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 is Section 8.5.

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 incorporate these changes while maintaining the informal and leisurely style of the first edition. New material was added, including a preliminary chapter, Chapter 0, about 100 new exercises and a 1000 new references, applications to statistics (omitted from the first edition because these were covered in the then recent books by Albert [**13**] and Rao and Mitra [**1250**]) and matrix volume. Otherwise, the old text is mostly unchanged.

Many colleagues helped this effort. Special thanks go to R. Bapat, R. Bhatia, S. Campbell, J. Miao, S. K. Mitra, 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, that marked the first edition, are sadly missed.

May 2001 *A. Ben–Israel*

Contents

Glossary of notation vii

Introduction 1
1. The inverse of a nonsingular matrix 1
2. Generalized inverses of matrices 1
3. Illustration: Solvability of linear systems 2
4. Diversity of generalized inverses 3
5. Preparation expected of the reader 4
6. Historical note 4
7. Remarks on notation 5

Suggested further reading 5

Chapter 0. Preliminaries 7
1. Scalars and vectors 7
2. Linear transformations and matrices 11
3. Elementary operations and determinants 21

Chapter 1. Existence and Construction of Generalized Inverses 27
1. The Penrose equations 27
2. Existence and construction of $\{1\}$–inverses 28
3. Properties of $\{1\}$–inverses 29
4. Bases for the range and null space of a matrix 31
5. Existence and construction of $\{1,2\}$–inverses 34
6. Existence and construction of $\{1,2,3\}$–, $\{1,2,4\}$– and $\{1,2,3,4\}$–inverses 35
7. Full–rank factorization 37
8. Explicit formula for $A^\dagger$ 37
9. Construction of $\{2\}$–inverses of prescrived rank 38
10. An application of $\{2\}$–inverses in iterative methods for solving nonlinear equations 40

Suggested further reading 43

Chapter 2. Linear Systems and Characterization of Generalized Inverses 45
1. Solutions of linear systems 45
2. Characterization of $A\{1,3\}$ and $A\{1,4\}$ 48
3. Characterization of $A\{2\}$, $A\{1,2\}$ and other subsets of $A\{2\}$. 49
4. Idempotent matrices and projectors 51
5. Generalized inverses with prescribed range and null space 57

6. Orthogonal projections and orthogonal projectors 60
7. Efficient characterization of classes of generalized inverses 69
8. Restricted generalized inverses 72
9. The Bott–Duffin inverse 75
10. An application of $\{1\}$–inverses in interval linear programming 78
11. A $\{1, 2\}$–inverse for the integral solution of linear equations 80
12. An application of the Bott–Duffin inverse to electrical networks 82
Suggested further reading 86

Chapter 3. Minimal Properties of Generalized Inverses 87
1. Least–squares solutions of inconsistent linear systems 87
2. Solutions of minimum norm 93
3. Weighted generalized inverses 98
4. Essentially strictly convex norms and the associated projectors and generalized inverses* 103
5. An extremal property of the Bott–Duffin inverse with application to electrical networks 122
Suggested further reading 124

Chapter 4. Spectral Generalized Inverses 125
1. Introduction 125
2. Spectral properties of a nonsingular matrix 125
3. Spectral inverse of a diagonable matrix 126
4. The group inverse 127
5. Spectral properties of the group inverse 130
6. The Drazin inverse 132
7. Spectral properties of the Drazin inverse 136
8. Index 1–nilpotent decomposition of a square matrix 137
9. Quasi–commuting inverses 139
10. Other spectral generalized inverses 139
Suggested further reading 143

Chapter 5. Generalized Inverses of Partitioned Matrices 145
1. Introduction 145
2. Partitioned matrices and linear equations 145
3. Intersection of manifolds 151
4. Common solutions of linear equations and generalized inverses of partitioned matrices 158
5. Greville's method and related results 165
6. Generalized inverses of bordered matrices 171
Suggested further reading 173

Chapter 6. A Spectral Theory for Rectangular Matrices 175
1. Introduction 175

2. The UDV^* decomposition 179
3. Partial isometries and the polar decomposition theorem 186
4. A spectral theory for rectangular matrices 199
Suggested further reading 208

Chapter 7. Computational Aspects of Generalized Inverses 209
1. Introduction 209
2. Computation of unrestricted $\{1\}$–inverses and $\{1,2\}$–inverses 209
3. Computation of unrestricted $\{1,3\}$–inverses 210
4. Computation of $\{2\}$–inverses with prescribed range and null space 211
5. Iterative methods for computing $A^\dagger$ 213
Suggested further reading 224

Chapter 8. Generalized Inverses of Linear Operators between Hilbert Spaces 225
1. Introduction 225
2. Hilbert spaces and operators: Preliminaries and notation 225
3. Generalized inverses of linear operators between Hilbert spaces 231
4. Minimal properties of generalized inverses 249
5. Series and integral representations and iterative computation of generalized inverses 256
Suggested further reading 260

Bibliography 261

Subject Index 309

Author Index 313

Glossary of notation

$(T_2)_{[D(T_1)]}$ – restriction of T to $D(T_1)$, 227
$A \otimes B$ – Kronecker product of A, B, 46
$A\{1\}_{T,S}$ – $\{1\}$ inverses of A associated with T, S, 58
$A\{i, j, \dots, k\}_s$ – matrices in $A\{i, j, \dots, k\}$ of rank s, 49
$A^{(-1)}_{(L)}$ – Bott–Duffin inverse of A w.r.t. L, 76
$A^{1/2}$ – square root of A, 190
$A_{(k)}$ – best rank-k approximation of A, 182
$A_{[S]}$ – restriction of A to S, 73
$A_{\{\mathcal{U},\mathcal{V}\}}$ – matrix representation of A w.r.t $\{\mathcal{U}, \mathcal{V}\}$, 12
$A_{\{\mathcal{V}\}}$ – matrix representation of A w.r.t $\{\mathcal{V}, \mathcal{V}\}$, 12
$C(T)$, 226
$C[a, b]$ – continuous functions on $[a, b]$, 242
$D(T)$, 226
$E^i(\alpha)$ – elementary operation of type 1, 21
E^{ij} – elementary operation of type 3, 21
$E^{ij}(\beta)$ – elementary operation of type 2, 21
$F(A)$ – functions $f : \mathbb{C} \to \mathbb{C}$ analytic on $\sigma(A)$, 205
$G(T)$, 226
$G^{-1}(T)$, 226
$H > O$ – H positive definite, 151
$H \geq O$ – H positive semidefinite, 151
$IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$, 78
$L \oplus M$ – direct sum of L, M, 7
$L^\perp$, 225
$L^\perp$ – orthogonal complement of L, 58
$M \oplus N$ – direct sum of M, N, 225
$N(A, B)$ – matrices X with $AXB = O$, 96
$N(T)$, 226
P_L – orthogonal projector on L, 60
P_π – permutation matrix corresponding to π, 21
$P_{L,M}$ – projector on L along M, 51
$P_{L,\phi}$ – ϕ–metric projector on L, 104
$P^{-1}_{L,\phi}(\mathbf{l})$ – inverse image of $\mathbf{l}$ under $P_{L,\phi}$, 106
$Q(\alpha)$ – projective bound of α, 116
$R(A, B)$ – matrices AXB for some X, 96
$R(T)$, 226
R_k – residual, 213
S_n – symmetric group (permutations of order n), 21
$T \geq O$, 228
T^*, 227
T_r – restriction of T, 235
$U^{n\times n}$ – $n \times n$ unitary matrices, 175
$A[j \leftarrow \mathbf{b}]$ – A with j th–column replaced by $\mathbf{b}$, 25
$\mathbb{C}$, 7
$\mathbb{C}^{m\times n}_r$ – $m \times n$ complex matrices with rank r, 22
$\mathbb{R}$, 7
$\Re$ – real part, 9
$\mathbb{R}^{m\times n}_r$ – $m \times n$ real matrices with rank r, 22
$W^{m\times n}_\ell$ – partial isometries in $\mathbb{C}^{m\times n}_\ell$, 196
$\mathbb{Z}$ – ring of integers, 80
$\mathbb{Z}^m$ – m–dimensional vector space over Z, 80
$\mathbb{Z}^{m\times n}_r$ – $m \times n$ matrices over $\mathbb{Z}$ with rank r, 80
$\mathbb{Z}^{m\times n}$ – $m \times n$ matrices over $\mathbb{Z}$, 80
$\alpha(A)$ – singular values of A, 176
$\alpha_j(A)$ – the jth singular value of A, 176
$\det A$ – determinant of A, 23
A^D – Drazin inverse of A, 134
$\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$, 79
$\gamma(T)$, 228
$A\{2\}_{S,T}$ – $\{2\}$–inverses with range T, null space S, 59
$A\{i, j, \dots, k\}$ – $\{i, j, \dots, k\}$–inverses of A, 27
$A^{(-1)}_{\alpha,\beta}$ – α–β generalized inverse of A, 107
$A^{(1)}_{T,S}$ – a $\{1\}$ inverse of A associated with T, S, 58
$A^{(1,2)}_{(W,U)}$ – $\{W, U\}$ weighted $\{1, 2\}$ inverse of A, 100
$A^{(N)}$ – nilpotent part of A, 138
$A^{(S)}$ – S–inverse of A, 140
$A^{(i,j,\dots,k)}$ – an $\{i, j, \dots, k\}$–inverse of A, 27
$A^{<k>}$ – generalized kth power of A, 206
$A^\#$ – group inverse of A, 127
$A^\dagger$ – Moore–Penrose inverse of A, 27
$T^\dagger_S$ – the $N(S)$–restricted pseudoinverse of T, 255
$T^\dagger_e$ – extremal g.i., 252
T^q – Tseng generalized inverse, 231
$\lambda^\dagger$ – Moore–Penrose inverse of the scalar λ, 29
$\langle X, Y\rangle$ – inner product on $\mathbb{C}^{m\times n}$, 96
$\langle \mathbf{x}, \mathbf{y}\rangle$ – inner product of $\mathbf{x}, \mathbf{y}$, 8
$\langle \mathbf{x}, \mathbf{y}\rangle_Q$ – the inner product $\mathbf{y}^*Q\mathbf{x}$, 9
$\langle x, y\rangle$, 225
$\langle \alpha\rangle$ – smallest integer $\geq \alpha$, 221
$\mathbb{C}^{m\times n}$ – $m \times n$ complex matrices, 27
$\mathbb{F}^n$, 7
$\mathbb{F}^{m\times n}$ – $m \times n$ matrices over $\mathbb{F}$, 11
$\mathbb{R}^{m\times n}$ – $m \times n$ real matrices, 27
$\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ – bounded operators in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, 226
$\mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ – closed operators in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, 226
$\mathcal{E}_n$ – standard basis of $\mathbb{C}^n$, 12
$\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$ – Hilbert spaces, 225
$\mathcal{L}(U, V)$ – linear transformations: $U \to V$, 11
$\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ – linear transformations: $\mathbb{C}^n \to \mathbb{C}^m$, 51
$\mathcal{S}$ – function space, 242
$\overline{1, n}$– the index set $\{1, 2, \dots, n\}$, 5
$\| A \|_{\alpha,\beta}$ – least upper bound of A corresponding to $\{\alpha, \beta\}$, 116
$\| \mathbf{x} \|$ – norm of $\mathbf{x}$, 8
$\| \mathbf{x} \|_Q$ – ellipsoidal norm of $\mathbf{x}$, 9

$M \overset{\perp}{\oplus} N$ – orthogonal direct sum of M, N, 225
π^{-1} – permutation inverse to π, 21
$\mathrm{cond}(A)$ – condition number of A, 178
$\mathrm{LHS}(i.j)$ – left–hand side of equation $(i.j)$, 5
$\mathrm{RHS}(i.j)$ – right–hand side of equation $(i.j)$, 5
$\mathrm{ext}\, B$ – extension of B to $\mathbb{C}^n$, 73
$\mathrm{lub}_{\alpha,\beta}(A)$, 116
$\mathrm{sign}\, \pi$ – sign of permutation π, 22
$\mathrm{vec}(X)$ – vector made of rows of X, 46
$\mathbf{e}\mathbf{e}^T$ – matrix of ones, 85
$\widehat{A}$, 81
fl – floating point, 89

EP – matrices A with $R(A) = R(A^*)$, 128
EP_r, 128

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 [**1296**]) 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 [**174**], Penrose [**1177**]) 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 8. Pseudoinverses on general rings and semigroups are not discussed; the interested reader is referred to Drazin [**423**], Foulis [**507**], and Munn [**1098**].

