|A^{-1}\|$ and the vector and matrix norms are compatible. Equivalently, letting $\hat{b} \triangleq A\hat{x} - b$ and $x \triangleq A^{-1}b$, it follows that

$$\frac{1}{\kappa(A)}\frac{\|\hat{b}\|}{\|b\|} \le \frac{\|\hat{x}-x\|}{\|x\|} \le \kappa(A)\frac{\|\hat{b}\|}{\|b\|}.$$

(Remark: This result estimates the accuracy of an approximate solution $\hat{x}$ to $Ax = b$. $\kappa(A)$ is the *condition number* of A.)

Fact 9.9.32. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular, let $\hat{A} \in \mathbb{R}^{n\times n}$, assume that $\|A^{-1}\hat{A}\| < 1$, and let $b, \hat{b} \in \mathbb{R}^n$. Furthermore, let $x \in \mathbb{R}^n$ satisfy $Ax = b$, and let $\hat{x} \in \mathbb{R}^n$ satisfy $(A+\hat{A})\hat{x} = b+\hat{b}$. Then,

$$\frac{\|\hat{x}-x\|}{\|x\|} \le \frac{\kappa(A)}{1-\|A^{-1}\hat{A}\|}\left(\frac{\|\hat{b}\|}{\|b\|}+\frac{\|\hat{A}\|}{\|A\|}\right),$$

where $\kappa(A) \triangleq \|A\|\|A^{-1}\|$ and the vector and matrix norms are compatible. If, in addition, $\|A^{-1}\|\|\hat{A}\| < 1$, then

$$\frac{1}{\kappa(A)+1}\frac{\|\hat{b}-\hat{A}x\|}{\|b\|} \le \frac{\|\hat{x}-x\|}{\|x\|} \le \frac{\kappa(A)}{1-\|A^{-1}\hat{A}\|}\frac{\|\hat{b}-\hat{A}x\|}{\|b\|}.$$

(Proof: See [174, 175].)

Fact 9.9.33. Let $A, \hat{A} \in \mathbb{R}^{n\times n}$ satisfy $\|A^+\hat{A}\| < 1$, let $b \in \mathcal{R}(A)$, let $\hat{b} \in \mathbb{R}^n$, and assume that $b+\hat{b} \in \mathcal{R}(A+\hat{A})$. Furthermore, let $\hat{x} \in \mathbb{R}^n$ satisfy $(A+\hat{A})\hat{x} = b+\hat{b}$. Then, $x \triangleq A^+b + (I-A^+A)\hat{x}$ satisfies $Ax = b$ and

$$\frac{\|\hat{x}-x\|}{\|x\|} \le \frac{\kappa(A)}{1-\|A^{+}\hat{A}\|}\left(\frac{\|\hat{b}\|}{\|b\|}+\frac{\|\hat{A}\|}{\|A\|}\right),$$

where $\kappa(A) \triangleq \|A\|\|A^{-1}\|$ and the vector and matrix norms are compatible. (Proof: See [174].) (Remark: See [175] for a lower bound.)

Fact 9.9.34. Let $A \in \mathbb{F}^{n\times m}$ be the partitioned matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{bmatrix},$$

where $A_{ij} \in \mathbb{F}^{n_i\times n_j}$ for all $i,j = 1,\ldots,k$. Then, the following statements hold:

i) If $p \in [1,2]$, then

$$\sum_{i,j=1}^{k}\|A_{ij}\|_{\sigma p}^2 \le \|A\|_{\sigma p}^2 \le k^{4/p-2}\sum_{i,j=1}^{k}\|A_{ij}\|_{\sigma p}^2.$$

ii) If $p \in [2,\infty]$, then

$$k^{4/p-2}\sum_{i,j=1}^{k}\|A_{ij}\|_{\sigma p}^2 \le \|A\|_{\sigma p}^2 \le \sum_{i,j=1}^{k}\|A_{ij}\|_{\sigma p}^2.$$

iii) If $p \in [1,2]$, then

$$\|A\|_{\sigma p}^p \le \sum_{i,j=1}^{k}\|A_{ij}\|_{\sigma p}^p \le k^{2-p}\|A\|_{\sigma p}^p.$$

iv) If $p \in [2, \infty)$, then

$$k^{2-p}\|A\|_{\sigma p}^p \le \sum_{i,j=1}^{k} \|A_{ij}\|_{\sigma p}^p \le \|A\|_{\sigma p}^p.$$

(Proof: See [95].) (Remark: Equality holds for $p = 1$.)

9.10 Facts on Matrix Norms and Eigenvalues

Fact 9.10.1. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$|\operatorname{tr} A| \le \sum_{i=1}^{n} |\lambda_i| \le \|A\|_{\sigma 1} = \operatorname{tr}\langle A\rangle.$$

If, in addition, A is nonnegative semidefinite, then

$$\|A\|_{\sigma 1} = \operatorname{tr} A.$$

Fact 9.10.2. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$|\operatorname{tr} A^2| \le \sum_{i=1}^{n} |\lambda_i|^2 \le \|A\|_{\sigma 2}^2 = \|A\|_{\mathrm{F}}^2 = \operatorname{tr} A^*A.$$

If, in addition, A is Hermitian, then

$$\|A\|_{\sigma 2} = \sqrt{\operatorname{tr} A^2}.$$

(Proof: $\operatorname{tr}(A + A^*)^2 \ge 0$ and $\operatorname{tr}(A - A^*)^2 \le 0$.) (Remark: See Fact 8.14.3.)

Fact 9.10.3. Let $A \in \mathbb{F}^{n\times n}$, let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$, and let $p \in (0, 2]$. Then,

$$|\operatorname{tr} A^p| \le \sum_{i=1}^{n} |\lambda_i|^p \le \|A\|_{\sigma p}^p \le \|A\|_p^p.$$

(Proof: See Fact 8.14.3 and Proposition 9.2.5.)

Fact 9.10.4. Let $A, B \in \mathbb{F}^{n\times m}$, let $\operatorname{mspec}(A^*B) = \{\lambda_1, \ldots, \lambda_m\}_{\mathrm{m}}$, and let $p, q \in [1, \infty]$ satisfy $1/p + 1/q = 1$. Then,

$$|\operatorname{tr} A^*B| \le \sum_{i=1}^{n} |\lambda_i| \le \sum_{i=1}^{n} \sigma_i(A^*B) = \|AB\|_{\sigma 1} \le \|A\|_{\sigma p}\|B\|_{\sigma q}.$$

In particular,

$$|\operatorname{tr} A^*B| \le \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}.$$

(Proof: Use Proposition 9.3.6.)

Fact 9.10.5. Let $A, B \in \mathbb{F}^{n\times m}$, and let $\operatorname{mspec}(A^*B) = \{\lambda_1, \ldots, \lambda_m\}_{\mathrm{m}}$. Then,

$$|\operatorname{tr}(A^*B)^2| \le \sum_{i=1}^n |\lambda_i|^2 \le \sum_{i=1}^n \sigma_i^2(A^*B) = \operatorname{tr} AA^*BB^* = \|A^*B\|_{\mathrm{F}}^2 \le \|A\|_{\mathrm{F}}^2 \|B\|_{\mathrm{F}}^2.$$

(Proof: Use Fact 8.14.3.)

Fact 9.10.6. Let $A \in \mathbb{R}^{n\times n}$, and let $\lambda \in \operatorname{spec}(A)$. Then, the following inequalities hold:

i) $|\lambda| \le n\|A\|_\infty$.

ii) $|\operatorname{Re}\lambda| \le \frac{n}{2}\left\|A + A^{\mathrm{T}}\right\|_\infty$.

iii) $|\operatorname{Im}\lambda| \le \frac{\sqrt{n^2-n}}{2\sqrt{2}}\left\|A - A^{\mathrm{T}}\right\|_\infty$.

(Proof: See [395, p. 140].) (Remark: *i)* and *ii)* are *Hirsch's theorems*, while *iii)* is *Bendixson's theorem.* See Fact 5.9.21.)

Fact 9.10.7. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are Hermitian, and let $\operatorname{mspec}(A + \jmath B) = \{\lambda_1, \ldots, \lambda_n\}_{\mathrm{m}}$. Then,

$$\sum_{i=1}^n |\operatorname{Re}\lambda_i|^2 \le \|B\|_{\mathrm{F}}^2$$

and

$$\sum_{i=1}^n |\operatorname{Im}\lambda_i|^2 \le \|C\|_{\mathrm{F}}^2.$$

(Proof: See [466, p. 146].)

Fact 9.10.8. Let $A \in \mathbb{F}^{n\times n}$, let $\|\cdot\|$ be the norm on $\mathbb{F}^{n\times n}$ induced by the norm $\|\cdot\|'$ on $\mathbb{F}^n$, and define

$$\mu(A) \triangleq \lim_{\varepsilon\to 0^+} \frac{\|I + \varepsilon A\| - 1}{\varepsilon},$$

and let $A, B \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $\mu(A) = \mathrm{D}_+ f(A; I)$, where $f\colon\ \mathbb{F}^{n\times n} \mapsto \mathbb{R}$ is defined by $f(A) \triangleq \|A\|$.

ii) $\mu(A) = \lim_{\varepsilon\to 0^+} \varepsilon^{-1} \log\|e^{\varepsilon A}\|$.

iii) $\mu(I) = 1$, $\mu(-I) = -1$, and $\mu(0) = 0$.

iv) $-\|A\| \le -\mu(-A) \le \operatorname{Re}\lambda_i(A) \le \mu(A) \le \|A\|$ for all $i = 1, \ldots, n$.

v) $\mu(\alpha A) = |\alpha|\mu[(\operatorname{sign}\alpha)A]$ for all $\alpha \in \mathbb{R}$.

vi) $\mu(A + \alpha I) = \mu(A) + \operatorname{Re}\alpha$ for all $\alpha \in \mathbb{F}$.

vii) $\max\{\mu(A)-\mu(-B), -\mu(-A)+\mu(B)\} \le \mu(A+B) \le \mu(A)+\mu(B)$.

viii) $\mu(\alpha A+(1-\alpha)B) \le \alpha\mu(A)+(1-\alpha)\mu(B)$ for all $\alpha \in [0,1]$.

ix) $|\mu(A)-\mu(B)| \le \max\{|\mu(A-B)|, |\mu(B-A)|\} \le \|A-B\|$.

x) $\max\{-\mu(-A), -\mu(A)\}\|x\|' \le \|Ax\|'$ for all $x \in \mathbb{F}^n$.

xi) If A is nonsingular, then $\max\{-\mu(-A), -\mu(A)\} \le 1/\|A^{-1}\|$.

xii) $\operatorname{spabs}(A) \le \mu(A)$.

xiii) $\|e^A\| \le e^{\mu(A)}$.

xiv) If $\|\cdot\| = \sigma_{\max}(\cdot)$, then

$$\mu(A) = \tfrac{1}{2}\lambda_{\max}(A+A^*).$$

xv) If $\|\cdot\|' = \|\cdot\|_1$ and thus $\|\cdot\| = \|\cdot\|_{\rm col}$, then

$$\mu(A) = \max_{j\in\{1,\ldots,n\}}\left(\operatorname{Re} a_{jj} + \sum_{\substack{i=1\\ i\ne j}}^{n} |a_{ij}|\right).$$

xvi) If $\|\cdot\|' = \|\cdot\|_\infty$ and thus $\|\cdot\| = \|\cdot\|_{\rm row}$, then

$$\mu(A) = \max_{i\in\{1,\ldots,n\}}\left(\operatorname{Re} a_{ii} + \sum_{\substack{j=1\\ j\ne i}}^{n} |a_{ij}|\right).$$

(Proof: See [171, 172, 448, 532].) (Remark: $\mu(\cdot)$ is the *matrix measure* or *logarithmic derivative.* For applications, see [576]. See Fact 9.10.8 for the logarithmic derivative of an asymptotically stable matrix.)

Fact 9.10.9. Let $A, B \in \mathbb{F}^{n\times n}$, assume that A and B are Hermitian, and let $\|\cdot\|$ be a weakly unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\left\|\begin{bmatrix}\lambda_1(A) & & 0\\ & \ddots & \\ 0 & & \lambda_n(A)\end{bmatrix} - \begin{bmatrix}\lambda_1(B) & & 0\\ & \ddots & \\ 0 & & \lambda_n(B)\end{bmatrix}\right\| \le \|A-B\|$$
$$\le \left\|\begin{bmatrix}\lambda_1(A) & & 0\\ & \ddots & \\ 0 & & \lambda_n(A)\end{bmatrix} - \begin{bmatrix}\lambda_n(B) & & 0\\ & \ddots & \\ 0 & & \lambda_1(B)\end{bmatrix}\right\|.$$

In particular,

$$\max_{i\in\{1,\ldots,n\}} |\lambda_i(A)-\lambda_i(B)| \le \sigma_{\max}(A-B) \le \max_{i\in\{1,\ldots,n\}} |\lambda_i(A)-\lambda_{n-i+1}(B)|$$

and

$$\sum_{i=1}^{n}[\lambda_i(A)-\lambda_i(B)]^2 \le \|A-B\|_{\mathrm{F}}^2 \le \sum_{i=1}^{n}[\lambda_i(A)-\lambda_{n-i+1}(B)]^2.$$

(Proof: See [24], [92, p. 38], [93, p. 63, 69], [324, p. 126], [356, p. 134], [368], or [525, p. 202].) (Remark: The first inequality is the *Lidskii-Mirsky-Wielandt theorem.* The result can be stated without norms using Fact 9.8.33. See [368].)

Fact 9.10.10. Let $A,B\in\mathbb{F}^{n\times n}$, and assume that A and B are normal. Then, there exists a permutation σ of $1,\ldots,n$ such that

$$\sum_{i=1}^{n}|\lambda_{\sigma(i)}(A)-\lambda_i(B)|^2 \le \|A-B\|_{\mathrm{F}}^2.$$

(Proof: See [287, p. 368] or [466, pp. 160–161].) (Remark: This inequality is the *Hoffman-Wielandt theorem.*)

Fact 9.10.11. Let $A,B\in\mathbb{F}^{n\times n}$, and assume that A is Hermitian and B is normal. Furthermore, let $\operatorname{mspec}(B)=\{\lambda_1(B),\ldots,\lambda_n(B)\}_{\mathrm{m}}$, where $\operatorname{Re}\lambda_1(B)\ge\cdots\ge\operatorname{Re}\lambda_n(B)$. Then,

$$\sum_{i=1}^{n}|\lambda_i(A)-\lambda_i(B)|^2 \le \|A-B\|_{\mathrm{F}}^2.$$

(Proof: See [287, p. 370].) (Remark: This result is a special case of Fact 9.10.10.)

9.11 Facts on Singular Values Involving One Matrix

Fact 9.11.1. Let $A\in\mathbb{F}^{n\times n}$. Then,

$$\sigma_{\min}(A)=\min_{x\in\mathbb{F}^n\backslash\{0\}}\left(\frac{x^*A^*Ax}{x^*x}\right)^{1/2}$$

and

$$\sigma_{\max}(A)=\max_{x\in\mathbb{F}^n\backslash\{0\}}\left(\frac{x^*A^*Ax}{x^*x}\right)^{1/2}.$$

(Proof: See Lemma 8.4.3.)

Fact 9.11.2. Let $A\in\mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}\sigma_{\max}(A)&=\max\{|y^*Ax|\colon\ x\in\mathbb{F}^m,\ y\in\mathbb{F}^n,\ \|x\|_2=\|y\|_2=1\}\\&=\max\{|y^*Ax|\colon\ x\in\mathbb{F}^m,\ y\in\mathbb{F}^n,\ \|x\|_2\le 1,\ \|y\|_2\le 1\}.\end{aligned}$$

(Remark: See Fact 9.8.27.)

Fact 9.11.3. Let $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^m$, and define $\mathcal{S} \triangleq \{A \in \mathbb{F}^{n\times m}\colon \sigma_{\max}(A) \le 1\}$. Then,

$$\max_{A\in\mathcal{S}} x^*Ay = \sqrt{x^*xy^*y}.$$

Fact 9.11.4. Let $\|\cdot\|$ be an equi-induced unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then, $\|\cdot\| = \sigma_{\max}(\cdot)$.

Fact 9.11.5. Let $\|\cdot\|$ be an equi-induced self-adjoint norm on $\mathbb{F}^{n\times n}$. Then, $\|\cdot\| = \sigma_{\max}(\cdot)$.

Fact 9.11.6. Let $A \in \mathbb{F}^{n\times n}$, and let $\lambda \in \operatorname{spec}(A)$. Then,

$$\sigma_{\min}(A) \le |\lambda| \le \sigma_{\max}(A).$$

Hence,

$$[\sigma_{\min}(A)]^n \le |\det A| \le [\sigma_{\max}(A)]^n.$$

(Proof: The second inequality follows from $|\lambda|\,\|x\|_2 \le \sigma_{\max}(A)\|x\|_2$ or Proposition 9.2.6.)

Fact 9.11.7. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$|\det A| \le \sigma_{\min}(A)\sigma_{\max}^{n-1}(A).$$

(Proof: Use $|\det A| = \prod_{i=1}^n \sigma_i(A)$.)

Fact 9.11.8. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\sigma_{\min}(A) - 1 \le \sigma_{\min}(A+I) \le \sigma_{\min}(A) + 1.$$

(Proof: Use Proposition 9.6.8.)

Fact 9.11.9. Let $A \in \mathbb{F}^{n\times n}$ be normal and let $r \in \mathbb{N}$. Then,

$$\sigma_{\max}(A^r) = \sigma_{\max}^r(A).$$

(Remark: Nonnormal matrices may also satisfy these conditions. Consider $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.)

Fact 9.11.10. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\sigma_{\max}^2(A) - \sigma_{\max}\big(A^2\big) \le \sigma_{\max}(A^*A - AA^*) \le \sigma_{\max}^2(A).$$

If $A^2 = 0$, then

$$\sigma_{\max}(A^*A - AA^*) = \sigma_{\max}^2(A).$$

If A is normal, then

$$\sigma_{\max}^2(A) = \sigma_{\max}\big(A^2\big).$$

(Proof: See [336].)

Fact 9.11.11. Let $A \in \mathbb{F}^{n \times n}$. Then, the following statements are equivalent:

i) $\operatorname{sprad}(A) = \sigma_{\max}(A)$.

ii) $\sigma_{\max}(A^i) = \sigma_{\max}^i(A)$ for all $i \in \mathbb{P}$.

iii) $\sigma_{\max}(A^n) = \sigma_{\max}^n(A)$.

(Proof: See [208] and [289, p. 44].) (Remark: The result *iii)* $\Longrightarrow$ *i)* is due to Ptak.)

Fact 9.11.12. Let $A \in \mathbb{F}^{n \times n}$. Then,

$$\sigma_{\max}(A) \le \sigma_{\max}(|A|) \le \sqrt{\operatorname{rank} A}\,\sigma_{\max}(A).$$

(Proof: See [280, p. 111].)

Fact 9.11.13. Let $A \in \mathbb{R}^{n \times n}$. Then,

$$\sqrt{\tfrac{1}{2(n^2-n)}\left(\|A\|_{\mathrm{F}}^2 + \operatorname{tr} A^2\right)} \le \sigma_{\max}(A).$$

Furthermore, if $\|A\|_{\mathrm{F}} \le \operatorname{tr} A$, then

$$\sigma_{\max}(A) \le \tfrac{1}{n}\operatorname{tr} A + \sqrt{\tfrac{n-1}{n}\left[\|A\|_{\mathrm{F}}^2 - \tfrac{1}{n}(\operatorname{tr} A)^2\right]}.$$

(Proof: See [410].) (Proof: The complex case is considered in [410].)

Fact 9.11.14. Let $A \in \mathbb{F}^{n \times n}$. Then, the polynomial $p \in \mathbb{R}[s]$ defined by

$$p(s) \triangleq s^n - \|A\|_{\mathrm{F}}^2 s + (n-1)|\det A|^{2/(n-1)} = 0$$

has either exactly one or exactly two positive roots $0 < \alpha \le \beta$. Furthermore,

$$\alpha^{(n-1)/2} \le \sigma_{\min}(A) \le \sigma_{\max}(A) \le \beta^{(n-1)/2}.$$

(Proof: See [491].)

Fact 9.11.15. Let $A \in \mathbb{F}^{n \times n}$. Then, for all $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} \sigma_i\big(A^2\big) \le \sum_{i=1}^{k} \sigma_i^2(A).$$

Hence,

$$\operatorname{tr}\left(A^{2*}A^2\right)^{1/2} \le \operatorname{tr} A^*A,$$

that is,

$$\operatorname{tr}\left\langle A^2\right\rangle \le \operatorname{tr}\langle A\rangle^2.$$

(Proof: Let $A = B$ in Proposition 9.6.3.)

Fact 9.11.16. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_n\}_{\rm m}$, where $\lambda_1, \ldots, \lambda_n$ are ordered such that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. Then, for all $k = 1, \ldots, n$,

$$\prod_{i=1}^{k} |\lambda_i(A)|^2 \le \prod_{i=1}^{k} \sigma_i\big(A^2\big) \le \prod_{i=1}^{k} \sigma_i^2(A)$$

and

$$\prod_{i=1}^{n} |\lambda_i(A)|^2 = \prod_{i=1}^{n} \sigma_i\big(A^2\big) = \prod_{i=1}^{n} \sigma_i^2(A) = |\det A|^2.$$

(Proof: See [289, p. 172] and use Fact 5.9.13.) (Remark: See Fact 5.9.13 and Fact 8.14.16.)

Fact 9.11.17. Let $A \in \mathbb{F}^{n\times n}$, and let $\operatorname{mspec}(A) = \{\lambda_1, \cdots, \lambda_n\}_{\rm m}$, where $\lambda_1, \ldots, \lambda_n$ are ordered such that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. Then, for all $i = 1, \ldots, n$,

$$\lim_{k\to\infty} \sigma_i^{1/k}\Big(A^k\Big) = |\lambda_i(A)|.$$

In particular,

$$\lim_{k\to\infty} \Big[\sigma_{\max}\Big(A^k\Big)\Big]^{1/k} = \operatorname{sprad}(A).$$

(Proof: See [287, p. 180].) (Remark: This identity is due to Yamamoto.) (Remark: The expression for $\operatorname{sprad}(A)$ is a special case of Proposition 9.2.6.)

9.12 Facts on Singular Values Involving Two or More Matrices

Fact 9.12.1. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{m\times n}$, and $p \in [1, \infty)$, and assume that AB is normal. Then,

$$\|AB\|_{\sigma p} \le \|BA\|_{\sigma p}.$$

In particular,

$$\operatorname{tr}\langle AB\rangle \le \operatorname{tr}\langle BA\rangle,$$
$$\|AB\|_{\rm F} \le \|BA\|_{\rm F},$$
$$\sigma_{\max}(AB) \le \sigma_{\max}(BA).$$

(Proof: This result is due to Simon. See [107].)

Fact 9.12.2. Let $A \in \mathbb{R}^{n\times n}$ be nonsingular and let $B \in \mathbb{R}^{n\times n}$ be singular. Then,

$$\sigma_{\min}(A) \le \sigma_{\max}(A - B).$$

Furthermore, if $\sigma_{\max}\big(A^{-1}\big) = \operatorname{sprad}\big(A^{-1}\big)$, then there exists a singular matrix $C \in \mathbb{R}^{n\times n}$ such that $\sigma_{\max}(A - C) = \sigma_{\min}(A)$. (Proof: See [466, p. 151].) (Remark: This result is due to Franck.)

Fact 9.12.3. Let $A \in \mathbb{C}^{n\times n}$, assume that A is nonsingular, let $\|\cdot\|$ and $\|\cdot\|'$ be norms on $\mathbb{C}^n$, let $\|\cdot\|''$ be the norm on $\mathbb{C}^{n\times n}$ induced by $\|\cdot\|$ and $\|\cdot\|'$, and let $\|\cdot\|'''$ be the norm on $\mathbb{C}^{n\times n}$ induced by $\|\cdot\|'$ and $\|\cdot\|$. Then,

$$\min\{\|B\|''\colon\ B \in \mathbb{C}^{n\times n} \text{ and } A+B \text{ is nonsingular}\} = 1/\|A^{-1}\|'''.$$

In particular,

$$\min\{\|B\|_{\mathrm{col}}\colon\ B \in \mathbb{C}^{n\times n} \text{ and } A+B \text{ is singular}\} = 1/\|A^{-1}\|_{\mathrm{col}},$$

$$\min\{\sigma_{\max}(B)\colon\ B \in \mathbb{C}^{n\times n} \text{ and } A+B \text{ is singular}\} = \sigma_{\min}(A),$$

$$\min\{\|B\|_{\mathrm{row}}\colon\ B \in \mathbb{C}^{n\times n} \text{ and } A+B \text{ is singular}\} = 1/\|A^{-1}\|_{\mathrm{row}}.$$

(Proof: See [280, p. 111] and [278].) (Remark: This result is due to Gastinel. See [278].) (Remark: The result involving $\sigma_{\max}(B)$ is equivalent to the inequality in Fact 9.12.2.)

Fact 9.12.4. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = \operatorname{rank} B$ and $\alpha \triangleq \sigma_{\max}(A^+)\sigma_{\max}(A-B) < 1$. Then,

$$\sigma_{\max}(B^+) < \frac{1}{1-\alpha}\sigma_{\max}(A^+).$$

If, in addition, $n = m$, A and B are nonsingular, and $\sigma_{\max}(A-B) < \sigma_{\min}(A)$, then

$$\sigma_{\max}\big(B^{-1}\big) < \frac{\sigma_{\min}(A)}{\sigma_{\min}(A) - \sigma_{\max}(A-B)}\sigma_{\max}\big(A^{-1}\big).$$

(Proof: See [280, p. 400].)

Fact 9.12.5. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\sigma_{\max}(I - [A,B]) \ge 1.$$

(Proof: Since $\operatorname{tr}[A,B] = 0$ it follows that there exists $\lambda \in \operatorname{spec}(I - [A,B])$ such that $\operatorname{Re}\lambda \ge 1$, and thus $|\lambda| \ge 1$. Hence, Corollary 9.4.5 implies that $\sigma_{\max}(I - [A,B]) \ge \operatorname{sprad}(I - [A,B]) \ge |\lambda| \ge 1$.)

Fact 9.12.6. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{k\times m}$, and $D \in \mathbb{F}^{k\times l}$. Then,

$$\sigma_{\max}\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) \le \sigma_{\max}\left(\begin{bmatrix} \sigma_{\max}(A) & \sigma_{\max}(B) \\ \sigma_{\max}(C) & \sigma_{\max}(D) \end{bmatrix}\right).$$

(Proof: See [337] and references given therein.) (Remark: This is a result of Tomiyama.)

Fact 9.12.7. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{n\times l}$, and $C \in \mathbb{F}^{k\times m}$. Then, for all

$X \in \mathbb{F}^{k\times l}$,

$$\max\left\{\sigma_{\max}\left(\begin{bmatrix} A & B \end{bmatrix}\right), \sigma_{\max}\left(\begin{bmatrix} A \\ C \end{bmatrix}\right)\right\} \le \sigma_{\max}\left(\begin{bmatrix} A & B \\ C & X \end{bmatrix}\right).$$

Furthermore, there exists $X \in \mathbb{F}^{k\times l}$ such that equality holds. (Remark: This result is *Parrott's theorem.* See [158].)

Fact 9.12.8. Let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{n\times l}$. Then,

$$\begin{aligned}\max\{\sigma_{\max}(A), \sigma_{\max}(B)\} &\le \sigma_{\max}\left(\begin{bmatrix} A & B \end{bmatrix}\right)\\ &\le \left[\sigma_{\max}^2(A) + \sigma_{\max}^2(B)\right]^{1/2}\\ &\le \sqrt{2}\max\{\sigma_{\max}(A), \sigma_{\max}(B)\}\end{aligned}$$

and

$$\left[\sigma_n^2(A) + \sigma_n^2(B)\right]^{1/2} \le \sigma_n\left(\begin{bmatrix} A & B \end{bmatrix}\right) \le \begin{cases}\left[\sigma_n^2(A) + \sigma_{\max}^2(B)\right]^{1/2}\\ \left[\sigma_{\max}^2(A) + \sigma_n^2(B)\right]^{1/2}\end{cases}.$$

Fact 9.12.9. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\alpha > 0$. Then,

$$\sigma_{\max}(A+B) \le \left[\left(1+\alpha^2\right)\sigma_{\max}^2(A) + \left(1+\alpha^{-2}\right)\sigma_{\max}^2(B)\right]^{1/2}$$

and

$$\sigma_{\min}(A+B) \le \left[\left(1+\alpha^2\right)\sigma_{\min}^2(A) + \left(1+\alpha^{-2}\right)\sigma_{\max}^2(B)\right]^{1/2}.$$

Fact 9.12.10. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\sigma_{\max}\left(\begin{bmatrix} A^*A & 0 \\ 0 & BB^* \end{bmatrix}\right) \le \sigma_{\max}(A^*A - BB^*) + \sigma_{\max}(AB).$$

(Proof: See [623].)

Fact 9.12.11. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}\sigma_{\min}(A) - \sigma_{\max}(B) &\le |\det(A+B)|^{1/n}\\ &\le \prod_{i=1}^{n}|\sigma_i(A) + \sigma_{n-i+1}(B)|^{1/n}\\ &\le \sigma_{\max}(A) + \sigma_{\max}(B).\end{aligned}$$

(Proof: See [297, p. 63] and [367].)

Fact 9.12.12. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $\sigma_{\max}(B) \le \sigma_{\min}(A)$.

Then,

$$\begin{aligned} 0 &\le [\sigma_{\min}(A) - \sigma_{\max}(B)]^n \\ &\le \prod_{i=1}^n |\sigma_i(A) - \sigma_{n-i+1}(B)| \\ &\le |\det(A+B)| \\ &\le \prod_{i=1}^n |\sigma_i(A) + \sigma_{n-i+1}(B)| \\ &\le [\sigma_{\max}(A) + \sigma_{\max}(B)]^n. \end{aligned}$$

Hence, if $\sigma_{\max}(B) < \sigma_{\min}(A)$, then A is nonsingular and $A + \alpha B$ is nonsingular for all $-1 \le \alpha \le 1$. (Proof: See [367].) (Remark: See Fact 11.14.15.)

Fact 9.12.13. Let $A, B \in \mathbb{F}^{n \times m}$. Then,

$$\begin{bmatrix} \sigma_1(A+B) & \cdots & \sigma_{\min\{n,m\}}(A+B) \end{bmatrix}$$

weakly majorizes

$$\begin{bmatrix} \sigma_1(A) + \sigma_{\min\{n,m\}}(B) & \cdots & \sigma_{\min\{n,m\}}(A) + \sigma_1(B) \end{bmatrix}.$$

Furthermore, if either $\sigma_{\max}(A) < \sigma_{\min}(B)$ or $\sigma_{\max}(B) < \sigma_{\min}(A)$, then

$$\begin{bmatrix} |\sigma_1(A) - \sigma_{\min\{n,m\}}(B)| & \cdots & |\sigma_{\min\{n,m\}}(A) - \sigma_1(B)| \end{bmatrix}$$

weakly majorizes

$$\begin{bmatrix} \sigma_1(A+B) & \cdots & \sigma_{\min\{n,m\}}(A+B) \end{bmatrix}.$$

(Proof: See [367].)

Fact 9.12.14. Let $A \in \mathbb{F}^{n \times n}$, let $k \in \mathbb{P}$ satisfy $k < \operatorname{rank} A$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n \times n}$. Then,

$$\min_{B \in \{X \in \mathbb{F}^{n \times n}:\ \operatorname{rank} X = k\}} \|A - B\| = \|A - B_0\|,$$

where B_0 is formed by replacing the $n-k$ smallest singular values in the singular value decomposition of A by zeros. Furthermore,

$$\sigma_{\max}(A - B_0) = \sigma_{k+1}(A)$$

and

$$\|A - B_0\|_{\mathrm{F}} = \sqrt{\sum_{i=k+1}^r \sigma_i^2(A)}.$$

(Proof: The result follows from Fact 9.12.15. See [236] and [525, p. 208].) (Remark: This result is due to Schmidt and Mirsky.)

Fact 9.12.15. Let $A, B \in \mathbb{F}^{n\times m}$, define $A_\sigma, B_\sigma \in \mathbb{F}^{n\times m}$ by

$$A_\sigma \triangleq \begin{bmatrix} \sigma_1(A) & & & \\ & \ddots & & \\ & & \sigma_r(A) & \\ & & & 0_{(n-r)\times(m-r)} \end{bmatrix},$$

where $r \triangleq \operatorname{rank} A$, and

$$B_\sigma \triangleq \begin{bmatrix} \sigma_1(A) & & & \\ & \ddots & & \\ & & \sigma_l(A) & \\ & & & 0_{(n-l)\times(m-l)} \end{bmatrix},$$

where $l \triangleq \operatorname{rank} B$, let $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$ be unitary, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times m}$. Then,

$$\|A_\sigma - B_\sigma\| \le \|A - S_1 B S_2\| \le \|A_\sigma + B_\sigma\|.$$

In particular,

$$\max_{i\in\{1,\ldots,\max\{r,l\}\}} |\sigma_i(A) - \sigma_i(B)| \le \sigma_{\max}(A-B) \le \sigma_{\max}(A) + \sigma_{\max}(B).$$

(Proof: See [579].) (Remark: In the case $S_1 = I_n$ and $S_2 = I_m$, the left-hand inequality is *Mirsky's theorem.* See [525, p. 204].)

Fact 9.12.16. Let $A, B \in \mathbb{F}^{n\times m}$, and assume that $\operatorname{rank} A = \operatorname{rank} B$. Then,

$$\begin{aligned}\sigma_{\max}[AA^+(I - BB^+)] &= \sigma_{\max}[BB^+(I - AA^+)] \\ &\le \min\{\sigma_{\max}(A^+), \sigma_{\max}(B^+)\}\sigma_{\max}(A-B).\end{aligned}$$

(Proof: See [280, p. 400] and [525, p. 141].)

Fact 9.12.17. Let $A, B \in \mathbb{F}^{n\times m}$. Then, for all $k = 1, \ldots, \min\{n, m\}$,

$$\sum_{i=1}^k \sigma_i(A\circ B) \le \sum_{i=1}^k \sigma_i(A)\sigma_i(B).$$

In particular,

$$\sigma_{\max}(A\circ B) \le \sigma_{\max}(A)\sigma_{\max}(B).$$

(Proof: See [289, p. 334].)

Fact 9.12.18. Let $A \in \mathbb{F}^{n\times m}$, $B \in \mathbb{F}^{l\times k}$, and $p \in [1,\infty]$. Then,

$$\|A\otimes B\|_{\sigma p} = \|A\|_{\sigma p}\|B\|_{\sigma p}.$$

In particular,

$$\sigma_{\max}(A\otimes B) = \sigma_{\max}(A)\sigma_{\max}(B)$$

and

$$\|A \otimes B\|_{\mathrm{F}} = \|A\|_{\mathrm{F}} \|B\|_{\mathrm{F}}.$$

Fact 9.12.19. Let $A \in \mathbb{F}^{n \times m}$ and $B \in \mathbb{F}^{l \times m}$, and let $p, q > 1$ satisfy $1/p + 1/q = 1$. Then, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_i(AB^*) \le \sigma_i\Big(\tfrac{1}{p}\langle A\rangle^p + \tfrac{1}{q}\langle B\rangle^q\Big).$$

Equivalently, there exists a unitary matrix $S \in \mathbb{F}^{m \times m}$ such that

$$\langle AB^*\rangle^{1/2} \le S^*\Big(\tfrac{1}{p}\langle A\rangle^p + \tfrac{1}{q}\langle B\rangle^q\Big)S.$$

(Proof: See [24] or [625, p. 28].) (Remark: This result is a matrix version of Young's inequality. See Fact 1.4.5 and [282].)

Fact 9.12.20. Let $A \in \mathbb{F}^{n \times m}$ and $B \in \mathbb{F}^{l \times m}$. Then, for all $i = 1, \ldots, \min\{n, m, l\}$,

$$\sigma_i(AB^*) \le \tfrac{1}{2}\sigma_i(A^*A + B^*B).$$

(Proof: Set $p = q = 2$ in Fact 9.12.19. See [96].) (Remark: See Fact 9.9.16.)

9.13 Notes

The equivalence of absolute and monotone norms given by Proposition 9.1.2 is due to [67]. More general monotonicity conditions are considered in [313]. Induced lower bounds are treated in [353, pp. 369, 370]; see also [525, pp. 33, 80]. The induced norms (9.4.11) and (9.4.12) are given in [280, p. 116] and [140]. The $\mathrm{d}_{\max}$ norm is related to alternative norms for the convolution operator given in [603]. Proposition 9.3.6 is given in [482, p. 97]. Norm-related topics are discussed in [73]. Spectral perturbation theory in finite and infinite dimensions is treated in [324], where the emphasis is on the regularity of the spectrum as a function of the perturbation rather than on bounds for finite perturbations.

Chapter Ten

Functions of Matrices and Their Derivatives

The notion of a norm on $\mathbb{F}^n$ discussed in Chapter 9 provides the foundation for the development of some basic results in topology and analysis. This chapter provides a brief review of some basic definitions and results.

10.1 Open and Closed Sets

Let $\|\cdot\|$ be a norm on $\mathbb{F}^n$ and, for $x \in \mathbb{F}^n$ and $\varepsilon > 0$, define the *open ball of radius* ε *centered at* x by

$$\mathbb{B}_\varepsilon(x) \triangleq \{y \in \mathbb{F}^n\colon\ \|x-y\| < \varepsilon\} \tag{10.1.1}$$

and the *sphere of radius* ε *centered at* x by

$$\mathbb{S}_\varepsilon(x) \triangleq \{y \in \mathbb{F}^n\colon\ \|x-y\| = \varepsilon\}. \tag{10.1.2}$$

Definition 10.1.1. Let $\mathcal{S} \subseteq \mathbb{F}^n$. The vector $x \in \mathcal{S}$ is an *interior point* of $\mathcal{S}$ if there exists $\varepsilon > 0$ such that $\mathbb{B}_\varepsilon(x) \subseteq \mathcal{S}$. The *interior* of $\mathcal{S}$ is the set

$$\operatorname{int}\mathcal{S} \triangleq \{x \in \mathcal{S}\colon\ x \text{ is an interior point of } \mathcal{S}\}. \tag{10.1.3}$$

Finally, $\mathcal{S}$ is *open* if every element of $\mathcal{S}$ is an interior point, that is, if $\mathcal{S} = \operatorname{int}\mathcal{S}$.

Definition 10.1.2. Let $\mathcal{S} \subseteq \mathcal{S}' \subseteq \mathbb{F}^n$. The vector $x \in \mathcal{S}$ is an *interior point* of $\mathcal{S}$ *relative* to $\mathcal{S}'$ if there exists $\varepsilon > 0$ such that $\mathbb{B}_\varepsilon(x) \cap \mathcal{S}' \subseteq \mathcal{S}$ or, equivalently, $\mathbb{B}_\varepsilon(x) \cap \mathcal{S} = \mathbb{B}_\varepsilon(x) \cap \mathcal{S}'$. The *interior* of $\mathcal{S}$ *relative* to $\mathcal{S}'$ is the set

$$\operatorname{int}_{\mathcal{S}'}\mathcal{S} \triangleq \left\{x \in \mathcal{S}\colon\ x \text{ is an interior point of } \mathcal{S} \text{ relative to } \mathcal{S}'\right\}. \tag{10.1.4}$$

Finally, $\mathcal{S}$ is *open relative* to $\mathcal{S}'$ if $\mathcal{S} = \operatorname{int}_{\mathcal{S}'}\mathcal{S}$.

Definition 10.1.3. Let $\mathcal{S} \subseteq \mathbb{F}^n$. The vector $x \in \mathbb{F}^n$ is a *closure point* of $\mathcal{S}$ if, for all $\varepsilon > 0$, the set $\mathcal{S} \cap \mathbb{B}_\varepsilon(x)$ is not empty. The *closure* of $\mathcal{S}$ is the set

$$\operatorname{cl} \mathcal{S} \triangleq \{x \in \mathbb{F}^n\colon\ x \text{ is a closure point of } \mathcal{S}\}. \tag{10.1.5}$$

Finally, the set $\mathcal{S}$ is *closed* if every closure point of $\mathcal{S}$ is an element of $\mathcal{S}$, that is, if $\mathcal{S} = \operatorname{cl} \mathcal{S}$.

Definition 10.1.4. Let $\mathcal{S} \subseteq \mathcal{S}' \subseteq \mathbb{F}^n$. The vector $x \in \mathcal{S}'$ is a *closure point* of $\mathcal{S}$ relative to $\mathcal{S}'$ if, for all $\varepsilon > 0$, the set $\mathcal{S} \cap \mathbb{B}_\varepsilon(x)$ is not empty. The *closure* of $\mathcal{S}$ *relative* to $\mathcal{S}'$ is the set

$$\operatorname{cl}_{\mathcal{S}'} \mathcal{S} \triangleq \big\{x \in \mathbb{F}^n\colon\ x \text{ is a closure point of } \mathcal{S} \text{ relative to } \mathcal{S}'\big\}. \tag{10.1.6}$$

Finally, $\mathcal{S}$ is *closed relative* to $\mathcal{S}'$ if $\mathcal{S} = \operatorname{cl}_{\mathcal{S}'} \mathcal{S}$.

It follows from Theorem 9.1.8 on the equivalence of norms on $\mathbb{F}^n$ that these definitions are independent of the norm assigned to $\mathbb{F}^n$.

Let $\mathcal{S} \subseteq \mathcal{S}' \subseteq \mathbb{F}^n$. Then,

$$\operatorname{cl}_{\mathcal{S}'} \mathcal{S} = (\operatorname{cl} \mathcal{S}) \cap \mathcal{S}', \tag{10.1.7}$$

$$\operatorname{int}_{\mathcal{S}'} \mathcal{S} = \mathcal{S}' \backslash \operatorname{cl}(\mathcal{S}' \backslash \mathcal{S}), \tag{10.1.8}$$

and

$$\operatorname{int} \mathcal{S} \subseteq \operatorname{int}_{\mathcal{S}'} \mathcal{S} \subseteq \mathcal{S} \subseteq \operatorname{cl}_{\mathcal{S}'} \mathcal{S} \subseteq \operatorname{cl} \mathcal{S}. \tag{10.1.9}$$

The set $\mathcal{S}$ is *solid* if $\operatorname{int} \mathcal{S}$ is not empty, and $\mathcal{S}$ is *completely solid* if $\operatorname{cl} \operatorname{int} \mathcal{S} = \operatorname{cl} \mathcal{S}$. Note that if $\mathcal{S}$ is completely solid, then $\mathcal{S}$ is solid. The *boundary* of $\mathcal{S}$ is the set

$$\operatorname{bd} \mathcal{S} \triangleq \operatorname{cl} \mathcal{S} \backslash \operatorname{int} \mathcal{S}, \tag{10.1.10}$$

while the *boundary* of $\mathcal{S}$ *relative to* $\mathcal{S}'$ is the set

$$\operatorname{bd}_{\mathcal{S}'} \mathcal{S} \triangleq \operatorname{cl}_{\mathcal{S}'} \mathcal{S} \backslash \operatorname{int}_{\mathcal{S}'} \mathcal{S}. \tag{10.1.11}$$

Note that the empty set is both open and closed, although it is not solid.

The set $\mathcal{S} \subset \mathbb{F}^n$ is *bounded* if there exists $\delta > 0$ such that, for all $x, y \in \mathcal{S}$,

$$\|x - y\| < \delta. \tag{10.1.12}$$

The set $\mathcal{S} \subset \mathbb{F}^n$ is *compact* if it is both closed and bounded.

10.2 Limits

Definition 10.2.1. A *sequence* $\{x_1, x_2, \ldots\}_{\mathrm{m}}$ is an ordered multiset with countably infinite elements. We write $\{x_i\}_{i=1}^{\infty}$ for $\{x_1, x_2, \ldots\}_{\mathrm{m}}$.

Definition 10.2.2. The sequence $\{\alpha_i\}_{i=1}^\infty \subset \mathbb{F}$ *converges* to $\alpha \in \mathbb{F}$ if, for all $\varepsilon > 0$, there exists $p \in \mathbb{P}$ such that $|\alpha_i - \alpha| < \varepsilon$ for all $i > p$. In this case, we write $\alpha = \lim_{i\to\infty} \alpha_i$ or $\alpha_i \to \alpha$ as $i \to \infty$, where $i \in \mathbb{P}$.

Definition 10.2.3. The sequence $\{x_i\}_{i=1}^\infty \subset \mathbb{F}^n$ *converges* to $x \in \mathbb{F}^n$ if $\lim_{i\to\infty} \|x - x_i\| = 0$, where $\|\cdot\|$ is a norm on $\mathbb{F}^n$. In this case, we write $x = \lim_{i\to\infty} x_i$ or $x_i \to x$ as $i \to \infty$, where $i \in \mathbb{P}$. Similarly, $\{A_i\}_{i=1}^\infty \subset \mathbb{F}^{n\times m}$ *converges* to $A \in \mathbb{F}^{n\times m}$ if $\lim_{i\to\infty} \|A - A_i\| = 0$, where $\|\cdot\|$ is a norm on $\mathbb{F}^{n\times m}$. In this case, we write $A = \lim_{i\to\infty} A_i$ or $A_i \to A$ as $i \to \infty$, where $i \in \mathbb{P}$.

It follows from Theorem 9.1.8 that convergence of a sequence is independent of the choice of norm.

Proposition 10.2.4. Let $\mathcal{S} \subseteq \mathbb{F}^n$. The vector $x \in \mathbb{F}^n$ is a closure point of $\mathcal{S}$ if and only if there exists a sequence $\{x_i\}_{i=1}^\infty \subseteq \mathcal{S}$ such that $x = \lim_{i\to\infty} x_i$.

Proof. Suppose that $x \in \mathbb{F}^n$ is a closure point of $\mathcal{S}$. Then, for all $i \in \mathbb{P}$, there exists $x_i \in \mathcal{S}$ such that $\|x - x_i\| < 1/i$. Hence, $x - x_i \to 0$ as $i \to \infty$. Conversely, suppose that $\{x_i\}_{i=1}^\infty \subseteq \mathcal{S}$ is such that $x_i \to x$ as $i \to \infty$, and let $\varepsilon > 0$. Then, there exists $p \in \mathbb{P}$ such that $\|x - x_i\| < \varepsilon$ for all $i > p$. Therefore, $x_{p+1} \in \mathcal{S} \cap \mathbb{B}_\varepsilon(x)$, and thus $\mathcal{S} \cap \mathbb{B}_\varepsilon(x)$ is not empty. Hence, x is a closure point of $\mathcal{S}$. □

Theorem 10.2.5. Let $\mathcal{S} \subset \mathbb{F}^n$ be compact and let $\{x_i\}_{i=1}^\infty \subseteq \mathcal{S}$. Then, there exists a convergent subsequence $\{x_{i_j}\}_{j=1}^\infty \subseteq \{x_i\}_{i=1}^\infty$ such that $\lim_{j\to\infty} x_{i_j}$ exists and $\lim_{j\to\infty} x_{i_j} \in \mathcal{S}$.

Proof. See [434, p. 145]. □

Next, we define convergence for the *series* $\sum_{i=1}^\infty x_i$ in terms of the *partial sums* $\sum_{i=1}^k x_i$.

Definition 10.2.6. The series $\sum_{i=1}^\infty x_i$, where $\{x_i\}_{i=1}^\infty \subset \mathbb{F}^n$, *converges* to $x \in \mathbb{F}^n$ if

$$x = \lim_{k\to\infty} \sum_{i=1}^k x_i. \tag{10.2.1}$$

Furthermore, $\sum_{i=1}^\infty x_i$ *converges absolutely* if $\sum_{i=1}^\infty \|x_i\|$ converges, where $\|\cdot\|$ is a norm on $\mathbb{F}^n$. Similarly, the series $\sum_{i=1}^\infty A_i$, where $\{A_i\}_{i=1}^\infty \subset \mathbb{F}^{n\times m}$, *converges* to $A \in \mathbb{F}^{n\times m}$ if

$$A = \lim_{k\to\infty} \sum_{i=1}^k A_i. \tag{10.2.2}$$

Finally, $\sum_{i=1}^{\infty} A_i$ *converges absolutely* if $\sum_{i=1}^{\infty} \|A_i\|$ converges, where $\|\cdot\|$ is a norm on $\mathbb{F}^{n\times m}$.

10.3 Continuity

Definition 10.3.1. Let $\mathcal{D} \subseteq \mathbb{F}^m$, $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$, and $x \in \mathcal{D}$. Then, f is *continuous* at x if, for every convergent sequence $\{x_i\}_{i=1}^{\infty} \subseteq \mathcal{D}$ such that $\lim_{i\to\infty} x_i = x$, it follows that $\lim_{i\to\infty} f(x_i) = f(x)$. Furthermore, let $\mathcal{D}_0 \subseteq \mathcal{D}$. Then, f is *continuous* on $\mathcal{D}_0$ if f is continuous at x for all $x \in \mathcal{D}_0$. Finally, f is *continuous* if it is continuous on $\mathcal{D}$.

Theorem 10.3.2. Let $\mathcal{D} \subseteq \mathbb{F}^n$ be convex and let $f\colon\ \mathcal{D} \to \mathbb{F}$ be convex. Then, f is continuous on $\operatorname{int}_{\operatorname{aff}\mathcal{D}} \mathcal{D}$.

Proof. See [68, p. 81] and [485, p. 82]. □

Corollary 10.3.3. Let $A \in \mathbb{F}^{n\times m}$, and define $f\colon\ \mathbb{F}^m \to \mathbb{F}^n$ by $f(x) \triangleq Ax$. Then, f is continuous.

Proof. The result is a consequence of Theorem 10.3.2. Alternatively, let $x \in \mathbb{F}^m$, and let $\{x_i\}_{i=1}^{\infty} \subset \mathbb{F}^m$ be such that $x_i \to x$ as $i \to \infty$. Furthermore, let $\|\cdot\|$ and $\|\cdot\|'$ be compatible norms on $\mathbb{F}^m$ and $\mathbb{F}^{m\times n}$, respectively. Since $\|Ax - Ax_i\| \le \|A\|'\|x - x_i\|$, it follows that $Ax_i \to Ax$ as $i \to \infty$. □

Theorem 10.3.4. Let $\mathcal{D} \subseteq \mathbb{F}^m$, and let $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$. Then, the following statements are equivalent:

i) f is continuous.

ii) For all open $\mathcal{S} \subseteq \mathbb{F}^n$, the set $f^{-1}(\mathcal{S})$ is open relative to $\mathcal{D}$.

iii) For all closed $\mathcal{S} \subseteq \mathbb{F}^n$, the set $f^{-1}(\mathcal{S})$ is closed relative to $\mathcal{D}$.

Proof. See [434, pp. 87, 110]. □

Corollary 10.3.5. Let $A \in \mathbb{F}^{n\times m}$ and $\mathcal{S} \subseteq \mathbb{F}^n$, and define $\mathcal{S}' \triangleq \{x \in \mathbb{F}^m\colon\ Ax \in \mathcal{S}\}$. If $\mathcal{S}$ is open, then $\mathcal{S}'$ is open. If $\mathcal{S}$ is closed, then $\mathcal{S}'$ is closed.

The following result is the *open mapping theorem.*

Theorem 10.3.6. Let $A \in \mathbb{F}^{n\times m}$ be right invertible and let $\mathcal{D} \subseteq \mathbb{F}^m$ be open. Then, $A\mathcal{D}$ is open.

Theorem 10.3.7. Let $\mathcal{D} \subset \mathbb{F}^m$ be compact and let $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$ be continuous. Then, $f(\mathcal{D})$ is compact.

Proof. See [434, p. 146]. □

Corollary 10.3.8. Let $\mathcal{D} \subset \mathbb{F}^m$ be compact and let $f\colon\ \mathcal{D} \mapsto \mathbb{R}$ be continuous. Then, there exists $x \in \mathcal{D}$ such that $f(\mathcal{D})$ is compact.

The following result is the *Schauder fixed point theorem.*

Theorem 10.3.9. Let $\mathcal{D} \subseteq \mathbb{F}^m$ be closed and convex, let $f\colon\ \mathcal{D} \to \mathcal{D}$ be continuous, and assume that $f(\mathcal{D})$ is bounded. Then, there exists $x \in \mathcal{D}$ such that $f(x) = x$.

Proof. See [586, p. 167]. □

10.4 Derivatives

Let $\mathcal{D} \subseteq \mathbb{F}^m$, and let $x_0 \in \mathcal{D}$. Then, the *variational cone of* $\mathcal{D}$ *with respect to* x_0 is the set

$$\operatorname{vcone}(\mathcal{D}, x_0) \triangleq \{\xi \in \mathbb{F}^m\colon\ \text{there exists } \alpha_0 > 0 \text{ such that } x_0 + \alpha\xi \in \mathcal{D}, \alpha \in [0, \alpha_0)\}. \tag{10.4.1}$$

Note that $\operatorname{vcone}(\mathcal{D}, x_0)$ is a pointed cone, although it may consist of only the origin as can be seen from the example $x_0 = 0$ and

$$\mathcal{D} = \left\{x \in \mathbb{R}^2\colon\ 0 \le x_{(1)} \le 1, x_{(1)}^3 \le x_{(2)} \le x_{(1)}^2\right\}.$$

Now, let $\mathcal{D} \subseteq \mathbb{F}^m$ and $f\colon\ \mathcal{D} \to \mathbb{F}^n$. If $\xi \in \operatorname{vcone}(\mathcal{D}, x_0)$, then the *one-sided directional differential of* f *at* x_0 *in the direction* ξ is given by

$$\mathrm{D}_+f(x_0;\xi) \triangleq \lim_{\alpha\to 0^+} \tfrac{1}{\alpha}[f(x_0 + \alpha\xi) - f(x_0)] \tag{10.4.2}$$

if the limit exists. Similarly, if $\xi \in \operatorname{vcone}(\mathcal{D}, x_0)$ and $-\xi \in \operatorname{vcone}(\mathcal{D}, x_0)$, then the *two-sided directional differential* $\mathrm{D}f(x_0;\xi)$ of f *at* x_0 *in the direction* ξ is defined by replacing "$\alpha \to 0^+$" in (10.4.2) by "$\alpha \to 0$." If $\xi = e_i$ so that the direction ξ is one of the coordinate axes, then the *partial derivative* of f *with respect to* $x_{(i)}$ *at* x_0, denoted by $\frac{\partial f(x_0)}{\partial x_{(i)}}$, is given by

$$\frac{\partial f(x_0)}{\partial x_{(i)}} \triangleq \lim_{\alpha\to 0} \tfrac{1}{\alpha}[f(x_0 + \alpha e_i) - f(x_0)], \tag{10.4.3}$$

that is,

$$\frac{\partial f(x_0)}{\partial x_{(i)}} = \mathrm{D}f(x_0; e_i), \tag{10.4.4}$$

when the two-sided directional differential $\mathrm{D}f(x_0; e_i)$ exists.

Proposition 10.4.1. Let $\mathcal{D} \subseteq \mathbb{F}^m$ be a convex set, let $f\colon\ \mathcal{D} \to \mathbb{F}^n$ be convex, and let $x_0 \in \operatorname{int} \mathcal{D}$. Then, $\mathrm{D}_+ f(x_0;\xi)$ exists for all $\xi \in \mathbb{F}^m$.

Proof. See [68, p. 83]. □

Note that $\mathrm{D}_+ f(x_0;\xi) = \pm\infty$ is possible if x_0 is an element of the boundary of $\mathcal{D}$ even if f is continuous at x_0. For example, consider $f\colon\ [0,\infty) \mapsto \mathbb{R}$ given by $f(x) = 1 - \sqrt{x}$.

Next, we consider a stronger form of differentiation.

Proposition 10.4.2. Let $\mathcal{D} \subseteq \mathbb{F}^m$ be solid and convex, let $f\colon\ \mathcal{D} \to \mathbb{F}^n$, and let $x_0 \in \mathcal{D}$. Then, there exists at most one matrix $F \in \mathbb{F}^{n\times m}$ satisfying

$$\lim_{\substack{x\to x_0\\ x\in\mathcal{D}\backslash\{x_0\}}} \|x-x_0\|^{-1}[f(x) - f(x_0) - F(x-x_0)] = 0. \tag{10.4.5}$$

Proof. See [586, p. 170]. □

In (9.5) the limit is taken over all sequences that are contained in $\mathcal{D}$, do not include x_0, and converge to x_0.

Definition 10.4.3. Let $\mathcal{D} \subseteq \mathbb{F}^m$ be solid and convex, let $f\colon\ \mathcal{D} \to \mathbb{F}^n$, let $x_0 \in \mathcal{D}$, and assume there exists $F \in \mathbb{F}^{n\times m}$ satisfying (9.5). Then, f is *differentiable at* x_0 and the matrix F is the *(Frechet) derivative of* f *at* x_0. In this case, we write $f'(x_0) = F$ and

$$\lim_{\substack{x\to x_0\\ x\in\mathcal{D}\backslash\{x_0\}}} \|x-x_0\|^{-1}\big[f(x) - f(x_0) - f'(x_0)(x-x_0)\big] = 0. \tag{10.4.6}$$

Note that Proposition 10.4.2 and Definition 10.4.3 do not require that x_0 lie in the interior of $\mathcal{D}$. Sometimes we write $\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}$ for $f'(x_0)$.

Proposition 10.4.4. Let $\mathcal{D} \subseteq \mathbb{F}^m$ be solid and convex, let $f\colon\ \mathcal{D} \to \mathbb{F}^n$, let $x \in \mathcal{D}$, and assume that f is differentiable at x_0. Then, f is continuous at x_0.

Let $\mathcal{D} \subseteq \mathbb{F}^m$ be solid and convex and let $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$. In terms of its scalar components, f can be written as $f = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}^{\mathrm{T}}$, where $f_i\colon\ \mathcal{D} \mapsto \mathbb{F}$ for all $i = 1,\ldots,n$ and $f(x) = \begin{bmatrix} f_1(x) & \cdots & f_n(x) \end{bmatrix}^{\mathrm{T}}$ for all $x \in \mathcal{D}$. With this notation, $f'(x_0)$ can be written as

$$f'(x_0) = \begin{bmatrix} f_1'(x_0) \\ \vdots \\ f_n'(x_0) \end{bmatrix}, \tag{10.4.7}$$

where $f_i'(x_0) \in \mathbb{F}^{1\times m}$ is the *gradient of* f_i *at* x_0 and $f'(x_0)$ is the *Jacobian of* f *at* x_0. Furthermore, if $x \in \operatorname{int}\mathcal{D}$ then $f'(x_0)$ is related to the partial derivatives of f by

$$f'(x_0) = \begin{bmatrix} \frac{\partial f(x_0)}{\partial x_{(1)}} & \cdots & \frac{\partial f(x_0)}{\partial x_{(m)}} \end{bmatrix}, \tag{10.4.8}$$

where $\frac{\partial f(x_0)}{\partial x_{(i)}} \in \mathbb{F}^{n\times 1}$ for all $i = 1,\ldots,m$. Note that the existence of the partial derivatives of f at x_0 does not imply that f is differentiable at x_0, that is, $f'(x_0)$ given by (10.4.8) may not satisfy (10.4.6). Finally, note that the (i,j) entry of the $n\times m$ matrix $f'(x_0)$ is $\frac{\partial f_i(x_0)}{\partial x_{(j)}}$. For example, if $x \in \mathbb{F}^n$ and $A \in \mathbb{F}^{n\times n}$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}Ax = A, \tag{10.4.9}$$

Let $\mathcal{D} \subseteq \mathbb{F}^m$ and $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$. If $f'(x)$ exists for all $x \in \mathcal{D}$ and $f'\colon\ \mathcal{D} \mapsto \mathbb{F}^{n\times n}$ is continuous, then f is *continuously differentiable*, or C^1. The *second derivative* of f at $x_0 \in \mathcal{D}$, denoted by $f''(x_0)$, is the derivative of $f'\colon\ \mathcal{D} \mapsto \mathbb{F}^{n\times n}$ at $x_0 \in \mathcal{D}$. For $x_0 \in \mathcal{D}$ it can be seen that $f''(x_0)\colon\ \mathbb{F}^m \times \mathbb{F}^m \mapsto \mathbb{F}^n$ is *bilinear*, that is, for all $\hat{\eta} \in \mathbb{F}^m$, the mapping $\eta \mapsto f''(x_0)(\eta,\hat{\eta})$ is linear and, for all $\eta \in \mathbb{F}^m$, the mapping $\hat{\eta} \mapsto f''(x_0)(\eta,\hat{\eta})$ is linear. Letting $f = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}^{\mathrm{T}}$, it follows that

$$f''(x_0)(\eta,\hat{\eta}) = \begin{bmatrix} \eta^{\mathrm{T}} f_1''(x_0)\hat{\eta} \\ \vdots \\ \eta^{\mathrm{T}} f_n''(x_0)\hat{\eta} \end{bmatrix}, \tag{10.4.10}$$

where, for all $i = 1,\ldots,n$, the matrix $f_i''(x_0)$ is the $m\times m$ *Hessian* of f_i at x_0. We write $f^{(2)}(x_0)$ for $f''(x_0)$ and $f^{(k)}(x_0)$ for the *kth* derivative of f at x_0. f is C^k if $f^{(k)}(x)$ exists and is continuous on $\mathcal{D}$.

The following result is the *inverse function theorem.*

Theorem 10.4.5. Let $\mathcal{D} \subseteq \mathbb{F}^n$ be open, let $f\colon\ \mathcal{D} \mapsto \mathbb{F}^n$, and assume that f is C^k. Furthermore, let $x_0 \in \mathcal{D}$ be such that $\det f'(x_0) \neq 0$. Then, there exists an open set $\mathcal{N} \subset \mathbb{F}^n$ containing $f(x_0)$ and a C^k function $g\colon\ \mathcal{N} \mapsto \mathcal{D}$ such that $f(g(y)) = y$ for all $y \in \mathcal{N}$.

10.5 Functions of a Matrix

Consider the function $f\colon\ \mathcal{D} \subseteq \mathbb{C} \mapsto \mathbb{C}$ defined by the power series

$$f(s) = \sum_{i=0}^{\infty} \beta_i s^i, \tag{10.5.1}$$

where $\beta_i \in \mathbb{C}$ for all $i \in \mathbb{N}$, and assume that this series converges for all $|s| < \gamma$. Then, for $A \in \mathbb{C}^{n\times n}$, we define

$$f(A) \triangleq \sum_{i=1}^{\infty} \beta_i A^i, \tag{10.5.2}$$

which converges for all $A \in \mathbb{C}^{n\times n}$ such that $\mathrm{sprad}(A) < \gamma$. Now, assume that $A = SBS^{-1}$, where $S \in \mathbb{C}^{n\times n}$ is nonsingular, $B \in \mathbb{C}^{n\times n}$, and $\mathrm{sprad}(B) < \gamma$. Then,

$$f(A) = Sf(B)S^{-1}. \tag{10.5.3}$$

If, in addition, $B = \mathrm{diag}(J_1, \ldots, J_r)$ is the Jordan form of A, then

$$f(A) = S\,\mathrm{diag}[f(J_1), \ldots, f(J_r)]S^{-1}. \tag{10.5.4}$$

Letting $J = \lambda I_k + N_k$ denote a Jordan block, $f(J)$ is the upper triangular Toeplitz matrix

$$f(J) = f(\lambda)N_k + f'(\lambda)N_k + \tfrac{1}{2}f''(\lambda)N_k^2 + \cdots + \frac{1}{(k-1)!}f^{(k-1)}(\lambda)N_k^{k-1}$$

$$= \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \cdots & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ 0 & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) \\ 0 & 0 & f(\lambda) & \cdots & \frac{1}{(k-3)!}f^{(k-3)}(\lambda) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f(\lambda) \end{bmatrix}. \tag{10.5.5}$$

Next, we extend the definition $f(A)$ to functions $f\colon\ \mathcal{D} \subseteq \mathbb{C} \mapsto \mathbb{C}$ that are not necessarily of the form (10.5.1). To do this, let $A \in \mathbb{C}^{n\times n}$, where $\mathrm{spec}(A) \subset \mathcal{D}$, and assume that, for all $\lambda_i \in \mathrm{spec}(A)$, f is $k_i - 1$ times differentiable at λ_i, where $k_i \triangleq \mathrm{ind}_A(\lambda_i)$ is the order of the largest Jordan block associated with λ_i as given by Theorem 5.3.3. In this case, f is *defined* at A, and $f(A)$ is given by (10.5.3) and (10.5.4) with $f(J_i)$ defined as in (10.5.5).

Theorem 10.5.1. Let $A \in \mathbb{F}^{n\times n}$, let $\mathrm{spec}(A) = \{\lambda_1, \ldots, \lambda_r\}$, and, for $i = 1, \ldots, r$, let $k_i \triangleq \mathrm{ind}_A(\lambda_i)$. Furthermore, suppose that $f\colon\ \mathcal{D} \subseteq \mathbb{C} \mapsto \mathbb{C}$ is defined at A. Then, there exists $p \in \mathbb{F}[s]$ such that $f(A) =$

$p(A)$. Furthermore, there exists a unique polynomial p of minimal degree $\sum_{i=1}^{r} k_i$ satisfying $f(A) = p(A)$ and such that, for all $i = 1,\ldots,r$ and $j = 0,1,\ldots,k_i-1$,

$$f^{(j)}(\lambda_i) = p^{(j)}(\lambda_i). \tag{10.5.6}$$

This polynomial is given by

$$p(s) = \sum_{i=1}^{r}\left(\left[\prod_{\substack{j=1\\ j\neq i}}^{r}(s-\lambda_j)^{n_j}\right]\sum_{k=0}^{k_i-1}\frac{1}{k!}\frac{\mathrm{d}^k}{\mathrm{d}s^k}\frac{f(s)}{\prod_{\substack{l=1\\ l\neq i}}^{r}(s-\lambda_l)^{k_l}}\Bigg|_{s=\lambda_i}(s-\lambda_i)^k\right). \tag{10.5.7}$$

If, in addition, A is diagonalizable, then p is given by

$$p(s) = \sum_{i=1}^{r} f(\lambda_i)\prod_{\substack{j=1\\ j\neq i}}^{r}\frac{s-\lambda_j}{\lambda_i-\lambda_j}. \tag{10.5.8}$$

Proof. See [155, pp. 263, 264]. □

The polynomial (10.5.7) is the *Lagrange-Hermite interpolation polynomial* for f.

The following result, which is known as the *identity theorem*, is a special case of Theorem 10.5.1.

