Generalized Inverses: Theory and Applications

Adi Ben-Israel

Thomas N.E. Greville†

RUTCOR–Rutgers Center for Operations Research, Rutgers University,
640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA
E-mail address: bisrael@rutcor.rutgers.edu

From the preface to the First Edition

This book is intended to provide a survey of generalized inverses from a unified point of view, illustrating the theory with applications in many areas. It contains more than 450 exercises at different levels of difficulty, many of which are solved in detail. This feature makes it suitable either for reference and self–study or for use as a classroom text. It can be used profitably by graduate students or advanced undergraduates, only an elementary knowledge of linear algebra being assumed.

The book consists of an introduction and eight chapters, seven of which treat generalized inverses of finite matrices, while the eighth introduces generalized inverses of operators between Hilbert spaces. Numerical methods are considered in Chapter 7 and in Section 9.6.

While working in the area of generalized inverses, the authors have had the benefit of conversations and consultations with many colleagues. We would like to thank especially A. Charnes, R. E. Cline, P. J. Erdelsky, I. Erdélyi, J. B. Hawkins, A. S. Householder, A. Lent, C. C. MacDuffee, M. Z. Nashed, P. L. Odell, D. W. Showalter, and S. Zlobec. However, any errors that may have occurred are the sole responsibility of the authors.

. . . .

Finally, we are deeply indebted to Beatrice Shube, Editor for Wiley–Interscience Division for her constant encouragement and patience during the long period of bringing the manuscript to completion.

September 1973 *A. Ben–Israel*
T. N. E. Greville

Preface to the Second Edition

The field of generalized inverses has grown much since the appearance of the first edition in 1974, and is still growing. I tried to account for these developments while maintaining the informal and leisurely style of the first edition. New material was added, including a preliminary chapter (Chapter 0), a chapter on applications (Chapter 8), an Appendix on the work of E. H. Moore, new exercises and applications.

While preparing this volume I compiled a bibliography on generalized inverses, posted in the webpage of the *International Linear Algebra Society*

`http://www.math.technion.ac.il/iic/research.html`

This on-line bibliography, containing 2000+ items, will be updated from time to time. For reasons of space, many important works that appear in the on-line bibliography are not included in the bibliography of this book. I apologize to the authors of these works.

Many colleagues helped this effort. Special thanks go to R. Bapat, S. Campbell, J. Miao, S. K. Mitra, Y. Nievergelt, R. Puystjens, A. Sidi, G. Wang and Y. Wei.

Tom Greville, my friend and co–author, passed away before this project started. His scholarship and style marked the first edition, and are sadly missed.

January 2002 *A. Ben–Israel*

Contents

Glossary of notation

$\Gamma(p)$ – Gamma function, 286
$\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$, 83
$\gamma(T)$, 298
$\lambda^\dagger$ – Moore–Penrose inverse of the scalar λ, 37
$\lambda(A)$ – spectrum of A, 11
$\langle \alpha \rangle$ – smallest integer $\geq \alpha$, 248
$\nu(\lambda)$ – index of eigenvalue λ, 31
$\overline{1,n}$– the index set $\{1, 2, \ldots, n\}$, 4
π^{-1} – permutation inverse to π, 19
$\rho(A)$ – spectral radius of A, 17
$\sigma(A)$ – singular values of A (see footnote, p. 11), 12
$\sigma_j(A)$ – the jth singular value of A, 12
$\succeq$ – Löwner ordering, 69, 256, 257
$\tau(i)$ – period of state i, 272

A/A_{11} – Schur complement of A_{11} in A, 26
$A \succ O$, 69
$A \succeq O$, 69
$A\{1\}_{T,S}$ – $\{1\}$ inverses of A associated with T, S, 62
$A\{i, j, \ldots, k\}_s$ – matrices in $A\{i, j, \ldots, k\}$ of rank s, 49
A^* – adjoint of A, 10
$A^{(-1)}_{(L)}$ – Bott–Duffin inverse of A w.r.t. L, 79
$A^{1/2}$ – square root of A, 198
A^D – Drazin inverse of A, 145, 146
$A\{2\}_{T,S}$ – $\{2\}$–inverses with range T, null space S, 63
$A\{i, j, \ldots, k\}$ – $\{i, j, \ldots, k\}$–inverses of A, 35
$A^{(-1)}_{\alpha,\beta}$ – α–β generalized inverse of A, 118
$A^{(1)}_{T,S}$ – a $\{1\}$ inverse of A associated with T, S, 62
$A^{(1,2)}_{(W,Q)}$ – $\{W, Q\}$ weighted $\{1, 2\}$ inverse of A, 104, 105
$A^{(i,j,\ldots,k)}$ – an $\{i, j, \ldots, k\}$–inverse of A, 35
$A^\#$ – group inverse of A, 138
$A^\dagger$ – Moore–Penrose inverse of A, 35
$\| A \|_\infty$ – ∞–norm of a matrix, 17
$\| A \|_{\alpha,\beta}$ – l.u.b. of A w.r.t. $\{\alpha, \beta\}$, 126
$\widehat{A}$, 84
$\widetilde{A}$ – perturbation of A, 212
$\| A \|_1$ – 1–norm of a matrix, 17
$\| A \|_2$ – spectral norm of a matrix, 17
$A : B$ – Anderson–Duffin parallel sum of A, B, 254
$A \otimes B$ – Kronecker product of A, B, 46
$A \succeq B$ – Löwner ordering, 69
$A \overset{*}{<} B$ – $*$–order, 73
$A \mp B$ – Rao–Mitra parallel sum of A, B, 254
$\| A \|_F$ – Frobenius norm, 16
$A[I, *]$, 8
A_{I*}, 8
$A[I, J]$, 8
A_{IJ}, 8
$A[*, J]$, 8
A_{*J}, 8
$A[j \leftarrow \mathbf{b}]$ – A with j th–column replaced by $\mathbf{b}$, 25
$A_{(k)}$ – best rank-k approximation of A, 189
$A^{<k>}$ – generalized kth power of A, 222
$A^{(N)}$ – nilpotent part of A, 150
$\| A \|_p$ – p–norm of a matrix, 17
$A_{[S]}$ – restriction of A to S, 76
$A^{(S)}$ – S–inverse of A, 153
$A_{\{\mathcal{U},\mathcal{V}\}}$ – matrix representation of A w.r.t $\{\mathcal{U}, \mathcal{V}\}$, 9
$A_{\{\mathcal{V}\}}$ – matrix representation of A w.r.t $\{\mathcal{V}, \mathcal{V}\}$, 10
$A^{(1,2)}_{(W,Q)}$ – $\{W, Q\}$ weighted $\{1, 2\}$ inverse of A, 228

$Beta(p, q)$ – Beta distribution, 293
$\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ – bounded operators in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, 296
$B(p, q)$ – Beta function, 286

$\mathbb{C}$, 5
$C[a, b]$ – continuous functions on $[a, b]$, 311
$\mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ – closed operators in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, 296
$C_k(A)$ – k–compound matrix, 28
$\mathbb{C}^{m \times n}$ – $m \times n$ complex matrices, 35
$\mathbb{C}^{m \times n}_r$ – $m \times n$ complex matrices with rank r, 20
$\operatorname{cond}(A)$ – condition number of A, 181
$\cos\{L, M\}$ – “cos” of angle between subspaces, 208
$\operatorname{Cov} \mathbf{x}$ – covariance of $\mathbf{x}$, 255
$C(T)$, 296

$\mathcal{D}_+$ – positive diagonal matrices, 110
$d(A)$ – diagonal elements in UDV^*–decomposition, 186
$\det A$ – determinant of A, 24
$\operatorname{diag}(a_{11}, \cdots, a_{pp})$ – diagonal matrix, 9
$\operatorname{dist}(L, M)$ – distance between L, M, 208
$D(T)$, 296

$\mathbf{e}$ – vector of ones, 271
$\mathbf{e}\mathbf{e}^T$ – matrix of ones, 87
$E^i(\alpha)$ – elementary operation of type 1, 18
E^{ij} – elementary operation of type 3, 18
$E^{ij}(\beta)$ – elementary operation of type 2, 18
$\mathcal{E}_n$ – standard basis of $\mathbb{C}^n$, 10
EP – matrices A with $R(A) = R(A^*)$, 139
EP_r, 139
$\operatorname{E} \mathbf{x}$ – expected value of $\mathbf{x}$, 255
$\operatorname{ext} B$ – extension of B to $\mathbb{C}^n$, 76

$f^{(n)}_{ij}$ – probability of 1 st transition $i \leftarrow j$ in n th step, 272
$\mathcal{F}(A)$ – functions $f : \mathbb{C} \to \mathbb{C}$ analytic on $\lambda(A)$, 58
fl – floating point, 93
$\mathbb{F}^{m \times n}$ – $m \times n$ matrices over $\mathbb{F}$, 8

$\mathbb{F}^n$, 5

$G(\mathbf{x}_1, \dots, \mathbf{x}_n)$ – Gram matrix, 25
$G(T)$, 296
$G^{-1}(T)$, 296

$\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$ – Hilbert spaces, 295
$\mathcal{H}_{\boldsymbol{\xi},p}$ – hyperplane, 282

$i \leftrightsquigarrow j$ – states i, j communicate, 272
$\mathcal{I}(A)$, 25
Ind A – index of A, 136
$IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$, 82

$\mathcal{J}(A)$, 25
$J_k(\lambda)$ – Jordan block, 30

$L^{\perp}$ – orthogonal complement of L, 10, 62, 295
$\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ – linear transformations: $\mathbb{C}^n \to \mathbb{C}^m$, 51
LHS$(i.j)$ – left–hand side of equation $(i.j)$, 4
$\text{lub}_{\alpha,\beta}(A)$, 126
$\mathcal{L}(U, V)$ – linear transformations: $U \to V$, 9

$M \oplus N$ – direct sum of M, N, 295
$M \overset{\perp}{\oplus} N$ – orthogonal direct sum of M, N, 10, 295

$\mathcal{N}(A)$, 25
$N(A, B)$ – matrices X with $AXB = O$, 96
$N(T)$ – null space of T, 9, 296

P_π – permutation matrix corresponding to π, 19
$p_{ij}^{(n)}$ – n–step transition probability, 271
PD_n – $n \times n$ PD matrices, 11, 70
$\pi_i^{(t)}$ – probability of $\mathbf{X}_t = i$, 273
P_L – orthogonal projector on L, 64
$P_{L,\phi}$ – ϕ–metric projector on L, 115
$P_{L,\phi}^{-1}(\mathbf{l})$ – inverse image of $\mathbf{l}$ under $P_{L,\phi}$, 117
$P_{L,M}$ – projector on L along M, 51
PSD_n – $n \times n$ PSD matrices, 11, 70

$Q(\alpha)$ – projective bound of α, 126
$Q_{k,n}$ – increasing k sequences in $\overline{1, n}$, 8

$R(\lambda, A)$ – resolvent of A, 60, 220
$\mathbb{R}$, 5
$\widehat{R}(\lambda, A)$ – generalized resolvent of A, 220
$R(A, B)$ – matrices AXB for some X, 96
$\Re$ – real part, 6
RHS$(i.j)$ – right–hand side of equation $(i.j)$, 4
R_k – residual, 241
$R(L, M)$ – coefficient of inclination between L, M, 205
$r(L, M)$ – dimension of inclination between L, M, 205
$\mathbb{R}^{m \times n}$ – $m \times n$ real matrices, 35
$\mathbb{R}_r^{m \times n}$ – $m \times n$ real matrices with rank r, 20
$\mathbb{R}_J^n$ – basic subspace, 210
$R(T)$ – range of T, 9, 296
RV – random variable, 289

$\mathcal{S}$ – function space, 311
sign π – sign of permutation π, 19
$\sin\{L, M\}$ – "sin" of angle between subspaces, 208
S_n – symmetric group (permutations of order n), 19

$(T_2)_{[D(T_1)]}$ – restriction of T_2 to $D(T_1)$, 297
$T \succcurlyeq O$, 298
T^*, 297
T_r – restriction of T, 305
$T_S^\dagger$ – the $N(S)$–restricted pseudoinverse of T, 323
$T_e^\dagger$ – extremal g.i., 320
T^q – Tseng generalized inverse, 300

$U^{n \times n}$ – $n \times n$ unitary matrices, 179

vec(X) – vector made of rows of X, 47
vol A – volume of matrix A, 25

$W_\ell^{m \times n}$ – partial isometries in $\mathbb{C}_\ell^{m \times n}$, 203

$\| \mathbf{x} \|$ – norm of $\mathbf{x}$, 6
$\| \mathbf{x} \|_Q$ – ellipsoidal norm of $\mathbf{x}$, 7
$\langle X, Y \rangle$ – inner product on $\mathbb{C}^{m \times n}$, 96
$\angle\{\mathbf{x}, \mathbf{y}\}$ – angle between $\mathbf{x}, \mathbf{y}$, 7
$\langle \mathbf{x}, \mathbf{y} \rangle$ – inner product of $\mathbf{x}, \mathbf{y}$, 5, 295
$\langle \mathbf{x}, \mathbf{y} \rangle_Q$ – the inner product $\mathbf{y}^* Q \mathbf{x}$, 7

$(\mathbf{y}, X\boldsymbol{\beta}, V^2)$ – linear model, 255

$\mathbb{Z}$ – ring of integers, 33
$\mathbb{Z}^m$ – m–dimensional vector space over Z, 33
$\mathbb{Z}^{m \times n}$ – $m \times n$ matrices over $\mathbb{Z}$, 33
$\mathbb{Z}_r^{m \times n}$ – $m \times n$ matrices over $\mathbb{Z}$ with rank r, 33

Introduction

1. The inverse of a nonsingular matrix

It is well known that every nonsingular matrix A has a unique inverse, denoted by A^{-1}, such that

$$A\,A^{-1} = A^{-1}A = I \;, \tag{1}$$

where I is the identity matrix. Of the numerous properties of the inverse matrix, we mention a few. Thus,

$$\begin{aligned} \left(A^{-1}\right)^{-1} &= A \;, \\ \left(A^{T}\right)^{-1} &= \left(A^{-1}\right)^{T} \;, \\ \left(A^{*}\right)^{-1} &= \left(A^{-1}\right)^{*} \;, \\ (AB)^{-1} &= B^{-1}A^{-1} \;, \end{aligned}$$

where A^T and A^*, respectively, denote the transpose and conjugate transpose of A. It will be recalled that a real or complex number λ is called an eigenvalue of a square matrix A, and a nonzero vector $\mathbf{x}$ is called an eigenvector of A corresponding to λ, if

$$A\mathbf{x} = \lambda\mathbf{x} \;.$$

Another property of the inverse A^{-1} is that its eigenvalues are the reciprocals of those of A.

2. Generalized inverses of matrices

A matrix has an inverse only if it is square, and even then only if it is nonsingular, or, in other words, if its columns (or rows) are linearly independent. In recent years needs have been felt in numerous areas of applied mathematics for some kind of partial inverse of a matrix that is singular or even rectangular. By a *generalized inverse* of a given matrix A we shall mean a matrix X associated in some way with A that (i) exists for a class of matrices larger than the class of nonsingular matrices, (ii) has some of the properties of the usual inverse, and (iii) reduces to the usual inverse when A is nonsingular. Some writers have used the term “pseudoinverse” rather than “generalized inverse”.

As an illustration of part (iii) of our description of a generalized inverse, consider a definition used by a number of writers (e.g., Rohde [**710**]) to the effect that a generalized inverse of A is any matrix satisfying

$$AXA = A \;. \tag{2}$$

If A were nonsingular, multiplication by A^{-1} both on the left and on the right would give at once

$$X = A^{-1} \;.$$

3. Illustration: Solvability of linear systems

Probably the most familiar application of matrices is to the solution of systems of simultaneous linear equations. Let

$$A\mathbf{x} = \mathbf{b} \tag{3}$$

be such a system, where $\mathbf{b}$ is a given vector and $\mathbf{x}$ is an unknown vector. If A is nonsingular, there is a unique solution for $\mathbf{x}$ given by

$$\mathbf{x} = A^{-1}\mathbf{b} \ .$$

In the general case, when A may be singular or rectangular, there may sometimes be no solutions or a multiplicity of solutions.

The existence of a vector $\mathbf{x}$ satisfying (3) is tantamount to the statement that $\mathbf{b}$ is some linear combination of the columns of A. If A is $m \times n$ and of rank less than m, this may not be the case. If it is, there is some vector $\mathbf{h}$ such that

$$\mathbf{b} = A\mathbf{h} \ .$$

Now, if X is some matrix satisfying (2), and if we take

$$\mathbf{x} = X\mathbf{b} \ ,$$

we have

$$A\mathbf{x} = AX\mathbf{b} = AXA\mathbf{h} = A\mathbf{h} = \mathbf{b} \ ,$$

and so this $\mathbf{x}$ satisfies (3).