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. In particular, applications of generalized inverses in statistics will not be treated here, since they are amply covered in the books by Rao and Mitra [**1250**] and Albert [**13**]. 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 [**1121**].

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 8, 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 [**645**] or Noble [**1145**], 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 [**515**], 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 [**761**], who used the finite dimensionality of the null spaces of the Fredholm operators to give a simple algebraic construction (see, e.g., Exercises 8.19–8.20). Generalized inverses of differential operators, already implicit in Hilbert's discussion in 1904 of generalized Green's functions, [**734**], were consequently studied by numerous authors, in particular Myller (1906), Westfall (1909), Bounitzky [**212**] in 1909, Elliott (1928), and Reid (1931). For a history of this subject see the excellent survey by Reid [**1263**].

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 [**1087**], 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, [**906**, p. 676], has assigned the date 1906. Details were published, [**1088**], only in 1935 after Moore's death. 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 [**1361**] in 1937, and for operators by Tseng ([**1461**]–1933, [**1466**],[**1464**],[**1465**]–1949), Murray and von Neumann [**1103**] in 1936, Atkinson ([**45**]–1952, [**46**]–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 [**173**], [**172**], [**174**]). In 1955 Penrose [**1177**] 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 [**1504**] and [**1508**, p. 90], Murray and von Neumann [**1103**, p. 299], McCoy [**1009**], and Hartwig [**668**].

Section 4. For generalized inverses in abstract geometric setting see also Davis and Robinson [**381**], Gabriel ([**520**], [**521**], [**522**]), Hansen and Robinson [**653**], Hartwig [**668**], Munn and Penrose [**1100**], Pearl [**1172**], Rabson [**1227**] and Rado [**1230**].

For applications in statistics see Albert ([**13**], [**14**]), Albert and Sittler [**16**], Banerjee [**65**], Banerjee and Federer [**66**], Chernoff [**325**], Chipman ([**326**], [**327**]), Chipman and Rao ([**331**], [**330**]), Drygas ([**429**], [**430**], [**431**]), Goldman and Zelen [**544**], Golub ([**549**], [**550**]), Golub and Styan [**558**], Good [**562**], Graybill and Marsaglia [**571**], J. A. John [**783**], P. W. M. John [**784**], Meyer and Painter [**1032**], Mitra [**1058**], Mitra and Rao ([**1075**], [**1076**]), Price [**1206**], Rao ([**1240**], [**1241**], [**1242**]), Rao and Mitra [**1250**], Rayner and Pringle [**1258**], Rohde ([**1296**], [**1299**]), Rohde and Harvey [**1300**], Tan [**1419**], Zacks [**1624**], Zyskind [**1669**], and Zyskind and Martin [**1670**].

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 8. 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 *ith 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. An *inner product* of $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, is defined as a function : $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ 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 every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^n$ and every $\alpha \in \mathbb{C}$.

Note:

(a) For all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ 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 (I3) states, in particular, that $\langle \mathbf{x}, \mathbf{x} \rangle$ is real for all $\mathbf{x} \in \mathbb{C}^n$.

(c) The *if* part in (I3) follows from (I1) with $\alpha = 0$, $\mathbf{y} = \mathbf{0}$.

The *standard inner product* of $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_i)$ is

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

See Exs. 2–4.

1.5. A (*vector*) *norm* of $\mathbf{x} \in \mathbb{C}^n$, denoted by $\|\mathbf{x}\|$, is defined as a function : $\mathbb{C}^n \to \mathbb{R}$ 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 every $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ and every $\alpha \in \mathbb{C}$.

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. 3–9.

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. Prove: 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 $\|\mathbf{x}\| = \langle x, x \rangle^{1/2}$, 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 $\mathbb{C}^n$, we define orthogonality by the same condition, $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, as in the real case.

Ex. 6. 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. 7. *The ℓ_p–norms.* for any $p \geq 1$ the function

$$\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} \tag{10}$$

is a norm on $\mathbb{C}^n$, called the *ℓ_p–norm.*

Hint: The statement that (10) satisfies (N3) for $p \geq 1$ is the classical Minkowski's inequality; see, e.g., Beckenbach and Bellman [**100**].

Ex. 8. The most popular ℓ_p–norms are the choices $p = 1, 2$, and ∞

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| \; , \text{ the } \ell_1\text{–norm} \; , \tag{10.1}$$

$$\|\mathbf{x}\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} = (\mathbf{x}^*\mathbf{x})^{1/2} \; , \text{ the } \ell_2\text{–}\textit{norm} \text{ or the } \textit{Euclidean norm} \; , \tag{10.2}$$

$$\|\mathbf{x}\|_\infty = \max\{|x_j| : j \in \overline{1, n}\} \; , \text{ the } \ell_\infty\text{–}\textit{norm} \text{ or the } \textit{Tchebycheff norm} \; . \tag{10.∞}$$

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

Ex. 9. Let $\| \ \|_{(1)}$, $\| \ \|_{(2)}$ be any two norms on $\mathbb{C}^n$. Show that there exist positive scalars α, β such that

$$\alpha \|\mathbf{x}\|_{(1)} \leq \|\mathbf{x}\|_{(2)} \leq \beta \|\mathbf{x}\|_{(1)} \; , \tag{11}$$

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

Hint:

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

and

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

REMARK 1. Two norms, $\| \ \|_{(1)}$ and $\| \ \|_{(2)}$ are called *equivalent* if there exist positive scalars α, β such that (11) holds for all $\mathbf{x} \in \mathbb{C}^n$. from Ex. 9, *any* two norms on $\mathbb{C}^n$ are equivalent. Therefore, if a sequence $\{\mathbf{x}_k\} \subset \mathbb{C}^n$ satisfies

$$\lim_{k\to\infty} \|\mathbf{x}_k\| = 0 \tag{12}$$

for some norm, then (12) holds for any norm. Topological concepts like convergence and continuity, defined by limiting expressions like (12), 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 $A[i, j]$. The matrix A is

diagonal if $A[i, j] = 0$ for $i \neq j$,
upper triangular if if $A[i, j] = 0$ for $i > j$,
lower triangular if if $A[i, j] = 0$ for $i < j$.

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

Hermitian if $A = A^*$,
normal if $AA^* = A^*A$.

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

$$\begin{aligned} T^{-1}(T\mathbf{u}) &= \mathbf{u} \text{ and } T(T^{-1}\mathbf{v}) = \mathbf{v} \ , \ \forall\, \mathbf{u} \in U, \mathbf{v} \in V \ , & \text{(13a)} \\ \text{or equivalently} \ , \ T^{-1}T &= I_U \ , \ TT^{-1} = I_V \ , & \text{(13b)} \end{aligned}$$

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 $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 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 $A_{\{\mathcal{U},\mathcal{V}\}} = [a_{ij}]$ determined (uniquely) by

$$A\mathbf{v}_j = \sum_{i=1}^{m} a_{ij}\, \mathbf{u}_i \ , \ j \in \overline{1, n} \ . \tag{14}$$

For any such pair of bases $\{\mathcal{U}, \mathcal{V}\}$, (14) 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 $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 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

$$A\mathbf{v}_j = \sum_{i=1}^{n} a_{ij}\, \mathbf{v}_i \ , \ j \in \overline{1, n} \ . \tag{15}$$

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

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

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. 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 [**753**, Section 2.2]) define a matrix norm as a function having all four properties (M1)–(M4).

Exercises and examples.

Ex. 10. 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. 11. Let P_n be the set of polynomials with real coefficients, of degree $\leq n$,

$$P_n = \{\mathbf{p} :\ \mathbf{p}(x) = p_0 + p_1 x + \cdots + p_n x^n\,,\ p_i \in \mathbb{R}\}\,. \tag{16}$$

The name x of the variable in (16) 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. 12. Let $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be two bases of $\mathbb{C}^n$. Show that there is a unique $n \times n$ matrix $S = \big[s_{ij}\big]$ such that

$$\mathbf{w}_j = \sum_{i=1}^{n} s_{ij}\, \mathbf{v}_i \ , \ \ j \in \overline{1,n} \ , \tag{17}$$

and S is nonsingular. Using the rules of matrix multiplication we rewrite (17) as

$$[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & & \vdots \\ s_{n1} & \cdots & s_{nn} \end{bmatrix} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] S \ , \tag{18}$$

i.e.

$$[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n] S^{-1} \ . \tag{19}$$

Ex. 13. *Similar matrices.* We recall that two square matrices A, B are called *similar* if

$$B = S^{-1} A S \tag{20}$$

for some nonsingular matrix S. If S in (20) 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 $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ and $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ 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 $S = \big[s_{ij}\big]$ satisfying (17). Rewriting (15) as

$$A[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] A_{\{\mathcal{V}\}} \ . \tag{21}$$

we conclude, by substituting (19) in (21), that

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

and by the uniqueness of the matrix representation,

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

Only if. Similarly proved. □

Ex. 14. *Schur triangularization.* Any $A \in \mathbb{C}^{n \times n}$ is unitarily similar to a triangular matrix.

Proof. See, e.g., Marcus and Minc [**996**, p. 67]. □

Ex. 15. *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

$$\sum_{ij} |b_{ij}| \leq \epsilon \quad \text{(Bellman [\textbf{101}, p. 205])} .$$

Ex. 16. 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. 17. 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. 18. Denote the transformation of bases (17) 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

$$A_{\{\widetilde{\mathcal{U}},\widetilde{\mathcal{V}}\}} = S^{-1}\, A_{\{\mathcal{U},\mathcal{V}\}}\, T \ . \tag{22}$$

Proof. Similar to thr proof of Ex. 13. □

Ex. 19. *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

$$B = S^{-1}AT \ . \tag{23}$$

If S and T in (23) are unitary matrices, then A, B are called *unitarily equivalent.*

It follows from Ex. 18 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. 20. 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. 21. 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. 22. 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. 20–21. □

Ex. 23. 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) the eigenvectors of A

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

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

Ex. 24. 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. 11.
(b) Calculate the representations of the differentiation operator (Ex. 11(b)) w.r.t. to the bases $\{\widetilde{\mathcal{U}}_n, \widetilde{\mathcal{U}}_{n-1}\}$, and verify (22).
(c) Same for the integration operator of Ex. 11(c).

Ex. 25. 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. 26. 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 $P_{L,M}\,\mathbf{y} = P_{L,M}^2\,\mathbf{x}$, the projection of $\mathbf{y}$ on L along M? Show, therefore, that the transformation $P_{L,M}$ is idempotent.

Ex. 27. *Matrix norms.* Show that the functions

$$\left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} = (\operatorname{trace} A^* A)^{1/2} \tag{24}$$

and

$$\max\{|a_{ij}| : i \in \overline{1,m}\ ,\ j \in \overline{1,n}\} \tag{25}$$

are matrix norms. Which of these functions is a multiplicative norm?