Theorem 10.5.2. Let $A \in \mathbb{F}^{n\times n}$, let $\operatorname{spec}(A) = \{\lambda_1,\ldots,\lambda_r\}$, and, for $i = 1,\ldots,r$, let $k_i \triangleq \operatorname{ind}_A(\lambda_i)$. Furthermore, let $f\colon\ \mathcal{D} \subseteq \mathbb{C} \mapsto \mathbb{C}$ and $g\colon\ \mathcal{D} \subseteq \mathbb{C} \mapsto \mathbb{C}$ be analytic on a neighborhood of $\operatorname{spec}(A)$. Then, $f(A) = g(A)$ if and only if, for all $i = 1,\ldots,r$ and $j = 0,1,\ldots,k_i-1$,

$$f^{(j)}(\lambda_i) = g^{(j)}(\lambda_i). \tag{10.5.9}$$

Corollary 10.5.3. Let $A \in \mathbb{F}^{n\times n}$, and let $f\colon\ \mathcal{D} \subset \mathbb{C} \to \mathbb{C}$ be analytic on a neighborhood of $\operatorname{mspec}(A)$. Then,

$$\operatorname{mspec}[f(A)] = f[\operatorname{mspec}(A)]. \tag{10.5.10}$$

10.6 Matrix Derivatives

In this section we consider derivatives of differentiable scalar-valued functions with matrix arguments. Consider the linear function $f\colon\ \mathbb{F}^{m\times n} \mapsto \mathbb{F}$ given by $f(X) = \operatorname{tr} AX$, where $A \in \mathbb{F}^{n\times m}$ and $X \in \mathbb{F}^{m\times n}$. In terms of vectors $x \in \mathbb{F}^{mn}$, we can define the linear function $\hat{f}(x) \triangleq (\operatorname{vec} A)^{\mathrm{T}}x$ so that

$\hat{f}(\operatorname{vec} X) = f(X) = (\operatorname{vec} A)^{\mathrm{T}}\operatorname{vec} X$. Consequently, for all $Y \in \mathbb{F}^{m\times n}$, $f'(X_0)$ can be represented by $f'(X_0)Y = \operatorname{tr} AY$.

These observations suggest that a convenient representation of the derivative $\frac{\mathrm{d}}{\mathrm{d}X}f(X)$ of a differentiable scalar-valued differentiable function $f(X)$ of a matrix argument $X \in \mathbb{F}^{m\times n}$ is the $n\times m$ matrix whose (i,j) entry is $\frac{\partial f(X)}{\partial X_{(j,i)}}$. Note the order of indices.

Proposition 10.6.1. Let $x \in \mathbb{F}^n$. Then, the following statements hold:

i) If $A \in \mathbb{F}^{n\times n}$, then
$$\frac{\mathrm{d}}{\mathrm{d}x}x^{\mathrm{T}}Ax = x^{\mathrm{T}}\big(A + A^{\mathrm{T}}\big). \tag{10.6.1}$$

ii) If $A \in \mathbb{F}^{n\times n}$ is symmetric, then
$$\frac{\mathrm{d}}{\mathrm{d}x}x^{\mathrm{T}}Ax = 2x^{\mathrm{T}}A. \tag{10.6.2}$$

iii) If $A \in \mathbb{F}^{n\times n}$ is Hermitian, then
$$\frac{\mathrm{d}}{\mathrm{d}x}x^{*}Ax = 2x^{*}A. \tag{10.6.3}$$

Proposition 10.6.2. Let $X \in \mathbb{F}^{m\times n}$. Then, the following statements hold:

i) If $A \in \mathbb{F}^{n\times m}$, then
$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AX = A. \tag{10.6.4}$$

ii) If $A \in \mathbb{F}^{l\times m}$ and $B \in \mathbb{F}^{n\times l}$, then
$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AXB = BA. \tag{10.6.5}$$

iii) If $A \in \mathbb{F}^{l\times n}$ and $B \in \mathbb{F}^{m\times l}$, then
$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AX^{\mathrm{T}}B = A^{\mathrm{T}}B^{\mathrm{T}}. \tag{10.6.6}$$

iv) If $A \in \mathbb{F}^{l\times m}$ and $B \in \mathbb{F}^{n\times l}$, then
$$\frac{\mathrm{d}}{\mathrm{d}X}\det AXB = B(AXB)^{\mathrm{A}}A. \tag{10.6.7}$$

v) If $A \in \mathbb{F}^{k\times l}$, $B \in \mathbb{F}^{l\times m}$, $C \in \mathbb{F}^{n\times l}$, $D \in \mathbb{F}^{l\times l}$, and $E \in \mathbb{F}^{l\times k}$, then
$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} A(D + BXC)^{-1}E = -C(D + BXC)^{-1}EA(D + BXC)^{-1}B. \tag{10.6.8}$$

vi) If $A \in \mathbb{F}^{k\times l}$, $B \in \mathbb{F}^{l\times n}$, $C \in \mathbb{F}^{m\times l}$, $D \in \mathbb{F}^{l\times l}$, and $E \in \mathbb{F}^{l\times k}$, then

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} A\big(D + BX^{\mathrm{T}}C\big)^{-1}E = -B^{\mathrm{T}}\big(D + BX^{\mathrm{T}}C\big)^{-\mathrm{T}}A^{\mathrm{T}}E^{\mathrm{T}}\big(D + BX^{\mathrm{T}}C\big)^{-\mathrm{T}}C^{\mathrm{T}}. \tag{10.6.9}$$

vii) If $A \in \mathbb{F}^{n\times m}$ and $A \in \mathbb{F}^{m\times n}$, then

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AXBX = AXB + BXA. \tag{10.6.10}$$

viii) If $A \in \mathbb{F}^{m\times m}$ and $B \in \mathbb{F}^{n\times n}$, then

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AXBX^{\mathrm{T}} = BX^{\mathrm{T}}A + B^{\mathrm{T}}X^{\mathrm{T}}A^{\mathrm{T}}. \tag{10.6.11}$$

Proposition 10.6.3. Let $X \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) For all $k \in \mathbb{P}$,

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} X^k = kX^{k-1}. \tag{10.6.12}$$

ii) If $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times m}$, then, for all $k \in \mathbb{P}$,

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AX^kB = \sum_{i=0}^{k-1} X^{k-1-i}BAX^i. \tag{10.6.13}$$

iii) If X is nonsingular, $A \in \mathbb{F}^{m\times n}$, and $B \in \mathbb{F}^{n\times m}$, then

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} AX^{-1}B = -X^{-1}BAX^{-1}. \tag{10.6.14}$$

iv) For all $X \in \mathbb{F}^{n\times n}$,

$$\frac{\mathrm{d}}{\mathrm{d}X}\det X = X^{\mathrm{A}}. \tag{10.6.15}$$

v) If X is nonsingular, then

$$\frac{\mathrm{d}}{\mathrm{d}X}\log\det X = X^{-1}. \tag{10.6.16}$$

10.7 Facts on Open, Closed, and Convex Sets

Fact 10.7.1. Let $x \in \mathbb{F}^n$ and $\varepsilon > 0$. Then, $\mathbb{B}_\varepsilon(x)$ is completely solid and convex.

Fact 10.7.2. Let $\mathcal{S} \subset \mathbb{F}^n$ be bounded, let $\delta > 0$ satisfy $\|x - y\| < \delta$ for all $x, y \in \mathcal{S}$, and let $x_0 \in \mathcal{S}$. Then, $\mathcal{S} \subseteq \mathbb{B}_\delta(x_0)$.

Fact 10.7.3. Let $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathbb{F}^n$. Then,

$$\operatorname{cl} \mathcal{S}_1 \subseteq \operatorname{cl} \mathcal{S}_2$$

and

$$\operatorname{int} \mathcal{S}_1 \subseteq \operatorname{int} \mathcal{S}_2.$$

Fact 10.7.4. Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then, $\operatorname{cl} \mathcal{S}$ is the smallest closed set containing $\mathcal{S}$, and $\operatorname{int} \mathcal{S}$ is the largest open set contained in $\mathcal{S}$.

Fact 10.7.5. Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then,

$$(\operatorname{int} \mathcal{S})^\sim = \operatorname{cl}(\mathcal{S}^\sim)$$

and

$$\operatorname{bd} \mathcal{S} = (\operatorname{cl} \mathcal{S}) \cap (\operatorname{cl} \mathcal{S}^\sim) = [(\operatorname{int} \mathcal{S}) \cup \operatorname{int}(\mathcal{S}^\sim)]^\sim.$$

Fact 10.7.6. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be convex. Then, $\operatorname{cl} \mathcal{S}$, $\operatorname{int} \mathcal{S}$, and $\operatorname{int}_{\operatorname{aff} \mathcal{S}} \mathcal{S}$ are also convex. (Proof: See [485, p. 45] and [486, p. 64].)

Fact 10.7.7. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be convex. Then, $\mathcal{S}$ is solid if and only if $\mathcal{S}$ is completely solid.

Fact 10.7.8. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be solid. Then, $\operatorname{co} \mathcal{S}$ is solid and completely solid.

Fact 10.7.9. Let $\mathcal{S} \subseteq \mathbb{F}^n$. Then, $\operatorname{co} \operatorname{cl} \mathcal{S} \subseteq \operatorname{cl} \operatorname{co} \mathcal{S}$. (Remark: Equality does not generally hold. Consider

$$\mathcal{S} = \left\{ x \in \mathbb{R}^2\colon\ x_{(1)}^2 x_{(2)}^2 = 1 \text{ for all } x_{(1)} > 0 \right\}.$$

Hence, if $\mathcal{S}$ is closed, then it does not necessarily follow that $\operatorname{co} \mathcal{S}$ is closed.)

Fact 10.7.10. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be either bounded or convex. Then,

$$\operatorname{co} \operatorname{cl} \mathcal{S} = \operatorname{cl} \operatorname{co} \mathcal{S}.$$

(Proof: Use Fact 10.7.6 and Fact 10.7.9.)

Fact 10.7.11. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be open. Then, $\operatorname{co} \mathcal{S}$ is also open.

Fact 10.7.12. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be compact. Then, $\operatorname{co} \mathcal{S}$ is also compact.

Fact 10.7.13. Let $\mathcal{S} \subset \mathbb{F}^n$ be symmetric, solid, convex, closed, and bounded, and, for all $x \in \mathbb{F}^n$, define

$$\|x\| \triangleq \min\{\alpha \ge 0\colon\ x \in \alpha\mathcal{S}\} = \max\{\alpha \ge 0\colon\ \alpha x \in \mathcal{S}\}.$$

Then, $\|\cdot\|$ is a norm on $\mathbb{F}^n$, and $\mathbb{B}_1(0) = \operatorname{int} \mathcal{S}$. Conversely, let $\|\cdot\|$ be a norm on $\mathbb{F}^n$. Then, $\mathbb{B}_1(0)$ is convex, bounded, symmetric, and solid. (Proof:

See [297, pp. 38, 39].) (Remark: In all cases, $\mathbb{B}_1(0)$ is defined with respect to $\|\cdot\|$. This result is due to Minkowski.)

Fact 10.7.14. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be solid. Then, $\dim \mathcal{S} = n$.

Fact 10.7.15. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be a subspace. Then, $\mathcal{S}$ is closed.

Fact 10.7.16. $\mathbf{N}^n$ is a closed and completely solid subset of $\mathbb{F}^{n(n+1)/2}$. Furthermore,

$$\operatorname{int} \mathbf{N}^n = \mathbf{P}^n.$$

Fact 10.7.17. Let $\mathcal{S} \subseteq \mathbb{F}^n$ be convex. Then,

$$\operatorname{int} \operatorname{cl} \mathcal{S} = \operatorname{int} \mathcal{S}.$$

Fact 10.7.18. Let $\mathcal{D} \subseteq \mathbb{F}^n$, and let x_0 belong to a solid, convex subset of $\mathcal{D}$. Then,

$$\dim \operatorname{vcone}(\mathcal{D}, x_0) = n.$$

Fact 10.7.19. Let $\|\cdot\|$ be a norm on $\mathbb{F}^n$, let $\mathcal{S}$ be a subspace in $\mathbb{F}^n$, let $y \in \mathbb{F}^n$, and define

$$\mu \triangleq \max_{x \in \{z \in \mathcal{S}\colon \|z\|=1\}} |y^*x|.$$

Then, there exists $z \in \mathcal{S}^\perp$ such that

$$\max_{x \in \{z \in \mathbb{F}^n\colon \|z\|=1\}} |(y+z)^*x| = \mu.$$

(Proof: See [525, p. 57].) (Remark: This result is the *Hahn-Banach theorem.*) (Problem: Find a simple interpretation in $\mathbb{R}^2$.)

Fact 10.7.20. Let $\mathcal{S} \subset \mathbb{R}^n$ be a convex cone, let $x \in \mathbb{R}^n$, and assume that $x \notin \operatorname{int} \mathcal{S}$. Then, there exists nonzero $\lambda \in \mathbb{R}^n$ such that $\lambda^{\mathrm{T}} x \le 0$ and $\lambda^{\mathrm{T}} z \ge 0$ for all $z \in \mathcal{S}$. (Remark: This result is a *separation theorem.* See [357, p. 37] and [465, p. 443].)

Fact 10.7.21. Let $\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{R}^n$ be convex. Then, the following statements are equivalent:

i) There exists a nonzero vector $\lambda \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that $\lambda^{\mathrm{T}} x \le \alpha$ for all $x \in \mathcal{S}_1$, $\lambda^{\mathrm{T}} x \ge \alpha$ for all $x \in \mathcal{S}_2$, and either $\mathcal{S}_1$ or $\mathcal{S}_2$ is not contained in the affine hyperplane $\{x \in \mathbb{R}^n\colon\ \lambda^{\mathrm{T}} x = \alpha\}$.

ii) $\operatorname{int}_{\operatorname{aff} \mathcal{S}_1} \mathcal{S}_1$ and $\operatorname{int}_{\operatorname{aff} \mathcal{S}_2} \mathcal{S}_2$ are disjoint.

(Proof: See [80, p. 82].) (Remark: This result is a *proper separation theorem.*)

10.8 Facts on Functions and Derivatives

Fact 10.8.1. Let $\{x_i\}_{i=1}^{\infty} \subset \mathbb{F}^n$. Then, $\lim_{i\to\infty} x_i = x$ if and only if $\lim_{i\to\infty} x_{i(j)} = x_{(j)}$ for all $j = 1, \ldots, n$.

Fact 10.8.2. Let $\mathcal{S}_1 \subseteq \mathbb{F}^n$ be compact, let $\mathcal{S}_2 \subset \mathbb{F}^m$, and let $f\colon\ \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}$ be continuous. Then, $g\colon\ \mathcal{S}_2 \to \mathbb{R}$ defined by $g(y) \triangleq \max_{x\in\mathcal{S}_1} f(x,y)$ is continuous.

Fact 10.8.3. Let $f\colon\ [0,\infty) \to \mathbb{R}$, and assume that $\lim_{t\to\infty} f(t)$ exists. Then,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\tau)\,\mathrm{d}\tau = \lim_{t\to\infty} f(t).$$

Fact 10.8.4. Let $f\colon\ \mathbb{R}^2 \to \mathbb{R}$, $g\colon\ \mathbb{R} \to \mathbb{R}$, and $h\colon\ \mathbb{R} \to \mathbb{R}$. Then, assuming each of the following integrals exists,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\int_{g(\alpha)}^{h(\alpha)} f(t,\alpha)\,\mathrm{d}t = f(h(\alpha),\alpha)h'(\alpha) - f(g(\alpha),\alpha)g'(\alpha) + \int_{g(\alpha)}^{h(\alpha)} \frac{\partial}{\partial\alpha} f(t,\alpha)\,\mathrm{d}t.$$

(Remark: This identity is *Leibniz' rule.*)

Fact 10.8.5. Let $\mathcal{D} \subseteq \mathbb{R}^m$ be a convex set and let $f\colon\ \mathcal{D} \to \mathbb{R}$. Then, f is convex if and only if the set $\{(x,y) \in \mathbb{R}^n \times \mathbb{R}\colon\ y \ge f(x)\}$ is convex.

Fact 10.8.6. Let $\mathcal{D} \subseteq \mathbb{R}^m$ be a convex set and let $f\colon\ \mathcal{D} \to \mathbb{R}$ be convex. Then, $f^{-1}((-\infty,\alpha]) = \{x \in \mathcal{D}\colon\ f(x) \le \alpha\}$ is convex.

Fact 10.8.7. Let $f\colon\ \mathcal{D} \subseteq \mathbb{F}^m \mapsto \mathbb{F}^n$, and assume that $\mathrm{D}_+ f(0;\xi)$ exists. Then, for all $\beta > 0$,

$$\mathrm{D}_+ f(0;\beta\xi) = \beta\mathrm{D}_+ f(0;\xi).$$

Fact 10.8.8. Define $f\colon\ \mathbb{R} \to \mathbb{R}$ by $f(x) \triangleq |x|$. Then, for all $\xi \in \mathbb{R}$,

$$\mathrm{D}_+ f(0;\xi) = |\xi|.$$

Now, define $f\colon\ \mathbb{R}^n \to \mathbb{R}^n$ by $f(x) \triangleq \sqrt{x^{\mathrm{T}}x}$. Then, for all $\xi \in \mathbb{R}^n$,

$$\mathrm{D}_+ f(0;\xi) = \sqrt{\xi^{\mathrm{T}}\xi}.$$

Fact 10.8.9. Let $A, B \in \mathbb{F}^{n\times n}$. Then, for all $s \in \mathbb{F}$,

$$\frac{\mathrm{d}}{\mathrm{d}s}(A + sB)^2 = AB + BA + 2sB.$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}s}(A+sB)^2\bigg|_{s=0} = AB + BA.$$

Fact 10.8.10. Let $A, B \in \mathbb{F}^{n\times n}$, and let $\mathcal{D} \triangleq \{s \in \mathbb{F}\colon \ \det(A+sB) \neq 0\}$. Then, for all $s \in \mathcal{D}$,

$$\frac{\mathrm{d}}{\mathrm{d}s}(A+sB)^{-1} = -(A+sB)^{-1}B(A+sB)^{-1}.$$

Hence, if A is nonsingular, then

$$\frac{\mathrm{d}}{\mathrm{d}s}(A+sB)^{-1}\bigg|_{s=0} = -A^{-1}BA^{-1}.$$

Fact 10.8.11. Let $\mathcal{D} \subseteq \mathbb{F}$, and let $A\colon \ \mathcal{D} \longrightarrow \mathbb{F}^{n\times n}$ be differentiable. Then,

$$\frac{\mathrm{d}}{\mathrm{d}s}\det A(s) = \operatorname{tr}\left[A^{\mathrm{A}}(s)\frac{\mathrm{d}}{\mathrm{d}s}A(s)\right] = \frac{1}{n-1}\operatorname{tr}\left[A(s)\frac{\mathrm{d}}{\mathrm{d}s}A^{\mathrm{A}}(s)\right] = \sum_{i=1}^{n}\det A_i(s),$$

where $A_i(s)$ is obtained by differentiating the entries of the ith row of $A(s)$. (Proof: See [155, p. 267], [466, pp. 199, 212], and [484, p. 430].)

Fact 10.8.12. Let $\mathcal{D} \subseteq \mathbb{F}$, let $A\colon \ \mathcal{D} \longrightarrow \mathbb{F}^{n\times n}$ be differentiable, and assume that $A(s)$ is nonsingular for all $x \in \mathcal{D}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}s}A^{-1}(s) = -A^{-1}(s)\left[\frac{\mathrm{d}}{\mathrm{d}s}A(s)\right]A^{-1}(s)$$

and

$$\operatorname{tr}\left[A^{-1}(s)\frac{\mathrm{d}}{\mathrm{d}s}A(s)\right] = -\operatorname{tr}\left[A(s)\frac{\mathrm{d}}{\mathrm{d}s}A^{-1}(s)\right].$$

(Proof: See [466, pp. 198, 212].)

Fact 10.8.13. Let $A, B \in \mathbb{F}^{n\times n}$. Then, for all $s \in \mathbb{F}$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\det(A+sB) = \operatorname{tr}\big[B(A+sB)^{\mathrm{A}}\big].$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}s}\det(A+sB)\bigg|_{s=0} = \operatorname{tr} BA^{\mathrm{A}} = \sum_{i=1}^{n}\det\left[A \stackrel{i}{\leftarrow} \operatorname{col}_i(B)\right].$$

(Proof: Use Fact 10.8.11 and Fact 2.13.8.) (Remark: This result generalizes Lemma 4.4.7.)

Fact 10.8.14. Let $A \in \mathbb{F}^{n\times n}$, $r \in \mathbb{R}$, and $k \in \mathbb{P}$. Then, for all $s \in \mathbb{C}$,

$$\frac{\mathrm{d}^k}{\mathrm{d}s^k}[\det(I+sA)]^r = (r\operatorname{tr}A)^k[\det(I+sA)]^r.$$

Hence,

$$\frac{\mathrm{d}^k}{\mathrm{d}s^k}[\det(I+sA)]^r\bigg|_{s=0} = (r\operatorname{tr}A)^k.$$

Fact 10.8.15. Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $X \in \mathbb{R}^{m\times n}$ be such that XAX^{T} is nonsingular. Then,

$$\left(\frac{\mathrm{d}}{\mathrm{d}X}\det XAX^{\mathrm{T}}\right) = 2\big(\det XAX^{\mathrm{T}}\big)A^{\mathrm{T}}X^{\mathrm{T}}\big(XAX^{\mathrm{T}}\big)^{-1}.$$

(Proof: See [153].)

10.9 Notes

An introductory treatment of limits and continuity is given in [434]. Frechet and directional derivatives are discussed in [209], while differentiation of matrix functions is considered in [269, 388, 403, 460, 488, 504]. In [485, 486] the set $\operatorname{int}_{\operatorname{aff}\mathcal{S}}\mathcal{S}$ is called the relative interior of $\mathcal{S}$. An extensive treatment of matrix functions is given in Chapter 6 of [289]; see also [294]. The identity theorem is discussed in [305]. The chain rule for matrix functions is considered in [388, 406]. Differentiation with respect to complex matrices is discussed in [317].

Chapter Eleven

The Matrix Exponential and Stability Theory

The matrix exponential function is fundamental to the study or linear ordinary differential equations. This chapter focuses on the properties of the matrix exponential as well as on stability theory.

11.1 Definition of the Matrix Exponential

The scalar initial value problem

$$\dot{x}(t) = ax(t), \tag{11.1.1}$$

$$x(0) = x_0, \tag{11.1.2}$$

where $t \in [0, \infty)$ and $a, x(t) \in \mathbb{R}$, has the solution

$$x(t) = e^{at}x_0, \tag{11.1.3}$$

where $t \in [0, \infty)$. We are interested in systems of linear differential equations of the form

$$\dot{x}(t) = Ax(t), \tag{11.1.4}$$

$$x(0) = x_0, \tag{11.1.5}$$

where $t \in [0, \infty)$, $x(t) \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n \times n}$. Here $\dot{x}(t)$ denotes $\frac{\mathrm{d}x(t)}{\mathrm{d}t}$, where the derivative is one sided for $t = 0$ and two sided for $t > 0$. The solution to (11.1.4), (11.1.5) is given by

$$x(t) = e^{tA}x_0, \tag{11.1.6}$$

where $t \in [0, \infty)$ and e^{tA} is the *matrix exponential.* The following definition is based on (10.5.2).

Definition 11.1.1. Let $A \in \mathbb{F}^{n \times n}$. Then, the *matrix exponential* $e^A \in$

$\mathbb{F}^{n\times n}$ or $\exp(A) \in \mathbb{F}^{n\times n}$ is the matrix

$$e^A \triangleq \sum_{k=0}^{\infty} \tfrac{1}{k!}A^k. \tag{11.1.7}$$

Note that $0! \triangleq 1$ and $e^{0_{n\times n}} = I_n$.

Proposition 11.1.2. The series (11.1.7) converges absolutely for all $A \in \mathbb{F}^{n\times n}$. Furthermore, let $\|\cdot\|$ be a normalized submultiplicative norm on $\mathbb{F}^{n\times n}$. Then,

$$\|e^A\| \le e^{\|A\|}. \tag{11.1.8}$$

Proof. Defining the partial sum $S_r \triangleq \sum_{k=0}^{r} \frac{1}{k!}A^k$, we need to show that $\lim_{r\to\infty} S_r = e^A$. We thus have, as $r \to \infty$,

$$\begin{aligned}\|e^A - S_r\| &= \left\| \sum_{k=r+1}^{\infty} \tfrac{1}{k!}A^k \right\| \le \sum_{k=r+1}^{\infty} \tfrac{1}{k!}\|A\|^k \\ &= e^{\|A\|} - \sum_{k=0}^{r} \tfrac{1}{k!}\|A\|^k \to 0.\end{aligned}$$

Furthermore, note that

$$\|e^A\| = \left\| \sum_{k=0}^{\infty} \tfrac{1}{k!}A^k \right\| \le \sum_{k=0}^{\infty} \tfrac{1}{k!}\|A^k\| \le \sum_{k=0}^{\infty} \tfrac{1}{k!}\|A\|^k = e^{\|A\|},$$

which verifies (11.1.8). □

The following result generalizes the well-known scalar result.

Proposition 11.1.3. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$e^A = \lim_{k\to\infty} \left(I + \tfrac{1}{k}A\right)^k. \tag{11.1.9}$$

Proof. It follows from the binomial theorem that

$$\left(I + \tfrac{1}{k}A\right)^k = \sum_{i=0}^{k} \alpha_i(k)A^i,$$

where

$$\alpha_i(k) \triangleq \frac{1}{k^i}\binom{k}{i} = \frac{1}{k^i}\frac{k!}{i!(k-i)!}.$$

For all $i \in \mathbb{P}$, it follows that $\alpha_i(k) \to 1/i!$ as $k \to \infty$. Hence,

$$\lim_{k\to\infty} \left(I + \tfrac{1}{k}A\right)^k = \lim_{k\to\infty} \sum_{i=0}^{k} \alpha_i(k) A^i = \sum_{i=0}^{\infty} \tfrac{1}{i!}A^i = e^A. \qquad \square$$

The following results are immediate consequences of Definition 11.1.1.

Proposition 11.1.4. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) $\left(e^A\right)^{\mathrm{T}} = e^{A^{\mathrm{T}}}$.

ii) e^A is nonsingular, and $\left(e^A\right)^{-1} = e^{-A}$.

iii) If $A = \operatorname{diag}(A_1, \ldots, A_k)$, where $A_i \in \mathbb{F}^{n_i\times n_i}$ for all $i = 1, \ldots, k$, then $e^A = \operatorname{diag}\left(e^{A_1}, \ldots, e^{A_k}\right)$.

iv) If $S \in \mathbb{F}^{n\times n}$ is nonsingular, then $e^{SAS^{-1}} = Se^AS^{-1}$.

v) If A and $B \in \mathbb{F}^{n\times n}$ are similar, then e^A and e^B are similar.

vi) If A and $B \in \mathbb{F}^{n\times n}$ are unitarily similar, then e^A and e^B are unitarily similar.

vii) If A is Hermitian, then e^A is positive definite.

viii) If A is skew Hermitian, then e^A is unitary.

The converse of *v)* is not true. For example, $A \triangleq \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix} 0 & 2\pi \\ -2\pi & 0 \end{smallmatrix}\right]$ satisfy $e^A = e^B = I$, although A and B are not similar. The converses of *vi)* and *vii)* are given by *x)* and *vi)* of Proposition 11.4.6.

Let $S\colon\ [t_0, t_1] \mapsto \mathbb{F}^{n\times m}$, and assume that every entry of $S(t)$ is differentiable. Then, define $\dot{S}(t) \triangleq \frac{\mathrm{d}S(t)}{\mathrm{d}t} \in \mathbb{F}^{n\times m}$ for all $t \in [t_0, t_1]$ entrywise, that is, for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$,

$$[\dot{S}(t)]_{(i,j)} \triangleq \frac{\mathrm{d}}{\mathrm{d}t} S_{(i,j)}(t). \tag{11.1.10}$$

If $t = t_0$ or $t = t_1$, then "$\mathrm{d}/\mathrm{d}t$" denotes a one-sided derivative. Similarly, define $\int_{t_0}^{t_1} S(t)\mathrm{d}t$ entrywise, that is, for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$,

$$\left[\int_{t_0}^{t_1} S(t)\,\mathrm{d}t\right]_{(i,j)} \triangleq \int_{t_0}^{t_1} [S(t)]_{(i,j)}\,\mathrm{d}t. \tag{11.1.11}$$

Proposition 11.1.5. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $t \in \mathbb{R}$,

$$e^{tA} - I = \int_0^t Ae^{\tau A}\,\mathrm{d}\tau \tag{11.1.12}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{tA} = Ae^{tA}. \tag{11.1.13}$$

Proof. Note that

$$\int_0^t Ae^{\tau A}\,\mathrm{d}\tau = \int_0^t \sum_{k=0}^{\infty} \tfrac{1}{k!}\tau^k A^{k+1}\,\mathrm{d}\tau = \sum_{k=0}^{\infty} \tfrac{1}{k!}\tfrac{t^{k+1}}{k+1}A^{k+1} = e^{tA} - I,$$

which yields (11.1.12), while differentiating (11.1.12) with respect to t yields (11.1.13). □

Proposition 11.1.6. Let $A, B \in \mathbb{F}^{n\times n}$. Then, $AB = BA$ if and only if, for all $t \in [0,\infty)$,

$$e^{tA}e^{tB} = e^{t(A+B)}. \tag{11.1.14}$$

Proof. Suppose $AB = BA$. By expanding e^{tA}, e^{tB}, and $e^{t(A+B)}$, it can be seen that the expansions of $e^{tA}e^{tB}$ and $e^{t(A+B)}$ are identical. Conversely, differentiating (11.1.14) twice with respect to t and setting $t = 0$ yields $AB = BA$. □

Corollary 11.1.7. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $AB = BA$. Then,

$$e^A e^B = e^B e^A = e^{A+B}. \tag{11.1.15}$$

The converse of Corollary 11.1.7 is not true. For example, if $A \triangleq \left[\begin{smallmatrix} 0 & \pi \\ -\pi & 0 \end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix} 0 & (7+4\sqrt{3})\pi \\ (-7+4\sqrt{3})\pi & 0 \end{smallmatrix}\right]$, then $e^A = e^B = -I$ and $e^{A+B} = I$, but $AB \neq BA$. A partial converse is given by Fact 11.11.2.

Proposition 11.1.8. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$e^{A\otimes I_m} = e^A \otimes I_m, \tag{11.1.16}$$

$$e^{I_n\otimes B} = I_n \otimes e^B, \tag{11.1.17}$$

$$e^{A\oplus B} = e^A \otimes e^B. \tag{11.1.18}$$

Proof. Note that

$$\begin{aligned} e^{A\otimes I_m} &= I_{nm} + A\otimes I_m + \tfrac{1}{2!}(A\otimes I_m)^2 + \cdots \\ &= I_n\otimes I_m + A\otimes I_m + \tfrac{1}{2!}(A^2\otimes I_m) + \cdots \\ &= (I_n + A + \tfrac{1}{2!}A^2 + \cdots)\otimes I_m \\ &= e^A\otimes I_m \end{aligned}$$

and similarly for (11.1.17). To prove (11.1.18) note that $(A\otimes I_m)(I_n\otimes B) = A\otimes B$ and $(I_n\otimes B)(A\otimes I_m) = A\otimes B$, which shows that $A\otimes I_m$ and $I_n\otimes B$ commute. Thus, by Corollary 11.1.7,

$$e^{A\oplus B} = e^{A\otimes I_m + I_n\otimes B} = e^{A\otimes I_m}e^{I_n\otimes B} = \big(e^A\otimes I_m\big)\big(I_n\otimes e^B\big) = e^A\otimes e^B. \quad \square$$

11.2 Structure of the Matrix Exponential

To elucidate the structure of the matrix exponential, recall that, by Theorem 4.6.1, every term A^k in (11.1.7) for $k > r \triangleq \deg \mu_A$ can be expressed as a linear combination of $I, A, \ldots, A^{r-1}$. The following result provides an expression for e^{tA} in terms of $I, A, \ldots, A^{r-1}$.

Proposition 11.2.1. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $t \in \mathbb{R}$,

$$e^{tA} = \oint_{\mathcal{C}} (zI - A)^{-1} e^{tz}\, \mathrm{d}z = \sum_{i=0}^{n-1} \psi_i(t) A^i, \tag{11.2.1}$$

where, for all $i = 0, \ldots, n-1$, $\psi_i(t)$ is given by

$$\psi_i(t) \triangleq \oint_{\mathcal{C}} \frac{\chi_A^{[i]}(z)}{\chi_A(z)} e^{tz}\, \mathrm{d}z, \tag{11.2.2}$$

where $\mathcal{C}$ is a simple, closed contour in the complex plane enclosing $\operatorname{spec}(A)$,

$$\chi_A(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0, \tag{11.2.3}$$

and, for all $i = 0, \ldots, n-1$, the polynomials $\chi_A^{[i]}$ satisfy the recursion

$$s\chi_A^{[i+1]}(s) = \chi_A^{[i]}(s) - \beta_i,$$

where $\chi_A^{[0]} \triangleq \chi_A$. Then, for all $i = 0, \ldots, n-1$ and $t \ge 0$, $\psi_i(t)$ satisfies

$$\psi_i^{(n)}(t) + \beta_{n-1}\psi_i^{(n-1)}(t) + \cdots + \beta_1\psi_i'(t) + \beta_0\psi_i(t) = 0, \tag{11.2.4}$$

where, for all $i = 0, \ldots, n-1$,

$$\psi_i^{(j)}(0) = \begin{cases} 1, & j = i-1, \\ 0, & j \ne i-1. \end{cases} \tag{11.2.5}$$

(Remark: See Fact 4.9.8.)

Proof. See [615, p. 31], [236, p. 381], [362, 379], and Fact 4.9.8. □

To further understand the structure of e^{tA}, where $A \in \mathbb{F}^{n\times n}$, let $A = SBS^{-1}$, where $B = \mathrm{diag}(B_1, \ldots, B_k)$ is the Jordan form of A. Hence, by Proposition 11.1.4,

$$e^{tA} = Se^{tB}S^{-1}, \tag{11.2.6}$$

where

$$e^{tB} = \mathrm{diag}\big(e^{tB_1}, \ldots, e^{tB_k}\big). \tag{11.2.7}$$

The structure of e^{tB} can thus be determined by considering the block $B_i \in \mathbb{F}^{\alpha_i\times\alpha_i}$, which, for all $i = 1, \ldots, k$, has the form

$$B_i = \lambda_i I_{\alpha_i} + N_{\alpha_i}. \tag{11.2.8}$$

Since $\lambda_i I_{\alpha_i}$ and N_{α_i} commute, it follows from Proposition 11.1.6 that

$$e^{tB_i} = e^{t(\lambda_i I_{\alpha_i} + N_{\alpha_i})} = e^{\lambda_i t I_{\alpha_i}} e^{tN_{\alpha_i}} = e^{\lambda_i t} e^{tN_{\alpha_i}}. \tag{11.2.9}$$

Since $N_{\alpha_i}^{\alpha_i} = 0$, it follows that $e^{tN_{\alpha_i}}$ is a finite sum of powers of tN_{α_i}. Specifically,

$$e^{tN_{\alpha_i}} = I_{\alpha_i} + tN_{\alpha_i} + \tfrac{1}{2}t^2N_{\alpha_i}^2 + \cdots + \tfrac{1}{(\alpha_i-1)!}t^{\alpha_i-1}N_{\alpha_i}^{\alpha_i-1} \tag{11.2.10}$$

$$= \begin{bmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{\alpha_i-1}}{(\alpha_i-1)!} \\ 0 & 1 & t & \ddots & \frac{t^{\alpha_i-2}}{(\alpha_i-2)!} \\ 0 & 0 & 1 & \ddots & \frac{t^{\alpha_i-3}}{(\alpha_i-3)!} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{11.2.11}$$

which is a Toeplitz matrix. Note that (11.2.9) follows from (10.5.5) with $f(\lambda) = e^{\lambda t}$. Furthermore, every entry of e^{tB_i} is of the form $\frac{1}{r!}t^r e^{\lambda_i t}$, where $r \in \{0, \alpha_i - 1\}$ and λ_i is an eigenvalue of A. Reconstructing A by means of $A = SBS^{-1}$ shows that every entry of A is a linear combination of the entries of the blocks e^{tB_i}. If A is real, then e^{tA} is also real. Thus, the term $e^{\lambda_i t}$ for complex $\lambda_i = \nu_i + \jmath\omega_i \in \mathrm{spec}(A)$, where ν_i and ω_i are real, yields terms of the form $e^{\nu_i t}\cos\omega_i t$ and $e^{\nu_i t}\sin\omega_i t$.

The following result follows from (11.2.11) or Corollary 10.5.3.

Proposition 11.2.2. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\mathrm{mspec}\big(e^A\big) = \big\{e^\lambda\colon\ \lambda \in \mathrm{mspec}(A)\big\}_{\mathrm{m}}. \tag{11.2.12}$$

Proof. It can be seen that every diagonal entry of the Jordan form of e^A is of the form e^λ, where $\lambda \in \operatorname{spec}(A)$. □

Corollary 11.2.3. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\det e^A = e^{\operatorname{tr} A}. \tag{11.2.13}$$

Corollary 11.2.4. Let $A \in \mathbb{F}^{n\times n}$, and assume that $\operatorname{tr} A = 0$. Then, $\det e^A = 1$.

11.3 Explicit Expressions

In this section we present explicit expressions for the exponential of a general 2×2 real matrix A. Expressions are given in terms of both the entries of A and the eigenvalues of A.

Lemma 11.3.1. Let $A \triangleq \left[\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right] \in \mathbb{C}^{2\times 2}$. Then,

$$e^A = \begin{cases} e^a \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}, & a = d, \\ \begin{bmatrix} e^a & b\frac{e^a - e^d}{a-d} \\ 0 & e^d \end{bmatrix}, & a \neq d. \end{cases} \tag{11.3.1}$$

The following result gives an expression for e^A in terms of the eigenvalues of A.

Proposition 11.3.2. Let $A \in \mathbb{C}^{2\times 2}$, and let $\operatorname{mspec}(A) = \{\lambda, \mu\}_{\mathrm{m}}$. Then,

$$e^A = \begin{cases} e^\lambda[(1-\lambda)I + A], & \lambda = \mu, \\ \frac{\mu e^\lambda - \lambda e^\mu}{\mu - \lambda} I + \frac{e^\mu - e^\lambda}{\mu - \lambda} A, & \lambda \neq \mu. \end{cases} \tag{11.3.2}$$

Proof. The result follows from Theorem 10.5.1. Alternatively, suppose that $\lambda = \mu$. Then, there exists a nonsingular matrix $S \in \mathbb{C}^{2\times 2}$ such that $A = S\left[\begin{smallmatrix} \lambda & \alpha \\ 0 & \lambda \end{smallmatrix}\right]S^{-1}$, where $\alpha \in \mathbb{C}$. Hence, $e^A = e^\lambda S\left[\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right]S^{-1} = e^\lambda[(1-\lambda)I + A]$. Now, suppose that $\lambda \neq \mu$. Then, there exists a nonsingular matrix $S \in \mathbb{C}^{2\times 2}$ such that $A = S\left[\begin{smallmatrix} \lambda & 0 \\ 0 & \mu \end{smallmatrix}\right]S^{-1}$. Hence, $e^A = S\left[\begin{smallmatrix} e^\lambda & 0 \\ 0 & e^\mu \end{smallmatrix}\right]S^{-1}$. Then, the identity $\left[\begin{smallmatrix} e^\lambda & 0 \\ 0 & e^\mu \end{smallmatrix}\right] = \frac{\mu e^\lambda - \lambda e^\mu}{\mu - \lambda} I + \frac{e^\mu - e^\lambda}{\mu - \lambda}\left[\begin{smallmatrix} \lambda & 0 \\ 0 & \mu \end{smallmatrix}\right]$ yields the given result. □

Next, we give an expression for e^A in terms of the entries of $A \in \mathbb{R}^{2\times 2}$.

Corollary 11.3.3. Let $A \triangleq \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathbb{R}^{2\times 2}$, and define $\gamma \triangleq (a-d)^2 + 4bc$ and $\delta \triangleq \frac{1}{2}|\gamma|^{1/2}$. Then,

$$e^A = \begin{cases} e^{\frac{a+d}{2}} \begin{bmatrix} \cos\delta + \frac{a-d}{2\delta}\sin\delta & \frac{b}{\delta}\sin\delta \\ \frac{c}{\delta}\sin\delta & \cos\delta - \frac{a-d}{2\delta}\sin\delta \end{bmatrix}, & \gamma < 0, \\ e^{\frac{a+d}{2}} \begin{bmatrix} 1 + \frac{a-d}{2} & b \\ c & 1 - \frac{a-d}{2} \end{bmatrix}, & \gamma = 0, \\ e^{\frac{a+d}{2}} \begin{bmatrix} \cosh\delta + \frac{a-d}{2\delta}\sinh\delta & \frac{b}{\delta}\sinh\delta \\ \frac{c}{\delta}\sinh\delta & \cosh\delta - \frac{a-d}{2\delta}\sinh\delta \end{bmatrix}, & \gamma > 0. \end{cases} \quad (11.3.3)$$

Proof. The eigenvalues of A are $\lambda \triangleq \frac{1}{2}(a + d - \sqrt{\gamma})$ and $\mu \triangleq \frac{1}{2}(a + d + \sqrt{\gamma})$. Hence, $\lambda = \mu$ if and only if $\gamma = 0$. The result now follows from Proposition 11.3.2. □

Example 11.3.4. Let $A \triangleq \left[\begin{smallmatrix} \nu & \omega \\ -\omega & \nu \end{smallmatrix}\right] \in \mathbb{R}^{2\times 2}$. Then,

$$e^{tA} = e^{\nu t}\begin{bmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{bmatrix}. \quad (11.3.4)$$

On the other hand, if $A \triangleq \left[\begin{smallmatrix} \nu & \omega \\ \omega & -\nu \end{smallmatrix}\right]$, then

$$e^{tA} = \begin{bmatrix} \cosh\delta t + \frac{\nu}{\delta}\sinh\delta t & \frac{\omega}{\delta}\sinh\delta t \\ \frac{\omega}{\delta}\sinh\delta t & \cosh\delta t - \frac{\nu}{\delta}\sinh\delta t \end{bmatrix}, \quad (11.3.5)$$

where $\delta \triangleq \sqrt{\omega^2 + \nu^2}$.

Example 11.3.5. Let $\alpha \in \mathbb{F}$, and define $A \triangleq \left[\begin{smallmatrix} 0 & 1 \\ 0 & \alpha \end{smallmatrix}\right]$. Then,

$$e^{tA} = \begin{cases} \begin{bmatrix} 1 & \alpha^{-1}\left(e^{\alpha t} - 1\right) \\ 0 & e^{\alpha t} \end{bmatrix}, & \alpha \neq 0, \\ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, & \alpha = 0. \end{cases}$$

Example 11.3.6. Let $A \triangleq \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} \in \mathbb{R}^{2\times 2}$. Then,

$$e^{tA} = \begin{cases} \begin{bmatrix} e^{\alpha t} & \beta\frac{(e^{\alpha t}-e^{\gamma t})}{\alpha-\gamma} \\ 0 & e^{\gamma t} \end{bmatrix}, & \alpha \neq \gamma, \\ \begin{bmatrix} e^{\alpha t} & \beta t e^{\alpha t} \\ 0 & e^{\gamma t} \end{bmatrix}, & \alpha = \gamma. \end{cases}$$

In particular,

$$e^{t\left[\begin{smallmatrix} 1 & -1 \\ 0 & 2 \end{smallmatrix}\right]} = \begin{bmatrix} e^t & e^t - e^{2t} \\ 0 & e^{2t} \end{bmatrix}.$$

Example 11.3.7. Let $\theta \in \mathbb{R}$, and define $A \triangleq \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix}$. Then,

$$e^A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Furthermore, define $B \triangleq \begin{bmatrix} 0 & \frac{\pi}{2}-\theta \\ \frac{-\pi}{2}+\theta & 0 \end{bmatrix}$. Then,

$$e^B = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix}.$$

Example 11.3.8. Consider the second-order mechanical vibration equation

$$m\ddot{q} + c\dot{q} + kq = 0, \tag{11.3.6}$$

where m is positive and c and k are nonnegative. Here m, c, and k denote mass, damping, and stiffness parameters, respectively. Equation (11.3.6) can be written in companion form as the system

$$\dot{x} = Ax, \tag{11.3.7}$$

where

$$x \triangleq \begin{bmatrix} q \\ \dot{q} \end{bmatrix}, \qquad A \triangleq \begin{bmatrix} 0 & 1 \\ -k/m & -c/m \end{bmatrix}. \tag{11.3.8}$$

The inelastic case $k = 0$ is the simplest one since A is upper triangular. In this case,

$$e^{tA} = \begin{cases} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, & k = c = 0, \\ \begin{bmatrix} 1 & \frac{m}{c}(1-e^{-ct/m}) \\ 0 & e^{-ct/m} \end{bmatrix}, & k = 0,\ c > 0, \end{cases} \tag{11.3.9}$$

where $c = 0$ and $c > 0$ correspond to a rigid body and a damped rigid body, respectively.

Next, we consider the elastic case $c \geq 0$ and $k > 0$. In this case, we define

$$\omega_{\rm n} \triangleq \sqrt{\frac{k}{m}}, \qquad \zeta \triangleq \frac{c}{2\sqrt{mk}}, \tag{11.3.10}$$

where $\omega_{\rm n} > 0$ denotes the (undamped) *natural frequency* of vibration and $\zeta \geq 0$ denotes the *damping ratio*. Now, A can be written as

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_{\rm n}^2 & -2\zeta\omega_{\rm n} \end{bmatrix}, \tag{11.3.11}$$

and Corollary 11.3.3 yields

$$e^{tA} \tag{11.3.12}$$

$$= \begin{cases} \begin{bmatrix} \cos\omega_{\rm n}t & \frac{1}{\omega_{\rm n}}\sin\omega_{\rm n}t \\ -\omega_{\rm n}\sin\omega_{\rm n}t & \cos\omega_{\rm n}t \end{bmatrix}, & \zeta = 0, \\ e^{-\zeta\omega_{\rm n}{\rm t}}\begin{bmatrix} \cos\omega_{\rm d}t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_{\rm d}t & \frac{1}{\omega_{\rm d}}\sin\omega_{\rm d}t \\ \frac{-\omega_{\rm d}}{1-\zeta^2}\sin\omega_{\rm d}t & \cos\omega_{\rm d}t - \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_{\rm d}t \end{bmatrix}, & 0 < \zeta < 1, \\ e^{-\omega_{\rm n}t}\begin{bmatrix} 1+\omega_{\rm n}t & t \\ -\omega_{\rm n}^2 t & 1-\omega_{\rm n}t \end{bmatrix}, & \zeta = 1, \\ e^{-\zeta\omega_{\rm n}t}\begin{bmatrix} \cosh\omega_{\rm d}t + \frac{\zeta}{\sqrt{\zeta^2-1}}\sinh\omega_{\rm d}t & \frac{1}{\omega_{\rm d}}\sinh\omega_{\rm d}t \\ \frac{-\omega_{\rm d}}{\zeta^2-1}\sinh\omega_{\rm d}t & \cosh\omega_{\rm d}t - \frac{\zeta}{\sqrt{\zeta^2-1}}\sinh\omega_{\rm d}t \end{bmatrix}, & \zeta > 1, \end{cases}$$

where $\zeta = 0$, $0 < \zeta < 1$, $\zeta = 1$, and $\zeta > 1$ correspond to *undamped, underdamped, critically damped*, and *overdamped oscillators*, respectively, and where the *damped natural frequency* $\omega_{\rm d}$ is the positive number

$$\omega_{\rm d} \triangleq \begin{cases} \omega_{\rm n}\sqrt{1-\zeta^2}, & 0 < \zeta < 1, \\ \omega_{\rm n}\sqrt{\zeta^2-1}, & \zeta > 1. \end{cases} \tag{11.3.13}$$

11.4 Logarithms

Let $A \in \mathbb{F}^{n\times n}$ be positive definite so that $A = SBS^* \in \mathbb{F}^{n\times n}$, where $S \in \mathbb{F}^{n\times n}$ is unitary and $B \in \mathbb{R}^{n\times n}$ is diagonal with positive diagonal entries. In Section 8.5, $\log A$ is defined as $\log A = S(\log B)S^* \in \mathbf{H}^n$, where $(\log B)_{(i,i)} \triangleq \log B_{(i,i)}$. It can be seen that $\log A$ satisfies $A = e^{\log A}$. The following definition is not restricted to positive-definite matrices A.

Definition 11.4.1. Let $A \in \mathbb{F}^{n\times n}$. Then, $B \in \mathbb{F}^{n\times n}$ is a *logarithm* of A if $e^B = A$.

Proposition 11.4.2. Let $\|\cdot\|$ be a normalized submultiplicative norm on $\mathbb{F}^{n\times n}$, and, for $A \in \mathbb{F}^{n\times n}$, define

$$\log A \triangleq \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i}(A-I)^i. \tag{11.4.1}$$

Then, the following statements hold:

i) The series (11.4.1) converges absolutely for all $A \in \mathbb{F}^{n\times n}$ such that $\|A-I\| < 1$.

ii) If $A \in \mathbb{F}^{n\times n}$ and $\|A-I\| < 1$, then $\|\log A\| \le \log(1+\|A-I\|)$.

iii) If $A \in \mathbb{F}^{n\times n}$ and $\|A-I\| < 1$, then $\log A$ is a logarithm of A, that is, $e^{\log A} = A$.

iv) If $B \in \mathbb{F}^{n\times n}$ and $\|e^B - I\| < 1$, then $\log e^B = B$.

v) $\exp\colon\ \mathbb{B}_{\log 2}(0) \mapsto \mathbb{F}^{n\times n}$ is one-to-one.

Proof. For $\alpha \triangleq \|A-I\| < 1$ it follows from (11.4.1) that $\|\log A\| \le \sum_{i=1}^{\infty}(-1)^{i-1}\alpha^i/i = \log(1+\alpha)$, which proves *i)* and *ii)*. Statements *iii)* and *iv)* can be confirmed by using the series representation of the matrix exponential. To prove *v)*, let $B \in \mathbb{B}_{\log 2}(0)$, so that $e^{\|B\|} < 2$, and thus $\|e^B - I\| \le \sum_{i=1}^{\infty}\|B\|^i = e^{\|B\|}-1 < 1$. Now, let $B_1, B_2 \in \mathbb{B}_{\log 2}(0)$, and assume that $e^{B_1} = e^{B_2}$. Then, it follows from *ii)* that $B_1 = \log e^{B_1} = \log e^{B_2} = B_2$. □

The following result shows that every complex, nonsingular matrix has a complex logarithm.

Proposition 11.4.3. Let $A \in \mathbb{C}^{n\times n}$. Then, there exists a matrix $B \in \mathbb{C}^{n\times n}$ such that $A = e^B$ if and only if A is nonsingular.

Proof. See [289, p. 474]. □

However, only certain real matrices have a real logarithm.

Proposition 11.4.4. Let $A \in \mathbb{R}^{n\times n}$. Then, there exists a matrix $B \in \mathbb{R}^{n\times n}$ such that $A = e^B$ if and only if A is nonsingular and, for every negative eigenvalue λ of A and for every positive integer k, the Jordan form of A has an even number of $k\times k$ blocks associated with λ.

Proof. See [289, p. 475]. □

Replacing A and B in Proposition 11.4.4 by e^A and A, respectively, yields the following result.

Corollary 11.4.5. Let $A \in \mathbb{R}^{n\times n}$. Then, for every negative eigenvalue λ of e^A and for every positive integer k, the Jordan form of e^A has an even number of $k \times k$ blocks associated with λ.

Since the matrix $A \triangleq \begin{bmatrix} -2\pi & 4\pi \\ -2\pi & 2\pi \end{bmatrix}$ satisfies $e^A = I$ it follows that a positive-definite matrix can have a logarithm that is not normal. However, the following result shows that every positive-definite matrix has at least one Hermitian logarithm. Analogous results are given for several sets of matrices.

Proposition 11.4.6. Let $n \geq 1$. Then, the following functions are onto:

i) exp: $\mathrm{gl}_{\mathbb{C}}(n) \mapsto \mathrm{GL}_{\mathbb{C}}(n)$.

ii) exp: $\mathrm{gl}_{\mathbb{R}}(1) \mapsto \mathrm{PL}_{\mathbb{R}}(1)$.

iii) exp: $\mathrm{pl}_{\mathbb{C}}(n) \mapsto \mathrm{PL}_{\mathbb{C}}(n)$.

iv) exp: $\mathrm{sl}_{\mathbb{C}}(n) \mapsto \mathrm{SL}_{\mathbb{C}}(n)$.

v) exp: $\mathbf{H}^n \mapsto \mathbf{P}^n$.

vi) exp: $\mathrm{u}(n) \mapsto \mathrm{U}(n)$.

vii) exp: $\mathrm{su}(n) \mapsto \mathrm{SU}(n)$.

viii) exp: $\mathrm{so}(n) \mapsto \mathrm{SO}(n)$.

Furthermore, the following functions are not onto:

ix) exp: $\mathrm{gl}_{\mathbb{R}}(n) \mapsto \mathrm{PL}_{\mathbb{R}}(n)$, where $n \geq 2$.

x) exp: $\mathrm{sl}_{\mathbb{R}}(n) \mapsto \mathrm{SL}_{\mathbb{R}}(n)$.

xi) exp: $\mathrm{so}(n) \mapsto \mathrm{O}(n)$.

xii) exp: $\mathrm{sp}(n) \mapsto \mathrm{Sp}(n)$.

Proof. Statement *i)* follows from Proposition 11.4.3, while *ii)* is immediate. Statements *iii)*-*viii)* can be verified by construction; see [466, pp. 199, 212] for the proof of *vi)* and *viii)*. The example $A \triangleq \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$ and Proposition 11.4.4 show that *ix)* is not onto. For $\lambda < 0$, $\lambda \neq -1$, Proposition 11.4.4 and the example $\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}$ given in [496, p. 39] show that *x)* is not onto. See also [45, pp. 84, 85]. Statement *viii)* shows that *xi)* is not onto. For *xii)*, see [173]. $\square$

Let $A \in \mathbb{R}^{n\times n}$. If there exists $B \in \mathbb{R}^{n\times n}$ such that $A = e^B$, then

Corollary 11.2.3 implies that $\det A = \det e^B = e^{\operatorname{tr} B} > 0$. However, the converse is not true. Consider, for example, $A \triangleq \left[\begin{smallmatrix} -1 & 0 \\ 0 & -2 \end{smallmatrix}\right]$, which satisfies $\det A > 0$. However, Proposition 11.4.4 implies that there does not exist $B \in \mathbb{R}^{2\times 2}$ such that $A = e^B$. On the other hand, note that $A = e^B e^C$, where $B \triangleq \left[\begin{smallmatrix} 0 & \pi \\ -\pi & 0 \end{smallmatrix}\right]$ and $C \triangleq \left[\begin{smallmatrix} 0 & 0 \\ 0 & \log 2 \end{smallmatrix}\right]$. While the product of two exponentials of real matrices has positive determinant, the following result shows that the converse is also true.

Proposition 11.4.7. Let $A \in \mathbb{R}^{n\times n}$. Then, there exist $B, C \in \mathbb{R}^{n\times n}$ such that $A = e^B e^C$ if and only if $\det A > 0$.

Proof. Suppose that there exist $B, C \in \mathbb{R}^{n\times n}$ such that $A = e^B e^C$. Then, $\det A = \left(\det e^B\right)\left(\det e^C\right) > 0$. Conversely, suppose that $\det A > 0$. If A has no negative eigenvalues, then it follows from Proposition 11.4.4 that there exists $B \in \mathbb{R}^{n\times n}$ such that $A = e^B$. Hence, $A = e^B e^{0_{n\times n}}$. Now, suppose that A has at least one negative eigenvalue. Then, Theorem 5.3.5 on the real Jordan form implies that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ and matrices $A_1 \in \mathbb{R}^{n_1\times n_1}$ and $A_2 \in \mathbb{R}^{n_2\times n_2}$ such that $A = S\left[\begin{smallmatrix} A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right]S^{-1}$, where all of the eigenvalues of A_1 are negative and where none of the eigenvalues of A_2 are negative. Since $\det A$ and $\det A_2$ are positive, it follows that n_1 is even. Now, write $A = S\left[\begin{smallmatrix} -I_{n_1} & 0 \\ 0 & I_{n_2} \end{smallmatrix}\right]\left[\begin{smallmatrix} -A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right]S^{-1}$. Since the eigenvalue -1 of $\left[\begin{smallmatrix} -I_{n_1} & 0 \\ 0 & I_{n_2} \end{smallmatrix}\right]$ appears in an even number of 1×1 Jordan blocks, it follows from Proposition 11.4.4 that there exists $\hat{B} \in \mathbb{R}^{n\times n}$ such that $\left[\begin{smallmatrix} -I_{n_1} & 0 \\ 0 & I_{n_2} \end{smallmatrix}\right] = e^{\hat{B}}$. Furthermore, since $\left[\begin{smallmatrix} -A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right]$ has no negative eigenvalues, it follows that there exists $\hat{C} \in \mathbb{R}^{n\times n}$ such that $\left[\begin{smallmatrix} -A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right] = e^{\hat{C}}$. Hence, $e^A = Se^{\hat{B}}e^{\hat{C}}S^{-1} = e^{S\hat{B}S^{-1}}e^{S\hat{C}S^{-1}}$. □

Although $e^A e^B$ is generally different from e^{A+B}, the following result, known as the *Baker-Campbell-Hausdorff series*, provides an expansion for a matrix function $C(t)$ that satisfies $e^{C(t)} = e^{tA}e^{tB}$.

Proposition 11.4.8. Let $A_1, \ldots, A_l \in \mathbb{F}^{n\times n}$. Then, there exists $\varepsilon > 0$ such that, for all $t \in (-\varepsilon, \varepsilon)$,

$$e^{tA_1} \cdots e^{tA_l} = e^{C(t)}, \tag{11.4.2}$$

where

$$C(t) \triangleq \sum_{i=1}^{l} tA_i + \sum_{1\le i<j\le l} \tfrac{1}{2}t^2[A_i, A_j] + O(t^3). \tag{11.4.3}$$

Proof. See [571, p. 97] or [496, p. 35]. □

To illustrate (11.4.2), let $l = 2$, $A = A_1$, and $B = A_2$. Then, the first two terms of the series are given by

$$e^{tA}e^{tB} = e^{tA+tB+\frac{t^2}{2}[A,B]+\frac{t^3}{12}[[B,A],A+B]+\cdots}. \tag{11.4.4}$$

The radius of convergence of this series is discussed in [438].

Corollary 11.4.9. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$e^{A+B} = \lim_{k\to\infty} \left(e^{\frac{1}{k}A}e^{\frac{1}{k}B}\right)^k. \tag{11.4.5}$$

Proof. Setting $l = 2$ and $k = 1/t$ in (11.4.2) yields, as $k \to \infty$,

$$\left(e^{\frac{1}{k}A}e^{\frac{1}{k}B}\right)^k = \left[e^{\frac{1}{k}(A+B)+O\left(\frac{1}{k^2}\right)}\right]^k = e^{A+B+O(1/k)} \to e^{A+B}. \qquad \square$$

11.5 Lyapunov Stability Theory

Consider the dynamical system

$$\dot{x}(t) = f(x(t)), \tag{11.5.1}$$

where $t \geq 0$, $x(t) \in \mathcal{D} \subseteq \mathbb{R}^n$, and $f\colon\ \mathcal{D} \to \mathbb{R}^n$ is continuous. We assume that, for all $x_0 \in \mathcal{D}$ and for all $T > 0$, there exists a unique C^1 solution $x\colon\ [0,T] \mapsto \mathcal{D}$ satisfying (11.5.1). If $x_\mathrm{e} \in \mathcal{D}$ satisfies $f(x_\mathrm{e}) = 0$, then $x(t) \equiv x_\mathrm{e}$ is an *equilibrium* of (11.5.1). The following definition concerns the stability of an equilibrium. Throughout this section, let $\|\cdot\|$ denote a norm on $\mathbb{R}^n$.

Definition 11.5.1. Let $x_\mathrm{e} \in \mathcal{D}$ be an equilibrium of (11.5.1). Then, x_e is *Lyapunov stable* if, for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $\|x(0) - x_\mathrm{e}\| < \delta$, then $\|x(t) - x_\mathrm{e}\| < \varepsilon$ for all $t \geq 0$. Furthermore, x_e is *asymptotically stable* if it is Lyapunov stable and there exists $\varepsilon > 0$ such that, if $\|x(0) - x_\mathrm{e}\| < \varepsilon$, then $\lim_{t\to\infty} x(t) = x_\mathrm{e}$. In addition, x_e is *globally asymptotically stable* if it is Lyapunov stable, $\mathcal{D} = \mathbb{R}^n$, and, for all $x(0) \in \mathbb{R}^n$, $\lim_{t\to\infty} x(t) = x_\mathrm{e}$. Finally, x_e is *unstable* if it is not Lyapunov stable.

Note that if $x_\mathrm{e} \in \mathbb{R}^n$ is a globally asymptotically stable equilibrium, then x_e is the only equilibrium of (11.5.1).

The following result, known as *Lyapunov's direct method*, gives sufficient conditions for Lyapunov stability and asymptotic stability of an equilibrium of (11.5.1).

Theorem 11.5.2. Let $x_\mathrm{e} \in \mathcal{D}$ be an equilibrium of the dynamical system (11.5.1) and assume that there exists a C^1 function $V\colon\ \mathcal{D} \mapsto \mathbb{R}$ such that

$$V(x_{\rm e}) = 0, \tag{11.5.2}$$

such that, for all $x \in \mathcal{D}\backslash\{x_{\rm e}\}$,

$$V(x) > 0, \tag{11.5.3}$$

and such that, for all $x \in \mathcal{D}$,

$$V'(x)f(x) \le 0. \tag{11.5.4}$$

Then, $x_{\rm e}$ is Lyapunov stable. If, in addition, for all $x \in \mathcal{D}\backslash\{x_{\rm e}\}$,

$$V'(x)f(x) < 0, \tag{11.5.5}$$

then $x_{\rm e}$ is asymptotically stable. Finally, if $\mathcal{D} = \mathbb{R}^n$ and

$$\lim_{\|x\|\to\infty} V(x) = \infty, \tag{11.5.6}$$

then $x_{\rm e}$ is globally asymptotically stable.

Proof. For convenience, let $x_{\rm e} = 0$. To prove Lyapunov stability, let $\varepsilon > 0$ be such that $\mathbb{B}_\varepsilon(0) \subseteq \mathcal{D}$. Since $\mathbb{S}_\varepsilon(0)$ is compact and $V(x)$ is continuous, it follows from Theorem 10.3.7 that $V(\mathbb{S}_\varepsilon(0))$ is compact. Since $0 \notin \mathbb{S}_\varepsilon(0)$, $V(x) > 0$ for all $x \in \mathcal{D}\backslash\{0\}$, and $V(\mathbb{S}_\varepsilon(0))$ is compact, it follows that $\alpha \triangleq \min V(\mathbb{S}_\varepsilon(0))$ is positive. Next, since V is continuous, it follows that there exists $\delta \in (0, \varepsilon]$ such that $V(x) < \alpha$ for all $x \in \mathbb{B}_\delta(0)$. Now, let $x(t)$ for all $t \ge 0$ satisfy (11.5.1), where $\|x(0)\| < \delta$. Hence, $V(x(0)) < \alpha$. It thus follows from (11.5.4) that, for all $t \ge 0$,

$$V(x(t)) - V(x(0)) = \int_0^t V'(x(s))f(x(s))\,{\rm d}s \le 0,$$

and hence, for all $t \ge 0$,

$$V(x(t)) \le V(x(0)) < \alpha.$$

Now, since $V(x) \ge \alpha$ for all $x \in \mathbb{S}_\varepsilon(0)$, it follows that $x(t) \notin \mathbb{S}_\varepsilon(0)$ for all $t \ge 0$. Hence, $\|x(t)\| < \varepsilon$ for all $t \ge 0$, which proves that $x_{\rm e} = 0$ is Lyapunov stable.

To prove that $x_{\rm e} = 0$ is asymptotically stable, let $\varepsilon > 0$ be such that $\mathbb{B}_\varepsilon(0) \subseteq \mathcal{D}$. Since (11.5.5) implies (11.5.4), it follows that there exists $\delta > 0$ such that, if $\|x(0)\| < \delta$, then $\|x(t)\| < \varepsilon$ for all $t \ge 0$. Furthermore, $\frac{\rm d}{{\rm d}t}V(x(t)) = V'(x(t))f(x(t)) < 0$ for all $t \ge 0$, and thus $V(x(t))$ is decreasing and bounded from below by zero. Now, suppose that $V(x(t))$ does not converge to zero. Therefore, there exists $L > 0$ such that $V(x(t)) \ge L$ for all $t \ge 0$. Now, let $\delta' > 0$ be such that $V(x) < L$ for all $x \in \mathbb{B}_{\delta'}(0)$. Therefore, $\|x(t)\| \ge \delta'$ for all $t \ge 0$. Next, define $\gamma < 0$ by $\gamma \triangleq \max_{\delta' \le \|x\| \le \varepsilon} V'(x)f(x)$.

Therefore, since $\|x(t)\| < \varepsilon$ for all $t \geq 0$, it follows that

$$V(x(t)) - V(x(0)) = \int_0^t V'(x(\tau))f(x(\tau))\,\mathrm{d}\tau \leq \gamma t,$$

and hence

$$V(x(t)) \leq V(x(0)) + \gamma t.$$

However, $t > -V(x(0))/\gamma$ implies that $V(x(t)) < 0$, which is a contradiction.

To prove that $x_{\mathrm{e}} = 0$ is globally asymptotically stable, let $x(0) \in \mathbb{R}^n$, and let $\beta \triangleq V(x(0))$. It follows from (11.5.6) that there exists $\varepsilon > 0$ such that $V(x) > 2\beta$ for all $x \in \mathbb{R}^n$ such that $\|x\| > \varepsilon$. Therefore, $\|x(0)\| < \varepsilon$, and, since $V(x(t))$ is decreasing, it follows that $\|x(t)\| < \varepsilon$ for all $t > 0$. The remainder of the proof is identical to the proof of asymptotic stability. □

11.6 Linear Stability Theory

We now specialize Definition 11.5.1 to the linear system

$$\dot{x}(t) = Ax(t), \tag{11.6.1}$$

where $t \geq 0$, $x(t) \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n\times n}$. Note that $x_{\mathrm{e}} = 0$ is an equilibrium of (11.6.1), and that $x_{\mathrm{e}} \in \mathbb{R}^n$ is an equilibrium of (11.6.1) if and only if $x_{\mathrm{e}} \in \mathcal{N}(A)$. Hence, if x_{e} is the globally asymptotically stable equilibrium of (11.6.1), then A is nonsingular and $x_{\mathrm{e}} = 0$.

We consider three types of stability for the linear system (11.6.1). Unlike Definition 11.5.1, these definitions are stated in terms of the dynamics rather than the equilibrium.

Definition 11.6.1. For $A \in \mathbb{F}^{n\times n}$, define the following classes of matrices:

i) A is *Lyapunov stable* if $\operatorname{spec}(A) \subset \mathrm{CLHP}$ and, if $\lambda \in \operatorname{spec}(A)$ and $\operatorname{Re}\lambda = 0$, then λ is semisimple.

ii) A is *semistable* if $\operatorname{spec}(A) \subset \mathrm{OLHP} \cup \{0\}$ and, if $0 \in \operatorname{spec}(A)$, then 0 is semisimple.

iii) A is *asymptotically stable* if $\operatorname{spec}(A) \subset \mathrm{OLHP}$.

The following result concerns Lyapunov stability, semistability, and asymptotic stability for (11.6.1).