In the general case, however, when (3) may have many solutions, we may desire not just one solution but a characterization of all solutions. It has been shown (Bjerhammar [**100**], Penrose [**637**]) that, if X is any matrix satisfying $AXA = A$, then $A\mathbf{x} = \mathbf{b}$ has a solution if and only if

$$AX\mathbf{b} = \mathbf{b} \ ,$$

in which case the most general solution is

$$\mathbf{x} = X\mathbf{b} + (I - XA)\mathbf{y} \ , \tag{4}$$

where $\mathbf{y}$ is arbitrary.

We shall see later that for every matrix A there exist one or more matrices satisfying (2).

Exercises.

EX. 1. If A is nonsingular and has an eigenvalue λ, and $\mathbf{x}$ is a corresponding eigenvector, show that λ^{-1} is an eigenvalue of A^{-1} with the same eigenvector $\mathbf{x}$.

EX. 2. For any square A, let a "generalized inverse" be defined as any matrix X satisfying $A^{k+1}X = A^k$ for some positive integer k. Show that $X = A^{-1}$ if A is nonsingular.

EX. 3. If X satisfies $AXA = A$, show that $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $AX\mathbf{b} = \mathbf{b}$.

EX. 4. Show that (4) is the most general solution of $A\mathbf{x} = \mathbf{b}$. [Hint: First show that it is a solution; then show that every solution can be expressed in this form. Let $\mathbf{x}$ be any solution; then write $\mathbf{x} = XA\mathbf{x} + (I - XA)\mathbf{x}$.]

EX. 5. If A is an $m \times n$ matrix of zeros, what is the class of matrices X satisfying $AXA = A$?

EX. 6. Let A be an $m \times n$ whose elements are all zeros except the (i, j)th element, which is equal to 1. What is the class of matrices X satisfying (2)?

EX. 7. Let A be given, and let X have the property that $\mathbf{x} = X\mathbf{b}$ is a solution of $A\mathbf{x} = \mathbf{b}$ for *all* $\mathbf{b}$ such that a solution exists. Show that X satisfies $AXA = A$.

4. Diversity of generalized inverses

From Exercises 3, 4 and 7 the reader will perceive that, for a given matrix A, the matrix equation $AXA = A$ alone characterizes those generalized inverses X that are of use in analyzing the solutions of the linear system $A\mathbf{x} = \mathbf{b}$. For other purposes, other relationships play an essential role. Thus, if we are concerned with least–squares properties, (2) is not enough and must be supplemented by further relations. There results a more restricted class of generalized inverses.

If we are interested in spectral properties (i.e., those relating to eigenvalues and eigenvectors), consideration is necessarily limited to square matrices, since only these have eigenvalues and eigenvectors. In this connection, we shall see that (2) plays a role only for a restricted class of matrices A and must be supplanted, in the general case, by other relations.

Thus, unlike the case of the nonsingular matrix, which has a single unique inverse for all purposes, there are different generalized inverses for different purposes. For some purposes, as in the examples of solutions of linear systems, there is not a unique inverse, but any matrix of a certain class will do.

This book does not pretend to be exhaustive, but seeks to develop and describe in a natural sequence the most interesting and useful kinds of generalized inverses and their properties. For the most part, the discussion is limited to generalized inverses of finite matrices, but extensions to infinite–dimensional spaces and to differential and integral operators are briefly introduced in Chapter 9. Pseudoinverses on general rings and semigroups are not discussed; the interested reader is referred to Drazin [**226**], Foulis [**278**], and Munn [**584**].

The literature on generalized inverses has become so extensive that it would be impossible to do justice to it in a book of moderate size. We have been forced to make a selection of topics to be covered, and it is inevitable that not everyone will agree with the choices we have made. We apologize to those authors whose work has been slighted. A virtually complete bibliography as of 1976 is found in Nashed and Rall [**594**]. An on-line bibliography is posted in the webpage of the *International Linear Algebra Society* `http://www.math.technion.ac.il/iic/research.html`

5. Preparation expected of the reader

It is assumed that the reader has a knowledge of linear algebra that would normally result from completion of an introductory course in the subject. In particular, vector spaces will be extensively utilized. Except in Chapter 9, which deals with Hilbert spaces, the vector spaces and linear transformations used are finite–dimensional, real or complex. Familiarity with these topics is assumed, say at the level of Halmos [**359**] or Noble [**613**], see also Chapter 0 below.

6. Historical note

The concept of a generalized inverse seems to have been first mentioned in print in 1903 by Fredholm [**284**], where a particular generalized inverse (called by him "pseudoinverse") of an integral operator was given. The class of all pseudoinverses was characterized in 1912 by Hurwitz [**430**], who used the finite dimensionality of the null spaces of the Fredholm operators to give a simple algebraic construction (see, e.g., Exercises 9.19–9.20). Generalized inverses of differential operators, already implicit in Hilbert's discussion in 1904 of generalized Green's functions, [**412**], were consequently studied by numerous authors, in particular Myller (1906), Westfall (1909), Bounitzky [**123**] in 1909, Elliott (1928), and Reid (1931). For a history of this subject see the excellent survey by Reid [**691**].

Generalized inverses of differential and integral operators thus antedated the generalized inverses of matrices, whose existence was first noted by E.H. Moore, who defined a unique inverse (called by him the "general reciprocal") for every finite matrix (square or rectangular). Although his first publication on the subject [**571**], an abstract of a talk given at a meeting of the American

Mathematical Society, appeared in 1920, his results are thought to have been obtained much earlier. One writer, [**491**, p. 676], has assigned the date 1906. Details were published, [**572**], only in 1935 after Moore's death. A summary of Moore's work on the *general reciprocal* is given in Appendix A. Little notice was taken of Moore's discovery for 30 years after its first publication, during which time generalized inverses were given for matrices by Siegel [**767**] in 1937, and for operators by Tseng ([**824**]–1933, [**827**],[**825**],[**826**]–1949), Murray and von Neumann [**586**] in 1936, Atkinson ([**28**]–1952, [**29**]–1953) and others. Revival of interest in the subject in the 1950s centered around the least squares properties (not mentioned by Moore) of certain generalized inverses. These properties were recognized in 1951 by Bjerhammar, who rediscovered Moore's inverse and also noted the relationship of generalized inverses to solutions of linear systems (Bjerhammar [**99**], [**98**], [**100**]). In 1955 Penrose [**637**] sharpened and extended Bjerhammar's results on linear systems, and showed that Moore's inverse, for a given matrix A is the unique matrix X satisfying the four equations (1)–(4) of the next chapter. The latter discovery has been so important and fruitful that this unique inverse (called by some writers *the* generalized inverse) is now commonly called the *Moore–Penrose inverse.*

Since 1955 thousands of papers on various aspects of generalized inverses and their applications have appeared. In view of the vast scope of this literature, we shall not attempt to trace the history of the subject further, but the subsequent chapters will include selected references on particular items.

7. Remarks on notation

Equation j of Chapter i is denoted by (j) in Chapter i, and by $(i.j)$ in other chapters. Theorem j of Chapter i is called Theorem j in Chapter i, and Theorem $i.j$ in other chapters. Similar conventions apply to corollaries, lemmas, exercises, definitions etc.. The left and right members of equation $(i.j)$ are denoted LHS$(i.j)$ and RHS$(i.j)$, respectively. The index set $\{1, 2, \dots, n\}$ is denoted $\overline{1, n}$.

Suggested further reading

Section 2. A ring $\mathcal{R}$ is called *regular* if for every $A \in \mathcal{R}$ there exists an $X \in \mathcal{R}$ satisfying $AXA = A$. See von Neumann ([**846**], [**849**, p. 90]), Murray and von Neumann [**586**, p. 299], McCoy [**534**], and Hartwig [**373**].

Section 4. For generalized inverses in abstract geometric setting see also Davis and Robinson [**211**], Gabriel ([**285**], [**286**], [**287**]), Hansen and Robinson [**367**], Hartwig [**373**], Munn and Penrose [**585**], Pearl [**636**], Rabson [**665**] and Rado [**666**].

CHAPTER 0

Preliminaries

For ease of reference we collect here facts, definitions and notations that are used in successive chapters. This chapter can be skipped in first reading.

1. Scalars and vectors

1.1. *Scalars* are denoted by low case letters: $x, y, \lambda, \ldots$. We use mostly the *complex field* $\mathbb{C}$, and specialize to the *real field* $\mathbb{R}$ as necessary. A generic field is denoted by $\mathbb{F}$.

1.2. *Vectors* are denoted by bold letters: $\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \ldots$. Vector spaces are finite–dimensional, except in Chapter 9. The n–dimensional vector space over a field $\mathbb{F}$ is denoted by $\mathbb{F}^n$, its elements by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \ , \text{ or } \mathbf{x} = (x_i) \ , \ i \in \overline{1, n} \ , \ , \ x_i \in \mathbb{F} \ .$$

The n–dimensional vector $\mathbf{e}_i$ with components

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

is called the i <u>th</u>–*unit vector* of $\mathbb{F}^n$. The set $\mathcal{E}_n$ of unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is called the *standard basis* of $\mathbb{F}^n$.

1.3. The *sum* of two sets L, M in $\mathbb{C}^n$, denoted by $L + M$, is defined as

$$L + M = \{\mathbf{y} + \mathbf{z} : \ \mathbf{y} \in L, \mathbf{z} \in M\} \ .$$

If L and M are subspaces of $\mathbb{C}^n$, then $L + M$ is also a subspace of $\mathbb{C}^n$. If, in addition, $L \cap M = \{\mathbf{0}\}$, i.e., the only vector common to L and M is the zero vector, then $L + M$ is called the *direct sum* of L and M, denoted by $L \oplus M$. Two subspaces L and M of $\mathbb{C}^n$ are called *complementary* if

$$\mathbb{C}^n = L \oplus M \ . \tag{1}$$

When this is the case (see Ex. 1 below), every $\mathbf{x} \in \mathbb{C}^n$ can be expressed uniquely as a sum

$$\mathbf{x} = \mathbf{y} + \mathbf{z} \quad (\mathbf{y} \in L, \mathbf{z} \in M) \ . \tag{2}$$

We shall then call $\mathbf{y}$ the *projection of* $\mathbf{x}$ *on* L *along* M.

1.4. *Inner product.* Let V be a complex vector space. An *inner product* is a function : $V \times V \mapsto \mathbb{C}$, denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, that satisfies

(I1) $\langle \alpha\mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ (*linearity*),

(I2) $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ (*Hermitian symmetry*),

(I3) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$ (positivity),

for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha \in \mathbb{C}$.

Note:

(a) For all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, $\langle \mathbf{x}, \alpha\mathbf{y} \rangle = \overline{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle$ by (I1)–(I2).

(b) Condition (I2) states, in particular, that $\langle \mathbf{x}, \mathbf{x} \rangle$ is real for all $\mathbf{x} \in V$.
(c) The *if* part in (I3) follows from (I1) with $\alpha = 0$, $\mathbf{y} = \mathbf{0}$.

The *standard inner product* in $\mathbb{C}^n$ is

$$\mathbf{y}^* \mathbf{x} = \sum_{i=1}^{n} x_i \, \overline{y_i} \, , \tag{3}$$

for all $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_i)$ in $\mathbb{C}^n$. See Exs. 2–4.

1.5. Let V be a complex vector space. A (*vector*) *norm* is a function : $V \mapsto \mathbb{R}$, denoted by $\|\mathbf{x}\|$, that satisfies

(N1) $\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$ (*positivity*),
(N2) $\|\alpha \mathbf{x}\| = |\alpha| \, \|\mathbf{x}\|$ (*positive homogeneity*),
(N3) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (*triangle inequality*),

for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$.

Note:
(a) The *if* part of (N1) follows from (N2).
(b) $\|\mathbf{x}\|$ is interpreted as the *length* of the vector $\mathbf{x}$. Inequality (N3) then states, in $\mathbb{R}^2$, that the length of any side of a triangle is no greater than the sum of lengths of the other two sides.

See Exs. 3–11.

Exercises and examples.

EX. 1. *Direct sums.* Let L and M be subspaces of a vector space V. Then the following statements are equivalent:
(a) $V = L \oplus M$.
(b) Every vector $\mathbf{x} \in V$ is uniquely represented as $\mathbf{x} = \mathbf{y} + \mathbf{z} \quad (\mathbf{y} \in L \, , \, \mathbf{z} \in M)$.
(c) $\dim V = \dim L + \dim M \, , \; L \cap M = \{\mathbf{0}\}$.
(d) If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ are bases for L and M, respectively, then

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$$

is a basis for V.

EX. 2. *The Cauchy–Schwartz inequality.* For any $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \, \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle} \tag{4}$$

with equality if and only if $\mathbf{x} = \lambda \mathbf{y}$ for some $\lambda \in \mathbb{C}$.

PROOF. For any complex z,

$$\begin{aligned} 0 &\leq \langle \mathbf{x} + z\mathbf{y}, \mathbf{x} + z\mathbf{y} \rangle \, , \text{ by (I3)} \, , \\ &= \langle \mathbf{y}, \mathbf{y} \rangle |z|^2 + z \langle \mathbf{y}, \mathbf{x} \rangle + \overline{z} \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{x} \rangle \, , \text{ by (I1)–(I2)} \\ &= \langle \mathbf{y}, \mathbf{y} \rangle |z|^2 + 2\Re \{ z \, \langle \mathbf{x}, \mathbf{y} \rangle \} + \langle \mathbf{x}, \mathbf{x} \rangle \, , \\ &\leq \langle \mathbf{y}, \mathbf{y} \rangle |z|^2 + 2|z| |\langle \mathbf{x}, \mathbf{y} \rangle| + \langle \mathbf{x}, \mathbf{x} \rangle \, . \end{aligned} \tag{5}$$

Here $\Re$ denotes *real part.* The quadratic equation RHS(5) $= 0$ can have at most one solution $|z|$, proving that $|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \, \langle \mathbf{y}, \mathbf{y} \rangle$, with equality if and only if $\mathbf{x} + z\mathbf{y} = \mathbf{0}$ for some $z \in \mathbb{C}$. □

EX. 3. If $\langle \mathbf{x}, \mathbf{y} \rangle$ is an inner product on $\mathbb{C}^n$, then

$$\|\mathbf{x}\| \, := \, \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{6}$$

is a norm on $\mathbb{C}^n$. The *Euclidean norm* in $\mathbb{C}^n$

$$\|\mathbf{x}\| = \sqrt{\sum_{j=1}^{n} |x|^2} \, , \tag{7}$$

corresponds to the standard inner–product. (HINT: Use (4) to verify the triangle inequality (N3) in § 1.5).

EX. 4. Show that to every inner product $f : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ there corresponds a unique positive definite $Q = [q_{ij}] \in \mathbb{C}^{n\times n}$ such that

$$f(\mathbf{x},\mathbf{y}) = \mathbf{y}^* Q \mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{y_i}\, q_{ij} x_j \ . \tag{8}$$

The inner product (8) is denoted by $\langle \mathbf{x},\mathbf{y}\rangle_Q$. It induces a norm, by Ex. 3,

$$\|\mathbf{x}\|_Q = \sqrt{\mathbf{x}^* Q \mathbf{x}}\ ,$$

called *ellipsoidal*, or *weighted Euclidean* norm. The standard inner product (3), and the Euclidean norm, correspond to the special case $Q = I$.

SOLUTION. The inner product f and the positive definite matrix $Q = [q_{ij}]$ completely determine each other by

$$f(\mathbf{e}_i,\mathbf{e}_j) = q_{ij}, \quad (i, j \in \overline{1,n})\ ,$$

where $\mathbf{e}_i$ is the ith unit vector. □

EX. 5. Given an inner product $\langle \mathbf{x},\mathbf{y}\rangle$ and the corresponding norm $\|\mathbf{x}\| = \langle \mathbf{x},\mathbf{x}\rangle^{1/2}$, the *angle* between two vectors $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, denoted by $\angle\{\mathbf{x},\mathbf{y}\}$, is defined by

$$\cos\angle\{\mathbf{x},\mathbf{y}\} = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|}\ . \tag{9}$$

Two vectors $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$ are *orthogonal* if $\langle \mathbf{x},\mathbf{y}\rangle = 0$. Although it is not obvious how to define angles between vectors in $\mathbb{C}^n$, see, e.g. Scharnhorst [**731**], we define orthogonality by the same condition, $\langle \mathbf{x},\mathbf{y}\rangle = 0$, as in the real case.