Ex. 28. *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,

$$\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|\ . \tag{26}$$

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

$$\|A\|_* = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*} \tag{27}$$

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

$$\|A\| \geq \|A\|_*\ ,\ \text{ for all } A\ . \tag{28}$$

The norm $\|A\mathbf{x}\|_*$ defined by (27), is called the *matrix norm corresponding to the vector norm* $\|\ \|_*$, or the *bound* of A with respect to $K = \{\mathbf{x} : \|\mathbf{x}\|_* \leq 1\}$; see, e.g. Householder [**753**, Section 2.2] and Ex. 3.63 below.

Ex. 29. Show that (27) is the same as

$$\|A\mathbf{x}\|_* = \sup_{\|\mathbf{x}\|_* \leq 1} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*} = \sup_{\|\mathbf{x}\|_* = 1} \|A\mathbf{x}\|_*\ . \tag{29}$$

Ex. 30. Given a multiplicative matrix norm, find a vector norm consistent with it.

Ex. 31. *Corresponding norms.*

(a) The matrix norm on $\mathbb{C}^{m\times n}$, corresponding to the vector norm

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| \tag{10.1}$$

is

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}| \ . \tag{30}$$

(b) The matrix norm on $\mathbb{C}^{m\times n}$, corresponding to the vector norm

$$\|\mathbf{x}\|_\infty = \max_{1\le j\le n} |x_j| \tag{10.∞}$$

is

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}| \ . \tag{31}$$

Proof. (a) Follows from (29) since for any $\mathbf{x} \in \mathbb{C}^n$

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

with equality if $\mathbf{x}$ is the kth unit vector, where k is any j for which the maximum in (30) is attained

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

(b) Similarly proved. □

Ex. 32. The matrix norm on $\mathbb{C}^{m\times n}$, corresponding to the Euclidean norm

$$\|\mathbf{x}\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} \tag{10.2}$$

is

$$\|A\|_2 = \max\{\sqrt{\lambda} : \ \lambda \text{ an eigenvalue of } A^*A\} \ . \tag{32}$$

Note that (32) is different from (24), which is the Euclidean norm of the mn–dimensional vector obtained by listing all components of A in some order. The norm $\| \ \|_2$ given by (32) is called the *spectral norm*.

Ex. 33. For any matrix norm $\| \ \|$ on $\mathbb{C}^{m\times n}$, consistent with some vector norm, the norm of the unit matrix satisfies

$$\|I_n\| \geq 1 \ .$$

In particular, if $\| \ \|_*$ is a matrix norm, computed by (27) from a corresponding vector norm, then

$$\|I_n\|_* = 1 \ . \tag{33}$$

Ex. 34. 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 (24) and (32) are unitarily invariant.

Ex. 35. *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,

$$\rho(A) = \max\{|\lambda| : \lambda \text{ an eigenvalue of } A\} \ . \tag{34}$$

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

$$\rho(A) \leq \|A\| \ . \tag{35}$$

PROOF. Let $\| \ \|$ denote both a given multiplicative matrix norm, and a vector norm consistent with it; see, e.g., Ex. 30. Then

$$A\mathbf{x} = \lambda\mathbf{x} \quad \text{implies} \quad |\lambda|\|\mathbf{x}\| = \|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\| \ .$$

$\square$

Ex. 36. For any $A \in \mathbb{C}^{n\times n}$ and any $\epsilon > 0$, there exists a multiplicative matrix norm $\| \ \|$ such that

$$\|A\| \leq \rho(A) + \epsilon \quad (\text{Householder } [\mathbf{753}, \text{ p. } 46]) \ .$$

Ex. 37. If A is a square matrix,

$$\rho(A^k) = \rho^k(A) \ , \ k = 0, 1, \dots \tag{36}$$

Ex. 38. For any $A \in \mathbb{C}^{m\times n}$, the spectral norm $\| \ \|_2$ of (32) equals

$$\|A\|_2 = \rho^{1/2}(A^*A) = \rho^{1/2}(AA^*) \ . \tag{37}$$

In particular, if A is Hermitian then

$$\|A\|_2 = \rho(A) \ . \tag{38}$$

In general the spectral norm $\|A\|_2$ and the spectral radius $\rho(A)$ may be quite apart; see, e.g., Noble [**1145**, p. 430].

Ex. 39. *Convergent matrices.* A square matrix A is called *convergent* if

$$A^k \to O \text{ as } k \to \infty \,. \tag{39}$$

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

$$\rho(A) < 1 \,. \tag{40}$$

Proof. If: From (40) and Ex. 36 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 (39).
Only if: If $\rho(A) \geq 1$, then by (36), so is $\rho(A^k)$ for $k = 0, 1, \dots$, contradicting (39). □

Ex. 40. 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.,

$$(I - A)^{-1} = I + A + A^2 + \cdots = \sum_{j=0}^{\infty} A^j \quad \text{(Householder [\textbf{753}, p. 54])} \,. \tag{41}$$

Ex. 41. Let A be convergent. Then

$$(I + A)^{-1} = I - A + A^2 - \cdots = \sum_{j=0}^{\infty} (-1)^j A^j \,. \tag{42}$$

Ex. 42. Let $A \in \mathbb{C}^{n\times n}$ be nonsingular, and let $\| \ \|$ be any multiplicative matrix norm. Then $A + B$ is nonsingular for any matrix B satisfying

$$\|B\| < \frac{1}{\|A^{-1}\|} \,. \tag{43}$$

Proof. From

$$A + B = A(I + A^{-1}B)$$

and Ex. 41, it follows that $A + B$ is nonsingular if $A^{-1}B$ is convergent which, by Ex. 35, is implied by

$$\|A^{-1}B\| < 1$$

and hence by

$$\|A^{-1}\|\|B\| < 1 \,.$$

See also Exs. 4.57 and 6.15. □

Ex. 43. *Stein's Theorem.* A square matrix is convergent if and only if there exists a positive definite matrix H such that $A - A^*HA$ is also positive definite (Stein [**1396**], Taussky [**1435**]).

3. Elementary operations and determinants

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[1].

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 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 $\{1, 2, \ldots, n\}$, 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 $\pi_0\{1, 2, \ldots, n\} = \{1, 2, \ldots, n\}$,

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

(c) with any permutation π, its *inverse*, mapping $\{\pi(1), \pi(2), \ldots, \pi(n)\}$ back to $\{1, 2, \ldots, n\}$. 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,

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

---

[1]Only operations of types 1,2 are necessary, see Ex. 44(b). Type 3 operations are introduced for convenience, because of their frequent use.

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,

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

3.3. *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 [**996**, § 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

$$1 \le c_1 < c_2 < \cdots < c_r \le n \ , \tag{44}$$

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

$$R = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} \tag{45}$$

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 $\{\mathbf{e}_k : k \neq c_j \ , \ j \in \overline{1,r}\}$ in any order.

In particular cases, the partitioned form (45) 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 $r = m$.

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

$$E\,A\,P = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} , \tag{46}$$

where

$$E = E_k E_{k-1} \cdots E_2 E_1 \ , \tag{47}$$

in which case A is determined to have rank r. A *Gaussian elimination* is a sequence of elementary row operations as above, that reduce a given matrix to its Hermite normal form. 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. 46 below.

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

$$\det A = \sum_{\pi\in S_n} \operatorname{sign}\pi \prod_{i=1}^{n} a_{\pi(i),i} \tag{48}$$

see, e.g. Marcus and Minc [**996**, § 2.4]. We use here an alternative definition.

Definition 2. (Cullen and Gale [**369**]). The *determinant* is a function $\det : \mathbb{C}^{n\times n} \to \mathbb{C}$ such that

(a) $\det\left(E^i(\alpha)\right) = \alpha$, 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 [**369**] for proof that Definition 2 is equivalent to (48). See also Exs. 48–49 below.

3.5. *Volume.*

Exercises and examples.

Ex. 44. *Elementary operations.*

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

$$E^i(\alpha)^{-1} = E^i(1/\alpha) , \quad E^{ij}(\beta)^{-1} = E^{ij}(-\beta) , \quad (E^{ij})^{-1} = E^{ij} . \tag{49}$$

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

$$E^{ij} = E^i(-1)E^{ji}(1)E^{ij}(-1)E^{ji}(1) . \tag{50}$$

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

Ex. 45. Describe a recursive method for listing all $n!$ permutations in S_n.
Hint: If π is a permutation in S_{n-1}, mapping $\{1, 2, \dots, n-1\}$ to

$$\{\pi(1), \pi(2), \dots, \pi(n-1)\} , \tag{51}$$

then π gives rise to n permutations in S_n obtained by placing n in the "gaps" $\{\sqcup\pi(1) \sqcup \pi(2) \sqcup \dots \sqcup \pi(n-1)\sqcup\}$ of (51).

Ex. 46. *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

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

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

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

EA is a Hermite normal form of A.

Ex. 47. If $A \in \mathbb{C}^{m\times m}$ is nonsingular, then the permutation matrix P in (46) can be taken as the identity (i.e., permutation is unnecessary). Therefore $E = A^{-1}$ and

$$A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} \ . \tag{52}$$

(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

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

and illustrate (52).

Ex. 48. (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 (50) and Definition 2).
(c) If A is nonsingular, and given as product (52) of elementary matrices, then

$$\det A = \det\left(E_1^{-1}\right) \det\left(E_2^{-1}\right) \cdots \det\left(E_{k-1}^{-1}\right) \det\left(E_k^{-1}\right) \ . \tag{53}$$

(d) Use (53) to compute the determinant of A in Ex. 47(b).

Ex. 49. (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

$$A\mathbf{x} = \mathbf{b} \tag{54}$$

is given by

$$x_j = \frac{\det A[j \leftarrow \mathbf{b}]}{\det A} \ , \ j \in \overline{1,n} \ . \tag{55}$$

PROOF. (Robinson [**1291**]). 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

$$\det A \ \det I_n[j \leftarrow \mathbf{x}] = \det A[j \leftarrow \mathbf{b}] \ . \tag{56}$$

Then (55) follows from (56) since

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

□

See an extension of Cramer's rule in Corollary 5.8.

CHAPTER 1

Existence and Construction of Generalized Inverses

1. The Penrose equations

In 1955 Penrose [**1177**] showed that, for every finite matrix A (square or rectangular) of real or complex elements, there is a unique matrix X satisfying the four equations (that we call the *Penrose equations*)

$$AXA = A \ , \tag{1}$$

$$XAX = X \ , \tag{2}$$

$$(AX)^* = AX \ , \tag{3}$$

$$(XA)^* = XA \ , \tag{4}$$

where A^* denotes the conjugate transpose of A. Because this unique generalized inverse had previously been studied (though defined in a different way) by E. H. Moore [**1087**],[**1088**], it is commonly known as the *Moore–Penrose inverse*, and is often denoted by $A^\dagger$.

If A is nonsingular, it is clear that $X = A^{-1}$ trivially satisfies the four equations. Since the Moore–Penrose inverse is known to be unique (as we shall prove shortly) it follows that the Moore–Penrose inverse of a nonsingular matrix is the same as the ordinary inverse.

Throughout this book we shall be much concerned with generalized inverses that satisfy some, but not all, of the four Penrose equations. As we shall wish to deal with a number of different subsets of the set of four equations, we need a convenient notation for a generalized inverse satisfying certain specified equations. Let $\mathbb{C}^{m\times n}$ [$\mathbb{R}^{m\times n}$] denote the class of $m \times n$ complex [real] matrices.

Definition 1. For any $A \in \mathbb{C}^{m\times n}$, let $A\{i, j, \dots, k\}$ denote the set of matrices $X \in \mathbb{C}^{n\times m}$ which satisfy equations $(i), (j), \cdots, (k)$ from among the equations (1)–(4). A matrix $X \in A\{i, j, \dots, k\}$ is called[1] an $\{i, j, \dots, k\}$–*inverse* of A, and also denoted by $A^{(i,j,\dots,k)}$.