Proposition 11.6.2. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements are equivalent:

i) $x_{\rm e} = 0$ is a Lyapunov stable equilibrium of (11.6.1).

ii) At least one equilibrium of (11.6.1) is Lyapunov stable.

iii) Every equilibrium of (11.6.1) is Lyapunov stable.

iv) A is Lyapunov stable.

v) For every initial condition $x(0) \in \mathbb{R}^n$, $x(t)$ is bounded for all $t \geq 0$.

vi) $\|e^{tA}\|$ is bounded for all $t \geq 0$, where $\|\cdot\|$ is a norm on $\mathbb{R}^{n\times n}$.

vii) For every initial condition $x(0) \in \mathbb{R}^n$, $e^{tA}x(0)$ is bounded for all $t \geq 0$.

The following statements are equivalent:

vii) A is semistable.

viii) $\lim_{t\to\infty} e^{tA}$ exists. In fact, $\lim_{t\to\infty} e^{tA} = I - AA^{\#}$.

ix) For every initial condition $x(0)$, $\lim_{t\to\infty} x(t)$ exists.

The following statements are equivalent:

x) $x_{\rm e} = 0$ is an asymptotically stable equilibrium of (11.6.1).

xi) A is asymptotically stable.

xii) $\mathrm{spabs}(A) < 0$.

xiii) For every initial condition $x(0) \in \mathbb{R}^n$, $\lim_{t\to\infty} x(t) = 0$.

xiv) For every initial condition $x(0) \in \mathbb{R}^n$, $e^{tA}x(0) \to 0$ as $t \to \infty$.

xv) $e^{tA} \to 0$ as $t \to \infty$.

The following definition concerns the stability of a polynomial.

Definition 11.6.3. Let $p \in \mathbb{R}[s]$. Then, define the following terminology:

i) p is *Lyapunov stable* if $\mathrm{roots}(p) \subset \mathrm{CLHP}$ and, if λ is an imaginary root of p, then $\mathrm{m}_p(\lambda) = 1$.

ii) p is *semistable* if $\mathrm{roots}(p) \subset \mathrm{OLHP} \cup \{0\}$ and, if $0 \in \mathrm{roots}(p)$, then $\mathrm{m}_p(0) = 1$.

iii) p is *asymptotically stable* if $\mathrm{roots}(p) \subset \mathrm{OLHP}$.

For the following result, recall Definition 11.6.1.

Proposition 11.6.4. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is Lyapunov stable if and only if μ_A is Lyapunov stable.

ii) A is semistable if and only if μ_A is semistable.

Furthermore, the following statements are equivalent:

iii) A is asymptotically stable

iv) μ_A is asymptotically stable.

v) χ_A is asymptotically stable.

Next, consider the factorization of the minimal polynomial μ_A of A given by

$$\mu_A = \mu_A^{\rm s}\mu_A^{\rm u}, \tag{11.6.2}$$

where $\mu_A^{\rm s}$ and $\mu_A^{\rm u}$ are monic polynomials such that

$$\operatorname{roots}(\mu_A^{\rm s}) \subset \text{OLHP} \tag{11.6.3}$$

and

$$\operatorname{roots}(\mu_A^{\rm u}) \subset \text{CRHP}. \tag{11.6.4}$$

Proposition 11.6.5. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1}, \tag{11.6.5}$$

where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ satisfies $\operatorname{spec}(A_2) \subset \text{CRHP}$. Then,

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} 0 & C_{12\rm s} \\ 0 & \mu_A^{\rm s}(A_2) \end{bmatrix}S^{-1}, \tag{11.6.6}$$

where $C_{12\rm s} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\rm s}(A_2)$ is nonsingular, and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & C_{12\rm u} \\ 0 & 0 \end{bmatrix}S^{-1}, \tag{11.6.7}$$

where $C_{12\rm u} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\rm u}(A_1)$ is nonsingular. Consequently,

$$\mathcal{N}[\mu_A^{\rm s}(A)] = \mathcal{R}[\mu_A^{\rm u}(A)] = \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right). \tag{11.6.8}$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\rm s}(A_2) \end{bmatrix}S^{-1} \tag{11.6.9}$$

and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1}. \tag{11.6.10}$$

Consequently,

$$\mathcal{R}[\mu_A^{\mathrm{s}}(A)] = \mathcal{N}[\mu_A^{\mathrm{u}}(A)] = \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right). \tag{11.6.11}$$

Corollary 11.6.6. Let $A \in \mathbb{R}^{n\times n}$. Then,

$$\mathcal{N}[\mu_A^{\mathrm{s}}(A)] = \mathcal{R}[\mu_A^{\mathrm{u}}(A)] \tag{11.6.12}$$

and

$$\mathcal{N}[\mu_A^{\mathrm{u}}(A)] = \mathcal{R}[\mu_A^{\mathrm{s}}(A)]. \tag{11.6.13}$$

Proof. It follows from Theorem 5.3.5 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that (11.6.5) is satisfied, where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable, $A_{12} = 0$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ satisfies $\operatorname{spec}(A_2) \subset$ CRHP. The result now follows from Proposition 11.6.5. □

In view of Corollary 11.6.6 we define the *asymptotically stable subspace* $\mathcal{S}_{\mathrm{s}}(A)$ of A by

$$\mathcal{S}_{\mathrm{s}}(A) \triangleq \mathcal{N}[\mu_A^{\mathrm{s}}(A)] = \mathcal{R}[\mu_A^{\mathrm{u}}(A)] \tag{11.6.14}$$

and the *unstable subspace* $\mathcal{S}_{\mathrm{u}}(A)$ of A by

$$\mathcal{S}_{\mathrm{u}}(A) \triangleq \mathcal{N}[\mu_A^{\mathrm{u}}(A)] = \mathcal{R}[\mu_A^{\mathrm{s}}(A)]. \tag{11.6.15}$$

Note that

$$\dim \mathcal{S}_{\mathrm{s}}(A) = \operatorname{def} \mu_A^{\mathrm{s}}(A) = \operatorname{rank} \mu_A^{\mathrm{u}}(A) = \sum_{\substack{\lambda\in\operatorname{spec}(A)\\ \operatorname{Re}\lambda<0}} \operatorname{am}_A(\lambda) \tag{11.6.16}$$

and

$$\dim \mathcal{S}_{\mathrm{u}}(A) = \operatorname{def} \mu_A^{\mathrm{u}}(A) = \operatorname{rank} \mu_A^{\mathrm{s}}(A) = \sum_{\substack{\lambda\in\operatorname{spec}(A)\\ \operatorname{Re}\lambda\ge 0}} \operatorname{am}_A(\lambda). \tag{11.6.17}$$

Lemma 11.6.7. Let $A \in \mathbb{R}^{n\times n}$, assume that $\operatorname{spec}(A) \subset$ CRHP, let $x \in \mathbb{R}^n$, and assume that $\lim_{t\to\infty} e^{tA}x = 0$. Then, $x = 0$.

For the following result, note Proposition 11.6.2, Proposition 5.5.8, Fact 3.5.12, Fact 11.14.3, and Proposition 6.1.7.

Proposition 11.6.8. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) $\mathcal{S}_{\mathrm{s}}(A) = \{x \in \mathbb{R}^n\colon\ \lim_{t\to\infty} e^{tA}x = 0\}$.

ii) $\mu_A^{\mathrm{s}}(A)$ and $\mu_A^{\mathrm{u}}(A)$ are group invertible.

iii) $P_{\mathrm{s}} \triangleq I - \mu_A^{\mathrm{s}}(A)[\mu_A^{\mathrm{s}}(A)]^{\#}$ and $P_{\mathrm{u}} \triangleq I - \mu_A^{\mathrm{u}}(A)[\mu_A^{\mathrm{u}}(A)]^{\#}$ are idempotent.

iv) $P_s + P_u = I$.

v) $P_{s\perp} = P_u$ and $P_{u\perp} = P_s$.

vi) $\mathcal{S}_s(A) = \mathcal{R}(P_s) = \mathcal{N}(P_u)$.

vii) $\mathcal{S}_u(A) = \mathcal{R}(P_u) = \mathcal{N}(P_s)$.

viii) $\mathcal{S}_s(A)$ and $\mathcal{S}_u(A)$ are invariant subspaces of A.

ix) $\mathcal{S}_s(A)$ and $\mathcal{S}_u(A)$ are complementary subspaces.

x) P_s is the idempotent matrix onto $\mathcal{S}_s(A)$ along $\mathcal{S}_u(A)$.

xi) P_u is the idempotent matrix onto $\mathcal{S}_u(A)$ along $\mathcal{S}_s(A)$.

Proof. To prove *i)* let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable and $\operatorname{spec}(A_2) \subset \mathrm{CRHP}$. It then follows from Proposition 11.6.5 that

$$\mathcal{S}_s(A) = \mathcal{N}[\mu_A^s(A)] = \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

Furthermore,

$$e^{tA} = S\begin{bmatrix} e^{tA_1} & 0 \\ 0 & e^{tA_2} \end{bmatrix} S^{-1}.$$

To prove $\mathcal{S}_s(A) \subseteq \{x \in \mathbb{R}^n\colon\ \lim_{t\to\infty} e^{tA}x = 0\}$, let $x \triangleq S\left[\begin{smallmatrix} x_1 \\ 0 \end{smallmatrix}\right] \in \mathcal{S}_s(A)$, where $x_1 \in \mathbb{R}^r$. Then, $e^{tA}x = S\left[\begin{smallmatrix} e^{tA_1}x_1 \\ 0 \end{smallmatrix}\right] \to 0$ as $t \to \infty$. Hence, $x \in \{x \in \mathbb{R}^n\colon\ \lim_{t\to\infty} e^{tA}x = 0\}$. Conversely, to prove $\{x \in \mathbb{R}^n\colon\ \lim_{t\to\infty} e^{tA}x = 0\} \subseteq \mathcal{S}_s(A)$, let $x \triangleq S\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^n$ satisfy $\lim_{t\to\infty} e^{tA}x = 0$. Hence, $e^{tA_2}x_2 \to 0$ as $t \to \infty$. Since $\operatorname{spec}(A_2) \subset \mathrm{CRHP}$, it follows from Lemma 11.6.7 that $x_2 = 0$. Hence, $x \in \mathcal{R}\left(S\left[\begin{smallmatrix} I_r \\ 0 \end{smallmatrix}\right]\right) = \mathcal{S}_s(A)$.

The remaining statements follow directly from Proposition 11.6.5. □

11.7 The Lyapunov Equation

In this section we specialize Theorem 11.5.2 to the linear system (11.6.1).

Corollary 11.7.1. Let $A \in \mathbb{R}^{n\times n}$ and assume that there exists a nonneg-ative-semidefinite matrix $R \in \mathbb{R}^{n\times n}$ and a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ satisfying

$$A^{\mathrm{T}}P + PA + R = 0. \tag{11.7.1}$$

Then, A is Lyapunov stable. If, in addition, for all nonzero $\omega \in \mathbb{R}$,

$$\operatorname{rank}\begin{bmatrix} \jmath\omega I - A \\ R \end{bmatrix} = n, \tag{11.7.2}$$

then A is semistable. Finally, if R is positive definite, then A is asymptotically stable.

Proof. Define $V(x) \triangleq x^{\mathrm{T}}Px$, which satisfies (11.5.2) with $x_{\mathrm{e}} = 0$ and satisfies (11.5.3) for all nonzero $x \in \mathcal{D} = \mathbb{R}^n$. Furthermore, Theorem 11.5.2 implies that $V'(x)f(x) = 2x^{\mathrm{T}}PAx = x^{\mathrm{T}}\big(A^{\mathrm{T}}P + PA\big)x = -x^{\mathrm{T}}Rx$, which satisfies (11.5.4) for all $x \in \mathbb{R}^n$. Thus, Theorem 11.5.2 implies that A is Lyapunov stable. If, in addition, R is positive definite, then (11.5.5) is satisfied for all $x \neq 0$, and thus A is asymptotically stable.

Alternatively, we shall prove the first and third statements without using Theorem 11.5.2. Letting $\lambda \in \operatorname{spec}(A)$ and $x \in \mathbb{C}^n$ be an associated eigenvector, it follows that $0 \geq -x^*Rx = x^*\big(A^{\mathrm{T}}P + PA\big)x = (\overline{\lambda} + \lambda)x^*Px$. Therefore, $\operatorname{spec}(A) \subset$ CLHP. Now, suppose that $\jmath\omega \in \operatorname{spec}(A)$, where $\omega \in \mathbb{R}$, and let $x \in \mathcal{N}\big[(\jmath\omega I - A)^2\big]$. Defining $y \triangleq (\jmath\omega I - A)x$, it follows that $(\jmath\omega I - A)y = 0$ and thus $Ay = \jmath\omega y$. Therefore, $-y^*Ry = y^*\big(A^{\mathrm{T}}P + PA\big)y = -\jmath\omega y^*Py + \jmath\omega y^*Py = 0$, and thus $Ry = 0$. Hence, $0 = x^*Ry = -x^*\big(A^{\mathrm{T}}P + PA\big)y = -x^*\big(A^{\mathrm{T}} + \jmath\omega I\big)Py = y^*Py$. Since P is positive definite, it follows that $y = 0$, that is, $(\jmath\omega I - A)x = 0$. Therefore, $x \in \mathcal{N}(\jmath\omega I - A)$. Now, Proposition 5.5.14 implies that $\jmath\omega$ is semisimple. Therefore, A is Lyapunov stable.

Next, to prove that A is asymptotically stable, let $\lambda \in \operatorname{spec}(A)$, and let $x \in \mathbb{C}^n$ be an associated eigenvector. Thus, $0 > -x^*Rx = (\overline{\lambda} + \lambda)x^*Px$, which implies that A is asymptotically stable.

Finally, to prove that A is semistable, let $\jmath\omega \in \operatorname{spec}(A)$, where $\omega \in \mathbb{R}$ is nonzero, and let $x \in \mathbb{C}^n$ be an associated eigenvector. Then,

$$-x^*Rx = x^*\big(A^{\mathrm{T}}P + PA\big)x = x^*[(\jmath\omega I - A)^*P + P(\jmath\omega I - A]x = 0.$$

Therefore, $Rx = 0$ and thus

$$\begin{bmatrix} \jmath\omega I - A \\ R \end{bmatrix} x = 0,$$

which implies that $x = 0$, which contradicts $x \neq 0$. Consequently, $\jmath\omega \notin \operatorname{spec}(A)$ for all nonzero $\omega \in \mathbb{R}$, and thus A is semistable. □

Equation (11.7.1) is a *Lyapunov equation.* Converse results for Corollary 11.7.1 are given by Corollary 11.7.4, Proposition 11.7.5, Proposition 11.7.6, Proposition 11.7.7, and Proposition 12.7.5. The following lemma will be useful for analyzing (11.7.1).

Lemma 11.7.2. Assume that $A \in \mathbb{F}^{n\times n}$ is asymptotically stable. Then,

$$\int_0^\infty e^{tA}\,\mathrm{d}t = -A^{-1}. \tag{11.7.3}$$

Proof. Proposition 11.1.5 implies that $\int_0^t e^{\tau A}\,\mathrm{d}\tau = A^{-1}\big(e^{tA} - I\big)$. Letting $t \to \infty$ yields (11.7.3). $\square$

The following result concerns Sylvester's equation. See Fact 5.8.11 and Proposition 7.2.4.

Proposition 11.7.3. Let $A, B, C \in \mathbb{R}^{n\times n}$. Then, there exists a unique matrix $X \in \mathbb{R}^{n\times n}$ satisfying

$$AX + XB + C = 0 \tag{11.7.4}$$

if and only if $B^{\mathrm{T}} \oplus A$ is nonsingular. In this case, X is given by

$$X = -\operatorname{vec}^{-1}\Big[\big(B^{\mathrm{T}} \oplus A\big)^{-1}\operatorname{vec} C\Big]. \tag{11.7.5}$$

If, in addition, $B^{\mathrm{T}} \oplus A$ is asymptotically stable, then X is given by

$$X = \int_0^\infty e^{tA} C e^{tB}\,\mathrm{d}t. \tag{11.7.6}$$

Proof. The first two statements follow from Proposition 7.2.4. If $B^{\mathrm{T}} \oplus A$ is asymptotically stable, then it follows from (11.7.5) using Lemma 11.7.2 and Proposition 11.1.8 that

$$\begin{aligned} X &= \int_0^\infty \operatorname{vec}^{-1}\Big(e^{t(B^{\mathrm{T}} \oplus A)}\operatorname{vec} C\Big)\,\mathrm{d}t = \int_0^\infty \operatorname{vec}^{-1}\Big(e^{tB^{\mathrm{T}}} \otimes e^{tA}\Big)\operatorname{vec} C\,\mathrm{d}t \\ &= \int_0^\infty \operatorname{vec}^{-1}\operatorname{vec}\big(e^{tA} C e^{tB}\big)\,\mathrm{d}t = \int_0^\infty e^{tA} C e^{tB}\,\mathrm{d}t. \end{aligned}$$

$\square$

The following result provides a converse to Corollary 11.7.1 for the case of asymptotic stability.

Corollary 11.7.4. Let $A \in \mathbb{R}^{n\times n}$, and let $R \in \mathbb{R}^{n\times n}$. Then, there exists a unique matrix $P \in \mathbb{R}^{n\times n}$ satisfying (11.7.1) if and only if $A \oplus A$ is nonsingular. In this case, if R is symmetric, then P is symmetric. Now, assume that A is asymptotically stable. Then, $P \in \mathbf{S}^n$ is given by

$$P = \int_0^\infty e^{tA^{\mathrm{T}}} R e^{tA}\, \mathrm{d}t. \tag{11.7.7}$$

Finally, if R is (nonnegative semidefinite, positive definite), then P is (nonnegative semidefinite, positive definite).

Proof. First note that $A \oplus A$ is nonsingular if and only if $(A \oplus A)^{\mathrm{T}} = A^{\mathrm{T}} \oplus A^{\mathrm{T}}$ is nonsingular. Now, the first statement follows from Proposition 11.7.3. To prove the second statement note that $A^{\mathrm{T}}(P - P^{\mathrm{T}}) + (P - P^{\mathrm{T}})A = 0$, which implies that P is symmetric. Now, suppose that A is asymptotically stable. Then, Fact 11.14.29 implies that $A \oplus A$ is asymptotically stable. Consequently, (11.7.7) follows from (11.7.6). □

The following result provides a converse to Corollary 11.7.1 for the case of Lyapunov stability.

Proposition 11.7.5. Let $A \in \mathbb{R}^{n \times n}$, and assume that A is Lyapunov stable. Then, there exist a positive-definite matrix P and a nonnegative-semidefinite matrix R satisfying (11.7.1).

Proof. Let $S \in \mathbb{R}^{n \times n}$ be a nonsingular matrix such that $SAS^{-1} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$ is in real Jordan form, where $A_1 \in \mathbb{R}^{n_1 \times n_1}$, $\mathrm{spec}(A_1) \subset \mathrm{OLHP}$, $\mathrm{spec}(A_2) \subset \jmath\mathbb{R}$, and A_2 is skew symmetric. Let $R_1 \in \mathbb{R}^{n_1 \times n_1}$ be positive definite and let $P_1 \in \mathbb{R}^{n_1 \times n_1}$ be the positive-definite solution to $A_1^{\mathrm{T}} P_1 + P_1 A_1 + R_1 = 0$. Since $A_2 + A_2^{\mathrm{T}} = 0$, it follows that $(SAS^{-1})^{\mathrm{T}} \hat{P} + \hat{P} SAS^{-1} + \hat{R} = 0$, where $\hat{P} \triangleq \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\hat{R} \triangleq \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix}$. Therefore, (11.7.1) is satisfied with $P \triangleq S^{\mathrm{T}} \hat{P} S$ and $R \triangleq S^{\mathrm{T}} \hat{R} S$. □

The following results also include converse statements. We first consider asymptotic stability.

Consider the Lyapunov equation

$$A^{\mathrm{T}} P + PA + R = 0. \tag{11.7.8}$$

Proposition 11.7.6. Let $A \in \mathbb{R}^{n \times n}$. The following statements are equivalent:

i) A is asymptotically stable.

ii) For all positive-definite matrices $R \in \mathbb{R}^{n \times n}$ there exists a positive-definite matrix $P \in \mathbb{R}^{n \times n}$ such that (11.7.8) is satisfied.

iii) There exists a positive-definite matrix $R \in \mathbb{R}^{n \times n}$ and a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that (11.7.8) is satisfied.

Proof. The result $i) \Longrightarrow ii)$ follows from Corollary 11.7.1. The implications $ii) \Longrightarrow iii)$ and $iii) \Longrightarrow iv)$ are immediate. To prove $iv) \Longrightarrow i)$ note that, since there exists a nonnegative-semidefinite matrix P satisfying (11.7.8), it follows from Proposition 12.4.4 that (A, C) is completely undetectable. Thus, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $A = S\left[\begin{smallmatrix} A_1 & 0 \\ A_{21} & A_2 \end{smallmatrix}\right]S^{-1}$ and $C = \left[\begin{array}{cc} C_1 & 0 \end{array}\right]S^{-1}$, where (C_1, A_1) is observable and A_1 is asymptotically stable. Furthermore, since $(S^{-1}AS, CS)$ is detectable, it follows that A_2 is also asymptotically stable. Consequently, A is asymptotically stable. $\square$

Next, we consider the case of Lyapunov stability.

Proposition 11.7.7. Let $A \in \mathbb{R}^{n\times n}$. Then, A is Lyapunov stable if and only if there exists a nonnegative-semidefinite matrix $R \in \mathbb{R}^{n\times n}$ and a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that (11.7.8) is satisfied.

Proof. Suppose that A is Lyapunov stable. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $A = S\left[\begin{smallmatrix} A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right]S^{-1}$, $A_1 \in \mathbb{R}^{n_1\times n_1}$, $A_2 \in \mathbb{R}^{n_2\times n_2}$, $\operatorname{spec}(A_1) \subset \jmath\mathbb{R}$, and $\operatorname{spec}(A_2) \subset \mathrm{OLHP}$. Let $S_1 \in \mathbb{R}^{n_1\times n_1}$ be such that $A_1 = S_1 J_1 S_1^{-1}$, where $J_1 \in \mathbb{R}^{n_1\times n_1}$ is skew symmetric. Then, it follows that $A_1^{\mathrm{T}}P_1 + P_1A_1 = 0$, where $P_1 = S_1^{-\mathrm{T}}S_1^{-1}$ is positive definite. Next, let $R_2 \in \mathbb{R}^{n_2\times n_2}$ be positive definite and let $P_2 \in \mathbb{R}^{n_2\times n_2}$ be the positive-definite solution to $A_2^{\mathrm{T}}P_2 + P_2A_2 + R_2 = 0$. Hence, (11.7.8) is satisfied with $P \triangleq S^{-\mathrm{T}}\left[\begin{smallmatrix} P_1 & 0 \\ 0 & P_2 \end{smallmatrix}\right]S^{-1}$ and $R \triangleq S^{-\mathrm{T}}\left[\begin{smallmatrix} 0 & 0 \\ 0 & R_2 \end{smallmatrix}\right]S^{-1}$.

Conversely, suppose that there exist a nonnegative-semidefinite matrix $R \in \mathbb{R}^{n\times n}$ and a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that (11.7.8) is satisfied. Let $\lambda \in \operatorname{spec}(A)$, and let $x \in \mathbb{R}^n$ be an eigenvector of A associated with λ. It thus follows from (11.7.8) that $0 = x^*A^{\mathrm{T}}Px + x^*PAx + x^*Rx = (\lambda + \overline{\lambda})x^*Px + x^*Rx$. Therefore, $\operatorname{Re}\lambda = -x^*Rx/(2x^*Px)$, which shows that $\operatorname{spec}(A) \subset \mathrm{CLHP}$. Now, let $\jmath\omega \in \operatorname{spec}(A)$, and suppose that $x \in \mathbb{R}^n$ satisfies $(\jmath\omega I - A)^2x = 0$. Then, $(\jmath\omega I - A)y = 0$, where $y = (\jmath\omega I - A)x$. Computing $0 = y^*\left(A^{\mathrm{T}}P + PA\right)y + y^*Ry$ yields $y^*Ry = 0$ and thus $Ry = 0$. Therefore, $\left(A^{\mathrm{T}}P + PA\right)y = 0$ and thus $y^*Py = (A^{\mathrm{T}} + \jmath\omega I)Py = 0$. Since P is positive definite, it follows that $y = (\jmath\omega I - A)x = 0$. Therefore, $\mathcal{N}(\jmath\omega I - A) = \mathcal{N}\left[(\jmath\omega I - A)^2\right]$. Hence, it follows from Corollary TBD that $\jmath\omega$ is semisimple, $\square$

Corollary 11.7.8. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is Lyapunov stable if and only if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that $A^{\mathrm{T}}P + PA$ is nonpositive semidefinite.

ii) A is asymptotically stable if and only if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that $A^{\mathrm{T}}P + PA$ is negative definite.

11.8 Discrete-Time Stability Theory

The theory of difference equations is concerned with the behavior of discrete-time dynamical systems of the form

$$x_{k+1} = f(x_k), \tag{11.8.1}$$

where $f\colon\ \mathbb{R}^n \to \mathbb{R}^n$, $k \in \mathbb{N}$, $x_k \in \mathbb{R}^n$, and x_0 is the initial condition. The solution $x_k \equiv x_{\mathrm{e}}$ is an equilibrium of (11.8.1) if $x_{\mathrm{e}} = f(x_{\mathrm{e}})$.

A linear discrete-time system has the form

$$x_{k+1} = Ax_k, \tag{11.8.2}$$

where $A \in \mathbb{R}^{n\times n}$. For $k \in \mathbb{N}$, x_k is given by

$$x_k = A^k x_0. \tag{11.8.3}$$

The behavior of $\{x_k\}_{k=0}^{\infty}$ is determined by the stability of A. To study the stability of discrete-time systems it is helpful to define the *open unit disk* (OUD) and the *closed unit disk* (CUD) by

$$\mathrm{OUD} \triangleq \{x \in \mathbb{C}\colon\ \ |x| < 1\} \tag{11.8.4}$$

and

$$\mathrm{CUD} \triangleq \{x \in \mathbb{C}\colon\ \ |x| \le 1\}. \tag{11.8.5}$$

Definition 11.8.1. For $A \in \mathbb{F}^{n\times n}$, define the following classes of matrices:

i) A is *discrete-time Lyapunov stable* if $\mathrm{spec}(A) \subset \mathrm{CUD}$ and, if $\lambda \in \mathrm{spec}(A)$ and $|\lambda| = 1$, then λ is semisimple.

ii) A is *discrete-time semistable* if $\mathrm{spec}(A) \subset \mathrm{OUD} \cup \{1\}$ and, if $1 \in \mathrm{spec}(A)$, then 1 is semisimple.

iii) A is *discrete-time asymptotically stable* if $\mathrm{spec}(A) \subset \mathrm{OUD}$.

Proposition 11.8.2. Let $A \in \mathbb{R}^{n\times n}$ and consider the linear discrete-time system (11.8.2). Then, the following statements are equivalent:

i) A is discrete-time Lyapunov stable.

ii) For every initial condition $x_0 \in \mathbb{R}^n$, x_k is bounded for all $k \in \mathbb{N}$.

iii) $\|A^k\|$ is bounded for all $k \in \mathbb{N}$, where $\|\cdot\|$ is a norm on $\mathbb{R}^{n\times n}$.

iv) For every initial condition $x_0 \in \mathbb{R}^n$, $A^k x_0$ is bounded for all $k \in \mathbb{N}$.

The following statements are equivalent:

v) A is discrete-time semistable.

vi) $\lim_{k\to\infty} A^k$ exists. In this case, $\lim_{k\to\infty} A^k = I - (I-A)^{\#}(I-A)$.

vii) For every initial condition $x_0 \in \mathbb{R}^n$, $\lim_{k\to\infty} x_k$ exists.

The following statements are equivalent:

viii) A is discrete-time asymptotically stable.

ix) $\mathrm{sprad}(A) < 1$.

x) For every initial condition $x_0 \in \mathbb{R}^n$, $\lim_{k\to\infty} x_k = 0$.

xi) For every initial condition $x_0 \in \mathbb{R}^n$, $A^k x_0 \to 0$ as $k \to \infty$.

xii) $A^k \to 0$ as $k \to \infty$.

The following definition concerns the discrete-time stability of a polynomial.

Definition 11.8.3. Let $p \in \mathbb{R}[s]$. Then, define the following terminology:

i) p is *discrete-time Lyapunov stable* if $\mathrm{roots}(p) \subset \mathrm{CUD}$ and, if λ is an imaginary root of p, then $\mathrm{m}_p(\lambda) = 1$.

ii) p is *discrete-time semistable* if $\mathrm{roots}(p) \subset \mathrm{OUD} \cup \{1\}$ and, if $1 \in \mathrm{roots}(p)$, then $\mathrm{m}_p(1) = 1$.

iii) p is *discrete-time asymptotically stable* if $\mathrm{roots}(p) \subset \mathrm{OUD}$.

Proposition 11.8.4. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is discrete-time Lyapunov stable if and only if μ_A is discrete-time Lyapunov stable.

ii) A is discrete-time semistable if and only if μ_A is discrete-time semistable.

Furthermore, the following statements are equivalent:

i) A is discrete-time asymptotically stable.

ii) μ_A is discrete-time asymptotically stable.

iii) χ_A is discrete-time asymptotically stable.

11.9 Facts on Matrix Exponential Formulas

Fact 11.9.1. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) If $A^2 = 0$, then $e^{tA} = I + tA$.

ii) If $A^2 = I$, then $e^{tA} = (\cosh t)I + (\sinh t)A$.

iii) If $A^2 = -I$, then $e^{tA} = (\cos t)I + (\sin t)A$.

iv) If $A^2 = A$, then $e^{tA} = I - A + e^t A$.

v) If $A^2 = -A$, then $e^{tA} = I + A - e^{-t}A$.

vi) If $\operatorname{rank} A = 1$ and $\operatorname{tr} A = 0$, then $e^{tA} = I + tA$.

vii) If $\operatorname{rank} A = 1$ and $\operatorname{tr} A \neq 0$, then $e^{tA} = I + \frac{e^{(\operatorname{tr} A)t}-1}{\operatorname{tr} A}A$.

(Remark: See [458].)

Fact 11.9.2. Let $A \triangleq \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}$. Then,

$$e^{tA} = (\cosh t)I_{2n} + (\sinh t)A.$$

Furthermore,

$$e^{tJ_{2n}} = (\cos t)I_{2n} + (\sin t)J_{2n}.$$

Fact 11.9.3. Let $A \in \mathbb{R}^{n\times n}$ be skew symmetric. Then, $\{e^{\theta A}\colon\ \theta \in \mathbb{R}\} \subseteq \mathrm{SO}(n)$ is a group. If, in addition, $n = 2$, then

$$\{e^{\theta J_2}\colon\ \theta \in \mathbb{R}\} = \mathrm{SO}(2).$$

(Remark: Note that $e^{\theta J_2} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$. See Fact 3.6.14.)

Fact 11.9.4. Let $A \in \mathbb{R}^{n\times n}$, where

$$A \triangleq \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & n-1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Then,

$$e^A = \begin{bmatrix} \binom{0}{0} & \binom{1}{0} & \binom{2}{0} & \binom{3}{0} & \cdots & \binom{n-1}{0} \\ 0 & \binom{1}{1} & \binom{2}{1} & \binom{3}{1} & \cdots & \binom{n-1}{1} \\ 0 & 0 & \binom{2}{2} & \binom{3}{2} & \cdots & \binom{n-1}{2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & \binom{n-1}{n-2} \\ 0 & 0 & 0 & 0 & \cdots & \binom{n-1}{n-1} \end{bmatrix}.$$

Furthermore, if $k \ge n$, then

$$\sum_{i=1}^{k} i^{n-1} = \begin{bmatrix} 1^{n-1} & 2^{n-1} & \cdots & n^{n-1} \end{bmatrix} e^{-A} \begin{bmatrix} \binom{k}{1} \\ \vdots \\ \binom{k}{n} \end{bmatrix}.$$

(Proof: See [35].)

Fact 11.9.5. Let $A \in \mathbb{F}^{3\times 3}$. If $\operatorname{spec}(A) = \{\lambda\}$, then

$$e^{tA} = e^{\lambda t}\big[I + t(A-\lambda I) + \tfrac{1}{2}t^2(A-\lambda I)^2\big].$$

If $\operatorname{mspec}(A) = \{\lambda, \lambda, \mu\}_{\mathrm{m}}$, where $\mu \ne \lambda$, then

$$e^{tA} = e^{\lambda t}[I + t(A-\lambda I)] + \left[\frac{e^{\mu t} - e^{\lambda t}}{(\mu-\lambda)^2} - \frac{te^{\lambda t}}{\mu - \lambda}\right](A-\lambda I)^2.$$

If $\operatorname{spec}(A) = \{\lambda, \mu, \nu\}$, then

$$e^{tA} = \frac{e^{\lambda t}}{(\lambda-\mu)(\lambda-\nu)}(A-\mu I)(A-\nu I) + \frac{e^{\mu t}}{(\mu-\lambda)(\mu-\nu)}(A-\lambda I)(A-\nu I)$$
$$+ \frac{e^{\nu t}}{(\nu-\lambda)(\nu-\mu)}(A-\lambda I)(A-\mu I).$$

(Proof: See [32].)

Fact 11.9.6. Let $z_1, z_2, z_3 \in \mathbb{R}$, and define

$$A \triangleq \begin{bmatrix} 0 & -z_3 & z_2 \\ z_3 & 0 & -z_1 \\ -z_2 & z_1 & 0 \end{bmatrix}.$$

Then,

$$\begin{aligned} e^A &= I + \frac{\sin\theta}{\theta}A + \frac{1-\cos\theta}{\theta^2}A^2 \\ &= I + \frac{\sin\theta}{\theta}A + \tfrac{1}{2}\left[\frac{\sin(\theta/2)}{\theta/2}\right]^2 A^2 \\ &= (\cos\theta)I + \frac{\sin\theta}{\theta}A + \frac{1-\cos\theta}{\theta^2}zz^{\mathrm{T}}, \end{aligned}$$

where $z \triangleq \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}^{\mathrm{T}}$ and $\theta \triangleq \|z\|_2$. (Remark: For $x \in \mathbb{R}^3$, $e^A x$ is the rotation of x about the vector $\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}^{\mathrm{T}}$ through the angle θ. See [89]. See Fact 11.9.8.) (Proof: The Cayley-Hamilton theorem implies $A^3 + \theta^2 A = 0$. Then, every term A^k in the expansion of e^A can be expressed in terms of A or A^2. Finally, $\theta^2 I + A^2 = zz^{\mathrm{T}}$.)

Fact 11.9.7. Let $A \in \mathbb{F}^{3\times 3}$ be unitary and assume there exists $\theta \in \mathbb{R}$ such that $\operatorname{tr} A = 1 + 2\cos\theta$ and $|\theta| < \pi$. Then,

$$e^{\frac{\theta}{2\sin\theta}(A - A^{\mathrm{T}})} = A.$$

(Proof: See [307, p. 364].)

Fact 11.9.8. Let $x, y \in \mathbb{R}^n$ satisfy $x^{\mathrm{T}}y = 0$, let $\theta \in [0, 2\pi]$, and define $A \in \mathbb{F}^{n\times n}$ by

$$A \triangleq I + (\sin\theta)\big(xy^{\mathrm{T}} - yx^{\mathrm{T}}\big) - (1 - \cos\theta)\big(xx^{\mathrm{T}} + yy^{\mathrm{T}}\big).$$

Then, A is orthogonal and $\det A = 1$. Now, let $n = 3$ and $z \triangleq y \times x$. Then,

$$A = (\cos\theta)I + (\sin\theta)C(z) + (1 - \cos\theta)zz^{\mathrm{T}},$$

where

$$C(z) \triangleq \begin{bmatrix} 0 & -z_{(3)} & z_{(2)} \\ z_{(3)} & 0 & -z_{(1)} \\ -z_{(2)} & z_{(1)} & 0 \end{bmatrix}.$$

If, in addition, $\theta \neq \pi$, then

$$A = (I - B)(I + B)^{-1},$$

where

$$B \triangleq -\tan(\theta/2)C(z).$$

(Remark: See Fact 11.9.6.) (Problem: Represent A as a matrix exponential.)

Fact 11.9.9. Let $x, y \in \mathbb{R}^3$ be nonzero. Then, there exists a skew-symmetric matrix $A \in \mathbb{R}^{3\times 3}$ such that $y = e^A x$ if and only if $x^{\mathrm{T}}x = y^{\mathrm{T}}y$. If $x \neq -y$, then one such matrix is $A = \phi C(z)$, where

$$z \triangleq \|x \times y\|_2^{-1} x \times y,$$

$$C(z) \triangleq \begin{bmatrix} 0 & -z_{(3)} & z_{(2)} \\ z_{(3)} & 0 & -z_{(1)} \\ -z_{(2)} & z_{(1)} & 0 \end{bmatrix},$$

and

$$\phi \triangleq \cos^{-1}\big(x^{\mathrm{T}}y\big).$$

If $x = -y$, then one such matrix is $A = \pi C(z)$, where $z \triangleq \nu \times y$ and $\nu \in \{y\}^\perp$ satisfies $\nu^{\mathrm{T}}\nu = 1$. (Remark: Since $\det e^A = e^{\operatorname{tr} A}$, it follows that vectors in $\mathbb{R}^3$ having the same Euclidean length are always related by a *proper rotation*. See Fact 3.6.17 and Fact 3.7.3.) (Problem: Extend this result to $\mathbb{R}^n$. See [58].)

Fact 11.9.10. Let $A \in \mathbb{R}^{4\times 4}$ be skew symmetric with $\operatorname{mspec}(A) = \{\jmath\omega, -\jmath\omega, \jmath\mu, -\jmath\mu\}_{\mathrm{m}}$. If $\omega \neq \mu$, then

$$e^A = a_3A^3 + a_2A^2 + a_1A + a_0I,$$

where

$$\begin{aligned} a_3 &= \left(\omega^2 - \mu^2\right)^{-1}\left(\tfrac{1}{\mu}\sin\mu - \tfrac{1}{\omega}\sin\omega\right),\\ a_2 &= \left(\omega^2 - \mu^2\right)^{-1}(\cos\mu - \cos\omega),\\ a_1 &= \left(\omega^2 - \mu^2\right)^{-1}\left(\tfrac{\omega^2}{\mu}\sin\mu - \tfrac{\mu^2}{\omega}\sin\omega\right),\\ a_0 &= \left(\omega^2 - \mu^2\right)^{-1}\left(\omega^2\cos\mu - \mu^2\cos\omega\right). \end{aligned}$$

If $\omega = \mu$, then

$$e^A = (\cos\omega)I + \frac{\sin\omega}{\omega}A.$$

(Proof: See [250, p. 18] and [459].) (Remark: There are typographical errors in [250, p. 18] and [459].)

Fact 11.9.11. Let $C \in \mathbb{R}^{n\times n}$ be nonsingular and let $k \in \mathbb{P}$. Then, there exists $B \in \mathbb{R}^{n\times n}$ such that $C^{2k} = e^B$. (Proof: Use Proposition 11.4.4 with $A = C^2$ and note that every negative eigenvalue $-\alpha < 0$ of C^2 arises as the square of complex conjugate eigenvalues $\pm\jmath\sqrt{\alpha}$ of C.)

11.10 Facts on Matrix Exponential Identities Involving One Matrix

Fact 11.10.1. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is (lower triangular, upper triangular). Then, so is e^A. If, in addition, A is Toeplitz, then so is e^A. (Remark: See Fact 3.12.7.)

Fact 11.10.2. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\operatorname{sprad}\left(e^A\right) = e^{\operatorname{spabs}(A)}.$$

Fact 11.10.3. Let $A \in \mathbb{R}^{n\times n}$. Then, the matrix differential equation

$$\dot{X}(t) = AX(t),$$

where $t \geq 0$ and $X(t) \in \mathbb{R}^{n \times n}$, has the solution

$$X(t) = e^{tA}X(0).$$

Fact 11.10.4. Let $A\colon\ [0,T] \to \mathbb{R}^{n \times n}$ be continuous and assume that the matrix differential equation

$$\dot{X}(t) = A(t)X(t),$$

has a solution $X(t) \in \mathbb{R}^{n \times n}$. Then,

$$\det X(t) = e^{\int_0^t \operatorname{tr} A(\tau)\,\mathrm{d}\tau}\det X(0).$$

(Remark: This result is *Jacobi's identity.*)

Fact 11.10.5. Let $A \in \mathbb{R}^{n \times n}$, let $\lambda \in \operatorname{spec}(A)$, and let $v \in \mathbb{C}^n$ be an eigenvector of A associated with λ. Then, for all $t \geq 0$,

$$x(t) \triangleq \operatorname{Re}\left(e^{\lambda t}v\right)$$

satisfies $\dot{x}(t) = Ax(t)$.

Fact 11.10.6. Let $S\colon\ [t_0, t_1] \to \mathbb{R}^{n \times n}$ be differentiable. Then, for all $t \in [t_0, t_1]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}S^2(t) = \dot{S}(t)S(t) + S(t)\dot{S}(t).$$

Let $S_1\colon\ [t_0, t_1] \to \mathbb{R}^{n \times m}$ and $S_2\colon\ [t_0, t_1] \to \mathbb{R}^{m \times l}$ be differentiable. Then, for all $t \in [t_0, t_1]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}S_1(t)S_2(t) = \dot{S}_1(t)S_2(t) + S_1(t)\dot{S}_2(t).$$

Fact 11.10.7. Let $A \in \mathbb{F}^{n \times n}$, and let $A_1 = \frac{1}{2}(A + A^*)$ and $A_2 = \frac{1}{2}(A - A^*)$. Then, $A_1A_2 = A_2A_1$ if and only if A is normal. In this case, $e^{A_1}e^{A_2}$ is the polar decomposition of e^A. (Remark: See Fact 3.4.22.) (Problem: Obtain the polar decomposition of e^A when A is not normal.)

Fact 11.10.8. Let $A \in \mathbb{F}^{n \times m}$, and assume that $\operatorname{rank} A = m$. Then,

$$A^+ = \int_0^\infty e^{-tA^*A}A^*\,\mathrm{d}t.$$

Fact 11.10.9. Let $A \in \mathbb{F}^{n \times n}$, and assume that A is nonsingular. Then,

$$A^{-1} = \int_0^\infty e^{-tA^*A}\,\mathrm{d}t A^*.$$

Fact 11.10.10. Let $A \in \mathbb{F}^{n\times n}$ and let $k \triangleq \operatorname{ind} A$. Then,

$$A^{\mathrm{D}} = \int_0^\infty e^{-tA^kA^{(2k+1)*}A^{k+1}}\,\mathrm{d}t A^kA^{(2k+1)*}A^k.$$

(Proof: See [237].)

Fact 11.10.11. Let $A \in \mathbb{F}^{n\times n}$ and assume that $\operatorname{ind} A = 1$. Then,

$$A^{\#} = \int_0^\infty e^{-tAA^{3*}A^2}\,\mathrm{d}t AA^{3*}A.$$

(Proof: See Fact 11.10.10.)

Fact 11.10.12. Let $A \in \mathbb{F}^{n\times n}$ and let $k \triangleq \operatorname{ind} A$. Then,

$$\int_0^t e^{\tau A}\,\mathrm{d}\tau = A^{\mathrm{D}}\big(e^{tA} - I\big) + \big(I - AA^{\mathrm{D}}\big)\Big(tI + \tfrac{1}{2!}t^2A + \cdots + \tfrac{1}{k!}t^kA^{k-1}\Big).$$

If, in particular, A is group invertible, then

$$\int_0^t e^{\tau A}\,\mathrm{d}\tau = A^{\#}\big(e^{tA} - I\big) + \Big(I - AA^{\#}\Big)t.$$

Fact 11.10.13. Let $A \in \mathbb{F}^{n\times n}$, let $\operatorname{mspec}(A) = \{\lambda_1, \ldots, \lambda_r, 0, \ldots, 0\}_{\mathrm{m}}$, where $\lambda_1, \ldots, \lambda_r$ are nonzero, and let $t > 0$. Then,

$$\det\int_0^t e^{\tau A}\,\mathrm{d}\tau = t^{n-r}\prod_{i=1}^r \lambda_i^{-1}\Big(e^{\lambda_i t} - 1\Big).$$

Hence, $\det\int_0^t e^{\tau A}\,\mathrm{d}\tau \neq 0$ if and only if $2\pi jk/t \notin \operatorname{spec}(A)$ for all $k \in \mathbb{P}$. Finally, $\det\big(e^{tA} - I\big) \neq 0$ if and only if $\det A \neq 0$ and $\det\int_0^t e^{\tau A}\,\mathrm{d}\tau \neq 0$.

Fact 11.10.14. Let $A \in \mathbb{F}^{n\times n}$, and assume that e^A is orthogonal. Then, either A is skew symmetric or two eigenvalues of A differ by a nonzero integer multiple of $2\pi j$. (Remark: See [620].)

11.11 Facts on Matrix Exponential Identities Involving Two or More Matrices

Fact 11.11.1. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, and $C \in \mathbb{F}^{m\times m}$. Then,

$$e^{t\left[\begin{smallmatrix} A & B \\ 0 & C \end{smallmatrix}\right]} = \begin{bmatrix} e^{tA} & \int_0^t e^{(t-\tau)A}Be^{\tau C}\,\mathrm{d}\tau \\ 0 & e^{tC} \end{bmatrix}.$$

Furthermore,

$$\int_0^t e^{\tau A}\,\mathrm{d}\tau = \begin{bmatrix} I & 0 \end{bmatrix} e^{t\left[\begin{smallmatrix} A & I \\ 0 & 0 \end{smallmatrix}\right]} \begin{bmatrix} 0 \\ I \end{bmatrix}.$$

(Remark: The result can be extended to block-$k \times k$ matrices. See [567]. For an application, see [445].)

Fact 11.11.2. Let $A, B \in \mathbb{F}^{n\times n}$, assume that $e^Ae^B = e^Be^A$, and assume that either A and B are Hermitian or all of the entries of A and B are algebraic numbers (roots of polynomials with rational coefficients). Then, $AB = BA$. (Proof. See [261, pp. 88, 89, 270–272] and [594].) (Remark: The matrices $A \triangleq \left[\begin{smallmatrix} 0 & 1 \\ 0 & 2\pi\jmath \end{smallmatrix}\right]$ and $B \triangleq \left[\begin{smallmatrix} 2\pi\jmath & 0 \\ 0 & -2\pi\jmath \end{smallmatrix}\right]$ do not commute but satisfy $e^A = e^B = e^{A+B} = I$.)

Fact 11.11.3. Let $A, B \in \mathbb{R}^{n\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{A+tB} = \int_0^1 e^{\tau(A+tB)}Be^{(1-\tau)(A+tB)}\,\mathrm{d}\tau.$$

Hence,

$$\mathrm{Dexp}\big(e^{tA};B\big) = \frac{\mathrm{d}}{\mathrm{d}t}e^{A+tB}\bigg|_{t=0} = \int_0^1 e^{\tau A}Be^{(1-\tau)A}\,\mathrm{d}\tau.$$

Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr} e^{A+tB} = \operatorname{tr}\big(e^{A+tB}B\big).$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr} e^{A+tB}\bigg|_{t=0} = \operatorname{tr}\big(e^{A}B\big).$$

(Proof: See [74, p. 175] and [358, 404, 433].)

Fact 11.11.4. Let $A, B \in \mathbb{R}^{n\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{A+tB}\bigg|_{t=0} = \sum_{k=0}^{\infty} \tfrac{1}{(k+1)!}\mathrm{ad}_A^k(B)e^A.$$

(Proof: See [45, p. 49].) (Remark: See Fact 2.14.5.)

Fact 11.11.5. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $e^A = e^B$. Then, the following statements hold:

i) If $\lambda - \mu \neq 2k\pi\jmath$ for all $\lambda \in \operatorname{spec}(A)$, $\mu \in \operatorname{spec}(B)$, and $k \in \mathbb{Z}$, then $[A, B] = 0$.

ii) If A is normal and $\sigma_{\max}(A) < \pi$, then $[A, B] = 0$.

iii) If A is normal and $\sigma_{\max}(A) = \pi$, then $[A^2, B] = 0$.

(Proof: See [499].) (Remark: If $[A, B] = 0$, then $[A^2, B] = 0$.)

Fact 11.11.6. Let $A, B \in \mathbb{F}^{n\times n}$ be skew Hermitian. Then, $e^{tA}e^{tB}$ is unitary and there exists a skew-Hermitian matrix $C(t)$ such that $e^{tA}e^{tB} = e^{C(t)}$. (Problem: Does (11.4.2) converge in this case? See [190].)

Fact 11.11.7. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then,

$$\lim_{p\to 0}\left(e^{\frac{p}{2}A}e^{pB}e^{\frac{p}{2}A}\right)^{1/p} = e^{A+B}.$$

(Proof: See [26].)

Fact 11.11.8. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then

$$\lim_{p\to\infty}\left[\tfrac{1}{2}\left(e^{pA} + e^{pB}\right)\right]^{1/p} = e^{\frac{1}{2}(A+B)}.$$

(Proof: See [90].)

Fact 11.11.9. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian, let $q, p > 0$, where $q \le p$, and let $\|\cdot\|$ be a unitarily invariant norm on $\mathbb{F}^{n\times n}$. Then,

$$\left\|\left(e^{\frac{q}{2}A}e^{qB}e^{\frac{q}{2}A}\right)^{1/q}\right\| \le \left\|\left(e^{\frac{p}{2}A}e^{pB}e^{\frac{p}{2}A}\right)^{1/p}\right\|.$$

(Proof: See [26].)

Fact 11.11.10. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\lim_{k\to\infty}\left(e^{\frac{1}{k}A}e^{\frac{1}{k}B}e^{-\frac{1}{k}A}e^{-\frac{1}{k}B}\right)^{k^2} = e^{[A,B]}.$$

Fact 11.11.11. Let $A \in \mathbb{F}^{n\times m}$, $X \in \mathbb{F}^{m\times l}$, and $B \in \mathbb{F}^{l\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}X}\operatorname{tr} e^{AXB} = Be^{AXB}A.$$

Fact 11.11.12. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tA}e^{tB}e^{-tA}e^{-tB}\bigg|_{t=0} = 0$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{\sqrt{t}A}e^{\sqrt{t}B}e^{-\sqrt{t}A}e^{-\sqrt{t}B}\bigg|_{t=0} = AB - BA.$$

Fact 11.11.13. Let $A, B, C \in \mathbb{F}^{n\times n}$, assume that there exists $\beta \in \mathbb{F}$ such that $[A,B] = \beta B + C$, and assume that $[A,C] = [B,C] = 0$. Then,

$$e^{A+B} = e^{A}e^{\phi(\beta)B}e^{\psi(\beta)C},$$

where

$$\phi(\beta) \triangleq \begin{cases} \frac{1}{\beta}\left(1 - e^{-\beta}\right), & \beta \neq 0, \\ 1, & \beta = 0, \end{cases}$$

and

$$\psi(\beta) \triangleq \begin{cases} \frac{1}{\beta^2}\left(1 - \beta - e^{-\beta}\right), & \beta \neq 0, \\ -\frac{1}{2}, & \beta = 0. \end{cases}$$

(Proof: See [228, 540].)

Fact 11.11.14. Let $A, B \in \mathbb{F}^{n\times n}$ and assume there exist $\alpha, \beta \in \mathbb{F}$ such that $[A,B] = \alpha A + \beta B$. Then,

$$e^{t(A+B)} = e^{\phi(t)A}e^{\psi(t)B},$$

where

$$\phi(t) \triangleq \begin{cases} t, & \alpha = \beta = 0, \\ \alpha^{-1}\log(1+\alpha t), & \alpha = \beta \neq 0,\ 1 + \alpha t > 0, \\ \int_0^t \frac{\alpha - \beta}{\alpha e^{(\alpha-\beta)\tau} - \beta}\,\mathrm{d}\tau, & \alpha \neq \beta, \end{cases}$$

and

$$\psi(t) \triangleq \int_0^t e^{-\beta\phi(\tau)}\,\mathrm{d}\tau.$$

(Proof: See [541].)

Fact 11.11.15. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that there exists nonzero $\beta \in \mathbb{F}$ such that $[A,B] = \alpha B$. Then, for all $t > 0$,

$$e^{t(A+B)} = e^{tA}e^{\frac{1-e^{-\alpha t}}{\alpha}B}.$$

(Proof: Apply Fact 11.11.13 with $[tA, tB] = \alpha t(tB)$ and $\beta = \alpha t$.)

Fact 11.11.16. Let $A, B \in \mathbb{F}^{n\times n}$ and assume that $[[A,B],A] = 0$ and $[[A,B],B] = 0$. Then,

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]} = e^{A+B} e^{\frac{1}{2}[A,B]}$$

and

$$e^B e^{2A} e^B = e^{2A+2B}.$$

(Proof: See [600].)

Fact 11.11.17. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that $[A,B] = B^2$. Then,

$$e^{A+B} = e^A(I+B).$$

Fact 11.11.18. Let $A, B \in \mathbb{F}^{n\times n}$. Then, for all $t \in [0,\infty)$,

$$e^{t(A+B)} = e^{tA}e^{tB} + \sum_{k=2}^{\infty} C_k t^k,$$

where, for all $k \in \mathbb{N}$,

$$C_{k+1} \triangleq \tfrac{1}{k+1}([A+B]C_k + [B, D_k]), \qquad C_0 \triangleq 0,$$
$$D_{k+1} \triangleq \tfrac{1}{k+1}(AD_k + D_k B), \qquad D_0 \triangleq I.$$

(Proof: See [481].)

Fact 11.11.19. Let $A \in \mathbb{F}^{n\times n}$ be positive definite and let $B \in \mathbb{F}^{n\times n}$ be nonnegative semidefinite. Then,

$$A + B \le A^{1/2} e^{A^{-1/2}BA^{-1/2}} A^{1/2}.$$

Hence,

$$\frac{\det(A+B)}{\det A} \le e^{\operatorname{tr} A^{-1}B}.$$

Furthermore, for each inequality, equality holds if and only if $B = 0$. (Proof: For nonnegative semi-definite A it follows that $e^A \le I + A$.)

Fact 11.11.20. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then,

$$I \circ (A+B) \le \log\left(e^A \circ e^B\right).$$

(Proof: See [23, 625].) (Remark: See Fact 8.15.21.)

Fact 11.11.21. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then,

$$\left(\operatorname{tr} e^A\right) e^{\operatorname{tr}(e^A B)/\operatorname{tr} e^A} \le \operatorname{tr} e^{A+B}.$$

(Proof: See [69].) (Remark: This inequality is equivalent to the thermodynamic inequality. See Fact 11.11.22.)

Fact 11.11.22. Let $A, B \in \mathbb{F}^{n\times n}$, and assume that A is positive definite, $\operatorname{tr} A = 1$, and B is Hermitian. Then,

$$\operatorname{tr} AB \le \operatorname{tr}(A \log A) + \log \operatorname{tr} e^B.$$

Furthermore, equality holds if and only if

$$A = \left(\operatorname{tr} e^B\right)^{-1} e^B.$$

(Proof: See [69].) (Remark: This result is the *thermodynamic inequality*. Equivalent forms are given by Fact 8.12.19 and Fact 11.11.21.)

Fact 11.11.23. Let $A, B \in \mathbb{F}^{n\times n}$ be skew Hermitian. Then, there exist unitary matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that

$$e^A e^B = e^{S_1 A S_1^{-1} + S_2 B S_2^{-1}}.$$

(Proof: See [515, 547].)

Fact 11.11.24. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, there exist unitary matrices $S_1, S_2 \in \mathbb{F}^{n\times n}$ such that

$$e^{\frac{1}{2}A} e^B e^{\frac{1}{2}A} = e^{S_1 A S_1^{-1} + S_2 B S_2^{-1}}.$$

(Proof: See [514, 515, 547].) (Problem: Determine the relationship between this result and Fact 11.11.23.)

Fact 11.11.25. Let $B \in \mathbb{F}^{n\times n}$ be Hermitian. Then, $\phi\colon\ \mathbf{P}^n \to [0,\infty)$ defined by

$$\phi(A) \triangleq -\operatorname{tr} e^{B + \log A}$$

is convex. (Proof: See [372, 381].)

Fact 11.11.26. Let $A, B, C \in \mathbb{F}^{n\times n}$ be positive definite. Then,

$$\operatorname{tr} e^{\log A - \log B + \log C} \le \operatorname{tr} \int_0^\infty A(B + xI)^{-1} C (B + xI)^{-1}\, \mathrm{d}x.$$

(Proof: See [372, 381].) (Remark: $-\log B$ is correct.) (Remark: $\operatorname{tr} e^{A+B+C} \le |\operatorname{tr} e^A e^B e^C|$ is not generally true.)

Fact 11.11.27. Let $A \in \mathbb{F}^{n\times n}$ and $B \in \mathbb{F}^{m\times m}$. Then,

$$\operatorname{tr} e^{A \oplus B} = \left(\operatorname{tr} e^A\right)\left(\operatorname{tr} e^B\right).$$

Fact 11.11.28. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, and $C \in \mathbb{F}^{l\times l}$. Then,

$$e^{A \oplus B \oplus C} = e^A \otimes e^B \otimes e^C.$$

Fact 11.11.29. Let $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times m}$, $C \in \mathbb{F}^{k\times k}$, and $D \in \mathbb{F}^{l\times l}$. Then,

$$\operatorname{tr} e^{A\otimes I\otimes B\otimes I + I\otimes C\otimes I\otimes C} = \operatorname{tr} e^{A\otimes B}\operatorname{tr} e^{C\otimes D}.$$

(Proof: By Fact 7.4.29, a similarity transformation involving the Kronecker permutation matrix can be used to reorder the inner two terms. See [519].)

11.12 Facts on Eigenvalues, Singular Values, and Norms

Fact 11.12.1. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_{\max}\big(e^{At}\big)\bigg|_{t=0^+} = \tfrac{1}{2}\lambda_{\max}(A + A^*).$$

Hence, $\sigma_{\max}\big(e^{tA}\big)$ is decreasing for all sufficiently small $t > 0$ if and only if A is dissipative. (Proof: See [585].)

Fact 11.12.2. Let $A \in \mathbb{R}^{n\times n}$. Then, for all $t \ge 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\|e^{tA}\|_{\mathrm{F}}^2 = \operatorname{tr}\, e^{tA}(A + A^*)e^{tA^*}.$$

Hence, if A is dissipative, then $\|e^{tA}\|_{\mathrm{F}}$ is decreasing for all $t > 0$. (Proof: See [585].)

Fact 11.12.3. Let $A \in \mathbb{F}^{n\times n}$. Then,

$$\big|\operatorname{tr} e^{2A}\big| \le \operatorname{tr} e^{A}e^{A^*} \le \operatorname{tr} e^{A+A^*} \le \Big[n \operatorname{tr} e^{2(A+A^*)}\Big]^{1/2} \le \tfrac{n}{2} + \tfrac{1}{2}\operatorname{tr} e^{2(A+A^*)}.$$

In addition, $\operatorname{tr} e^{A}e^{A^*} = \operatorname{tr} e^{A+A^*}$ if and only if A is normal. (Proof: See [83], [289, p. 515], and [513].) (Remark: $\operatorname{tr} e^{A}e^{A^*} \le \operatorname{tr} e^{A+A^*}$ is *Bernstein's inequality*. See [24].)

Fact 11.12.4. Let $A \in \mathbb{F}^{n\times n}$. Then, for all $k = 1, \ldots, n$,

$$\prod_{i=1}^{k}\sigma_i\big(e^{A}\big) \le \prod_{i=1}^{k}\lambda_i\Big(e^{\frac{1}{2}(A+A^*)}\Big) = \prod_{i=1}^{k}e^{\lambda_i\left(\frac{1}{2}(A+A^*)\right)} \le \prod_{i=1}^{k}e^{\sigma_i(A)}.$$

Furthermore, for all $k = 1, \ldots, n$,

$$\sum_{i=1}^{k}\sigma_i\big(e^{A}\big) \le \sum_{i=1}^{k}\lambda_i\Big(e^{\frac{1}{2}(A+A^*)}\Big) = \sum_{i=1}^{k}e^{\lambda_i\left(\frac{1}{2}(A+A^*)\right)} \le \sum_{i=1}^{k}e^{\sigma_i(A)}.$$

In particular,

$$\sigma_{\max}\big(e^{A}\big) \le \lambda_{\max}\Big(e^{\frac{1}{2}(A+A^*)}\Big) = e^{\frac{1}{2}\lambda_{\max}(A+A^*)} \le e^{\sigma_{\max}(A)}$$

or, equivalently,

$$\lambda_{\max}\big(e^{A}e^{A^*}\big) \le \lambda_{\max}\big(e^{A+A^*}\big) = e^{\lambda_{\max}(A+A^*)} \le e^{2\sigma_{\max}(A)}.$$

Furthermore,

$$\big|\det e^{A}\big| = \big|e^{\operatorname{tr} A}\big| \le e^{|\operatorname{tr} A|} \le e^{\operatorname{tr}\langle A\rangle}$$

and

$$\operatorname{tr}\big\langle e^{A}\big\rangle \le \sum_{i=1}^{n} e^{\sigma_i(A)}.$$

(Proof: See [516], Fact 8.14.2, Fact 8.14.3, and Fact 8.16.5.)

Fact 11.12.5. Let $A \in \mathbb{F}^{n\times n}$, and let $\|\cdot\|$ be a unitarily invariant norm. Then,

$$\big\|e^{A}e^{A^*}\big\| \le \big\|e^{A+A^*}\big\|.$$

In particular,

$$\lambda_{\max}\big(e^{A}e^{A^*}\big) \le \lambda_{\max}\big(e^{A+A^*}\big)$$

and

$$\operatorname{tr} e^{A}e^{A^*} \le \operatorname{tr} e^{A+A^*}.$$

(Proof: See [150].)

Fact 11.12.6. Let $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}\big|\operatorname{tr} e^{A+B}\big| &\le \operatorname{tr} e^{\frac{1}{2}(A+B)}e^{\frac{1}{2}(A+B)^*} \le \operatorname{tr} e^{\frac{1}{2}(A+A^*+B+B^*)} \le \operatorname{tr} e^{\frac{1}{2}(A+A^*)}e^{\frac{1}{2}(B+B^*)}\\ &\le \big(\operatorname{tr} e^{A+A^*}\big)^{1/2}\big(\operatorname{tr} e^{B+B^*}\big)^{1/2} \le \tfrac{1}{2}\operatorname{tr}\big(e^{A+A^*}+e^{B+B^*}\big)\end{aligned}$$

and

$$\left.\begin{array}{c}\operatorname{tr} e^{A}e^{B}\\ \tfrac{1}{2}\operatorname{tr}\big(e^{2A}+e^{2B}\big)\end{array}\right\} \le \tfrac{1}{2}\operatorname{tr}\big(e^{A}e^{A^*}+e^{B}e^{B^*}\big) \le \tfrac{1}{2}\operatorname{tr}\big(e^{A+A^*}+e^{B+B^*}\big).$$

(Proof: See [83, 151, 454] and [289, p. 514].)

Fact 11.12.7. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian. If $\|\cdot\|$ is a unitarily invariant norm on $\mathbb{F}^{n\times n}$, then

$$\big\|e^{A+B}\big\| \le \Big\|e^{\frac{1}{2}A}e^{B}e^{\frac{1}{2}A}\Big\| \le \big\|e^{A}e^{B}\big\|.$$

Furthermore, for all $k = 1, \ldots, n$,

$$\prod_{i=1}^{k}\lambda_i\big(e^{A+B}\big) \le \prod_{i=1}^{k}\lambda_i\big(e^{A}e^{B}\big) \le \prod_{i=1}^{k}\sigma_i\big(e^{A}e^{B}\big)$$

with equality for $k = n$, that is,

$$\prod_{i=1}^{n}\lambda_i\big(e^{A+B}\big) = \prod_{i=1}^{n}\lambda_i\big(e^{A}e^{B}\big) = \prod_{i=1}^{n}\sigma_i\big(e^{A}e^{B}\big) = \det\big(e^{A}e^{B}\big).$$

Furthermore, for all $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} \lambda_i\big(e^{A+B}\big) \le \sum_{i=1}^{k} \lambda_i\big(e^{A}e^{B}\big) \le \sum_{i=1}^{k} \sigma_i\big(e^{A}e^{B}\big).$$

In particular,

$$\lambda_{\max}\big(e^{A+B}\big) \le \lambda_{\max}\big(e^{A}e^{B}\big) \le \sigma_{\max}\big(e^{A}e^{B}\big)$$

and

$$\operatorname{tr} e^{A+B} \le \operatorname{tr} e^{A}e^{B} \le \operatorname{tr} \big\langle e^{A}e^{B} \big\rangle .$$

(Proof: See [26], Fact 5.9.13, and Fact 8.16.5.) (Remark: $\operatorname{tr} e^{A+B} \le \operatorname{tr} e^{A}e^{B}$ is the *Golden-Thompson inequality.*)

Fact 11.12.8. Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and let $\|\cdot\|$ be a unitarily invariant norm. Then,

$$\big\|e^{A+B}\big\| \le \Big\|e^{\frac{1}{2}A}e^{B}e^{\frac{1}{2}A}\Big\| \le \big\|e^{A}e^{B}\big\| .$$

(Remark: The left-hand inequality is *Segal's inequality.* See [24].)

Fact 11.12.9. Let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$, and let $A, B \in \mathbb{F}^{n\times n}$. Then, for all $t \ge 0$,

$$\big\|e^{tA} - e^{tB}\big\| \le e^{\|A\|t}\Big(e^{\|A-B\|t} - 1\Big).$$

Fact 11.12.10. Let $A, B \in \mathbb{R}^{n\times n}$, and assume that A is normal. Then, for all $t \ge 0$,

$$\sigma_{\max}\big(e^{tA} - e^{tB}\big) \le \sigma_{\max}\big(e^{tA}\big)\Big[e^{\sigma_{\max}(A-B)t} - 1\Big].$$

(Proof: See [594].)

Fact 11.12.11. Let $A \in \mathbb{F}^{n\times n}$, and define $f_i\colon\ \mathbb{R} \mapsto \mathbb{R}$ by $f_i(t) \triangleq \log \sigma_i\big(e^{tA}\big)$. Then, A is normal if and only if, for all $i = 1, \ldots, n$, f_i is convex. (Proof: See [43].)

11.13 Facts on Stable Polynomials

Fact 11.13.1. Let $p \in \mathbb{R}[s]$ be asymptotically stable and let $p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$. Then, $\beta_i > 0$ for all $i = 0, \ldots, n-1$.

Fact 11.13.2. Let $p \in \mathbb{R}[s]$, where $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_0$. If p is asymptotically stable, then $a_0, \ldots, a_{n-1}$ are positive. Now, assume that $a_0, \ldots, a_{n-1}$ are positive. Then, the following statements hold:

i) If $n = 1$ or $n = 2$, then p is asymptotically stable.

ii) If $n = 3$, then p is asymptotically stable if and only if

$$a_0 < a_1 a_2.$$

iii) If $n = 4$, then p is asymptotically stable if and only if

$$a_1^2 + a_0 a_3^2 < a_1 a_2 a_3.$$

iv) If $n = 5$, then p is asymptotically stable if and only if

$$a_2 < a_3 a_4,$$
$$a_2^2 + a_1 a_4^2 < a_0 a_4 + a_2 a_3 a_4,$$
$$a_0^2 + a_1 a_2^2 + a_1^2 a_4^2 + a_0 a_3^2 a_4 < a_0 a_2 a_3 + 2 a_0 a_1 a_4 + a_1 a_2 a_3 a_4.$$

(Remark: These results are special cases of the *Routh criterion*, which provides stability criteria for polynomials of arbitrary degree n. See [135].)