EX. 6. Let $\langle\cdot,\cdot\rangle$ be an inner product on $\mathbb{C}^n$. A set $\{\mathbf{v}_1,\dots,\mathbf{v}_k\}$ of $\mathbb{C}^n$ is called *orthonormal* (abbreviated *o.n.*) if

$$\langle \mathbf{v}_i,\mathbf{v}_j\rangle = \delta_{ij}\ , \text{ for all } i,j \in \overline{1,k}\ . \tag{10}$$

(a) An o.n. set is linearly independent.
(b) If $\mathcal{B} = \{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ is an *o.n. basis* of $\mathbb{C}^n$, then for all $\mathbf{x} \in \mathbb{C}^n$,

$$\mathbf{x} = \sum_{j=1}^{n} \xi_j \mathbf{v}_j\ , \text{ with } \xi_j = \langle \mathbf{x},\mathbf{v}_j\rangle\ , \tag{11}$$

and

$$\langle \mathbf{x},\mathbf{x}\rangle = \sum_{j=1}^{n} |\xi_j|^2\ . \tag{12}$$

EX. 7. *Gram–Schmidt orthonormalization.* Let $\mathcal{A} = \{\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_n\} \subset \mathbb{C}^m$ be a set of vectors spanning a subspace L, $L = \left\{\sum_{i=1}^{n} \alpha_i\,\mathbf{a}_i : \alpha_i \in \mathbb{C}\right\}$. Then an o.n. basis $\mathcal{Q} = \{\mathbf{q}_1,\mathbf{q}_2,\dots,\mathbf{q}_r\}$ of L is computed using the *Gram–Schmidt orthonormalization* process (abbreviated GSO) as follows.

$$\mathbf{q}_1 = \frac{\mathbf{a}_{c_1}}{\|\mathbf{a}_{c_1}\|} \quad \text{if } \mathbf{a}_{c_1} \neq \mathbf{0} = \mathbf{a}_j \text{ for } 1 \le j < c_1 \tag{13a}$$

$$\mathbf{x}_j = \mathbf{a}_j - \sum_{\ell=1}^{k-1} \langle \mathbf{a}_j,\mathbf{q}_\ell\rangle\,\mathbf{q}_\ell\ , \quad j = c_{k-1}+1, c_{k-1}+2,\dots,c_k \tag{13b}$$

and

$$\mathbf{q}_k = \frac{\mathbf{x}_{c_k}}{\|\mathbf{x}_{c_k}\|} \quad \text{if } \mathbf{x}_{c_k} \neq \mathbf{0} = \mathbf{x}_j \text{ for } c_{k-1}+1 \le j < c_k\ ,\ k = 2,\dots,r\ . \tag{13c}$$

The integer r found by the GSO process is the dimension of the subspace L. The integers $\{c_1,\dots,c_r\}$ are the indices of a maximal linearly independent subset $\{\mathbf{a}_{c_1},\dots,\mathbf{a}_{c_r}\}$ of $\mathcal{A}$.

EX. 8. Let $\|\ \|_{(1)}$, $\|\ \|_{(2)}$ be two norms on $\mathbb{C}^n$ and let α_1 , α_2 be positive scalars. Show that the following functions

(a) $\max\{\|\mathbf{x}\|_{(1)}, \|\mathbf{x}\|_{(2)}\}$ (b) $\alpha_1\|\mathbf{x}\|_{(1)} + \alpha_2\|\mathbf{x}\|_{(2)}$

are norms on $\mathbb{C}^n$.

Ex. 9. *The ℓ_p–norms.* for any $p \geq 1$ the function

$$\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} \tag{14}$$

is a norm on $\mathbb{C}^n$, called the *ℓ_p–norm.*
Hint: The statement that (14) satisfies (N3) for $p \geq 1$ is the classical Minkowski's inequality; see, e.g., Beckenbach and Bellman [**54**].

Ex. 10. The most popular ℓ_p–norms are the choices $p = 1, 2$, and ∞

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| \ , \text{ the } \ell_1\text{–norm} \ , \tag{14.1}$$

$$\|\mathbf{x}\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} = (\mathbf{x}^*\mathbf{x})^{1/2} \ , \text{ the } \ell_2\text{–}norm \text{ or the } Euclidean\ norm \ , \tag{14.2}$$

$$\|\mathbf{x}\|_\infty = \max\{|x_j| : j \in \overline{1,n}\} \ , \text{ the } \ell_\infty\text{–}norm \text{ or the } Tchebycheff\ norm \ . \tag{14.∞}$$

Is $\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p$?

Ex. 11. Let $\| \ \|_{(1)}$, $\| \ \|_{(2)}$ be any two norms on $\mathbb{C}^n$. Show that there exist positive scalars α, β such that

$$\alpha\|\mathbf{x}\|_{(1)} \leq \|\mathbf{x}\|_{(2)} \leq \beta\|\mathbf{x}\|_{(1)} \ , \tag{15}$$

for all $\mathbf{x} \in \mathbb{C}^n$.
Hint:

$$\alpha = \inf\{\|\mathbf{x}\|_{(2)} : \|\mathbf{x}\|_{(1)} = 1\} \ , \ \beta = \sup\{\|\mathbf{x}\|_{(2)} : \|\mathbf{x}\|_{(1)} = 1\} \ .$$

Remark 1. Two norms, $\| \ \|_{(1)}$ and $\| \ \|_{(2)}$ are called *equivalent* if there exist positive scalars α, β such that (15) holds for all $\mathbf{x} \in \mathbb{C}^n$. from Ex. 11, *any* two norms on $\mathbb{C}^n$ are equivalent. Therefore, if a sequence $\{\mathbf{x}_k\} \subset \mathbb{C}^n$ satisfies

$$\lim_{k\to\infty} \|\mathbf{x}_k\| = 0 \tag{16}$$

for some norm, then (16) holds for any norm. Topological concepts like convergence and continuity, defined by limiting expressions like (16), are therefore independent of the norm used in their definition. Thus we say that a sequence $\{\mathbf{x}_k\} \subset \mathbb{C}^n$ converges to a point $\mathbf{x}_\infty$ if

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}_\infty\| = 0$$

for some norm.

2. Linear transformations and matrices

2.1. The set of $m \times n$ matrices with elements in $\mathbb{F}$ is denoted $\mathbb{F}^{m\times n}$. A matrix $A \in \mathbb{F}^{m\times n}$ is *square* if $m = n$, *rectangular* otherwise.

The elements of a matrix $A \in \mathbb{F}^{m\times n}$ are denoted by a_{ij} or $A[i, j]$.

We denote by

$$Q_{k,n} = \{(i_1, i_2, \ldots , i_k) : 1 \leq i_1 < i_2 < \cdots < i_k \leq n\} \ .$$

the set of increasing sequences of k elements from $\overline{1,n}$, for given integers $0 < k \leq n$. For $A \in \mathbb{C}^{m\times n}$, $I \in Q_{p,m}$, $J \in Q_{q,n}$ we denote

A_{IJ} (or $A[I, J]$)	the $p \times q$ submatrix $(A[i, j])$, $i \in I$, $j \in J$,
A_{I*} (or $A[I, *]$)	the $p \times n$ submatrix $(A[i, j])$, $i \in I$, $j \in \overline{1,n}$,
A_{*J} (or $A[*, J]$)	the $m \times q$ submatrix $(A[i, j])$, $i \in \overline{1,m}$, $j \in J$.

The matrix A is

diagonal if $A[i, j] = 0$ for $i \neq j$,

upper triangular if $A[i,j]=0$ for $i>j$,
lower triangular if $A[i,j]=0$ for $i<j$.
An $m\times n$ diagonal matrix $A=\big[a_{ij}\big]$ is denoted $A=\operatorname{diag}(a_{11},\cdots,a_{pp})$ where $p=\min\{m,n\}$.
Given a matrix $A\in\mathbb{C}^{m\times n}$, its
transpose is the matrix $A^T\in\mathbb{C}^{n\times m}$ with $A^T[i,j]=A[j,i]$ for all i,j,
conjugate transpose is the matrix $A^*\in\mathbb{C}^{n\times m}$ with $A^*[i,j]=\overline{A[j,i]}$ for all i,j.
A square matrix is:
Hermitian [*symmetric*] if $A=A^*$ [A is real, $A=A^T$],
normal if $AA^*=A^*A$,
unitary [*orthogonal*] if $A^*=A^{-1}$ [A is real, $A^T=A^{-1}$].

2.2. Given vector spaces U,V over a field $\mathbb{F}$, and a mapping $T:U\longmapsto V$, we say that T is *linear*, or a *linear transformation*, if $T(\alpha\mathbf{x}+\mathbf{y})=\alpha T\mathbf{x}+T\mathbf{y}$, for all $\alpha\in\mathbb{F}$ and $\mathbf{x},\mathbf{y}\in U$. The set of linear transformations from U to V is denoted $\mathcal{L}(U,V)$. It is a vector space with operations T_1+T_2 and αT defined by

$$(T_1+T_2)\mathbf{u}=T_1\mathbf{u}+T_2\mathbf{u}\ ,\ (\alpha T)\mathbf{u}=\alpha(T\mathbf{u})\ ,\ \forall\ \mathbf{u}\in U\ .$$

The zero element of $\mathcal{L}(U,V)$ is the transformation O mapping every $\mathbf{u}\in U$ into $\mathbf{0}\in V$. The identity mapping $I_U\in\mathcal{L}(U,U)$ is defined by $I_U\mathbf{u}=\mathbf{u}\ ,\ \forall\ \mathbf{u}\in U$. We usually omit the subscript U, writing the identity as I.

2.3. Let $T\in\mathcal{L}(U,V)$. For any $\mathbf{u}\in U$, the point $T\mathbf{u}$ in V is called the *image* of $\mathbf{u}$ (under T). The *range* of T, denoted $R(T)$ is the set of all its images

$$R(T)=\{\mathbf{v}\in V:\ \mathbf{v}=T\mathbf{u}\text{ for some }\mathbf{u}\in U\}\ .$$

For any $\mathbf{v}\in R(T)$, the *inverse image* $T^{-1}(\mathbf{v})$ is the set

$$T^{-1}(\mathbf{v})=\{\mathbf{u}\in U:\ T\mathbf{u}=\mathbf{v}\}\ .$$

In particular, the *null space* of T, denoted by $N(T)$, is the inverse image of the zero vector $\mathbf{0}\in V$,

$$N(T)=\{\mathbf{u}\in U:\ T\mathbf{u}=\mathbf{0}\}\ .$$

2.4. $T\in\mathcal{L}(U,V)$ is *one–to–one* if for all $\mathbf{x},\mathbf{y}\in U$, $\mathbf{x}\neq\mathbf{y}\implies T\mathbf{x}\neq T\mathbf{y}$, or equivalently, if for every $\mathbf{v}\in R(T)$ the inverse image $T^{-1}\mathbf{v}$ is a singleton. T is *onto* if $R(T)=V$. If T is one–to–one and onto, it has an *inverse* $T^{-1}\in\mathcal{L}(V,U)$ such that

$$T^{-1}(T\mathbf{u})\ =\ \mathbf{u}\text{ and }T(T^{-1}\mathbf{v})=\mathbf{v}\ ,\ \forall\ \mathbf{u}\in U,\ \mathbf{v}\in V\ , \tag{17a}$$

$$\text{or equivalently}\ ,\ T^{-1}T\ =\ I_U\ ,\ TT^{-1}=I_V\ , \tag{17b}$$

in which case T is called *invertible* or *nonsingular*.

2.5. Given
- a linear transformation $A\in\mathcal{L}(\mathbb{C}^n,\mathbb{C}^m)$ and
- two bases $\mathcal{U}=\{\mathbf{u}_1,\dots,\mathbf{u}_m\}$ and $\mathcal{V}=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ of $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively,

the *matrix representation of A relative to the bases* $\{U,V\}$ is the $m\times n$ matrix $A_{\{\mathcal{U},\mathcal{V}\}}=\big[a_{ij}\big]$ determined (uniquely) by

$$A\mathbf{v}_j=\sum_{i=1}^{m}a_{ij}\,\mathbf{u}_i\ ,\ j\in\overline{1,n}\ . \tag{18}$$

For any such pair of bases $\{\mathcal{U},\mathcal{V}\}$, (18) is a one–to–one correspondence between the linear transformations $\mathcal{L}(\mathbb{C}^n,\mathbb{C}^m)$ and the matrices $\mathbb{C}^{m\times n}$, allowing the customary practice of using the same symbol A to denote both the linear transformation $A : \mathbb{C}^n \to \mathbb{C}^m$ and its matrix representation $A_{\{\mathcal{U},\mathcal{V}\}}$.

If A is a linear transformation from $\mathbb{C}^n$ to itself, and $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of $\mathbb{C}^n$, then the matrix representation $A_{\{\mathcal{V},\mathcal{V}\}}$ is denoted simply by $A_{\{\mathcal{V}\}}$. It is the (unique) matrix $A_{\{\mathcal{V}\}} = [a_{ij}] \in \mathbb{C}^{n\times n}$ satisfying

$$A\mathbf{v}_j = \sum_{i=1}^{n} a_{ij}\,\mathbf{v}_i \ , \ \ j \in \overline{1,n} \ . \tag{19}$$

The *standard basis* of $\mathbb{C}^n$ is the basis $\mathcal{E}_n$ consisting of the n unit vectors

$$\mathcal{E}_n = \{\mathbf{e}_1, \dots, \mathbf{e}_n\} \ .$$

Unless otherwise noted, linear transformations $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ are represented in terms of the standard bases $\{\mathcal{E}_m, \mathcal{E}_n\}$.

2.6. Let $\langle\cdot,\cdot\rangle$ denote the standard inner product. If $A \in \mathbb{C}^{m\times n}$ then

$$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^*\mathbf{y}\rangle \ , \quad \text{for all } \mathbf{x} \in \mathbb{C}^n, \ \mathbf{y} \in \mathbb{C}^m \ . \tag{20}$$

$H \in \mathbb{C}^{n\times n}$ is Hermitian if and only if

$$\langle H\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, H\mathbf{y}\rangle \ , \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \ . \tag{21}$$

If $\langle A\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A\mathbf{x}\rangle$ for all $\mathbf{x}$ then A need not be Hermitian. Example: $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.

2.7. Let $\langle\cdot,\cdot\rangle_{\mathbb{C}^n}$ and $\langle\cdot,\cdot\rangle_{\mathbb{C}^m}$ be inner products on C^n and $\mathbb{C}^m$, respectively, and let $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$. The *adjoint* of A, denoted by A^*, is the linear transformation $A^* \in \mathcal{L}(\mathbb{C}^m, \mathbb{C}^n)$ defined by

$$\langle A\mathbf{v}, \mathbf{u}\rangle_{\mathbb{C}^m} = \langle \mathbf{v}, A^*\mathbf{u}\rangle_{\mathbb{C}^n} \tag{22}$$

for all $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{u} \in \mathbb{C}^m$. Unless otherwise stated, we use the standard inner product, in which case *adjoint* = *conjugate transpose*.

2.8. Given a subspace L of $\mathbb{C}^n$, define

$$L^\perp := \{\mathbf{x} \in \mathbb{C}^n : \mathbf{x} \text{ is orthogonal to every vector in } L\} \ . \tag{23}$$

Then $L^\perp$ is a subspace complementary to L. $L^\perp$ is called the *orthogonal complement* of L, and $\mathbb{C}^n$ is called an *orthogonal direct sum* of L, $L^\perp$, a fact denoted by

$$\mathbb{C}^n = L \overset{\perp}{\oplus} L^\perp \ . \tag{24}$$

With any matrix $A \in \mathbb{C}^{m\times n}$ there are associated four subspaces

$$N(A),\ R(A^*) \quad \text{in } \mathbb{C}^n \ ,$$
$$N(A^*),\ R(A) \quad \text{in } \mathbb{C}^m \ .$$

An important result is that these pairs form orthogonal complements:

THEOREM 1. For any $A \in \mathbb{C}^{m\times n}$,

$$N(A) = R(A^*)^\perp \ , \tag{25}$$

$$N(A^*) = R(A)^\perp \ . \tag{26}$$

Two proofs of (25) are given below. The first, using inner products, is immediately generalizable to transformations on Hilbert space, which will be discussed in Chapter 9. The second proof shows that, in the more restricted context of finite matrices, (25) is a consequence of the equation $A\mathbf{x} = \mathbf{0}$, which defines $N(A)$. The dual relation (26) follows by reversing the roles of A, A^*.

FIRST PROOF. Let $\mathbf{x} \in N(A)$. Then LHS(20) vanishes for all $\mathbf{y} \in \mathbb{C}^m$. It follows then that $\mathbf{x} \perp A^*\mathbf{y}$ for all $\mathbf{y} \in \mathbb{C}^m$, or, in other words, $\mathbf{x} \perp R(A^*)$. This proves that $N(A) \subset R(A^*)^\perp$.