[1]Some writers have adopted descriptive names to designate various classes of generalized inverses. However there is a notable lack of uniformity and consistency in the use of these terms by different writers. Thus, $X \in A\{1\}$ is called a *generalized inverse* (Rao [**1241**]), *pseudoinverse* (Sheffield [**1347**]), *inverse* (Bjerhammar [**174**]). $X \in A\{1, 2\}$ is called a *semi–inverse* (Frame [**508**]), *reciprocal inverse* (Bjerhammar), *reflexive generalized inverse* (Rohde [**1297**]). $X \in A\{1, 2, 3\}$ is called a weak generalized inverse (Goldman and Zelen [**544**]). $X \in A\{1, 2, 3, 4\}$ is called the *general reciprocal* (Moore [**1087, 1088**]), *generalized inverse* (Penrose [**1177**]), *pseudoinverse* (Greville [**579**]), the *Moore–Penrose inverse* (Ben-Israel and Charnes [**126**]). In view of this diversity of terminology, the unambiguous notation adopted here is considered preferable. This notation also emphasizes the lack of uniqueness of many of the generalized inverses considered.

In Chapter 4 we shall extend the scope of this notation by enlarging the set of four matrix equations to include several further equations, applicable only to square matrices, that will play an essential role in the study of generalized inverses having spectral properties.

Exercises.

EX. 1. If $A\{1,2,3,4\}$ is nonempty, then it consists of a single element (Penrose [**1177**]).

PROOF. Let $X, Y \in A\{1,2,3,4\}$. Then

$$\begin{aligned} X &= X(AX)^* = XX^*A^* = X(AX)^*(AY)^* \\ &= XAY = (XA)^*(YA)^*Y = A^*Y^*Y \\ &= (YA)^*Y = Y\ . \end{aligned}$$

□

EX. 2. By means of a (trivial) example, show that $A\{2,3,4\}$ is nonempty.

2. Existence and construction of $\{1\}$–inverses

It is easy to construct a $\{1\}$–inverse of the matrix $R \in \mathbb{C}^{m\times n}r$ given by

$$R = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} \tag{0.45}$$

For any $L \in \mathbb{C}^{(n-r)\times(m-r)}$, the $n \times m$ matrix

$$S = \begin{bmatrix} I_r & O \\ O & L \end{bmatrix}$$

is a $\{1\}$–inverse of (0.45). If R is of full column [row] rank, the two lower [right–hand] submatrices are interpreted as absent.

The construction of $\{1\}$–inverses for an arbitrary $A \in \mathbb{C}^{m\times n}$ is simplified by transforming A into a Hermite normal form, as shown in the following theorem, where E is the product of elementary matrices (0.47), and P is a permutation matrix.

THEOREM 1. Let $A \in \mathbb{C}_r^{m\times n}$, and let $E \in \mathbb{C}_m^{m\times m}$ and $P \in \mathbb{C}_n^{n\times n}$ be such that

$$EAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}\ . \tag{0.46}$$

Then for any $L \in \mathbb{C}^{(n-r)\times(m-r)}$, the $n \times m$ matrix

$$X = P \begin{bmatrix} I_r & O \\ O & L \end{bmatrix} E \tag{5}$$

is a $\{1\}$–inverse of A. The partitioned matrices in (0.46) and (5) must be suitably interpreted in case $r = m$ or $r = n$.

PROOF. Rewriting (0.46) as

$$A = E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1} , \tag{6}$$

it is easily verified that any X given by (5) satisfies $AXA = A$. □

In the trivial case of $r = 0$, when A is therefore the $m \times n$ null matrix, *any* $n \times m$ matrix is a $\{1\}$–inverse.

We note that since P and E are both nonsingular, the rank of X as given by (5) is the rank of the partitioned matrix in the right member. In view of the form of the latter matrix,

$$\operatorname{rank} X = r + \operatorname{rank} L . \tag{7}$$

Since L is arbitrary, it follows that a $\{1\}$–inverse of A exists having any rank between r and $\min\{m, n\}$, inclusive (see also Fisher [**494**]).

Theorem 1 shows that every finite matrix with elements in the complex field has a $\{1\}$–inverse, and suggests how such an inverse can be constructed.

Exercises.

EX. 3. What is the Hermite normal form of a nonsingular matrix A? In this case, whet is the matrix E, and what is its relationship to A? What is the permutation matrix P? What is the matrix X given by (5)?

EX. 4. An $m \times n$ matrix A has all its elements equal to 0 except for the (i, j)th element, which is 1. What is the Hermite normal form? Show that E can be taken as a permutation matrix. What are the simplest choices of E and P? (By "simplest" we mean having the smallest number of elements different from the corresponding elements of the unit matrix of the same order.) Using these choices of E and P, but regarding L as entirely arbitrary, what is the form of the resulting matrix X given by (5)? Is this X the most general $\{1\}$–inverse of A? (See Exercise 6, Introduction.)

EX. 5. Show that every square matrix has a nonsingular $\{1\}$–inverse.

3. Properties of $\{1\}$–inverses

Certain properties of $\{1\}$–inverses are given in Lemma 1. For a given matrix A, we denote any $\{1\}$–inverse by $A^{(1)}$. Note that, in general, $A^{(1)}$ is not a uniquely defined matrix (see Ex. 8 below). For any scalar λ we define $\lambda^\dagger$ by

$$\lambda^\dagger = \begin{cases} \lambda^{-1} & (\lambda \neq 0) \\ 0 & (\lambda = 0) \end{cases} . \tag{8}$$

It will be recalled that a square matrix E is called *idempotent* if $E^2 = E$. Idempotent matrices are intimately related to generalized inverses, and their properties are considered in some detail in Chapter 2.

Lemma 1. Let $A \in \mathbb{C}_r^{m\times n}$, $\lambda \in \mathbb{C}$. Then,
(a) $(A^{(1)})^* \in A^*\{1\}$.
(b) If A is nonsingular, $A^{(1)} = A^{-1}$ uniquely (see also Ex. 7 below).
(c) $\lambda^\dagger A^{(1)} \in (\lambda A)\{1\}$.
(d) $\operatorname{rank} A^{(1)} \geq \operatorname{rank} A$.
(e) If S and T are nonsingular, $T^{-1}A^{(1)}S^{-1} \in SAT\{1\}$.
(f) $AA^{(1)}$ and $A^{(1)}A$ are idempotent and have the same rank as A.

Proof. These are immediate consequences of the defining relation (1); (d) and the latter part of (f) depend on the fact that the rank of a product of matrices does not exceed the rank of any factor. □

If an $m \times n$ matrix A is of full column rank, its $\{1\}$–inverses are its left inverses. If it is of full row rank, its $\{1\}$–inverses are its right inverses.

Lemma 2. Let $A \in \mathbb{C}_r^{m\times n}$. Then,
(a) $A^{(1)}A = I_n$ if and only if $r = n$.
(b) $AA^{(1)} = I_m$ if and only if $r = m$.

Proof. (a) *If*: Let $A \in \mathbb{C}_r^{m\times n}$. Then the $n \times n$ matrix $A^{(1)}A$ is, by Lemma 1(f), idempotent and nonsingular. Multiplying $(A^{(1)}A)^2 = A^{(1)}A$ by $(A^{(1)}A)^{-1}$ gives $A^{(1)}A = I_n$.
Only if: $A^{(1)}A = I_n \implies \operatorname{rank} A^{(1)}A = n \implies \operatorname{rank} A = n$, by Lemma 1(f).
(b) Similarly proved. □

Exercises and examples.

Ex. 6. *Computing a $\{1\}$–inverse.* This is demonstrated for the matrix A of Ex. 0.46, using (5) with E as computed in Ex. 0.46 and an arbitrary $L \in \mathbb{C}^{(n-r)\times(m-r)}$. A permutation matrix P such that

$$EAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}$$

is

$$P = [e_2\ e_4\ e_1\ e_3\ e_5\ e_6]$$

where e_j denotes the jth column of I_6. (A different permutation of the last $n - r = 4$ columns of P results in a different permutation of the columns of K.) Thus,

$$EAP = \left[\begin{array}{cc|cccc} 1 & 0 & 0 & \frac{1}{2} & 1-2i & -\frac{1}{2}i \\ 0 & 1 & 0 & 0 & 2 & 1+i \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We take

$$L = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix}$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{C}$, since $m = 3$, $n = 6$, $r = 2$. Equation (5) then gives

$$X = P \begin{bmatrix} I_r & O \\ O & L \end{bmatrix} E$$

$$= \left[\begin{array}{cc:cccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc:c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hdashline 0 & 0 & \alpha \\ 0 & 0 & \beta \\ 0 & 0 & \gamma \\ 0 & 0 & \delta \end{array}\right] \begin{bmatrix} -\frac{1}{2}i & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ i & \frac{1}{3} & 1 \end{bmatrix} = \begin{bmatrix} i\alpha & \frac{1}{3}\alpha & \alpha \\ -\frac{1}{2}i & 0 & 0 \\ i\beta & \frac{1}{3}\beta & \beta \\ 0 & -\frac{1}{3} & 0 \\ i\gamma & \frac{1}{3}\gamma & \gamma \\ i\delta & \frac{1}{3}\delta & \delta \end{bmatrix}$$

Note that, in general, the scalars $i\alpha, i\beta, i\gamma, i\delta$ are not pure imaginaries since $\alpha, \beta, \gamma, \delta$ are complex.

Ex. 7. Let $A = FHG$ where F is of full column rank and G is of full row rank. Then $\operatorname{rank} A = \operatorname{rank} H$. (*Hint*: Use Lemma 2.)

4. 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 $c_1\,\underline{\text{th}}$, $c_2\,\underline{\text{th}}$,…,$c_r\,\underline{\text{th}}$–columns of A, where the $\{c_j : j \in \overline{1,r}\}$ are as in Definition 0.1 (Willner [**1601**]). To see this, let P_1 denote the submatrix consisting of the first r columns of the permutation matrix P of (0.46) and (5). Then, because of the way in which these r columns of P were chosen,

$$EAP_1 = \begin{bmatrix} I_r \\ O \end{bmatrix} . \tag{9}$$

Now, AP_1 is an $m \times r$ matrix, and is of rank r, since RHS(9) is of rank r. But AP_1 is merely the submatrix of A consisting of the c_1th, c_2th, $\dots$,c_rth columns.

It follows from (6) that the columns of the $n \times (n-r)$ matrix

$$P \begin{bmatrix} -K \\ I_{n-r} \end{bmatrix} \tag{10}$$

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

$$EA = \begin{bmatrix} G \\ O \end{bmatrix} , \tag{11}$$

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

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

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

As an example, for the matrix A of Exs. 0.46 and 6, we note that in its Hermite normal form EA (as exhibited in the left–hand portion of the matrix T_2 of Ex. 0.46) 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 (10) with K and P computed as in Ex. 6, we find that the columns of the following matrix form a basis for $N(A)$:

$$P \begin{bmatrix} -K \\ I_{n-r} \end{bmatrix} = \left[\begin{array}{cc:cccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} 0 & -\frac{1}{2} & -1+2i & \frac{1}{2}i \\ 0 & 0 & -2 & -1-i \\ \cdots & \cdots & \cdots\cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

Exercises.

Ex. 8. Show that A is nonsingular if and only if it has a unique $\{1\}$–inverse, which then coincides with A^{-1}.