Fact 11.13.3. Let $p \in \mathbb{R}[s]$ be monic and define $q(s) \triangleq s^n p(1/s)$, where $n \triangleq \deg p$. Then, p is asymptotically stable if and only if q is asymptotically stable. (Remark: See Fact 4.8.1 and Fact 11.13.4.)

Fact 11.13.4. Let $p \in \mathbb{R}[s]$ be monic and assume that p is semistable. Then, $q(s) \triangleq p(s)/s$ and $\hat{q}(s) \triangleq s^n p(1/s)$ are asymptotically stable. (Remark: See Fact 4.8.1 and Fact 11.13.3.)

Fact 11.13.5. Let $p \in \mathbb{R}[s]$ be asymptotically stable and let $p(s) = \beta_n s^n + \beta_{n-1} s^{n-1} + \cdots + \beta_1 s + \beta_0$, where $\beta_n > 0$. Then, for all $i = 1, \ldots, n-2$,

$$\beta_{i-1}\beta_{i+2} < \beta_i \beta_{i+1}.$$

(Remark: This result is a necessary condition for asymptotic stability, which can be used to show that a given polynomial with positive coefficients is unstable.) (Remark: This result is due to Xie. See [621].)

Fact 11.13.6. Let $n \in \mathbb{P}$ be even, let $m \triangleq n/2$, let $p \in \mathbb{R}[s]$, where $p(s) = \beta_n s^n + \beta_{n-1} s^{n-1} + \cdots + \beta_1 s + \beta_0$ and $\beta_n > 0$, and assume that p is asymptotically stable. Then, for all $i = 1, \ldots, m-1$,

$$\binom{m}{i} \beta_0^{(m-i)/m} \beta_n^{i/m} \le \beta_{2i}.$$

(Remark: This result is a necessary condition for asymptotic stability, which can be used to show that a given polynomial with positive coefficients is unstable.) (Remark: This result is due to Borobia and Dormido. See [621] for extensions to polynomials of odd degree.)

Fact 11.13.7. Let $p, q \in \mathbb{R}[s]$, where $p(s) = \alpha_n s^n + \alpha_{n-1} s^{n-1} + \cdots +$

$\alpha_1 s+\alpha_0$ and $q(s) = \beta_m s^m+\beta_{m-1}s^{m-1}+\cdots+\beta_1 s+\beta_0$. If p and q are (Lyapunov, asymptotically) stable, then $r(s) \triangleq \alpha_l\beta_l s^l+\alpha_{l-1}\beta_{l-1}s^{l-1}+\cdots+\alpha_1\beta_1 s+\alpha_0\beta_0$, where $l \triangleq \min\{m,n\}$, is (Lyapunov, asymptotically) stable. (Proof: See [224].) (Remark: The polynomial r is the *Schur product* of p and q. See [39, 311].)

Fact 11.13.8. Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable over $\mathbb{R}$. Then, χ_A has all positive coefficients if and only if χ_A (equivalently, A) is asymptotically stable. (Proof: Sufficiency follows from Fact 11.13.1. For necessity, note that χ_A has only real roots and that $\chi_A(\lambda) > 0$ for all $\lambda \geq 0$. Hence, $\operatorname{roots}(\chi_A) \subset (-\infty, 0)$.)

Fact 11.13.9. Let $A \in \mathbb{R}^{n\times n}$. Then, $\chi_{A\oplus A}$ has all positive coefficients if and only if $\chi_{A\oplus A}$ (equivalently, A) is asymptotically stable. (Proof: If A is not asymptotically stable, then Fact 11.14.28 implies that $A \oplus A$ has a positive eigenvalue λ. Since $\chi_{A\oplus A}(\lambda) = 0$, it follows that $\chi_{A\oplus A}$ cannot have all positive coefficients. See [217, Theorem 5].)

11.14 Facts on Stable Matrices

Fact 11.14.1. Let $A \in \mathbb{F}^{n\times n}$ be semistable. Then, A is Lyapunov stable.

Fact 11.14.2. Let $A \in \mathbb{F}^{n\times n}$ be Lyapunov stable. Then, A is group invertible.

Fact 11.14.3. Let $A \in \mathbb{F}^{n\times n}$ be semistable. Then, A is group invertible.

Fact 11.14.4. Let $A \in \mathbb{F}^{n\times n}$ be semistable. Then,

$$\lim_{t\to\infty} e^{tA} = I - AA^{\#}$$

and thus

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t e^{\tau A}\,\mathrm{d}\tau = I - AA^{\#}.$$

(Remark: See Fact 11.14.1, Fact 11.14.2, and Fact 10.8.3.)

Fact 11.14.5. Let $A \in \mathbb{R}^{n\times n}$ be Lyapunov stable. Then,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t e^{\tau A}\,\mathrm{d}\tau = I - AA^{\#}.$$

(Remark: See Fact 11.14.2.)

Fact 11.14.6. Let $A, B \in \mathbb{F}^{n\times n}$. Then, $\lim_{\alpha\to\infty} e^{A+\alpha B}$ exists if and only if B is semistable. In this case,

$$\lim_{\alpha\to\infty} e^{A+\alpha B} = e^{(I-BB^{\#})A}\Big(I - BB^{\#}\Big) = \Big(I - BB^{\#}\Big)e^{A(I-BB^{\#})}.$$

(Proof: See [125].)

Fact 11.14.7. Let $A \in \mathbb{R}^{n\times n}$. Then, e^{tA} is nonnegative for all $t \ge 0$ if and only if

$$A_{(i,j)} \ge 0$$

for all $i, j = 1, \ldots, n$ such that $i \ne j$. In this case, A is asymptotically stable if and only if, for all $i = 1, \ldots, n$, the sign of the ith leading principal subdeterminant of A is $(-1)^i$. (Proof: See [88] and [223, p. 74].) (Remark: A is *essentially nonnegative*.)

Fact 11.14.8. Let $\|\cdot\|$ be a submultiplicative norm on $\mathbb{F}^{n\times n}$, let $A \in \mathbb{F}^{n\times n}$ be asymptotically stable, and let $\beta > \operatorname{spabs}(A)$. Then, there exists $\gamma > 0$ such that, for all $t \ge 0$,

$$\left\|e^{tA}\right\| \le \gamma e^{\beta t}.$$

(Remark: See [229, pp. 201–206] and [320].)

Fact 11.14.9. let $A \in \mathbb{F}^{n\times n}$ be asymptotically stable, let $R \in \mathbb{F}^{n\times n}$ be positive definite, and let $P \in \mathbb{F}^{n\times n}$ be the positive-definite solution of $A^*P + PA + R = 0$. Then,

$$\sigma_{\max}\big(e^{tA}\big) \le \sqrt{\frac{\sigma_{\max}(P)}{\sigma_{\min}(P)}}\, e^{-t\lambda_{\min}(RP^{-1})/2}$$

and

$$\|e^{tA}\|_{\mathrm{F}} \le \sqrt{\|P\|_{\mathrm{F}}\|P^{-1}\|_{\mathrm{F}}}\, e^{-t\lambda_{\min}(RP^{-1})/2}.$$

If, in addition, $A + A^*$ is negative definite, then

$$\|e^{tA}\|_{\mathrm{F}} \le e^{-t\lambda_{\min}(-A-A^*)/2}.$$

(Proof: See [390].)

Fact 11.14.10. let $A \in \mathbb{R}^{n\times n}$ be asymptotically stable, let $R \in \mathbb{R}^{n\times n}$ be positive definite, and let $P \in \mathbb{R}^{n\times n}$ be the positive-definite solution of $A^{\mathrm{T}}P + PA + R = 0$. Furthermore, define the vector norm $\|x\|' \triangleq \sqrt{x^{\mathrm{T}}Px}$ on $\mathbb{R}^n$, let $\|\cdot\|$ denote the induced norm on $\mathbb{R}^{n\times n}$, and let $\mu(\cdot)$ denote the corresponding logarithmic norm. Then,

$$\mu(A) = -\lambda_{\min}\big(RP^{-1}\big)/2.$$

Consequently,

$$\|e^{tA}\| \le e^{-t\lambda_{\min}(RP^{-1})/2}.$$

(Proof: See [300] and use *xiii*) of Fact 9.10.8.) (Remark: See Fact 9.10.8 for the definition and properties of the logarithmic derivative.)

Fact 11.14.11. Let $A \in \mathbb{F}^{n\times n}$. Then, A is similar to a skew-Hermitian matrix if and only if there exists a positive-definite matrix $P \in \mathbb{F}^{n\times n}$ such that $A^*P + PA = 0$.

Fact 11.14.12. Let $A \in \mathbb{R}^{n\times n}$. Then, A and A^2 are asymptotically stable if and only if, for all $\lambda = re^{\jmath\theta} \in \operatorname{spec}(A)$, where $\theta \in [0, 2\pi]$, it follows that $\theta \in \left(\frac{\pi}{2}, \frac{3\pi}{4}\right) \cup \left(\frac{5\pi}{4}, \frac{3\pi}{2}\right)$.

Fact 11.14.13. Let $A \in \mathbb{R}^{n\times n}$. Then, A is group invertible and $2\pi k\jmath \notin \operatorname{spec}(A)$ for all $k \in \mathbb{P}$ if and only if

$$AA^{\#} = \left(e^A - I\right)\left(e^A - I\right)^{\#}.$$

In particular, if A is semistable, then this identity holds. (Proof: Use *ii*) of Fact 11.15.16 and *ix*) of Proposition 11.6.2.)

Fact 11.14.14. Let $A \in \mathbb{F}^{n\times n}$. Then, A is asymptotically stable if and only if A^{-1} is asymptotically stable. Hence, $e^{tA} \to 0$ as $t \to \infty$ if and only if $e^{tA^{-1}} \to 0$ as $t \to \infty$.

Fact 11.14.15. Let $A, B \in \mathbb{R}^{n\times n}$, assume A is asymptotically stable, and assume that $\sigma_{\max}(B \oplus B) < \sigma_{\min}(A \oplus A)$. Then, $A+B$ is asymptotically stable. (Proof: Since $A \oplus A$ is nonsingular, Fact 9.12.12 implies that $A \oplus A + \alpha(B \oplus B) = (A + \alpha B) \oplus (A + \alpha B)$ is nonsingular for all $0 \le \alpha \le 1$. Now, suppose that $A + B$ is not asymptotically stable. Then, there exists $\alpha_0 \in (0, 1]$ such that $A + \alpha_0 B$ has an imaginary eigenvalue, and thus $(A + \alpha_0 B) \oplus (A + \alpha_0 B) = A \oplus A + \alpha_0(B \oplus B)$ is singular, which is a contradiction.) (Remark: This result provides a suboptimal solution to a nearness problem. See [278, Section 7] and Fact 9.12.12.)

Fact 11.14.16. Let $A \in \mathbb{C}^{n\times n}$ be asymptotically, let $\|\cdot\|$ denote either $\sigma_{\max}(\cdot)$ or $\|\cdot\|_{\mathrm{F}}$, and define

$$\beta(A) \triangleq \{\|B\|\colon\ B \in \mathbb{C}^{n\times n} \text{ and } A + B \text{ is not asymptotically stable}\}.$$

Then,

$$\begin{aligned}\tfrac{1}{2}\sigma_{\min}(A \otimes A) \le \beta(A) &= \min_{\gamma\in\mathbb{R}} \sigma_{\min}(A + \gamma\jmath I)\\ &\le \min\{\operatorname{spabs}(A), \sigma_{\min}(A), \tfrac{1}{2}\sigma_{\max}(A + A^*)\}.\end{aligned}$$

Furthermore, let $R \in \mathbb{F}^{n\times n}$ be positive definite, and let $P \in \mathbb{F}^{n\times n}$ be the positive-definite solution of $A^*P + PA + R = 0$. Then,

$$\tfrac{1}{2}\sigma_{\min}(R)/\|P\| \le \beta(A).$$

If, in addition, $A + A^*$ is negative definite, then

$$-\tfrac{1}{2}\lambda_{\min}(A + A^*) \le \beta(A).$$

(Proof: See [278, 568].) (Remark: The analogous problem for real matrices and real perturbations is discussed in [471].)

Fact 11.14.17. Let $A \in \mathbb{F}^{n\times n}$ be asymptotically stable, let $V \in \mathbb{F}^{n\times n}$ be positive definite, and let $Q \in \mathbf{P}^n$ satisfy $AQ + QA^* + V = 0$. Then, for all $t \ge 0$,

$$e^{tA}e^{tA^*} \le \kappa(Q)\,\mathrm{tr}\,e^{-tS^{-1}VS^{-*}} \le \kappa(Q)e^{-(t/\sigma_{\max}(Q))V},$$

where $S \in \mathbb{F}^{n\times n}$ satisfies $Q = SS^*$ and $\kappa(Q) \triangleq \sigma_{\max}(Q)/\sigma_{\min}(Q)$. (Proof: See [620].) (Remark: Fact 11.12.3 yields $e^{tA}e^{tA^*} \le e^{t(A+A^*)}$. However, $A+A^*$ may not be asymptotically stable. See [84].)

Fact 11.14.18. Let $A \in \mathbb{R}^{n\times n}$, and assume that every entry of $A \in \mathbb{R}^{n\times n}$ is positive. Then, A is unstable. (Proof: See Fact 4.11.1.)

Fact 11.14.19. Let $A \in \mathbb{R}^{n\times n}$. Then, A is asymptotically stable if and only if there exist $B, C \in \mathbb{R}^{n\times n}$ such that B is positive definite, C is dissipative, and $A = BC$. (Proof: $A = P^{-1}\big(-A^{\mathrm{T}}P - R\big)$.) (Remark: To reverse the order of factors, consider A^{T}.)

Fact 11.14.20. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements hold:

i) All of the real eigenvalues of A are positive if and only if A is the product of two dissipative matrices.

ii) A is nonsingular and $A \ne \alpha I$ for all $\alpha < 0$ if and only if A is the product of two asymptotically stable matrices.

iii) A is nonsingular if and only if A is the product of three or fewer asymptotically stable matrices.

(Proof: See [56, 618].)

Fact 11.14.21. Let $p \in \mathbb{R}[s]$, where $p(s) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0$

and $\beta_0,\ldots,\beta_n > 0$. Furthermore, define $A \in \mathbb{R}^{n\times n}$ by

$$A \triangleq \begin{bmatrix} \beta_{n-1} & \beta_{n-3} & \beta_{n-5} & \beta_{n-7} & \cdots & \cdots & 0 \\ 1 & \beta_{n-2} & \beta_{n-4} & \beta_{n-6} & \cdots & \cdots & 0 \\ 0 & \beta_{n-1} & \beta_{n-3} & \beta_{n-5} & \cdots & \cdots & 0 \\ 0 & 1 & \beta_{n-2} & \beta_{n-4} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \beta_1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & \beta_2 & \beta_0 \end{bmatrix}.$$

If p is Lyapunov stable, then every subdeterminant of A is nonnegative. (Remark: A is *totally nonnegative*.) Furthermore, p is asymptotically stable if and only if every leading principal subdeterminant of A is positive. (Proof: See [39].) (Remark: The second statement is due to Hurwitz.) (Remark: The diagonal entries of A are $\beta_{n-1},\ldots,\beta_0$.) (Problem: Show that this condition for stability is equivalent to the condition given in [202, p. 183] in terms of an alternative matrix $\hat{A}$.)

Fact 11.14.22. Let $A \in \mathbb{R}^{n\times n}$ be tridiagonal and assume that $A_{(i,i)} > 0$ for all $i = 1,\ldots,n$ and $A_{(i,i+1)}A_{(i+1,i)} > 0$ for all $i = 1,\ldots,n-1$. Then, A is asymptotically stable. (Proof: See [127].) (Remark: This result is due to Barnett and Storey.)

Fact 11.14.23. Let $A \in \mathbb{R}^{n\times n}$ be cyclic. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $A_{\mathrm{S}} = SAS^{-1}$ is given by the tridiagonal matrix

$$A_{\mathrm{S}} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -\alpha_n & 0 & 1 & \cdots & 0 & 0 & \\ 0 & -\alpha_{n-1} & 0 & \cdots & 0 & 0 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & 1 & \\ 0 & 0 & 0 & \cdots & -\alpha_2 & -\alpha_1 & \end{bmatrix},$$

where $\alpha_1,\ldots,\alpha_n$ are real numbers. If $\alpha_1\alpha_2\cdots\alpha_n \neq 0$, then the number of eigenvalues of A in the OLHP is equal to the number of positive elements in $\{\alpha_1, \alpha_1\alpha_2, \ldots, \alpha_1\alpha_2\cdots\alpha_n\}_{\mathrm{m}}$. Furthermore, $A_{\mathrm{S}}^{\mathrm{T}}P + PA_{\mathrm{S}} + R = 0$, where

$$P \triangleq \operatorname{diag}(\alpha_1\alpha_2\cdots\alpha_n, \alpha_1\alpha_2\cdots\alpha_{n-1}, \ldots, \alpha_1\alpha_2, \alpha_1)$$

and

$$R \triangleq \operatorname{diag}\big(0,\ldots,0,2\alpha_1^2\big).$$

(Remark: A_{S} is in *Schwarz form*.) (Proof: See [66, pp. 52, 95].)

Fact 11.14.24. Let $\alpha_1,\alpha_2,\alpha_3 > 0$, and define $A, P, R \in \mathbb{R}^{3\times 3}$ by the tridiagonal matrix

$$A_{\mathrm{R}} \triangleq \begin{bmatrix} -\alpha_1 & \alpha_2^{1/2} & 0 \\ -\alpha_2^{1/2} & 0 & \alpha_3^{1/2} \\ 0 & -\alpha_3^{1/2} & 0 \end{bmatrix}$$

and the diagonal matrices

$$P \triangleq I, \quad R \triangleq \operatorname{diag}(2\alpha_1, 0, 0).$$

Then, $A_{\mathrm{R}}^{\mathrm{T}}P + PA_{\mathrm{R}} + R = 0$. (Remark: The matrix A_{R} is in *Routh form.* The Routh form A_{R} and the Schwarz form A_{S} are related by $A_{\mathrm{R}} = S_{\mathrm{RS}}A_{\mathrm{S}}S_{\mathrm{RS}}^{-1}$, where

$$S_{\mathrm{RS}} \triangleq \begin{bmatrix} 0 & 0 & \alpha_1^{1/2} \\ 0 & -(\alpha_1\alpha_2)^{1/2} & 0 \\ (\alpha_1\alpha_2\alpha_3)^{1/2} & 0 & 0 \end{bmatrix}.)$$

Fact 11.14.25. Let $\alpha_1, \alpha_2, \alpha_3 > 0$, and define $A_{\mathrm{C}}, P, R \in \mathbb{R}^{3\times 3}$ by the tridiagonal matrix

$$A_{\mathrm{C}} \triangleq \begin{bmatrix} 0 & 1/a_3 & 0 \\ -1/a_2 & 0 & 1/a_2 \\ 0 & -1/a_1 & -1/a_1 \end{bmatrix}$$

and the diagonal matrices

$$P \triangleq \operatorname{diag}(a_3, a_2, a_1), \quad R \triangleq \operatorname{diag}(0, 0, 2),$$

where $a_1 \triangleq 1/\alpha_1$, $a_2 \triangleq \alpha_1/\alpha_2$, and $a_3 \triangleq \alpha_2/(\alpha_1\alpha_3)$. Then, $A_{\mathrm{C}}^{\mathrm{T}}P + PA_{\mathrm{C}} + R = 0$. (Remark: The matrix A_{C} is in *Chen form.*) The Schwarz form A_{S} and the Chen form A_{C} are related by $A_{\mathrm{S}} = S_{\mathrm{SC}}A_{\mathrm{C}}S_{\mathrm{SC}}^{-1}$, where

$$S_{\mathrm{SC}} \triangleq \begin{bmatrix} 1/(\alpha_1\alpha_3) & 0 & 0 \\ 0 & 1/\alpha_2 & 0 \\ 0 & 0 & 1/\alpha_1 \end{bmatrix}.)$$

(Proof: See [141, p. 346].) (Remark: The Schwarz, Routh, and Chen forms provide the basis for the Routh criterion. See [15, 115, 141, 452].)

Fact 11.14.26. Let $A \in \mathbb{F}^{n\times n}$. Then, the following statements are equivalent:

i) A is asymptotically stable.

ii) There exist a negative-definite matrix $B \in \mathbb{F}^{n\times n}$, a skew-Hermitian matrix $C \in \mathbb{F}^{n\times n}$, and a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that $A = B + SCS^{-1}$.

iii) There exist a negative-definite matrix $B \in \mathbb{F}^{n\times n}$, a skew-Hermitian matrix $C \in \mathbb{F}^{n\times n}$, and a nonsingular matrix $S \in \mathbb{F}^{n\times n}$ such that

$$A = S(B+C)S^{-1}.$$

(Proof: See [160].)

Fact 11.14.27. Let $A \in \mathbb{R}^{n\times n}$, and let $k \geq 2$. Then, there exist asymptotically stable matrices $A_1, \ldots, A_k \in \mathbb{R}^{n\times n}$ such that $A = \sum_{i=1}^k A_i$ if and only if $\operatorname{tr} A < 0$. (Proof: See [308].)

Fact 11.14.28. Let $A \in \mathbb{R}^{n\times n}$. Then, A is (Lyapunov stable, semistable, asymptotically stable) if and only if $A \oplus A$ is. (Proof: Use Fact 7.4.27 and the fact that $\operatorname{vec}\big(e^{tA}Ve^{tA^*}\big) = e^{t(A\oplus\overline{A})}\operatorname{vec} V$.)

Fact 11.14.29. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$. Then, the following statements hold:

i) If A and B are (Lyapunov stable, semistable, asymptotically stable), then $A \oplus B$ is (Lyapunov stable, semistable, asymptotically stable).

ii) If $A \oplus B$ is (Lyapunov stable, semistable, asymptotically stable), then either A or B is (Lyapunov stable, semistable, asymptotically stable).

(Proof: Use Fact 7.4.27.)

Fact 11.14.30. Let $A \in \mathbb{R}^{2\times 2}$. Then, A is asymptotically stable if and only if $\operatorname{tr} A < 0$ and $\det A > 0$.

Fact 11.14.31. Let $A \in \mathbb{C}^{n\times n}$. Then, there exists a unique asymptotically stable matrix $B \in \mathbb{C}^{n\times n}$ such that $B^2 = -A$. (Remark: This result is stated in [526]. The uniqueness of the square root for complex matrices that have no eigenvalues in $(-\infty, 0]$ is implicitly assumed in [527].) (Remark: See Fact 5.13.16.)

Fact 11.14.32. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) If A is semidissipative, then A is Lyapunov stable.

ii) If A is dissipative, then A is asymptotically stable.

iii) If A is Lyapunov stable and normal, then A is semidissipative.

iv) If A is asymptotically stable and normal, then A is dissipative.

v) If A is discrete-time Lyapunov stable and normal, then A is semicontractive.

Fact 11.14.33. Let $A \in \mathbb{R}^{n\times n}$, and assume that $A_{(i,j)} \leq 0$ for all $i,j = 1,\ldots,n$ such that $i \neq j$. (Remark: A is a *Z-matrix*.) Then, the following conditions are equivalent:

i) $-A$ is asymptotically stable.

ii) There exists $B \in \mathbb{R}^{n\times n}$ such that $B \geq\geq 0$, $A = \alpha I - B$, and $\alpha > \operatorname{sprad}(B)$.

iii) If $\lambda \in \operatorname{spec}(A)$ is real, then $\lambda > 0$.

iv) $A + \alpha I$ is nonsingular for all $\alpha \geq 0$.

v) $A + B$ is nonsingular for all nonnegative, diagonal matrices $B \in \mathbb{R}^{n\times n}$.

vi) Every principal subdeterminant of A is positive.

vii) Every leading principal subdeterminant of A is positive.

viii) For all $k \in \{1, \ldots, n\}$, the sum of all $k\times k$ principal subdeterminants of A is positive.

ix) There exists $x \in \mathbb{R}^n$ such that $x >> 0$ and $Ax >> 0$.

x) If $x \in \mathbb{R}^n$ and $Ax \geq\geq 0$, then $x \geq\geq 0$.

xi) A is nonsingular and $A^{-1} \geq\geq 0$.

(Proof: See [81, pp. 134–140] or [289, pp. 114–116].) (Remark: A is an *M-matrix.*)

11.15 Facts on Discrete-Time Stability

Fact 11.15.1. Let $p \in \mathbb{R}[s]$, where $p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_0$. Then, the following statements hold:

i) If $n = 1$, then p is discrete-time asymptotically stable if and only if $|a_0| < 1$.

ii) If $n = 2$, then p is discrete-time asymptotically stable if and only if $|a_0| < 1$ and $|a_1| < 1 + a_0$.

iii) If $n = 3$, then p is discrete-time asymptotically stable if and only if $|a_0| < 1$, $|a_2| < 1 + a_0$, and $|a_2 - a_0a_1| < 1$.

(Remark: These results are special cases of the *Jury test*, which provides stability criteria for polynomials of arbitrary degree n. See [141, 319].)

Fact 11.15.2. Let $A \in \mathbb{R}^{2\times 2}$. Then, A is discrete-time asymptotically stable if and only if $|\operatorname{tr} A| < 1 + \det A$ and $|\det A| < 1$.

Fact 11.15.3. Let $A \in \mathbb{F}^{n\times n}$. Then, A is discrete-time asymptotically stable if and only if A^2 is discrete-time asymptotically stable.

Fact 11.15.4. Let $A \in \mathbb{R}^{n\times n}$. Then, for all $k \geq 0$,

$$A^k = x_1(k)I + x_2(k)A + \cdots + x_n(k)A^{n-1},$$

where, for all $i = 1,\ldots,n$ and for all $k \geq 0$, x_i satisfies

$$x(n+k) + \beta_{n-1}x(n+k-1) + \cdots + c_1x(k+1) + c_0x(k) = 0,$$

with, for all $i,j = 1,\ldots,n$, the initial conditions

$$x_i(j-1) = \delta_{ij}.$$

(Proof: See [346].)

Fact 11.15.5. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) If A is semicontractive, then A is discrete-time Lyapunov stable.

ii) If A is contractive, then A is discrete-time asymptotically stable.

iii) If A is discrete-time Lyapunov stable and normal, then A is semicontractive.

iv) If A is discrete-time asymptotically stable and normal, then A is contractive.

(Problem: Prove these results by using Fact 11.12.5.)

Fact 11.15.6. Let $A \in \mathbb{F}^{n\times n}$. Then, A is discrete-time (Lyapunov stable, semistable, asymptotically stable) if and only if $A \otimes A$ is. (Proof: Use Fact 7.4.24.)

Fact 11.15.7. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$. Then, the following statements hold:

i) If A and B are discrete-time (Lyapunov stable, semistable, asymptotically stable), then $A \otimes B$ is discrete-time (Lyapunov stable, semistable, asymptotically stable).

ii) If $A \otimes B$ is discrete-time (Lyapunov stable, semistable, asymptotically stable), then either A or B is discrete-time (Lyapunov stable, semistable, asymptotically stable).

(Proof: Use Fact 7.4.24.)

Fact 11.15.8. Let $A \in \mathbb{R}^{n\times n}$ be (Lyapunov stable, semistable, asymptotically stable). Then, e^A is discrete-time (Lyapunov stable, semistable, asymptotically stable). (Problem: If $B \in \mathbb{R}^{n\times n}$ is discrete-time (Lyapunov stable, semistable, asymptotically stable), when does there exist (Lyapunov stable, semistable, asymptotically stable) $A \in \mathbb{R}^{n\times n}$ such that $B = e^A$? See Proposition 11.4.4.)

Fact 11.15.9. Let $A \in \mathbb{R}^{n\times n}$. If A is discrete-time asymptotically stable, then $B \triangleq (A+I)^{-1}(A-I)$ is asymptotically stable. Conversely, if $B \in \mathbb{R}^{n\times n}$ is asymptotically stable, then $A \triangleq (I+B)(I-B)^{-1}$ is discrete-time asymptotically stable. (Proof: See [271].) (Remark: For additional results on the Cayley transform, see Fact 3.6.23, Fact 3.6.24, Fact 3.6.25, Fact 3.9.8, and Fact 8.7.18.) (Problem: Obtain analogous results for Lyapunov-stable and semistable matrices.)

Fact 11.15.10. Let $\begin{bmatrix} P_1 & P_{12} \\ P_{12}^{\mathrm{T}} & P_2 \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$ be positive definite, where P_1, $P_{12}, P_2 \in \mathbb{R}^{n\times n}$. If $P_1 \ge P_2$, then $A \triangleq P_1^{-1}P_{12}^{\mathrm{T}}$ is discrete-time asymptotically stable, while if $P_2 \ge P_1$, then $A \triangleq P_2^{-1}P_{12}$ is discrete-time asymptotically stable. (Proof: If $P_1 \ge P_2$, then $P_1 - P_{12}P_1^{-1}P_1P_1^{-1}P_{12}^{\mathrm{T}} \ge P_1 - P_{12}P_2^{-2}P_{12}^{\mathrm{T}} > 0$. See [145].)

Fact 11.15.11. Let $A \in \mathbb{F}^{n\times n}$. Then, A is discrete-time semistable if and only if

$$A_\infty \triangleq \lim_{k\to\infty} A^k$$

exists. In this case, A_∞ is idempotent and is given by

$$A_\infty = I - (A-I)(A-I)^{\#}.$$

(Proof: See [416, p. 640].) (Remark: See Fact 11.15.16 and Fact 11.15.15.)

Fact 11.15.12. Let $A \in \mathbb{F}^{n\times n}$. Then, A is discrete-time Lyapunov stable if and only if

$$A_\infty \triangleq \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} A^i$$

exists. In this case,

$$A_\infty = I - (A-I)(A-I)^{\#}.$$

(Proof: See [416, p. 633].) (Remark: A is *Cesaro summable.*) (Remark: See Fact 6.3.17.)

Fact 11.15.13. Let $A \in \mathbb{F}^{n\times n}$. Then, A is discrete-time asymptotically stable if and only if

$$\lim_{k\to\infty} A^k = 0.$$

Fact 11.15.14. Let $A \in \mathbb{F}^{n\times n}$, and assume that A is unitary. Then, A is discrete-time Lyapunov stable.

Fact 11.15.15. Let $A, B \in \mathbb{R}^{n\times n}$, assume that A is discrete-time semistable, and let $A_\infty \triangleq \lim_{k\to\infty} A^k$. Then,

$$\lim_{k\to\infty} \left(A + \tfrac{1}{k}B\right)^k = A_\infty e^{A_\infty B A_\infty}.$$

(Proof: See [101, 598].) (Remark: If A is idempotent, then $A_\infty = A$. The existence of A_∞ is guaranteed by either Fact 11.15.11 or Fact 11.15.16.)

Fact 11.15.16. Let $A \in \mathbb{R}^{n\times n}$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is discrete-time Lyapunov stable if and only if $\{\|A^k\|\}_{k=0}^\infty$ is bounded.

ii) A is discrete-time semistable if and only if $A_\infty \triangleq \lim_{k\to\infty} A^k$ exists. In this case, $A_\infty = I - (A-I)(A-I)^\#$ is idempotent.

iii) A is discrete-time asymptotically stable if and only if $\lim_{k\to\infty} A^k = 0$.

(Remark: *ii)* is given by Fact 11.15.11. See Fact 11.15.15.)

Fact 11.15.17. Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements hold:

i) A is discrete-time Lyapunov stable if and only if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that $P - A^{\mathrm{T}}PA$ is nonnegative semidefinite.

ii) A is discrete-time asymptotically stable if and only if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that $P - A^{\mathrm{T}}PA$ is positive definite.

(Remark: The *discrete-time Lyapunov equation* or the *Stein equation* is $P = A^{\mathrm{T}}PA + R$.)

Fact 11.15.18. Let $\{A_k\}_{k=0}^\infty \subset \mathbb{R}^{n\times n}$ and, for $k \in \mathbb{N}$, consider the discrete-time, time-varying system

$$x_{k+1} = A_k x_k.$$

Furthermore, assume that there exist real numbers $\beta \in (0,1)$, $\gamma > 0$, and $\varepsilon > 0$ such that, for all $k \in \mathbb{N}$,

$$\begin{aligned} \operatorname{sprad}(A_k) &< \beta, \\ \|A_k\| &< \gamma, \\ \|A_{k+1} - A_k\| &< \varepsilon, \end{aligned}$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^{n\times n}$. Then, $x_k \to 0$ as $k \to \infty$. (Proof: See [265, pp. 170–173].) (Remark: This result arises from the theory of *infinite matrix products*.

11.16 Facts on Subspace Decomposition

Fact 11.16.1. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1}, \tag{11.16.1}$$

where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$. Then,

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} 0 & B_{12{\rm s}} \\ 0 & \mu_A^{\rm s}(A_2) \end{bmatrix}S^{-1},$$

where $B_{12{\rm s}} \in \mathbb{R}^{r\times(n-r)}$, and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & B_{12{\rm u}} \\ 0 & \mu_A^{\rm u}(A_2) \end{bmatrix}S^{-1},$$

where $B_{12{\rm u}} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\rm u}(A_1)$ is nonsingular. Consequently,

$$\mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right) \subseteq \mathcal{S}_{\rm s}(A).$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\rm s}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & 0 \\ 0 & \mu_A^{\rm u}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{S}_{\rm u}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

(Proof: The result follows from Fact 4.10.8.)

Fact 11.16.2. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ satisfies $\mathrm{spec}(A_2) \subset$ CRHP. Then,

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} \mu_A^{\rm s}(A_1) & C_{12{\rm s}} \\ 0 & \mu_A^{\rm s}(A_2) \end{bmatrix}S^{-1},$$

where $C_{12{\rm s}} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\rm s}(A_2)$ is nonsingular, and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & C_{12{\rm u}} \\ 0 & 0 \end{bmatrix}S^{-1},$$

where $C_{12\mathrm{u}} \in \mathbb{R}^{r\times(n-r)}$. Consequently,

$$\mathcal{S}_{\mathrm{s}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1},$$

$$\mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right) \subseteq \mathcal{S}_{\mathrm{u}}(A).$$

Fact 11.16.3. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ satisfies $\mathrm{spec}(A_1) \subset \mathrm{CRHP}$, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & B_{12\mathrm{s}} \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

where $\mu_A^{\mathrm{s}}(A_1)$ is nonsingular and $B_{12\mathrm{s}} \in \mathbb{R}^{r\times(n-r)}$, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & B_{12\mathrm{u}} \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

where $B_{12\mathrm{u}} \in \mathbb{R}^{r\times(n-r)}$. Consequently,

$$\mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right) \subseteq \mathcal{S}_{\mathrm{u}}(A).$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{S}_{\mathrm{s}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

Fact 11.16.4. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ is asymptotically stable. Then,

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} \mu_A^{\rm s}(A_1) & C_{12{\rm s}} \\ 0 & 0 \end{bmatrix}S^{-1},$$

where $C_{12{\rm s}} \in \mathbb{R}^{r\times(n-r)}$, and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & C_{12{\rm u}} \\ 0 & \mu_A^{\rm u}(A_2) \end{bmatrix}S^{-1},$$

where $\mu_A^{\rm u}(A_2)$ is nonsingular and $C_{12{\rm u}} \in \mathbb{R}^{r\times(n-r)}$. Consequently,

$$\mathcal{S}_{\rm u}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} \mu_A^{\rm s}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1},$$

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} \mu_A^{\rm u}(A_1) & 0 \\ 0 & \mu_A^{\rm u}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right) \subseteq \mathcal{S}_{\rm s}(A).$$

Fact 11.16.5. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ satisfies $\operatorname{spec}(A_1) \subset \mathrm{CRHP}$, $A_{12} \in \mathbb{R}^{r\times(n-r)}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ is asymptotically stable. Then,

$$\mu_A^{\rm s}(A) = S\begin{bmatrix} \mu_A^{\rm s}(A_1) & C_{12{\rm s}} \\ 0 & 0 \end{bmatrix}S^{-1},$$

where $C_{12{\rm s}} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\rm s}(A_1)$ is nonsingular, and

$$\mu_A^{\rm u}(A) = S\begin{bmatrix} 0 & C_{12{\rm u}} \\ 0 & \mu_A^{\rm u}(A_2) \end{bmatrix}S^{-1},$$

where $C_{12\mathrm{u}} \in \mathbb{R}^{r\times(n-r)}$ and $\mu_A^{\mathrm{u}}(A_2)$ is nonsingular. Consequently,

$$\mathcal{S}_{\mathrm{u}}(A) = \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

If, in addition, $A_{12} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1}$$

and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

Consequently,

$$\mathcal{S}_{\mathrm{s}}(A) = \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

Fact 11.16.6. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} 0 & 0 \\ B_{21\mathrm{s}} & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

where $B_{21\mathrm{s}} \in \mathbb{R}^{(n-r)\times r}$, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ B_{21\mathrm{u}} & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

where $B_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{u}}(A_1)$ is nonsingular. Consequently,

$$\mathcal{S}_{\mathrm{u}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right) \subseteq \mathcal{S}_{\mathrm{s}}(A).$$

Fact 11.16.7. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ satisfies $\operatorname{spec}(A_2) \subset \mathrm{CRHP}$. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ C_{21\mathrm{s}} & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{s}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{s}}(A_2)$ is nonsingular, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ C_{21\mathrm{u}} & 0 \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$. Consequently,

$$\mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right) \subseteq \mathcal{S}_{\mathrm{u}}(A).$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1},$$

$$\mathcal{S}_{\mathrm{s}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

Fact 11.16.8. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ is asymptotically stable, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ satisfies $\operatorname{spec}(A_2) \subset \mathrm{CRHP}$. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} 0 & 0 \\ C_{21\mathrm{s}} & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{s}} \in \mathbb{R}^{n-r\times r}$ and $\mu_A^{\mathrm{s}}(A_2)$ is nonsingular, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ C_{21\mathrm{u}} & 0 \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{u}}(A_1)$ is nonsingular. Consequently,

$$\mathcal{S}_{\mathrm{u}}(A) = \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1}$$

and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1}.$$

Consequently,

$$\mathcal{S}_{\mathrm{s}}(A) = \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

Fact 11.16.9. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ is asymptotically stable. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ B_{21\mathrm{s}} & 0 \end{bmatrix}S^{-1},$$

where $B_{21\mathrm{s}} \in \mathbb{R}^{(n-r)\times r}$, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ B_{21\mathrm{u}} & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

where $B_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{u}}(A_2)$ is nonsingular. Consequently,

$$\mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right) \subseteq \mathcal{S}(A)$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{u}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{S}_{\mathrm{u}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

Fact 11.16.10. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ satisfies $\mathrm{spec}(A_1) \subset \mathrm{CRHP}$, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ C_{12\mathrm{s}} & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{s}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{s}}(A_1)$ is nonsingular, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ C_{21\mathrm{u}} & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$. Consequently,

$$\mathcal{S}_{\mathrm{s}}(A) \subseteq \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix}S^{-1},$$

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

$$\mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right) \subseteq \mathcal{S}_{\mathrm{u}}(A).$$

Fact 11.16.11. Let $A \in \mathbb{R}^{n\times n}$, and let $S \in \mathbb{R}^{n\times n}$ be a nonsingular matrix such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1},$$

where $A_1 \in \mathbb{R}^{r\times r}$ satisfies $\operatorname{spec}(A_1) \subset \mathrm{CRHP}$, $A_{21} \in \mathbb{R}^{(n-r)\times r}$, and $A_2 \in \mathbb{R}^{(n-r)\times(n-r)}$ is asymptotically stable. Then,

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ C_{21\mathrm{s}} & 0 \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{s}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{s}}(A_1)$ is nonsingular, and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ C_{21\mathrm{u}} & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1},$$

where $C_{21\mathrm{u}} \in \mathbb{R}^{(n-r)\times r}$ and $\mu_A^{\mathrm{u}}(A_2)$ is nonsingular. Consequently,

$$\mathcal{S}_{\mathrm{s}}(A) = \mathcal{R}\left(S\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix}\right).$$

If, in addition, $A_{21} = 0$, then

$$\mu_A^{\mathrm{s}}(A) = S\begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & 0 \\ 0 & 0 \end{bmatrix}S^{-1}$$

and

$$\mu_A^{\mathrm{u}}(A) = S\begin{bmatrix} 0 & 0 \\ 0 & \mu_A^{\mathrm{u}}(A_2) \end{bmatrix}S^{-1}.$$

Consequently,

$$\mathcal{S}_{\mathrm{u}}(A) = \mathcal{R}\left(S\begin{bmatrix} I_r \\ 0 \end{bmatrix}\right).$$

11.17 Notes

Explicit formulas for the matrix exponential are given in [32, 89, 142, 264,458,459]. Computational methods are discussed in [426]. An arithmetic-mean-geometric-mean iteration for computing the matrix exponential and matrix logarithm is given in [527].

The exponential function plays a central role in the theory of Lie groups, see [72, 132, 299, 304, 496, 571]. Applications to robotics and kinematics are given in [432,450]. Additional applications are discussed in [131].

The real logarithm is discussed in [156, 274, 441, 469].

An asymptotically stable polynomial is traditionally called *Hurwitz.* Semistability was first defined in [124]. Stability theory is treated in [257, 361, 463]. Solutions of the Lyapunov equation under weak conditions are considered in [512].

Chapter Twelve

Linear Systems and Control Theory

This chapter considers linear state space systems with inputs and outputs. These systems are considered in both the time domain and frequency (Laplace) domain. Some basic results in control theory are also considered.

12.1 State Space and Transfer Function Models

Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, and, for $t \geq t_0$, consider the *state equation*

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.1.1}$$

with the *initial condition*

$$x(t_0) = x_0. \tag{12.1.2}$$

In (12.1.1), $x(t) \in \mathbb{R}^n$ is the *state* and $u(t) \in \mathbb{R}^m$ is the *input*.

Proposition 12.1.1. For $t \geq t_0$ the state $x(t)$ of the dynamical equation (12.1.1) with initial condition (12.1.2) is given by

$$x(t) = e^{(t-t_0)A}x_0 + \int_{t_0}^{t} e^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau. \tag{12.1.3}$$

Proof. Multiplying (12.1.1) by e^{-tA} yields

$$e^{-tA}[\dot{x}(t) - Ax(t)] = e^{-tA}Bu(t),$$

which is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t}\big[e^{-tA}x(t)\big] = e^{-tA}Bu(t).$$

Integrating over $[t_0, t]$ yields

$$e^{-tA}x(t) = e^{-t_0A}x(t_0) + \int_{t_0}^{t} e^{-\tau A}Bu(\tau)\,\mathrm{d}\tau.$$

Now, multiplying by e^{tA} yields (12.1.3).

Alternatively, let $x(t)$ be given by (12.1.3). Then, it follows from Liebniz' rule Fact 10.8.4 that

$$\begin{aligned}\dot{x}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}e^{(t-t_0)A}x_0 + \frac{\mathrm{d}}{\mathrm{d}t}\int_{t_0}^{t} e^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau \\ &= Ae^{(t-t_0)A}x_0 + \int_{t_0}^{t} Ae^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau + Bu(t) \\ &= Ax(t) + Bu(t).\end{aligned}$$

□

For convenience, we can reset the clock and assume without loss of generality that $t_0 = 0$. In this case, $x(t)$ for all $t \geq 0$ is given by

$$x(t) = e^{tA}x_0 + \int_{0}^{t} e^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau. \tag{12.1.4}$$

If $u(t) = 0$ for all $t \geq 0$, then, for all $t \geq 0$, $x(t)$ is given by

$$x(t) = e^{tA}x_0. \tag{12.1.5}$$

Now, let $u(t) = \delta(t)v$, where $\delta(t)$ is the *unit impulse* at $t = 0$ and $v \in \mathbb{R}^m$. Then, for all $t \geq 0$, $x(t)$ is given by

$$x(t) = e^{tA}x_0 + e^{tA}Bv. \tag{12.1.6}$$

Let $a < b$. Then, $\delta(t)$, which has physical dimensions of 1/time, satisfies

$$\int_{a}^{b} \delta(\tau)\,\mathrm{d}\tau = \begin{cases} 0, & a > 0 \text{ or } b \leq 0, \\ 1, & a \leq 0 < b. \end{cases} \tag{12.1.7}$$

More generally, if $g\colon\ \mathcal{D} \to \mathbb{R}^n$, where $[a, b] \subseteq \mathcal{D} \subseteq \mathbb{R}$, $t_0 \in \mathcal{D}$, and g is continuous at t_0, then

$$\int_{a}^{b} \delta(\tau - t_0)g(\tau)\,\mathrm{d}\tau = \begin{cases} 0, & a > t_0 \text{ or } b \leq t_0, \\ g(t_0), & a \leq t_0 < b. \end{cases} \tag{12.1.8}$$

Alternatively, if the input $u(t)$ is constant, that is, $u(t) = v$ for all $t \geq 0$, where $v \in \mathbb{R}^m$, then, by a change of variable of integration, it follows that, for all $t \geq 0$,

$$x(t) = e^{tA}x_0 + \int_0^t e^{\tau A}\,\mathrm{d}\tau Bv. \tag{12.1.9}$$

Using Fact 11.10.12, (12.1.9) can be written for all $t \geq 0$ as

$$x(t) = e^{tA}x_0 + \left[A^{\mathrm{D}}\big(e^{tA} - I\big) + \big(I - AA^{\mathrm{D}}\big)\sum_{i=1}^{\operatorname{ind} A}(i!)^{-1}t^iA^{i-1}\right]Bv. \tag{12.1.10}$$

If A is group invertible, then, for all $t \geq 0$, (12.1.10) becomes

$$x(t) = e^{tA}x_0 + \left[A^{\#}\big(e^{tA} - I\big) + t\big(I - AA^{\#}\big)\right]Bv. \tag{12.1.11}$$

If, in addition, A is nonsingular, then, for all $t \geq 0$, (12.1.11) becomes

$$x(t) = e^{tA}x_0 + A^{-1}\big(e^{tA} - I\big)Bv. \tag{12.1.12}$$

Next, consider the *output equation*

$$y(t) = Cx(t) + Du(t), \tag{12.1.13}$$

where $t \geq 0$, $y(t) \in \mathbb{R}^l$ is the *output*, $C \in \mathbb{R}^{l\times n}$, and $D \in \mathbb{R}^{l\times m}$. Then, for all $t \geq 0$,

$$y(t) = Ce^{tA}x_0 + \int_0^t Ce^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau + Du(t). \tag{12.1.14}$$

If $u(t) = 0$ for all $t \geq 0$, then the *free response* is given by

$$y(t) = Ce^{tA}x_0, \tag{12.1.15}$$

while, if $x_0 = 0$, then the *forced response* is given by

$$y(t) = \int_0^t Ce^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau + Du(t). \tag{12.1.16}$$

In particular, setting $u(t) = \delta(t)v$ yields, for all $t > 0$,

$$y(t) = Ce^{tA}x_0 + H(t)v, \tag{12.1.17}$$

where, for all $t \geq 0$, the *impulse response function* $H(t)$ is defined by

$$H(t) \triangleq Ce^{tA}B + \delta(t)D, \tag{12.1.18}$$

and the *impulse response* is

$$y(t) = H(t)v. \tag{12.1.19}$$

Alternatively, if $u(t) = v$ for all $t \geq 0$, then

$$y(t) = Ce^{tA}x_0 + \int_0^t Ce^{\tau A}\,\mathrm{d}\tau Bv + Dv, \tag{12.1.20}$$

and the *step response* is

$$y(t) = \int_0^t H(\tau)\,\mathrm{d}\tau v = \int_0^t Ce^{\tau A}\,\mathrm{d}\tau Bv + Dv. \tag{12.1.21}$$

In general, the forced response can be written as

$$y(t) = \int_0^t H(t-\tau)u(\tau)\,\mathrm{d}\tau. \tag{12.1.22}$$

Proposition 12.1.2. Let $D = 0$ and $m = 1$, and assume that $x_0 = Bv$. Then, the free response and the impulse response are equal and given by

$$y(t) = Ce^{tA}x_0 = Ce^{tA}Bv. \tag{12.1.23}$$

Now, consider the linear system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.1.24}$$

$$y(t) = Cx(t) + Du(t), \tag{12.1.25}$$

with state $x(t) \in \mathbb{R}^n$, input $u(t) \in \mathbb{R}^m$, and *output* $y(t) \in \mathbb{R}^l$, where $t \geq 0$ and $x(0) = x_0$. Taking Laplace transforms yields

$$s\hat{x}(s) - x_0 = A\hat{x}(s) + B\hat{u}(s), \tag{12.1.26}$$

$$\hat{y}(s) = C\hat{x}(s) + D\hat{u}(s), \tag{12.1.27}$$

where

$$\hat{x}(s) \triangleq \mathcal{L}\{x(t)\} \triangleq \int_0^\infty e^{-st}x(t)\,\mathrm{d}t, \tag{12.1.28}$$

$$\hat{u}(s) \triangleq \mathcal{L}\{u(t)\}, \tag{12.1.29}$$

and

$$\hat{y}(s) \triangleq \mathcal{L}\{y(t)\}. \tag{12.1.30}$$

Hence,

$$\hat{x}(s) = (sI - A)^{-1}x_0 + (sI - A)^{-1}B\hat{u}(s), \tag{12.1.31}$$

and thus

$$\hat{y}(s) = C(sI - A)^{-1}x_0 + \big[C(sI - A)^{-1}B + D\big]\hat{u}(s). \tag{12.1.32}$$

We can also obtain (12.1.32) from the time-domain expression for $y(t)$ given by (12.1.14). To do this, the following result will be needed.

Lemma 12.1.3. Let $A \in \mathbb{R}^{n\times n}$. Then, for all $s \in \mathbb{C}\backslash\mathrm{spec}(A)$,

$$\mathcal{L}\{e^{tA}\} = (sI - A)^{-1}. \tag{12.1.33}$$

Proof. Let $s \in \mathbb{C}$ satisfy $\mathrm{Re}\, s > \mathrm{spabs}(A)$ so that $A - sI$ is asymptotically stable. Thus, it follows from Lemma 11.7.2 that

$$\mathcal{L}\{e^{tA}\} = \int_0^\infty e^{-st}e^{tA}\,\mathrm{d}t = \int_0^\infty e^{t(A-sI)}\,\mathrm{d}t = (sI - A)^{-1}.$$

By analytic continuation, $\mathcal{L}\{e^{tA}\}$ is given by (12.1.33) for all $s \in \mathbb{C}\backslash\mathrm{spec}(A)$. □

Using Lemma 12.1.3, it follows from (12.1.14) that

$$\begin{aligned}\hat{y}(s) &= \mathcal{L}\{Ce^{tA}x_0\} + \mathcal{L}\left\{\int_0^t Ce^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau\right\} + D\hat{u}(s)\\ &= C\mathcal{L}\{e^{tA}\}x_0 + C\mathcal{L}\{e^{tA}\}B\hat{u}(s) + D\hat{u}(s)\\ &= C(sI - A)^{-1}x_0 + \left[C(sI - A)^{-1}B + D\right]\hat{u}(s),\end{aligned} \tag{12.1.34}$$

which coincides with (12.1.32). We define

$$G(s) \triangleq C(sI - A)^{-1}B + D. \tag{12.1.35}$$

Note that $G \in \mathbb{R}^{l\times m}(s)$, that is, by Definition 4.7.2, G is a rational transfer function. Since $\mathcal{L}\{\delta(t)\} = 1$ it follows that

$$G(s) = \mathcal{L}\{H(t)\}. \tag{12.1.36}$$

Using (4.7.2), G can be written as

$$G(s) = \frac{1}{\chi_A(s)}C(sI - A)^{\mathrm{A}}B + D. \tag{12.1.37}$$

It follows from (4.7.3) that G is a proper rational transfer function. Furthermore, G is a strictly proper rational transfer function if and only if $D = 0$, whereas G is an exactly proper rational transfer function if and only if $D \neq 0$. Finally, if A is nonsingular, then

$$G(0) = -CA^{-1}B + D. \tag{12.1.38}$$

Let $A \in \mathbb{R}^{n\times n}$. If $|s| > \operatorname{sprad}(A)$, then Proposition 9.4.10 implies that

$$(sI-A)^{-1} = \tfrac{1}{s}\big(I - \tfrac{1}{s}A\big)^{-1} = \sum_{k=0}^{\infty} \tfrac{1}{s^{k+1}}A^k, \tag{12.1.39}$$

where the series is absolutely convergent, and thus

$$G(s) = \sum_{k=-1}^{\infty} \tfrac{1}{s^{k+1}}H_k, \tag{12.1.40}$$

where, for $k \ge -1$, the *Markov parameter* $H_k \in \mathbb{R}^{l\times m}$ is defined by

$$H_k \triangleq \begin{cases} D, & k=-1,\\ CA^kB, & k \ge 0. \end{cases} \tag{12.1.41}$$

It follows from (12.1.39) that $\lim_{s\to\infty}(sI-A)^{-1} = 0$, and thus

$$\lim_{s\to\infty} G(s) = D. \tag{12.1.42}$$

Finally, it follows from Definition 4.7.2 that

$$\operatorname{reldeg} G = \min\{k \ge -1\colon\ H_k \ne 0\}. \tag{12.1.43}$$

12.2 Observability

Let $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{l\times n}$, and, for $t \ge 0$, consider the linear system

$$\dot{x}(t) = Ax(t), \tag{12.2.1}$$

$$x(0) = x_0, \tag{12.2.2}$$

$$y(t) = Cx(t). \tag{12.2.3}$$

Definition 12.2.1. The *unobservable subspace* $\mathcal{U}_{t_{\rm f}}(A,C)$ of (A,C) at time $t_{\rm f} > 0$ is the subspace

$$\mathcal{U}_{t_{\rm f}}(A,C) \triangleq \{x_0 \in \mathbb{R}^n\colon\ y(t) = 0 \text{ for all } t \in [0,t_{\rm f}]\}. \tag{12.2.4}$$

Let $t_{\rm f} > 0$. Since $y(t) = 0$ for all $t \in [0,t_{\rm f}]$ is the free response corresponding to $x_0 = 0$, it follows that $0 \in \mathcal{U}_{t_{\rm f}}(A,C)$. Hence, if $x_0 \ne 0$ and $x_0 \in \mathcal{U}_{t_{\rm f}}(A,C)$, then x_0 cannot be determined from knowledge of $y(t)$ for all $t \in [0,t_{\rm f}]$.

The following result provides explicit expressions for $\mathcal{U}_{t_{\rm f}}(A,C)$.

Lemma 12.2.2. Let $t_{\rm f} > 0$. Then, the following subspaces are equal:

i) $\mathcal{U}_{t_{\rm f}}(A,C)$

ii) $\bigcap_{t\in[0,t_{\rm f}]} \mathcal{N}\big(Ce^{tA}\big)$

iii) $\bigcap_{i=0}^{n-1} \mathcal{N}\big(CA^i\big)$

iv) $\mathcal{N}\left(\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}\right)$

v) $\mathcal{N}\Big(\int_0^{t_{\rm f}} e^{tA^{\rm T}}C^{\rm T}Ce^{tA}\,{\rm d}t\Big)$

Proof. The proof is dual to the proof of Lemma 12.5.2. □

Lemma 12.2.2 shows that $\mathcal{U}_{t_{\rm f}}(A,C)$ is independent of $t_{\rm f}$. Hence, we write $\mathcal{U}(A,C)$ for $\mathcal{U}_{t_{\rm f}}(A,C)$ and call $\mathcal{U}(A,C)$ the *unobservable subspace* of (A,C). (A,C) is *observable* if $\mathcal{U}(A,C) = \{0\}$. For convenience, define the *observability matrix*

$$\mathcal{O}(A,C) \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{12.2.5}$$

so that

$$\mathcal{U}(A,C) = \mathcal{N}[\mathcal{O}(A,C)]. \tag{12.2.6}$$

Define

$$p \triangleq n - \dim \mathcal{U}(A,C) = n - \operatorname{def} \mathcal{O}(A,C). \tag{12.2.7}$$

The following result shows that the unobservable subspace $\mathcal{U}(A,C)$ is unchanged by replacing $\dot{x}(t) = Ax(t)$ by $\dot{x}(t) = Ax(t) + Fy(t)$.

Proposition 12.2.3. Let $F \in \mathbb{R}^{n\times l}$. Then,

$$\mathcal{U}(A+FC,C) = \mathcal{U}(A,C). \tag{12.2.8}$$

In particular, (A,C) is observable if and only if $(A+FC,C)$ is observable.

Proof. The proof is dual to the proof of Proposition 12.5.3. □

Let $\tilde{\mathcal{U}}(A,C) \subseteq \mathbb{R}^n$ be a subspace that is complementary to $\mathcal{U}(A,C)$. Then, $\tilde{\mathcal{U}}(A,C)$ is an *observable subspace* in the sense that if $x_0 = x_0' + x_0''$, where $x_0' \in \tilde{\mathcal{U}}(A,C)$ and $x_0'' \in \mathcal{U}(A,C)$, then it is possible to determine x_0' from knowledge of $y(t)$ for $t \in [0,t_{\rm f}]$. The following result uses $y(t)$ to determine x_0' for $\tilde{\mathcal{U}}(A,C) \triangleq \mathcal{U}(A,C)^\perp$.

Lemma 12.2.4. Let $t_{\mathrm{f}} > 0$, and define $\mathcal{P} \in \mathbb{R}^{n\times n}$ by

$$\mathcal{P} \triangleq \left(\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t\right)^{+}\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t. \tag{12.2.9}$$

Then, $\mathcal{P}_{\perp}$ is the projector onto $\mathcal{U}(A,C)$, and $\mathcal{P}$ is the projector onto $\mathcal{U}(A,C)^{\perp}$. Hence,

$$\mathcal{U}(A,C) = \mathcal{N}(\mathcal{P}) = \mathcal{R}(\mathcal{P}_{\perp}), \tag{12.2.10}$$

$$\mathcal{U}(A,C)^{\perp} = \mathcal{R}(\mathcal{P}) = \mathcal{N}(\mathcal{P}_{\perp}), \tag{12.2.11}$$

$$n - p = \dim\mathcal{U}(A,C) = \operatorname{def}\mathcal{P} = \operatorname{rank}\mathcal{P}_{\perp}, \tag{12.2.12}$$

$$p = \dim\mathcal{U}(A,C)^{\perp} = \operatorname{rank}\mathcal{P} = \operatorname{def}\mathcal{P}_{\perp}. \tag{12.2.13}$$

If $x_0 = x_0' + x_0''$, where $x_0' \in \mathcal{U}(A,C)^{\perp}$ and $x_0'' \in \mathcal{U}(A,C)$, then

$$x_0' = \mathcal{P}x_0 = \left(\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t\right)^{+}\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}y(t)\,\mathrm{d}t. \tag{12.2.14}$$

Finally, (A,C) is observable if and only if $\mathcal{P} = I_n$. In this case, for all $x_0 \in \mathbb{R}^n$,

$$x_0 = \left(\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t\right)^{-1}\int_0^{t_{\mathrm{f}}} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}y(t)\,\mathrm{d}t. \tag{12.2.15}$$

Lemma 12.2.5. Let $\alpha \in \mathbb{R}$. Then,

$$\mathcal{U}(A+\alpha I, C) = \mathcal{U}(A,C). \tag{12.2.16}$$

The following result uses a coordinate transformation to characterize $\mathcal{U}(A,C)$.

Theorem 12.2.6. There exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that A and C have the form

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1}, \qquad C = \begin{bmatrix} C_1 & 0 \end{bmatrix}S^{-1}, \tag{12.2.17}$$

where $A_1 \in \mathbb{R}^{p\times p}$, $C_1 \in \mathbb{R}^{l\times p}$, and (A_1, C_1) is observable.

Proof. The proof is dual to the proof of Theorem 12.5.6. □

Proposition 12.2.7. Let $S \in \mathbb{R}^{n\times n}$ be orthogonal. Then, the following conditions are equivalent:

i) A and C have the form (12.2.17), where $A_1 \in \mathbb{R}^{p\times p}$, $C_1 \in \mathbb{R}^{l\times p}$, and (A_1, C_1) is observable.

ii) $\mathcal{U}(A,C) = \mathcal{R}\big(S\big[\begin{smallmatrix}0\\ I_{n-p}\end{smallmatrix}\big]\big)$.

iii) $\mathcal{U}(A,C)^\perp = \mathcal{R}\big(S\big[\begin{smallmatrix}I_p\\ 0\end{smallmatrix}\big]\big)$.

iv) $\mathcal{P} = S\begin{bmatrix} I_p & 0\\ 0 & 0\end{bmatrix}S^{\mathrm{T}}$.

Proposition 12.2.8. Let $S \in \mathbb{R}^{n\times n}$ be nonsingular. Then, the following conditions are equivalent:

i) A and C have the form (12.2.17), where $A_1 \in \mathbb{R}^{p\times p}$, $C_1 \in \mathbb{R}^{l\times p}$, and (A_1, C_1) is observable.

ii) $\mathcal{U}(A,C) = \mathcal{R}\big(S\big[\begin{smallmatrix}0\\ I_{n-p}\end{smallmatrix}\big]\big)$.

iii) $\mathcal{U}(A,C)^\perp = \mathcal{R}\big(S^{-\mathrm{T}}\big[\begin{smallmatrix}I_p\\ 0\end{smallmatrix}\big]\big)$.

Definition 12.2.9. Let $\lambda \in \operatorname{spec}(A)$. Then, λ is an *observable eigenvalue* of (A,C) if

$$\operatorname{rank}\begin{bmatrix}\lambda I - A\\ C\end{bmatrix} = n. \tag{12.2.18}$$

Otherwise, λ is an *unobservable eigenvalue* of (A,C).

Proposition 12.2.10. Let $\lambda \in \operatorname{mspec}(A)$ and $F \in \mathbb{R}^{n\times l}$. Then, λ is an observable eigenvalue of (A,C) if and only if λ is an observable eigenvalue of $(A+FC, C)$.

Lemma 12.2.11. Let $\lambda \in \operatorname{mspec}(A)$. Then,

$$\operatorname{Re}\mathcal{N}\left(\begin{bmatrix}\lambda I - A\\ C\end{bmatrix}\right) \subseteq \mathcal{U}(A,C). \tag{12.2.19}$$

Proof. Let $x \in \mathcal{N}\big(\big[\begin{smallmatrix}\lambda I - A\\ C\end{smallmatrix}\big]\big)$ so that $Ax = \lambda x$ and $Cx = 0$. Let $x_0 \triangleq \operatorname{Re} x$. Then, for all $t \ge 0$, $y(t) = Ce^{tA}x_0 = Ce^{tA}\operatorname{Re} x = \operatorname{Re} Ce^{tA}x = \operatorname{Re} Ce^{\lambda t}x = \operatorname{Re} e^{\lambda t}Cx = 0$. Hence, $\operatorname{Re} x = x_0 \in \mathcal{U}(A,C)$. □

The next result characterizes observability in several equivalent ways.

Theorem 12.2.12. The following statements are equivalent:

i) (A,C) is observable.

ii) There exists $t > 0$ such that $\int_0^t e^{\tau A^{\mathrm{T}}}C^{\mathrm{T}}Ce^{\tau A}\,\mathrm{d}\tau$ is positive definite.

iii) $\int_0^t e^{\tau A^{\mathrm{T}}}C^{\mathrm{T}}Ce^{\tau A}\,\mathrm{d}\tau$ is positive definite for all $t > 0$.

iv) $\operatorname{rank}\mathcal{O}(A,C) = n$.

v) Every eigenvalue of (A,C) is observable.

vi) For every self-conjugate multiset $\{\lambda_1, \ldots, \lambda_n\}_{\rm m} \subset \mathbb{C}$, there exists a matrix $F \in \mathbb{R}^{n \times l}$ such that $\operatorname{mspec}(A + FC) = \{\lambda_1, \ldots, \lambda_n\}_{\rm m}$.

Proof. The proof is dual to the proof of Theorem 12.5.12. □

12.3 Detectability

Let $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{l \times n}$, and $p \triangleq n - \dim \mathcal{U}(A, C)$.

Definition 12.3.1. (A, C) is *detectable* if

$$\mathcal{U}(A, C) \subseteq \mathcal{S}_{\rm s}(A). \tag{12.3.1}$$

Proposition 12.3.2. Let $F \in \mathbb{R}^{n \times l}$. Then, (A, C) is detectable if and only if $(A + FC, C)$ is detectable.

Proposition 12.3.3. The following statements are equivalent:

i) (A, C) is detectable.

ii) There exists a nonsingular matrix $S \in \mathbb{R}^{n \times n}$ such that

$$A = S \begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix} S^{-1}, \quad C = \begin{bmatrix} C_1 & 0 \end{bmatrix} S^{-1}, \tag{12.3.2}$$

where $A_1 \in \mathbb{R}^{p \times p}$, $C_1 \in \mathbb{R}^{l \times p}$, (A_1, C_1) is observable, and $A_2 \in \mathbb{R}^{(n-p) \times (n-p)}$ is asymptotically stable.

iii) Every CRHP eigenvalue of (A, C) is observable.

iv) $(A + FC, C)$ is detectable for all $F \in \mathbb{R}^{n \times l}$.

Proof. The proof is dual to the proof of Proposition 12.6.3. □

Lemma 12.3.4. Assume that (A, C) is detectable and that

$$P \triangleq \int_0^\infty e^{tA^{\rm T}} C^{\rm T} C e^{tA^{\rm T}} \, {\rm d}t$$

exists. Then, A is asymptotically stable.

12.4 Observable Asymptotic Stability

Definition 12.4.1. (A, C) is *observably asymptotically stable* if

$$\mathcal{S}_{\rm u}(A) \subseteq \mathcal{U}(A, C). \tag{12.4.1}$$

Proposition 12.4.2. Let $F \in \mathbb{R}^{n\times l}$. Then, (A, C) is observably asymptotically stable if and only if $(A+FC, C)$ is observably asymptotically stable.

Lemma 12.4.3. Assume that the nonnegative-semidefinite matrix $P \in \mathbb{R}^{n\times n}$ defined by

$$P \triangleq \int_0^\infty e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t \tag{12.4.2}$$

exists. Then, P satisfies

$$A^{\mathrm{T}}P + PA + C^{\mathrm{T}}C = 0. \tag{12.4.3}$$

The matrix P defined by (12.4.2) is the *observability Gramian*, and equation (12.4.3) is the *observation Lyapunov equation.* If $A \in \mathbb{R}^{n\times n}$ is asymptotically stable, Then, Corollary 11.7.4 implies that the P defined by (12.4.2) exists and is the unique solution to (12.4.3).

Proposition 12.4.4. The following statements are equivalent:

i) (A, C) is observably asymptotically stable.

ii) There exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ and $k \in \mathbb{N}$ such that

$$A = S\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}S^{-1}, \qquad C = \begin{bmatrix} C_1 & 0 \end{bmatrix}S^{-1}, \tag{12.4.4}$$

where $A_1 \in \mathbb{R}^{k\times k}$ is asymptotically stable and $C_1 \in \mathbb{R}^{l\times k}$.

iii) $\lim_{t\to\infty} Ce^{tA} = 0$.

iv) $P \triangleq \int_0^\infty e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t$ exists.

v) There exists a nonnegative-semidefinite matrix $P \in \mathbb{R}^{n\times n}$ satisfying (12.4.3).