Conversely, let $\mathbf{x} \in R(A^*)^\perp$, so that RHS(20) vanishes for all $\mathbf{y} \in \mathbb{C}^m$. This implies that $A\mathbf{x} \perp \mathbf{y}$ for all $\mathbf{y} \in \mathbb{C}^m$. Therefore $A\mathbf{x} = \mathbf{0}$. This proves that $R(A^*)^\perp \subset N(A)$, and completes the proof. □

SECOND PROOF. By definition of matrix multiplication, $A\mathbf{x} = \mathbf{0}$ is equivalent to the statement that each row of A postmultiplied by $\mathbf{x}$ gives the product 0. Now, the rows of A are the conjugate transposes of the columns of A^*, and therefore $\mathbf{x} \in N(A)$ if and only if it is orthogonal to every column of A^*, i.e., if and only if it is orthogonal to the subspace spanned by these columns, namely $R(A^*)$. □

2.9. A (*matrix*) *norm* of $A \in \mathbb{C}^{m\times n}$, denoted by $\|A\|$, is defined as a function : $\mathbb{C}^{m\times n} \to \mathbb{R}$ that satisfies

$$\|A\| \geq 0 \ , \ \|A\| = 0 \text{ only if } A = O \ , \tag{M1}$$

$$\|\alpha A\| = |\alpha| \|A\| \ , \tag{M2}$$

$$\|A + B\| \leq \|A\| + \|B\| \ , \tag{M3}$$

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

$$\|AB\| \leq \|A\|\|B\| \tag{M4}$$

whenever the matrix product AB is defined, then $\|\ \|$ is called a *multiplicative norm*. Some authors (see, e.g., Householder [**427**, Section 2.2]) define a matrix norm as a function having all four properties (M1)–(M4).

2.10. If $A \in \mathbb{C}^{n\times n}$, $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ are such that

$$A\mathbf{x} = \lambda\mathbf{x} \tag{27}$$

then λ is an *eigenvalue* of A corresponding to the *eigenvector* $\mathbf{x}$. The set of eigenvalues of A is called its *spectrum*, and is denoted by[1] $\lambda(A)$. If λ is an eigenvalue of A, the subspace

$$\{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \lambda\mathbf{x}\} \tag{28}$$

is the corresponding *eigenspace* of A, its dimension is called the *geometric multiplicity* of the eigenvalue λ.

2.11. If $H \in \mathbb{C}^{n\times n}$ is Hermitian, then:
(a) the eigenvalues of H are real,
(b) eigenvectors corresponding to different eigenvalues are orthogonal,
(c) there is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of H, and
(d) the eigenvalues of H, ordered by

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \ ,$$

and corresponding eigenvectors,

$$H\mathbf{x}_j = \lambda_j \mathbf{x}_j \ , \ j \in \overline{1,n} \ ,$$

can be computed recursively as

$$\lambda_1 = \max\{\langle H\mathbf{x}, \mathbf{x}\rangle : \|\mathbf{x}\| = 1\} = \langle H\mathbf{x}_1, \mathbf{x}_1\rangle \ ,$$
$$\lambda_j = \max\{\langle H\mathbf{x}, \mathbf{x}\rangle : \|\mathbf{x}\| = 1 \ , \ \mathbf{x} \perp \{\mathbf{x}_1, \mathbf{x}_2, \ldots , \mathbf{x}_{j-1}\}\} = \langle H\mathbf{x}_j, \mathbf{x}_j\rangle \ , \ j \in \overline{2,n} \ .$$

2.12. A Hermitian matrix $H \in \mathbb{C}^{n\times n}$ is *positive semi–definite* (*PSD* for short) if $\langle H\mathbf{x}, \mathbf{x}\rangle \geq 0$ for all $\mathbf{x}$ or, equivalently if its eigenvalues are nonnegative. Similarly, H is called *positive definite* (*PD* for short), if $\langle H\mathbf{x}, \mathbf{x}\rangle > 0$ for all $\mathbf{x} \neq \mathbf{0}$, or equivalently if its eigenvalues are positive. The set of $n \times n$ PSD [PD] matrices is denoted by PSD_n [PD_n].

[1]The spectrum of A is often denoted elsewhere by $\sigma(A)$, a symbol reserved here for the *singular values* of A.

2.13. Let $A \in \mathbb{C}^{m\times n}$, let the eigenvalues $\lambda_j(AA^*)$ of AA^* (which is PSD) be ordered by

$$\lambda_1(AA^*) \geq \lambda_2(AA^*) \geq \cdots \geq \lambda_r(AA^*) > \lambda_{r+1}(AA^*) = \cdots = \lambda_n(AA^*) = 0\,. \tag{29}$$

The *singular values* of A, denoted by $\sigma_j(A)$ or σ_j, $j \in \overline{1,r}$, are defined as

$$\sigma_j(A) = +\sqrt{\lambda_j(AA^*)}\,, \quad j \in \overline{1,r}\,, \tag{30a}$$

$$\text{or equivalently,} \quad \sigma_j(A) = +\sqrt{\lambda_j(A^*A)}\,, \quad j \in \overline{1,r}\,, \tag{30b}$$

and are ordered, by (29),

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0\,. \tag{31}$$

The *set of singular values* of A is denoted by $\sigma(A)$.

2.14. Let A and $\{\sigma_j\}$ be as above, let $\{\mathbf{u}_i : i \in \overline{1,m}\}$ be an o.n. basis of $\mathbb{C}^m$ made of eigenvectors of AA^*,

$$AA^*\mathbf{u}_i = \sigma_i^2\mathbf{u}_i\,,\ i \in \overline{1,r}\,, \tag{32a}$$

$$AA^*\mathbf{u}_i = \mathbf{0}\,,\ i \in \overline{r+1,m}\,, \tag{32b}$$

let

$$\mathbf{v}_j = \frac{1}{\sigma_j}A^*\mathbf{u}_j\,,\ j \in \overline{1,r}\,, \tag{33}$$

and let $\{\mathbf{v}_j : j \in \overline{r+1,n}\}$ be an o.n. set of vectors, orthogonal to $\{\mathbf{v}_j : j \in \overline{1,r}\}$. Then the set $\{\mathbf{v}_j : j \in \overline{1,n}\}$ is an o.n. basis of $\mathbb{C}^n$ consisting of eigenvectors of A^*A,

$$A^*A\mathbf{v}_j = \sigma_i^2\mathbf{v}_j\,,\ j \in \overline{1,r}\,, \tag{34a}$$

$$A^*A\mathbf{v}_j = \mathbf{0}\,,\ j \in \overline{r+1,n}\,. \tag{34b}$$

Conversely, starting from an o.n. basis $\{\mathbf{v}_j : j \in \overline{1,n}\}$ of $\mathbb{C}^n$ satisfying (34a)–(34b), we construct an o.n. basis of $\mathbb{C}^m$ with (32a)–(32b) by

$$\mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_j\,,\ i \in \overline{1,r}\,, \tag{35}$$

and completing to an o.n. set. See Theorem 6.1.

2.15. Let A, $\{\mathbf{u}_i : i \in \overline{1,m}\}$ and $\{\mathbf{v}_j : j \in \overline{1,n}\}$ be as above. Then (35) can be written as

$$A\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_r & \vdots & \mathbf{v}_{r+1} & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_r & \vdots & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \sigma_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O\end{bmatrix}$$

or

$$AV = U\Sigma\,,\ \text{where } \Sigma = \begin{bmatrix}\sigma_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \sigma_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O\end{bmatrix}, \tag{36}$$

and $U = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix}$, $V = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$. Since V is unitary,

$$A = U \Sigma V^* \ , \tag{37}$$

called a *singular value decomposition* (abbreviated *SVD*) of A. See Theorem 6.2.

Exercises and examples.

EX. 12. Let L, M be subspaces of $\mathbb{C}^n$, with $\dim L \geq (k+1)$, $\dim M \leq k$. Then $L \cap M^\perp \neq \{\mathbf{0}\}$.

PROOF. Otherwise $L + M^\perp$ is a direct sum with dimension $= \dim L + \dim M^\perp \geq (k+1) + (n-k) > n$. □

EX. 13. (*The QR factorization*). Let the o.n. set $\{\mathbf{q}_1, \dots, \mathbf{q}_r\}$ be obtained from the set of vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ by the GSO process described in Ex. 7, and let $\widetilde{Q} = \begin{bmatrix}\mathbf{q}_1, \dots, \mathbf{q}_r\end{bmatrix} \in \mathbb{C}^{m\times r}$ and $A = \begin{bmatrix}\mathbf{a}_1, \dots, \mathbf{a}_n\end{bmatrix} \in \mathbb{C}^{m\times n}$ be the corresponding matrices. Then

$$A = \widetilde{Q}\widetilde{R} \tag{38}$$

where the columns of $\widetilde{Q}$ are an o.n. basis of $R(A)$, and $\widetilde{R} \in \mathbb{C}_r^{r\times n}$ is an upper triangular matrix. If $r < m$ it is possible to complete $\widetilde{Q}$ to a unitary matrix $Q = \begin{bmatrix}\widetilde{Q} & Z\end{bmatrix}$, where the columns of Z are an o.n. basis of $N(A^*)$. Then (38) can be written as

$$A = QR \tag{39}$$

where $R = \begin{bmatrix}\widetilde{R} \\ O\end{bmatrix}$ is upper triangular. The expression (39) is called a *QR factorization* of A. By analogy, we call (38) a $\widetilde{Q}\widetilde{R}$ *factorization* of A.

EX. 14. Let U and V be finite–dimensional vector spaces over a field $\mathbb{F}$, and let $T \in \mathcal{L}(U, V)$. Then the null space $N(T)$ and range $R(T)$ are subspaces of U and V respectively.

PROOF. L is a subspace of U if and only if

$$\mathbf{x}, \mathbf{y} \in L \ , \ \alpha \in \mathbb{F} \implies \alpha\mathbf{x} + \mathbf{y} \in L \ .$$

If $\mathbf{x}, \mathbf{y} \in N(T)$ then $T(\mathbf{x} + \alpha\mathbf{y}) = T\mathbf{x} + \alpha T\mathbf{y} = \mathbf{0}$ for all $\alpha \in \mathbb{F}$, proving that $N(T)$ is a subspace of U. The proof that $R(T)$ is a subspace is similar. □

EX. 15. Let P_n be the set of polynomials with real coefficients, of degree $\leq n$,

$$P_n = \{\mathbf{p} : \mathbf{p}(x) = p_0 + p_1 x + \cdots + p_n x^n \ , \ p_i \in \mathbb{R}\} \ . \tag{40}$$

The name x of the variable in (40) is immaterial.

(a) Show that P_n is a vector space with the operations

$$\mathbf{p} + \mathbf{q} = \sum_{i=0}^{n} p_i x^i + \sum_{i=0}^{n} q_i x^i = \sum_{i=0}^{n} (p_i + q_i) x^i \ , \ \alpha\,\mathbf{p} = \sum_{i=0}^{n} (\alpha\, p_i)\, x^i$$

and the dimension of P_n is $n+1$.

(b) The set of monomials $\mathcal{U}_n = \{1, x, x^2, \cdots, x^n\}$ is a basis of P_n. Let T be the differentiation operator, mapping a function $f(x)$ into its derivative $f'(x)$. Show that $T \in \mathcal{L}(P_n, P_{n-1})$. What are the range and null space of T? Find the representation of T w.r.t. the bases $\{\mathcal{U}_n, \mathcal{U}_{n-1}\}$.

(c) Let S be the integration operator, mapping a function $f(x)$ into its integral $\int f(x)dx$ with zero constant of integration. Show that $S \in \mathcal{L}(P_{n-1}, P_n)$. What are the range and null space of S? Find the representation of S w.r.t. $\{\mathcal{U}_{n-1}, \mathcal{U}_n\}$.

(d) Let $T_{\mathcal{U}_n, \mathcal{U}_{n-1}}$ and $S_{\mathcal{U}_{n-1}, \mathcal{U}_n}$ be the matrix representations in parts (b) and (c). What are the matrix products $T_{\{\mathcal{U}_n, \mathcal{U}_{n-1}\}}\, S_{\{\mathcal{U}_{n-1}, \mathcal{U}_n\}}$ and $S_{\{\mathcal{U}_{n-1}, \mathcal{U}_n\}}\, T_{\{\mathcal{U}_n, \mathcal{U}_{n-1}\}}$? Interpret these results in view of the fact that integration and differentiation are *inverse operations*.

EX. 16. Let $\mathbb{C}^m$ and $\mathbb{C}^n$ have o.n. bases $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ and $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, respectively, and let $\langle \cdot, \cdot \rangle$ be an inner product on $\mathbb{C}^m$. Then for any $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$:

(a) The matrix representation $A_{\{\mathcal{U},\mathcal{V}\}} = \begin{bmatrix}a_{ij}\end{bmatrix}$ is given by $a_{ij} = \langle A\mathbf{v}_j, \mathbf{u}_i \rangle$, $\forall\ i, j$.

(b) The adjoint A^* is represented by the matrix $A^*_{\{\mathcal{V},\mathcal{U}\}} = \begin{bmatrix}b_{k\ell}\end{bmatrix}$ where $b_{k\ell} = \overline{a_{\ell k}}$, i.e., the matrix $A^*_{\{\mathcal{V},\mathcal{U}\}}$ is the conjugate transpose of $A_{\{\mathcal{U},\mathcal{V}\}}$.

Ex. 17. *The simplest matrix representation.* Let $O \neq A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$. Then there exist o.n. bases $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively such that

$$A_{\{\mathcal{U},\mathcal{V}\}} = \operatorname{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0) \in \mathbb{R}^{n\times m} , \tag{41}$$

a diagonal matrix, whose nonzero diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A.

Ex. 18. Let $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be two bases of $\mathbb{C}^n$. Show that there is a unique $n \times n$ matrix $S = [s_{ij}]$ such that

$$\mathbf{w}_j = \sum_{i=1}^{n} s_{ij}\, \mathbf{v}_i \ , \ j \in \overline{1,n} \ , \tag{42}$$

and S is nonsingular. Using the rules of matrix multiplication we rewrite (42) as

$$[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n] = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & & \vdots \\ s_{n1} & \cdots & s_{nn} \end{bmatrix} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]S \ , \tag{43}$$

i.e.

$$[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n]S^{-1} \ . \tag{44}$$

Ex. 19. *Similar matrices.* We recall that two square matrices A, B are called *similar* if

$$B = S^{-1}AS \tag{45}$$

for some nonsingular matrix S. If S in (45) is *unitary* [*orthogonal*] then A, B are called called called *unitarily similar* [*orthogonally similar*].

Show that two $n \times n$ complex matrices are similar if and only if each is a matrix representation of the same linear transformation relative to a basis of $\mathbb{C}^n$.

Proof. *If.* Let $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be two bases of $\mathbb{C}^n$ and let $A_{\{\mathcal{V}\}}$ and $A_{\{\mathcal{W}\}}$ be the corresponding matrix representations of a given linear transformation $A : \mathbb{C}^n \to \mathbb{C}^n$. The bases $\mathcal{V}$ and $\mathcal{W}$ determine a (unique) nonsingular matrix $S = [s_{ij}]$ satisfying (42). Rewriting (19) as

$$A[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]A_{\{\mathcal{V}\}} \ . \tag{46}$$

we conclude, by substituting (44) in (46), that

$$A[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n] = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n]S^{-1}A_{\{\mathcal{V}\}}S \ ,$$

and by the uniqueness of the matrix representation,

$$A_{\{\mathcal{W}\}} = S^{-1}A_{\{\mathcal{V}\}}S \ .$$

Only if. Similarly proved. □

Ex. 20. *Schur triangularization.* Any $A \in \mathbb{C}^{n\times n}$ is unitarily similar to a triangular matrix.
(For proof see, e.g., Marcus and Minc [**530**, p. 67]).

Ex. 21. *Perron's approximate diagonalization.* Let $A \in \mathbb{C}^{n\times n}$. Then for any $\epsilon > 0$ there is a nonsingular matrix S such that $S^{-1}AS$ is a triangular matrix

$$S^{-1}AS = \begin{bmatrix} \lambda_1 & b_{12} & \cdots & \cdots & b_{1n} \\ 0 & \lambda_2 & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \lambda_n \end{bmatrix}$$

with the off–diagonal elements satisfying

$$\sum_{ij} |b_{ij}| \leq \epsilon \quad \text{(Bellman [55, p. 205])} \ .$$

Ex. 22. A matrix in $\mathbb{C}^{n\times n}$ is:

(a) normal if and only if it is unitarily similar to a diagonal matrix,

(b) Hermitian if and only if it is unitarily similar to a real diagonal matrix.

Ex. 23. For any $n \geq 2$ there is an $n \times n$ real matrix which is not similar to a triangular matrix in $\mathbb{R}^{n\times n}$.