PROOF. For any $\mathbf{x} \in N(A)$ $[\mathbf{y} \in N(A^*)]$, adding $\mathbf{x}$ $[\mathbf{y}^*]$ to any column [row] of an $X \in A\{1\}$ gives another $\{1\}$–inverse of A. The uniqueness of the $\{1\}$–inverse is therefore equivalent to

$$N(A) = \{\mathbf{0}\}\ ,\ N(A^*) = \{\mathbf{0}\}\ ,$$

i.e., to the nonsingularity of A. □

EX. 9. Show that if $A^{(1)} \in A\{1\}$, then $R(AA^{(1)}) = R(A)$, $N(AA^{(1)}) = N(A)$, and $R((A^{(1)}A)^*) = R(A^*)$.

PROOF. We have

$$R(A) \supset R(AA^{(1)}) \supset R(AA^{(1)}A) = R(A)\ ,$$

from which the first result follows.

Similarly,

$$N(A) \subset N(A^{(1)}A) \subset N(AA^{(1)}A) = N(A)$$

yields the second equation.

Finally, by Lemma 1(a),

$$R(A^*) \supset R(A^*(A^{(1)})^*) = R((A^{(1)}A)^*) \supset R(A^*(A^{(1)})^*A^*) = R(A^*)\ .$$

□

EX. 10. More generally, show that $R(AB) = R(A)$ if and only if $\operatorname{rank} AB = \operatorname{rank} A$, and $N(AB) = N(B)$ if and only if $\operatorname{rank} AB = \operatorname{rank} B$.

PROOF. Evidently, $R(A) \supset R(AB)$, and these two subspaces are identical if and only if they have the same dimension. But, the rank of any matrix is the dimension of its range.

Similarly, $N(B) \subset N(AB)$. Now, the nullity of any matrix is the dimension of its null space, and also the number of columns minus the rank. Thus, $N(B) = N(AB)$ if and only if B and AB have the same nullity, which is equivalent, in this case, to having the same rank, since the two matrices have the same number of columns. □

EX. 11. The answer to the last question in Ex. 4 indicates that, for particular choices of E and P, one does not get all the $\{1\}$–inverses of A merely by varying L in (5). Note, however, that Theorem 1 does not require P to be a permutation matrix. Could one get all the $\{1\}$–inverses by considering all nonsingular P and Q such that

$$QAP = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\ ? \tag{12}$$

Given $A \in \mathbb{C}_r^{m\times n}$, show that $X \in A\{1\}$ if and only if

$$X = P \begin{bmatrix} I_r & O \\ O & L \end{bmatrix} Q \tag{13}$$

for some L and for some nonsingular P and Q satisfying (12).

PROOF. If (12) and (13) hold, X is a $\{1\}$–inverse of A by Theorem 1.

On the other hand, let $AXA = A$. Then, both AX and XA are idempotent and of rank r, by Lemma 1(f). Since any idempotent matrix E satisfies $E(E-I) = O$, its only eigenvalues are 0 and 1. Thus, the Jordan canonical forms of both AX and XA are of the form

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\ ,$$

being of orders m and n, respectively. Therefore, there exist nonsingular P and R such that

$$R^{-1}AXR = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\ ,\ P^{-1}XAP = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\ .$$

Thus,

$$\begin{aligned} R^{-1}AP &= R^{-1}AXAXAP = (R^{-1}AXR)R^{-1}AP(P^{-1}XAP) \\ &= \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} R^{-1}AP \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} . \end{aligned}$$

It follows that $R^{-1}AP$ is of the form

$$R^{-1}AP = \begin{bmatrix} H & O \\ O & O \end{bmatrix} ,$$

where $H \in \mathbb{C}_r^{r\times r}$, i.e., nonsingular. Let

$$Q = \begin{bmatrix} H^{-1} & O \\ O & I_{m-r} \end{bmatrix} R^{-1} .$$

Then (12) is satisfied. Consider the matrix $P^{-1}XQ^{-1}$. We have,

$$\begin{aligned} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} (P^{-1}XQ^{-1}) &= (QAP)(P^{-1}XQ^{-1}) = QAXQ^{-1} \\ &= \begin{bmatrix} H^{-1} & O \\ O & I_{m-r} \end{bmatrix} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \begin{bmatrix} H & O \\ O & I_{m-r} \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} (P^{-1}XQ^{-1}) \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} &= (P^{-1}XQ^{-1})(QAP) = P^{-1}XAP \\ &= \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} . \end{aligned}$$

From the latter two equations it follows that

$$P^{-1}XQ^{-1} = \begin{bmatrix} I_r & O \\ O & L \end{bmatrix}$$

for some L. But this is equivalent to (13). □

5. Existence and construction of $\{1,2\}$–inverses

It was first noted by Bjerhammar [**174**] that the existence of a $\{1\}$–inverse of a matrix A implies the existence of a $\{1,2\}$–inverse. This easily verified observation is stated as a lemma for convenience of reference.

Lemma 3. Let $Y, Z \in A\{1\}$, and let

$$X = YAZ .$$

Then $X \in A\{1,2\}$.

Since the matrices A and X occur symmetrically in (1) and (2), $X \in A\{1,2\}$ and $A \in X\{1,2\}$ are equivalent statements, and in either case we can say that A and X are $\{1,2\}$–inverses *of each other*.

From (1) and (2) and the fact that the rank of a product of matrices does not exceed the rank of any factor, it follows at once that if A and X are $\{1,2\}$–inverses of each other, they have the same rank. Less obvious is the fact, first noted by Bjerhammar [**174**], that if X is a $\{1\}$–inverse of A and of the same rank as A, it is a $\{1,2\}$–inverse of A.

Theorem 2. (Bjerhammar) Given A and $X \in A\{1\}$, $X \in A\{1,2\}$ if and only if $\operatorname{rank} X = \operatorname{rank} A$.

PROOF. *If*: Clearly $R(XA) \subset R(X)$. But $\operatorname{rank} XA = \operatorname{rank} A$ by Lemma 1(f), and so, if $\operatorname{rank} X = \operatorname{rank} A$, $R(XA) = R(X)$ by Ex. 10. Thus,

$$XAY = X$$

for some Y. Premultiplication by A gives

$$AX = AXAY = AY \ ,$$

and therefore

$$XAX = X \ .$$

Only if: This follows at once from (1) and (2). □

An equivalent statement is the following:

COROLLARY 1. Any two of the following three statement imply the third:

$$\begin{aligned} X &\in A\{1\} \ , \\ X &\in A\{2\} \ , \\ \operatorname{rank} X &= \operatorname{rank} A \ . \end{aligned}$$ □

In view of Theorem 2, (7) shows that the $\{1\}$–inverse obtained from the Hermite normal form is a $\{1,2\}$–inverse if we take $L = O$. In other words,

$$X = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} E$$

is a $\{1,2\}$–inverse of A is E and O are nonsingular and satisfy (0.46).

6. Existence and construction of $\{1,2,3\}$–, $\{1,2,4\}$– and $\{1,2,3,4\}$–inverses

Just as Bjerhammar [**174**] showed that the existence of a $\{1\}$–inverse implies the existence of a $\{1.2\}$–inverse, Urquhart [**1480**] has shown that the existence of a $\{1\}$–inverse of every finite matrix with elements in $\mathbb{C}$ implies the existence of a $\{1,2,3\}$–inverse and a $\{1,2,4\}$–inverse of every such matrix. However, in order to show the nonemptiness of $A\{1,2,3\}$ and $A\{1,2,4\}$ for any given A, we shall utilize the $\{1\}$–inverse not of A itself but of a related matrix. For that purpose we shall need the following lemma.

LEMMA 4. For any finite matrix A,

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

PROOF. If $A \in \mathbb{C}^{m\times n}$, both A and AA^* have m rows. Now, the rank of any m–rowed matrix is equal to m minus the number of independent linear relations among its rows. To show that $\operatorname{rank} AA^* = \operatorname{rank} A$, it is sufficient, therefore, to show that every linear relation among the rows of A holds for the corresponding rows of AA^*, and vice versa. Any nontrivial linear relation among the rows of a matrix H is equivalent to the existence of a nonzero row vector $\mathbf{x}^*$ such that $\mathbf{x}^*H = \mathbf{0}$. Now evidently,

$$\mathbf{x}^*A = \mathbf{0} \implies \mathbf{x}^*AA^* = \mathbf{0} \ ,$$

and, conversely,

$$\begin{aligned} \mathbf{x}^*AA^* = \mathbf{0} &\implies \mathbf{0} = \mathbf{x}^*AA^*\mathbf{x} = (A^*\mathbf{x})^*A^*\mathbf{x} \\ &\implies A^*\mathbf{x} = \mathbf{0} \implies \mathbf{0} = (A^*\mathbf{x})^* = \mathbf{x}^*A \ . \end{aligned}$$

Here we have used the fact that, for any column vector $\mathbf{y}$ of complex elements $\mathbf{y}^*\mathbf{y}$ is the sum of squares of the absolute values of the elements, and this sum vanishes only if every element is zero.

Finally, applying this result to the matrix A^* gives $\operatorname{rank} A^*A = \operatorname{rank} A^*$, and, of course, $\operatorname{rank} A^* = \operatorname{rank} A$. □

COROLLARY 2. For any finite matrix A, $R(AA^*) = R(A)$ and $N(AA^*) = N(A)$.

PROOF. This follows from Lemma 4 and Ex. 10. □

Using the preceding lemma, we can now prove the following theorem.

THEOREM 3. (Urquhart) For every finite matrix A with complex elements,

$$Y = (A^*A)^{(1)}A^* \in A\{1,2,3\} \tag{14}$$

and

$$Z = A^*(AA^*)^{(1)} \in A\{1,2,4\} \ . \tag{15}$$

PROOF. Applying Corollary 2 to A^* gives

$$R(A^*A) = R(A^*) \ ,$$

and so,

$$A^* = A^*AU \tag{16}$$

for some U. Taking conjugate transpose gives

$$A = U^*A^*A \ . \tag{17}$$

Consequently,

$$AYA = U^*A^*A(A^*A)^{(1)}A^*A = U^*A^*A = A \ .$$

Thus, $Y \in A\{1\}$. But $\operatorname{rank} Y \geq \operatorname{rank} A$ by Lemma 1(d), and $\operatorname{rank} Y \leq \operatorname{rank} A^* = \operatorname{rank} A$ by the definition of Y. Therefore

$$\operatorname{rank} Y = \operatorname{rank} A \ ,$$

and, by Theorem 2, $Y \in A\{1,2\}$. Finally, (16) and (17) give

$$AY = U^*A^*A(A^*A)^{(1)}A^*AU = U^*A^*AU \ ,$$

which is clearly Hermitian. Thus, (14) is established.

Relation (15) is similarly proved. □

A $\{1,2\}$–inverse of a matrix A is, of course, a $\{2\}$–inverse, and similarly, a $\{1,2,3\}$–inverse is also a $\{1,3\}$–inverse and a $\{2,3\}$–inverse. Thus, if we can establish the existence of a $\{1,2,3,4\}$–inverse, we will have demonstrated the existence of an $\{i,j,\dots,k\}$–inverse for all possible choices of one, two or three integers $i,j,\dots,k$ from the set $\{1,2,3,4\}$. It was shown in Ex. 1 that if a $\{1,2,3,4\}$–inverse exists, it is unique. We know, as a matter of fact, that it does exist, because it is the well–known Moore–Penrose inverse, $A^\dagger$. However, we have not yet proved this. This is done in the next theorem.

THEOREM 4. (Urquhart) For any finite matrix A of complex elements,

$$A^{(1,4)}AA^{(1,3)} = A^\dagger \tag{18}$$

PROOF. Let X denote LHS(18). It follows at once from Lemma 3 that $X \in A\{1,2\}$. Moreover, (18) gives

$$AX = AA^{(1,3)} \ , \ XA = A^{(1,4)}A \ .$$

But, both $AA^{(1,3)}$ and $A^{(1,4)}A$ are Hermitian, by the definition of $A^{(1,3)}$ and $A^{(1,4)}$. Thus

$$X \in A\{1,2,3,4\} \ .$$

However, by Ex. 1, $A\{1,2,3,4\}$ contains at most a single element. Therefore, it contains exactly one element, which we denote by $A^\dagger$, and $X = A^\dagger$. □

7. Full–rank factorization

A non–null matrix that is not of full (column or row) rank 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 5. 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

$$A = FG \ . \tag{19}$$

Proof. 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 (19), since every column of A is uniquely representable as a linear combination of the columns of F. Finally, $\operatorname{rank} G = r$, since

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

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 (19).