In this case, one nonnegative-semidefinite solution of (12.4.3) is given by (12.4.2). Furthermore,

$$\mathcal{P} = PP^{+}, \tag{12.4.5}$$

$$\mathcal{R}(\mathcal{P}) = \mathcal{R}(P) = \mathcal{U}(A, C)^{\perp}, \tag{12.4.6}$$

$$\mathcal{N}(\mathcal{P}) = \mathcal{N}(P) = \mathcal{U}(A, C), \tag{12.4.7}$$

$$\operatorname{rank}\mathcal{P} = \operatorname{rank}P = p. \tag{12.4.8}$$

Proof. The proof is dual to the proof of Proposition 12.7.4. □

Proposition 12.4.5. The following statements are equivalent:

i) A is asymptotically stable.

ii) (A, C) is detectable and observably asymptotically stable.

Furthermore, if two of the following three conditions are satisfied, then the third condition is satisfied:

iii) A is asymptotically stable.

iv) (A, C) is observable.

v) $P \triangleq \int_0^\infty e^{tA^{\mathrm{T}}} C^{\mathrm{T}} C e^{tA}\, \mathrm{d}t$ exists and is positive definite.

12.5 Controllability

Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$, and, for $t \geq 0$, consider the linear system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.5.1}$$

$$x(0) = 0. \tag{12.5.2}$$

Definition 12.5.1. The *controllable subspace* $\mathcal{C}_{t_\mathrm{f}}(A, B)$ of (A, B) at time $t_\mathrm{f} > 0$ is the subspace

$$\mathcal{C}_{t_\mathrm{f}}(A, B) \triangleq \{x_\mathrm{f} \in \mathbb{R}^n\colon \text{ there exists a continuous control } u\colon [0, t_\mathrm{f}] \mapsto \mathbb{R}^m \text{ such that the solution } x(\cdot) \text{ of (12.5.1), (12.5.2) satisfies } x(t_\mathrm{f}) = x_\mathrm{f}\}. \tag{12.5.3}$$

Let $t_\mathrm{f} > 0$. Then, Definition 12.5.1 states that $x_\mathrm{f} \in \mathcal{C}_{t_\mathrm{f}}(A, B)$ if and only if there exists a continuous control $u\colon [0, t_\mathrm{f}] \mapsto \mathbb{R}^m$ such that

$$x_\mathrm{f} = \int_0^{t_\mathrm{f}} e^{(t_\mathrm{f} - t)A} Bu(t)\, \mathrm{d}t. \tag{12.5.4}$$

The following result provides explicit expressions for $\mathcal{C}_{t_\mathrm{f}}(A, B)$.

Lemma 12.5.2. Let $t_\mathrm{f} > 0$. Then, the following subspaces are equal:

i) $\mathcal{C}_{t_\mathrm{f}}(A, B)$

ii) $\left[\bigcap_{t \in [0, t_\mathrm{f}]} \mathcal{N}\left(B^{\mathrm{T}} e^{tA^{\mathrm{T}}}\right)\right]^\perp$

iii) $\left[\bigcap_{i=0}^{n-1} \mathcal{N}\left(B^{\mathrm{T}} A^{i\mathrm{T}}\right)\right]^\perp$

iv) $\mathcal{R}\left(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}\right)$

v) $\mathcal{R}\left(\int_0^{t_\mathrm{f}} e^{tA} B B^{\mathrm{T}} e^{tA^{\mathrm{T}}}\, \mathrm{d}t\right)$

Proof. To prove that $i) \subseteq ii)$, let $\eta \in \bigcap_{t\in[0,t_{\rm f}]}\mathcal{N}\Big(B^{\rm T}e^{tA^{\rm T}}\Big)$ so that $\eta^{\rm T}e^{tA}B = 0$ for all $t \in [0, t_{\rm f}]$. Now, let $u\colon\ [0, t_{\rm f}] \mapsto \mathbb{R}^m$ be continuous. Then, $\eta^{\rm T}\int_0^{t_{\rm f}} e^{(t_{\rm f}-t)A}Bu(t)\,{\rm d}t = 0$, which implies that $\eta \in \mathcal{C}_{t_{\rm f}}(A,B)^\perp$.

To prove that $ii) \subseteq iii)$, let $\eta \in \bigcap_{i=0}^{n-1}\mathcal{N}\big(B^{\rm T}A^{i{\rm T}}\big)$ so that $\eta^{\rm T}A^iB = 0$ for all $i = 0, 1, \ldots, n-1$. It follows from Theorem 4.4.6 that $\eta^{\rm T}A^iB = 0$ for all $i \ge 0$. Now, let $t \in [0, t_{\rm f}]$. Then, $\eta^{\rm T}e^{tA}B = \sum_{i=0}^{\infty} t^i(i!)^{-1}\eta^{\rm T}A^iB = 0$, and thus $\eta \in \mathcal{N}\Big(B^{\rm T}e^{tA^{\rm T}}\Big)$.

To show that $iii) \subseteq iv)$, let $\eta \in \mathcal{R}\big(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}\big)^\perp$. Then, $\eta \in \mathcal{N}\Big(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}^{\rm T}\Big)$, which implies that $\eta^{\rm T}A^iB = 0$ for all $i = 0, 1, \ldots, n-1$.

To prove that $iv) \subseteq v)$, let $\eta \in \mathcal{N}\Big(\int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}{\rm d}t\Big)$. Then,

$$\eta^{\rm T}\int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}{\rm d}t\eta = 0,$$

which implies that $\eta^{\rm T}e^{tA}B = 0$ for all $t \in [0, t_{\rm f}]$. Differentiating with respect to t and setting $t = 0$ implies that $\eta^{\rm T}A^iB = 0$ for all $i = 0, 1, \ldots, n-1$. Hence, $\eta \in \mathcal{R}\big(\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}\big)^\perp$.

To prove that $v) \subseteq i)$, let $\eta \in \mathcal{C}_{t_{\rm f}}(A,B)^\perp$. Then, $\eta^{\rm T}\int_0^{t_{\rm f}} e^{(t_{\rm f}-t)A}Bu(t)\,{\rm d}t = 0$ for all continuous $u\colon\ [0, t_{\rm f}] \mapsto \mathbb{R}^m$. Letting $u(t) = B^{\rm T}e^{(t_{\rm f}-t)A^{\rm T}}\eta^{\rm T}$, it follows that $\eta^{\rm T}\int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}{\rm d}t\eta = 0$, which implies that $\eta \in \mathcal{N}\Big(\int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}{\rm d}t\Big)$. □

Lemma 12.5.2 shows that the controllable subspace $\mathcal{C}_{t_{\rm f}}(A,B)$ at time $t_{\rm f} > 0$ is independent of $t_{\rm f}$. Hence, we write $\mathcal{C}(A,B)$ for $\mathcal{C}_{t_{\rm f}}(A,B)$ and call $\mathcal{C}(A,B)$ the *controllable subspace* of (A,B). (A,B) is *controllable* if $\mathcal{C}(A,B) = \mathbb{R}^n$. For convenience, define the *controllability matrix*

$$\mathcal{K}(A,B) \triangleq \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \tag{12.5.5}$$

so that

$$\mathcal{C}(A,B) = \mathcal{R}[\mathcal{K}(A,B)]. \tag{12.5.6}$$

Define

$$q \triangleq \dim \mathcal{C}(A,B) = \operatorname{rank}\mathcal{K}(A,B). \tag{12.5.7}$$

The following result shows that the controllable subspace $\mathcal{C}(A,B)$ is unchanged by full-state feedback $u(t) = Kx(t) + v(t)$.

Proposition 12.5.3. Let $K \in \mathbb{R}^{m\times n}$. Then,

$$\mathcal{C}(A+BK,B) = \mathcal{C}(A,B). \tag{12.5.8}$$

In particular, (A,B) is controllable if and only if $(A+BK,B)$ is controllable.

Proof. Note that

$$\begin{aligned}
&\mathcal{C}(A+BK,B)\\
&\quad = \mathcal{R}[\mathcal{K}(A+BK,B)]\\
&\quad = \mathcal{R}\big(\big[\ B \quad AB+BKB \quad A^2B+ABKB+BKAB+BKBKB \quad \cdots\ \big]\big)\\
&\quad = \mathcal{R}[\mathcal{K}(A,B)] = \mathcal{C}(A,B).
\end{aligned}$$

□

Let $\tilde{\mathcal{C}}(A,B) \subseteq \mathbb{R}^n$ be a subspace that is complementary to $\mathcal{C}(A,B)$. Then, $\tilde{\mathcal{C}}(A,B)$ is an *uncontrollable subspace* in the sense that if $x_{\rm f} = x'_{\rm f} + x''_{\rm f} \in \mathbb{R}^n$, where $x'_{\rm f} \in \mathcal{C}(A,B)$ and $x''_{\rm f} \in \tilde{\mathcal{C}}(A,B)$ is nonzero, then there is a continuous control $u\colon\ [0,t_{\rm f}] \to \mathbb{R}^m$ such that $x(t_{\rm f}) = x'_{\rm f}$ but no continuous control such that $x(t_{\rm f}) = x_{\rm f}$. The following result provides a continuous control $u(\cdot)$ that yields $x(t_{\rm f}) = x'_{\rm f}$ for $\tilde{\mathcal{C}}(A,B) \triangleq \mathcal{C}(A,B)^\perp$.

Lemma 12.5.4. Let $t_{\rm f} > 0$, and define $\mathcal{Q} \in \mathbb{R}^{n\times n}$ by

$$\mathcal{Q} \triangleq \left(\int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}\,{\rm d}t\right)^{+} \int_0^{t_{\rm f}} e^{tA}BB^{\rm T}e^{tA^{\rm T}}\,{\rm d}t. \tag{12.5.9}$$

Then, $\mathcal{Q}$ is the projector onto $\mathcal{C}(A,B)$, and $\mathcal{Q}_\perp$ is the projector onto $\mathcal{C}(A,B)^\perp$. Hence,

$$\mathcal{C}(A,B) = \mathcal{R}(\mathcal{Q}) = \mathcal{N}(\mathcal{Q}_\perp), \tag{12.5.10}$$

$$\mathcal{C}(A,B)^\perp = \mathcal{N}(\mathcal{Q}) = \mathcal{R}(\mathcal{Q}), \tag{12.5.11}$$

$$q = \dim \mathcal{C}(A,B) = \operatorname{rank} \mathcal{Q} = \operatorname{def} \mathcal{Q}_\perp, \tag{12.5.12}$$

$$n-q = \dim \mathcal{C}(A,B)^\perp = \operatorname{def} \mathcal{Q} = \operatorname{rank} \mathcal{Q}_\perp. \tag{12.5.13}$$

Now, define $u\colon\ [0,t_{\rm f}] \mapsto \mathbb{R}^m$ by

$$u(t) \triangleq B^{\rm T}e^{(t_{\rm f}-t)A^{\rm T}}\left(\int_0^{t_{\rm f}} e^{\tau A}BB^{\rm T}e^{\tau A^{\rm T}}\,{\rm d}\tau\right)^{+} x_{\rm f}. \tag{12.5.14}$$

If, in addition, $x_{\rm f} = x'_{\rm f} + x''_{\rm f}$, where $x'_{\rm f} \in \mathcal{C}(A,B)$ and $x''_{\rm f} \in \mathcal{C}(A,B)^\perp$, then

$$x'_{\rm f} = \mathcal{Q}x_{\rm f} = \int_0^{t_{\rm f}} e^{(t_{\rm f}-t)A}Bu(t)\,{\rm d}t. \tag{12.5.15}$$

Finally, (A,B) is controllable if and only if $\mathcal{Q} = I_n$. In this case, for all $x_\mathrm{f} \in \mathbb{R}^n$,

$$x_\mathrm{f} = \int_0^{t_\mathrm{f}} e^{(t_\mathrm{f}-t)A} Bu(t)\,\mathrm{d}t, \tag{12.5.16}$$

where $u\colon\ [0,t_\mathrm{f}] \mapsto \mathbb{R}^m$ is defined by

$$u(t) \triangleq B^\mathrm{T} e^{(t_\mathrm{f}-t)A^\mathrm{T}} \left(\int_0^{t_\mathrm{f}} e^{\tau A} BB^\mathrm{T} e^{\tau A^\mathrm{T}}\,\mathrm{d}\tau \right)^{-1} x_\mathrm{f}. \tag{12.5.17}$$

Lemma 12.5.5. Let $\alpha \in \mathbb{R}$. Then,

$$\mathcal{C}(A+\alpha I, B) = \mathcal{C}(A,B). \tag{12.5.18}$$

The following result uses a coordinate transformation to characterize $\mathcal{C}(A,B)$.

Theorem 12.5.6. There exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix} S^{-1}, \qquad B = S\begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \tag{12.5.19}$$

where $A_1 \in \mathbb{R}^{q\times q}$, $B_1 \in \mathbb{R}^{q\times m}$, and (A_1, B_1) is controllable.

Proof. Let $\alpha > 0$ be such that $A_\alpha \triangleq A - \alpha I$ is asymptotically stable, and let $Q \in \mathbb{R}^{n\times n}$ be the nonnegative-semidefinite solution to

$$A_\alpha Q + QA_\alpha^\mathrm{T} + BB^\mathrm{T} = 0 \tag{12.5.20}$$

given by

$$Q = \int_0^\infty e^{tA_\alpha} BB^\mathrm{T} e^{tA_\alpha^\mathrm{T}}\,\mathrm{d}t.$$

It now follows from Lemma 12.5.2 with $t_\mathrm{f} = \infty$ and Lemma 12.5.5

$$\operatorname{rank} Q = \operatorname{rank} \int_0^\infty e^{tA_\alpha} BB^\mathrm{T} e^{tA_\alpha^\mathrm{T}}\,\mathrm{d}t = \dim \mathcal{C}(A_\alpha, B) = \dim \mathcal{C}(A,B) = q,$$

and let $S \in \mathbb{R}^{n\times n}$ be an orthogonal matrix such that $Q = S\left[\begin{smallmatrix} Q_1 & 0 \\ 0 & 0 \end{smallmatrix}\right] S^\mathrm{T}$, where $Q_1 \in \mathbb{R}^{q\times q}$ is positive definite. Writing $A_\alpha = S\left[\begin{smallmatrix} \hat{A}_1 & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_2 \end{smallmatrix}\right] S^{-1}$ and $B = S\left[\begin{smallmatrix} B_1 \\ B_2 \end{smallmatrix}\right]$, where $\hat{A}_1 \in \mathbb{R}^{q\times q}$ and $B_1 \in \mathbb{R}^{q\times m}$, it follows from (12.5.20) that

$$\hat{A}_1 Q_1 + Q_1 \hat{A}_1^{\mathrm{T}} + B_1 B_1^{\mathrm{T}} = 0,$$
$$\hat{A}_{21} Q_1 + B_2 B_1^{\mathrm{T}} = 0,$$
$$B_2 B_2^{\mathrm{T}} = 0.$$

Therefore, $B_2 = 0$ and $\hat{A}_{21} = 0$, and thus

$$A_\alpha = S\begin{bmatrix} \hat{A}_1 & \hat{A}_{12} \\ 0 & \hat{A}_2 \end{bmatrix} S^{-1}, \quad B = S\begin{bmatrix} B_1 \\ 0 \end{bmatrix}.$$

Hence,

$$A = S\begin{bmatrix} \hat{A}_1 & \hat{A}_{12} \\ 0 & \hat{A}_2 \end{bmatrix} S^{-1} + \alpha I = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix} S^{-1},$$

where $A_1 \triangleq \hat{A}_1 + \alpha I_q$, $A_{12} \triangleq \hat{A}_{12}$, and $A_2 \triangleq \hat{A}_2 + \alpha I_{n-q}$. □

Proposition 12.5.7. Let $S \in \mathbb{R}^{n\times n}$ be orthogonal. Then, the following conditions are equivalent:

i) A and B have the form (12.5.19), where $A_1 \in \mathbb{R}^{q\times q}$, $B_1 \in \mathbb{R}^{q\times m}$, and (A_1, B_1) is controllable.

ii) $\mathcal{C}(A,B) = \mathcal{R}\big(S\big[\begin{smallmatrix} I_q \\ 0 \end{smallmatrix}\big]\big)$.

iii) $\mathcal{C}(A,B)^\perp = \mathcal{R}\big(S\big[\begin{smallmatrix} 0 \\ I_{n-q} \end{smallmatrix}\big]\big)$.

iv) $\mathcal{Q} = S\begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix} S^{\mathrm{T}}$.

Proposition 12.5.8. Let $S \in \mathbb{R}^{n\times n}$ be nonsingular. Then, the following conditions are equivalent:

i) A and B have the form (12.5.19), where $A_1 \in \mathbb{R}^{q\times q}$, $B_1 \in \mathbb{R}^{q\times m}$, and (A_1, B_1) is controllable.

ii) $\mathcal{C}(A,B) = \mathcal{R}\big(S\big[\begin{smallmatrix} I_q \\ 0 \end{smallmatrix}\big]\big)$.

iii) $\mathcal{C}(A,B)^\perp = \mathcal{R}\big(S^{-\mathrm{T}}\big[\begin{smallmatrix} 0 \\ I_{n-q} \end{smallmatrix}\big]\big)$.

Definition 12.5.9. Let $\lambda \in \mathrm{spec}(A)$. Then, λ is a *controllable eigenvalue* of (A,B) if

$$\mathrm{rank}\begin{bmatrix} \lambda I - A & B \end{bmatrix} = n. \tag{12.5.21}$$

Otherwise, λ is an *uncontrollable eigenvalue* of (A,B).

Proposition 12.5.10. Let $\lambda \in \mathrm{mspec}(A)$ and $K \in \mathbb{R}^{n\times m}$. Then, λ is a controllable eigenvalue of (A,B) if and only if λ is a controllable eigenvalue of $(A+BK, B)$.

Proposition 12.5.11. Let $\lambda \in \operatorname{mspec}(A)$. Then,

$$\mathcal{C}(A,B) \subseteq \mathcal{R}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right). \tag{12.5.22}$$

Proof. First, note that (12.5.22) is equivalent to

$$\operatorname{Re}\mathcal{R}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right)^{\perp} \subseteq \mathcal{C}(A,B)^{\perp}.$$

Let $x \in \mathcal{R}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right)^{\perp} = \mathcal{N}\left(\begin{bmatrix} \overline{\lambda} I - A^{\mathrm{T}} \\ B^{\mathrm{T}} \end{bmatrix}\right)$ so that $\overline{\lambda}x = A^{\mathrm{T}}x$ and $B^{\mathrm{T}}x = 0$. Now, let $u(\cdot)$ be given by (12.5.14) with $x_{\mathrm{f}} \triangleq \operatorname{Re} x$. Then,

$$\int_0^{t_{\mathrm{f}}} e^{(t_{\mathrm{f}}-t)A}Bu(t)\,\mathrm{d}t = \mathcal{Q}x_{\mathrm{f}} = 0,$$

which implies that $\operatorname{Re} x = x_{\mathrm{f}} \in \mathcal{C}(A,B)^{\perp}$. □

The next result characterizes controllability in several equivalent ways.

Theorem 12.5.12. The following statements are equivalent:

i) (A,B) is controllable.

ii) There exists $t > 0$ such that $\int_0^t e^{tA}BB^{\mathrm{T}}e^{tA^{\mathrm{T}}}\,\mathrm{d}t$ is positive definite.

iii) $\int_0^{t_{\mathrm{f}}} e^{tA}BB^{\mathrm{T}}e^{tA^{\mathrm{T}}}\,\mathrm{d}t$ is positive definite for all $t > 0$.

iv) $\operatorname{rank}\mathcal{K}(A,B) = n$.

v) Every eigenvalue of (A,B) is controllable.

vi) For every self-conjugate multiset $\{\lambda_1,\ldots,\lambda_n\}_{\mathrm{m}} \subset \mathbb{C}$ there exists a matrix $K \in \mathbb{R}^{m\times n}$ such that $\operatorname{mspec}(A+BK) = \{\lambda_1,\ldots,\lambda_n\}_{\mathrm{m}}$.

Proof. The equivalence of *i)*–*iv)* follows from Lemma 12.5.2. To prove that *iv)* implies *v)*, suppose that *v)* does not hold, that is, there exists $\lambda \in \operatorname{spec}(A)$ and a nonzero vector $x \in \mathbb{C}^n$ such that $x^{\mathrm{T}}A = \lambda x^{\mathrm{T}}$ and $x^{\mathrm{T}}B = 0$. It thus follows that $x^{\mathrm{T}}AB = \lambda x^{\mathrm{T}}B = 0$. Similarly, we obtain $x^{\mathrm{T}}A^iB = 0$ for all $i = 0,1,\ldots,n-1$. Hence, $\dim\mathcal{C}(A,B) < n$.

Conversely, to show that *v)* $\Longrightarrow$ *iv)*, suppose that $\operatorname{rank}\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} < n$. Then, there exists nonzero $x \in \mathbb{R}^n$ such that $x^{\mathrm{T}}A^iB = 0$ for all $i = 0,\ldots,n-1$. Now, let $p \in \mathbb{R}[s]$ be a nonzero polynomial of minimal degree such that $x^{\mathrm{T}}p(A) = 0$. Note that p is not a constant polynomial and that $x^{\mathrm{T}}\mu_A(A) = 0$. Thus, $1 \le \deg p \le \deg \mu_A$. Now, let $\lambda \in \mathbb{C}$ be such that $p(\lambda) = 0$, and let $q \in \mathbb{R}[s]$ be such that $p(s) = q(s)(s-\lambda)$ for all $s \in \mathbb{C}$. Since $\deg q < \deg p$, it follows that $x^{\mathrm{T}}q(A) \ne 0$. Therefore, $\eta \triangleq q(A)x$ is nonzero. Furthermore, $\eta^{\mathrm{T}}(A-\lambda I) = x^{\mathrm{T}}p(A) = 0$. Since $x^{\mathrm{T}}A^iB = 0$ for all $i = 0,\ldots,n-1$, it follows that $\eta^{\mathrm{T}}B = x^{\mathrm{T}}q(A)B = 0$. Consequently, *v)* does not hold.

The equivalence of *v*) and *vi*) is immediate.

To prove that *i*) implies *vi*), assume that $m = 1$, and let $A_c = C(\chi_A)$ and $B_c = e_n$. Then, Proposition 12.8.3 implies that $\mathcal{K}(A_c, B_c)$ is nonsingular, while Proposition 12.8.6 implies that $A_c = S^{-1}AS$ and $B_c = S^{-1}B$. Now, let $\{\lambda_1, \ldots, \lambda_n\}_m \subset \mathbb{C}$ be self conjugate and define $p \in \mathbb{R}[s]$ by $p(s) \triangleq \Pi_{i=1}^n (s - \lambda_i)$. Letting $K \triangleq e_n^T[C(p) - A_c]S^{-1}$ it follows that

$$\begin{aligned} A + BK &= S(A_c + B_c KS)S^{-1} \\ &= S(A_c + B_c e_n^T[C(p) - A_c])S^{-1} \\ &= SC(p)S^{-1}. \end{aligned}$$

See [494, p. 248] for the case $m > 1$. See wonham/kailath.

Conversely, to show that *vii*) implies *i*), suppose that (A, B) is not controllable. Then, it follows from Proposition 12.5.8 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that A and B have the form given by *i*) of Proposition 12.5.8. Since the eigenvalues of A_2 are not affected by $K \in \mathbb{R}^{m\times n}$, it follows that *vi*) does not hold. □

12.6 Stabilizability

Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $q \triangleq \dim \mathcal{C}(A, C)$.

Definition 12.6.1. (A, B) is *stabilizable* if

$$\mathcal{S}_u(A) \subseteq \mathcal{C}(A, B). \tag{12.6.1}$$

Proposition 12.6.2. Let $K \in \mathbb{R}^{m\times n}$. Then, (A, B) is stabilizable if and only if $(A + BK, B)$ is stabilizable.

Proposition 12.6.3. The following statements are equivalent:

i) (A, B) is stabilizable.

ii) There exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}S^{-1}, \quad B = S\begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \tag{12.6.2}$$

where $A_1 \in \mathbb{R}^{q\times q}$, $B_1 \in \mathbb{R}^{q\times m}$, (A_1, B_1) is controllable, and $A_2 \in \mathbb{R}^{(n-q)\times(n-q)}$ is asymptotically stable.

iii) Every CRHP eigenvalue of (A, B) is controllable.

iv) $(A + BK, B)$ is stabilizable for all $K \in \mathbb{R}^{m \times n}$.

Proof. First assume that (A, B) is stabilizable so that $\mathcal{S}_{\mathrm{u}}(A) = \mathcal{N}[\mu_A^{\mathrm{u}}(A)] = \mathcal{R}[\mu_A^{\mathrm{s}}(A)] \subseteq \mathcal{C}(A, B)$. Using Proposition 12.5.8 it follows that there exists an orthogonal matrix $S \in \mathbb{R}^{n \times n}$ such that (12.6.2) is satisfied, where $A_1 \in \mathbb{R}^{q \times q}$ and (A_1, B_1) is controllable. Thus, $\mathcal{R}[\mu_A^{\mathrm{s}}(A)] \subseteq \mathcal{C}(A, B) = \mathcal{R}\big(S\left[\begin{smallmatrix} I_q \\ 0 \end{smallmatrix}\right]\big)$.

Next, note that

$$\mu_A^{\mathrm{s}}(A) = S \begin{bmatrix} \mu_A^{\mathrm{s}}(A_1) & B_{12\mathrm{s}} \\ 0 & \mu_A^{\mathrm{s}}(A_2) \end{bmatrix} S^{-1},$$

where $B_{12\mathrm{s}} \in \mathbb{R}^{q \times (n-q)}$, and suppose that A_2 is not asymptotically stable with CRHP eigenvalue λ. Then, $\lambda \notin \mathrm{roots}(\mu_A^{\mathrm{s}})$, and thus $\mu_A^{\mathrm{s}}(A_2) \neq 0$. Let $x_2 \in I_{n-q}$ satisfy $\mu_A^{\mathrm{s}}(A_2)x_2 \neq 0$. Then,

$$\mu_A^{\mathrm{s}}(A)S \begin{bmatrix} 0 \\ x_2 \end{bmatrix} \notin \mathcal{R}\left(S \begin{bmatrix} I_q \\ 0 \end{bmatrix}\right),$$

which implies that $\mathcal{S}_{\mathrm{u}}(A)$ is not contained in $\mathcal{C}(A, B)$. Hence, A_2 is asymptotically stable.

Conversely, assume that there exists a nonsingular matrix $S \in \mathbb{R}^{n \times n}$ such that (12.6.2) is satisfied, where $A_1 \in \mathbb{R}^{q \times q}$, $q = \dim \mathcal{C}(A, B)$, and $A_2 \in \mathbb{R}^{(n-q) \times (n-q)}$ is asymptotically stable. Using Fact 11.16.4 it follows that $\mathcal{S}_{\mathrm{u}}(A) \subseteq \mathcal{R}\big(S\left[\begin{smallmatrix} I_q \\ 0 \end{smallmatrix}\right]\big) = \mathcal{C}(A, B)$, which implies that (A, B) is stabilizable. $\square$

Lemma 12.6.4. Assume that (A, B) is stabilizable and

$$Q \triangleq \int_0^\infty e^{tA} B B^{\mathrm{T}} e^{tA^{\mathrm{T}}} \, \mathrm{d}t$$

exists. Then, A is asymptotically stable.

Proof. Since (A, B) is stabilizable, it follows from Proposition 12.3.3 that there exists a nonsingular matrix $S \in \mathbb{R}^{n \times n}$ such that $A = S\left[\begin{smallmatrix} A_1 & 0 \\ A_{21} & A_2 \end{smallmatrix}\right]S^{-1}$ and $C = \begin{bmatrix} C_1 & 0 \end{bmatrix} S^{-1}$, where $A_1 \in \mathbb{R}^{p \times p}$, (A_1, C_1) is observable, and A_2 is asymptotically stable. Thus, the integral

$$\int_0^\infty e^{tA^{\mathrm{T}}} C^{\mathrm{T}} C e^{tA} \, \mathrm{d}t = S \begin{bmatrix} \int_0^\infty e^{tA_1^{\mathrm{T}}} C_1^{\mathrm{T}} C_1 e^{tA_1} \, \mathrm{d}t & 0 \\ 0 & 0 \end{bmatrix} S^{-1}$$

exists. Now, suppose that A is not asymptotically stable so that A_1 is not asymptotically stable. Let $\lambda \in \mathrm{spec}(A_1) \cap \mathrm{CRHP}$, and let $x_1 \in \mathbb{C}^p$ be an associated eigenvector. Since (A_1, C_1) is observable, it follows from Proposition 8.5.3 and *iii)* of Theorem 12.2.12 that $\int_0^\infty e^{tA_1^{\mathrm{T}}} C_1^{\mathrm{T}} C_1 e^{tA_1} \, \mathrm{d}t$ is positive definite. Consequently,

$$\alpha \triangleq x_1^* \int_0^\infty e^{tA_1^{\mathrm{T}}} C_1^{\mathrm{T}} C_1 e^{tA_1}\, \mathrm{d}t x_1$$

is positive. However, we also have that

$$\alpha = x_1^* \int_0^\infty e^{\overline{\lambda}t} C_1^{\mathrm{T}} C_1 e^{\lambda t}\, \mathrm{d}t x_1 = x_1^* C_1^{\mathrm{T}} C_1 x_1 \int_0^\infty e^{2(\mathrm{Re}\,\lambda)t}\, \mathrm{d}t.$$

Since $\mathrm{Re}\,\lambda \geq 0$, it follows that $\int_0^\infty e^{2(\mathrm{Re}\,\lambda)t}\,\mathrm{d}t = \infty$, which contradicts the fact that α is finite. □

12.7 Controllable Asymptotic Stability

Definition 12.7.1. (A,B) is *controllably asymptotically stable* if

$$\mathcal{C}(A,B) \subseteq \mathcal{S}_{\mathrm{s}}(A). \tag{12.7.1}$$

Proposition 12.7.2. Let $K \in \mathbb{R}^{m\times n}$. Then, (A,B) is controllably asymptotically stable if and only if $(A+BK, B)$ is controllably asymptotically stable.

Lemma 12.7.3. Assume that the nonnegative-semidefinite matrix $P \in \mathbb{R}^{n\times n}$ defined by

$$Q \triangleq \int_0^\infty e^{tA} BB^{\mathrm{T}} e^{tA^{\mathrm{T}}}\, \mathrm{d}t \tag{12.7.2}$$

exists. Then, Q satisfies

$$AQ + QA^{\mathrm{T}} + BB^{\mathrm{T}} = 0. \tag{12.7.3}$$

Proposition 12.7.4. The following statements are equivalent:

i) (A,B) is controllably asymptotically stable.

ii) There exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ and $k \in \mathbb{N}$ such that

$$A = S\begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix} S^{-1}, \quad B = S\begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \tag{12.7.4}$$

where $A_1 \in \mathbb{R}^{k\times k}$ is asymptotically stable and $B_1 \in \mathbb{R}^{k\times m}$.

iii) $\lim_{t\to\infty} e^{tA}B = 0$.

iv) $Q \triangleq \int_0^\infty e^{tA} BB^{\mathrm{T}} e^{tA^{\mathrm{T}}}\mathrm{d}t$ exists.

v) There exists a nonnegative-semidefinite matrix $Q \in \mathbb{R}^{n\times n}$ satisfying (12.7.3).

In this case, one nonnegative-semidefinite solution is given by (12.7.2). Furthermore,

$$\mathcal{Q} = QQ^+, \tag{12.7.5}$$

$$\mathcal{R}(\mathcal{Q}) = \mathcal{R}(Q) = \mathcal{C}(A,B), \tag{12.7.6}$$

$$\mathcal{R}(\mathcal{Q}) = \mathcal{R}(Q) = \mathcal{C}(A,B)^\perp, \tag{12.7.7}$$

$$\operatorname{rank}\mathcal{Q} = \operatorname{rank}Q = q. \tag{12.7.8}$$

Proof. To prove that *i*) implies *ii*), assume that (A,C) is controllably asymptotically stable. It then follows that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $A = S\big[\begin{smallmatrix} A_1 & 0 \\ A_{21} & A_2 \end{smallmatrix}\big]S^{-1}$ and $C = \begin{bmatrix} C_1 & 0 \end{bmatrix} S^{-1}$, where A_1 is asymptotically stable. Thus,

$$Ce^{tA} = \begin{bmatrix} C_1 e^{tA_1} & 0 \end{bmatrix} S \to 0$$

as $t \to \infty$. Next, to prove that *ii*) implies *iii*), assume that $Ce^{tA} \to 0$ as $t \to \infty$. Then, every entry of Ce^{tA} involves exponentials of t, where the coefficients of t have negative real part. Hence, so does every entry of $e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}$, which implies that $\int_0^\infty e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t$ exists. To prove that *iii*) implies *iv*), assume that $P = \int_0^\infty e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t$ exists. Then, $e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA} \to 0$ as $t \to \infty$, and thus

$$\begin{aligned}
A^{\mathrm{T}}P + PA &= \int_0^\infty \Big[A^{\mathrm{T}}e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA} + e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}A\Big]\,\mathrm{d}t \\
&= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA}\,\mathrm{d}t \\
&= \lim_{t\to\infty} e^{tA^{\mathrm{T}}}C^{\mathrm{T}}Ce^{tA} - C^{\mathrm{T}}C \\
&= -C^{\mathrm{T}}C,
\end{aligned}$$

which shows that P satisfies (12.4.3).

To prove that *iv*) implies *i*), suppose that there exists a nonnegative-semidefinite matrix $P \in \mathbb{R}^{n\times n}$ satisfying (12.4.3). Then,

$$\begin{aligned}\int_0^t e^{\tau A^{\mathrm{T}}} C^{\mathrm{T}} C e^{\tau A}\,\mathrm{d}\tau &= -\int_0^t e^{\tau A^{\mathrm{T}}}\big(A^{\mathrm{T}}P + PA\big)e^{\tau A}\,\mathrm{d}\tau \\ &= -\int_0^t \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\tau A^{\mathrm{T}}} P e^{\tau A}\,\mathrm{d}\tau \\ &= P - e^{tA^{\mathrm{T}}} P e^{tA} \\ &\le P.\end{aligned}$$

Next, it follows from Proposition 12.5.7 that there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that $A = S\left[\begin{smallmatrix} A_1 & 0 \\ A_{21} & A_2 \end{smallmatrix}\right]S^{\mathrm{T}}$ and $C = \left[\begin{array}{cc} C_1 & 0 \end{array}\right] S^{\mathrm{T}}$, where (A_1, C_1) is observable. Consequently, we have

$$\begin{aligned}\int_0^t e^{\tau A_1^{\mathrm{T}}} C_1^{\mathrm{T}} C_1 e^{\tau A_1}\,\mathrm{d}\tau &= \left[\begin{array}{cc} I & 0 \end{array}\right] S \int_0^t e^{\tau A^{\mathrm{T}}} C^{\mathrm{T}} C e^{\tau A}\,\mathrm{d}\tau S^{\mathrm{T}} \left[\begin{array}{c} I \\ 0 \end{array}\right] \\ &\le \left[\begin{array}{cc} I & 0 \end{array}\right] S P S^{\mathrm{T}} \left[\begin{array}{c} I \\ 0 \end{array}\right].\end{aligned}$$

Thus, it follows from Proposition 8.5.3 that $\int_0^\infty e^{tA_1^{\mathrm{T}}} C_1^{\mathrm{T}} C_1 e^{tA_1}\,\mathrm{d}t$ exists. Since (A_1, C_1) is observable, it follows from Lemma 12.4.3 that A_1 is asymptotically stable. Therefore, (A_1, C_1) is controllably asymptotically stable. □

Proposition 12.7.5. The following statements are equivalent:

i) A is asymptotically stable.

ii) (A, B) is stabilizable and controllably asymptotically stable.

Furthermore, if two of the following three conditions are satisfied, then the third condition is satisfied:

iii) A is asymptotically stable.

iv) (A, B) is controllable.

v) $\int_0^\infty e^{tA} B B^{\mathrm{T}} e^{tA^{\mathrm{T}}}\,\mathrm{d}t$ exists and is positive definite.

12.8 Realization Theory

Given a proper rational transfer function G we wish to determine (A, B, C, D) such that (12.1.35) holds. The following terminology is standard.

Definition 12.8.1. Let $G \in \mathbb{R}^{l\times m}(s)$. If $l = m = 1$, then G is a *single-input/single-output (SISO)* rational transfer function; if $l = 1$ and $m > 1$, then G is a *multiple-input/single-output (MISO)* rational transfer function; if $l > 1$ and $m = 1$, then G is a *single-input/multiple-output (SIMO)* rational transfer function; and, if $l > 1$ and $m > 1$, then G is a *multiple-input/multiple output (MIMO)* rational transfer function.

Definition 12.8.2. Let $G \in \mathbb{R}^{l\times m}(s)$ be proper, and assume that $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{l\times n}$, and $D \in \mathbb{R}^{l\times m}$ satisfy $G(s) = C(sI-A)^{-1}B+D$. Then, $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is a *realization* of G, which is written as

$$G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{12.8.1}$$

The *order* of the realization $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is the order of A.

Although realizations are not unique, the matrix D is unique and is given by

$$D = G(\infty). \tag{12.8.2}$$

Furthermore, note that $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ if and only if $\hat{G} \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$, where $\hat{G} \triangleq G - D$. Therefore, it suffices to construct realizations for strictly proper transfer functions.

Suppose that $n = 0$. Then, A, B, and C are empty matrices, and G is given by

$$G(s) = 0_{l\times 0}(sI_{0\times 0} - 0_{0\times 0})^{-1}0_{0\times m} + D = 0_{l\times m} + D = D. \tag{12.8.3}$$

Therefore, the order of the realization $\left[\begin{array}{c|c} 0_{0\times 0} & 0_{0\times m} \\ \hline 0_{l\times 0} & D \end{array}\right]$ is zero.

The following result shows that every strictly proper, SISO rational transfer function has a realization. In fact, two realizations are the *controllable canonical form* and the *observable canonical form* given by (12.8.6) and (12.8.8), respectively.

Proposition 12.8.3. Let $G \in \mathbb{R}(s)$ be strictly proper and given by

$$G(s) = \frac{\alpha_{n-1}s^{n-1} + \alpha_{n-2}s^{n-2} + \cdots + \alpha_1 s + \alpha_0}{s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0}. \tag{12.8.4}$$

Then, $G \sim \left[\begin{array}{c|c} A_{\rm c} & B_{\rm c} \\ \hline C_{\rm c} & 0 \end{array}\right]$, where $A_{\rm c}, B_{\rm c}, C_{\rm c}$ are given by

$$A_{\rm c} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\beta_0 & -\beta_1 & -\beta_2 & \cdots & -\beta_{n-1} \end{bmatrix}, \quad B_{\rm c} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \tag{12.8.5}$$

$$C_{\rm c} = \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} \end{bmatrix}, \tag{12.8.6}$$

and $G \sim \left[\begin{array}{c|c} A_{\rm o} & B_{\rm o} \\ \hline C_{\rm o} & 0 \end{array}\right]$, where $A_{\rm o}, B_{\rm o}, C_{\rm o}$ are given by

$$A_{\rm o} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\beta_0 \\ 1 & 0 & \cdots & 0 & -\beta_1 \\ 0 & 1 & \cdots & 0 & -\beta_2 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -\beta_{n-1} \end{bmatrix}, \quad B_{\rm o} = \begin{bmatrix} \alpha_0 \\ \alpha_2 \\ \vdots \\ \alpha_{n-1} \end{bmatrix}, \tag{12.8.7}$$

$$C_{\rm o} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{12.8.8}$$

Furthermore, $(A_{\rm c}, B_{\rm c})$ is controllable and $(A_{\rm o}, C_{\rm o})$ is observable.

Proof. The realizations can be verified directly. Furthermore, note that

$$\mathcal{C}(A_{\rm c}, B_{\rm c}) = \mathcal{O}(A_{\rm o}, C_{\rm o}) = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \iddots & \iddots & -\beta_{n-1} \\ 0 & 0 & 0 & \iddots & & \vdots \\ 0 & 0 & 1 & \cdots & \cdots & -\beta_2 \\ 0 & 1 & -\beta_{n-1} & \cdots & \cdots & -\beta_1 \\ 1 & -\beta_{n-1} & -\beta_{n-2} & \cdots & \cdots & -\beta_0 \end{bmatrix}.$$

Using Fact 2.12.20 it follows that $\det \mathcal{C}(A_{\rm c}, B_{\rm c}) = \det \mathcal{O}(A_{\rm o}, C_{\rm o}) = (-1)^{\lfloor n/2 \rfloor}$, which shows that $(A_{\rm c}, B_{\rm c})$ is controllable and $(A_{\rm o}, C_{\rm o})$ is observable. $\square$

The following result shows that every proper rational transfer function has a realization.

Theorem 12.8.4. Let $G \in \mathbb{R}^{l \times m}(s)$ be proper. Then, there exist $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{l \times n}$, and $D \in \mathbb{R}^{l \times m}$ such that $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$.

Proof. By Proposition 12.8.3, every entry $G_{(i,j)}$ of G has a realization $G_{(i,j)} \sim \left[\begin{array}{c|c} A_{ij} & B_{ij} \\ \hline C_{ij} & D_{ij} \end{array}\right]$. Combining these realizations yields a realization of

$G.$ $\square$

Let $G \in \mathbb{R}^{l\times m}(s)$, and let $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ be a realization of G, where $A \in \mathbb{R}^{n\times n}$. If $S \in \mathbb{R}^{n\times n}$ is nonsingular, then $\left[\begin{array}{c|c} SAS^{-1} & SB \\ \hline CS^{-1} & D \end{array}\right]$ is also a realization of G.

Definition 12.8.5. Let $G \in \mathbb{R}^{l\times m}(s)$ be proper, and let $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ and $\left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & D \end{array}\right]$ be nth-order realizations of G. Then, $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ and $\left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & D \end{array}\right]$ are *equivalent* if there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $\hat{A} = SAS^{-1}$, $\hat{B} = SB$, and $\hat{C} = CS^{-1}$.

Proposition 12.8.6. Let $G \in \mathbb{R}(s)$ be SISO and strictly proper with nth-order realization $\left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$. If (A, B) is controllable, then there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $\left[\begin{array}{c|c} SAS^{-1} & SB \\ \hline CS^{-1} & 0 \end{array}\right]$ is in controllable companion form. Furthermore, if (A, C) is observable, then there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $\left[\begin{array}{c|c} SAS^{-1} & SB \\ \hline CS^{-1} & 0 \end{array}\right]$ is in observable companion form.

Proof. Defining $S \triangleq \mathcal{K}(A, B)[\mathcal{K}(A_{\rm c}, B_{\rm c})]^{-1}$, it follows that $SAS^{-1} = C(\chi_A)$ and $S^{-1}B = e_n$. Alternatively, defining $S \triangleq [\mathcal{O}(A_{\rm o}, C_{\rm o})]^{-1}\mathcal{O}(A_{\rm o}, C_{\rm o})$, it follows that $SAS^{-1} = C(\chi_A)^{\rm T}$ and $CS^{-1} = e_n^{\rm T}$. $\square$

Proposition 12.8.7. Let $G \in \mathbb{R}^{l\times m}(s)$ be proper and have controllable and observable realizations $G \sim \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D \end{array}\right]$ and $G \sim \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D \end{array}\right]$. Then, these realizations are equivalent.

Proof. For the SISO case $l = m = 1$, the result is an immediate consequence of Proposition 12.8.6. In the MIMO case, for $i = 1, 2$ define $\mathcal{K}_i \triangleq \mathcal{K}(A_i, B_i)$, $\mathcal{O}_i \triangleq \mathcal{O}(A_i, C_i)$, and $S \triangleq (\mathcal{O}_2^{\rm T}\mathcal{O}_2)^{-1}\mathcal{O}_2^{\rm T}\mathcal{O}_1$. Then, $S^{-1} = \mathcal{K}_1\mathcal{K}_2^{\rm T}(\mathcal{K}_2\mathcal{K}_2^{\rm T})^{-1}$ and it follows that $A_2 = SA_1S^{-1}$, $B_2 = SB_1$, and $C_2 = C_1S^{-1}$. NEEDS TO BE CHECKED $\square$

A rational transfer function $G \in \mathbb{R}^{l\times m}(s)$ can have realizations of different orders. For example, letting

$$A = 1, \qquad B = 1, \qquad C = 1, \qquad D = 0,$$

and

$$\hat{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \hat{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \hat{D} = 0,$$

it follows that

$$G(s) = C(sI - A)^{-1}B + D = \hat{C}(sI - \hat{A})^{-1}\hat{B} + \hat{D} = \frac{1}{s-1}.$$

Generally, it is desirable to find realizations whose order is as small as possible.

Definition 12.8.8. Let $G \in \mathbb{R}^{l\times m}(s)$ be proper. Then, the realization $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is a *minimal realization* of G if its order is less than or equal to the order of every realization of G. In this case, we write

$$G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{12.8.9}$$

Note that minimality of a realization is independent of D. The following result is useful for constructing minimal realizations.

Proposition 12.8.9. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$A = S\begin{bmatrix} A_1 & 0 & A_{13} & 0 \\ A_{21} & A_2 & A_{23} & A_{24} \\ 0 & 0 & A_3 & 0 \\ 0 & 0 & A_{43} & A_4 \end{bmatrix}S^{-1}, \quad B = S\begin{bmatrix} B_1 \\ B_2 \\ 0 \\ 0 \end{bmatrix}, \tag{12.8.10}$$

$$C = \begin{bmatrix} C_1 & 0 & C_3 & 0 \end{bmatrix}S^{-1}, \tag{12.8.11}$$

where (A_1, B_1) and (A_2, B_2) are controllable, and (A_1, C_1) and (A_3, C_3) are observable. Furthermore, $G \sim \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$.

Proof. The result is obtained by combining Proposition 12.5.7 and Proposition 12.5.8. More directly, it follows from Theorem 8.3.4 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that the controllability and observability Gramians (12.7.2) and (12.4.2) satisfy

$$Q = S\begin{bmatrix} Q_1 & & & 0 \\ & Q_2 & & \\ & & 0 & \\ 0 & & & 0 \end{bmatrix}S^{\mathrm{T}}, \quad P = S^{-\mathrm{T}}\begin{bmatrix} P_1 & & & 0 \\ & 0 & & \\ & & P_2 & \\ 0 & & & 0 \end{bmatrix}S^{-1},$$

where Q_1, Q_2, P_1 and P_2 are positive definite and diagonal. The form of SAS^{-1}, SB, and CS^{-1} given by (12.8.11) now follows from (12.4.3) and (12.7.3). Finally, it can be verified directly that $\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$ is a realization of G. □

The following result show that the controllable and observable realization $\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$ of G in Proposition 12.8.9 is, in fact, minimal.

Corollary 12.8.10. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \in \mathbb{R}^{l\times m}(s)$. Then, $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is minimal if and only if it is controllable and observable.

Proof. To prove necessity, suppose that $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is either not controllable or not observable. Then, Proposition 12.8.3 can be used to construct a realization of G of order less than n. Hence, $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is not minimal. Sufficiency is proved in [494, pp. 172, 173] or [572, p. 50]. □

Theorem 12.8.11. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$. Then, the McMillan degree of G is equal to the order of A.

Proof. See ????. □

Definition 12.8.12. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Then, G is (asymptotically stable, semistable, Lyapunov stable) if A is.

Proposition 12.8.13. Let $G \in \mathbb{R}^{l\times m}(s)$. Then, G is (asymptotically stable, semistable, Lyapunov stable) if and only if every entry of G has the same property.

Definition 12.8.14. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ and A is asymptotically stable. Then, the realization $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ is *semi-balanced* if the controllability and observability Gramians (12.4.2) and (12.7.2) are diagonal, and *balanced* if they are diagonal and equal.

Proposition 12.8.15. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ and A is asymptotically stable. If, in addition, $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, then there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that the realization $G \sim \left[\begin{array}{c|c} SAS^{-1} & SB \\ \hline CS^{-1} & D \end{array}\right]$ is semi-balanced.

Proof. It follows from Corollary 8.3.7 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that SQS^{T} and $S^{-\mathrm{T}}PS^{-1}$ are diagonal, where Q and P are the controllability and observability Gramians. Hence, the realization $\left[\begin{array}{c|c} SAS^{-1} & SB \\ \hline CS^{-1} & D \end{array}\right]$ is semi-balanced. □

12.9 System Zeros

Recall Definition 4.2.4 on the rank of a matrix polynomial.

Definition 12.9.1. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Then, the *Rosenbrock system matrix* $\mathcal{Z} \in \mathbb{R}^{(n+l)\times(n+m)}[s]$ is the polynomial matrix

$$\mathcal{Z}(s) \triangleq \begin{bmatrix} sI - A & B \\ C & D \end{bmatrix}. \tag{12.9.1}$$

Furthermore, $z \in \mathbb{C}$ is an *invariant zero* of the realization $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ if

$$\operatorname{rank} \mathcal{Z}(z) < \operatorname{rank} \mathcal{Z}. \tag{12.9.2}$$

It is easy to see that equivalent realizations have the same invariant zeros. Furthermore, invariant zeros are not changed by full-state feedback. To see this, let $u = Kx + v$, which leads to the rational transfer function

$$G_K \sim \left[\begin{array}{c|c} A + BK & B \\ \hline C + DK & D \end{array}\right]. \tag{12.9.3}$$

Since

$$\begin{bmatrix} zI - (A + BK) & B \\ -(C + DK) & D \end{bmatrix} = \begin{bmatrix} zI - A & B \\ -C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -K & I \end{bmatrix}, \tag{12.9.4}$$

it follows that $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ and $\left[\begin{array}{c|c} A + BK & B \\ \hline C + DK & D \end{array}\right]$ have the same invariant zeros.

Proposition 12.9.2. Let $G \in \mathbb{R}^{l\times m}(s)$, where $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, and assume that $C^{\mathrm{T}}D = 0$ and $D^{\mathrm{T}}D$ is positive definite. Then, the following statements hold:

i) $\operatorname{rank} \mathcal{Z} = n + m$.

ii) $z \in \mathbb{C}$ is an invariant zero of $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ if and only if z is an unobservable eigenvalue of (A, C).

Proof. To prove *i)*, assume that $\operatorname{rank} \mathcal{Z} < n + m$. Then, for every $s \in \mathbb{C}$, there exists nonzero $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathcal{N}(\mathcal{Z}(s))$, that is,

$$\begin{bmatrix} sI - A & B \\ -C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0.$$

Consequently, $-Cx + Dy = 0$, which implies that $-D^{\mathrm{T}}Cx + D^{\mathrm{T}}Dy = 0$ and thus $y = 0$. Furthermore, since $(sI - A)x = 0$, it follows that choosing $s \notin \operatorname{spec}(A)$ yields $x = 0$, which is a contradiction. To prove *ii)*, note that

z is an invariant zero of $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ if and only if $\operatorname{rank} \mathcal{Z}(z) < n+m$, which holds if and only if there exists nonzero $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathcal{N}(\mathcal{Z}(z))$. This condition is equivalent to $y = 0$ and $\left[\begin{smallmatrix} zI-A \\ -C \end{smallmatrix}\right] x = 0$. Since $x \neq 0$, this last condition is equivalent to the fact that z is an unobservable eigenvalue of (A, C). □

Corollary 12.9.3. Let (A, C) be observable and assume that $C^{\mathrm{T}}D = 0$ and $D^{\mathrm{T}}D$ is positive definite. Then, $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$ has no invariant zeros.

Definition 12.9.4. Let $G \in \mathbb{R}^{p \times m}$. Then, $z \in \mathbb{C}$ is a *transmission zero* of G if $\operatorname{rank} G(z) < \operatorname{rank} G$.

Proposition 12.9.5. Let $G \in \mathbb{R}^{p \times m}(s)$, where $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. If $s \notin \operatorname{spec}(A)$, then

$$\operatorname{rank} \mathcal{Z}(s) = n + \operatorname{rank} G(s). \tag{12.9.5}$$

Furthermore,

$$\operatorname{rank} \mathcal{Z} = n + \operatorname{rank} G. \tag{12.9.6}$$

Proof. Since $s \notin \operatorname{spec}(A)$, it follows that

$$\begin{bmatrix} sI-A & B \\ -C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ -C(sI-A)^{-1} & I \end{bmatrix} \begin{bmatrix} sI-A & B \\ 0 & G(s) \end{bmatrix},$$

which implies (12.9.5) and (12.9.6). □

Theorem 12.9.6. Let $G \in \mathbb{R}^{p \times m}(s)$, where $G \overset{\min}{\sim} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, and let $z \notin \operatorname{spec}(A)$. Then, z is a transmission zero of G if and only if z is an invariant zero of $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$.

Proof. Let $z \notin \operatorname{spec}(A)$ be a transmission zero of G. Then,

$$\operatorname{rank} \mathcal{Z}(z) = n + \operatorname{rank} G(z) < n + \operatorname{rank} G = \operatorname{rank} \mathcal{Z},$$

which implies that z is an invariant zero of $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Conversely, let $z \notin \operatorname{spec}(A)$ be an invariant zero of $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Then,

$$\operatorname{rank} G(z) = \operatorname{rank} \mathcal{Z}(z) - n < \operatorname{rank} \mathcal{S} - n = \operatorname{rank} G,$$

which implies that z is a transmission zero of G. □

12.10 H_2 System Norm

Consider the system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.10.1}$$

$$y(t) = Cx(t), \tag{12.10.2}$$

where $A \in \mathbb{R}^{n\times n}$ is asymptotically stable, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{l\times n}$. Then, for all $t \ge 0$, the *impulse response function* is given by $H(t) = Ce^{tA}B$. The L_2 *norm* of $H(\cdot)$ is given by

$$\|H\|_{\mathrm{L}_2} \triangleq \left[\int_0^\infty \|H(t)\|_{\mathrm{F}}^2 \,\mathrm{d}t\right]^{1/2}. \tag{12.10.3}$$

The following result provides expressions for $\|H(\cdot)\|_{\mathrm{L}_2}$ in terms of the controllability and observability Gramians.

Theorem 12.10.1. Let $H(t) = Ce^{tA}B$, where A is asymptotically stable. Then, the L_2 norm of H is given by

$$\|H\|_{\mathrm{L}_2}^2 = \operatorname{tr} CQC^{\mathrm{T}} = \operatorname{tr} B^{\mathrm{T}}PB, \tag{12.10.4}$$

where $Q, P \in \mathbb{R}^{n\times n}$ satisfy

$$AQ + QA^{\mathrm{T}} + BB^{\mathrm{T}} = 0, \tag{12.10.5}$$

$$A^{\mathrm{T}}P + PA + C^{\mathrm{T}}C = 0. \tag{12.10.6}$$

Proof. Note that

$$\|H\|_{\mathrm{L}_2}^2 = \int_0^\infty \operatorname{tr} Ce^{tA}BB^{\mathrm{T}}e^{tA^{\mathrm{T}}}C^{\mathrm{T}}\mathrm{d}t = \operatorname{tr} CQC^{\mathrm{T}},$$

where Q satisfies (12.10.5). The dual expression follows in a similar manner or by noting that

$$\begin{aligned}\operatorname{tr} CQC^{\mathrm{T}} &= \operatorname{tr} C^{\mathrm{T}}CQ = -\operatorname{tr}\left(A^{\mathrm{T}}P + PA\right)Q \\ &= -\operatorname{tr}\left(AQ + QA^{\mathrm{T}}\right)P = \operatorname{tr} BB^{\mathrm{T}}P = \operatorname{tr} B^{\mathrm{T}}PB.\end{aligned}$$

□

For the following definition note that

$$\|G(s)\|_{\mathrm{F}} = [\operatorname{tr} G(s)G^*(s)]^{1/2}. \tag{12.10.7}$$

Definition 12.10.2. The H_2 *norm* of $G \in \mathbb{R}^{l\times m}(s)$ is the nonnegative number

$$\|G\|_{\mathrm{H}_2} \triangleq \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \|G(\jmath\omega)\|_{\mathrm{F}}^2 \,\mathrm{d}\omega\right]^{1/2}. \tag{12.10.8}$$

The following result is *Parseval's theorem*, which relates the L_2 norm of the impulse response function to the H_2 norm of its transform.

Theorem 12.10.3. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$, where $A \in \mathbb{R}^{n\times n}$ is asymptotically stable, and let $H(t) = Ce^{tA}B$. Then,

$$\int_0^{\infty} H(t)H^{\mathrm{T}}(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\jmath\omega)G^*(\jmath\omega)\,\mathrm{d}\omega. \tag{12.10.9}$$

Therefore,

$$\|H\|_{\mathrm{L}_2} = \|G\|_{\mathrm{H}_2}. \tag{12.10.10}$$

Proof. First note that

$$G(s) = \mathcal{L}\{H(t)\} = \int_0^{\infty} H(t)e^{-st}\,\mathrm{d}t$$

and that

$$H(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\jmath\omega)e^{\jmath\omega t}\,\mathrm{d}\omega.$$

Hence,

$$\begin{aligned}\int_0^{\infty} H(t)H^{\mathrm{T}}(t)e^{-st}\,\mathrm{d}t &= \int_0^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} G(\jmath\omega)e^{\jmath\omega t}\,\mathrm{d}\omega\right]H^{\mathrm{T}}(t)e^{-st}\,\mathrm{d}t \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\jmath\omega)\left[\int_0^{\infty} H^{\mathrm{T}}(t)e^{-(s-\jmath\omega)t}\,\mathrm{d}t\right]\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\jmath\omega)G^{\mathrm{T}}(s-\jmath\omega)\,\mathrm{d}\omega.\end{aligned}$$

Setting $s = 0$ yields (12.10.6), while taking the trace of (12.10.9) yields (12.10.10). $\square$

Corollary 12.10.4. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$, where $A \in \mathbb{R}^{n\times n}$ is asymptotically stable, and let $H(t) = Ce^{tA}B$. Then,

$$\|G\|_{\mathrm{H}_2}^2 = \|H\|_{\mathrm{L}_2}^2 = CQC^{\mathrm{T}} = B^{\mathrm{T}}PB, \tag{12.10.11}$$

where $Q, P \in \mathbb{R}^{n\times n}$ satisfy (12.10.5) and (12.10.6), respectively.

The following corollary of Theorem 12.10.3 provides a frequency domain expression for the solution of the Lyapunov equation.

Corollary 12.10.5. Let $A \in \mathbb{R}^{n\times n}$ be asymptotically stable and let $B \in \mathbb{R}^{n\times m}$. Then, the matrix $Q \in \mathbb{R}^{n\times n}$ given by

$$Q = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\jmath\omega I - A)^{-1}BB^{\mathrm{T}}(\jmath\omega I - A)^{-*}\,\mathrm{d}\omega \tag{12.10.12}$$

satisfies

$$AQ + QA^{\mathrm{T}} + BB^{\mathrm{T}} = 0. \tag{12.10.13}$$

Proof. The result follows directly from Theorem 12.10.3 with $H(t) = e^{tA}B$ and $G(s) = (sI - A)^{-1}B$. Alternatively, it follows from (12.10.13) that

$$\int_{-\infty}^{\infty}(\jmath\omega I - A)^{-1}\,\mathrm{d}\omega Q + Q\int_{-\infty}^{\infty}(\jmath\omega I - A)^{-*}\,\mathrm{d}\omega = \int_{-\infty}^{\infty}(\jmath\omega I - A)^{-1}BB^{\mathrm{T}}(\jmath\omega I - A)^{-*}\,\mathrm{d}\omega.$$

Assuming A is diagonalizable with eigenvalues $\lambda_i = -\sigma_i + \jmath\omega_i$, it follows that

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{\jmath\omega - \lambda_i} = \int_{-\infty}^{\infty}\frac{\sigma_i - \jmath\omega}{\sigma_i^2 + \omega^2}\,\mathrm{d}\omega = \frac{\sigma_i\pi}{|\sigma_i|} - \jmath\lim_{r\to\infty}\int_{-r}^{r}\frac{\omega}{\sigma_i^2 + \omega^2}\,\mathrm{d}\omega = \pi,$$

which implies that

$$\int_{-\infty}^{\infty}(\jmath\omega I - A)^{-1}\,\mathrm{d}\omega = \pi I_n,$$

which yields (12.10.12). See [139] for a proof of the general case. □

Proposition 12.10.6. Let $G_1, G_2 \in \mathbb{R}^{l\times m}(s)$ be asymptotically stable rational transfer functions. Then,

$$\|G_1 + G_2\|_{\mathrm{H}_2} \le \|G_1\|_{\mathrm{H}_2} + \|G_2\|_{\mathrm{H}_2}. \tag{12.10.14}$$

Proof. Let $G_1 \overset{\min}{\sim} \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & 0 \end{array}\right]$ and $G_2 \overset{\min}{\sim} \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & 0 \end{array}\right]$, where $A_1 \in \mathbb{R}^{n_1\times n_1}$ and $A_2 \in \mathbb{R}^{n_2\times n_2}$. It thus follows that $G_1 + G_2 \sim \left[\begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & 0 \end{array}\right]$. It follows from Theorem 12.10.3 that $\|G_1\|_{\mathrm{H}_2} = \sqrt{\operatorname{tr} C_1Q_1C_1^{\mathrm{T}}}$ and $\|G_2\|_{\mathrm{H}_2} = \sqrt{\operatorname{tr} C_2Q_2C_2^{\mathrm{T}}}$, where $Q_1 \in \mathbb{R}^{n_1\times n_1}$ and $Q_2 \in \mathbb{R}^{n_2\times n_2}$ are the unique positive-definite matrices satisfying $A_1Q_1 + Q_1A_1^{\mathrm{T}} + B_1B_1^{\mathrm{T}} = 0$ and $A_2Q_2 + Q_2A_2^{\mathrm{T}} + B_2B_2^{\mathrm{T}} = 0$. Furthermore,

$$\|G_2 + G_2\|_{\mathrm{H}_2}^2 = \operatorname{tr} \begin{bmatrix} C_1 & C_2 \end{bmatrix} Q \begin{bmatrix} C_1^{\mathrm{T}} \\ C_2^{\mathrm{T}} \end{bmatrix},$$

where $Q \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ is the unique, nonnegative-semidefinite matrix satisfying

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} Q + Q \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}^{\mathrm{T}} = 0.$$

It can be seen that $Q = \left[\begin{smallmatrix} Q_1 & Q_{12} \\ Q_{12}^{\mathrm{T}} & Q_2 \end{smallmatrix}\right]$, where Q_1 and Q_2 are as given above and where Q_{12} satisfies $A_1Q_{12} + Q_{12}A_2^{\mathrm{T}} + B_1B_2^{\mathrm{T}} = 0$. Now, using the Cauchy-Schwarz inequality (9.3.15) and *iii*) of Proposition 8.2.3, it follows that

$$\begin{aligned}
\|G_1 + G_2\|_{\mathrm{H}_2}^2 &= \operatorname{tr}\big(C_1Q_1C_1^{\mathrm{T}} + C_2Q_2C_2^{\mathrm{T}} + C_2Q_{12}^{\mathrm{T}}C_1^{\mathrm{T}} + C_1Q_{12}C_2^{\mathrm{T}}\big) \\
&= \|G_1\|_{\mathrm{H}_2}^2 + \|G_2\|_{\mathrm{H}_2}^2 + 2\operatorname{tr} C_1Q_{12}Q_2^{-1/2}Q_2^{1/2}C_2^{\mathrm{T}} \\
&\le \|G_1\|_{\mathrm{H}_2}^2 + \|G_2\|_{\mathrm{H}_2}^2 + 2\operatorname{tr}\big(C_1Q_{12}Q_2^{-1}Q_{12}^{\mathrm{T}}C_1^{\mathrm{T}}\big)\operatorname{tr}\big(C_2Q_2C_2^{\mathrm{T}}\big) \\
&\le \|G_1\|_{\mathrm{H}_2}^2 + \|G_2\|_{\mathrm{H}_2}^2 + 2\operatorname{tr}\big(C_1Q_1C_1^{\mathrm{T}}\big)\operatorname{tr}\big(C_2Q_2C_2^{\mathrm{T}}\big) \\
&= (\|G_1\|_{\mathrm{H}_2} + \|G_2\|_{\mathrm{H}_2})^2.
\end{aligned}$$

□

12.11 Harmonic Steady-State Response

The following result, which is the *fundamental theorem of linear systems theory*, concerns the response of a linear system to a harmonic input.

Theorem 12.11.1. For $t \ge 0$, consider the linear system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.11.1}$$

with harmonic input

$$u(t) = \operatorname{Re} u_0 e^{\jmath\omega_0 t}, \tag{12.11.2}$$

where $u_0 \in \mathbb{C}^m$ and $\omega_0 \in \mathbb{R}$ is such that $\jmath\omega_0 \notin \operatorname{spec}(A)$. Then, $x(t)$ is given by

$$x(t) = e^{tA}\big(x(0) - \operatorname{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0\big]\big) + \operatorname{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0 e^{\jmath\omega_0 t}\big]. \tag{12.11.3}$$

Proof. We have

$$
\begin{aligned}
x(t) &= e^{tA}x(0) + \int_0^t e^{(t-\tau)A}B\mathrm{Re}(u_0e^{\jmath\omega_0\tau})\,\mathrm{d}\tau \\
&= e^{tA}x(0) + e^{tA}\mathrm{Re}\left[\int_0^t e^{-\tau A}e^{\jmath\omega_0\tau}\,\mathrm{d}\tau Bu_0\right] \\
&= e^{tA}x(0) + e^{tA}\mathrm{Re}\left[\int_0^t e^{\tau(\jmath\omega_0 I - A)}\,\mathrm{d}\tau Bu_0\right] \\
&= e^{tA}x(0) + e^{tA}\mathrm{Re}\Big[(\jmath\omega_0 I - A)^{-1}\Big(e^{\tau(\jmath\omega_0 I-A)} - I\Big)Bu_0\Big] \\
&= e^{tA}x(0) + \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}\big(e^{\jmath\omega_0 tI} - e^{tA}\big)Bu_0\big] \\
&= e^{tA}x(0) + \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}\big(-e^{tA}\big)\,Bu_0\big] + \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}e^{\jmath\omega_0 t}Bu_0\big] \\
&= e^{tA}\big(x(0) - \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0\big]\big) + \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0e^{\jmath\omega_0 t}\big]. \qquad \square
\end{aligned}
$$

Theorem 12.11.1 shows that the response of a linear system to a harmonic input consists of two components, namely, a transient component

$$x_{\mathrm{trans}}(t) \triangleq e^{tA}\big(x(0) - \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0\big]\big), \tag{12.11.4}$$

which depends on both the initial condition and input, and a harmonic steady-state component

$$x_{\mathrm{hss}}(t) = \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0e^{\jmath\omega_0 t}\big], \tag{12.11.5}$$

which depends only on the input.