Hint. The diagonal elements of a triangular matrix are its eigenvalues.

Ex. 24. Denote the transformation of bases (42) by $\mathcal{W} = \mathcal{V}S$. Let $\{\mathcal{U}, \mathcal{V}\}$ be bases of $\{\mathbb{C}^m, \mathbb{C}^n\}$, respectively, and let $\{\widetilde{\mathcal{U}}, \widetilde{\mathcal{V}}\}$ be another pair of bases, obtained by

$$\widetilde{\mathcal{U}} = \mathcal{U}S\ ,\ \widetilde{\mathcal{V}} = \mathcal{V}T\ ,$$

where S and T are $m \times m$ and $n \times n$ matrices, respectively. Show that for any $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$, the representations $A_{\{\mathcal{U},\mathcal{V}\}}$ and $A_{\{\widetilde{\mathcal{U}},\widetilde{\mathcal{V}}\}}$ are related by

$$A_{\{\widetilde{\mathcal{U}},\widetilde{\mathcal{V}}\}} = S^{-1} A_{\{\mathcal{U},\mathcal{V}\}} T\ . \tag{47}$$

Proof. Similar to thr proof of Ex. 19. □

Ex. 25. *Equivalent matrices.* Two matrices A, B in $\mathbb{C}^{m\times n}$ are called *equivalent* if there are nonsingular matrices $S \in \mathbb{C}^{m\times m}$ and $T \in \mathbb{C}^{n\times n}$ such that

$$B = S^{-1}AT\ . \tag{48}$$

If S and T in (48) are unitary matrices, then A, B are called *unitarily equivalent.*

It follows from Ex. 24 that two matrices in $\mathbb{C}^{m\times n}$ are equivalent if, and only if, each is a matrix representation of the same linear transformation relative to a pair of bases of $\mathbb{C}^m$ and $\mathbb{C}^n$.

Ex. 26. Let $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ and $B \in \mathcal{L}(\mathbb{C}^p, \mathbb{C}^n)$, and let $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ be bases of $\mathbb{C}^m, \mathbb{C}^n$ and $\mathbb{C}^p$, respectively. The *product* (or *composition*) of A and B, denoted by AB, is the transformation $\mathbb{C}^p \to \mathbb{C}^m$ defined by

$$(AB)\mathbf{w} = A(B\mathbf{w}) \quad \text{for all } \mathbf{w} \in \mathbb{C}^p\ .$$

(a) The transformation AB is linear, i.e., $(AB) \in \mathcal{L}(\mathbb{C}^p, \mathbb{C}^m)$.

(b) The matrix representation of AB relative to $\{\mathcal{U}, \mathcal{W}\}$ is

$$(AB)_{\{\mathcal{U},\mathcal{W}\}} = A_{\{\mathcal{U},\mathcal{V}\}}B_{\{\mathcal{V},\mathcal{W}\}}\ ,$$

the (matrix) product of the corresponding matrix representations of A and B.

Ex. 27. The matrix representation of the identity transformation I in $\mathbb{C}^n$, relative to any basis, is the $n \times n$ identity matrix I.

Ex. 28. For any invertible $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$ and any two bases $\{\mathcal{U}, \mathcal{V}\}$ of $\mathbb{C}^n$, the matrix representation of A^{-1} relative to $\{\mathcal{V}, \mathcal{U}\}$ is the inverse of the matrix $A_{\{\mathcal{U},\mathcal{V}\}}$,

$$\left(A^{-1}\right)_{\{\mathcal{V},\mathcal{U}\}} = \left(A_{\{\mathcal{U},\mathcal{V}\}}\right)^{-1}$$

Proof. Follows from Exs. 26–27. □

Ex. 29. The real matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has the complex eigenvalue $\lambda = i$, with geometric multiplicity $= 2$, i.e., every nonzero $\mathbf{x} \in \mathbb{R}^2$ is an eigenvector.

Proof.

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \implies \lambda^2 = -1\ ,$$

unless $x_1 = x_2 = 0$. □

Ex. 30. Let $A \in \mathcal{L}(\mathbb{C}^m, \mathbb{C}^n)$. A property shared by all matrix representations $A_{\{\mathcal{U},\mathcal{V}\}}$ of A, as $\mathcal{U}$ and $\mathcal{V}$ range over all bases of $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, is an *intrinsic property* of the linear transformation A. Example: If A, B are similar matrices, they have the same determinant. The determinant is thus intrinsic to the linear transformation represented by A and B.

Given a matrix $A = (a_{ij}) \in \mathbb{C}^{m\times n}$, which of the following items are intrinsic properties of a linear transformation represented by A?

(a) if $m = n$

(a1) the eigenvalues of A (a2) their geometric multiplicities (a3) the eigenvectors of A

(b) if m, n are not necessarily equal,

(b1) the rank of A (b2) the null–space of A (b3) $\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2$

Ex. 31. Let $\widetilde{\mathcal{U}}_n = \{\widetilde{p}_1, \dots, \widetilde{p}_n\}$ be the set of partial sums of monomials

$$\widetilde{p}_k(x) = \sum_{i=0}^{k} x^i \ , \ k \in \overline{1,n} \ .$$

(a) Show that $\widetilde{\mathcal{U}}_n$ is a basis of P_n, and determine the matrix A, such that $\widetilde{\mathcal{U}}_n = A\mathcal{U}_n$, where $\mathcal{U}_n$ is the basis of monomials, see Ex. 15.
(b) Calculate the representations of the differentiation operator (Ex. 15(b)) w.r.t. to the bases $\{\widetilde{\mathcal{U}}_n, \widetilde{\mathcal{U}}_{n-1}\}$, and verify (47).
(c) Same for the integration operator of Ex. 15(c).

Ex. 32. Let L and M be complementary subspaces of $\mathbb{C}^n$. Show that the projector $P_{L,M}$, which carries $\mathbf{x} \in \mathbb{C}^n$ into its projection on L along M, is a linear transformation (from $\mathbb{C}^n$ to L).

Ex. 33. Let L and M be complementary subspaces of $\mathbb{C}^n$, let $\mathbf{x} \in \mathbb{C}^n$, and let $\mathbf{y}$ be the projection of $\mathbf{x}$ on L along M. What is the unique expression for $\mathbf{x}$ as the sum of a vector in L and a vector in M? What, therefore, is $P_{L,M}\,\mathbf{y} = P^2_{L,M}\,\mathbf{x}$, the projection of $\mathbf{y}$ on L along M? Show, therefore, that the transformation $P_{L,M}$ is idempotent.

Ex. 34. *Matrix norms.* Show that the functions

$$\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = (\operatorname{trace} A^*A)^{1/2} \tag{49}$$

and

$$\max\{|a_{ij}| : i \in \overline{1,m}\ ,\ j \in \overline{1,n}\} \tag{50}$$

are matrix norms. The norm (49) is called the *Frobenius norm*, and denoted $\|A\|_F$. Which of these norms is multiplicative?

Ex. 35. *Consistent norms.* A vector norm $\|\ \|$ and a matrix norm $\|\ \|$ are called *consistent* if for any vector $\mathbf{x}$ and matrix A such that $A\mathbf{x}$ is defined,

$$\|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\| \ . \tag{51}$$

Given a vector norm $\|\ \|_*$ show that

$$\|A\|_* = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*} \tag{52}$$

is a multiplicative matrix norm consistent with $\|\mathbf{x}\|_*$, and that any other matrix norm $\|\ \|$ consistent with $\|\mathbf{x}\|_*$, satisfies

$$\|A\| \ge \|A\|_* \ , \text{ for all } A \ . \tag{53}$$

The norm $\|A\|_*$ defined by (52), is called the *matrix norm corresponding to the vector norm* $\|\mathbf{x}\|_*$, or the *bound* of A with respect to $K = \{\mathbf{x} : \|\mathbf{x}\|_* \le 1\}$; see, e.g. Householder [**427**, Section 2.2] and Ex. 3.66 below.

Ex. 36. Show that (52) is the same as

$$\|A\|_* = \max_{\|\mathbf{x}\|_*\le 1} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*} = \max_{\|\mathbf{x}\|_*=1} \|A\mathbf{x}\|_* \ . \tag{54}$$

Ex. 37. Given a multiplicative matrix norm, find a vector norm consistent with it.

EX. 38. *Corresponding norms.*

vector norm on $\mathbb{C}^n$	corresponding matrix norm on $\mathbb{C}^{m\times n}$	
(14) $\|\mathbf{x}\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$	$\|A\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$	(55)
(14.1) $\|\mathbf{x}\|_1 = \sum_{j=1}^n |x_j|$	$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^m |a_{ij}|$	(55.1)
(14.∞) $\|\mathbf{x}\|_\infty = \max_{1\le j\le n} |x_j|$	$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}|$	(55.∞)
(14.2) $\|\mathbf{x}\|_2 = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2}$	$\|A\|_2 = \max\{\sqrt{\lambda} :\ \lambda \text{ an eigenvalue of } A^*A\}$	(55.2)

Note that (55.2) is different from the Frobenius norm (49), which is the Euclidean norm of the mn–dimensional vector obtained by listing all components of A. The norm $\|\ \|_2$ given by (55.2) is called the *spectral norm.*

PROOF. (55.1) follows from (54) since for any $\mathbf{x} \in \mathbb{C}^n$

$$\begin{aligned}\|A\mathbf{x}\|_1 &= \sum_{i=1}^m \left|\sum_{j=1}^n a_{ij}x_j\right| \le \sum_{i=1}^m\sum_{j=1}^n |a_{ij}||x_j| \\ &\le \sum_{j=1}^n |x_j| \sum_{i=1}^m |a_{ij}| \\ &\le \left(\max_{1\le j\le n}\sum_{i=1}^m |a_{ij}|\right)(\|\mathbf{x}\|_1)\end{aligned}$$

with equality if $\mathbf{x}$ is the kth–unit vector, where k is any j for which the maximum in (55) is attained

$$\sum_{i=1}^m |a_{ik}| = \max_{1\le j\le n}\sum_{i=1}^m |a_{ij}| \ .$$

(55.∞) is similarly proved, and (55.2) is left as exercise. □

EX. 39. For any matrix norm $\|\ \|$ on $\mathbb{C}^{m\times n}$, consistent with some vector norm, the norm of the unit matrix satisfies

$$\|I_n\| \ge 1 \ .$$

In particular, if $\|\ \|_*$ is a matrix norm, computed by (52) from a corresponding vector norm, then

$$\|I_n\|_* = 1 \ . \tag{56}$$

EX. 40. A matrix norm $\|\ \|$ on $\mathbb{C}^{m\times n}$ is called *unitarily invariant* if for any two unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$

$$\|UAV\| = \|A\| \quad \text{for all } A \in \mathbb{C}^{m\times n} \ .$$

Show that the matrix norms (49) and (55.2) are unitarily invariant.

EX. 41. *Spectral radius.* The *spectral radius* $\rho(A)$ of a square matrix $A \in \mathbb{C}^{n\times n}$ is the maximal value among the n moduli of the eigenvalues of A,

$$\rho(A) = \max\{|\lambda| :\ \lambda \in \lambda(A)\} \ . \tag{57}$$

Let $\|\ \|$ be any multiplicative norm on $\mathbb{C}^{n\times n}$. Then for any $A \in \mathbb{C}^{n\times n}$,

$$\rho(A) \le \|A\| \ . \tag{58}$$

PROOF. Let $\|\ \|$ denote both a given multiplicative matrix norm, and a vector norm consistent with it. Then $A\mathbf{x} = \lambda\mathbf{x}$ implies $|\lambda|\|\mathbf{x}\| = \|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\|$. □

Ex. 42. For any $A \in \mathbb{C}^{n\times n}$ and any $\epsilon > 0$, there exists a multiplicative matrix norm $\| \;\|$ such that

$$\|A\| \leq \rho(A) + \epsilon \quad \text{(Householder [\textbf{427}, p. 46])} .$$

Ex. 43. If A is a square matrix,

$$\rho(A^k) = \rho^k(A) \ , \ k = 0, 1, \dots \tag{59}$$

Ex. 44. For any $A \in \mathbb{C}^{m\times n}$, the spectral norm $\| \;\|_2$ of (55.2) equals

$$\|A\|_2 = \rho^{1/2}(A^*A) = \rho^{1/2}(AA^*) \ . \tag{60}$$

In particular, if A is Hermitian then

$$\|A\|_2 = \rho(A) \ . \tag{61}$$

In general the spectral norm $\|A\|_2$ and the spectral radius $\rho(A)$ may be quite apart; see, e.g., Noble [**613**, p. 430].

Ex. 45. *Convergent matrices.* A square matrix A is called *convergent* if

$$A^k \to O \text{ as } k \to \infty \ . \tag{62}$$

Show that $A \in \mathbb{C}^{n\times n}$ is convergent if and only if

$$\rho(A) < 1 \ . \tag{63}$$

PROOF. If: From (63) and Ex. 42 it follows that there exists a multiplicative matrix norm $\| \;\|$ such that $\|A\| < 1$. Then

$$\|A^k\| \leq \|A\|^k \to 0 \text{ as } k \to \infty \ ,$$

proving (62).
Only if: If $\rho(A) \geq 1$, then by (59), so is $\rho(A^k)$ for $k = 0, 1, \dots$, contradicting (62). □

Ex. 46. A square matrix A is convergent if and only if the sequence of partial sums

$$S_k = I + A + A^2 + \cdots + A^k = \sum_{j=0}^{k} A^j$$

converges, in which case it converges to $(I - A)^{-1}$, i.e.,

$$(I - A)^{-1} = I + A + A^2 + \cdots = \sum_{j=0}^{\infty} A^j \quad \text{(Householder [\textbf{427}, p. 54])} \ . \tag{64}$$

Ex. 47. Let A be convergent. Then

$$(I + A)^{-1} = I - A + A^2 - \cdots = \sum_{j=0}^{\infty} (-1)^j A^j \ . \tag{65}$$

Ex. 48. *Stein's Theorem.* A square matrix is convergent if and only if there exists a positive definite matrix H such that $H - A^*HA$ is also positive definite (Stein [**783**], Taussky [**808**]).

3. Elementary operations and permutations

3.1. *Elementary operations.* The following operations on a matrix,
(1) multiplying row i by a nonzero scalar α, denoted by $E^i(\alpha)$,
(2) adding β times row j to row i, denoted by $E^{ij}(\beta)$ (here β is any scalar), and
(3) interchanging rows i and j, denoted by E^{ij}, (here $i \neq j$),
are called *elementary row operations* of types 1,2 and 3 respectively[2].

Applying an elementary row operation to the identity matrix I_m results in *elementary row matrix* of the same type. We denote these elementary matrices also by $E^i(\alpha)$, $E^{ij}(\beta)$, and E^{ij}. Elementary row

[2]Only operations of types 1,2 are necessary, see Ex. 49(b). Type 3 operations are introduced for convenience, because of their frequent use.

matrices of types 1,2 have only one row that is different from the corresponding row of the identity matrix of the same order. Examples for $m = 4$,

$$E^2(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} , \quad E^{42}(\beta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \beta & 0 & 1 \end{bmatrix} , \quad E^{13} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Elementary column operations, and the corresponding elementary matrices, are defined analogously.

Performing an elementary row [column] operation is the same as multiplying on the left [right] by the corresponding elementary matrix. For example, $E^{25}(-3)A$ is the matrix obtained from A by subtracting $3 \times$ row 5 from row 2.

3.2. *Permutations.* Given a positive integer n, a *permutation* of order n is a rearrangement of $\{1, 2, \dots, n\}$, i.e. a mapping: $\overline{1, n} \longrightarrow \overline{1, n}$. The set of such permutations is denoted by S_n. It contains:

(a) the identity permutation $\pi_0\{1, 2, \dots, n\} = \{1, 2, \dots, n\}$,

(b) with any two permutations π_1, π_2, their *product* $\pi_1\pi_2$, defined as π_1 applied to $\{\pi_2(1), \pi_2(2), \dots, \pi_2(n)\}$,

(c) with any permutation π, its *inverse*, mapping $\{\pi(1), \pi(2), \dots, \pi(n)\}$ back to $\{1, 2, \dots, n\}$. The inverse of π is denoted by π^{-1}.