We shall call a factorization (19) with the properties stated in Lemma 5 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. Nevertheless, the lemma still holds in this case.

A full–rank factorization of any matrix is easily read off from its Hermite normal form. Indeed, it was pointed out in Section 4 above that the first r rows of the Hermite normal form EA (i.e., the rows of the matrix G of (11)) form a basis for the space spanned by the rows of A. Thus, this G can serve also as the matrix G of (19). Consequently, (19) holds for some F. As in Section 4, 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 (19) on ths right by P_1 gives

$$F = AP_1 \ ,$$

and so (19) becomes

$$A = (AP_1)G \ , \tag{20}$$

where P_1 and G are as in Section 4. (Indeed it was already noted there that the columns of AP_1 are a basis for $R(A)$.)

For example, for the matrix A of Exs. 0.46 and 6, (20) gives

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

8. Explicit formula for $A^\dagger$

C. C. MacDuffee apparently was the first to point out, about 1959, that a full–rank factorization of a matrix A leads to an explicit formula for its Moore–Penrose inverse, $A^\dagger$. However, he did so in private communications, and there is no published work that can be cited.

Theorem 5. (MacDuffee). If $A \in \mathbb{C}_r^{m\times n}$, $r > 0$, has a full–rank factorization

$$A = FG \ , \tag{21}$$

then

$$A^\dagger = G^*(F^*AG^*)^{-1}F^* \ . \tag{22}$$

PROOF. First, we must show that F^*AG^* is nonsingular. By (21),

$$F^*AG^* = (F^*F)(GG^*) , \tag{23}$$

and both factors of the right member are $r \times r$ matrices. Also, by Lemma 4, both are of rank r. Thus, F^*AG^* is the product of two nonsingular matrices, and therefore nonsingular. Moreover, (23) gives

$$(F^*AG^*)^{-1} = (G^*G)^{-1}(F^*F)^{-1} .$$

Denoting by X the right member of (22), we now have

$$X = G^*(GG^*)^{-1}(F^*F)^{-1}F^* , \tag{24}$$

and it is easily verified that this expression for X satisfies the Penrose equations (1)–(4). As $A^\dagger$ is the sole element of $A\{1,2,3,4\}$, (22) is therefore established. $\square$

Exercises.

EX. 12. Theorem 5 provides an alternative proof of the existence of the $\{1,2,3,4\}$–inverse (previously established by Theorem 4). However, Theorem 5 excludes the case $r = 0$. Complete the alternative existence proof by showing that if $r = 0$, (2) has a unique solution for X, and this X satisfies (1), (3) and (4).

EX. 13. Compute $A^\dagger$ for the matrix A of Exs. 0.46 and 6.

EX. 14. What is the most general $\{1,2\}$–inverse of the special matrix A of Ex. 4? What is its Moore–Penrose inverse?

EX. 15. Show that if $A = FG$ is a full–rank factorization, then

$$A^\dagger = G^\dagger F^\dagger .$$

EX. 16. Show that for every matrix A,

(a) $(A^\dagger)^\dagger = A$ (b) $(A^*)^\dagger = (A^\dagger)^*$
(c) $(A^T)^\dagger = (A^\dagger)^T$ (d) $A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger$

EX. 17. If $\mathbf{a}$ and $\mathbf{b}$ are column vectors, then

(a) $\mathbf{a}^\dagger = (\mathbf{a}^*\mathbf{a})^\dagger\mathbf{a}^*$ (b) $(\mathbf{a}\mathbf{b}^*)^\dagger = (\mathbf{a}^*\mathbf{a})^\dagger(\mathbf{b}^*\mathbf{b})^\dagger\mathbf{b}\mathbf{a}^*$.

EX. 18. Show that if H is Hermitian and idempotent, $H^\dagger = H$.

EX. 19. Show that $H^\dagger = H$ if and only if H^2 is Hermitian and idempotent and $\operatorname{rank} H^2 = \operatorname{rank} H$.

EX. 20. If $D = \operatorname{diag}(d_1, d_2, \dots , d_n)$, show that $D^\dagger = \operatorname{diag}\left(d_1^\dagger, d_2^\dagger, \dots , d_n^\dagger\right)$.

EX. 21. If U and V are unitary matrices, show that

$$(UAV)^\dagger = V^*A^\dagger U^*$$

for any matrix A for which the product UAV is defined.

9. Construction of $\{2\}$–inverses of prescrived rank

Following the proof of Theorem 1, we desribed A.G. Fisher's construction of a $\{1\}$–inverse of a given $A \in \mathbb{C}_r^{m\times n}$ having any prescribed rank between r and $\min(m, n)$, inclusive. From (2) it is easily deduced that

$$\operatorname{rank} A^{(2)} \le r .$$

We note also that the $n \times m$ null matrix is a $\{2\}$–inverse of rank 0, and any $A^{(1,2)}$ is a $\{2\}$–inverse of rank r, by Theorem 2. For $r > 1$, is there a construction analogous to Fisher's for a $\{2\}$–inverse of rank s for arbitrary s between 0 anr r? Using the principle of full–rank factorization, we can readily answer the question in the affirmative.

Let $X_0 \in A\{1,2\}$ have a full–rank factorization

$$X_0 = YZ .$$

Then, $Y \in \mathbb{C}_r^{m\times r}$ and $Z \in \mathbb{C}_r^{r\times n}$, and (2) becomes

$$YZAYZ = YZ\ .$$

In view of Lemma 2, multiplication on the left by $Y^{(1)}$ and on the right by $Z^{(1)}$ gives (see Stewart [**1400**])

$$ZAY = I_r\ . \tag{25}$$

Let Y_s denote the submatrix of Y consisting of the first s columns, and Z_s the submatrix of Z consisting of the first s rows. Then, both Y_s and Z_s are of full rank s, and it follows from (25) that

$$Z_sAY_s = I_s\ . \tag{26}$$

Now, let

$$X_s = Y_sZ_s\ .$$

Then, $\operatorname{rank} X_s = s$, by Ex. 7 and (26) gives

$$X_sAX_s = X_s\ .$$

Exercises.

EX. 22. For

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

find elements of $A\{2\}$ of ranks 1,2, and 3, respectively.

EX. 23. With A as in Ex. 22, find a {2}–inverse of rank 2 having zero elements in the last two rows and the last two columns.

EX. 24. Show that there is at most one matrix X satisfying the three equations $AX = B$, $XA = D$, $XAX = X$ (Cline; see Cline and Greville [**356**]).

EX. 25. Let $A = FG$ be a full–rank factorization of $A \in \mathbb{C}_r^{m\times n}$, i.e., $F \in \mathbb{C}_r^{m\times r}$, $G \in \mathbb{C}_r^{r\times n}$. Then

(a) $G^{(i)}F^{(1)} \in A\{i\}\ ,\ (i = 1, 2, 4)\ ,$ (b) $G^{(1)}F^{(j)} \in A\{j\}\ ,\ (j = 1, 2, 3)\ .$

PROOF.
(a) $i = 1$:

$$FGG^{(1)}F^{(1)}FG = FG\ ,$$

since

$$F^{(1)}F = GG^{(1)} = I_r$$

by Lemma 2.
$i = 2$:

$$G^{(2)}F^{(1)}FGG^{(2)}F^{(1)} = G^{(2)}F^{(1)}$$

since

$$F^{(1)}F = I_r\ ,\ G^{(2)}GG^{(2)} = G^{(2)}\ .$$

$i = 4$:

$$G^{(4)}F^{(1)}FG = G^{(4)}G = (G^{(4)}G)^*\ .$$

(b) Similarly proved, with the roles of F and G interchanged. □

Ex. 26. Let A, F, G be as in Ex. 25. Then

$$A^\dagger = G^\dagger F^{(1,3)} = G^{(1,4)} F^\dagger \ .$$

10. An application of $\{2\}$–inverses in iterative methods for solving nonlinear equations

One of the best–known methods for solving a single equation in a single variable, say

$$f(x) = 0 \ , \tag{27}$$

is *Newton's* (also *Newton–Raphson*) method

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \ , \ (k = 0, 1, \dots) \ . \tag{28}$$

Under suitable conditions on the function f and the initial approximation x_0, the sequence (28) converges to a solution of (27); see, e.g., Ortega and Rheinboldt [**1153**]. The *modified Newton method* uses the iteration

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_0)} \ , \ (k = 0, 1, \dots) \ . \tag{29}$$

instead of (28).

Newton's method for solving a system of m equations in n variables

$$\begin{array}{rcl} f_1(x_1, \dots, x_n) & = & 0 \\ \cdots & & \\ f_m(x_1, \dots, x_n) & = & 0 \end{array} \qquad \text{or} \quad \mathbf{f}(\mathbf{x}) = \mathbf{0} \tag{30}$$

is similarly given, for the case $m = n$, by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{f}'(\mathbf{x}_k)^{-1} \mathbf{f}(x_k) \ , \ (k = 0, 1, \dots) \ , \tag{31}$$

where $\mathbf{f}'(\mathbf{x}_k)$ is the *derivative* of $\mathbf{f}$ at $\mathbf{x}_k$, represented by the matrix of partial derivatives (the *Jacobian* matrix)

$$\mathbf{f}'(\mathbf{x}_k) = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}_k) \right) \ . \tag{32}$$

The reader is referred to the excellent texts by Ortega and Rheinboldt [**1153**] and Rall [**1236**], for iterative methods in nonlinear analysis, and in particular, for the many variations and extensions of Newton's method (31).

If the nonsingularity of $\mathbf{f}'(\mathbf{x}_k)$ cannot be assumed for every $\mathbf{x}_k$, and in particular, if the number of equations (30) is different from the number of unknowns, then it is natural to inquire whether a generalized inverse of $\mathbf{f}'(\mathbf{x}_k)$ can be used in (31), still resulting in a sequence converging to a solution of (30).

In this section we illustrate the use of $\{2\}$–inverses in a modified Newton method (Theorem 6 below) and in a Newton method (Ex. 27 below) for solving the nonlinear equations (30). Other applications of generalized inverses in the iterative methods of nonlinear analysis are Leach [**920**], Altman [**19**] and [**20**], Ben-Israel [**106**], [**107**] and [**113**], Rheinboldt ([**1268**] especially Theorem 3.5), and Fletcher [**495**].

Readers not familiar with vector and matrix norms used in this section may consult Exs. 0.6–0.32 for a brief introduction to norms.