If A is asymptotically stable, then $\lim_{t\to\infty} x_{\mathrm{trans}}(t) = 0$ and thus $x(t)$ approaches its harmonic steady-state component $x_{\mathrm{hss}}(t)$ for large t. Since the response is sinusoidal, it follows that $x(t)$ does not converge in the usual sense. If A is semistable, then it follows from *vii*) of Proposition 11.6.2 that

$$\lim_{t\to\infty} x_{\mathrm{trans}}(t) = \Big(I - AA^{\#}\Big)\big(x(0) - \mathrm{Re}\big[(\jmath\omega_0 I - A)^{-1}Bu_0\big]\big), \tag{12.11.6}$$

which represents a constant offset to the harmonic steady-state component. Finally, note that the complex amplitude of $x_{\mathrm{hss}}(t)$ involves $G(\jmath\omega_0) = (\jmath\omega_0 I - A)^{-1}B$, that is, the value of the rational transfer function $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$ evaluated at $s = \jmath\omega_0$, where ω_0 is the input frequency.

12.12 System Interconnections

Let $G \in \mathbb{R}^{l\times m}(s)$. We define the *parahermitian conjugate* $G^\sim$ of G by $G^\sim \triangleq G^{\mathrm{T}}(-s)$. The following result provides realizations for $G^{\mathrm{T}}, G^\sim$ and G^{-1}.

Proposition 12.12.1. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. Then,

$$G^{\mathrm{T}} \sim \left[\begin{array}{c|c} A^{\mathrm{T}} & C^{\mathrm{T}} \\ \hline B^{\mathrm{T}} & D^{\mathrm{T}} \end{array}\right] \tag{12.12.1}$$

and

$$G^\sim \sim \left[\begin{array}{c|c} -A^{\mathrm{T}} & -C^{\mathrm{T}} \\ \hline B^{\mathrm{T}} & D^{\mathrm{T}} \end{array}\right]. \tag{12.12.2}$$

Furthermore, if G is square and D is nonsingular, then

$$G^{-1} \sim \left[\begin{array}{c|c} A - BD^{-1}C & -BD^{-1} \\ \hline D^{-1}C & D^{-1} \end{array}\right]. \tag{12.12.3}$$

Proof. Since $y = Gu$, it follows that G^{-1} must satisfy $u = G^{-1}y$. Since $\dot{x} = Ax + Bu$ and $y = Cx + Du$, it follows that $u = -D^{-1}Cx + D^{-1}y$, and thus $\dot{x} = Ax + B\left(-D^{-1}Cx + D^{-1}y\right) = \left(A - BD^{-1}C\right)x + BD^{-1}y$. □

Note that if G is a SISO rational transfer function and $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, then $G \sim \left[\begin{array}{c|c} A^{\mathrm{T}} & B^{\mathrm{T}} \\ \hline C^{\mathrm{T}} & D \end{array}\right]$.

Let G_1 and G_2 be $l_1 \times m_1$ and $l_2 \times m_2$ rational transfer functions, respectively. Then, the *cascade interconnection* of G_1 and G_2 is the product G_2G_1, while the *parallel interconnection* is the sum $G_1 + G_1$. Note that G_2G_1 is defined only if $m_2 = l_1$ while $G_1 + G_2$ requires that $m_1 = m_2$ and $l_1 = l_2$.

Proposition 12.12.2. Let $G_1 \in \mathbb{R}^{l_1\times m_1}(s)$ and $G_1 \in \mathbb{R}^{l_2\times m_2}(s)$, and let $G_1 \sim \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$ and $G_2 \sim \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right]$. If $m_2 = l_1$, then

$$G_2G_1 \sim \left[\begin{array}{cc|c} A_1 & 0 & B_1 \\ B_2C_1 & A_2 & B_2D_1 \\ \hline D_2C_1 & C_2 & D_2D_1 \end{array}\right]. \tag{12.12.4}$$

If $m_1 = m_2$ and $l_1 = l_2$, then

$$G_1 + G_2 \sim \left[\begin{array}{c|cc} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & D_1 + D_2 \end{array}\right]. \tag{12.12.5}$$

Proof. Consider the state space equations

$$\dot{x}_1 = A_1x_1 + B_1u_1, \quad \dot{x}_2 = A_2x_2 + B_2u_2,$$
$$y_1 = C_1x_1 + D_1u_1, \quad y_2 = C_2x_2 + D_2u_2.$$

Since $u_2 = y_1$, it follows that

$$\dot{x}_2 = A_2x_2 + B_2C_1x_1 + B_2D_1u_1,$$
$$y_2 = C_2x_2 + D_2C_1x_1 + D_2D_1u_1,$$

and thus

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2D_1 \end{bmatrix} u_1,$$
$$y_2 = \begin{bmatrix} D_2C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + D_2D_1u_1,$$

which yields the realization (12.12.4) of G_2G_1. The realization (12.12.5) for $G_1 + G_2$ can be obtained by similar techniques. □

It is sometimes useful to combine systems by concatenating them in row, column, or block-diagonal forms.

Proposition 12.12.3. Let $G_1 \sim \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$ and $G_2 \sim \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right]$. Then,

$$\begin{bmatrix} G_1 & G_2 \end{bmatrix} \sim \left[\begin{array}{cc|cc} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ \hline C_1 & C_2 & D_1 & D_2 \end{array}\right], \tag{12.12.6}$$

$$\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \sim \left[\begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & 0 & D_1 \\ 0 & C_2 & D_2 \end{array}\right], \tag{12.12.7}$$

$$\begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \sim \left[\begin{array}{cc|cc} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ \hline C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{array}\right]. \tag{12.12.8}$$

Next, we interconnect a pair of systems G_1, G_2 by means of feedback as shown in Figure 2. It can be seen that u and y are related by

$$\hat{y} = (I + G_1G_2)^{-1}G_1\hat{u} \tag{12.12.9}$$

or

$$\hat{y} = G_1(I + G_2G_1)^{-1}\hat{u}. \tag{12.12.10}$$

The equivalence of (12.12.9) and (12.12.10) follows from the *push-through identity* Fact 2.13.15

$$(I + G_1G_2)^{-1}G_1 = G_1(I + G_2G_1)^{-1}. \tag{12.12.11}$$

A realization of this rational transfer function is given by the following result.

Proposition 12.12.4. Let $G_1 \sim \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$ and $G_2 \sim \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right]$. Then,

$$[I + G_1G_2]^{-1}\, G_1$$

$$\sim \left[\begin{array}{cc|c} A_1 - B_1(I + D_2D_1)^{-1}D_2C_1 & -B_1(I + D_2D_1)^{-1}C_2 & B_1(I + D_2D_1)^{-1} \\ B_2(I + D_1D_2)^{-1}C_1 & A_2 - B_2(I + D_1D_2)^{-1}D_1C_2 & B_2(I + D_1D_2)^{-1}D_1 \\ \hline (I + D_1D_2)^{-1}C_1 & -(I + D_1D_2)^{-1}D_1C_2 & (I + D_1D_2)^{-1}D_1 \end{array}\right]. \tag{12.12.12}$$

12.13 H_2 Standard Problem

The standard problem of feedback control involves four distinct signals, namely, an *exogenous input* w, a *control input* u, a *performance variable* z, and a *feedback signal* y. This system can be written as

$$\begin{bmatrix} \hat{z}(s) \\ \hat{y}(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} \hat{w}(s) \\ \hat{u}(s) \end{bmatrix}, \tag{12.13.1}$$

where G_{11}, G_{12}, G_{21}, and G_{22} are rational transfer functions.

Now, define the two-vector-input, two-vector-output transfer function

$$\mathcal{G} \triangleq \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}, \tag{12.13.2}$$

which has a realization

$$\mathcal{G} \sim \left[\begin{array}{c|cc} A & D_1 & B \\ \hline E_1 & E_0 & E_2 \\ C & D_2 & D \end{array}\right]. \tag{12.13.3}$$

Consequently, it can be seen that

$$\mathcal{G}(s) = \begin{bmatrix} E_1(sI - A)^{-1}D_1 + E_0 & E_1(sI - A)^{-1}B + E_2 \\ C(sI - A)^{-1}D_1 + D_2 & C(sI - A)^{-1}B + D \end{bmatrix}, \tag{12.13.4}$$

which shows that G_{11}, G_{12}, G_{21}, and G_{22} have the realizations

$$G_{11} \sim \left[\begin{array}{c|c} A & D_1 \\ \hline E_1 & E_0 \end{array}\right], \qquad G_{12} \sim \left[\begin{array}{c|c} A & B \\ \hline E_1 & E_2 \end{array}\right], \tag{12.13.5}$$

$$G_{21} \sim \left[\begin{array}{c|c} A & D_1 \\ \hline C & D_2 \end{array}\right], \qquad G_{22} \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{12.13.6}$$

Letting $G_{\rm c}$ denote the feedback controller, we interconnect G and $G_{\rm c}$ according to

$$\hat{u}(s) = G_{\rm c}(s)\hat{y}(s). \tag{12.13.7}$$

The resulting rational transfer function $\tilde{\mathcal{G}}$ satisfying $\hat{z}(s) = \tilde{\mathcal{G}}(s)\hat{w}(s)$ is thus given by

$$\tilde{\mathcal{G}} = G_{11} + G_{12}G_{\rm c}(I - G_{22}G_{\rm c})^{-1}G_{21} \tag{12.13.8}$$

or

$$\tilde{\mathcal{G}} = G_{11} + G_{12}(I - G_{\rm c}G_{22})^{-1}G_{\rm c}G_{21}. \tag{12.13.9}$$

A realization of $\tilde{G}$ is given by the following result.

Proposition 12.13.1. Let $\tilde{\mathcal{G}} \sim \left[\begin{array}{c|cc} A & D_1 & B \\ \hline E_1 & E_0 & E_2 \\ C & D_2 & D \end{array}\right]$ and $G_{\rm c} \sim \left[\begin{array}{c|c} A_{\rm c} & B_{\rm c} \\ \hline C_{\rm c} & D_{\rm c} \end{array}\right]$. If $\det(I - DD_c) \neq 0$, then

$$\tilde{\mathcal{G}} \sim \left[\begin{array}{cc|c} A + BD_{\rm c}(I - DD_{\rm c})^{-1}C & BC_{\rm c} + BD_{\rm c}(I - DD_{\rm c})^{-1}DC_{\rm c} & D_1 + BD_{\rm c}(I + DD_{\rm c})^{-1}D_2 \\ B_{\rm c}(I - DD_{\rm c})^{-1}C & A_{\rm c} + B_{\rm c}(I - DD_{\rm c})^{-1}DC_{\rm c} & B_{\rm c}(I - DD_{\rm c})^{-1}D_2 \\ \hline E_1 + E_2D_{\rm c}(I - DD_{\rm c})^{-1}C & E_2C_{\rm c} + E_2D_{\rm c}(I - DD_{\rm c})^{-1}DC_{\rm c} & E_0 + E_2D_{\rm c}(I - DD_{\rm c})^{-1}D_2 \end{array}\right]. \tag{12.13.10}$$

The realization (12.13.10) can be simplified when $DD_{\rm c} = 0$. For example, if $D = 0$, then

$$\tilde{\mathcal{G}} \sim \left[\begin{array}{cc|c} A + BD_{\rm c}C & BC_{\rm c} & D_1 + BD_{\rm c}D_2 \\ B_{\rm c}C & A_{\rm c} & B_{\rm c}D_2 \\ \hline E_1 + E_2D_{\rm c}C & E_2C_{\rm c} & E_0 + E_2D_{\rm c}D_2 \end{array}\right], \tag{12.13.11}$$

while if $D_{\rm c} = 0$, then

$$\tilde{\mathcal{G}} \sim \left[\begin{array}{cc|c} A & BC_{\rm c} & D_1 \\ B_{\rm c}C & A_{\rm c} + B_{\rm c}DC_{\rm c} & B_{\rm c}D_2 \\ \hline E_1 & E_2C_{\rm c} & E_0 \end{array}\right]. \tag{12.13.12}$$

Finally, if both $D = 0$ and $D_c = 0$, then

$$\tilde{\mathcal{G}} \sim \left[\begin{array}{cc|c} A & BC_c & D_1 \\ B_cC & A_c & B_cD_2 \\ \hline E_1 & E_2C_c & E_0 \end{array}\right]. \tag{12.13.13}$$

The feedback interconnection shown in Figure 4 forms the basis for the *standard problem* in feedback control. For this problem the signal w is interpreted as a *disturbance*, while the signal z represents the *performance variables*, that is, variables whose behavior reflects the performance of the closed-loop system. The performance variables need not be physically measured. The *controlled input* or the *control* u is driven by the feedback controller G_c, while the *measurement* signal y serves as the input to the *feedback controller* G_c. The standard problem in feedback control theory is the following: Given knowledge of w, determine G_c to minimize a performance criterion $J(G_c)$.

12.14 Linear-Quadratic Control

Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, and, for all $t \in [0,\infty)$, consider the system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.14.1}$$

$$x(0) = x_0. \tag{12.14.2}$$

Furthermore, let $K \in \mathbb{R}^{m\times n}$ and consider the full-state-feedback control law

$$u(t) = Kx(t). \tag{12.14.3}$$

The objective of the *linear-quadratic control problem* is to minimize the quadratic performance measure

$$J(K, x_0) = \int_0^\infty [x^{\mathrm{T}}(t)R_1x(t) + x^{\mathrm{T}}(t)R_{12}u(t) + u^{\mathrm{T}}(t)R_2u(t)]\,\mathrm{d}t, \tag{12.14.4}$$

where $R_1 \in \mathbb{R}^{n\times n}$, $R_{12} \in \mathbb{R}^{n\times m}$, and $R_2 \in \mathbb{R}^{m\times m}$. We assume that $\begin{bmatrix} R_1 & R_{12} \\ R_{12}^{\mathrm{T}} & R_2 \end{bmatrix}$ is nonnegative semidefinite and R_2 is positive definite.

The performance measure (12.14.4) indicates the desire to maintain the state vector $x(t)$ close to the zero equilibrium without an excessive expenditure of control effort. Specifically, the term $x^{\mathrm{T}}(t)R_1x(t)$ is a measure of the deviation of the state $x(t)$ from the zero state, where the $n \times n$ nonnegative-semidefinite matrix R_1 determines how much weighting is associated with every component of the state. Likewise, the $m \times m$ positive-definite matrix

R_2 weights the magnitude of the control input.

Using (12.14.1) and (12.14.3) the closed-loop dynamic system can be written as

$$\dot{x}(t) = (A + BK)x(t) \tag{12.14.5}$$

so that

$$x(t) = e^{t\tilde{A}}x_0, \tag{12.14.6}$$

where $\tilde{A} \triangleq A + BK$. Thus, the performance measure (12.14.4) becomes

$$\begin{aligned} J(K, x_0) &= \int_0^\infty x^{\mathrm{T}}(t)\left(R_1 + 2R_{12}K + K^{\mathrm{T}}R_2K\right)x(t)\,\mathrm{d}t \\ &= \int_0^\infty x_0^{\mathrm{T}} e^{t\tilde{A}^{\mathrm{T}}}\tilde{R}e^{t\tilde{A}}x_0\,\mathrm{d}t \\ &= \operatorname{tr} x_0^{\mathrm{T}} \int_0^\infty e^{t\tilde{A}^{\mathrm{T}}}\tilde{R}e^{t\tilde{A}^{\mathrm{T}}}\,\mathrm{d}t x(0), \\ &= \operatorname{tr} \int_0^\infty e^{t\tilde{A}^{\mathrm{T}}}\tilde{R}e^{t\tilde{A}}\,\mathrm{d}t x_0 x_0^{\mathrm{T}}, \end{aligned} \tag{12.14.7}$$

where $\tilde{R} \triangleq R_1 + 2R_{12}K + K^{\mathrm{T}}R_2K$.

Consider the standard problem with plant

$$\mathcal{G} \sim \left[\begin{array}{c|cc} A & D_1 & B \\ \hline E_1 & 0 & E_2 \\ I_n & 0 & 0 \end{array}\right]. \tag{12.14.8}$$

and full-state feedback $u = Kx$. Then, the closed-loop transfer function is given by

$$\tilde{\mathcal{G}} \sim \left[\begin{array}{c|c} A + BK & D_1 \\ \hline E_1 + E_2K & 0 \end{array}\right]. \tag{12.14.9}$$

The following result shows that the quadratic performance measure (12.14.4) is equivalent to an H_2 norm.

Proposition 12.14.1. Assume that $m = 1$, $D_1 = x_0$, and

$$\begin{bmatrix} R_1 & R_{12} \\ R_{12}^{\mathrm{T}} & R_2 \end{bmatrix} = \begin{bmatrix} E_1^{\mathrm{T}} \\ E_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} E_1 & E_2 \end{bmatrix}, \tag{12.14.10}$$

and let $\tilde{\mathcal{G}}$ be given by (12.14.9). Then,

$$J(K, x_0) = \|\tilde{\mathcal{G}}\|_{\mathrm{H}_2}^2. \tag{12.14.11}$$

Proof. The result is a consequence of Proposition 12.1.2. □

To develop necessary conditions for the linear-quadratic control problem, we restrict K to the set of stabilizing gains

$$\mathcal{S} \triangleq \{K \in \mathbb{R}^{m\times n}\colon\ A + BK \text{ is asymptotically stable}\}. \tag{12.14.12}$$

Obviously, $\mathcal{S}$ is nonempty if and only if (A, B) is stabilizable. The following result gives necessary conditions for characterizing a stabilizing solution K of the linear-quadratic control problem.

Theorem 12.14.2. Assume that (A, B) is stabilizable and assume that $K \in \mathcal{S}$ solves the linear-quadratic control problem. Then, there exists an $n \times n$ nonnegative-semidefinite matrix P such that K is given by

$$K = -R_2^{-1}B^{\mathrm{T}}P \tag{12.14.13}$$

and such that P satisfies

$$A^{\mathrm{T}}P + PA + R_1 - PBR_2^{-1}B^{\mathrm{T}}P = 0. \tag{12.14.14}$$

Furthermore, the minimal cost is given by

$$J(K) = \operatorname{tr} PV. \tag{12.14.15}$$

Proof. Since $K \in \mathcal{S}$, it follows that $\tilde{A}$ is asymptotically stable. It then follows that $J(K)$ is given by (12.14.15), where $P \triangleq \int_0^\infty e^{t\tilde{A}^{\mathrm{T}}}\tilde{R}e^{t\tilde{A}}\,\mathrm{d}t$ is nonnegative semidefinite and satisfies the Lyapunov equation

$$\tilde{A}^{\mathrm{T}}P + P\tilde{A} + \tilde{R} = 0. \tag{12.14.16}$$

Note that (12.14.16) can be written as

$$(A + BK)^{\mathrm{T}}P + P(A + BK) + R_1 + K^{\mathrm{T}}R_2K = 0. \tag{12.14.17}$$

To optimize (12.14.15) subject to the constraint (12.14.16) over the open set $\mathcal{S}$, form the Lagrangian

$$\mathcal{L}(K, P, Q, \lambda_0) \triangleq \operatorname{tr}\Big[\lambda_0 PV + Q\left(\tilde{A}^{\mathrm{T}}P + P\tilde{A} + \tilde{R}\right)\Big], \tag{12.14.18}$$

where the Lagrange multipliers $\lambda_0 \ge 0$ and $Q \in \mathbb{R}^{n\times n}$ are not both zero. Note that the $n \times n$ Lagrange multiplier Q accounts for the $n \times n$ constraint equation (12.14.16).

Next, setting $\partial\mathcal{L}/\partial P = 0$ yields

$$\tilde{A}Q + Q\tilde{A}^{\mathrm{T}} + \lambda_0 V = 0. \tag{12.14.19}$$

Since $\tilde{A}$ is asymptotically stable, it follows from Proposition 11.7.3 that, for all $\lambda_0 \geq 0$, (12.14.19) has a unique solution Q and, furthermore, Q is nonnegative-semidefinite. In particular, if $\lambda_0 = 0$, then $Q = 0$. Since λ_0 and Q are not both zero, we can set $\lambda_0 = 1$ so that (12.14.19) becomes

$$\tilde{A}Q + Q\tilde{A}^{\mathrm{T}} + V = 0, \tag{12.14.20}$$

Since V is positive definite, it follows that Q is positive definite.

Next, evaluating $\partial\mathcal{L}/\partial K$ yields

$$R_2KQ + B^{\mathrm{T}}PQ = 0. \tag{12.14.21}$$

Since Q is positive definite, it follows from (12.14.21) that (12.14.13) is satisfied. Furthermore, using (12.14.13), it follows that (12.14.16) is equivalent to (12.14.14). □

Note that with K given by (12.14.13) the closed-loop dynamics matrix $\tilde{A} = A + BK$ is given by

$$\tilde{A} = A - BR_2^{-1}B^{\mathrm{T}}P, \tag{12.14.22}$$

where P is the solution of the *Riccati equation* (12.14.14). For convenience we define $\Sigma \triangleq BR_2^{-1}B^{\mathrm{T}}$ so that $\tilde{A} = A - \Sigma P$ and (12.14.14) can be written as

$$A^{\mathrm{T}}P + PA + R_1 - P\Sigma P = 0. \tag{12.14.23}$$

Note that (12.14.23) can be written in the form of the Lyapunov equation

$$(A - \Sigma P)^{\mathrm{T}}P + P(A - \Sigma P) + R_1 + P\Sigma P = 0, \tag{12.14.24}$$

which is equivalent to (12.14.16) with $\tilde{R} = R_1 + P\Sigma P$.

Next, we consider solutions of the Riccati equation (12.14.23). For convenience we let $R_1 = E_1^{\mathrm{T}}E_1$, where $E_1 \in \mathbb{R}^{q\times n}$ characterizes a performance variable $z(t) = E_1x(t)$. The following examples help to clarify conditions under which (12.14.23) has a solution.

Example 12.14.3. Let $A = 0$, $B = 0$, $E_1 \neq 0$, and $R_2 = I$. In this case (A, B) is not stabilizable, and (12.14.23) becomes $R_1 = 0$. Thus, (12.14.23) has no solution.

Example 12.14.4. Let $A = I$, $B = 0$, $E_1 = I$, and $R_2 = I$. In this case (A, B) is not stabilizable. Furthermore, (12.14.23) becomes $2P + I = 0$

so that $P = -\frac{1}{2}I$ is the only solution. Thus, (12.14.23) does not have a nonnegative-semidefinite solution.

Example 12.14.5. Let $n > 1$, $A = 0$, $B = I$, $E_1 = I$ and $R_2 = I$. In this case (A, B) is stabilizable. Furthermore, (12.14.23) becomes $P^2 = I$, which is satisfied by infinitely many real symmetric matrices P given by $P = S\left[\begin{smallmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right]S^{\mathrm{T}}$, where $S \in \mathbb{R}^{2\times 2}$ is orthogonal. However, $P = I$ is the only nonnegative-semidefinite solution. In fact, P is positive definite.

Example 12.14.6. Let $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, $E_1 = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and $R_2 = 1$ so that (A, B) is controllable but neither of the states is weighted. In this case (12.14.23) has four nonnegative-semidefinite solutions given by

$$P_1 = \begin{bmatrix} 18 & -24 \\ -24 & 36 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}, \quad P_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The corresponding feedback matrices are given by $K_1 = \begin{bmatrix} 6 & -12 \end{bmatrix}$, $K_2 = \begin{bmatrix} -2 & 0 \end{bmatrix}$, $K_3 = \begin{bmatrix} 0 & -4 \end{bmatrix}$, and $K_4 = \begin{bmatrix} 0 & 0 \end{bmatrix}$. Letting $\tilde{A}_i = A - \Sigma P_i$, it follows that $\operatorname{spec}(\tilde{A}_1) = \{-1, -2\}$, $\operatorname{spec}(\tilde{A}_2) = \{-1, 2\}$, $\operatorname{spec}(\tilde{A}_3) = \{1, -2\}$, and $\operatorname{spec}(\tilde{A}_4) = \{1, 2\}$. Thus, P_1 is the only solution that stabilizes the closed-loop system, while the solutions P_2 and P_3 partially stabilize the closed-loop system. Note also that the closed-loop poles that differ from those of the open-loop system are mirror images of the open-loop poles as reflected across the imaginary axis. Finally, note that these solutions satisfy the partial ordering $P_1 \geq P_2 \geq P_4$ and $P_1 \geq P_3 \geq P_4$, and that "larger" solutions have a more stabilizing effect than "smaller" solutions. Moreover, letting $J(K_i) = \operatorname{tr} P_i V$, it can be seen that larger solutions incur a greater closed-loop cost, with the greatest cost incurred by the stabilizing solution P_4. However, the expression $J(K) = \operatorname{tr} PV$ requires justification when $A + BK$ is not asymptotically stable.

Example 12.14.7. Let $A = \left[\begin{smallmatrix} -1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, $E_1 = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and $R_2 = 1$ so that (A, B) is stabilizable, while only the asymptotically stable eigenvalue is weighted. Now, $P = 0$ is the only nonnegative-semidefinite solution of (12.14.23). This solution is not asymptotically stabilizing since reflecting the eigenvalue at the origin across the imaginary axis fails to move it into the open left half plane.

Example 12.14.8. Let $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, $E_1 = \begin{bmatrix} 0 & 0 \end{bmatrix}$, and $R_2 = 1$. Taking the trace of (12.14.23) yields $\operatorname{tr} P^2 = 0$. Thus, the only nonnegative-semidefinite matrix P satisfying (12.14.23) is $P = 0$, which implies that $K = 0$ and $\tilde{A} = A$. Consequently, the open-loop eigenvalues $\pm \jmath$ are unmoved by the feedback gain (12.14.13) even though (A, B) is controllable. As in the previous example, reflecting these unweighted poles across the imaginary axis fails to move them into the open left half plane.

12.15 Solutions of the Riccati Equation

The following definitions will be useful in studying the various solutions to the Riccati equation.

Definition 12.15.1. A matrix $P \in \mathbb{R}^{n\times n}$ is a *solution* of the Riccati equation (12.14.23) if P is symmetric and satisfies (12.14.23). Furthermore, P is the *stabilizing solution* to (12.14.23) if $A - \Sigma P$ is asymptotically stable. Finally, a solution P is the *maximal solution* to (12.14.23) if $P \geq P'$ for every solution P' to (12.14.23).

Theorem 12.15.2. There exists a nonnegative-semidefinite solution to (12.14.23) if and only if (A, B, E_1) has no CRHP eigenvalues that are uncontrollable and observable.

Proof. To prove necessity, suppose that (12.14.23) has a nonnegative-semidefinite solution P, let $\tilde{A} = A - \Sigma P$, and suppose that (A, B, E_1) has a CRHP eigenvalue that is uncontrollable and observable. It thus follows from Proposition 12.8.9 that there exists an invertible matrix $S \in \mathbb{R}^{n\times n}$ such that

$$SAS^{-1} = \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad E_1S^{-1} = \begin{bmatrix} E_{11} & E_{12} \end{bmatrix},$$

where (A_2, E_{12}) is observable and A_2 is not asymptotically stable. Next, note that

$$\int_0^t e^{\tau\tilde{A}^{\mathrm{T}}} E_1^{\mathrm{T}} E_1 e^{\tau\tilde{A}}\,\mathrm{d}\tau \leq \int_0^t e^{\tau\tilde{A}^{\mathrm{T}}} \tilde{R} e^{\tau\tilde{A}}\,\mathrm{d}\tau = -\int_0^t e^{\tau\tilde{A}^{\mathrm{T}}} \left(\tilde{A}^{\mathrm{T}}P + P\tilde{A}\right) e^{\tau\tilde{A}}\,\mathrm{d}\tau$$
$$= -\int_0^t \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\tau\tilde{A}^{\mathrm{T}}} P e^{\tau\tilde{A}}\,\mathrm{d}\tau = P - e^{t\tilde{A}^{\mathrm{T}}} P e^{t\tilde{A}} \leq P.$$

Next, it can be seen that the (2,2) block of this inequality in the transformed basis is given by

$$\int_0^t e^{\tau A_2^{\mathrm{T}}} E_{12}^{\mathrm{T}} E_{12} e^{\tau A_2}\,\mathrm{d}\tau \leq \begin{bmatrix} 0 & I \end{bmatrix} S^{\mathrm{T}} P S \begin{bmatrix} 0 \\ I \end{bmatrix}.$$

Since (A_2, E_{12}) is observable and the integral is bounded, it follows from Proposition 12.4.3 that A_2 is asymptotically stable, which is a contradiction.

Conversely, suppose that (A, B, E_1) has no CRHP eigenvalues that are uncontrollable and observable. Then, it follows from Theorem 5.4.1

that there exists an invertible matrix $S \in \mathbb{F}^{n\times n}$ such that

$$SAS^{-1} = \begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, E_1S^{-1} = \begin{bmatrix} E_{11} & 0 \end{bmatrix},$$

where (A_1, B_1) is stabilizable and (A_1, E_{11}) is observable. Theorem XXX thus implies that the reduced Riccati equation $A_1^{\mathrm{T}}P_1 + P_1A_1 + E_{11}^{\mathrm{T}}E_{11} - P_1B_1R_2^{-1}B_1^{\mathrm{T}}P_1 = 0$ has a nonnegative-semidefinite solution P_1. Finally, the Riccati equation (12.14.23) is now satisfied by $P = S^{\mathrm{T}}\left[\begin{smallmatrix} P_1 & 0 \\ 0 & 0 \end{smallmatrix}\right]S$, which is nonnegative semidefinite.

where (A_2, E_{12}) is observable and $\lambda \in \operatorname{spec}(A_2)$. Since (12.14.23) has a nonnegative-semidefinite solution, it follows from Proposition XXX that $(\hat{A} - \hat{\Sigma}\hat{P}, \hat{E}_1)$ is controllably asymptotically stable, where $\hat{\Sigma} = \left[\begin{smallmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $\Sigma_1 = B_1R_2^{-1}B_1^{\mathrm{T}}$, and $\hat{P} = S^{\mathrm{T}}PS = \left[\begin{smallmatrix} P_1 & P_{12} \\ P_{12}^{\mathrm{T}} & P_2 \end{smallmatrix}\right]$. Therefore, Proposition XXX implies that $\hat{E}_1e^{t(\hat{A}-\hat{\Sigma}\hat{P})} \to 0$ as $t \to \infty$. Consequently, $E_{11}e^{t(A_1-\Sigma_1P_1)} \to 0$ as $t \to \infty$ and

$$E_{11}\int_0^t e^{t(A_1-\Sigma_1P_1)}(A_{12} - \Sigma_1P_{12})e^{(t-\tau)\hat{A}_2}\,\mathrm{d}t + E_{12}e^{tA_2} \to 0 \text{ as } t \to \infty. \tag{12.15.1}$$

For large $t > 0$, the first term has norm proportional to $|e^{\lambda_1 t}|$, where $\operatorname{Re}\lambda_1 < \operatorname{Re}\lambda$, and the second term has norm proportional to $|e^{\lambda t}|$. However, $\operatorname{Re}\lambda \geq 0$ contradicts (12.15.1).

Conversely, suppose that (A, B, E_1) has no ORHP eigenvalues that are uncontrollable and observable. Then, it follows from Proposition 12.8.9 that there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$SAS^{-1} = \begin{bmatrix} A_1 & A_{13} & 0 & 0 \\ 0 & A_3 & 0 & 0 \\ A_{21} & A_{23} & A_2 & A_{24} \\ 0 & A_{43} & 0 & A_4 \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ 0 \\ B_2 \\ 0 \end{bmatrix},$$
$$E_1S^{-1} = \begin{bmatrix} E_{11} & E_{13} & 0 & 0 \end{bmatrix}. \tag{12.15.2}$$

where (A_1, B_1, E_{11}) is controllable and observable, (A_2, B_2) is controllable, (A_3, E_{13}) is observable, and A_3 is asymptotically stable. Therefore,

$$\left(\begin{bmatrix} A_1 & A_{13} \\ 0 & A_3 \end{bmatrix}, \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \begin{bmatrix} E_{11} & E_{13} \end{bmatrix}\right)$$

is stabilizable and detectable, and thus Theorem 1 implies that there exists a nonnegative-semidefinite solution $\hat{P}_1$ to

$$\begin{bmatrix} A_1 & A_{13} \\ 0 & A_3 \end{bmatrix}^{\mathrm{T}} \hat{P}_1 + \hat{P}_1 \begin{bmatrix} A_1 & A_{13} \\ 0 & A_3 \end{bmatrix} + \begin{bmatrix} E_{11}^{\mathrm{T}} E_{11} & E_{11}^{\mathrm{T}} E_{13} \\ E_{13}^{\mathrm{T}} E_{11} & E_{13}^{\mathrm{T}} E_{13} \end{bmatrix} - \hat{P}_1 \begin{bmatrix} B_1 R_2^{-1} B_1^{\mathrm{T}} & 0 \\ 0 & 0 \end{bmatrix} \hat{P}_1 = 0. \tag{12.15.3}$$

Consequently, $P = S^{\mathrm{T}} \operatorname{diag}(\hat{P}_1, 0, 0) S$ is a nonnegative-semidefinite solution of (12.14.23). □

Corollary 12.15.3. Suppose that (A, B) is stabilizable. Then, (12.14.23) has a nonnegative-semidefinite solution.

Theorem 12.15.4. Let P be a nonnegative-semidefinite solution to (12.14.23). Then, P is maximal if and only if $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$.

Proof. See □

Note that, since the ordering "$\leq$" is antisymmetric, there exists at most one maximal solution to (12.14.23). Therefore, it follows from Theorem 12.15.2 that (12.14.23) has at most one nonnegative-semidefinite solution P such that $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$.

Corollary 12.15.5. There exists at most one stabilizing solution (12.14.23). If P is the stabilizing solution to (12.14.23), then P is nonnegative-semidefinite and it is also the maximal solution (12.14.23).

Proof. Suppose there exist two stabilizing solutions P_1 and P_2 to (12.14.23). Then,

$$A^{\mathrm{T}} P_1 + P_1 A + R_1 - P_1 \Sigma P_1 = 0,$$
$$A^{\mathrm{T}} P_2 + P_2 A + R_1 - P_2 \Sigma P_2 = 0.$$

Subtracting these equations and rearranging yields

$$(A - \Sigma P_1)^{\mathrm{T}} (P_1 - P_2) + (P_1 - P_2)(A - \Sigma P_2) = 0.$$

Since $A - \Sigma P_1$ and $A - \Sigma P_2$ are asymptotically stable, it follows from Proposition 7.2.3 and Proposition 11.7.3 that $P_1 - P_2 = 0$. Hence, there exists at most one stabilizing solution to (12.14.23).

Next, suppose that there exists a stabilizing P to (12.14.23). Then, it follows from (12.14.23) that

$$P = \int_0^\infty e^{t(A - \Sigma P)^{\mathrm{T}}} (R_1 + P \Sigma P) e^{t(A - \Sigma P)} \, \mathrm{d}t,$$

which shows that P is nonnegative semidefinite. Next, let P' be a solution to (12.14.23). Then, it follows that

$$(A-\Sigma P)^{\mathrm{T}}(P-P')+(P-P')(A-\Sigma P)+(P-P')\Sigma(P-P')=0,$$

which implies that $P' \le P$. Thus, P is also the maximal solution to (12.14.23). □

Next, we consider the existence of a maximal solution to (12.14.23). The following lemma is needed.

Lemma 12.15.6. Let (A,B) be controllable, let $t_1>0$, and define

$$P=\left(\int_0^{t_1} e^{-tA}\Sigma e^{-tA^{\mathrm{T}}}\,\mathrm{d}t\right)^{-1}. \tag{12.15.4}$$

Then, $A-\Sigma P$ is asymptotically stable.

Proof. It can be seen that P satisfies

$$(A-\Sigma P)^{\mathrm{T}}P+P(A-\Sigma P)+P\left(\Sigma+e^{t_1A}\Sigma e^{t_1A^{\mathrm{T}}}\right)P=0.$$

Since $\left(A-\Sigma P,\Sigma+e^{t_1A}\Sigma e^{t_1A^{\mathrm{T}}}\right)$ is observable and P is positive definite, it follows from Proposition 11.7.6 that $A-\Sigma P$ is asymptotically stable. □

Theorem 12.15.7. Suppose that (A,B) is stabilizable. Then, there exists a maximal solution P to (12.14.23). Furthermore, $\operatorname{spec}(A-\Sigma P)\subset$ CLHP.

Proof. Since (A,B) is stabilizable, it follows from Corollary 12.6.3 that there exists an invertible matrix $S\in\mathbb{R}^{n\times n}$ such that

$$SAS^{-1}=\begin{bmatrix} A_1 & A_{12}\\ 0 & A_2\end{bmatrix},\quad SB=\begin{bmatrix} B_1\\ 0\end{bmatrix}, \tag{12.15.5}$$

where $A_1\in\mathbb{R}^{r\times r}$, (A_1,B_1) is controllable, and A_2 is asymptotically stable. Next, since the pair (A_1,B_1) is controllable, there exists a positive-definite matrix $\hat{P}_0\in\mathbb{R}^{r\times r}$ such that $A_1-B_1R_2^{-1}B_1^{\mathrm{T}}\hat{P}_0$ is asymptotically stable. It follows from Lemma 12.15.6 that one such matrix is given by

$$\hat{P}_0=\left(\int_0^{1} e^{-tA_1}B_1\Sigma B_1^{\mathrm{T}}e^{-tA_1^{\mathrm{T}}}\,\mathrm{d}t\right)^{-1}. \tag{12.15.6}$$

Thus, for the nonnegative-semidefinite matrix

$$P_0\triangleq\begin{bmatrix}\hat{P}_0 & 0\\ 0 & 0\end{bmatrix}$$

it follows that $A - \Sigma P_0$ is asymptotically stable.

Next, it follows from Proposition 12.1? that there exists a nonnegative-semidefinite matrix $P \in \mathbb{F}^{n\times n}$ satisfying (12.14.23), that is,

$$(A - \Sigma P)^{\mathrm{T}} P + P(A - \Sigma P) + P\Sigma P + R_1 = 0. \tag{12.15.7}$$

Now, define a sequence of nonnegative-semidefinite matrices $\{P_k\}_{k=0}^{\infty}$ satisfying

$$(A - \Sigma P_k)^{\mathrm{T}} P_{k+1} + P_{k+1}(A - \Sigma P_k) + P_k\Sigma P_k + R_1' = 0, \tag{12.15.8}$$

where $R_1' \in \mathbb{R}^{n\times n}$ is symmetric and satisfies $R_1' \ge R_1$. Assuming now that $A - MSigP_k$ is asymptotically stable, we show that $A - \Sigma P_{k+1}$ is asymptotically stable. To do this, first note that (12.20??) and (12.21??) imply that

$$(A - \Sigma P_k)^{\mathrm{T}} P + P(A - \Sigma P_k) + P_k\Sigma P_k - (P - P_k)\Sigma(P - P_k) + R_1 = 0. \tag{12.15.9}$$

Subtracting (12.15.9) from (12.15.8) yields

$$\begin{aligned}(A - \Sigma P_k)^{\mathrm{T}}(P_{k+1} - P) + (P_{k+1} - P)(A - \Sigma P_k)\\ +(P - P_k)\Sigma(P - P_k) + R_1' - R_1 = 0,\end{aligned} \tag{12.15.10}$$

which, since $A - \Sigma P_k$ is asymptotically stable, implies that

$$P_{k+1} - P = \int_0^\infty e^{t(A-\Sigma P_k)^{\mathrm{T}}}[(P - P_k)\Sigma(P - P_k) + R_1' - R_1]e^{t(A-\Sigma P_k)}\,\mathrm{d}t \ge 0. \tag{12.15.11}$$

Hence, $P_{k+1} \ge P$.

Next, note that (12.15.8) is equivalent to

$$\begin{aligned}(A - \Sigma P_{k+1})^{\mathrm{T}} P_{k+1} + P_{k+1}(A - \Sigma P_{k+1}) + P_{k+1}\Sigma P_{k+1}\\ +(P_{k+1} - P_k)\Sigma(P_{k+1} - P_k) + R_1' = 0.\end{aligned} \tag{12.15.12}$$

Subtracting (12.15.9) with k replaced by $k+1$ from (12.15.12) yields

$$(A - \Sigma P_{k+1})^{\mathrm{T}}(P_{k+1} - P) + (P_{k+1} - P)(A - \Sigma P_{k+1}) = M, \tag{12.15.13}$$

where $M \triangleq -(P_{k+1}-P_k)\Sigma(P_{k+1}-P_k)-(P_{k+1}-P)\Sigma(P_{k+1}-P)-R_1'+R_1 \le 0$.

Now, let $\lambda \in \mathbb{C}$ and nonzero $x \in \mathbb{C}^n$ satisfy $(A-\Sigma P_{k+1})x = \lambda x$. Then, it follows from (12.15.13) that

$$(\lambda + \overline{\lambda})x^*(P_{k+1} - P)x = x^*Mx. \tag{12.15.14}$$

Since $\lambda + \overline{\lambda} \geq 0$ and $P_{k+1} \geq P$, it follows from (12.15.14) that $x^*Mx = 0$, which in turn implies that

$$x^*(P_{k+1} - P_k)\Sigma(P_{k+1} - P_k)x = 0. \tag{12.15.15}$$

Furthermore, since Σ is nonnegative semidefinite, it follows that $\Sigma(P_{k+1} - P_k)x = 0$, which implies that

$$(A - \Sigma P_k)x = (A - \Sigma P_{k+1})x = \lambda x. \tag{12.15.16}$$

However, $A - \Sigma P_k$ is asymptotically stable, which implies that $\operatorname{Re}\lambda < 0$, which is a contradiction. Hence, $A - \Sigma P_{k+1}$ is asymptotically stable.

Next, subtract (12.15.12) with k replaced by $k-1$ from (12.15.8) to obtain

$$\begin{aligned}(A - \Sigma P_k)^{\mathrm{T}}(P_k - P_{k+1}) + (P_k - P_{k+1})(A - \Sigma P_k)&\\ +(P_k - P_{k-1})\Sigma(P_k - P_{k-1}) = 0,&\end{aligned} \tag{12.15.17}$$

which, since $A - \Sigma P_k$ is asymptotically stable, implies that

$$P_k - P_{k+1} = \int_0^\infty e^{t(A-\Sigma P_k)^{\mathrm{T}}}(P_k - P_{k+1})\Sigma(P_k - P_{k+1})e^{t(A-\Sigma P_k)}\,\mathrm{d}t \geq 0. \tag{12.15.18}$$

Hence, $\{P_k\}_{k=0}^\infty$ is a nonincreasing sequence of nonnegative-semidefinite matrices bounded from below by P. Thus, $P_+ \triangleq \lim_{k\to\infty} P_k$ exists.

Now, let $R_1' = R_1$. Letting $k \to \infty$it follows from (12.15.8) that P_+ is a solution to (12.14.23). Furthermore, since $A - \Sigma P_k$ is asymptotically stable for all $k \in \mathbb{P}$ it follows that $\operatorname{spabs}(A - \Sigma P_+) \leq 0$. Also note that $P_+ \geq P$ for every solution P of (12.14.23), which implies that P_+ is the maximal solution of (12.14.23). $\square$

Proposition 12.15.8. Suppose that (A, B) is stabilizable, let $R_1' \in \mathbb{N}^n$ satisfy $R_1' \geq R_1$, and let P_+ and P_+' denote, respectively, the maximal solutions of (12.14.23) and

$$A^{\mathrm{T}}P + PA + R_1' - P\Sigma P = 0. \tag{12.15.19}$$

Then, $P_+' \geq P_+$.

Proof. Letting $k \to \infty$ in (12.15.8), it follows that $P_0 \triangleq \lim_{k\to\infty} P_k$ is a solution of (12.15.19) and satisfies $P_0 \geq P$ for every solution of (12.15.8). Hence, $P_0 = P_+'$ and thus $P_0 \geq P_+$. $\square$

Proposition 12.15.9. Suppose that (A, B) is stabilizable and (A, E_1) is detectable. Then, there exists a nonnegative-semidefinite solution P to

(12.14.23) such that $A-\Sigma P$ is asymptotically stable. If, in addition, (A,E_1) is observable, then P is positive definite.

Proof. Define a sequence of nonnegative-semidefinite matrices $\{P_k\}_{k=0}^{\infty}$ satisfying

$$(A+BK_k)^{\mathrm{T}}P_k + P_k(A+BK_k) + R_1 + K_k^{\mathrm{T}}R_2K_k = 0, \tag{12.15.20}$$

where $K_0 \in \mathbb{R}^{m\times n}$ is such that $A+BK_0$ is asymptotically stable, and, for all $k\in\mathbb{P}$, K_k is given by

$$K_{k+1} = -R_2^{-1}B^{\mathrm{T}}P_k. \tag{12.15.21}$$

Therefore, P_0 is nonnegative semidefinite.

Next, note the identity

$$\begin{aligned}(A-\Sigma P_k)^{\mathrm{T}}P_k &+ P_k(A-\Sigma P_k) + P_k\Sigma P_k\\ &= (A-\Sigma P_{k-1})^{\mathrm{T}}P_k + P_k(A-\Sigma P_{k-1}) + P_{k-1}\Sigma P_{k-1}\\ &- (P_k-P_{k-1})\Sigma(P_k-P_{k-1}),\end{aligned} \tag{12.15.22}$$

or, equivalently,

$$\begin{aligned}(A+BK_{k+1})^{\mathrm{T}}P_k &+ P_k(A+BK_{k+1}) + R_1 + K_{k+1}^{\mathrm{T}}R_2K_{k+1}\\ &= (A+BK_k)^{\mathrm{T}}P_k + P_k(A+BK_k) + R_1 + K_k^{\mathrm{T}}R_2K_k -\\ &(K_k-K_{k+1})^{\mathrm{T}}R_2(K_k-K_{k+1}).\end{aligned} \tag{12.15.23}$$

Next, using (12.15.20) it follows from (12.15.23) that

$$(A+BK_{k+1})^{\mathrm{T}}P_k + P_k(A+BK_{k+1}) + R_1 + N_k + K_{k+1}^{\mathrm{T}}R_2K_{k+1} = 0, \tag{12.15.24}$$

where

$$N_k \triangleq (K_k-K_{k+1})^{\mathrm{T}}R_2(K_k-K_{k+1}) \ge 0. \tag{12.15.25}$$

Since, by assumption, (A,E_1) is detectable, it follows from Lemma 12.17.33, that the pair $\left(A+BK_{k+1}, \left[R_1+N_k+K_{k+1}^{\mathrm{T}}R_2K_{k+1}\right]^{1/2}\right)$ is also detectable for all $k\in\mathbb{N}$. Now, assume that P_k is nonnegative semidefinite so that Proposition 12.12.4 implies that $A+BK_{k+1}$ is asymptotically stable. Next, replacing k by $k+1$ in (12.15.20) yields

$$(A+BK_{k+1})^{\mathrm{T}}P_{k+1} + P_{k+1}(A+BK_{k+1}) + R_1 + K_{k+1}^{\mathrm{T}}R_2K_{k+1} = 0. \tag{12.15.26}$$

Since $A+BK_{k+1}$ is asymptotically stable, it follows that P_{k+1} is nonnegative semidefinite.

Next, subtracting (12.15.26) from (12.15.24) yields

$$\left(A + BK_{k+1}^{\mathrm{T}}\right)(P_k - P_{k+1}) + (P_k - P_{k+1})(A + BK_{k+1}) + N_k = 0, \tag{12.15.27}$$

which, since $A + BK_{k+1}$ is asymptotically stable, implies that $\{P_k\}_{k=0}^{\infty}$ is a nonincreasing sequence of nonnegative-semidefinite matrices. Thus, $P \triangleq \lim_{k\to\infty} P_k$ exists and satisfies

$$(A + BK)^{\mathrm{T}}P + P(A + BK) + R_1 + K^{\mathrm{T}}R_2K = 0. \tag{12.15.28}$$

Furthermore, $K \triangleq \lim_{k\to\infty} K_k = -R_2^{-1}B^{\mathrm{T}}P$ also exists. Next, since $\left(A + BK, \left[R_1 + K^{\mathrm{T}}R_2K\right]^{1/2}\right)$ is detectable, Proposition 12.12.4 implies that $A + BK$ is asymptotically stable.

Next, assume that (A, E_1) is observable so that $\left(A + BK, \left[R_1 + K^T R_2K\right]^{1/2}\right)$ is observable. Since $A+BK$ is asymptotically stable, it follows from (12.15.28) that P is positive definite. □

Theorem 12.15.10. (12.14.23) has a nonnegative-semidefinite solution if and only if every CRHP eigenvalue of (A, B, E_1) is either controllable or unobservable.

Theorem 12.15.11. The following statements hold:

i) (12.14.23) has at most one maximal solution.

ii) (12.14.23) has a nonnegative-semidefinite maximal solution if and only if it has a nonnegative-semidefinite solution and every unobservable eigenvalue of (A, B, E_1) is controllable.

Proof. To prove *i)*, suppose that P_1 and P_2 are maximal solutions of (12.14.23). Then, $P_1 \le P_2$ and $P_2 \le P_1$. Since "$\le$" is antisymmetric, it follows that $P_1 = P_2$.

To prove the necessity part of *ii)*, suppose that (12.14.23) has a nonnegative-semidefinite solution P and (A, B, E_1) has an imaginary eigenvalue that is unobservable and uncontrollable. Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that

$$SAS^{-1} = \begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad E_1S^{-1} = \begin{bmatrix} E_{11} & 0 \end{bmatrix},$$

where $\mathrm{spec}(A_2) \subset \jmath\mathbb{R}$. Therefore, $P = S^{\mathrm{T}}\left[\begin{smallmatrix} P_1 & P_{12} \\ P_{12}^{\mathrm{T}} & P_2 \end{smallmatrix}\right]S$, where P_2 satisfies $A_2^{\mathrm{T}}P_2 + P_2A_2 = 0$. Letting $\hat{P}_2$ be a nonzero nonnegative-semidefinite solution of $A_2^{\mathrm{T}}\hat{P}_2 + \hat{P}_2A_2 = 0$, it follows that $\hat{P} = S^{\mathrm{T}}\left[\begin{smallmatrix} P_1 & P_{12} \\ P_{12}^{\mathrm{T}} & P_2 + \alpha\hat{P}_2 \end{smallmatrix}\right]S$ is a solution of

(12.14.23) for all $\alpha > 0$. Therefore, (12.14.23) does not have a maximal solution.

To prove the sufficiency part of *ii*) suppose that (12.14.23) has a nonnegative-semidefinite solution and every unobservable imaginary eigenvalue of (A, B, E_1) is controllable. Then, (TBD). □

Theorem 12.15.12. (12.14.23) has a solution such that $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$ if and only if (TBD).

Theorem 12.15.13. The following statements are equivalent:

i) (12.14.23) has a maximal solution P satisfying $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$.

ii) (12.14.23) has a unique nonnegative-semidefinite solution P satisfying $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$.

iii) (A, B) is stabilizable.

Proof. To prove that $ii \Longrightarrow iii$), suppose that (A, B) is not stabilizable. If (A, B, E_1) has a CRHP eigenvalue that is uncontrollable and observable, then (12.14.23) does not have a nonnegative-semidefinite solution. If (12.14.23) has a nonnegative-semidefinite solution but (A, B, E_1) has an imaginary eigenvalue that is uncontrollable and unobservable, then (TBD).

Since (A, B) is stabilizable, it follows from Proposition 12.6.3 that there exists an invertible matrix $S \in \mathbb{R}^{n \times n}$ such that

$$SAS^{-1} = \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix},$$

where $A_1 \in \mathbb{R}^{n \times n}$, (A_1, B_1) is controllable, and A_2 is asymptotically stable. Next, since the pair (A_1, B_1) is controllable, it follows that there exists a positive-definite matrix $\hat{P}_0 \in \mathbb{R}^{n \times n}$ such that $A_1 - B_1 R_2^{-1} B_1^{\mathrm{T}} \hat{P}_0$ is asymptotically stable. It follows from Lemma 12.1 that one such matrix is given by

$$\hat{P}_0 = \left(\int_0^1 e^{-tA_1} B_1 R_2^{-1} B_1^{\mathrm{T}} e^{-tA_1^{\mathrm{T}}} \, \mathrm{d}t \right)^{-1}.$$

Thus, $A - \Sigma P_0$ is asymptotically stable where,

$$P_0 \triangleq \begin{bmatrix} \hat{P}_0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Next, it follows from Proposition 12.1 that there exists a nonnegative-semidefinite matrix $P \in \mathbb{R}^{n\times n}$ satisfying (12.14.23), that is,

$$(A - \Sigma P)^{\mathrm{T}} P + P(A - \Sigma P) + P\Sigma P + R_1 = 0. \tag{12.15.29}$$

Now, define a sequence of nonnegative-semidefinite matrices $\{P_k\}_{k=0}^{\infty}$ satisfying

$$(A - \Sigma P_k)^{\mathrm{T}} P_{k+1} + P_{k+1}(A - \Sigma P_k) + P_k\Sigma P_k + R_1' = 0, \quad k = 0, 1 \ldots, \tag{12.15.30}$$

where $R_1' \in \mathbb{R}^{n\times n}$ is symmetric and satisfies $R_1' \geq R_1$. Assuming now that $A - \Sigma P_k$ is asymptotically stable, we show that $A - \Sigma P_{k+1}$ is stable. To do this, first note that 12.15.40 and 12.15.41 imply that

$$(A - \Sigma P_k)^{\mathrm{T}} P + P(A - \Sigma P_k) + P_k\Sigma P_k - (P - P_k)\Sigma(P - P_k) + R_1 = 0. \tag{12.15.31}$$

Subtracting (12.15.31) from (12.15.30) yields

$$\begin{aligned}(A - \Sigma P_k)^{\mathrm{T}}(P_{k+1} - P) + (P_{k+1} - P)(A - \Sigma P_k)\\ + (P - P_k)\Sigma(P - P_k) + R_1' - R_1 = 0,\end{aligned} \tag{12.15.32}$$

which, since $A - \Sigma P_k$ is asymptotically stable, implies that

$$P_{k+1} - P = \int_0^\infty e^{t(A-\Sigma P_k)^{\mathrm{T}}}[(P - P_k)\Sigma(P - P_k) + R_1' - R_1]e^{t(A-\Sigma P_k)}\,\mathrm{d}t \geq 0. \tag{12.15.33}$$

Hence, $P_{k+1} \geq P$. Next, note that (12.15.30) is equivalent to

$$\begin{aligned}(A - \Sigma P_{k+1})^{\mathrm{T}}(P_{k+1} + P_{k+1}(A - \Sigma P_{k+1}) + P_{k+1}\Sigma P_{k+1} +\\ (P_{k+1} - P_k)\Sigma(P_{k+1} - P_k) + R_1' = 0.\end{aligned} \tag{12.15.34}$$

Subtracting (12.15.31) with k replaced by $k+1$ from (12.15.34) yields

$$\begin{aligned}(A - \Sigma P_k)^{\mathrm{T}}(P_{k+1} - P) + (P_{k+1} - P)(A - \Sigma P_{k+1}) =\\ -(P_{k+1} - P_k)\Sigma(P_{k+1} - P_k) - (P_{k+1} - P)\Sigma(P_{k+1} - P) - R_1' + R_1.\end{aligned} \tag{12.15.35}$$

Now, let $(A - \Sigma P_{k+1})x = \lambda x$ for $\lambda \in \mathbb{C}$ where $\operatorname{Re}\lambda \geq 0$ and nonzero $x \in \mathbb{C}^n$. Then, it follows from (12.15.35) that

$$(\lambda + \overline{\lambda})x^*(P_{k+1} - P)x = x^*Mx \tag{12.15.36}$$

where $M \leq 0$ denotes the right hand side of (12.15.35). Since $\lambda + \overline{\lambda} \geq 0$ and $P_{k+1} \geq P$, it follows from (12.15.36) that $x^*Mx = 0$, which in turn implies

$$x^*(P_{k+1} - P_k)\Sigma(P_{k+1} - P_k)x = 0. \tag{12.15.37}$$

Furthermore, since Σ is nonnegative-semidefinite, it follows that $\Sigma(P_{k+1} - P_k)x = 0$, which implies that

$$(A - \Sigma P_k)x = (A - \Sigma P_{k+1})x = \lambda x. \tag{12.15.38}$$

However, $A - \Sigma P_k$ is asymptotically stable, which implies that $\operatorname{Re}\lambda < 0$, which is a contradiction. Hence, $A - \Sigma P_{k+1}$ is asymptotically stable.

Next, subtract (12.15.34) with k replaced by $k-1$ from (12.15.30) to obtain

$$\begin{aligned}(A - \Sigma P_k)^{\mathrm{T}}(P_k - P_{k+1}) + (P_k - P_{k+1})(A - \Sigma P_k) + \\ (P_k - P_{k-1})\Sigma(P_k - P_{k-1}) = 0,\end{aligned} \tag{12.15.39}$$

which, since $A - \Sigma P_k$ is asymptotically stable, implies that

$$P_k - P_{k+1} = \int_0^\infty e^{t(A-\Sigma P_k)^{\mathrm{T}}}(P_k - P_{k+1})\Sigma(P_k - P_{k+1})e^{t(A-\Sigma P_k)}\,\mathrm{d}t \ge 0. \tag{12.15.40}$$

Hence, $\{P_k\}_{k=0}^\infty$ is a nonincreasing sequence of nonnegative-semidefinite matrices bounded from below by P. Thus, $P_+ \triangleq \lim_{k\to\infty} P_k$ exists.

Now, let $R_1' = R_1$. Letting $k \to \infty$, it follows from 12.21 that P_+ is a solution to (12.14.23). Furthermore, since $A - \Sigma P_k$ is asymptotically stable for all $k = 0, 1, 2, \ldots$, it follows that $\operatorname{Re}\lambda(A - \Sigma P_+) \le 0$. Also note that $P_+ \ge P$ for every solution P of (12.14.23), which implies that P_+ is the maximal solution of (12.14.23). □

Theorem 12.15.14. The following statements hold:

i) (12.14.23) has at most one stabilizing solution. If it exists, then it is nonnegative-semidefinite and maximal.

ii) (12.14.23) has a stabilizing solution if and only if (A, B) is stabilizable and every imaginary eigenvalue of (A, E_1) is observable.

Proof. To prove *i)*, suppose there exist two stabilizing solutions P_1 and P_2 to (12.14.23). Then,

$$A^{\mathrm{T}}P_1 + P_1A + R_1 - P_1\Sigma P_1 = 0, A^{\mathrm{T}}P_2 + P_2A + R_1 - P_2\Sigma P_2 = 0.$$

Subtracting these equations and rearranging yields

$$(A - \Sigma P_1)^{\mathrm{T}}(P_1 - P_2) + (P_1 - P_2)(A - \Sigma P_2) = 0.$$

Since $A - \Sigma P_1$ and $A - \Sigma P_2$ are asymptotically stable, it follows from Proposition 11.2 that $P_1 - P_2 = 0$. Hence, there exists at most one stabilizing solution to (12.14.23).

Next, suppose that there exists a stabilizing solution P to (12.14.23). Then, it follows from (12.14.23) that

$$P = \int_0^\infty e^{t(A-\Sigma P)^{\mathrm{T}}}(R_1 + P\Sigma P)e^{t(A-\Sigma P)}\,\mathrm{d}t,$$

which shows that P is nonnegative semidefinite. Next, let P' be a solution to (12.14.23). Then, it follows that

$$(A-\Sigma P)^{\mathrm{T}}(P-P') + (P-P')(A-\Sigma P) + (P-P')\Sigma(P-P') = 0,$$

which implies that $P' \le P$. Thus, P is the maximal solution to (12.14.23).

Finally, statement *ii*) follows from Theorem 1.5. □

Proposition 12.15.15. Suppose that (A,B) is stabilizable, let $R_1' \in \mathbb{R}^{n\times n}$ satisfy $R_1' \ge R_1$, and let P_+ and P_+' denote, respectively, the maximal solutions of (12.14.23) and

$$A^{\mathrm{T}}P + PA + R_1' - P\Sigma P = 0. \tag{12.15.41}$$

Then, $P_+' \ge P_+$.

Proof. Letting $k \to \infty$ in (12.15.39) it follows that $P_+' \triangleq \lim_{k\to\infty} P_k$ is a solution of (12.15.41) and satisfies $P_+' \ge P$ for every solution P of (12.14.23). In particular, $P_+' \ge P_+$. □

Proposition 12.15.16. Let $R_1 = 0$, and let $P \in \mathbb{R}^{n\times n}$ be a nonnegative-semidefinite solution to (12.15.38). Then, P is the maximal solution to (12.14.23) if and only if

$$\operatorname{mspec}(A-\Sigma P) = [\operatorname{mspec}(-A)\cap \mathrm{OLHP}] \cup [\operatorname{mspec}(A)\cap \mathrm{CLHP}]. \tag{12.15.42}$$

Proof. To prove necessity, let P be the maximal solution to (12.14.23) with $R_1 = 0$. Therefore, P satisfies

$$(A-\Sigma P)^{\mathrm{T}}P + PA = 0.$$

Next, let S be such that

$$\hat{P} = S^{\mathrm{T}}PS = \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix},$$

where P_1 is positive definite. Now, define $\hat{A} = S^{-1}AS$ and $\hat{\Sigma} = S^{-1}\Sigma S^{-\mathrm{T}}$

so that

$$(\hat{A} - \hat{\Sigma}\hat{P})^{\mathrm{T}}\hat{P} + \hat{P}\hat{A} = 0. \tag{12.15.43}$$

Letting

$$\hat{A} = \begin{bmatrix} \hat{A}_1 & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_2 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_1 & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{12}^{\mathrm{T}} & \hat{\Sigma}_2 \end{bmatrix},$$

(12.15.43) implies that

$$(\hat{A}_1 - \hat{\Sigma}_1\hat{P}_1)^{\mathrm{T}}P_1 + P_1\hat{A}_1 = 0, \tag{12.15.44}$$

$$P_1\hat{A}_{12} = 0. \tag{12.15.45}$$

Since P_1 is positive definite it follows from (12.15.44) and (12.15.45) that

$$(\hat{A}_1 - \hat{\Sigma}_1 P_1)^{\mathrm{T}} = -P_1\hat{A}_1P_1^{-1}, \tag{12.15.46}$$

$$\hat{A}_{12} = 0. \tag{12.15.47}$$

Hence,

$$\hat{A} - \hat{\Sigma}\hat{P} = \begin{bmatrix} -P_1^{-1}\hat{A}_1^{\mathrm{T}}P_1 & 0 \\ \hat{A}_{21} - \hat{\Sigma}_{21}^{\mathrm{T}}P_1 & \hat{A}_2 \end{bmatrix}, \tag{12.15.48}$$

where

$$\hat{A} = \begin{bmatrix} \hat{A}_1 & 0 \\ \hat{A}_{21} & \hat{A}_2 \end{bmatrix}. \tag{12.15.49}$$

Next, it follows from (12.15.48) that

$$\operatorname{mspec}(A - \Sigma P) = \operatorname{mspec}(-\hat{A}_1) \cup \operatorname{mspec}(\hat{A}_2). \tag{12.15.50}$$

Furthermore, Theorem 4.3.2 implies that $\operatorname{spec}(A - \Sigma P) \subset \mathrm{CLHP}$. Therefore,

$$\operatorname{mspec}(-\hat{A}_1) \subset \mathrm{CLHP} \tag{12.15.51}$$

and

$$\operatorname{mspec}(\hat{A}_2) \subset \mathrm{CLHP}, \tag{12.15.52}$$

and thus

$$\operatorname{mspec}(-\hat{A}_1) = \{-\lambda \in \operatorname{mspec}(\hat{A}_1)\colon\ \lambda \in \mathrm{ORHP}\} \cup \left[\operatorname{mspec}(\hat{A}_1) \cap \jmath\mathbb{R}\right]. \tag{12.15.53}$$

Next, it follows from (12.15.48) that

$$\operatorname{mspec}(A) = \operatorname{mspec}(\hat{A}_1) \cup \operatorname{mspec}(\hat{A}_2). \tag{12.15.54}$$

Now, combining (12.15.50)–(12.15.54) yields (12.15.42). Finally, sufficiency follows from Theorem 12.15.14. □

Corollary 12.15.17. Let $R_1 = 0$, and assume that $\operatorname{spec}(A) \subset \mathrm{CLHP}$. Then, $P = 0$ is the only nonnegative-semidefinite solution to (12.14.23).

12.16 Hamiltonian-Based Analysis of the Riccati Equation

We now analyze the Riccati equation by means of the $2n \times 2n$ Hamiltonian matrix

$$\mathcal{H} \triangleq \begin{bmatrix} A & \Sigma \\ R_1 & -A^{\mathrm{T}} \end{bmatrix}.$$

The Hamiltonian matrix is closely linked to the Riccati equation due to the fact that P is a solution to (12.14.23) if and only if P is symmetric and

$$\begin{bmatrix} P & I \end{bmatrix} \mathcal{H} \begin{bmatrix} P \\ I \end{bmatrix} = 0. \tag{12.16.1}$$

It is also useful to note that if P is a solution to (12.14.23), then

$$\mathcal{H} = \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix} \begin{bmatrix} A - \Sigma P & \Sigma \\ 0 & -(A - \Sigma P)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} I & 0 \\ P & I \end{bmatrix}. \tag{12.16.2}$$

It thus follows that

$$\operatorname{mspec}(\mathcal{H}) = \operatorname{mspec}(A - \Sigma P) \cup \operatorname{mspec}(-(A - \Sigma P)) \tag{12.16.3}$$

and

$$\chi_{\mathcal{H}}(s) = (-1)^n \chi_{A-\Sigma P}(s) \chi_{A-\Sigma P}(-s). \tag{12.16.4}$$

The factorization (12.16.4) of the characteristic polynomial of $\mathcal{H}$ is a *spectral decomposition.* It can be seen that the existence of a spectral decomposition requires that *i*) if λ is an element of the spectrum of $\mathcal{H}$, then $-\lambda$ is also an element of the spectrum of $\mathcal{H}$ with the same algebraic multiplicity, and *ii*) if λ is an element of the spectrum of $\mathcal{H}$ with $\operatorname{Re} \lambda = 0$, then λ must have even algebraic multiplicity. Note that the spectral decomposition (12.16.4) was obtained under the assumption that (12.14.23) has a solution.

We now show that the characteristic polynomial of the Hamiltonian matrix associated with the Riccati equation (12.14.23) has a spectral decomposition.

Corollary 12.16.1. Every imaginary eigenvalue of $\mathcal{H}$ has even algebraic multiplicity.

It is important to keep in mind that spectral decompositions are not unique. For example, if $\chi_{\mathcal{H}}(s) = (s+1)(s+2)(-s+1)(-s+2)$, then $\chi_{\mathcal{H}}(s) = p(s)p(-s) = \hat{p}(s)\hat{p}(-s)$, where $p(s) = (s+1)(s+2)$ and $\hat{p}(s) = (s+1)(s-2)$. Thus, the spectral factors $p(s)$ and $p(-s)$ can "trade" roots. These roots are the eigenvalues of $\mathcal{H}$.

Lemma 12.16.2. Let $\lambda \in \operatorname{spec}(A)$ be an uncontrollable eigenvalue of (A, B). Then, $\lambda \in \operatorname{spec}(\mathcal{H})$.