Thus S_n is a *group*, called the *symmetric group.*

Given a permutation $\pi \in S_n$, the corresponding *permutation matrix* P_π is defined as $P_\pi = \left[\delta_{\pi(i),j}\right]$, and the correspondence $\pi \longleftrightarrow P_\pi$ is one–to–one. For example,

$$\pi\{1, 2, 3\} = \{2, 3, 1\} \quad \longleftrightarrow \quad P_\pi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Products of permutations correspond to matrix products:

$$P_{\pi_1\pi_2} = P_{\pi_1}P_{\pi_2} , \quad \forall \pi_1, \pi_2 \in S_n .$$

A *transposition* is a permutation that switches only a pair of elements, for example, $\pi\{1, 2, 3, 4\} = \{1, 4, 3, 2\}$. Every permutation $\pi \in S_n$ is a product of transpositions, generally in more than one way. However, the number of transpositions in such a product is always even or odd, depending only on π. Accordingly, a permutation π is called *even* or *odd*, if it is the product of an even or odd number of transpositions, respectively. The *sign* of the permutation π, denoted $\operatorname{sign} \pi$, is defined as

$$\operatorname{sign} \pi = \begin{cases} +1 & \text{if } \pi \text{ is even,} \\ -1 & \text{if } \pi \text{ is odd .} \end{cases}$$

The following table summarizes the situation for permutations of order 3.

permutation π		inverse π^{-1}	product of transpositions	sign π
π_0	$\{1, 2, 3\}$	π_0	$\pi_1\pi_1$, $\pi_2\pi_2$, etc.	$+1$
π_1	$\{1, 3, 2\}$	π_1	π_1	-1
π_2	$\{2, 1, 3\}$	π_2	π_2	-1
π_3	$\{2, 3, 1\}$	π_4	$\pi_1\pi_2$	$+1$
π_4	$\{3, 1, 2\}$	π_3	$\pi_2\pi_1$	$+1$
π_5	$\{3, 2, 1\}$	π_5	π_5	-1

Multiplying a matrix A by a permutation matrix P_π on the left [right] results in a permutation π $[\pi^{-1}]$ of the rows [columns] of A. For example,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{11} & a_{12} \end{bmatrix} , \quad \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} b_{13} & b_{11} & b_{12} \\ b_{23} & b_{21} & b_{22} \end{bmatrix} .$$

Exercises and examples.

Ex. 49. *Elementary operations.*

(a) The elementary matrices are nonsingular, and their inverses are

$$E^i(\alpha)^{-1} = E^i(1/\alpha)\ , \quad E^{ij}(\beta)^{-1} = E^{ij}(-\beta)\ , \quad (E^{ij})^{-1} = E^{ij}\ . \tag{66}$$

(b) Type 3 elementary operations are expressible in terms of the other two types:

$$E^{ij} = E^i(-1)E^{ji}(1)E^{ij}(-1)E^{ji}(1)\ . \tag{67}$$

(c) Conclude from (b) that any permutation matrix is a product of elementary matrices of types 1,2.

Ex. 50. Describe a recursive method for listing all $n!$ permutations in S_n.
Hint: If π is a permutation in S_{n-1}, mapping $\{1, 2, \ldots, n-1\}$ to

$$\{\pi(1), \pi(2), \ldots, \pi(n-1)\}\ , \tag{68}$$

then π gives rise to n permutations in S_n obtained by placing n in the "gaps" $\{\sqcup\pi(1) \sqcup \pi(2) \sqcup \ldots \sqcup \pi(n-1)\sqcup\}$ of (68).

4. The Hermite normal form and related items

4.1. *Hermite normal form.* Let $\mathbb{C}_r^{m\times n}$ $[\mathbb{R}_r^{m\times n}]$ denote the class of $m \times n$ complex [real] matrices of rank r.

Definition 1. (Marcus and Minc [**530**, § 3.6]) A matrix in $\mathbb{C}_r^{m\times n}$ is said to be in *Hermite normal form* (also called *reduced row–echelon form*) if:

(a) the first r rows contain at least one nonzero element; the remaining rows contain only zeros,

(b) there are r integers

$$1 \le c_1 < c_2 < \cdots < c_r \le n\ , \tag{69}$$

such that the first nonzero element in row $i \in \overline{1,r}$, appears in column c_i, and

(c) all other elements in column c_i are zero, $i \in \overline{1,r}$. □

By a suitable permutation of its columns, a matrix $H \in \mathbb{C}_r^{m\times n}$ in Hermite normal form can be brought into the partitioned form

$$R = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} \tag{70}$$

where O denotes a null matrix. Such a permutation of the columns of H can be interpreted as multiplication of H on the right by a suitable permutation matrix P. If P_j denotes the j th–column of P, and $\mathbf{e}_j$ the j th–column of I_n, we have

$$P_j = \mathbf{e}_k \text{ where } k = c_j\ ,\ j \in \overline{1,r}\ ,$$

the remaining columns of P are the remaining unit vectors $\{\mathbf{e}_k : k \ne c_j\ ,\ j \in \overline{1,r}\}$ in any order. In general, there are $(n-r)!$ different pairs $\{P, K\}$, corresponding to all arrangements of the last $n-r$ columns of P.

In particular cases, the partitioned form (70) may be suitably interpreted. If $R \in \mathbb{C}_r^{m\times n}$, then the two right–hand submatrices are absent in case $r = n$, and the two lower submatrices are absent if $r = m$.

4.2. *Gaussian elimination.* A *Gaussian elimination* is a sequence of elementary row operations, that transform a given matrix to a desired form.

The Hermite normal form of a given matrix can be computed by Gaussian elimination. Transpositions of rows (i.e., elementary operations of type 3) are used, if necessary, to bring the nonzero rows to the top. The *pivots* of the elimination are the leading nonzeros in these rows. This is illustrated in Ex. 51 below.

Let $A \in \mathbb{C}^{m\times n}$, and let $E_k, E_{k-1}, \ldots, E_2, E_1$ be elementary row operations, and P a permutation matrix such that

$$E\,A\,P = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}\ , \tag{71}$$

where

$$E = E_k E_{k-1} \cdots E_2 E_1 \; , \tag{72}$$

in which case A is determined to have rank r. Equation (71) can be rewritten as

$$A \; = \; E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1} \; . \tag{73}$$

4.3. *Bases for the range and null space of a matrix* For any $A \in \mathbb{C}^{m\times n}$ we denote by

$$\begin{aligned} R(A) \; &= \; \{\mathbf{y} \in \mathbb{C}^m : \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{C}^n\} \; , \text{ the } \textit{range} \text{ of } A \; , \\ N(A) \; &= \; \{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \mathbf{0}\} \; , \text{ the } \textit{null space} \text{ of } A \; . \end{aligned}$$

A basis for $R(A)$ is useful in a number of applications, such as, for example, in the numerical computation of the Moore–Penrose inverse, and the group inverse to be discussed in Chapter 4.

The need for a basis of $N(A)$ is illustrated by the fact that the general solution of the linear inhomogeneous equation

$$A\mathbf{x} \; = \; \mathbf{b}$$

is the sum of any particular solution $\mathbf{x}_0$ and the general solution of the homogeneous equation

$$A\mathbf{x} \; = \; \mathbf{0} \; .$$

The latter general solution consists of all linear combinations of the elements of any basis for $N(A)$.

A further advantage of the Hermite normal form EA of A (and its column–permuted form EAP) is that from them bases for $R(A), N(A)$, and $R(A^*)$ can be read off directly.

A basis for $R(A)$ consists of the c_1 th, c_2 th, … , c_r th columns of A, where the $\{c_j : j \in \overline{1,r}\}$ are as in Definition 1. To see this, let P_1 denote the submatrix consisting of the first r columns of the permutation matrix P of (71). Then, because of the way in which these r columns of P were chosen,

$$EAP_1 \; = \; \begin{bmatrix} I_r \\ O \end{bmatrix} \; . \tag{74}$$

Now, AP_1 is an $m \times r$ matrix, and is of rank r, since RHS(74) is of rank r. But AP_1 is merely the submatrix of A consisting of the c_1 th, c_2 th, … , c_r th columns.

It follows from (73) that the columns of the $n \times (n-r)$ matrix

$$P \begin{bmatrix} -K \\ I_{n-r} \end{bmatrix} \tag{75}$$

are a basis for $N(A)$. (The reader should verify this.)

Moreover, it is evident that the first r rows of the Hermite normal form EA are linearly independent, and each is some linear combination of the rows of A. Thus, they are a basis for the space spanned by the rows of A. Consequently, if

$$EA = \begin{bmatrix} G \\ O \end{bmatrix} \; , \tag{76}$$

then the columns of the $n \times r$ matrix

$$G^* \; = \; P \begin{bmatrix} I_r \\ K^* \end{bmatrix}$$

are a basis for $R(A^*)$.

4.4. *Full–rank factorization.* A nonzero matrix can be expressed as the product of a matrix of full column rank and a matrix of full row rank. Such factorizations turn out to be a powerful tool in the study of generalized inverses.

LEMMA 1. Let $A \in \mathbb{C}_r^{m\times n}$, $r > 0$. Then there exist matrices $F \in \mathbb{C}_r^{m\times r}$ and $G \in \mathbb{C}_r^{r\times n}$, such that

$$A = FG \,. \tag{77}$$

PROOF. The $\widetilde{Q}\widetilde{R}$ factorization, Ex. 13, is a case in point. Let F be any matrix whose columns are a basis for $R(A)$. Then $F \in \mathbb{C}_r^{m\times r}$. The matrix $G \in \mathbb{C}^{r\times n}$ is then uniquely determined by (77), since every column of A is uniquely representable as a linear combination of the columns of F. Finally, $\operatorname{rank} G = r$, since

$$\operatorname{rank} G \geq \operatorname{rank} FG = r \,.$$

□

The columns of F can, in particular, be chosen as any maximal linearly independent set of columns of A. Also, G could be chosen first as any matrix whose rows are a basis for the space spanned by the rows of A, and then F is uniquely determined by (77).

A factorization (77) with the properties stated in Lemma 1 is called a (*full*) *rank factorization* of A. When A is of full (column or row) rank, the most obvious factorization is a trivial one, one factor being a unit matrix.

A rank factorization of any matrix is easily read off from its Hermite normal form. Indeed, it was pointed out in § 4.3 above that the first r rows of the Hermite normal form EA (i.e., the rows of the matrix G of (76)) form a basis for the space spanned by the rows of A. Thus, this G can serve also as the matrix G of (77). Consequently, (77) holds for some F. As in § 4.3, let P_1 denote the submatrix of P consisting of the first r columns. Because of the way in which these r columns were constructed,

$$GP_1 = I_r \,.$$

Thus, multiplying (77) on the right by P_1 gives

$$F = AP_1 \,,$$

and so (77) becomes

$$A = (AP_1)G \,, \tag{78}$$

where P_1 and G are as in § 4.3. (Indeed it was already noted there that the columns of AP_1 are a basis for $R(A)$.)

Exercises and examples.

EX. 51. *Transforming a matrix into Hermite normal form.* Let $A \in \mathbb{C}^{m\times n}$, and $T_0 = [A \,\vdots\, I_m]$. A matrix E transforming A into a Hermite normal form EA can be found by Gaussian elimination on T_0, where, after the elimination is completed,

$$ET_0 = [EA \,\vdots\, E] \,,$$

E being recorded as the right–hand $m \times m$ submatrix of ET_0. We illustrate this procedure for the matrix

$$A = \begin{bmatrix} 0 & 2i & i & 0 & 4+2i & 1 \\ 0 & 0 & 0 & -3 & -6 & -3-3i \\ 0 & 2 & 1 & 1 & 4-4i & 1 \end{bmatrix} , \tag{79}$$

marking the pivots by square brackets.

$$T_0 = \left[\begin{array}{cccccc:ccc} 0 & [2i] & i & 0 & 4+2i & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -3 & -6 & -3-3i & 0 & 1 & 0 \\ 0 & 2 & 1 & 1 & 4-4i & 1 & 0 & 0 & 1 \end{array}\right],$$

$$T_1 = E^{31}(-2)E^{1}(1/2i)T_0 = \left[\begin{array}{cccccc:ccc} 0 & 1 & \frac{1}{2} & 0 & 1-2i & -\frac{1}{2}i & -\frac{1}{2}i & 0 & 0 \\ 0 & 0 & 0 & [-3] & -6 & -3-3i & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1+i & i & 0 & 1 \end{array}\right],$$

$$T_2 = E^{32}(-1)E^{2}(-1/3)T_1 = \left[\begin{array}{cccccc:ccc} 0 & 1 & \frac{1}{2} & 0 & 1-2i & -\frac{1}{2}i & -\frac{1}{2}i & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1+i & 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & i & \frac{1}{3} & 1 \end{array}\right].$$

From $T_2 = [EA \vdots E]$ we read the Hermite normal form

$$EA = \begin{bmatrix} 0 & 1 & \frac{1}{2} & 0 & 1-2i & -\frac{1}{2}i \\ 0 & 0 & 0 & 1 & 2 & 1+i \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{80}$$

where

$$E = E^{32}(-1)E^{2}(-1/3)E^{31}(-2)E^{1}(1/2i) = \begin{bmatrix} -\frac{1}{2}i & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ i & \frac{1}{3} & 1 \end{bmatrix}, \quad \text{and } r = \operatorname{rank} A = 2\ . \tag{81}$$

Ex. 52. (Ex. 51 continued). To bring the Hermite normal form (80) to the standard form (71), use

$$P = \left[\begin{array}{cc:cccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right], \quad \text{to get} \quad E\,A\,P = \begin{bmatrix} I_2 & K \\ O & O \end{bmatrix} \text{ with } K = \begin{bmatrix} 0 & \frac{1}{2} & 1-2i & -\frac{1}{2}i \\ 0 & 0 & 2 & 1+i \end{bmatrix}. \tag{82}$$

In this example there are 4! different pairs $\{P, K\}$, corresponding to all arrangements of the last 4 columns of P.

Ex. 53. (Ex. 51 continued). Consider the matrix A of (79), and its Hermite normal form (80) where the two unit vectors of $\mathbb{C}^2$ appear in the *second* and *fourth* columns. Therefore, the second and fourth columns of A form a basis for $R(A)$.

Using (75) with P and K selected by (82), we find that the columns of the following matrix form a basis for $N(A)$:

$$P\begin{bmatrix}-K\\ I_{n-r}\end{bmatrix} = \begin{bmatrix}0&0&\vdots&1&0&0&0\\ 1&0&\vdots&0&0&0&0\\ 0&0&\vdots&0&1&0&0\\ 0&1&\vdots&0&0&0&0\\ 0&0&\vdots&0&0&1&0\\ 0&0&\vdots&0&0&0&1\end{bmatrix}\begin{bmatrix}0&-\frac{1}{2}&-1+2i&\frac{1}{2}i\\ 0&0&-2&-1-i\\ \cdots&\cdots&\cdots\cdots&\cdots\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\end{bmatrix}$$

$$= \begin{bmatrix}1&0&0&0\\ 0&-\frac{1}{2}&-1+2i&\frac{1}{2}i\\ 0&1&0&0\\ 0&0&-2&-1-i\\ 0&0&1&0\\ 0&0&0&1\end{bmatrix}.$$

EX. 54. If $A \in \mathbb{C}^{m\times m}$ is nonsingular, then the permutation matrix P in (71) can be taken as the identity (i.e., permutation is unnecessary). Therefore $E = A^{-1}$ and

$$A = E_1^{-1}E_2^{-1}\cdots E_{k-1}^{-1}E_k^{-1}\,. \tag{83}$$

(a) Conclude that A is nonsingular if and only if it is a product of elementary row matrices.
(b) Compute the Hermite normal forms of

$$A = \begin{bmatrix}1&2&3\\ 4&5&6\\ 7&8&8\end{bmatrix} \text{ and } B = \begin{bmatrix}1&2&3\\ 4&5&6\\ 7&8&9\end{bmatrix},$$

and illustrate (83).

EX. 55. Consider the matrix A of (79). Using the results of Exs. 51–52, a rank factorization is computed by (78),

$$A = (AP_1)G = \begin{bmatrix}2i&0\\ 0&-3\\ 2&1\end{bmatrix}\begin{bmatrix}0&1&\frac{1}{2}&0&1-2i&-\frac{1}{2}i\\ 0&0&0&1&2&1+i\end{bmatrix}. \tag{84}$$

EX. 56. Given a nonzero $m \times n$ matrix $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, a full column submatrix $F = [\mathbf{a}_{c_1}, \ldots, \mathbf{a}_{c_r}]$ can be found by Gram–Schmidt orthogonalization, see Ex. 7. Indeed, applying the GSO process to the columns of A gives the integers $\{c_1, \ldots, c_r\}$.

EX. 57. Use the GSO process to compute a $\widetilde{Q}\widetilde{R}$ factorization of the matrix A of (79).

5. Determinants and volume

5.1. *Determinants.* The *determinant* of an $n \times n$ matrix $A = [a_{ij}]$, denoted $\det A$, is customarily defined as

$$\det A = \sum_{\pi\in S_n} \operatorname{sign}\pi \prod_{i=1}^{n} a_{\pi(i),i} \tag{85}$$

see, e.g. Marcus and Minc [**530**, § 2.4]. We use here an alternative definition.