Throughout this section we denote by $\| \ \|$ both a given (but arbitrary) vector norm in $\mathbb{C}^n$, and a matrix norm in $\mathbb{C}^{m \times n}$ consistent with it; see, e.g.., Ex. 0.28. For a given point $\mathbf{x}_0 \in \mathbb{C}^n$ and a positive scalar r we denote by

$$B(\mathbf{x}_0, r) = \{\mathbf{x} \in \mathbb{C}^n : \| \mathbf{x} - \mathbf{x}_0 \| < r\}$$

the *open ball* with *center* $\mathbf{x}_0$ and *radius* r. The *closed ball* with the same center and radius is

$$\overline{B(\mathbf{x}_0, r)} = \{\mathbf{x} \in \mathbb{C}^n : \| \mathbf{x} - \mathbf{x}_0 \| \leq r\} ,$$

THEOREM 6. Let the following be given:

$$\begin{aligned} &\mathbf{x}_0 \in \mathbb{C}^n \ , \ r > 0 \ , \\ &\mathbf{f} : B(\mathbf{x}_0, r) \to \mathbb{C}^m \text{ a function;} , \\ &A \in \mathbb{C}^{m\times n} \ , \ T \in \mathbb{C}^{n\times m} \text{ matrices} , \\ &\epsilon > 0 \ , \ \delta > 0 \text{ positive scalars} , \end{aligned}$$

such that:

$$\begin{aligned} \| \mathbf{f}(\mathbf{u}) - \mathbf{f}(\mathbf{v}) - A(\mathbf{u} - \mathbf{v}) \| &\leq \epsilon \| \mathbf{u} - \mathbf{v} \| \\ &\qquad \text{for all } \mathbf{u}, \mathbf{v} \in B(\mathbf{x}_0, r) , && (33) \\ TAT &= T , && (34) \\ \epsilon \| T \| = \delta &< 1 , && (35) \\ \| T \| \| \mathbf{f}(\mathbf{x}_0) \| &< (1 - \delta) r . && (36) \end{aligned}$$

Then the sequence

$$\mathbf{x}_{k+1} = \mathbf{x}_k - T\,\mathbf{f}(\mathbf{x}_k) \tag{37}$$

converges to a point

$$\mathbf{x}_\infty \in \overline{B(\mathbf{x}_0, r)} \tag{38}$$

satisfying

$$T\mathbf{f}(\mathbf{x}) = \mathbf{0} . \tag{39}$$

PROOF. Using induction on k we prove that the sequence (37) satisfies for $k = 0, 1, \ldots$

$$\begin{aligned} \mathbf{x}_k &\in B(\mathbf{x}_0, r) , && (40) \\ \| \mathbf{x}_{k+1} - \mathbf{x}_k \| &\leq \delta^k (1 - \delta) r . && (41) \end{aligned}$$

We denote by $(40.k)$ and $(41.k)$ the validity of (40) and (41), respectively, for the given value of k. Now (41.0) and (40.1) follow from (36). Assuming $(41.j)$ for $0 \leq j \leq k-1$ we get

$$\| \mathbf{x}_k - \mathbf{x}_0 \| \leq \sum_{j=0}^{k-1} \| \mathbf{x}_{j+1} - \mathbf{x}_j \| \leq (1 - \delta) r \sum_{j=0}^{k-1} \delta^j = (1 - \delta^k) r ,$$

which proves $(40.k)$. To prove $(41.k)$ we write

$$\begin{aligned} \mathbf{x}_{k+1} - \mathbf{x}_k &= -T\mathbf{f}(\mathbf{x}_k) \\ &= -T\mathbf{f}(\mathbf{x}_{k-1}) - T\left[\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{x}_{k-1})\right] \\ &= T\left[A(\mathbf{x}_k - \mathbf{x}_{k-1} - \mathbf{f}(\mathbf{x}_k) + \mathbf{f}(\mathbf{x}_{k-1})\right] \ , \text{ by (34) and (37)} . \end{aligned}$$

From (33) and (35) it therefore follows that

$$\begin{aligned} \| \mathbf{x}_{k+1} - \mathbf{x}_k \| &\leq \| T \| \| \mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{x}_{k-1}) - A(\mathbf{x}_k - \mathbf{x}_{k-1} \| \\ &\leq \delta \| \mathbf{x}_k - \mathbf{x}_{k-1} \| , \end{aligned}$$

proving $(41.k)$. □

REMARK 2. $\mathbf{f}$ is *differentiable* at $\mathbf{x}_0$ and the linear transformation A is its *derivative* at $\mathbf{x}_0$, if

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\| \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - A(\mathbf{x} - \mathbf{x}_0) \|}{\| \mathbf{x} - \mathbf{x}_0 \|} = 0 .$$

Comparing this with (33) we conclude that the linear transformation A in Theorem 6 is an "approximate derivative," and can be chosen as the derivative of $\mathbf{f}$ at $\mathbf{x}_0$ if $\mathbf{f}$ is continuously differentiable in $B(\mathbf{x}_0, r)$.

REMARK 3. The limit $\mathbf{x}_\infty$ of the sequence (37) is a solution of (39), but in general not of (30), unless T is of full column rank in which case (39) and (30) are equivalent. Thus, the choice of the $\{2\}$–inverse T in Theorem 6, which by Section 9 can have any rank between 0 and rank A, will determine the extent to which $\mathbf{x}_\infty$ can be called a solution of (30). The "worst" choice of T is the trivial choice $T = O$, in which case any $\mathbf{x}$ is a solution of (39) and the iterations (37) stop at $\mathbf{x}_0$.

REMARK 4. For any nontrivial T, the inequality (36) bounds the value of $\mathbf{f}$ at $\mathbf{x}_0$ as follows

$$\| \mathbf{f}(\mathbf{x}_0) \| < \frac{(1-\delta)r}{\| T \|} .$$

REMARK 5. Note that (33) needs to hold only for $\mathbf{u}, \mathbf{v} \in B(\mathbf{x}_0, r)$ such that $\mathbf{u} - \mathbf{v} \in R(T)$, and the limit $\mathbf{x}_\infty$ of (37) lies in

$$\overline{B(\mathbf{x}_0, r)} \cap \{\mathbf{x}_0 + R(T)\} .$$

Exercises.

EX. 27. *A Newton's method using* $\{2\}$*–inverses.* Let the following be given

$$\begin{aligned}
&\mathbf{x}_0 \in \mathbb{C}^n \ , \ r > 0 \ , \\
&\mathbf{f} : B(\mathbf{x}_0, r) \to \mathbb{C}^m \text{ a function;} , \\
&\epsilon > 0 \ , \ \delta > 0 \ , \ \eta > 0 \text{ positive scalars} ,
\end{aligned}$$

and for any $\mathbf{x} \in \overline{B(\mathbf{x}_0, r)}$ let

$$A_{\mathbf{x}} \in \mathbb{C}^{m \times n} \ , \ T_{\mathbf{x}} \in \mathbb{C}^{n \times m}$$

be matrices satisfying for all $\mathbf{u}, \mathbf{v} \in B(\mathbf{x}_0, r)$:

$$\begin{aligned}
\| \mathbf{f}(\mathbf{u}) - \mathbf{f}(\mathbf{v}) - A_{\mathbf{v}}(\mathbf{u} - \mathbf{v}) \| &\le \epsilon \| \mathbf{u} - \mathbf{v} \| && (42)\\
T_{\mathbf{u}} A_{\mathbf{u}} T_{\mathbf{u}} &= T_{\mathbf{u}} , && (43)\\
\| (T_{\mathbf{u}} - T_{\mathbf{v}})\mathbf{f}(\mathbf{v}) \| &\le \eta \| \mathbf{u} - \mathbf{v} \| , && (44)\\
\epsilon \| T_{\mathbf{u}} \| + \eta \le \delta &< 1 , && (45)\\
\| T_{\mathbf{x}_0} \| \| \mathbf{f}(\mathbf{x}_0) \| &< (1-\delta) r . && (46)
\end{aligned}$$

Then the sequence

$$\mathbf{x}_{k+1} = \mathbf{x}_k - T_{\mathbf{x}_k} \mathbf{f}(\mathbf{x}_k) \quad (k = 0, 1, \dots) \tag{47}$$

converges to a point

$$\mathbf{x}_\infty \in \overline{B(\mathbf{x}_0, r)} \tag{38}$$

which is a solution of

$$T_{\mathbf{x}_\infty} \mathbf{f}(\mathbf{x}) = \mathbf{0} . \tag{48}$$

PROOF. As in the proof of Theorem 6 we use induction on k to prove that the sequence (47) satisfies

$$\begin{aligned}
\mathbf{x}_k &\in B(\mathbf{x}_0, r) , && (40.k)\\
\| \mathbf{x}_{k+1} - \mathbf{x}_k \| &\le \delta^k (1-\delta) r . && (41.k)
\end{aligned}$$

Again (41.0) and (40.1) follow from (46), and assuming ($41.j$) for $0 \leq j \leq k-1$ we get ($40.k$). To prove ($41.k$) we write

$$\begin{aligned}
\mathbf{x}_{k+1} - \mathbf{x}_k &= -T_{\mathbf{x}_k}\mathbf{f}(\mathbf{x}_k) \\
&= \mathbf{x}_k - \mathbf{x}_{k-1} - T_{\mathbf{x}_k}\mathbf{f}(\mathbf{x}_k) + T_{\mathbf{x}_{k-1}}\mathbf{f}(\mathbf{x}_{k-1}) \ , \ \text{by (47)} \ , \\
&= T_{\mathbf{x}_{k-1}}A_{\mathbf{x}_{k-1}}(\mathbf{x}_k - \mathbf{x}_{k-1}) - T_{\mathbf{x}_k}\mathbf{f}(\mathbf{x}_k) + T_{\mathbf{x}_{k-1}}\mathbf{f}(\mathbf{x}_{k-1}) \ , \\
&\qquad \text{since } TAT = T \text{ implies } TA\mathbf{x} = \mathbf{x} \text{ for every } \mathbf{x} \in R(T) \\
&= T_{\mathbf{x}_{k-1}}\left[A_{\mathbf{x}_{k-1}}(\mathbf{x}_k - \mathbf{x}_{k-1}) - \mathbf{f}(\mathbf{x}_k) + \mathbf{f}(\mathbf{x}_{k-1})\right] + (T_{\mathbf{x}_{k-1}} - T_{\mathbf{x}_k})\mathbf{f}(\mathbf{x}_k) \ .
\end{aligned}$$

Therefore

$$\begin{aligned}
\| \mathbf{x}_{k+1} - \mathbf{x}_k \| &\leq (\epsilon \| T_{\mathbf{x}_{k-1}} \| + \eta) \| \mathbf{x}_k - \mathbf{x}_{k-1} \| \ , \ \text{by (42) and (44)} \ , \\
&\leq \delta \| \mathbf{x}_k - \mathbf{x}_{k-1} \| \ , \ \text{by (45)} \ ,
\end{aligned}$$

which proves ($41.k$). □

Suggested further reading

Section 2. Rao [**1241**], Sheffield [**1347**].
Section 3. Rao ([**1240**], [**1243**]).
Section 5. Deutsch [**400**], Frame [**508**], Greville [**586**], Hartwig [**666**], Przeworska–Rolewicz and Rolewicz [**1209**].
Section 6. Hearon and Evans [**711**], Rao [**1243**], Sibuya [**1355**].
Section 7. Hartwig [**670**].
Section 10. See also Burmeister [**244**], Fletcher [**497**], Golub and Pereyra [**555**].

CHAPTER 2

Linear Systems and Characterization of Generalized Inverses

1. Solutions of linear systems

As already indicated in Section 3, Introduction, the principal application of $\{1\}$–inverses is to the solution of linear systems, where they are used in much the same way as ordinary inverses in the nonsingular case. The main result of this section is the following theorem of Penrose [**1177**], to whom the proof is also due.

THEOREM 1. Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, $D \in \mathbb{C}^{m\times q}$. Then the matrix equation

$$AXB = D \tag{1}$$

is consistent if and only if for some $A^{(1)}, B^{(1)}$,

$$AA^{(1)}DB^{(1)}B = D \ , \tag{2}$$

in which case the general solution is

$$X = A^{(1)}DB^{(1)} + Y - A^{(1)}AYBB^{(1)} \tag{3}$$

for arbitrary $Y \in \mathbb{C}^{n\times p}$.