Proof. Since

$$\operatorname{rank}\begin{bmatrix} A^{\mathrm{T}} - \lambda I \\ B^{\mathrm{T}} \end{bmatrix} < n,$$

it follows that there exists nonzero $x \in \mathbb{R}^n$ such that $A^{\mathrm{T}}x = \lambda x$ and $B^{\mathrm{T}}x = 0$, and thus $\Sigma x = 0$. Now, note that

$$\mathcal{H}\begin{bmatrix} 0 \\ x \end{bmatrix} = \begin{bmatrix} \Sigma x \\ -A^{\mathrm{T}}x \end{bmatrix} = \begin{bmatrix} 0 \\ -\lambda x \end{bmatrix} = -\lambda \begin{bmatrix} 0 \\ x \end{bmatrix}.$$

Thus, $-\lambda \in \operatorname{spec}(\mathcal{H})$. Since $\mathcal{H}$ is Hamiltonian, it follows from Fact 4.9.14 that $\lambda \in \operatorname{spec}(\mathcal{H})$. □

Lemma 12.16.3. Let $\lambda \in \operatorname{spec}(A)$ be an unobservable eigenvalue of (A, E_1). Then, $\lambda \in \operatorname{spec}(\mathcal{H})$.

Proof. Since

$$\operatorname{rank}\begin{bmatrix} A - \lambda I \\ E_1 \end{bmatrix} < n,$$

it follows that there exists nonzero $y \in \mathbb{R}^n$ such that $Ay = \lambda y$ and $E_1 y = 0$. Now, note that

$$\mathcal{H}\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} Ay \\ E_1^{\mathrm{T}}E_1 y \end{bmatrix} = \begin{bmatrix} \lambda y \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} y \\ 0 \end{bmatrix}.$$

Thus, $\lambda \in \operatorname{spec}(\mathcal{H})$. □

Next, we present a partial converse of Lemma 12.16.2 and Lemma 12.16.3.

Lemma 12.16.4. Suppose $\lambda \in \operatorname{spec}(\mathcal{H})$ is such that $\operatorname{Re}\lambda = 0$. Then, λ is either an uncontrollable eigenvalue of (A, B) or an unobservable eigenvalue of (A, E_1).

Proof. Suppose that $\lambda = \jmath\omega$ is an eigenvalue of $\mathcal{H}$, where $\omega \in \mathbb{R}$. Then, there exist $x, y \in \mathbb{C}^n$ such that $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \neq 0$ and $\mathcal{H}\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \jmath\omega\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. Consequently,

$$Ax + \Sigma y = \jmath\omega x, \quad R_1 x - A^{\mathrm{T}}y = \jmath\omega y.$$

Rewriting these identities as

$$(A - \jmath\omega I)x = -\Sigma y, \quad (A - \jmath\omega I)^* y = R_1 x,$$

yields

$$y^*(A - \jmath\omega I)x = -y^*\Sigma y, \quad x^*(A - \jmath\omega I)^* y = x^* R_1 x.$$

Hence, $-y^*\Sigma y = x^*R_1x$, and thus $y^*\Sigma y = x^*R_1x = 0$, which implies that $B^{\mathrm{T}}y = 0$ and $E_1x = 0$. Consequently, we have

$$(A - \jmath\omega I)x = 0, \quad (A - \jmath\omega I)^*y = 0,$$

and hence

$$\begin{bmatrix} A - \jmath\omega I \\ E_1 \end{bmatrix} x = 0, \quad y^* \begin{bmatrix} A - \jmath\omega I & B \end{bmatrix} = 0.$$

Since $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \neq 0$, it follows that either $x \neq 0$ or $y \neq 0$, and thus either $\operatorname{rank} \left[\begin{smallmatrix} A-\jmath\omega I \\ E_1 \end{smallmatrix}\right] < n$ or $\operatorname{rank} \begin{bmatrix} A - \jmath\omega I & B \end{bmatrix} < n$. □

Combining Lemmas 12.16.2, 12.16.3, and 12.16.4 yields the following result.

Proposition 12.16.5. Suppose that $\lambda \in \mathbb{C}$ and $\operatorname{Re}\lambda = 0$. Then, λ is an eigenvalue of $\mathcal{H}$ if and only if λ is either an uncontrollable eigenvalue of (A, B) or an unobservable eigenvalue of (A, E_1).

Corollary 12.16.6. Suppose that (A, B) is stabilizable and every imaginary eigenvalue of (A, B, E_1) is either uncontrollable or observable. Then, $\mathcal{H}$ has no imaginary eigenvalues.

Theorem 12.16.7. The following statements are equivalent:

i) (A, B) is stabilizable, and every imaginary eigenvalue of (A, E_1) is observable.

ii) $\mathcal{H}$ has no imaginary eigenvalues, and, if $S = \left[\begin{smallmatrix} S_1 & S_{12} \\ S_{21} & S_2 \end{smallmatrix}\right] \in \mathbb{R}^{2n\times 2n}$ is an invertible matrix such that $\mathcal{H} = SZS^{-1}$, where $Z = \left[\begin{smallmatrix} Z_1 & Z_{12} \\ 0 & Z_2 \end{smallmatrix}\right] \in \mathbb{R}^{2n\times 2n}$ and $Z_1 \in \mathbb{R}^{n\times n}$ is asymptotically stable, then S_1 is invertible and $P \triangleq -S_{21}S_1^{-1}$ is the nonnegative-semidefinite stabilizing solution to (12.14.23).

In this case, the following statements hold:

iii) If (A, E_1) is detectable, then P is the only nonnegative-semidefinite solution to (12.14.23).

iv) $\operatorname{rank} P$ is equal to the number of OLHP observable eigenvalues of (A, E_1).

v) If all of the OLHP eigenvalues of (A, E_1) are observable, then P is positive definite.

Proof. To prove that *i)* implies *ii)*, first note that Corollary ??? implies that $\mathcal{H}$ has no imaginary eigenvalues. Since $\mathcal{H}$ is Hamiltonian, it follows that there exists $S = \left[\begin{smallmatrix} S_1 & S_{12} \\ S_{21} & S_2 \end{smallmatrix}\right] \in \mathbb{R}^{2n\times 2n}$ such that S is invertible and

$\mathcal{H} = SZS^{-1}$, where $Z = \begin{bmatrix} Z_1 & Z_{12} \\ 0 & Z_2 \end{bmatrix}$ and $Z_1 \in \mathbb{R}^{n\times n}$ is asymptotically stable.

Next, note that $\mathcal{H}S = SZ$ implies that $\mathcal{H}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} = S\begin{bmatrix} Z_1 \\ 0 \end{bmatrix} = \begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} Z_1$. Therefore,

$$\begin{aligned}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix}^{\mathrm{T}} J_n\mathcal{H}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} &= \begin{bmatrix} S_1 \\ S_{21} \end{bmatrix}^{\mathrm{T}} J_n\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} Z_1 \\ &= \begin{bmatrix} S_1^{\mathrm{T}} & S_{21}^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} S_{21} \\ -S_1 \end{bmatrix} Z_1 \\ &= LZ_1,\end{aligned}$$

where $L \triangleq S_1^{\mathrm{T}}S_{21} - S_{21}^{\mathrm{T}}S_1$. Since $J_n\mathcal{H} = (J_n\mathcal{H})^{\mathrm{T}}$, it follows that LZ_1 is symmetric, that is, $LZ_1 = Z_1^{\mathrm{T}}L^{\mathrm{T}}$. Since, in addition, L is skew symmetric, it follws that $0 = Z_1^{\mathrm{T}}L + LZ_1$. Now, since Z_1 is asymptotically stable, it follows that $L = 0$. Hence, $S_1^{\mathrm{T}}S_{21} = S_{21}^{\mathrm{T}}S_1$, which shows that $S_{21}^{\mathrm{T}}S_1$ is symmetric.

To show that S_1 is invertible, note that it follows from the identity $\begin{bmatrix} I & 0 \end{bmatrix}\mathcal{H}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} Z_1$ that $AS_1 + \Sigma S_{21} = S_1Z_1$. Now, let $x \in \mathbb{R}^n$ satisfy $S_1x = 0$. We thus have

$$\begin{aligned}x^{\mathrm{T}}S_{21}\Sigma S_{21}x = x^{\mathrm{T}}S_{21}^{\mathrm{T}}[AS_1 + \Sigma S_{21}]x &= x^{\mathrm{T}}S_{21}^{\mathrm{T}}S_1Z_1x \\ &= x^{\mathrm{T}}S_1^{\mathrm{T}}S_{21}Z_1x = 0,\end{aligned}$$

which implies that $B^{\mathrm{T}}S_{21}x = 0$. Hence, $S_1Z_1x = (AS_1 + \Sigma S_{21})x = 0$. Thus, $Z_1\colon\ \mathcal{N}(S_1) \mapsto \mathcal{N}(S_1)$.

Now, suppose that S_1 is singular. Since $Z_1\colon\ \mathcal{N}(S_1) \mapsto \mathcal{N}(S_1)$, it follows that there exists $\lambda \in \operatorname{spec}(Z_1)$ and $x \in \mathbb{C}^n$ such that $Z_1x = \lambda x$ and $S_1x = 0$. Forming $\begin{bmatrix} 0 & I \end{bmatrix}\mathcal{H}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} x = \begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} Z_1x$ yields $-A^{\mathrm{T}}S_{21}x = S_{21}\lambda Z$ and thus $\left(\lambda I + A^{\mathrm{T}}\right)S_{21}x = 0$. Since, in addition, as shown above, $B^{\mathrm{T}}S_{21}x = 0$, it follows that $x^*S_{21}^{\mathrm{T}}\begin{bmatrix} -\overline{\lambda}I - A & B \end{bmatrix} = 0$. Since $\lambda \in \operatorname{spec}(Z_1)$, it follows that $\operatorname{Re}(-\overline{\lambda}) > 0$. Furthermore, since, by assumption (A, B) is stabilizable, it follows that $\operatorname{rank}\begin{bmatrix} \overline{\lambda}I - A & B \end{bmatrix} = n$. Therefore, $S_{21}x = 0$. Combining this fact with $S_1x = 0$ yields $\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} x = 0$. Since x is nonzero, it follows that S is singular, which is a contradiction. Consequently, S_1 is invertible. Next, define $P \triangleq -S_{21}S_1^{-1}$ and note that, since $S_1^{\mathrm{T}}S_{21}$ is symmetric, it follows that $P = -S_1^{-\mathrm{T}}(S_1^{\mathrm{T}}S_{21})S_1^{-1}$ is also symmetric.

Since $\mathcal{H}\begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} = \begin{bmatrix} S_1 \\ S_{21} \end{bmatrix} Z_1$, it follows that

$$\mathcal{H}\begin{bmatrix} I \\ S_{21}S_1^{-1} \end{bmatrix} = \begin{bmatrix} I \\ S_{21}S_1^{-1} \end{bmatrix} S_1Z_1S_1^{-1},$$

and thus

$$\mathcal{H}\begin{bmatrix} I \\ -P \end{bmatrix} = \begin{bmatrix} I \\ -P \end{bmatrix} S_1 Z_1 S_1^{-1}.$$

Multiplying on the left by $\begin{bmatrix} P & I \end{bmatrix}$ yields

$$0 = \begin{bmatrix} P & I \end{bmatrix}\mathcal{H}\begin{bmatrix} I \\ -P \end{bmatrix} = A^{\mathrm{T}}P + PA + R_1 - P\Sigma P,$$

which shows that P is a solution to (12.14.23). Similarly, multiplying on the left by $\begin{bmatrix} I & 0 \end{bmatrix}$ yields $A - \Sigma P = S_1 Z_1 S_1^{-1}$. Since Z_1 is asymptotically stable, it follows that $A - \Sigma P$ is also asymptotically stable.

Conversely, to prove that *ii*) implies *i*), note that, since $A - \Sigma P$ is asymptotically stable, it follows that (A, B) is stabilizable. Furthermore, since P is a solution to (12.14.23), it follows that $\operatorname{mspec}(\mathcal{H}) = \operatorname{mspec}(A - \Sigma P) \cup \operatorname{mspec}(-(A - \Sigma P))$, which implies that $\mathcal{H}$ has no imaginary eigenvalues. Thus, Lemma 12.16.4 implies that (A, E_1) has no unobservable imaginary eigenvalues. Therefore, (A, B, E_1) has no imaginary eigenvalues that are controllable and unobservable.

To prove *iii*), (TO BE ADDED). □

Theorem 12.16.8. Suppose (A, B) is stabilizable. Then, there exists a solution to (12.14.23). Furthermore, the maximal solution P to (12.14.23) exists, is unique, and is nonnegative semidefinite. If $\lambda \in \operatorname{spec}(\mathcal{H})$ is imaginary, then λ has even-dimensional Jordan blocks. In addition,the following statements hold:

i) (A, E_1) observable implies P is positive definite.

ii) (A, E_1) is detectable if and only if P is nonnegative semidefinite.

iii) If $\lambda \in \operatorname{spec}(A)$ is imaginary, then λ is E_1-observable.

iv) λ is E_1-observable if and only if there are no eigenvalues of $\mathcal{H}$.

v) λ is E_1-observable if and only if $\operatorname{Re}\lambda < 0$.

12.17 Facts on Linear System Theory

Fact 12.17.1. If two of the following three conditions are satisfied, then the third condition is also satisfied:

i) A is asymptotically stable.

ii) (A, C) is observable.

iii) There exists a positive-definite solution $P \in \mathbb{R}^{n \times n}$ to (12.4.3).

Fact 12.17.2. The step response $y(t) = \int_0^t Ce^{tA}\,\mathrm{d}\tau Bv + Dv$ is bounded for all $v \in \mathbb{F}^m$ if and only if A is Lyapunov stable and nonsingular.

Fact 12.17.3. Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and assume that A is skew symmetric and (A,B) is controllable. Then, $A - \alpha BB^{\mathrm{T}}$ is asymptotically stable for all $\alpha > 0$.

Fact 12.17.4. Let $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, assume that (A,C) is detectable, and assume that $y(t) \to 0$ as $t \to \infty$, where $\dot{x}(t) = Ax(t)$ and $y(t) = Cx(t)$. Then, $x(t) \to 0$ as $t \to \infty$.

Fact 12.17.5. Let $x(0) = x_0$, and let $x_{\mathrm{f}} - e^{t_{\mathrm{f}}A}x_0 \in \mathcal{C}(A,B)$. Then, for all $t \in [0,t_{\mathrm{f}}]$, the control $u\colon\ [0,t_{\mathrm{f}}] \mapsto \mathbb{R}^m$ defined by

$$u(t) \triangleq B^{\mathrm{T}}e^{(t_{\mathrm{f}}-t)A^{\mathrm{T}}}\left(\int_0^{t_{\mathrm{f}}} e^{\tau A}BB^{\mathrm{T}}e^{\tau A^{\mathrm{T}}}\,\mathrm{d}\tau\right)^{+}\big(x_{\mathrm{f}} - e^{t_{\mathrm{f}}A}x_0\big)$$

yields $x(t_{\mathrm{f}}) = x_{\mathrm{f}}$.

Fact 12.17.6. Let $x(0) = x_0$, let $x_{\mathrm{f}} \in \mathbb{R}^n$, and assume that (A,B) is controllable. Then, for all $t \in [0,t_{\mathrm{f}}]$, the control $u\colon\ [0,t_{\mathrm{f}}] \mapsto \mathbb{R}^m$ defined by

$$u(t) \triangleq B^{\mathrm{T}}e^{(t_{\mathrm{f}}-t)A^{\mathrm{T}}}\left(\int_0^{t_{\mathrm{f}}} e^{\tau A}BB^{\mathrm{T}}e^{\tau A^{\mathrm{T}}}\,\mathrm{d}\tau\right)^{-1}\big(x_{\mathrm{f}} - e^{t_{\mathrm{f}}A}x_0\big)$$

yields $x(t_{\mathrm{f}}) = x_{\mathrm{f}}$.

Fact 12.17.7. Let $A \in \mathbb{R}^{n\times n}$ be asymptotically stable, let $V \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite, and let $Q \in \mathbb{R}^{n\times n}$ be the unique, positive-definite solution to $AQ + QA^{\mathrm{T}} + V = 0$. Furthermore, let $C \in \mathbb{R}^{p\times n}$, and assume that CVC^{T} is positive definite. Then, CQC^{T} is positive definite.

Fact 12.17.8. Let $A \in \mathbb{R}^{n\times n}$ be asymptotically stable, let $R \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite, and let $P \in \mathbb{R}^{n\times n}$ satisfy $A^{\mathrm{T}}P + PA + R = 0$. Then, there exist $\alpha_{ij} \in \mathbb{R}$ for all $i,j = 1,\ldots,n$, such that

$$P = \sum_{i,j=1}^{n} \alpha_{ij} A^{(i-1)\mathrm{T}} R A^{j-1}.$$

In particular, $\alpha_{ij} = \hat{P}_{(i,j)}$, where $\hat{P} \in \mathbb{R}^{n\times n}$ satisfies $\hat{A}^{\mathrm{T}}\hat{P} + \hat{P}\hat{A} + \hat{R} = 0$, where $\hat{A} = C(\chi_A)$ and $\hat{R} = E_{1,1}$. (Proof: See [511].) (Remark: This identity is *Smith's method*. See [178] for finite series solutions of linear matrix equations.)

Fact 12.17.9. Let $A \in \mathbb{R}^{n\times n}$ be asymptotically stable. Then,

$$(A \oplus A)^{-1} = \int_{-\infty}^{\infty} (\jmath\omega I - A)^{-1} \otimes (\jmath\omega I - A)^{-1}\, \mathrm{d}\omega$$

and

$$\int_{-\infty}^{\infty} (\omega^2 I + A^2)\, \mathrm{d}\omega = -\pi A^{-1}.$$

(Hint: Use $(\jmath\omega I - A)^{-1} + (-\jmath\omega I - A)^{-1} = -2A(\omega^2 I + A^2)^{-1}$.)

Fact 12.17.10. Let $G_1 \in \mathbb{R}^{p_1\times m}(s)$ and $G_2 \in \mathbb{R}^{p_2\times m}(s)$ be strictly proper. Then,

$$\left\| \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \right\|_{\mathrm{H}_2}^2 = \|G_1\|_{\mathrm{H}_2}^2 + \|G_2\|_{\mathrm{H}_2}^2.$$

Fact 12.17.11. Let $G_1, G_2 \in \mathbb{R}^{m\times m}(s)$ be strictly proper. Then,

$$\left\| \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \right\|_{\mathrm{H}_2} = \left\| \begin{bmatrix} G_1 & G_2 \end{bmatrix} \right\|_{\mathrm{H}_2}.$$

Fact 12.17.12. Let $H(t) = Ce^{tA}B$, where $C(sI - A)^{-1}B = \frac{\alpha}{s+\beta}$ and $\beta > 0$. Then,

$$\|H\|_{\mathrm{L}_2} = \frac{\alpha}{\sqrt{2\beta}}.$$

Fact 12.17.13. Let $H(t) = Ce^{tA}B$, where $C(sI - A)^{-1}B = \frac{\alpha_1 s + \alpha_0}{s^2 + \beta_1 s + \beta_0}$ and $\beta_1, \beta_0 > 0$. Then,

$$\|H\|_{\mathrm{L}_2} = \sqrt{\frac{\alpha_0^2}{2\beta_0\beta_1} + \frac{\alpha_1^2}{2\beta_1}}.$$

Fact 12.17.14. Let $G_1(s) = \frac{\alpha_1 s}{s+\beta_1}$ and $G_2(s) = \frac{\alpha_2 s}{s+\beta_2}$, where $\beta_1 > 0$ and $\beta_2 > 0$. Then,

$$\|G_1 G_2\|_{\mathrm{H}_2} \le \|G_1\|_{\mathrm{H}_2} \|G_2\|_{\mathrm{H}_2}$$

if and only if $\beta_1 + \beta_2 \ge 2$. (Remark: The H_2 norm is not submultiplicative.)

Fact 12.17.15. Let $A \in \mathbb{R}^{n\times n}$. Then, there exists a symmetric matrix $P \in \mathbb{R}^{n\times n}$ such that $A^{\mathrm{T}}P + PA$ is positive definite if and only if A has no eigenvalues on the imaginary axis. (Proof: See [446].)

Fact 12.17.16. Let $A, P \in \mathbb{R}^{n\times n}$, and assume that all of the eigenvalues of A are on the imaginary axis and P is nonnegative semidefinite. Then, $A^{\mathrm{T}}P + PA$ is either zero or has at least one positive eigenvalue and one negative eigenvalue. (Proof: See [561].)

Fact 12.17.17. Let $A \in \mathbb{R}^{n\times n}$, let $P \in \mathbb{R}^{n\times n}$ be symmetric, let $R \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite, and assume that $A^{\mathrm{T}}P + PA + R = 0$. Then,

$$|\nu_+(A) - \nu_+(P)| \le n - \operatorname{rank} \mathcal{O}(A, R)$$

and

$$|\nu_0(A) - \nu_0(P)| \le n - \operatorname{rank} \mathcal{O}(A, R).$$

(Proof: See [380].) (Remark: For related results, see [446] and references given in [380]. See also [162].)

Fact 12.17.18. Let $A_1, A_2 \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^n$, $C \in \mathbb{R}^{1\times n}$, assume that $A_1 \oplus A_2$ is nonsingular, and let $P \in \mathbb{R}^{n\times n}$ satisfy $A_1P + PA_2 + BC = 0$. If (A_1, B) is controllable and (A_2, C) is observable, then P is nonsingular.

Fact 12.17.19. Let $A_1 \in \mathbb{R}^{n_1\times n_1}$, $A_2 \in \mathbb{R}^{n_2\times n_2}$, $B \in \mathbb{R}^{n_1\times m}$, and $C \in \mathbb{R}^{m\times n_2}$, assume that $A_1 \oplus A_2$ is nonsingular, and assume that $\operatorname{rank} B = \operatorname{rank} C = m$. Furthermore, let $X \in \mathbb{R}^{n_1\times n_2}$ be the unique solution to $A_1X + XA_2 + BC = 0$. Then,

$$\operatorname{rank} X \le \min\{\operatorname{rank} \mathcal{K}(A_1, B), \operatorname{rank} \mathcal{O}(A_2, C)\}.$$

Finally, equality holds if $m = 1$. (Proof: See [167].) (Remark: Related results are given in [604, 608].)

Fact 12.17.20. Let $A \in \mathbb{R}^{n\times n}$, and assume that there exist nonnegative-semidefinite matrices $P, R \in \mathbb{R}^{n\times n}$ such that $A^{\mathrm{T}}P + PA + R = 0$ is satisfied and such that $\mathcal{N}(\mathcal{O}(A, R)) = \mathcal{N}(A)$. Then, A is semistable. (Proof: See [91].)

Fact 12.17.21. Let $A \in \mathbb{R}^{n\times n}$, let $R \in \mathbb{R}^{n\times n}$ be nonnegative semidefinite, and let $q, r \in \mathbb{R}$, where $r > 0$. If there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ satisfying

$$[A - (q + r)I]^{\mathrm{T}}P + P[A - (q + r)I] + \tfrac{1}{r}A^{\mathrm{T}}PA + R = 0,$$

then the spectrum of A is contained in disk centered at $q + \jmath 0$ with radius r. (Remark: See [61, 255] for related results concerning elliptical and parabolic regions.)

Fact 12.17.22. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, let $a, b \in \mathbb{R}$, where $a \ne 0$, and define $H(s) \triangleq G(as + b)$. Then,

$$H \sim \left[\begin{array}{c|c} a^{-1}(A - bI) & B \\ \hline a^{-1}C & D \end{array}\right].$$

Fact 12.17.23. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$, where A is nonsingular, and define

$H(s) \triangleq G(1/s)$. Then,

$$H \sim \left[\begin{array}{c|c} A^{-1} & -A^{-1}B \\ \hline CA^{-1} & D - CA^{-1}B \end{array}\right].$$

Fact 12.17.24. Let $G(s) = C(sI - A)^{-1}B$. Then,

$$G(\jmath\omega) = -CA(\omega^2 I + A^2)^{-1}B - \jmath\omega C(\omega^2 I + A^2)^{-1}B.$$

Fact 12.17.25. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$ and $H(s) = sG(s)$. Then,

$$H \sim \left[\begin{array}{c|c} A & B \\ \hline CA & CB \end{array}\right].$$

Consequently,

$$sC(sI - A)^{-1}B = CA(sI - A)^{-1}B + CB.$$

Fact 12.17.26. Let $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$, where $G_{ij} \sim \left[\begin{array}{c|c} A_{ij} & B_{ij} \\ \hline C_{ij} & D_{ij} \end{array}\right]$ for all $i, j = 1, 2$. Then,

$$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \sim \left[\begin{array}{cccc|cc} A_{11} & 0 & 0 & 0 & B_{11} & 0 \\ 0 & A_{12} & 0 & 0 & 0 & B_{12} \\ 0 & 0 & A_{21} & 0 & B_{21} & 0 \\ 0 & 0 & 0 & A_{22} & 0 & B_{22} \\ \hline C_{11} & C_{12} & 0 & 0 & D_{11} & D_{12} \\ 0 & 0 & C_{21} & C_{22} & D_{21} & D_{22} \end{array}\right].$$

Fact 12.17.27. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$, where $G \in \mathbb{R}^{l \times m}(s)$, and let $M \in \mathbb{R}^{m \times p}$. Then,

$$[I + GM]^{-1} \sim \left[\begin{array}{c|c} A - BMC & B \\ \hline -C & I \end{array}\right]$$

and

$$[I + GM]^{-1}G \sim \left[\begin{array}{c|c} A - BMC & B \\ \hline C & 0 \end{array}\right].$$

Fact 12.17.28. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$. If D has a left inverse D^{L}, then

$$G^{\mathrm{L}} \sim \left[\begin{array}{c|c} A - BD^{\mathrm{L}}C & BD^{\mathrm{L}} \\ \hline -D^{\mathrm{L}}C & D^{\mathrm{L}} \end{array}\right]$$

satisfies $G^{\mathrm{L}}G = I$. If D has a right inverse D^{R}, then

$$G^{\mathrm{R}} \sim \left[\begin{array}{c|c} A - BD^{\mathrm{R}}C & BD^{\mathrm{R}} \\ \hline -D^{\mathrm{R}}C & D^{\mathrm{R}} \end{array}\right]$$

satisfies $GG^{\mathrm{R}} = I$.

Fact 12.17.29. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, (A,B) is (controllable, stabilizable) if and only if (A, BB^{T}) is (controllable, stabilizable). In particular, if $A, B \in \mathbb{R}^{n\times n}$, where B is nonnegative semidefinite, then (A,B) is (controllable, stabilizable) if and only if $(A, B^{1/2})$ is (controllable, stabilizable).

Fact 12.17.30. Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $\hat{B} \in \mathbb{R}^{n\times \hat{m}}$, and assume that (A,B) is (controllable, stabilizable) and $\mathcal{R}(B) \subseteq \mathcal{R}(\hat{B})$. Then, $(A,\hat{B})$ is also (controllable, stabilizable).

Fact 12.17.31. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, the following statements are equivalent:

i) (A,B) is controllable.

ii) There exists $\alpha \in \mathbb{R}$ such that $(A+\alpha I, B)$ is controllable.

iii) $(A+\alpha I, B)$ is controllable for all $\alpha \in \mathbb{R}$.

Fact 12.17.32. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, the following statements are equivalent:

i) (A,B) is stabilizable.

ii) There exists $\alpha \le \max\{0, -\operatorname{spabs}(A)\}$ such that $(A+\alpha I, B)$ is stabilizable.

iii) $(A+\alpha I, B)$ is stabilizable for all $\alpha \le \max\{0, -\operatorname{spabs}(A)\}$.

Fact 12.17.33. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, assume that (A,B) is (controllable, stabilizable), and let $D \in \mathbb{R}^{n\times l}$, $E \in \mathbb{R}^{l\times n}$ and $R \in \mathbb{R}^{l\times l}$, where R is positive definite. Then, $\left(A+DE, \left[BB^{\mathrm{T}} + DRD^{\mathrm{T}}\right]^{1/2}\right)$ is also (controllable, stabilizable). (Proof: See [615, p. 79].)

Fact 12.17.34. Let $A \in \mathbb{R}^{n\times n}$ be diagonal and let $B \in \mathbb{R}^{n\times 1}$. Then, (A,B) is controllable if and only if the diagonal entries of A are distinct and all of the entries of B are nonzero. (Proof: Note that

$$\det\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = \det\left(\begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{bmatrix}\begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ \vdots & \vdots & \therefore & \vdots \\ 1 & a_n & \cdots & a_n^{n-1} \end{bmatrix}\right)$$

$$= \left(\prod_{i=1}^n b_i\right)\prod_{i<j}(a_i - a_j).)$$

Fact 12.17.35. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times 1}$, and assume that (A, B) is controllable. Then, A is cyclic.

Fact 12.17.36. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, the following conditions are equivalent:

i) (A, B) is (controllable, stabilizable) and A is nonsingular.

ii) (A, AB) is (controllable, stabilizable).

Fact 12.17.37. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, and assume that (A, B) is controllable. Then, $(A, B^{\mathrm{T}}S^{-\mathrm{T}})$ is observable, where $S \in \mathbb{R}^{n\times n}$ is a nonsingular matrix satisfying $A^{\mathrm{T}} = S^{-1}AS$.

Fact 12.17.38. Let $G \sim \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right]$ be a SISO rational transfer function, and let $\lambda \in \mathbb{C}$. Then, there exists a rational function H such that

$$G(s) = \frac{1}{(s+\lambda)^r}H(s)$$

and such that λ is neither a pole nor a zero of H if and only if the Jordan form of A has exactly one block associated with λ, which is of size $r \times r$.

Fact 12.17.39. Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{m\times n}$. Then,

$$\det[sI - (A + BC)] = [I - C(sI - A)^{-1}B]\det(sI - A).$$

(Proof: Note that

$$\begin{aligned}
[I - C(sI - A)^{-1}B]\det(sI - A) &= \det\begin{bmatrix} sI - A & B \\ C & I \end{bmatrix} \\
&= \det\begin{bmatrix} sI - A & B \\ C & I \end{bmatrix}\begin{bmatrix} I & 0 \\ -C & I \end{bmatrix} \\
&= \det\begin{bmatrix} sI - A - BC & B \\ 0 & I \end{bmatrix} \\
&= \det(sI - A - BC).)
\end{aligned}$$

Fact 12.17.40. Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, and $K \in \mathbb{R}^{m\times n}$, and assume that $A + BK$ is nonsingular. Then,

$$\det\begin{bmatrix} A & B \\ C & 0 \end{bmatrix} = (-1)^m \det(A + BK)\det\left[C(A + BK)^{-1}B\right].$$

Hence, $\left[\begin{smallmatrix} A & B \\ C & 0 \end{smallmatrix}\right]$ is nonsingular if and only if $C(A + BK)^{-1}B$ is nonsingular.

(Proof:

$$\begin{aligned}\det\begin{bmatrix} A & B \\ C & 0 \end{bmatrix} &= \det\begin{bmatrix} A & B \\ C & 0 \end{bmatrix}\begin{bmatrix} I & 0 \\ K & I \end{bmatrix} \\ &= \det\begin{bmatrix} A+BK & B \\ C & 0 \end{bmatrix} \\ &= \det(A+BK)\det\big[-C(A+BK)^{-1}B\big].)\end{aligned}$$

Fact 12.17.41. Let $A, B \in \mathbb{C}^{n\times n}$, and assume that the $2n \times 2n$ matrix

$$\begin{bmatrix} A & -2I \\ 2B - \frac{1}{2}A^2 & A \end{bmatrix}$$

is simple. Then, there exists $X \in \mathbb{C}^{n\times n}$ satisfying

$$X^2 + AX + B = 0.$$

(Proof: See [557].)

Fact 12.17.42. Let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite and, for all $t \ge 0$, let $P(t) \in \mathbb{R}^{n\times n}$ satisfy

$$\begin{aligned}\dot{P}(t) &= A^{\mathrm{T}}P(t) + P(t)A + P(t)VP(t), \\ P(0) &= P_0.\end{aligned}$$

Then, for all $t \ge 0$,

$$P(t) = e^{tA^{\mathrm{T}}}\left[P_0^{-1} - \int_0^t e^{\tau A}Ve^{\tau A^{\mathrm{T}}}\,\mathrm{d}\tau\right]^{-1}e^{tA}.$$

(Remark: $P(t)$ satisfies a Riccati differential equation.)

12.18 Notes

Linear system theory is treated in [112, 556, 611]. The PBH test is proved in [270]. Spectral factorization results are given in [146].

Zeros are treated in [199, 321, 385, 453, 495, 501].

Matrix-based methods for linear system identification are developed in [570].

Solutions of the LQR problem under weak conditions are given in [225]. Solutions of the Riccati equation are considered in [341, 343, 351, 352, 402, 480, 602, 607, 609]. There are numerous extensions to the results given in this chapter to various generalizations of (12.14.23). These include the case in which R_1 is indefinite [232, 605, 606] as well as the case in which Σ is indefinite [497]. The latter case is relevant to H_∞ optimal control theory [86]. Additional extensions include the Riccati inequality $A^{\mathrm{T}}P+PA+R_1-P\Sigma P \geq 0$ [475] as well as the discrete-time Riccati equation [306] and extensions to fixed-order controllers [302]. Monotonicity properties are studied in [607]. Riccati equations for discrete-time systems are discussed in [1].

Bibliography

[1] C. D. Ahlbrandt and A. C. Peterson, *Discrete Hamiltonian Systems: Difference Equations, Continued Fractions, and Riccati Equations.* Dordrecht: Kluwer, 1996.

[2] A. C. Aitken, *Determinants and Matrices*, 9th ed. Edinburgh: Oliver and Boyd, 1956.

[3] M. Al Ahmar, "An Identity of Jacobi," *Amer. Math. Monthly*, Vol. 103, pp. 78–79, 1996.

[4] A. A. Albert and B. Muckenhoupt, "On Matrices of Trace Zero," *Michigan Math. J.*, Vol. 4, pp. 1–3, 1957.

[5] A. E. Albert, "Conditions for Positive and Nonnegative Definiteness in Terms of Pseudoinverses," *SIAM J. Appl. Math.*, Vol. 17, pp. 434–440, 1969.

[6] A. E. Albert, *Regression and the Moore-Penrose Pseudoinverse.* New York: Academic Press, 1972.

[7] R. Aldrovandi, *Special Matrices of Mathematical Physics: Stochastic, Circulant and Bell Matrices.* Singapore: World Scientific, 2001.

[8] M. Alic, B. Mond, J. E. Pecaric, and V. Volenec, "Bounds for the Differences of Matrix Means," *SIAM J. Matrix Anal. Appl.*, Vol. 18, pp. 119–123, 1997.

[9] G. Alpargu and G. P. H. Styan, "Some Remarks and a Bibliography on the Kantorovich Inequality," in *Multidimensional Statistical Analysis and Theory of Random Matrices.* Utrecht: VSP, 1996, pp. 1–13.

[10] R. C. Alperin, "The Matrix of a Rotation," *College Math. J.*, Vol. 20, p. 230, 1989.

[11] S. L. Altmann, *Rotations, Quaternions, and Double Groups.* New York: Oxford University Press, 1986.

[12] H. Alzer, "A Lower Bound for the Difference Between the Arithmetic and Geometric Means," *Nieuw. Arch. Wisk.*, Vol. 8, pp. 195–197, 1990.

[13] B. D. O. Anderson, "Orthogonal Decomposition Defined by a Pair of Skew-Symmetric Forms," *Lin. Alg. Appl.*, Vol. 8, pp. 91–93, 1974.

[14] B. D. O. Anderson, "Weighted Hankel-Norm Approximation: Calculation of Bounds," *Sys. Contr. Lett.*, Vol. 7, pp. 247–255, 1986.

[15] B. D. O. Anderson, E. I. Jury, and M. Mansour, "Schwarz Matrix Properties for Continuous and Discrete Time Systems," *Int. J. Contr.*, Vol. 23, pp. 1–16, 1976.

[16] B. D. O. Anderson and J. B. Moore, "A Matrix Kronecker Lemma," *Lin. Alg. Appl.*, Vol. 15, pp. 227–234, 1976.

[17] W. N. Anderson, "Shorted Operators," *SIAM J. Appl. Math.*, Vol. 20, pp. 520–525, 1971.

[18] W. N. Anderson and R. J. Duffin, "Series and Parallel Addition of Matrices," *J. Math. Anal. Appl.*, Vol. 26, pp. 576–594, 1969.

[19] W. N. Anderson, E. J. Harner, and G. E. Trapp, "Eigenvalues of the Difference and Product of Projections," *Lin. Multilin. Alg.*, Vol. 17, pp. 295–299, 1985.

[20] W. N. Anderson and M. Schreiber, "On the Infimum of Two Projections," *Acta Sci. Math.*, Vol. 33, pp. 165–168, 1972.

[21] W. N. Anderson and G. E. Trapp, "Shorted Operators II," *SIAM J. Appl. Math.*, Vol. 28, pp. 60–71, 1975.

[22] W. N. Anderson and G. E. Trapp, "Symmetric Positive Definite Matrices," *Amer. Math. Monthly*, Vol. 95, pp. 261–262, 1988.

[23] T. Ando, "Concavity of Certain Maps on Positive Definite Matrices and Applications to Hadamard Products," *Lin. Alg. Appl.*, Vol. 26, pp. 203–241, 1979.

[24] T. Ando, "Majorizations and Inequalities in Matrix Theory," *Lin. Alg. Appl.*, Vol. 199, pp. 17–67, 1994.

[25] T. Ando, "Majorization Relations for Hadamard Products," *Lin. Alg. Appl.*, Vol. 223–224, pp. 57–64, 1995.

[26] T. Ando and F. Hiai, "Log-Majorization and Complementary Golden-Thompson Type Inequalities," *Lin. Alg. Appl.*, Vol. 197/198, pp. 113–131, 1994.

[27] T. Ando and F. Hiai, "Holder Type Inequalities for Matrices," *Math. Ineq. Appl.*, Vol. 1, pp. 1–30, 1998.

[28] T. Ando, C.-K. Li, and R. Mathias, "Geometric Means."

[29] T. Ando and X. Zhan, "Norm Inequalities Related to Operator Monotone Functions," *Math. Ann.*, Vol. 315, pp. 771–780, 1999.

[30] E. Andruchow, G. Corach, and D. Stojanoff, "Geometric Operator Inequalities," *Lin. Alg. Appl.*, Vol. 258, pp. 295–310, 1997.

[31] J. D. Aplevich, *Implicit Linear Systems.* Berlin: Springer, 1991.

[32] T. M. Apostol, "Some Explicit Formulas for the Exponential Matrix," *Amer. Math. Monthly*, Vol. 76, pp. 289–292, 1969.

[33] H. Araki, "On an Inequality of Lieb and Thirring," *Lett. Math. Phys.*, Vol. 19, pp. 167–170, 1990.

[34] A. Arimoto, "A Simple Proof of the Classification of Normal Toeplitz Matrices," *Elec. J. Lin. Alg.*, Vol. 9, pp. 108–111, 2002.

[35] T. Arponen, "A Matrix Approach to Polynomials," *Lin. Alg. Appl.*, Vol. 359, pp. 181–196, 2003.

[36] M. Artin, *Algebra.* Englewood Cliffs: Prentice-Hall, 1991.

[37] H. Aslaksen, "Laws of Trigonometry on SU(3)," *Trans. Amer. Math. Soc.*, Vol. 317, pp. 127–142, 1990.

[38] H. Aslaksen, "Quaternionic Determinants," *Math. Intell.*, Vol. 18(3), pp. 57–65, 1996.

[39] B. A. Asner, "On the Total Nonnegativity of the Hurwitz Matrix," *SIAM J. Appl. Math.*, Vol. 18, pp. 407–414, 1970.

[40] Y.-H. Au Yeung, "Some Inequalities for the Rational Power of a Nonnegative Definite Matrix," *Lin. Alg. Appl.*, Vol. 7, pp. 347–350, 1973.

[41] J. S. Aujla, "Some Norm Inequalities for Completely Monotone Functions," *SIAM J. Matrix Anal. Appl.*, Vol. 22, pp. 569–573, 2000.

[42] J. S. Aujla and F. C. Silva, "Weak Majorization Inequalities and Convex Functions," *Lin. Alg. Appl.*, Vol. 369, pp. 217–233, 2003.

[43] B. Aupetit and J. Zemanek, "A Characterization of Normal Matrices by Their Exponentials," *Lin. Alg. Appl.*, Vol. 132, pp. 119–121, 1990.

[44] O. Axelsson, *Iterative Solution Methods.* Cambridge: Cambridge University Press, 1994.

[45] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory.* New York: Springer, 2001.

[46] J. K. Baksalary, F. Pukelsheim, and G. P. H. Styan, "Some Properties of Matrix Partial Orderings," *Lin. Alg. Appl.*, Vol. 119, pp. 57–85, 1989.

[47] J. K. Baksalary and G. P. H. Styan, "Generalized Inverses of Bordered Matrices," *Lin. Alg. Appl.*, Vol. 354, pp. 41–47, 2002.

[48] C. S. Ballantine, "Products of Positive Semidefinite Matrices," *Pac. J. Math.*, Vol. 23, pp. 427–433, 1967.

[49] C. S. Ballantine, "Products of Positive Definite Matrices II," *Pac. J. Math.*, Vol. 24, pp. 7–17, 1968.

[50] C. S. Ballantine, "Products of Positive Definite Matrices III," *J. Algebra*, Vol. 10, pp. 174–182, 1968.

[51] C. S. Ballantine, "Products of Positive Definite Matrices IV," *Lin. Alg. Appl.*, Vol. 3, pp. 79–114, 1970.

[52] C. S. Ballantine, "Products of EP Matrices," *Lin. Alg. Appl.*, Vol. 12, pp. 257–267, 1975.

[53] C. S. Ballantine, "Some Involutory Similarities," *Lin. Multilin. Alg.*, Vol. 3, pp. 19–23, 1975.

[54] C. S. Ballantine, "Products of Involutory Matrices I," *Lin. Multilin. Alg.*, Vol. 5, pp. 53–62, 1977.

[55] C. S. Ballantine, "Products of Idempotent Matrices," *Lin. Alg. Appl.*, Vol. 19, pp. 81–86, 1978.

[56] C. S. Ballantine and C. R. Johnson, "Accretive Matrix Products," *Lin. Multilin. Alg.*, Vol. 3, pp. 169–185, 1975.

[57] R. B. Bapat and B. Zheng, "Generalized Inverses of Bordered Matrices," *Elec. J. Lin. Alg.*, Vol. 10, pp. 16–30, 2003.

[58] I. Y. Bar Itzhack, D. Hershkowitz, and L. Rodman, "Pointing in Real Euclidean Space," *J. Guid. Contr. Dyn.*, Vol. 20, pp. 916–922, 1997.

[59] S. Barnett, "A Note on the Bezoutian Matrix," *SIAM J. Applied Math.*, Vol. 22, pp. 84–86, 1972.

[60] S. Barnett, "Inversion of Partitioned Matrices with Patterned Blocks," *Int. J. Sys. Sci.*, Vol. 14, pp. 235–237, 1983.

[61] S. Barnett, *Polynomials and Linear Control Systems.* New York: Marcel Dekker, 1983.

[62] S. Barnett, *Matrices in Control Theory*, revised ed. Malabar: Krieger, 1984.

[63] S. Barnett, "Leverrier's Algorithm: A New Proof and Extensions," *SIAM J. Matrix Anal. Appl.*, Vol. 10, pp. 551–556, 1989.

[64] S. Barnett, *Matrices: Methods and Applications.* Oxford: Clarendon Press, 1990.

[65] S. Barnett and P. Lancaster, "Some Properties of the Bezoutian for Polynomial Matrices," *Lin. Multilin. Alg.*, Vol. 9, pp. 99–110, 1980.

[66] S. Barnett and C. Storey, *Matrix Methods in Stability Theory.* New York: Barnes and Noble, 1970.

[67] F. L. Bauer, J. Stoer, and C. Witzgall, "Absolute and Monotonic Norms," *Numer. Math.*, Vol. 3, pp. 257–264, 1961.

[68] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*, 2nd ed. John Wiley and Sons, 1993.

[69] N. Bebiano, J. da Providencia, and R. Lemos, "Matrix Inequalities in Statistical Mechanics," *Lin. Alg. Appl.*, 2003.

[70] E. F. Beckenbach and R. Bellman, *Inequalities.* Berlin: Springer, 1965.

[71] P. A. Bekker, "The Positive Semidefiniteness of Partitioned Matrices," *Lin. Alg. Appl.*, Vol. 111, pp. 261–278, 1988.

[72] J. G. Belinfante, B. Kolman, and H. A. Smith, "An Introduction to Lie Groups and Lie Algebras with Applications," *SIAM Rev.*, Vol. 8, pp. 11–46, 1966.

[73] G. R. Belitskii and Y. I. Lyubich, *Matrix Norms and Their Applications.* Basel: Birkhauser, 1988.

[74] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed. New York: McGraw-Hill, 1960, reprinted by SIAM, Philadelphia, 1995.

[75] R. Bellman, "Some Inequalities for the Square Root of a Positive Definite Matrix," *Lin. Alg. Appl.*, Vol. 1, pp. 321–324, 1968.

[76] A. Ben Israel, "A Note on Partitioned Matrices and Equations," *SIAM Rev.*, Vol. 11, pp. 247–250, 1969.

[77] A. Ben-Israel, "The Moore of the Moore-Penrose Inverse," *Elect. J. Lin. Alg.*, Vol. 9, pp. 150–157, 2002.

[78] A. Ben Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications.* New York: Wiley, 1973, reprinted by Krieger, Malabar.

[79] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization.* Philadelphia: SIAM, 2001.

[80] L. D. Berkovitz, *Convexity and Optimization in* $\mathbb{R}^n$. New York: John Wiley & Sons, 2002.

[81] A. Berman, M. Neumann, and R. J. Stern, *Nonnegative Matrices in Dynamic Systems*, 1994th ed., ser. Philadelphia. New York: John Wiley and Sons, 1989, reprinted by SIAM.

[82] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences.* New York: Academic Press, 1979, reprinted by SIAM, Philadelphia, 1979.

[83] D. S. Bernstein, "Inequalities for the Trace of Matrix Exponentials," *SIAM J. Matrix Anal. Appl.*, Vol. 9, pp. 156–158, 1988.

[84] D. S. Bernstein, "Some Open Problems in Matrix Theory Arising in Linear Systems and Control," *Lin. Alg. Appl.*, Vol. 162–164, pp. 409–432, 1992.

[85] D. S. Bernstein and S. P. Bhat, "Lyapunov Stability, Semistability, and Asymptotic Stability of Matrix Second-Order Systems," *ASME Trans. J. Vibr. Acoustics*, Vol. 117, pp. 145–153, 1995.

[86] D. S. Bernstein and W. M. Haddad, "LQG Control With an H_∞ Performance Bound: A Riccati Equation Approach," *IEEE Trans. Autom. Contr.*, Vol. 34, pp. 293–305, 1989.

[87] D. S. Bernstein, W. M. Haddad, D. C. Hyland, and F. Tyan, "Maximum Entropy-Type Lyapunov Functions for Robust Stability and Performance Analysis," *Sys. Contr. Lett.*, Vol. 21, pp. 73–87, 1993.

[88] D. S. Bernstein and D. C. Hyland, "Compartmental Modeling and Second-Moment Analysis of State Space Systems," *SIAM J. Matrix Anal. Appl.*, Vol. 14, pp. 880–901, 1993.

[89] D. S. Bernstein and W. So, "Some Explicit Formulas for the Matrix Exponential," *IEEE Trans. Autom. Contr.*, Vol. 38, pp. 1228–1232, 1993.

[90] K. V. Bhagwat and R. Subramanian, "Inequalities Between Means of Positive Operators," *Math. Proc. Camb. Phil. Soc.*, Vol. 83, pp. 393–401, 1978.

[91] S. P. Bhat and D. S. Bernstein, "Lyapunov Analysis of Semistability," *Proc. Amer. Contr. Conf.*, pp. 1608–1612, 1999, San Diego, CA.

[92] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues.* Essex: Longman Scientific and Technical, 1987.

[93] R. Bhatia, *Matrix Analysis.* New York: Springer, 1997.

[94] R. Bhatia and C. Davis, "More Matrix Forms of the Arithmetic-Geometric Mean Inequality," *SIAM J. Matrix Anal. Appl.*, Vol. 14, pp. 132–136, 1993.

[95] R. Bhatia and F. Kittaneh, "Norm Inequalities for Partitioned Operators and an Application," *Math. Anal.*, Vol. 287, pp. 719–726, 1990.

[96] R. Bhatia and F. Kittaneh, "On the Singular Values of a Product of Operators," *SIAM J. Matrix Anal. Appl.*, Vol. 11, pp. 272–277, 1990.

[97] R. Bhatia and F. Kittaneh, "Notes on Matrix Arithmetic-Geometric Mean Inequalities," *Lin. Alg. Appl.*, Vol. 308, pp. 203–211, 2000.

[98] R. Bhatia and P. Semrl, "Orthogonality of Matrices and Some Distance Problems," *Lin. Alg. Appl.*, Vol. 287, pp. 77–85, 1999.

[99] M. R. Bicknell, "The Lambda Number of a Matrix: The Sum of Its n^2 Cofactors," *Amer. Math. Monthly*, Vol. 72, pp. 260–264, 1965.

[100] A. Bjorck, *Numerical Methods for Least Squares Problems.* Philadelphia: SIAM, 1996.

[101] W. Boehm, "An Operator Limit," *SIAM Rev.*, Vol. 36, p. 659, 1994.

[102] A. Borck, *Numerical Methods for Least Squares Problems.* Philadelphia: SIAM, 1996.

[103] J. M. Borwein and A. M. Lewis, *Convex Analysis and Nonlinear Optimization.* New York: Springer, 2000.

[104] A. J. Bosch, "The Factorization of a Square Matrix Into Two Symmetric Matrices," *Amer. Math. Monthly*, Vol. 93, pp. 462–464, 1986.

[105] A. J. Bosch, "Note on the Factorization of a Square Matrix into Two Hermitian or Symmetric Matrices," *SIAM Rev.*, Vol. 29, pp. 463–468, 1987.

[106] T. L. Boullion and P. L. Odell, *Generalized Inverse Matrices.* New York: John Wiley and Sons, 1971.

[107] J.-C. Bourin, "Some Inequalities for Norms on Matrices and Operators," *Lin. Alg. Appl.*, Vol. 292, pp. 139–154, 1999.

[108] S. Boyd, "Entropy and Random Feedback," in *Open Problems in Mathematical Systems and Control Theory*, V. D. Blondel and et al, Eds. New York: Springer, 1998, pp. 71–74.

[109] J. L. Brenner, "Expanded Matrices from Matrices with Complex Elements," *SIAM Rev.*, Vol. 3, pp. 165–166, 1961.

[110] J. L. Brenner and J. S. Lim, "The Matrix Equations $A = XYZ$ and $B = ZYX$ and Related Ones," *Canad. Math. Bull.*, Vol. 17, pp. 179–183, 1974.

[111] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory," *IEEE Trans. Circ. Sys.*, Vol. CAS–25, pp. 772–781, 1978, Correction: CAS–26:360, 1979.

[112] R. Brockett, *Finite Dimensional Linear Systems.* New York: Wiley, 1970.

[113] E. T. Browne, *Introduction to the Theory of Determinants and Matrices.* Chapel Hill: The University of North Carolina Press, 1958.

[114] R. A. Brualdi and J. Q. Massey, "Some Applications of Elementary Linear Algebra in Combinatorics," *College Math. J.*, Vol. 24, pp. 10–19, 1993.

[115] R. A. Brualdi and S. Mellendorf, "Regions in the Complex Plane Containing the Eigenvalues of a Matrix," *Amer. Math. Monthly*, Vol. 101, pp. 975–985, 1994.

[116] R. A. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory.* Cambridge: Cambridge University Press, 1991.

[117] P. S. Bullen, *A Dictionary of Inequalities.* Longman, 1998, Essex.

[118] P. S. Bullen, *Handbook of Means and Their Inequalities.* Dordrecht: Kluwer Academic Publishers, 2003.

[119] P. S. Bullen, D. S. Mitrinovic, and P. M. Vasic, *Means and Their Inequalities.* Dordrecht: Reidel Publ. Co., 1988.

[120] A. Bultheel and M. Van Barel, *Linear Algebra, Rational Approximation and Orthogonal Polynomials.* Amsterdam: Elsevier, 1997.

[121] F. Burns, D. Carlson, E. V. Haynsworth, and T. L. Markham, "Generalized Inverse Formulas Using the Schur-Complement," *SIAM J. Appl. Math*, Vol. 26, pp. 254–259, 1974.

[122] P. J. Bushell and G. B. Trustrum, "Trace Inequalities for Positive Definite Matrix Power Products," *Lin. Alg. Appl.*, Vol. 132, pp. 173–178, 1990.

[123] S. L. Campbell, *Singular Systems.* London: Pitman, 1980.

[124] S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations.* Pitman, 1979, reprinted by Dover, Mineola, 1991.

[125] S. L. Campbell and N. J. Rose, "Singular Perturbation of Autonomous Linear Systems," *SIAM J. Math. Anal.*, Vol. 10, pp. 542–551, 1979.

[126] E. A. Carlen and E. H. Lieb, "A Minkowski Type Trace Inequality and Strong Subadditivity of Quantum Entropy," *Amer. Math. Soc. Transl.*, Vol. 189, pp. 59–62, 1999.

[127] D. Carlson, "Controllability, Inertia, and Stability for Tridiagonal Matrices," *Lin. Alg. Appl.*, Vol. 56, pp. 207–220, 1984.

[128] D. Carlson, E. V. Haynsworth, and T. L. Markham, "A Generalization of the Schur Complement by Means of the Moore-Penrose Inverse," *SIAM J. Appl. Math.*, Vol. 26, pp. 169–175, 1974.

[129] D. Carlson, C. R. Johnson, D. C. Lay, and A. D. Porter, Eds., *Linear Algebra Gems: Assets for Undergraduate Mathematics.* Washington, DC: The Mathematical Association of America, 2002.

[130] D. Carlson, C. R. Johnson, D. C. Lay, A. D. Porter, A. E. Watkins, and W. Watkins, Eds., *Resources for Teaching Linear Algebra.* Washington, DC: The Mathematical Association of America, 1997.

[131] P. Cartier, "Mathemagics, A Tribute to L. Euler and R. Feynman," in *Noise, Oscillators and Algebraic Randomness*, M. Planat, Ed. New York: Springer, 2000, pp. 6–67.

[132] D. I. Cartwright and M. J. Field, "A Refinement of the Arithmetic Mean-Geometric Mean Inequality," *Proc. Amer. Math. Soc.*, Vol. 71, pp. 36–38, 1978.

[133] F. S. Cater, "Products of Central Collineations," *Lin. Alg. Appl.*, Vol. 19, pp. 251–274, 1978.

[134] N. N. Chan and M. K. Kwong, "Hermitan Matrix Inequalities and a Conjecture," *Amer. Math. Monthly*, Vol. 92, pp. 533–541, 1985.

[135] H. Chapellat, M. Mansour, and S. P. Bhattacharyya, "Elementary Proofs of Some Classical Stability Criteria," *IEEE Trans. Educ.*, Vol. 33, pp. 232–239, 1990.

[136] F. Chatelin, *Eigenvalues of Matrices.* New York: John Wiley and Sons, 1993.

[137] J.-J. Chattot, *Computational Aerodynamics and Fluid Dynamics.* Berlin: Springer, 2002.

[138] V.-S. Chellaboina and W. M. Haddad, "Is the Frobenius Matrix Norm Induced?" *IEEE Trans. Autom. Contr.*, Vol. 40, pp. 2137–2139, 1995.

[139] V.-S. Chellaboina and W. M. Haddad, "Solution to 'Some Matrix Integral Identities'," *SIAM Rev.*, Vol. 39, pp. 763–765, 1997.

[140] V.-S. Chellaboina, W. M. Haddad, D. S. Bernstein, and D. A. Wilson, "Induced Convolution Operator Norms of Linear Dynamical Systems," *Math. Contr. Sig. Sys.*, Vol. 13, pp. 216–239, 2000.

[141] C.-T. Chen, *Linear System Theory and Design.* New York: Holt, Rhinehart, Winston, 1984.

[142] H.-W. Cheng and S. S.-T. Yau, "More Explicit Formulas for the Matrix Exponential," *Lin. Alg. Appl.*, Vol. 262, pp. 131–163, 1997.

[143] J. Chollet, "Some Inequalities for Principal Submatrices," *Amer. Math. Monthly*, Vol. 104, pp. 609–617, 1997.

[144] M. T. Chu, R. E. Funderlic, and G. H. Golub, "A Rank-One Reduction Formula and Its Application to Matrix Factorizations," *SIAM Rev.*, Vol. 37, pp. 512–530, 1995.

[145] N. L. C. Chui and J. M. Maciejowski, "Realization of Stable Models with Subspace Methods," *Automatica*, Vol. 32, pp. 1587–1595, 1996.

[146] D. J. Clements, B. D. O. Anderson, A. J. Laub, and J. B. Matson, "Spectral Factorization with Imaginary-Axis Zeros," *Lin. Alg. Appl.*, Vol. 250, pp. 225–252, 1997.

[147] R. E. Cline, "Representations for the Generalized Inverse of a Partitioned Matrix," *SIAM J. Appl. Math.*, Vol. 12, pp. 588–600, 1964.

[148] R. E. Cline and R. E. Funderlic, "The Rank of a Difference of Matrices and Associated Generalized Inverses," *Lin. Alg. Appl.*, Vol. 24, pp. 185–215, 1979.

[149] M. J. Cloud and B. C. Drachman, *Inequalities With Applications to Engineering.* New York: Springer, 1998.

[150] J. E. Cohen, "Spectral Inequalities for Matrix Exponentials," *Lin. Alg. Appl.*, Vol. 111, pp. 25–28, 1988.

[151] J. E. Cohen, S. Friedland, T. Kato, and F. P. Kelly, "Eigenvalue Inequalities for Products of Matrix Exponentials," *Lin. Alg. Appl.*, Vol. 45, pp. 55–95, 1982.

[152] D. K. Cohoon, "Sufficient Conditions for the Zero Matrix," *Amer. Math. Monthly*, Vol. 96, pp. 448–449, 1989.

[153] P. J. Costa and S. Rabinowitz, "Matrix Differentiation Identities," *SIAM Rev.*, Vol. 36, pp. 657–659, 1994.

[154] C. G. Cullen, "A Note on Normal Matrices," *Amer. Math. Monthly*, Vol. 72, pp. 643–644, 1965.

[155] C. G. Cullen, *Matrices and Linear Transformations*, 2nd ed. Reading: Addison-Wesley, 1972, reprinted by Dover, Mineola, 1990.

[156] W. J. Culver, "On the Existence and Uniqueness of the Real Logarithm of a Matrix," *Proc. Amer. Math. Soc.*, Vol. 17, pp. 1146–1151, 1966.

[157] M. L. Curtis, *Matrix Groups*, 2nd ed. New York: Springer-Verlag, 1984.

[158] R. D'Andrea, "Extension of Parrott's Theorem to Nondefinite Scalings," *IEEE Trans. Autom. Contr.*, Vol. 45, pp. 937–940, 2000.

[159] F. M. Dannan, "Matrix and Operator Inequalities," *J. Inequal. Pure. Appl. Math.*, Vol. 2, no. 3/34, pp. 1–7, 2001.

[160] R. Datko and V. Seshadri, "A Characterization and a Canonical Decomposition of Hurwitzian Matrices," *Amer. Math. Monthly*, Vol. 77, pp. 732–733, 1970.

[161] B. N. Datta, *Numerical Linear Algebra and Applications.* Pacific Grove: Brooks/Cole, 1995.

[162] B. N. Datta, "Stability and Inertia," *Lin. Alg. Appl.*, Vol. 302–303, pp. 563–600, 1999.

[163] P. J. Davis, *Circulant Matrices*, 2nd ed. New York: Chelsea, 1994.

[164] P. P. N. de Groen, "A Counterexample on Vector Norms and the Subordinate Matrix Norms," *Amer. Math. Monthly*, Vol. 97, pp. 406–407, 1990.

[165] J. de Pillis, "Transformations on Partitioned Matrices," *Duke Math. J.*, Vol. 36, pp. 511–515, 1969.

[166] J. de Pillis, "Inequalities for Partitioned Semidefinite Matrices," *Lin. Alg. Appl.*, Vol. 4, pp. 79–94, 1971.

[167] E. de Souza and S. P. Bhattacharyya, "Controllability, Observability and the Solution of $AX - XB = C$," *Lin. Alg. Appl.*, Vol. 39, pp. 167–188, 1981.

[168] H. P. Decell, "An Application of the Cayley-Hamilton Theorem to Generalized Matrix Inversion," *SIAM Rev.*, Vol. 7, pp. 526–528, 1965.

[169] J. W. Demmel, *Applied Numerical Linear Algebra.* Philadelphia: SIAM, 1997.

[170] E. D. Denman and A. N. Beavers, "The Matrix Sign Function and Computations in Systems," *Appl. Math. Computation*, Vol. 2, pp. 63–94, 1976.

[171] C. A. Desoer and H. Haneda, "The Measure of a Matrix as a Tool to Analyze Computer Algorithms for Circuit Analysis," *IEEE Trans. Circ. Thy.*, Vol. 19, pp. 480–486, 1972.

[172] E. Deutsch and M. Mlynarski, "Matricial Logarithmic Derivatives," *Lin. Alg. Appl.*, Vol. 19, pp. 17–31, 1978.

[173] L. Dieci, "Real Hamiltonian Logarithm of a Symplectic Matrix," *Lin. Alg. Appl.*, Vol. 281, pp. 227–246, 1998.

[174] J. Ding, "Perturbation of Systems in Linear Algebraic Equations," *Lin. Multilin. Alg.*, Vol. 47, pp. 119–127, 2000.

[175] J. Ding, "Lower and Upper Bounds for the Perturbation of General Linear Algebraic Equations," *Appl. Math. Lett.*, Vol. 14, pp. 49–52, 2001.

[176] J. Ding and W. C. Pye, "On the Spectrum and Pseudoinverse of a Special Bordered Matrix," *Lin. Alg. Appl.*, Vol. 331, pp. 11–20, 2001.

[177] A. Dittmer, "Cross Product Identities in Arbitrary Dimension," *Amer. Math. Monthly*, Vol. 101, pp. 887–891, 1994.

[178] T. E. Djaferis and S. K. Mitter, "Algebraic Methods for the Study of Some Linear Matrix Equations," *Lin. Alg. Appl.*, Vol. 44, pp. 125–142, 1982.

[179] D. Z. Djokovic, "Product of Two Involutions," *Arch. Math.*, Vol. 18, pp. 582–584, 1967.

[180] D. Z. Djokovic, "On Some Representations of Matrices," *Lin. Multilin. Alg.*, Vol. 4, pp. 33–40, 1976.

[181] D. Z. Djokovic and O. P. Lossers, "A Determinant Inequality," *Amer. Math. Monthly*, Vol. 83, pp. 483–484, 1976.

[182] D. Z. Djokovic and J. Malzan, "Products of Reflections in the Unitary Group," *Proc. Amer. Math. Soc.*, Vol. 73, pp. 157–160, 1979.

[183] D. Z. Dokovic, "On the Product of Two Alternating Matrices," *Amer. Math. Monthly*, Vol. 98, pp. 935–936, 1991.

[184] W. F. Donoghue, *Monotone Matrix Functions and Analytic Continuation.* New York: Springer, 1974.

[185] H. G. Eggleston, *Convexity.* Cambridge University Press, Cambridge, 1958.

[186] L. Elsner and K. D. Ikramov, "Normal Matrices: An update," *Lin. Alg. Appl.*, Vol. 285, pp. 291–303, 1998.

[187] L. Elsner, C. R. Johnson, and J. A. D. DaSilva, "The Perron Root of a Weighted Geometric Mean of Nonnegative Matrices," *Lin. Multilin. Alg.*, Vol. 24, pp. 1–13, 1988.

[188] L. Elsner and M. H. C. Paardekooper, "On Measures of Nonnormality of Matrices," *Lin. Alg. Appl.*, Vol. 92, pp. 107–124, 1987.

[189] L. Elsner and T. Szulc, "Convex Sets of Schur Stable and Stable Matrices," *Lin. Multilin. Alg.*, Vol. 48, pp. 1–19, 2000.

[190] K. Engo, "On the BCH formula in so(3)," *Numerical Mathematics BIT*, Vol. 41, pp. 629–632, 2001.

[191] S. Fallat and M. J. Tsatsomeros, "On the Cayley Transform of Positivity Classes of Matrices," *Elec. J. Lin. Alg.*, Vol. 9, pp. 190–196, 2002.

[192] K. Fan, "Generalized Cayley Transforms and Strictly Dissipative Matrices," *Lin. Alg. Appl.*, Vol. 5, pp. 155–172, 1972.

[193] K. Fan, "On Real Matrices with Positive Definite Symmetric Component," *Lin. Multilinear Alg.*, Vol. 1, pp. 1–4, 1973.

[194] K. Fan, "On Strictly Dissipative Matrices," *Lin. Alg. Appl.*, Vol. 9, pp. 223–241, 1974.

[195] Y. Fang, K. A. Loparo, and X. Feng, "Inequalities for the Trace of Matrix Product," *IEEE Trans. Autom. Contr.*, Vol. 39, pp. 2489–2490, 1994.

[196] A. Fassler and E. Stiefel, *Group Theoretical Methods and Their Applications.* Boston: Birkhauser, 1992.

[197] A. E. Fekete, *Real Linear Algebra.* New York: Marcel Dekker, 1985.

[198] B. Q. Feng, "Equivalence Constants for Certain Matrix Norms," *Lin. Alg. Appl.*, Vol. 374, pp. 247–254, 2003.

[199] P. G. Ferreira and S. P. Bhattacharyya, "On Blocking Zeros," *IEEE Trans. Autom. Contr.*, Vol. AC–22, pp. 258–259, 1977.

[200] J. H. Ferziger and M. Peric, *Computational Methods for Fluid Dynamics*, 3rd ed. Berlin: Springer, 2002.

[201] M. Fiedler, "A Note on the Hadamard Product of Matrices," *Lin. Alg. Appl.*, Vol. 49, pp. 233–235, 1983.

[202] M. Fiedler, *Special Matrices and Their Applications in Numerical Mathematics.* Dordrecht: Martinus Nijhoff, 1986.

[203] M. Fiedler and T. L. Markham, "A Characterization of the Moore-Penrose Inverse," *Lin. Alg. Appl.*, Vol. 179, pp. 129–133, 1993.

[204] M. Fiedler and T. L. Markham, "An Observation on the Hadamard Product of Hermitian Matrices," *Lin. Alg. Appl.*, Vol. 215, pp. 179–182, 1995.

[205] M. Fiedler and V. Ptak, "A New Positive Definite Geometric Mean of Two Positive Definite Matrices," *Lin. Alg. Appl.*, Vol. 251, pp. 1–20, 1997.

[206] P. A. Fillmore, "On Similarity and the Diagonal of a Matrix," *Amer. Math. Monthly*, Vol. 76, pp. 167–169, 1969.

[207] H. Flanders, "Methods of Proof in Linear Algebra," *Amer. Math. Monthly*, Vol. 63, pp. 1–15, 1956.

[208] H. Flanders, "On the Norm and Spectral Radius," *Lin. Multilin. Alg.*, Vol. 2, pp. 239–240, 1974.

[209] T. M. Flett, *Differential Analysis.* Cambridge: Cambridge University Press, 1980.

[210] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer Graphics Principles and Practice*, 2nd ed. Reading: Addison-Wesley, 1990.

[211] E. Formanek, "Polynomial Identities and the Cayley-Hamilton Theorem," *Mathematical Intelligencer*, Vol. 11, pp. 37–39, 1989.