DEFINITION 2. The *determinant* is a function $\det : \mathbb{C}^{n\times n} \to \mathbb{C}$ such that
(a) $\det(E^i(\alpha)) = \alpha$, for all $\alpha \in \mathbb{C}$, $i \in \overline{1,n}$, and
(b) $\det(AB) = \det(A)\det(B)$, for all $A, B \in \mathbb{C}^{n\times n}$.

The reader is referred to Cullen and Gale [**209**] for proof that Definition 2 is equivalent to (85). See also Exs. 58–59 below.

The Binet–Cauchy formula. If $A \in \mathbb{C}^{k\times n}$, $B \in \mathbb{C}^{n\times k}$ then

$$\det(AB) = \sum_{I\in Q_{k,n}} \det A_{I*} \det B_{*I} \ . \tag{86}$$

For proof see, e.g., Gantmacher [**290**, Vol. I, p. 9].

5.2. *Gram matrices.* The *Gram matrix* of $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \subset \mathbb{C}^n$ is the $k \times k$ matrix of inner products (unless otherwise noted, $\langle\cdot,\cdot\rangle$ is the standard inner–product)

$$G(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) = (\langle \mathbf{x}_i, \mathbf{x}_j\rangle) \ . \tag{87}$$

The determinant $\det G(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$ is called the *Gramian* of $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$.

If $X \in \mathbb{C}^{n\times k}$ is the matrix with columns $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, then by the Binet–Cauchy formula,

$$\det G(\mathbf{x}_1, \dots, \mathbf{x}_n) = \det X^*X = \sum_{I\in Q_{k,n}} |\det X_{I*}|^2 \ .$$

5.3. *Volume.* The matrices in this section are real. Given $A \in \mathbb{R}^{m\times n}_r$, we denote

$$\begin{aligned}
\mathcal{I}(A) &= \{I \in Q_{r,m} : \ \operatorname{rank} A(I,*) = r\} \ , \\
\mathcal{J}(A) &= \{J \in Q_{r,n} : \ \operatorname{rank} A(*,J) = r\} \ , \\
\mathcal{N}(A) &= \{(I,J) : \ I \in Q_{r,m}, J \in Q_{r,n}, \ A(I,J) \text{ is nonsingular}\} \ ,
\end{aligned}$$

or $\mathcal{I}$, $\mathcal{J}$, $\mathcal{N}$ if A is understood. $\mathcal{I}$ and $\mathcal{J}$ are the index sets of maximal submatrices of full row rank and full column rank, respectively, and $\mathcal{N}$ is the index set of maximal nonsingular submatrices.

The (*r–dimensional*) *volume* of a matrix $A \in \mathbb{R}^{m\times n}_r$, denoted $\operatorname{vol} A$ or $\operatorname{vol}_r A$, is defined as 0 if $r = 0$, and otherwise

$$\operatorname{vol} A := \sqrt{\sum_{(I,J)\in\mathcal{N}(A)} \det^2 A_{IJ}} \ . \tag{88}$$

see also Ex. 69 below.

Exercises and examples.

Ex. 58. (Properties of determinants).

(a) $\det E^{ij}(\beta) = 1$, for all $\beta \in \mathbb{C}$, $i, j \in \overline{1,n}$, and

(b) $\det E^{ij} = -1$, for all $i, j \in \overline{1,n}$. (*Hint.* Use (67) and Definition 2).

(c) If A is nonsingular, and given as product (83) of elementary matrices, then

$$\det A = \det\left(E_1^{-1}\right) \det\left(E_2^{-1}\right) \cdots \det\left(E_{k-1}^{-1}\right) \det\left(E_k^{-1}\right) \ . \tag{89}$$

(d) Use (89) to compute the determinant of A in Ex. 54(b).

Ex. 59. (The *Cramer rule*). Given a matrix A and a compatible vector $\mathbf{b}$, we denote by $A[j \leftarrow \mathbf{b}]$ the matrix obtained from A by replacing the j th–column by $\mathbf{b}$.

Let $A \in \mathbb{C}^{n\times n}$ be nonsingular. Then for any $\mathbf{b} \in \mathbb{C}^n$, the solution $\mathbf{x} = [x_j]$ of

$$A\mathbf{x} = \mathbf{b} \tag{90}$$

is given by

$$x_j = \frac{\det A[j \leftarrow \mathbf{b}]}{\det A} \ , \ j \in \overline{1,n} \ . \tag{91}$$

Proof. (Robinson [**706**]). Write $A\mathbf{x} = \mathbf{b}$ as

$$AI_n[j \leftarrow \mathbf{x}] = A[j \leftarrow \mathbf{b}] \ , \quad j \in \overline{1,n} \ ,$$

and take determinants

$$\det A \det I_n[j \leftarrow \mathbf{x}] = \det A[j \leftarrow \mathbf{b}] \ . \tag{92}$$

Then (91) follows from (92) since

$$\det I_n[j \leftarrow \mathbf{x}] = x_j \ .$$

□

See an extension of Cramer's rule in Corollary 5.6.

Ex. 60. *The Hadamard inequality.* Let $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n] \in \mathbb{C}^{n \times n}$. Then

$$|\det A| \leq \prod_{i=1}^{n} \|\mathbf{a}_i\|_2 \ , \tag{93}$$

with equality if and only if the set of columns $\{\mathbf{a}_i\}$ is orthogonal or if one of the columns is zero.

Proof. LHS(93) = the volume of the parallelepiped defined by the columns of $A \leq$ the volume of the cube with sides of lengths $\|\mathbf{a}_i\|_2 =$ RHS(93). □

Ex. 61. *The Schur complement* (Schur [**738**]). Let the matrix A be partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \ , \quad \text{with } A_{11} \text{ nonsingular} \ . \tag{94}$$

The *Schur complement* of A_{11} in A, denoted A/A_{11}, is defined by

$$A/A_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12} \ . \tag{95}$$

(a) If A is square, its determinant is

$$\det A = \det A_{11} \det(A/A_{11}) \ . \tag{96}$$

(b) *The quotient property.* If A_{11} is further partitioned as

$$A_{11} = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \ , \quad \text{with } E \text{ nonsingular} \ , \tag{97}$$

then

$$A/A_{11} = (A/E)/(A_{11}/E) \quad \text{(Crabtree and Haynsworth [\textbf{207}])} \ . \tag{98}$$

(c) If A, A_{11} and A_{22} are nonsingular, then

$$A^{-1} = \begin{bmatrix} (A/A_{22})^{-1} & -A_{11}^{-1}A_{12}(A/A_{11})^{-1} \\ -A_{22}^{-1}A_{21}(A/A_{22})^{-1} & (A/A_{11})^{-1} \end{bmatrix} \ . \tag{99}$$

Proof. (a) follows from

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ O & I \end{bmatrix} = \begin{bmatrix} A_{11} & O \\ A_{21} & A/A_{11} \end{bmatrix} \ .$$

(b) is left for the reader, after noting that A_{11}/E is nonsingular (since A_{11} and E are nonsingular, and $\det A_{11} = \det E \det(A_{11}/E)$ by (96)). (c) is verified by multiplying A and RHS(99). □

Ex. 62. The set $S = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ are linearly independent if, and only if, the Gram matrix $G(S)$ is nonsingular.

Proof. If S is linearly dependent, then

$$\sum_{i=1}^{k} \alpha_i \, \mathbf{x}_i = \mathbf{0} \tag{100}$$

has a nonzero solution $(\alpha_1, \dots, \alpha_k)$. Therefore

$$\sum_{i=1}^{k} \alpha_i \langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0 \ , \ j \in \overline{1,k} \tag{101}$$

showing $G(S)$ is singular. Conversely, writing (101) as

$$\langle \sum_{i=1}^{k} \alpha_i \, \mathbf{x}_i, \mathbf{x}_j \rangle = 0 \ , \ j \in \overline{1,k}$$

multiplying by α_j and summing we get

$$\langle \sum_{i=1}^{k} \alpha_i \, \mathbf{x}_i, \sum_{i=1}^{k} \alpha_i \, \mathbf{x}_i \rangle = 0 \ .$$

proving (100). ☐

Ex. 63. For any set of vectors $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subset \mathbb{C}^n$, det $G(S)$ is independent of the order of the vectors. Moreover, det $G(S) \geq 0$ and det $G(S) = 0$ if and only if the set S is linearly dependent.

Ex. 64. Let $A \in \mathbb{C}_r^{m \times n}$ have a rank–factorization $A = CR$, $C \in \mathbb{C}_r^{m \times r}$, $R \in \mathbb{C}_r^{r \times n}$. Then:
(a) $\mathcal{I}(A) = \mathcal{I}(C)$.
(b) $\mathcal{J}(A) = \mathcal{J}(R)$.
(c) $\mathcal{N}(A) = \mathcal{I}(A) \times \mathcal{J}(A)$.

Proof. (a) and (b) are obvious, as is $\mathcal{N}(A) \subset \mathcal{I}(A) \times \mathcal{J}(A)$. The converse

$$\mathcal{N}(A) \supset \mathcal{I}(A) \times \mathcal{J}(A) \ ,$$

follows since every A_{IJ} is the product

$$A_{IJ} = C_{I*} R_{*J} \ . \tag{102}$$

☐

Ex. 65. If the matrix C is of full column–rank r then, by the Binet–Cauchy Theorem,

$$\text{vol}_r^2\,(C) = \det\, C^T C \ , \tag{103}$$

the Gramian of the columns of C. Similarly, if R is of full row–rank r,

$$\text{vol}_r^2\,(R) = \det\, R R^T \ . \tag{104}$$

Ex. 66. Let $A \in \mathbb{R}_r^{m \times n}$, $r > 0$, and let $A = CR$ be any rank factorization. Then

$$\text{vol}_r^2\,(A) = \sum_{I \in \mathcal{I}} \text{vol}_r^2\,(A_{I*}) \ , \tag{105a}$$

$$= \sum_{J \in \mathcal{J}} \text{vol}_r^2\,(A_{*J}) \ , \tag{105b}$$

$$= \text{vol}_r^2\,(C)\, \text{vol}_r^2\,(R) \ . \tag{105c}$$

Proof. Follows from Definition (88), Ex. 64(c) and (102). ☐

Ex. 67. *A generalized Hadamard inequality.* Let $A \in \mathbb{R}_r^{m \times n}$ be partitioned into two matrices $A = (A_1, A_2)$, $A_i \in \mathbb{R}_{r_i}^{m \times n_i}$, $i = 1, 2$, with $r_1 + r_2 = r$. Then

$$\text{vol}_r A \leq \text{vol}_{r_1} A_1 \ \text{vol}_{r_2} A_2 \ , \tag{106}$$

with equality iff the columns of A_1 are orthogonal to those of A_2.

Proof. The full-rank case $n_i = r_i$, $i = 1, 2$, was proved in [**290**, Vol. I, p. 254]. The general case follows since every $m \times r$ submatrix of rank r has r_1 columns from A_1 and r_2 columns from A_2. ☐

A statement of (106) in terms of the principal angles [**3**] between $R(A_1)$ and $R(A_2)$ is given in [**545**].

6. Some multilinear algebra

The setting of multilinear algebra is natural for the matrix volume, allowing simplification of statements and proofs.

Let $V = \mathbb{R}^n$, $U = \mathbb{R}^m$. We use the same letter to denote both a linear transformation in $\mathcal{L}(V, U)$ and its matrix representation with respect to fixed bases in V and U.

Let $\overset{k}{\wedge} V$ be the k th–*exterior space* over V, spanned by *exterior products* $\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k$ of elements $\mathbf{x}_i \in V$, see, e.g., [**527**], [**528**] and [**581**]. For $A \in \mathbb{R}_r^{m\times n}$, $r > 0$ and $k = 1, \cdots, r$, the *k–compound matrix* $C_k(A)$ is an element of $\mathcal{L}(\overset{k}{\wedge} V, \overset{k}{\wedge} U)$, given by

$$A\mathbf{x}_1 \wedge \cdots \wedge A\mathbf{x}_k = C_k(A)\,(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k)\ , \quad \forall\ \mathbf{x}_i \in V\ , \tag{107}$$

see, e.g., [**528**, § 4.2, p. 94]. Then $C_k(A)$ is an $\binom{m}{k} \times \binom{n}{k}$ matrix of rank $\binom{r}{k}$, see Ex. 6.22.

To any r-dimensional subspace $W \subset V$ there corresponds a unique 1-dimensional subspace $\overset{r}{\wedge} W \subset \overset{r}{\wedge} V$, spanned by the exterior product

$$\mathbf{w}^{\wedge} = \mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_r \tag{108}$$

where $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ is any basis of W, e.g. [**581**]. The $\binom{n}{r}$ components of $\mathbf{w}^{\wedge}$ (determined up to a multiplicative constant) are the *Plücker coordinates* of W.

Results relating volumes, compound matrices and full–rank factorizations are collected in the following lemma. The proofs are omitted.

LEMMA 2 (Volume and compounds). Let $r > 0$, $A \in \mathbb{R}_r^{m\times n}$, $C \in \mathbb{R}_r^{m\times r}$ have columns $\mathbf{c}^{(j)}$ and $R \in \mathbb{R}_r^{r\times n}$ have rows $\mathbf{r}_{(i)}$. Then:

$$C_r(R)\,(\mathbf{r}_{(1)} \wedge \cdots \wedge \mathbf{r}_{(r)}) = \mathrm{vol}^2\, R\ , \tag{109a}$$

$$C_r(C^T)\ (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)}) = \mathrm{vol}^2\, C\ . \tag{109b}$$

If $A = CR$ is a full rank factorization of A, then

$$C_r(A) = (\mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)})(\mathbf{r}_{(1)} \wedge \cdots \wedge \mathbf{r}_{(r)}) \tag{109c}$$

is a full rank factorization[3] of $C_r(A)$. Moreover, the volume of A is given by the inner product,

$$\langle \mathbf{c}^{(1)} \wedge \cdots \wedge \mathbf{c}^{(r)}, A\mathbf{r}_{(1)} \wedge \cdots \wedge A\mathbf{r}_{(r)} \rangle = \mathrm{vol}^2\, A\ , \tag{110}$$

and

$$\mathrm{vol}_r^2\, A = \mathrm{vol}_1^2\, C_r(A) = \mathrm{vol}_1^2 C_r(C)\ \mathrm{vol}_1^2\, C_r(R)\ . \tag{111}$$

Exercises and examples.

EX. 68. Consider the 3×3 matrix of rank 2, with a full rank factorization

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = CR = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \tag{112}$$

Then the 2-compound matrix is

$$\begin{aligned} C_2(A) &= \begin{bmatrix} -3 & -6 & -3 \\ -6 & -12 & -6 \\ -3 & -6 & -3 \end{bmatrix} = C_2(C)\, C_2(R) \\ &= \begin{bmatrix} -3 \\ -6 \\ -3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \end{aligned}$$

[3] This is a restatement of (102).

a full–rank factorization. The volume of A is calculated by (111)

$$\text{vol}_2^2\, A = \text{vol}_1^2(C)\ \text{vol}_1^2(R) = (9+36+9)(1+4+1) = 324.$$

EX. 69. For $k = 1, \dots, r$, the k–*volume* of A is defined as the Frobenius norm of the k-th compound matrix $C_k(A)$,

$$\text{vol}_k\, A := \sqrt{\sum_{I\in Q_{k,m},\, J\in Q_{k,n}} |\det A_{IJ}|^2} \tag{113a}$$

or equivalently,

$$\text{vol}_k\, A = \sqrt{\sum_{I\in Q_{k,r}} \left(\prod_{i\in I} \sigma_i^2(A)\right)} \tag{113b}$$

the square root of the k-th symmetric function of $\{\sigma_1^2(A), \cdots, \sigma_r^2(A)\}$. We use the convention

$$\text{vol}_k\, A := 0\ , \quad \text{for } k = 0 \text{ or } k > \text{rank}\, A\ . \tag{113c}$$

EX. 70. Let $S \in \mathbb{R}^{m\times m}$, $A \in \mathbb{R}_m^{m\times n}$. Then

$$\text{vol}_m\,(SA) = |\det S|\,\text{vol}\, A\ . \tag{114}$$

PROOF. If S is singular, then both sides of (114) are zero. Let S be nonsingular. Then $\text{rank}\,(SA) = m$, and

$$\begin{aligned}\text{vol}_m\,(SA) &= \sqrt{\sum_{J\in Q_{m,n}} \det{}^2\,(SA)_{*J}} \\ &= \sqrt{\sum_{J\in Q_{m,n}} \det{}^2 S \det{}^2 A_{*J}} \\ &= |\det S|\,\text{vol}\, A\ .\end{aligned}$$

□

7. The Jordan normal form

Let the matrices $A \in \mathbb{C}^{n\times n}$, $X \in \mathbb{C}_k^{n\times k}$ and the scalar $\lambda \in \mathbb{C}$ satisfy

$$AX = XJ_k(\lambda)\ , \quad \text{where} \quad J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{k\times k}\ , \tag{115}$$

or, writing X by its columns, $X = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k\end{bmatrix}$,

$$A\mathbf{x}_1 = \lambda\, \mathbf{x}_1\ , \tag{116}$$

$$A\mathbf{x}_j = \lambda\, x_j + \mathbf{x}_{j-1}\ ,\ j = 2, \dots, k\ . \tag{117}$$

It follows, for $j \in \overline{1,k}$, that

$$(A - \lambda\, I)^j\, \mathbf{x}_j = \mathbf{0}\ , \quad (A - \lambda\, I)^{j-1}\, \mathbf{x}_j = \mathbf{x}_1 \neq \mathbf{0} \tag{118}$$

where we interpret $(A - \lambda\, I)^0$ as I. The vector $\mathbf{x}_1$ is therefore an *eigenvector* of A corresponding to the *eigenvalue* λ. We call $\mathbf{x}_j$ a λ–*vector* of A of grade j, or, following Wilkinson [**881**, p. 43], a *principal vector*[4] of A of *grade* j associated with the eigenvalue λ. Evidently principal vectors are a generalization of eigenvectors. In fact, an eigenvector is a principal vector of grade 1.