[212] E. Formanek, *The Polynomial Identities and Invariants of $n \times n$ Matrices.* Providence, RI: American Mathematical Society, 1991.

[213] B. A. Francis, *A Course in H_∞ Control Theory.* New York: Springer-Verlag, 1987.

[214] J. Franklin, *Matrix Theory.* Englewood Cliffs: Prentice-Hall, 1968.

[215] M. Frazier, *An Introduction to Wavelets Through Linear Algebra.* New York: Springer, 1999.

[216] P. A. Fuhrmann, *A Polynomial Approach to Linear Algebra.* New York: Springer, 1996.

[217] A. T. Fuller, "Conditions for a Matrix to Have Only Characteristic Roots with Negative Real Parts," *J. Math. Anal. Appl.*, Vol. 23, pp. 71–98, 1968.

[218] T. Furuta, "$A \geq B \geq 0$ Assures $(B^r A^p B^r)^{1/q} \geq B^{(p+2r)/q}$ for $r \geq 0,\ p \geq 0,\ q \geq 1$ with $(1+2r)q \geq p+2r$," *Proc. Amer. Math. Soc.*, Vol. 101, pp. 85–88, 1987.

[219] T. Furuta, "Norm Inequalities Equivalent to Loewner-Heinz Theorem," *Rev. Math. Phys.*, Vol. 1, pp. 135–137, 1989.

[220] F. Gaines, "A Note on Matrices with Zero Trace," *Amer. Math. Monthly*, Vol. 73, pp. 630–631, 1966.

[221] F. Gaines, "A Note on Matrices with Zero Trace," *Amer. Math. Monthly*, Vol. 73, pp. 630–631, 1966.

[222] F. R. Gantmacher, *The Theory of Matrices.* New York: Chelsea, 1959, Vol. I.

[223] F. R. Gantmacher, *The Theory of Matrices.* New York: Chelsea, 1959, Vol. II.

[224] J. Garloff and D. G. Wagner, "Hadamard Products of Stable Polynomials Are Stable," *J. Math. Anal. Appl.*, Vol. 202, pp. 797–809, 1996.

[225] T. Geerts, "A Necessary and Sufficient Condition for Solvability of the Linear-Quadratic Control Problem without Stability," *Sys. Contr. Lett.*, Vol. 11, pp. 47–51, 1988.

[226] A. Gerrard and J. M. Burch, *Introduction to Matrix Methods in Optics.* New York: Wiley, 1975.

[227] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications.* New York: John Wiley and Sons, 1974.

[228] M. L. Glasser, "Exponentials of Certain Hilbert Space Operators," *SIAM Rev.*, Vol. 34, pp. 498–500, 1992.

[229] S. K. Godunov, *Modern Aspects of Linear Algebra.* Providence: American Mathematical Society, 1998.

[230] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials.* New York: Academic Press, 1982.

[231] I. Gohberg, P. Lancaster, and L. Rodman, *Invariant Subspaces of Matrices with Applications.* New York: John Wiley and Sons, 1986.

[232] I. Gohberg, P. Lancaster, and L. Rodman, "On Hermitian Solutions of the Symmetric Algebraic Riccati Equation," *SIAM J. Contr. Optim.*, Vol. 24, pp. 1323–1334, 1986.

[233] M. Goldberg, "Mixed Multiplicativity and l_p Norms for Matrices," *Lin. Alg. Appl.*, Vol. 73, pp. 123–131, 1986.

[234] M. Goldberg, "Equivalence Constants for l_p Norms of Matrices," *Lin. Multilin. Alg.*, Vol. 21, pp. 173–179, 1987.

[235] M. Goldberg, "Multiplicativity Factors and Mixed Multiplicativity," *Lin. Alg. Appl.*, Vol. 97, pp. 45–56, 1987.

[236] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: The Johns Hopkins University Press, 1996.

[237] N. C. Gonzalez, J. J. Koliha, and Y. Wei, "Integral Representation of the Drazin Inverse," *Electronic J. Lin. Alg.*, Vol. 9, pp. 129–131, 2002.

[238] N. Gordon and D. Salmond, "Bayesian Pattern Matching Technique for Target Acquisition," *J. Guid. Contr. Dyn.*, Vol. 22, pp. 68–77, 1999.

[239] W. Govaerts and B. Sijnave, "Matrix Manifolds and the Jordan Structure of the Bialternate Matrix Product," *Lin. Alg. Appl.*, Vol. 292, pp. 245–266, 1999.

[240] R. Gow, "The Equivalence of an Invertible Matrix to its Transpose," *Lin. Alg. Appl.*, Vol. 8, pp. 329–336, 1980.

[241] R. Gow and T. J. Laffey, "Pairs of Alternating Forms and Products of Two Skew-Symmetric Matrices," *Lin. Alg. Appl.*, Vol. 63, pp. 119–132, 1984.

[242] A. Graham, *Kronecker Products and Matrix Calculus With Applications.* Chichester: Ellis Horwood, 1981.

[243] J. F. Grcar, "A Matrix Lower Bound."

[244] W. Greub, *Linear Algebra.* New York: Springer, 1981.

[245] T. N. E. Greville, "Solutions of the Matrix Equation $XAX = X$ and Relations Between Oblique and Orthogonal projectors," *SIAM J. Appl. Math*, Vol. 26, pp. 828–832, 1974.

[246] R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, "Normal Matrices," *Lin. Alg. Appl.*, Vol. 87, pp. 213–225, 1987.

[247] J. Gross, "On the Product of Orthogonal Projectors," *Lin. Alg. Appl.*, Vol. 289, pp. 141–150, 1999.

[248] J. Gross, G. Trenkler, and S.-O. Troschke, "Quaternions: Further Contributions to a Matrix Oriented Approach," *Lin. Alg. Appl.*, Vol. 326, pp. 205–213, 2001.

[249] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions.* Boca Raton: CRC Press, 1999.

[250] K. Gurlebeck and W. Sprossig, *Quaternionic and Clifford Calculus for Physicists and Engineers.* New York: Chichester, 1997.

[251] K. E. Gustafson, "Matrix Trigonometry," *Lin. Alg. Appl.*, Vol. 217, pp. 117–140, 1995.

[252] K. E. Gustafson and D. K. M. Rao, *Numerical Range.* New York: Springer, 1997.

[253] W. H. Gustafson, P. R. Halmos, and H. Radjavi, "Products of Involutions," *Lin. Alg. Appl.*, Vol. 13, pp. 157–162, 1976.

[254] W. M. Haddad and D. S. Bernstein, "Robust Stabilization with Positive Real Uncertainty: Beyond the Small Gain Theorem," *Sys. Contr. Lett.*, Vol. 17, pp. 191–208, 1991.

[255] W. M. Haddad and D. S. Bernstein, "Controller Design with Regional Pole Constraints," *IEEE Trans. Autom. Contr.*, Vol. 37, pp. 54–69, 1992.

[256] W. W. Hager, "Updating the Inverse of a Matrix," *SIAM Rev.*, Vol. 31, pp. 221–239, 1989.

[257] W. Hahn, *Stability of Motion.* Berlin: Springer-Verlag, 1967.

[258] P. R. Halmos, *Finite-Dimensional Vector Spaces.* Princeton: Van Nostrand, 1958, reprinted by Springer, New York, 1974.

[259] P. R. Halmos, *A Hilbert Space Problem Book.* New York: Springer, 1980.

[260] P. R. Halmos, "Bad Products of Good Matrices," *Lin. Alg. Appl.*, Vol. 29, pp. 1–20, 1991.

[261] P. R. Halmos, *Problems for Mathematicians Young and Old.* Washington, D.C.: The Mathematical Association of America, 1991.

[262] P. R. Halmos, *Linear Algebra Problem Book.* Washington, D.C.: The Mathematical Association of America, 1995.

[263] L. A. Harris, "The Inverse of a Block Matrix," *Amer. Math. Monthly*, Vol. 102, pp. 656–657, 1995.

[264] W. A. Harris, J. P. Fillmore, and D. R. Smith, "Matrix Exponentials–Another Approach," *SIAM Rev.*, Vol. 43, pp. 694–706, 2001.

[265] D. J. Hartfiel, *Nonhomogeneous Matrix Products.* Singapore: World Scientific, 2002.

[266] R. E. Hartwig, "Block Generalized Inverses," *Arch. Rat. Mech. Anal.*, Vol. 61, pp. 197–251, 1976.

[267] R. E. Hartwig, "A Note on the Partial Ordering of Positive Semi-Definite Matrices," *Lin. Multilinear Alg.*, Vol. 6, pp. 223–226, 1978.

[268] R. E. Hartwig and I. J. Katz, "On Products of EP Matrices," *Lin. Alg. Appl.*, Vol. 252, pp. 339–345, 1997.

[269] D. A. Harville, *Matrix Algebra from a Statistician's Perspective.* New York: Springer, 1997.

[270] M. L. J. Hautus, "Controllability and Observability Conditions of Linear Autonomous Systems," *Proc. Koniklijke Akademic Van Wetenshappen*, Vol. 72, pp. 443–448, 1969.

[271] T. Haynes, "Stable Matrices, theCayley Transform, and convergent Matrices," *Int. J. Math. Math. Sci.*, Vol. 14, pp. 77–81, 1991.

[272] E. V. Haynsworth, "Applications of an Inequality for the Schur Complement," *Proc. Amer. Math. Soc.*, Vol. 24, pp. 512–516, 1970.

[273] U. Helmke and P. A. Fuhrmann, "Bezoutians," *Lin. Alg. Appl.*, Vol. 122–124, pp. 1039–1097, 1989.

[274] B. W. Helton, "Logarithms of Matrices," *Proc. Amer. Math. Soc.*, Vol. 19, pp. 733–738, 1968.

[275] H. V. Henderson, F. Pukelsheim, and S. R. Searle, "On the History of the Kronecker Product," *Lin. Multilin. Alg.*, Vol. 14, pp. 113–120, 1983.

[276] H. V. Henderson and S. R. Searle, "The Vec-Permutation Matrix, The Vec Operator and Kronecker Products: A Review," *Lin. Multilin. Alg.*, Vol. 9, pp. 271–288, 1981.

[277] N. J. Higham, "Newton's Method for the Matrix Square Root," *Math. Computation*, Vol. 46, pp. 537–549, 1986.

[278] N. J. Higham, "Matrix Nearness Problems and Applications," in *Applications of Matrix Theory*, M. J. C. Gover and S. Barnett, Eds. Oxford: Oxford University Press, 1989, pp. 1–27.

[279] N. J. Higham, "Estimating the Matrix p-Norm," *Numer. Math.*, Vol. 62, pp. 539–555, 1992.

[280] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. Philadelphia: SIAM, 2002.

[281] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra.* New York: Academic Press, 1974.

[282] O. Hirzallah and F. Kittaneh, "Matrix Young Inequalities for the Hilbert-Schmidt Norm," *Lin. Alg. Appl.*, Vol. 308, pp. 77–84, 2000.

[283] O. Hirzallah and F. Kittaneh, "Non-commutative Clarkson Inequalities for Unitarily Invariant Norms," *Pacific J. Math.*, Vol. 202, pp. 363–369, 2002.

[284] A. Hmamed, “A Matrix Inequality,” *Int. J. Contr.*, Vol. 49, pp. 363–365, 1989.

[285] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. Englewood Cliffs: Prentice-Hall, 1971.

[286] Y. Hong and R. A. Horn, “The Jordan Canonical form of a Product of a Hermitian and a Positive Semidefinite Matrix,” *Lin. Alg. Appl.*, Vol. 147, pp. 373–386, 1991.

[287] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge: Cambridge University Press, 1985.

[288] R. A. Horn and C. R. Johnson, “Hadamard and Conventional Submultiplicativity for Unitarily Invariant Norms on Matrices,” *Lin. Multilin. Alg.*, Vol. 20, pp. 91–106, 1987.

[289] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis.* Cambridge: Cambridge University Press, 1991.

[290] R. A. Horn and R. Mathias, “An Analog of the Cauchy-Schwarz Inequality for Hadamard Products and Unitarily Invariant Norms,” *SIAM J. Matrix Anal. Appl.*, Vol. 11, pp. 481–498, 1990.

[291] R. A. Horn and R. Mathias, “Cauchy-Schwarz Inequalities Associated with Positive Semidefinite Matrices,” *Lin. Alg. Appl.*, Vol. 142, pp. 63–82, 1990.

[292] R. A. Horn and R. Mathias, “Block-Matrix Generalizations of Schur’s Basic Theorems on Hadamard Products,” *Lin. Alg. Appl.*, Vol. 172, pp. 337–346, 1992.

[293] R. A. Horn and I. Olkin, “When Does $a^*a = b^*b$ and Why Does One Want to Know?” *Amer. Math. Monthly*, Vol. 103, pp. 470–482, 1996.

[294] R. A. Horn and G. G. Piepmeyer, “Two Applications of the Theory of Primary Matrix Functions,” *Lin. Alg. Appl.*, Vol. 361, pp. 99–106, 2003.

[295] B. G. Horne, “Lower Bounds for the Spectral Radius of a Matrix,” *Lin. Alg. Appl.*, Vol. 263, pp. 261–273, 1997.

[296] S.-H. Hou, “A Simple Proof of the Leverrier-Faddeev Characteristic Polynomial Algorithm,” *SIAM Rev.*, Vol. 40, pp. 706–709, 1998.

[297] A. S. Householder, *The Theory of Matrices in Numerical Analysis.* New York: Blaisdell Publishing Company, 1964, reprinted by Dover, New York, 1975.

[298] A. S. Householder, “Bezoutiants, Elimination and Localization,” *SIAM Rev.*, Vol. 12, pp. 73–78, 1970.

[299] R. Howe, “Very Basic Lie Theory,” *Amer. Math. Monthly*, Vol. 90, pp. 600–623, 1983.

[300] G.-D. Hu and G.-H. Hu, "A Relation between the Weighted Logarithmic Norm of a Matrix and the Lyapunov Equation," *Numerical Mathematics BIT*, Vol. 40, pp. 606–610, 2000.

[301] C. H. Hung and T. L. Markham, "The Moore-Penrose Inverse of a Partitioned Matrix," *Lin. Alg. Appl.*, Vol. 11, pp. 73–86, 1975.

[302] D. C. Hyland and D. S. Bernstein, "The Optimal Projection Equations for Fixed-Order Dynamic Compensation," *IEEE Trans. Autom. Contr.*, Vol. AC–29, pp. 1034–1037, 1984.

[303] D. C. Hyland and E. G. Collins, "Block Kronecker Products and Block Norm Matrices in Large-Scale Systems Analysis," *SIAM J. Matrix Anal. Appl.*, Vol. 10, pp. 18–29, 1989.

[304] N. H. Ibragimov, *Elementary Lie Group Analysis and Ordinary Differential Equations*. Chichester: Wiley, 1999.

[305] Y. Ikebe and T. Inagaki, "An Elementary Approach to the Functional Calculus for Matrices," *Amer. Math. Monthly*, Vol. 93, pp. 390–392, 1986.

[306] V. Ionescu, C. Oar, and M. Weiss, *Generalized Riccati Theory and Robust Control*. Chichester: John Wiley and Sons, 1999.

[307] A. Iserles, H. Z. Munthe-Kaas, S. P. Norsett, and A. Zanna, "Lie-Group Methods," *Acta Numerica*, Vol. 9, pp. 215–365, 2000.

[308] Y. Ito, S. Hattori, and H. Maeda, "On the Decomposition of a Matrix into the Sum of Stable Matrices," *Lin. Alg. Appl.*, Vol. 297, pp. 177–182, 1999.

[309] A. Jennings and J. J. McKeown, *Matrix Computation*, 2nd ed. New York: John Wiley and Sons, 1992.

[310] C. R. Johnson, "An Inequality for Matrices Whose Symmetric Part is Positive Definite," *Lin. Alg. Appl.*, Vol. 6, pp. 13–18, 1973.

[311] C. R. Johnson, "Closure Properties of Certain Positivity Classes of Matrices under Various Algebraic Operations," *Lin. Alg. Appl.*, Vol. 97, pp. 243–247, 1987.

[312] C. R. Johnson, M. Neumann, and M. J. Tsatsomeros, "Conditions for the Positivity of Determinants," *Lin. Multilin. Alg.*, Vol. 40, pp. 241–248, 1996.

[313] C. R. Johnson and P. Nylen, "Monotonicity Properties of Norms," *Lin. Alg. Appl.*, Vol. 148, pp. 43–58, 1991.

[314] C. R. Johnson, K. Okubo, and R. Beams, "Uniqueness of Matrix Square Roots," *Lin. Alg. Appl.*, Vol. 323, pp. 51–60, 2001.

[315] C. R. Johnson and R. Schreiner, "The Relationship Between AB and BA," *Amer. Math. Monthly*, Vol. 103, pp. 578–582, 1996.

[316] C. R. Johnson and H. Shapiro, "The Relative Gain Array $A \circ A^{-\mathrm{T}}$," *SIAM J. Alg. Disc. Meth.*, Vol. 7, pp. 627–644, 1986.

[317] M. Jolly, "On the Calculus of Complex Matrices," *Int. J. Contr.*, Vol. 61, pp. 749–755, 1995.

[318] A. Joseph, A. Melnikov, and R. Rentschler, Eds., *Studies in Memory of Issai Schur*. Cambridge: Birkhauser, 2002.

[319] E. I. Jury, *Inners and Stability of Dynamic Systems*, 2nd ed. Malabar: Krieger Publishing Co., 1982.

[320] J. B. Kagstrom, "Bounds and Perturbation Bounds for the Matrix Exponential," *BIT*, Vol. 17, pp. 39–57, 1977.

[321] T. Kailath, *Linear Systems*. Englewood Cliffs: Prentice-Hall, 1980.

[322] S. Karlin and F. Ost, "Some Monotonicity Properties of Schur Powers of Matrices and Related Inequalities," *Lin. Alg. Appl.*, Vol. 68, pp. 47–65, 1985.

[323] T. Kato, "Spectral Order and a Matrix Limit Theorem," *Lin. Multilin. Alg.*, Vol. 8, pp. 15–19, 1979.

[324] T. Kato, *Perturbation Theory for Linear Operators*. Berlin: Springer-Verlag, 1980.

[325] J. Y. Kazakia, "Orthogonal Transformation of a Trace Free Symmetric Matrix Into One With Zero diagonal Elements," *Int. J. Eng. Sci.*, Vol. 26, pp. 903–906, 1988.

[326] C. Kenney and A. J. Laub, "Controllability and Stability Radii for Companion Form Systems," *Math. Contr. Sig. Sys.*, Vol. 1, pp. 239–256, 1988.

[327] C. Kenney and A. J. Laub, "Rational Iteration Methods for the Matrix Sign Function," *SIAM J. Matrix Anal. Appl.*, Vol. 12, pp. 273–291, 1991.

[328] H. Kestelman, "Eigenvectors of a Cross-Diagonal Matrix," *Amer. Math. Monthly*, Vol. 93, p. 566, 1986.

[329] N. Keyfitz, *Introduction to the Mathematics of Population*. Reading: Addison-Wesley, 1968.

[330] C. G. Khatri and S. K. Mitra, "Hermitian and Nonnegative Definite Solutions of Linear Matrix Equations," *SIAM J. Appl. Math.*, Vol. 31, pp. 579–585, 1976.

[331] F. Kittaneh, "Inequalities for the Schatten p-norm III," *Comm. Math. Phys.*, Vol. 104, pp. 307–310, 1986.

[332] F. Kittaneh, "Inequalities for the Schatten $p-$Norm. IV," *Commun. Math. Phys.*, Vol. 106, pp. 581–585, 1986.

[333] F. Kittaneh, "On Zero-Trace Matrices," *Lin. Alg. Appl.*, Vol. 151, pp. 119–124, 1991.

[334] F. Kittaneh, "Singular Values of Companion Matrices and Bounds on Zeros of Polynomials," *SIAM J. Matrix Anal. Appl.*, Vol. 16, pp. 333–340, 1995.

[335] F. Kittaneh, "Norm Inequalities for Certain Operator Sums," *J. Funct. Anal.*, Vol. 143, pp. 337–348, 1997.

[336] F. Kittaneh, "Commutator Inequalities Associated with the Polar Decomposition," *Proc. Amer. Math. Soc.*, Vol. 130, pp. 1279–1283, 2001.

[337] F. Kittaneh, "Norm Inequalities for Sums of Positive Operators," *J. Operator Theory*, Vol. 48, pp. 95–103, 2002.

[338] R. H. Koning, H. Neudecker, and T. Wansbeek, "Block Kronecker Products and the vecb Operator," *Lin. Alg. Appl.*, Vol. 149, pp. 165–184, 1991.

[339] T. Koshy, *Fibonacci and Lucas Numbers with Applications.* New York: John Wiley & Sons, 2001.

[340] O. Krafft, "An Arithmetic-Harmonic-Mean Inequality for Nonnegative Definite Matrices," *Lin. Alg. Appl.*, Vol. 268, pp. 243–246, 1998.

[341] W. Kratz, *Quadratic Functionals in Variational Analysis and Control Theory.* New York: Wiley, 1995.

[342] E. Kreindler and A. Jameson, "Conditions for Nonnegativeness of Partitioned Matrices," *IEEE Trans. Autom. Contr.*, Vol. AC–17, pp. 147–148, 1972.

[343] V. Kucera, "On Nonnegative Definite Solutions to Matrix Quadratic Equations," *Automatica*, Vol. 8, pp. 413–423, 1972.

[344] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality.* Princeton: Princeton University Press, 1999.

[345] K. Kwakernaak and R. Sivan, *Linear Optimal Control Systems.* New York: John Wiley and Sons, 1972.

[346] M. Kwapisz, "The Power of a Matrix," *SIAM Rev.*, Vol. 40, pp. 703–705, 1998.

[347] K. R. Laberteaux, "Hermitian Matrices," *Amer. Math. Monthly*, Vol. 104, p. 277, 1997.

[348] T. J. Laffey, "Products of Skew-Symmetric Matrices," *Lin. Alg. Appl.*, Vol. 68, pp. 249–251, 1985.

[349] S. Lakshminarayanan, S. L. Shah, and K. Nandakumar, "Cramer's Rule for Non-Square Matrices," *Amer. Math. Monthly*, Vol. 106, p. 865, 1999.

[350] P. Lancaster, *Lambda-matrices and Vibrating Systems*. Oxford: Pergamon Press, 1966, reprinted by Dover Publications, Mineola, 2002.

[351] P. Lancaster and L. Rodman, "Solutions of the Continuous and Discrete Time Algebraic Riccati Equations: A Review," in *The Riccati Equation*, S. Bittanti, J. C. Willems, and A. Laub, Eds. New York: Springer, 1991, pp. 11–51.

[352] P. Lancaster and L. Rodman, *Algebraic Riccati Equations*. Oxford: Clarendon Press, 1995.

[353] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. Orlando: Academic Press, 1985.

[354] A. J. Laub and K. Meyer, "Canonical Forms for Symplectic and Hamiltonian Matrices," *Celestial Mechanics*, Vol. 9, pp. 213–238, 1974.

[355] C. L. Lawson, *Solving Least Squares Problems*. Englewood Cliffs: Prentice-Hall, 1974, reprinted by SIAM, Philadelphia,1995.

[356] P. D. Lax, *Linear Algebra*. New York: John Wiley and Sons, 1997.

[357] S. R. Lay, *Convex Sets and Their Applications*. New York: John Wiley and Sons, 1982.

[358] K. J. LeCouteur, "Representation of the Function $\mathrm{Tr}(\exp(A\text{-}\lambda B))$ as a Laplace Transform with Positive Weight and Some Matrix Inequalities," *J. Phys. A*, Vol. 13, pp. 3147–3159, 1980.

[359] A. Lee, "Centrohermitian and Skew-Centrohermitian Matrices," *Lin. Alg. Appl.*, Vol. 29, pp. 205–210, 1980.

[360] J. M. Lee and D. A. Weinberg, "A Note on Canonical Forms for Matrix Congruence," *Lin. Alg. Appl.*, Vol. 249, pp. 207–215, 1996.

[361] S. H. Lehnigk, *Stability Theorems for Linear Motions*. Englewood Cliffs: Prentice-Hall, 1966.

[362] E. Leonard, "The Matrix Exponential," *SIAM Rev.*, Vol. 38, pp. 507–512, 1996.

[363] G. Letac, "A Matrix and Its Matrix of Reciprocals Both Positive Semidefinite," *Amer. Math. Monthly*, Vol. 82, pp. 80–81, 1975.

[364] J. S. Lew, "The Cayley Hamilton Theorem in n Dimensions," *Z. Angew. Math. Phys.*, Vol. 17, pp. 650–653, 1966.

[365] D. C. Lewis, "A Qualitative Analysis of S-Systems: Hopf Bifurcations," in *Canonical Nonlinear Modeling*, E. O. Voit, Ed. New York: Van Nostrand Reinhold, 1991, pp. 304–344.

[366] A.-L. Li and C.-K. Li, "Isometries for the Vector (p, q) Norm and the Induced (p, q) Norm," *Lin. Multilin. Alg.*, Vol. 21, pp. 315–332, 1995.

[367] C.-K. Li and R. Mathias, "The Determinant of the Sum of Two Matrices," *Bull. Austral. Math. Soc.*, Vol. 52, pp. 425–429, 1995.

[368] C.-K. Li and R. Mathias, "The Lidskii-Mirsky-Wielandt Theorem–Additive and Multiplicative Versions," *Numer. Math.*, Vol. 81, pp. 377–413, 1999.

[369] C.-K. Li and R. Mathias, "Extremal Characterizations of the Schur Complement and Resulting Inequalities," *SIAM Rev.*, Vol. 42, pp. 233–246, 2000.

[370] C.-K. Li and R. Mathias, "Inequalities on Singular Values of Block Triangular Matrices," *SIAM J. Matrix Anal. Appl.*, Vol. 24, pp. 126–131, 2002.

[371] C.-K. Li and H. Schneider, "Orthogonality of Matrices," *Lin. Alg. Appl.*, Vol. 347, pp. 115–122, 2002.

[372] E. H. Lieb, "Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture," *Advances Math.*, Vol. 11, pp. 267–288, 1973.

[373] E. H. Lieb and M. B. Ruskai, "Some Operator Inequalities of the Schwarz Type," *Adv. Math.*, Vol. 12, pp. 269–273, 1974.

[374] E. H. Lieb and W. E. Thirring, "Inequalities for the Moments of the Eigenvalues of the Schrodinger Hamiltonian and Their Relation to Sobolev Inequalities," in *Studies in Mathematical Physics*, E. Lieb, B. Simon, and A. Wightman, Eds. Princeton: Princeton University Press, 1976, pp. 269–303.

[375] T.-P. Lin, "The Power Mean and the Logarithmic Mean," *Amer. Math. Monthly*, Vol. 81, pp. 879–883, 1974.

[376] R.-W. Liu and R. J. Leake, "Exhaustive Equivalence Classes of Optimal Systems with Separable Controls," *SIAM Rev.*, Vol. 4, pp. 678–685, 1966.

[377] S. Liu, "Several Inequalities Involving Khatri-Rao Products of Positive Semidefinite Matrices," *Lin. Alg. Appl.*, Vol. 354, pp. 175–186, 2002.

[378] S. Liu and H. Neudecker, "Several Matrix Kantorovich-Type Inequalities," *Math. Anal. Appl.*, Vol. 197, pp. 23–26, 1996.

[379] E. Liz, "A Note on the Matrix Exponential," *SIAM Rev.*, Vol. 40, pp. 700–702, 1998.

[380] R. Loewy, "An Inertia Theorem for Lyapunov's Equation and the Dimension of a Controllability Space," *Lin. Alg. Appl.*, Vol. 260, pp. 1–7, 1997.

[381] M. Loss and M. B. Ruskai, Eds., *Inequalities: Selecta of Elliott H. Lieb.* New York: Springer, 2002.

[382] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley and Sons, 1969.

[383] H. Lutkepohl, *Handbook of Matrices.* Chichester: John Wiley and Sons, 1996.

[384] C. C. MacDuffee, *The Theory of Matrices.* New York: Chelsea, 1956.

[385] A. G. J. Macfarlane and N. Karcanias, "Poles and Zeros of Linear Multivariable Systems: A Survey of the Algebraic, Geometric, and Complex-Variable Theory," *Int. J. Contr.*, Vol. 24, pp. 33–74, 1976.

[386] D. S. Mackey, N. Mackey, and F. Tisseur, "Structured Tools for Structured Matrices," *Elec. J. Lin. Alg.*, Vol. 10, pp. 106–145, 2003.

[387] J. R. Magnus, *Linear Structures.* London: Griffin, 1988.

[388] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: John Wiley and Sons, 1988.

[389] K. N. Majindar, "On Simultaneous Hermitian Congruence Transformations of Matrices," *Amer. Math. Monthly*, Vol. 70, pp. 842–844, 1963.

[390] A. N. Malyshev and M. Sadkane, "On the Stability of Large Matrices," *J. Computational Appl. Math.*, Vol. 102, pp. 303–313, 1999.

[391] L. E. Mansfield, *Linear Algebra With Geometric Application.* New York: Marcel-Dekker, 1976.

[392] M. Marcus, "An Eigenvalue Inequality for the Product of Normal Matrices," *Amer. Math. Monthly*, Vol. 63, pp. 173–174, 1956.

[393] M. Marcus, "Two Determinant Condensation Formulas," *Lin. Multilinear Alg.*, Vol. 22, pp. 95–102, 1987.

[394] M. Marcus and N. A. Khan, "A Note on the Hadamard Product," *Canad. Math. J.*, Vol. 2, pp. 81–83, 1959.

[395] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities.* Boston: Prindle, Weber, and Schmidt, 1964, reprinted by Dover, New York, 1992.

[396] T. L. Markham, "An Application of Theorems of Schur and Albert," *Proc. Amer. Math. Soc.*, Vol. 59, pp. 205–210, 1976.

[397] T. L. Markham, "Oppenheim's Inequality for Positive Definite Matrices," *Amer. Math. Monthly*, Vol. 93, pp. 642–644, 1986.

[398] G. Marsaglia and G. P. H. Styan, "Equalities and Inequalities for Ranks of Matrices," *Lin. Multilin. Alg.*, Vol. 2, pp. 269–292, 1974.

[399] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry.* Springer, 1994.

[400] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press, 1979.

[401] A. W. Marshall and I. Olkin, "Matrix Versions of the Cauchy and Kantorovich Inequalities," *Aequationes Math.*, Vol. 40, pp. 89–93, 1990.

[402] K. Martensson, "On the Matrix Riccati Equation," *Information Sciences*, Vol. 3, pp. 17–49, 1971.

[403] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Argument.* Singapore: World Scientific, 1997.

[404] R. Mathias, "Evaluating the Frechet Derivative of the Matrix Exponential," *Numer. Math.*, Vol. 63, pp. 213–226, 1992.

[405] R. Mathias, "An Arithmetic-Geometric-Harmonic Mean Inequality Involving Hadamard Products," *Lin. Alg. Appl.*, Vol. 184, pp. 71–78, 1993.

[406] R. Mathias, "A Chain Rule for Matrix Functions and Applications," *SIAM J. Matrix. Anal. Appl.*, Vol. 17, pp. 610–620, 1996.

[407] J. P. McCloskey, "Characterizations of r-Potent Matrices," *Math. Proc. Camb. Phil. Soc.*, Vol. 96, pp. 213–222, 1984.

[408] A. R. Meenakshi and C. Rajian, "On a Product of Positive Semidefinite Matrices," *Lin. Alg. Appl.*, Vol. 295, pp. 3–6, 1999.

[409] Y. A. Melnikov, "Influence Functions and Matrices," *Marcel Dekker*, 1998.

[410] J. K. Merikoski, H. Sarria, and P. Tarazaga, "Bounds for Singular Values Using Traces," *Lin. Alg. Appl.*, Vol. 210, pp. 227–254, 1994.

[411] R. Merris, "Inequalities Involving the Inverses of Positive Definite Matrices," *Proc. Edinburgh Math. Soc.*, Vol. 22, pp. 11–15, 1979.

[412] R. Merris, *Multilinear Algebra.* Amsterdam: Gordon and Breach Science Publishers, 1997.

[413] R. Merris and S. Pierce, "Monotonicity of Positive Semidefinite Hermitian Matrices," *Proc. Amer. Math. Soc.*, Vol. 31, pp. 437–440, 1972.

[414] C. D. Meyer, "The Moore-Penrose Inverse of a Bordered Matrix," *Lin. Alg. Appl.*, Vol. 5, pp. 375–382, 1972.

[415] C. D. Meyer, "Generalized Inverses and Ranks of Block Matrices," *SIAM J. Appl. Math*, Vol. 25, pp. 597–602, 1973.

[416] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra.* Philadelphia: SIAM, 2000.

[417] J.-M. Miao, "General Expressions for the Moore-Penrose Inverse of a 2×2 Block Matrix," *Lin. Alg. Appl.*, Vol. 151, pp. 1–15, 1991.

[418] L. Mihalyffy, "An Alternative Representation of the Generalized Inverse of Partitioned Matrices," *Lin. Alg. Appl.*, Vol. 4, pp. 95–100, 1971.

[419] K. S. Miller, *Some Eclectic Matrix Theory.* Malabar: Krieger, 1987.

[420] G. A. Milliken and F. Akdeniz, "A Theorem on the Difference of the Generalized Inverses of Two Nonnegative Marices," *Comm. Statist. Theory Methods*, Vol. 6, pp. 73–79, 1977.

[421] N. Minamide, "An Extension of the Matrix Inversion Lemma," *SIAM J. Alg. Disc. Meth.*, Vol. 6, pp. 371–377, 1985.

[422] H. Miranda and R. C. Thompson, "A Trace Inequality With a Subtracted Term," *Lin. Alg. Appl.*, Vol. 185, pp. 165–172, 1993.

[423] L. Mirsky, *An Introduction to Linear Algebra.* Oxford: Clarendon Press, 1972, reprinted by Dover, Mineola, 1990.

[424] D. S. Mitrinovic, J. E. Pecaric, and A. M. Fink, *Classical and New Inequalities in Analysis.* Dordrecht: Kluwer, 1993.

[425] B. Mityagin, "An Inequality in Linear Algebra," *SIAM Rev.*, Vol. 33, pp. 125–127, 1991.

[426] C. Moler and C. F. Van Loan, "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," *SIAM Rev.*, Vol. 45, pp. 3–49, 2000.

[427] B. Mond and J. E. Pecaric, "Inequalities for the Hadamard Product of Matrices," *SIAM J. Matrix Anal. Appl.*, Vol. 19, pp. 66–70, 1998.

[428] V. V. Monov, "On the Spectrum of Convex Sets of Matrices," *IEEE Trans. Autom. Contr.*, Vol. 44, pp. 1009–1012, 1992.

[429] T. Mori, "Comments on "A Matrix Inequality Associated with Bounds on Solutions of Algebraic Riccati and Lyapunov Equation"," *IEEE Trans. Autom. Contr.*, Vol. 33, p. 1088, 1988.

[430] T. Muir, *The Theory of Determinants in the Historical Order of Development.* New York: Dover, 1966.

[431] W. W. Muir, "Inequalities Concerning the Inverses of Positive Definite Matrices," *Proc. Edinburgh Math. Soc.*, Vol. 19, pp. 109–113, 1974–75.

[432] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation.* Boca Raton: CRC Press, 1994.

[433] I. Najfeld and T. F. Havel, "Derivatives of the Matrix Exponential and Their Computation," *Adv. Appl. Math.*, Vol. 16, pp. 321–375, 1995.

[434] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science.* New York: Springer, 1986.

[435] C. N. Nett and W. M. Haddad, "A System-Theoretic Appropriate Realization of the Empty Matrix Concept," *IEEE Trans. Autom. Contr.*, Vol. 38, pp. 771–775, 1993.

[436] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models.* Baltimore: Johns Hopkins University Press, 1981, reprinted by Dover, Mineola, 1994.

[437] R. W. Newcomb, "On the Simultaneous Diagonalization of Two Semidefinite Matrices," *Quart. Appl. Math.*, Vol. 19, pp. 144–146, 1961.

[438] M. Newman, W. So, and R. C. Thompson, "Convergence Domains for the Campbell-Baker-Hausdorff Formula," *Lin. Multilin. Alg.*, Vol. 24, pp. 301–310, 1989.

[439] K. Nishio, "The Structure of a Real Linear Combination of Two Projections," *Lin. Alg. Appl.*, Vol. 66, pp. 169–176, 1985.

[440] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 3rd ed. Englewood Cliffs: Prentice-Hall, 1988.

[441] J. Nunemacher, "Which Real Matrices Have Real Logarithms?" *Math. Mag.*, Vol. 62, pp. 132–135, 1989.

[442] H. Ogawa, "An Operator Pseudo-Inversion Lemma," *SIAM J. Appl. Math.*, Vol. 48, pp. 1527–1531, 1988.

[443] I. Olkin, "An Inequality for a Sum of Forms," *Lin. Alg. Appl.*, Vol. 52–53, pp. 529–532, 1983.

[444] J. M. Ortega, *Matrix Theory, A Second Course.* New York: Plenum Press, 1987.

[445] S. L. Osburn and D. S. Bernstein, "An Exact Treatment of the Achievable Closed-Loop H_2 Performance of Sampled-Data Controllers: From Continuous-Time to Open-Loop," *Automatica*, Vol. 31, pp. 617–620, 1995.

[446] A. Ostrowski and H. Schneider, "Some Theorems on the Inertia of General Matrices," *J. Math. Anal. Appl.*, Vol. 4, pp. 72–84, 1962.

[447] D. A. Overdijk, "Skew-symmetric Matrices in Classical Mechanics," Eindhoven University, Memorandum COSOR 89–23, 1989.

[448] C. V. Pao, "Logarithmic Derivatives of a Square Matrix," *Lin. Alg. Appl.*, Vol. 6, pp. 159–164, 1973.

[449] J. G. Papastravridis, *Tensor Calculus and Analytical Dynamics.* Boca Raton: CRC Press, 1998.

[450] F. C. Park, "Computational Aspects of the Product-of-Exponentials Formula for Robot Kinematics," *IEEE Trans. Autom. Contr.*, Vol. 39, pp. 643–647, 1994.

[451] P. Park, "On theTrace Bound of a Matrix Product," *IEEE Trans. Autom. Contr.*, Vol. 41, pp. 1799–1802, 1996.

[452] P. C. Parks, "A New Proof of the Routh-Hurwitz Stability Criterion Using the Second Method of Liapunov," *Proc. Camb. Phil. Soc.*, Vol. 58, pp. 694–702, 1962.

[453] R. V. Patel, "On Blocking Zeros in Linear Multivariable Systems," *IEEE Trans. Autom. Contr.*, Vol. AC–31, pp. 239–241, 1986.

[454] R. V. Patel and M. Toda, "Trace Inequalities Involving Hermitian Matrices," *Lin. Alg. Appl.*, Vol. 23, pp. 13–20, 1979.

[455] M. C. Pease III, *Methods of Matrix Algebra.* Academic Press, New York, 1965.

[456] S. Perlis, *Theory of Matrices.* Reading: Addison-Wesley, 1952, reprinted by Dover, New York, 1991.

[457] I. R. Petersen and C. V. Hollot, "A Riccati Equation Approach to the Stabilization of Uncertain Systems," *Automatica*, Vol. 22, pp. 397–411, 1986.

[458] L. A. Pipes, "Applications of Laplace Transforms of Matrix Functions," *J. Franklin Inst.*, Vol. 285, pp. 436–451, 1968.

[459] T. Politi, "A Formula for the Exponential of a Real Skew-Symmetric Matrix of Order 4," *Numerical Mathematics BIT*, Vol. 41, pp. 842–845, 2001.

[460] D. S. G. Pollock, "Tensor Products and Matrix Differential Calculus," *Lin. Alg. Appl.*, Vol. 67, pp. 169–193, 1985.

[461] B. Poonen, "A Unique $(2k+1)$-th Root of a Matrix," *Amer. Math. Monthly*, Vol. 98, p. 763, 1991.

[462] B. Poonen, "Positive Deformations of the Cauchy Matrix," *Amer. Math. Monthly*, Vol. 102, pp. 842–843, 1995.

[463] V. M. Popov, *Hyperstability of Control Systems.* Berlin: Springer-Verlag, 1973.

[464] G. J. Porter, "Linear Algebra and Affine Planar Transformations," *College Math. J.*, Vol. 24, pp. 47–51, 1993.

[465] B. H. Pourciau, "Modern Multiplier Rules," *American Mathematical Monthly*, Vol. 87, pp. 433–452, 1980.

[466] V. V. Prasolov, *Problems and Theorems in Linear Algebra.* Providence: American Mathematical Society, 1994.

[467] J. S. Przemieniecki, *Theory of Matrix Structural Analysis.* New York: McGraw-Hill, 1968.

[468] P. J. Psarrakos, "On the mth Roots of a Complex Matrix," *Elec. J. Lin. Alg.*, Vol. 9, pp. 32–41, 2002.

[469] N. J. Pullman, *Matrix Theory and Its Applications: Selected Topics.* New York: Marcel Dekker, 1976.

[470] R. X. Qian and C. L. DeMarco, "An Approach to Robust Stability of Matrix Polytopes Through Copositive Homogeneous Polynomials," *IEEE Trans. Autom. Contr.*, Vol. 37, pp. 848–852, 1992.

[471] L. Qiu, B. Bernhardsson, A. Rantzer, E. J. Davison, P. M. Young, and J. C. Doyle, "A Formula for Computation of the Real Stability Radius," *Automatica*, Vol. 31, pp. 879–890, 1995.

[472] H. Radjavi, "Decomposition of Matrices into Simple Involutions," *Lin. Alg. Appl.*, Vol. 12, pp. 247–255, 1975.

[473] H. Radjavi and P. Rosenthal, *Simultaneous Triangularization.* New York: Springer, 2000.

[474] H. Radjavi and J. P. Williams, "Products of Self-Adjoint Operators," *Michigan Math. J.*, Vol. 16, pp. 177–185, 1969.

[475] A. C. M. Ran and R. Vreugdenhil, "Existence and Comparison Theorems for Algebraic Riccati Equations for Continuous- and Discrete-Time Systems," *Lin. Alg. Appl.*, Vol. 99, pp. 63–83, 1988.

[476] A. Rantzer, "On the Kalman-Yakubovich-Popov Lemma," *Sys. Contr. Lett.*, Vol. 28, pp. 7–10, 1996.

[477] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications.* New York: John Wiley and Sons, 1971.

[478] J. V. Rao, "Some More Representations for the Generalized Inverse of a Partitioned Matrix," *SIAM J. Appl. Math.*, Vol. 24, pp. 272–276, 1973.

[479] P. A. Regalia and S. K. Mitra, "Kronecker Products, Unitary Matrices and Signal Processing Applications," *SIAM Rev.*, Vol. 31, pp. 586–613, 1989.

[480] T. J. Richardson and R. H. Kwong, "On Positive Definite Solutions to the Algebraic Riccati Equation," *Sys. Contr. Lett.*, Vol. 7, pp. 99–104, 1986.

[481] A. N. Richmond, "Expansions for the Exponential of a Sum of Matrices," in *Applications of Matrix Theory*, M. J. C. Gover and S. Barnett, Eds. Oxford: Oxford University Press, 1989, pp. 283–289.

[482] J. R. Ringrose, *Compact Non-Self-Adjoint Operators.* Van Nostrand Reinhold, 1971.

[483] R. S. Rivlin, "Further Remarks on the Stress Deformation Relations for Isotropic Materials," *J. Rational Mech. Anal.*, Vol. 4, pp. 681–702, 1955.

[484] J. W. Robbin, *Matrix Algebra Using MINImal MATlab.* Wellesley: A. K. Peters, 1995.

[485] R. T. Rockafellar, *Convex Analysis.* Princeton: Princeton University Press, 1990.

[486] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis.* Berlin: Springer, 1998.

[487] L. Rodman, "Products of Symmetric and Skew Symmetric Matrices," *Lin. Multilin. Alg.*, Vol. 43, pp. 19–34, 1997.

[488] G. S. Rogers, *Matrix Derivatives.* New York: Marcel Dekker, 1980.

[489] C. A. Rohde, "Generalized Inverses of Partitioned Matrices," *SIAM J. Appl. Math.*, Vol. 13, pp. 1033–1035, 1965.

[490] J. Rohn, "Computing the Norm $||A||_{\infty,1}$ is NP-Hard," *Lin. Multilin. Alg.*, Vol. 47, pp. 195–204, 2000.

[491] O. Rojo, "Further Bounds for the Smallest Singular Value and the Spectral Condition Number," *Computers Math. Appl.*, Vol. 38, pp. 215–228, 1999.

[492] K. H. Rosen, Ed., *Handbook of Discrete and Combinatorial Mathematics.* Boca Raton: CRC, 2000.

[493] M. Rosenfeld, "A Sufficient Condition for Nilpotence," *Amer. Math. Monthly*, Vol. 103, pp. 907–909, 1996.

[494] W. J. Rugh, *Linear System Theory*, 2nd ed. Upper Saddle River: Prentice Hall, 1996.

[495] M. K. Sain and C. B. Schrader, "The Role of Zeros in the Performance of Multiinput, Multioutput Feedback Systems," *IEEE Trans. Educ.*, Vol. 33, pp. 244–257, 1990.

[496] D. H. Sattinger and O. L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics.* New York: Springer, 1986.

[497] C. W. Scherer, "The Algebraic Riccati Equation and Inequality for Systems with Uncontrollable Modes on the Imaginary Axis," *SIAM. J. Matrix Anal. Appl.*, pp. 1308–1327, 1995.

[498] P. Scherk, "On the Decomposition of Orthogonalities into Symmetries," *Proc. Amer. Math. Soc.*, Vol. 1, pp. 481–491, 1950.

[499] C. Schmoeger, "On the Operator Equation $e^a = e^b$," *Lin. Alg. Appl.*, Vol. 359, pp. 169–179, 2003.

[500] H. Schneider, "Olga Taussky-Todd's Influence on Matrix Theory and Matrix Theorists," *Lin. Multilin. Alg.*, Vol. 5, pp. 197–224, 1977.

[501] C. B. Schrader and M. K. Sain, "Research on System Zeros: A Survey," *Int. J. Contr.*, Vol. 50, pp. 1407–1433, 1989.

[502] A. J. Schwenk, "Tight Bounds on the Spectral Radius of Asymmetric Nonnegative Matrices," *Lin. Alg. Appl.*, Vol. 75, pp. 257–265, 1986.

[503] S. R. Searle, *Matrix Algebra Useful for Statistics.* New York: John Wiley and Sons, 1982.

[504] P. Sebastian, "On the Derivatives of Matrix Powers," *SIAM J. Matrix Anal. Appl.*, Vol. 17, pp. 640–648, 1996.

[505] D. Serre, *Matrices: Theory and Applications.* New York: Springer, 2002.

[506] C. Shafroth, "A Generalization of the Formula for Computing the Inverse of a Matrix," *Amer. Math. Monthly*, Vol. 88, pp. 614–616, 1981.

[507] H. Shapiro, "Notes from Math 223: Olga Taussky Todd's Matrix Theory Course, 1976–1977," *Mathematical Intelligencer*, Vol. 19, no. 1, pp. 21–27, 1997.

[508] R. Shaw and F. I. Yeadon, "On $(a \times b) \times c$," *Amer. Math. Monthly*, Vol. 96, pp. 623–629, 1989.

[509] G. E. Shilov, *Linear Algebra.* Englewood Cliffs: Prentice-Hall, 1971. Reprinted by Dover, New York, 1977.

[510] D. D. Siljak, *Large-Scale Dynamic Systems: Stability and Structure.* New York: North-Holland, 1978.

[511] R. A. Smith, "Matrix Calculations for Lyapunov Quadratic Forms," *J. Diff. Eqns.*, Vol. 2, pp. 208–217, 1966.

[512] J. Snyders and M. Zakai, "On Nonnegative Solutions of the Equation $AD + DA' = C$," *SIAM J. Appl. Math.*, Vol. 18, pp. 704–714, 1970.

[513] W. So, "Equality Cases in Matrix Exponential Inequalities," *SIAM J. Matrix Anal. Appl.*, Vol. 13, pp. 1154–1158, 1992.

[514] W. So, "The High Road to an Exponential Formula," *Lin. Alg. Appl.*, Vol. TBD, pp. TBD–TBD, 2003.

[515] W. So and R. C. Thompson, "Product of Exponentials of Hermitian and Complex Symmetric Matrices," *Lin. Multilin. Alg.*, Vol. 29, pp. 225–233, 1991.

[516] W. So and R. C. Thompson, "Singular Values of Matrix Exponentials," *Lin. Multilin. Alg.*, Vol. 47, pp. 249–258, 2000.

[517] A. R. Sourour, "A Factorization Theorem for Matrices," *Lin. Multilin. Alg.*, Vol. 19, pp. 141–147, 1986.

[518] W.-H. Steeb, *Matrix Calculus and Kronecker Product with Applications and C++ Programs.* Singapore: World Scientific, 2001.

[519] W.-H. Steeb and F. Wilhelm, "Exponential Functions of Kronecker Products and Trace Calculation," *Lin. Multilin. Alg.*, Vol. 9, pp. 345–346, 1981.

[520] C. Stepniak, "Ordering of Nonnegative Definite Matrices with Application to Comparison of Linear Models," *Lin. Alg. Appl.*, Vol. 70, pp. 67–71, 1985.

[521] G. W. Stewart, *Introduction to Matrix Computations.* New York: Academic Press, 1973.

[522] G. W. Stewart, *Matrix Algorithms Volume I: Basic Decompositions.* Philadelphia: SIAM, 1998.

[523] G. W. Stewart, "On the Adjugate Matrix," *Lin. Alg. Appl.*, Vol. 283, pp. 151–164, 1998.

[524] G. W. Stewart, *Matrix Algorithms Volume II: Eigensystems.* Philadelphia: SIAM, 2001.

[525] G. W. Stewart and J. Sun, *Matrix Perturbation Theory.* Boston: Academic Press, 1990.

[526] E. U. Stickel, "Fast Computation of Matrix Exponential and Logarithm," *Analysis*, Vol. 5, pp. 163–173, 1985.

[527] E. U. Stickel, "An Algorithm for Fast High Precision Computation of Matrix Exponential and Logarithm," *Analysis*, Vol. 10, pp. 85–95, 1990.

[528] J. Stoer, "On the Characterization of Least Upper Bound Norms in Matrix Space," *Numer. Math*, Vol. 6, pp. 302–314, 1964.

[529] M. G. Stone, "A Mnemonic for Areas of Polygons," *Amer. Math. Monthly*, Vol. 93, pp. 479–480, 1986.

[530] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. San Diego: Harcourt, Brace, Jovanovich, 1988.

[531] G. Strang, "The Fundamental Theorem of Linear Algebra," *Amer. Math. Monthly*, Vol. 100, pp. 848–855, 1993.

[532] T. Strom, "On Logarithmic Norms," *SIAM J. Numer. Anal.*, Vol. 12, pp. 741–753, 1975.

[533] K. N. Swamy, "On Sylvester's Criterion for Positive-Semidefinite Matrices," *IEEE Trans. Autom. Contr.*, Vol. AC–18, p. 306, 1973.

[534] O. Taussky, "Positive-Definite Matrices and Their Role in the Study of the Characteristic Roots of General Matrices," *Adv. Math.*, Vol. 2, pp. 175–186, 1968.

[535] O. Taussky, "The Role of Symmetric Matrices in the Study of General Matrices," *Lin. Alg. Appl.*, Vol. 5, pp. 147–154, 1972.

[536] O. Taussky, "How I Became a Torchbearer for Matrix Theory," *Amer. Math. Monthly*, Vol. 95, pp. 801–812, 1988.

[537] O. Taussky and J. Todd, "Another Look at a Matrix of Mark Kac," *Lin. Alg. Appl.*, Vol. 150, pp. 341–360, 1991.

[538] O. Taussky and H. Zassenhaus, "On the Similarity Transformation Between a Matrix and Its Transpose," *Pacific J. Math.*, Vol. 9, pp. 893–896, 1959.

[539] W. Tempelman, "The Linear Algebra of Cross Product Operations," *J. Astron. Sciences*, Vol. 36, pp. 447–461, 1988.

[540] R. E. Terrell, "Solution to 'Exponentials of Certain Hilbert Space Operators'," *SIAM Rev.*, Vol. 34, pp. 498–500, 1992.

[541] R. E. Terrell, "Matrix Exponentials," *SIAM Rev.*, Vol. 38, pp. 313–314, 1996.

[542] R. C. Thompson, "On Matrix Commutators," *J. Washington Academy of Sciences*, Vol. 48, pp. 306–307, 1958.

[543] R. C. Thompson, "A Determinantal Inequality for Positive Definite Matrices," *Canad. Math. Bull.*, Vol. 4, pp. 57–62, 1961.

[544] R. C. Thompson, "Some Matrix Factorization Theorems," *Pacific J. Math.*, Vol. 33, pp. 763–810, 1970.

[545] R. C. Thompson, "A Matrix Inequality," *Comment. Math. Univ. Carolinae*, Vol. 17, pp. 393–397, 1976.

[546] R. C. Thompson, "Matrix Type Metric Inequalities," *Lin. Multilin. Alg.*, Vol. 5, pp. 303–319, 1978.

[547] R. C. Thompson, "Proof of a Conjectured Exponential Formula," *Lin. Multilin. Alg.*, Vol. 19, pp. 187–197, 1986.

[548] R. C. Thompson, "Pencils of Complex and Real Symmetric and Skew Matrices," *Lin. Alg. Appl.*, Vol. 147, pp. 323–371, 1991.

[549] R. C. Thompson, "High, Low, and Quantitative Roads in Linear Algebra," *Lin. Alg. Appl.*, Vol. 162–164, pp. 23–64, 1992.

[550] Y. Tian, "The moore-penrose Inverse of $m \times n$ Block Matrices and Their Applications," *Lin. Alg. Appl.*, Vol. 283, pp. 35–60, 1998.

[551] Y. Tian and G. P. H. Styan, "How to Establish Universal Block-Matrix Factorizations," *Electr. J. Lin. Alg.*, Vol. 8, pp. 115–127, 2001.

[552] A. Tonge, "Equivalence Constants for Matrix Norms: A Problem of Goldberg," *Lin. Alg. Appl.*, Vol. 306, pp. 1–13, 2000.

[553] G. E. Trapp, "Hermitian Semidefinite Matrix Means and Related Matrix Inequalities-An Introduction," *Lin. Multilinear Alg.*, Vol. 16, pp. 113–123, 1984.

[554] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. Philadelphia: SIAM, 1997.

[555] G. Trenkler, "A Trace Inequality," *Amer. Math. Monthly*, Vol. 102, pp. 362–363, 1995.

[556] H. L. Trentelman, A. A. Stoorvogel, and M. L. J. Hautus, *Control Theory for Linear Systems*. New York: Springer, 2001.

[557] P. Treuenfels, "The Matrix Equation $X^2 - 2AX + B = 0$," *Amer. Math. Monthly*, Vol. 66, pp. 145–146, 1959.

[558] S. H. Tung, "On Lower and Upper Bounds of the Difference Between the Arithmetic and the Geometric Mean," *Math. Comput.*, Vol. 29, pp. 834–836, 1975.

[559] D. A. Turkington, *Matrix Calculus and Zero-One Matrices*. Cambridge: Cambridge University Press, 2002.

[560] H. W. Turnbull, *The Theory of Determinants, Matrices and Invariants*. London: Blackie, 1950.

[561] F. Tyan and D. S. Bernstein, "Global Stabilization of Systems Containing a Double Integrator Using a Saturated Linaer Controller," *Int. J. Robust Nonlinear Contr.*, Vol. 9, pp. 1143–1156, 1999.

[562] F. E. Udwadia and R. E. Kalaba, *Analytical Dynamics: A New Approach*. Cambridge: Cambridge University Press, 1996.

[563] F. Uhlig, "A Recurring Theorem About Pairs of Quadratic Forms and Extensions: A Survey," *Lin. Alg. Appl.*, Vol. 25, pp. 219–237, 1979.

[564] F. Uhlig, "Constructive Ways for Generating (Generalized) Real Orthogonal Matrices as Products of (Generalized) Symmetries," *Lin. Alg. Appl.*, Vol. 332–334, pp. 459–467, 2001.

[565] F. A. Valentine, *Convex Sets*. New York: McGraw-Hill, 1964.

[566] M. Van Barel, V. Ptak, and Z. Vavrin, "Bezout and Hankel Matrices Associated with Row Reduced Matrix Polynomials, Barnett-Type Formulas," *Lin. Alg. Appl.*, Vol. 332–334, pp. 583–606, 2001.

[567] C. F. Van Loan, "Computing Integrals Involving the Matrix Exponential," *IEEE Trans. Autom. Contr.*, Vol. AC–23, pp. 395–404, 1978.

[568] C. F. Van Loan, "How Near is a Stable Matrix to an Unstable Matrix," *Contemporary Math.*, Vol. 47, pp. 465–478, 1985.

[569] C. F. Van Loan, *Computational Frameworks for the Fast Fourier Transform.* Philadelphia: SIAM, 1992.

[570] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications.* Dordrecht: Kluwer, 1996.

[571] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations.* New York: Springer-Verlag, 1984.

[572] A. I. G. Vardulakis, *Linear Multivariable Control: Algebraic Analysis and Synthesis Methods.* Chichester: John Wiley and Sons, 1991.

[573] R. S. Varga, *Matrix Iterative Analysis.* Englewood Cliffs: Prentice-Hall, 1962.

[574] R. Vein and P. Dale, *Determinants and Their Applications in Mathematical Physics.* New York: Springer, 1999.

[575] W. J. Vetter, "Matrix calculus Operations and Taylor Expansions," *SIAM Rev.*, Vol. 15, pp. 352–369, 1973.

[576] M. Vidyasagar, "On Matrix Measures and Convex Liapunov Functions," *J. Math. Anal. Appl.*, Vol. 62, pp. 90–103, 1978.

[577] G. Visick, "A Quantitative Version of the Observation That the Hadamard Product is a Principal Submatrix of the Kronecker Product," *Lin. Alg. Appl.*, Vol. 304, pp. 45–68, 2000.

[578] B.-Y. Wang and M.-P. Gong, "Some Eigenvalue Inequalities for Positive Semidefinite Matrix Power Products," *Lin. Alg. Appl.*, Vol. 184, pp. 249–260, 1993.

[579] B.-Y. Wang, B.-Y. Xi, and F. Zhang, "Some Inequalities for Sum and Product of Positive Semidefinite Matrices," *Lin. Alg. Appl.*, Vol. 293, pp. 39–49, 1999.

[580] B.-Y. Wang and F. Zhang, "A Trace Inequality for Unitary Matrices," *Amer. Math. Monthly*, Vol. 101, pp. 453–455, 1994.

[581] B.-Y. Wang and F. Zhang, "Trace and Eigenvalue Inequalities for Ordinary and Hadamard Products of Positive Semidefinite Hermitian Matrices," *SIAM J. Matrix Anal. Appl.*, Vol. 16, pp. 1173–1183, 1995.

[582] B.-Y. Wang and F. Zhang, "Schur Complements and Matrix Inequalities of Hadamard Products," *Lin. Multilin. Alg.*, Vol. 43, pp. 315–326, 1997.

[583] D. Wang, "The Polar Decomposition and a Matrix Inequality," *Amer. Math. Monthly*, Vol. 96, pp. 517–519, 1989.

[584] Q.-G. Wang, "Necessary and Sufficient Conditions for Stability of a Matrix Polytope with Normal Vertex Matrices," *Automatica*, Vol. 27, pp. 887–888, 1991.

[585] Y. W. Wang and D. S. Bernstein, "L_2 Controller Synthesis with L_∞-Bounded Closed-Loop Impulse Response," *Int. J. Contr.*, Vol. 60, pp. 1295–1306, 1994.

[586] J. Warga, *Optimal Control of Differential and Functional Equations.* New York: Academic Press, 1972.

[587] W. E. Waterhouse, "A Determinant Identity with Matrix Entries," *Amer. Math. Monthly*, Vol. 97, pp. 249–250, 1990.

[588] W. Watkins, "Convex Matrix Functions," *Proc. Amer. Math. Soc.*, Vol. 44, pp. 31–34, 1974.

[589] W. Watkins, "A Determinantal Inequality for Correlation Matrices," *Lin. Alg. Appl.*, Vol. 104, pp. 59–63, 1988.

[590] J. R. Weaver, "Centrosymmetric (Cross-Symmetric) Matrices, Their Basic Properties, Eigenvalues, and Eigenvectors," *Amer. Math. Monthly*, Vol. 92, pp. 711–717, 1985.

[591] R. Webster, *Convexity.* Oxford: Oxford University Press, 1994.

[592] M. Wei, "Reverse Order laws for Generalized Inverses of Multiple Matrix Products," *Lin. Alg. Appl.*, Vol. 293, pp. 273–288, 1999.

[593] Y. Wei, "Expressions for the Drazin Inverse of a 2×2 Block Matrix," *Lin. Multilin. Alg.*, Vol. 45, pp. 131–146, 1998.

[594] E. M. E. Wermuth, "Two Remarks on Matrix Exponentials," *Lin. Alg. Appl.*, Vol. 117, pp. 127–132, 1989.

[595] P. Wesseling, *Principles of Computational Fluid Dynamics.* Berlin: Springer, 2001.

[596] J. R. Westlake, *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations.* New York: John Wiley, 1968.

[597] N. A. Wiegmann, "Normal Products of Matrices," *Duke Math. J.*, Vol. 15, pp. 633–638, 1948.

[598] Z. Wiener, "An Interesting Matrix Exponent Formula," *Lin. Alg. Appl.*, Vol. 257, pp. 307–310, 1997.

[599] E. P. Wigner and M. M. Yanase, "On the Positive Semidefinite Nature of a Certain Matrix Expression," *Canad. J. Math.*, Vol. 16, pp. 397–406, 1964.

[600] R. M. Wilcox, "Exponential Operators and Parameter Differentiation in Quantum Physics," *J. Math. Physics*, Vol. 8, pp. 962–982, 1967.

[601] J. H. Wilkinson, *The Algebraic Eigenvalue Problem.* London: Oxford University Press, 1965.

[602] J. C. Willems, "Least Squares Stationary Optimal Control and the Algebraic Riccati Equation," *IEEE Trans. Autom. Contr.*, Vol. AC–16, pp. 621–634, 1971.

[603] D. A. Wilson, "Convolution and Hankel Operator Norms for Linear Systems," *IEEE Trans. Autom. Contr.*, Vol. AC–34, pp. 94–97, 1989.

[604] H. K. Wimmer, "Inertia Theorems for Matrices, Controllability and Linear Vibrations," *Lin. Alg. Appl.*, Vol. 8, pp. 337–343, 1974.

[605] H. K. Wimmer, "The Algebraic Riccati Equation Without Complete Controllability," *SIAM J. Alg. Disc. Math.*, Vol. 3, pp. 1–12, 1982.

[606] H. K. Wimmer, "The Algebraic Riccati Equation: Conditions for the Existence and Uniqueness of Solutions," *Lin. Alg. Appl.*, Vol. 58, pp. 441–452, 1984.

[607] H. K. Wimmer, "Monotonicity of Maximal Solutions of Algebraic Riccati Equations," *Sys. Contr. Lett.*, Vol. 5, pp. 317–319, 1985.

[608] H. K. Wimmer, "Linear Matrix Equations, Controllability and Observability, and the Rank of Solutions," *SIAM J. Matrix Anal. Appl.*, Vol. 9, pp. 570–578, 1988.

[609] H. K. Wimmer, "Lattice Properties of Sets of Semidefinite Solutions of Continuous-time Algebraic Riccati Equations," *Automatica*, Vol. 31, pp. 173–182, 1995.

[610] H. Wolkowicz and G. P. H. Styan, "Bounds for Eigenvalues Using Traces," *Lin. Alg. Appl.*, Vol. 29, pp. 471–506, 1980.

[611] W. A. Wolovich, *Linear Multivariable Systems.* New York: Springer-Verlag, 1974.

[612] M. J. Wonenburger, "A Decomposition of Orthogonal Transformations," *Canad. Math. Bull.*, Vol. 7, pp. 379–383, 1964.

[613] M. J. Wonenburger, "Transformations Which are Products of Two Involutions," *J. Math. Mech.*, Vol. 16, pp. 327–338, 1966.

[614] C. S. Wong, "Characterizations of Products of Symmetric Matrices," *Lin. Alg. Appl.*, Vol. 42, pp. 243–251, 1982.

[615] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*, 2nd ed. New York: Springer, 1979.

[616] P. Y. Wu, "Products of Nilpotent Matrices," *Lin. Alg. Appl.*, Vol. 96, pp. 227–232, 1987.

[617] P. Y. Wu, "Products of Positive Semidefinite Matrices," *Lin. Alg. Appl.*, Vol. 111, pp. 53–61, 1988.

[618] P. Y. Wu, "The Operator Factorization Problems," *Lin. Alg. Appl.*, Vol. 117, pp. 35–63, 1989.

[619] Z.-G. Xiao and Z.-H. Zhang, "The Inequalities $g \leq l \leq i \leq a$ in n Variables," *J. Inequal. Pure. Appl. Math.*, Vol. 4, no. 2/39, pp. 1–6, 2003.

[620] H. Xu, "Two Results About the Matrix Exponential," *Lin. Alg. Appl.*, Vol. 262, pp. 99–109, 1997.

[621] X. Yang, "Necessary Conditions of Hurwitz Polynomials," *Lin. Alg. Appl.*, Vol. 359, pp. 21–27, 2003.

[622] Z. P. Yang and X. X. Feng, "A Note on the Trace Inequality for Products of Hermitian Matrix Power," *J. Inequal. Pure. Appl. Math.*, Vol. 3, no. 5/78, pp. 1–12, 2002.

[623] X. Zhan, "On Some Matrix Inequalities," *Lin. Alg. Appl.*

[624] X. Zhan, "Inequalities for Unitarily Invariant Norms," *SIAM J. Matrix Anal. Appl.*, Vol. 20, pp. 466–470, 1998.

[625] X. Zhan, *Matrix Inequalities.* New York: Springer, 2002.

[626] F. Zhang, *Linear Algebra: Challenging Problems for Students.* Baltimore: Johns Hopkins University Press, 1996.

[627] F. Zhang, "Quaternions and Matrices of Quaternions," *Lin. Alg. Appl.*, Vol. 251, pp. 21–57, 1997.

[628] F. Zhang, "A Compound Matrix with Positive Determinant," *Amer. Math. Monthly*, Vol. 105, p. 958, 1998.

[629] F. Zhang, *Matrix Theory: Basic Results and Techniques.* New York: Springer, 1999.

[630] F. Zhang, "Schur Complements and Matrix Inequalities in the Lowner Ordering," *Lin. Alg. Appl.*, Vol. 321, pp. 399–410, 2000.

[631] L. Zhang, "A Characterization of the Drazin Inverse," *Lin. Alg. Appl.*, Vol. 335, pp. 183–188, 2001.

[632] K. Zhou, *Robust and Optimal Control.* Upper Saddle River: Prentice-Hall, 1996.

[633] D. Zwillinger, *Standard Mathematical Tables and Formulae*, 31st ed. Boca Raton: Chapman and Hall/CRC, 2003.

Author Index

Index

B

C

E

G

H

I

J

K

L

P

Q

T

V