[4]The vectors $\mathbf{x}_j$, $j \in \overline{2,k}$, are sometimes called *generalized eigenvectors* associated with λ, see, e.g., [**684**, p. 74].

The matrix $J_k(\lambda)$ in (115) is called a $k \times k$–*Jordan block* corresponding to the eigenvalue λ. The following theorem, stated without proof (that can be found in linear algebra texts, see, e.g., [**490**, Chapter 6]) is of central importance:

THEOREM 2 (The Jordan normal form). Any matrix $A \in \mathbb{C}^{n\times n}$ is similar to a block diagonal matrix J with Jordan blocks on its diagonal, i.e., there exists a nonsingular matrix X such that

$$X^{-1}AX = J = \begin{bmatrix} J_{k_1}(\lambda_1) & O & \cdots & O \\ O & J_{k_2}(\lambda_2) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & J_{k_p}(\lambda_p) \end{bmatrix}, \tag{119}$$

and the matrix J is unique up to a rearrangement of its blocks. □

The matrix J is the *Jordan normal form* of A. The scalars $\{\lambda_1, \ldots, \lambda_p\}$ in (119) are the eigenvalues of A. Each Jordan block $J_{k_j}(\lambda_j)$ corresponds to k_j principal vectors, of which one is an eigenvector.

Writing (119) as

$$A = XJX^{-1}, \tag{120}$$

we verify

$$A^s = XJ^sX^{-1}, \quad \text{for all integers } s \geq 0, \tag{121}$$

and, for any polynomial p

$$p(A) = Xp(J)X^{-1}. \tag{122}$$

Using (119) and (126) below we verify

$$(A - \lambda_1 I)^{k_1} = X \begin{bmatrix} O & O & \cdots & O \\ O & ? & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & ? \end{bmatrix} X^{-1},$$

where exact knowledge of the ? blocks is not needed. Continuing in this fashion we prove

THEOREM 3 (The Cayley–Hamilton Theorem). For A as above,

$$(A - \lambda_1 I)^{k_1}(A - \lambda_2 I)^{k_2} \cdots (A - \lambda_p I)^{k_p} = O. \tag{123}$$

PROOF.

$$\text{LHS}(123) = X \begin{bmatrix} O & O & \cdots & O \\ O & ? & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & ? \end{bmatrix} \begin{bmatrix} ? & O & \cdots & O \\ O & O & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & ? \end{bmatrix} \cdots \begin{bmatrix} ? & O & \cdots & O \\ O & ? & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & O \end{bmatrix} X^{-1}$$

□

This result can be stated in terms of the polynomial

$$c(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_p)^{k_p}, \tag{124}$$

called the *characteristic polynomial* of A, see Ex. 72. Indeed, LHS(123) is obtained by substituting A for the variable λ, and replacing λ_i by $\lambda_i I$. The n roots of the characteristic polynomial of A are the eigenvalues of A, counting multiplicities. The Cayley–Hamilton Theorem states that a matrix A satisfies the polynomial equation $c(A) = O$, where $c(\lambda)$ is its characteristic polynomial.

If an eigenvalue λ_i is repeated in q Jordan blocks,

$$J_{j_1}(\lambda_i), J_{j_2}(\lambda_i), \cdots, J_{j_q}(\lambda_i), \quad \text{with} \quad j_1 \geq j_2 \geq \cdots \geq j_q,$$

then the characteristic polynomial is the product of factors

$$c(\lambda) = c_1(\lambda)\, c_2(\lambda) \cdots c_p(\lambda)\ , \quad \text{with } c_i(\lambda) = (\lambda - \lambda_i)^{a_i}\ , \quad \text{where } a_i = j_1 + j_2 + \cdots + j_q\ .$$

The exponent a_i is called the *algebraic multiplicity* of the eigenvalue λ_i. It is the sum of dimensions of the Jordan blocks corresponding to λ_i. The dimension of the largest block, j_1, is called the *index* of the eigenvalue λ_i, and denoted $\nu(\lambda_i)$. By (126) below it is the smallest integer k such that

$$(J_{j_1}(\lambda_i) - \lambda_i I_{j_1})^k = O,\ (J_{j_2}(\lambda_i) - \lambda_i I_{j_2})^k = O,\ \cdots,\ (J_{j_q}(\lambda_i) - \lambda_i I_{j_q})^k = O\ ,$$

see also Ex. 82.

Let $m_i(\lambda) = (\lambda - \lambda_i)^{\nu(\lambda_i)}$. Then the polynomial

$$m(\lambda) = m_1(\lambda)\, m_2(\lambda) \cdots m_p(\lambda)$$

satisfies $m(A) = O$, and has the smallest degree among such polynomials. It is called the *minimal polynomial* of A.

Exercises and examples.

EX. 71. Let $A \in \mathbb{C}^{n\times n}$ have the Jordan form (119). Then the following statements are equivalent:
(a) A is nonsingular,
(b) J is nonsingular,
(c) 0 is not an eigenvalue of A.
If these hold, then

$$A^{-1} = XJ^{-1}X^{-1}\ , \tag{125}$$

and (121) holds for all integers s, if we interpret A^{-s} as $\left(A^{-1}\right)^s$.

EX. 72. Let $A \in \mathbb{C}^{n\times n}$. Then the characteristic polynomial of A is $c(\lambda) = (-1)^n \det(A - \lambda I)$.

EX. 73. Let $A \in \mathbb{C}^{n\times n}$. Then λ is an eigenvalue of A if and only if $\overline{\lambda}$ is an eigenvalue of A^* .

PROOF. $\det(A^* - \overline{\lambda} I) = \overline{\det(A - \lambda I)}$. □

EX. 74. Let A be given in Jordan form

$$A = X \begin{bmatrix} J_3(\lambda_1) & O & O & O & O & O \\ O & J_2(\lambda_1) & O & O & O & O \\ O & O & J_2(\lambda_1) & O & O & O \\ O & O & O & J_1(\lambda_1) & O & O \\ O & O & O & O & J_2(\lambda_2) & O \\ O & O & O & O & O & J_2(\lambda_2) \end{bmatrix} X^{-1}\ .$$

Then the characteristic polynomial of A is $c(\lambda) = (\lambda - \lambda_1)^8\, (\lambda - \lambda_2)^4$ and the minimal polynomial is $m(\lambda) = (\lambda-\lambda_1)^3\, (\lambda-\lambda_2)^2$. The algebraic multiplicity of λ_1 is 8, its geometric multiplicity is 4 (every Jordan block contributes an eigenvector), and its index is 3.

EX. 75. A matrix N is *nilpotent* if $N^k = O$ for some integer $k \geq 0$. The smallest such k is called the *index of nilpotency* of N. Let $J_k(\lambda)$ be a Jordan block, and let $j \in \overline{1,k}$. Then

$$(J_k(\lambda) - \lambda I_k)^j = J_k(0)^j = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}^j = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & 1 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}\ , \tag{126}$$

with ones in positions $\{(i, i+j) : i \in \overline{1, k-j}\,\}$, zeros elsewhere. In particular, $(J_k(\lambda)-\lambda I_k)^{k-1} \neq O = (J_k(\lambda)-\lambda I_k)^k$, showing that $(J_k(\lambda) - \lambda I_k)$ is nilpotent with index k.

Ex. 76. Let $J_k(\lambda)$ be a Jordan block and let m be a nonnegative integer. Show that the power $(J_k(\lambda))^m$ is

$$(J_k(\lambda))^m = \begin{bmatrix} \lambda^m & m\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \binom{m}{k-1}\lambda^{m-k+1} \\ 0 & \lambda^m & m\lambda^{m-1} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \binom{m}{2}\lambda^{m-2} \\ \vdots & & & \lambda^m & m\lambda^{m-1} \\ 0 & \cdots & \cdots & 0 & \lambda^m \end{bmatrix} \tag{127a}$$

$$= \sum_{j=0}^{k-1} \binom{m}{j}\lambda^{m-j}(J_k(\lambda) - \lambda\, I_k)^j \tag{127b}$$

$$= \sum_{j=0}^{k-1} \frac{p^{(j)}(\lambda)}{j!}\,(J_k(\lambda) - \lambda\, I_k)^j\ , \tag{127c}$$

where $p(\lambda) = \lambda^m$, $p^{(j)}$ is j th–derivative, and in (127), $\binom{m}{\ell}$ is interpreted as zero if $m < \ell$.

Ex. 77. Let $J_k(\lambda)$ be a Jordan block, and let $p(\lambda)$ be a polynomial. Then $p(J_k(\lambda)$ is defined by using (127)

$$p(J_k(\lambda)) = \begin{bmatrix} p(\lambda) & \frac{1}{1!}p'(\lambda) & \frac{1}{2!}p''(\lambda) & \cdots & \frac{1}{(k-1)!}p^{(k-1)}(\lambda) \\ 0 & p(\lambda) & \frac{1}{1!}p'(\lambda) & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \frac{1}{2!}p''(\lambda) \\ \vdots & & & p(\lambda) & \frac{1}{1!}p'(\lambda) \\ 0 & \cdots & \cdots & 0 & p(\lambda) \end{bmatrix} \tag{128}$$

$$= \sum_{j=0}^{k-1} \frac{p^{(j)}(\lambda)}{j!}\,(J_k(\lambda) - \lambda\, I_k)^j\ . \tag{127c}$$

Ex. 78. ([**359**, p. 104]). Let P_n be the set of polynomials with real coefficients, of degree $\le n$, and let T be the differentiation operator $Tf(x) = f'(x)$, see Ex. 15(b). The solution of $f'(x) = \lambda f(x)$ is $f(x) = e^{\lambda x}$, which is a polynomial only if $\lambda = 0$, the only eigenvalue of T. The geometric multiplicity of this eigenvalue is 1, its algebraic multiplicity is $n + 1$.

Ex. 79. Let the $n \times n$ matrix A have the characteristic polynomial

$$c(\lambda) = \lambda^n + c_{n-1}\,\lambda^{n-1} + \cdots + c_2\,\lambda^2 + c_1\,\lambda + c_0\ .$$

Then A is nonsingular if and only if $c_0 \ne 0$, in which case

$$A^{-1} = -\frac{1}{c_0}\left(A^{n-1} + c_{n-1}\,A^{n-2} + \cdots + c_2\,A + c_1\,I\right)\ .$$

Ex. 80. Let A be a nonsingular matrix. Show that its minimal polynomial can be written in the form

$$m(\lambda) = c(1 - \lambda\, q(\lambda)) \tag{129}$$

where $c \ne 0$ and q is a polynomial, in which case

$$A^{-1} = q(A)\ . \tag{130}$$

See also Ex. 6.87 below.

Ex. 81. Let the 2×2 matrix A have eigenvalues $\pm i$. Find A^{-1}.

Hint. Here $c(\lambda) = m(\lambda) = \lambda^2 + 1$. Use Ex. 79 or Ex. 80. □

Ex. 82. For a given eigenvalue λ, the maximal grade of the λ–vectors of A is the index of λ.

8. The Smith normal form

Let $\mathbb{Z}$ denote the *ring of integers* $0, \pm 1, \pm 2, \dots$ and let:
$\mathbb{Z}^m$ be the m–dimensional vector space over Z,
$\mathbb{Z}^{m\times n}$ be the $m \times n$ matrices over $\mathbb{Z}$,
$\mathbb{Z}_r^{m\times n}$ be the same with rank r.
Any vector in $\mathbb{Z}^m$ will be called an *integral vector*. Similarly, any element of $\mathbb{Z}^{m\times n}$ will be called an *integral matrix*.

A nonsingular matrix $A \in \mathbb{Z}^{n\times n}$ whose inverse A^{-1} is also in $\mathbb{Z}^{n\times n}$ is called a *unit matrix*; e.g. Marcus and Minc [**530**, p. 42].

Two matrices $A, S \in \mathbb{Z}^{m\times n}$ are said to be *equivalent over* $\mathbb{Z}$ if there exist two unit matrices $P \in \mathbb{Z}^{m\times m}$ and $Q \in \mathbb{Z}^{n\times n}$ such that

$$PAQ = S \ . \tag{131}$$

THEOREM 4. Let $A \in \mathbb{Z}_r^{m\times n}$. Then A is equivalent over $\mathbb{Z}$ to a matrix $S = [s_{ij}] \in \mathbb{Z}_r^{m\times n}$ such that:
(a) $s_{ii} \neq 0$, $\quad i \in \overline{1, r}$,
(b) $s_{ij} = 0$ otherwise, and
(c) s_{ii} divides $s_{i+1,i+1}$ for $i \in \overline{1, r-1}$.

Remark. S is called the *Smith normal form* of A, and its nonzero elements s_{ii} $(i \in \overline{1, r})$ are the *invariant factors* of A; see, e.g., Marcus and Minc [**530**, pp. 42–44].

PROOF. The proof given in Marcus and Minc [**530**, p. 44] is constructive and describes an algorithm to
(i) find the greatest common divisor of the elements of A,
(ii) bring it to position $(1, 1)$, and
(iii) make zeros of all other elements in the first row and column.
This is done, in an obvious way, by using a sequence of elementary row and column operations consisting of

$$\text{interchanging two rows [columns]} \tag{132}$$

$$\text{subtracting an integer multiple of one row [column] from another row [column]} \tag{133}$$

The matrix $B = [b_{ij}]$ so obtained is equivalent over $\mathbb{Z}$ to A, and
b_{11} divides b_{ij} $\quad (i > 1, j > 1)$,
$b_{i1} = b_{1j} = 0$ $\quad (i > 1, j > 1)$.
Setting $s_{11} = b_{11}$, one repeats the algorithm for $(m-1) \times (n-1)$ matrix $[b_{ij}]$ $\quad (i > 1, j > 1)$, etc.

The algorithm is repeated r times and stops when the bottom right $(m-r) \times (n-r)$ submatrix is zero, giving the Smith normal form.

The unit matrix $P[Q]$ in (131) is the product of all the elementary row [column] operators, in the right order. □

Exercises.

EX. 83. Two matrices $A, B \in \mathbb{Z}^{m\times n}$ are equivalent over $\mathbb{Z}$ if and only if B can be obtained from A by a sequence of elementary row and column operations (132)–(133).

EX. 84. Describe in detail the algorithm mentioned in the proof of Theorem 4.

9. Nonnegative matrices

A matrix $A = \left[a_{ij}\right] \in \mathbb{R}^{n\times n}$ is:
(a) *nonnegative* if all $a_{ij} \geq 0$, and
(b) *reducible* if there is a permutation matrix Q such that

$$Q^T A Q = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix} \tag{134}$$

where the submatrices A_{11}, A_{22} are square, and is otherwise *irreducible*.

THEOREM 5 (The Perron–Frobenius Theorem). If $A \in \mathbb{R}^{n\times n}$ is nonnegative and irreducible then:

(a) A has a positive eigenvalue, ρ, equal to the spectral radius of A.

(b) ρ has algebraic multiplicity 1.

(c) There is a positive eigenvector corresponding to ρ.

Suggested further reading

§ 4.4. Bhimasankaram [**94**], Hartwig [**374**].

Section 5. Schur complements: Carlson [**165**], Cottle [**204**], Horn and Johnson [**423**] and Ouellette [**623**].

Section 6. Finzel [**267**], Marcus ([**527**],[**528**]), Mostow and Sampson [**580**], Mostow, Sampson and Meyer [**581**], Niu [**611**].

Section 9. Berman and Plemmons [**90**], Lancaster and Tismenetsky [**490**, Chapter 15].