PROOF. If (2) holds, then $X = A^{(1)}DB^{(1)}$ is a solution of (1). Conversely, if X is any solution of (1), then

$$D = AXB = AA^{(1)}AXBB^{(1)}B = AA^{(1)}DB^{(1)}B \ .$$

Moreover, it follows from (2) and the definition of $A^{(1)}$ and $B^{(1)}$ that every matrix X of the form (3) satisfies (1). On the other hand, let X be any solution of (1). Then, clearly

$$X = A^{(1)}DB^{(1)} + X - A^{(1)}AXBB^{(1)} \ ,$$

which is of the form (3). $\square$

The following characterization of the set $A\{1\}$ in terms of an arbitrary element $A^{(1)}$ of the set is due essentially to Bjerhammar [**174**].

COROLLARY 1. Let $A \in \mathbb{C}^{m\times n}$, $A \in A\{1\}$. Then

$$A\{1\} = \{A^{(1)} + Z - A^{(1)}AZAA^{(1)} : \ Z \in \mathbb{C}^{n\times m}\} \tag{4}$$

PROOF. The set described in RHS(4) is obtained by writing $Y = A^{(1)} + Z$ in the set of solutions of $AXA = A$ as given by Theorem 1. $\square$

Specializing Theorem 1 to ordinary systems of linear equations gives:

COROLLARY 2. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^{m}$. Then the equation

$$A\mathbf{x} = \mathbf{b} \tag{5}$$

is consistent if and only if for some $A^{(1)}$

$$AA^{(1)}\mathbf{b} = \mathbf{b} \ , \tag{6}$$

in which case the general solution of (5) is

$$\mathbf{x} = A^{(1)}\mathbf{b} + (I - A^{(1)}A)\mathbf{y} \tag{7}$$

for arbitrary $\mathbf{y} \in \mathbb{C}^{n}$. $\square$

The following theorem appears in the doctoral dissertation of C. A. Rohde [**1296**], who attributes it to R. C. Bose. It is an alternative characterization of $A\{1\}$.

THEOREM 2. Let $A \in \mathbb{C}^{m\times n}$, $X \in \mathbb{C}^{n\times m}$. Then $X \in A\{1\}$ if and only if for all $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ is consistent, $\mathbf{x} = X\mathbf{b}$ is a solution.

PROOF. *If*: Let $\mathbf{a}_j$ denote the jth column of A. Then

$$A\mathbf{x} = \mathbf{a}_j$$

is consistent, and $X\mathbf{a}_j$ is a solution, i.e.,

$$AX\mathbf{a}_j = \mathbf{a}_j \quad (j \in \overline{1,n}) \ .$$

Therefore

$$AXA = A \ .$$

Only if: This follows from (6). □

Exercises and examples.

EX. 1. Show that the general solution of $A\mathbf{x} = \mathbf{b}$, where A is the matrix of Ex. 0.46 and

$$\mathbf{b} = \begin{bmatrix} 14+5i \\ -15+3i \\ 10-15i \end{bmatrix}$$

can be written in the form

$$\mathbf{x} = \begin{bmatrix} 0 \\ \frac{5}{2} - 7i \\ 0 \\ 5-i \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & -1+2i & \frac{1}{2}i \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1-i \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}$$

where $y_1, y_2, \ldots, y_6$ are arbitrary.

EX. 2. *Kronecker products.* The *Kronecker product* $A \otimes B$ of the two matrices $A = (a_{ij}) \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$ is the $mp \times nq$ matrix expressible in partitioned form as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

The properties of this product (e.g., Marcus and Minc [**996**]) include

$$(A \otimes B)^* = A^* \otimes B^* \ , \ (A \otimes B)^T = A^T \otimes B^T \ , \tag{8}$$

and

$$(A \otimes B)(P \otimes Q) = AP \otimes BQ \tag{9}$$

for every A, B, P, Q for which the above products are refined.

An important application of the Kronecker product is rewriting a matrix equation

$$AXB = D \tag{1}$$

as a vector equation. For any $X = (x_{ij}) \in \mathbb{C}^{m\times n}$, let the vector $\mathrm{vec}(X) = (v_k) \in \mathbb{C}^{mn}$ be the vector obtained by listing the elements of X by rows. In other words,

$$v_{n(i-1)+j} = x_{ij} \quad (i \in \overline{1,m}; j \in \overline{1,n})$$

For example,

$$\operatorname{vec}\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{bmatrix}1\\2\\3\\4\end{bmatrix}$$

The energetic reader should now verify that

$$\operatorname{vec}(AXB) = (A \otimes B^T)\operatorname{vec}(X)\ . \tag{10}$$

By using (10), the matrix equation (1) can be rewritten as the vector equation

$$(A \otimes B^T)\operatorname{vec}(X) = \operatorname{vec}(D) \tag{11}$$

Theorem 1 must therefore be equivalent to Corollary 2 applied to the vector equation (11). To demonstrate this we need the following two results

$$A^{(1)} \otimes B^{(1)} \ \in\ (A \otimes b)\{1\} \quad \text{(follows from (9))}\ , \tag{12}$$

$$(A^{(1)})^T \ \in\ A^T\{1\}\ . \tag{13}$$

Now (1) is consistent if and only if (11) is consistent, and the latter statement

$$\begin{aligned}
&\iff (A \otimes B^T)(A \otimes B^T)^{(1)}\operatorname{vec}(D) = \operatorname{vec}(D) \quad \text{(by Corollary 2)}\\
&\iff (A \otimes B^T)(A^{(1)} \otimes (B^{(1)})^T)\operatorname{vec}(D) = \operatorname{vec}(D) \quad \text{(by (12),(13))}\\
&\iff (AA^{(1)} \otimes (B^{(1)}B)^T)\operatorname{vec}(D) = \operatorname{vec}(D) \quad \text{(by (9))}\\
&\iff AA^{(1)}DB^{(1)}B = D \quad \text{(by (10))}\ .
\end{aligned}$$

The other statements of Theorem 1 can be shown similarly to follow from their counterparts in Corollary 2. The two results are thus equivalent.

Ex. 3. $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$ (Greville [**581**]).

Proof. Upon replacing A by $A \otimes B$ and X by $A^\dagger \otimes B^\dagger$ in (1.1)–(1.4) and making use of (8) and (9), it is easily verified that (1.1)–(1.4) are satisfied. □

Ex. 4. The matrix equations

$$AX = B\ ,\ XD = E \tag{14}$$

have a common solution if and only if each equation separately has a solution and

$$AE = BD\ .$$

Proof. (Penrose [**1177**]). *If*: For any $A^{(1)}, D^{(1)}$,

$$X = A^{(1)}B + ED^{(1)} - A^{(1)}AED^{(1)}$$

is a common solution of both equations (14) provided $AE = BD$ and

$$AA^{(1)}B = B\ ,\ ED^{(1)}D = E\ .$$

By Theorem 1, the latter two equations are equivalent to the consistency of equations (14) considered separately.

Only if: Obvious. □

Ex. 5. Let equations (14) have a common solution $X_0 \in \mathbb{C}^{m\times n}$. Then, show that the general solution is

$$X = X_0 + (I - A^{(1)}A)Y(I - DD^{(1)}) \tag{15}$$

for arbitrary $A^{(1)} \in A\{1\}, D^{(1)} \in D\{1\}, Y \in \mathbb{C}^{m\times n}$.

Hint: First, show that RHS(15) is a common solution. Then, if X is any common solution, evaluate RHS(15) for $Y = X - X_0$.

2. Characterization of $A\{1,3\}$ and $A\{1,4\}$

The set $A\{1\}$ is completely characterized in Corollary 1. Let us now turn our attention to $A\{1,3\}$. The key to its characterization is the following theorem.

Theorem 3. The set $A\{1,3\}$ consists of all solutions for X of

$$AX = AA^{(1,3)} , \tag{16}$$

where $A^{(1,3)}$ is an arbitrary element of $A\{1,3\}$.

Proof. If X satisfies (16), then clearly

$$AXA = AA^{(1,3)}A = A ,$$

and, moreover, AX is Hermitian since $AA^{(1,3)}$ is Hermitian by definition. Thus, $X \in A\{1,3\}$.

On the other hand, if $X \in A\{1,3\}$, then

$$\begin{aligned} AA^{(1,3)} &= AXAA^{(1,3)} = (AX)^*AA^{(1,3)} = X^*A^*(A^{(1,3)})^*A^* \\ &= X^*A^* = AX , \end{aligned}$$

where we have used Lemma 1.1(a). □

Corollary 3. Let $A \in \mathbb{C}^{m\times n}$, $A^{(1,3)} \in A\{1,3\}$. Then

$$A\{1,3\} = \{A^{(1,3)} + (I - A^{(1,3)}A)Z : Z \in \mathbb{C}^{n\times m}\} . \tag{17}$$

Proof. Applying Theorem 1 to (16) and substituting $Z + A^{(1,3)}$ for Y gives (17). □

The following theorem and its corollary are obtained in a manner analogous to the proofs of Theorem 3 and Corollary 3.

Theorem 4. The set $A\{1,4\}$ consists of all solutions for X of

$$XA = A^{(1,4)}A .$$

Corollary 4. Let $A \in \mathbb{C}^{m\times n}$, $A^{(1,4)} \in A\{1,4\}$. Then

$$A\{1,4\} = \{A^{(1,4)} + Y(I - AA^{(1,4)}) : Y \in \mathbb{C}^{n\times m}\} .$$

Other characterizations of $A\{1,3\}$ and $A\{1,4\}$ based on their least squares properties will be given in Chapter 3.

Exercises.

Ex. 6. Prove Theorem 4 and Corollary 4.

Ex. 7. If A is the matrix of Ex. 0.46, show that $A\{1,3\}$ is the set of matrices X of the form

$$X = \frac{1}{38}\begin{bmatrix} 0 & 0 & 0 \\ -10i & 3 & 9 \\ 0 & 0 & 0 \\ 2i & -12 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & -1+2i & \frac{1}{2}i \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1-i \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} Z ,$$

where Z is an arbitrary element of $\mathbb{C}^{6\times 3}$.

Ex. 8. For the matrix A of Ex. 0.46, show that $A\{1,4\}$ is the set of matrices Y of the form

$$Y = \frac{1}{276}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 20-18i & 42 \\ 0 & 10-9i & 21 \\ 0 & -29-9i & -9-27i \\ 0 & -2+4i & 24+30i \\ 0 & -29+30i & -36+3i \end{bmatrix} + Z\begin{bmatrix} 1 & -\frac{1}{3}i & -i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} ,$$

where Z is an arbitrary element of $\mathbb{C}^{6\times 3}$.

Ex. 9. Using Theorem 1.4 and the results of Exs. 7 and 8, calculate $A^\dagger$. (Since *any* $A^{(1,4)}$ and $A^{(1,3)}$ will do, choose the simplest.)

Ex. 10. Give an alternative proof of Theorem 1.4, using Theorem 3 and 4. (*Hint*: Take $X = A^\dagger$).

Ex. 11. By applying Ex. 5 show that if $A \in \mathbb{C}^{m\times n}$ and $A^{(1,3,4)} \in A\{1,3,4\}$, then

$$A\{1,3,4\} = \{A^{(1,3,4)} + (I - A^{(1,3,4)}A)Y(I - AA^{(1,3,4)}) : Y \in \mathbb{C}^{n\times m}\} .$$

Ex. 12. Show that if $A \in \mathbb{C}^{m\times n}$ and $A^{(1,2,3)} \in A\{1,2,3\}$, then

$$A\{1,2,3\} = \{A^{(1,2,3)} + (I - A^{(1,2,3)}A)ZA^{(1,2,3)} : Z \in \mathbb{C}^{n\times m}\} .$$

Ex. 13. Similarly, show that if $A \in \mathbb{C}^{m\times n}$ and $A^{(1,2,4)} \in A\{1,2,4\}$, then

$$A\{1,2,4\} = \{A^{(1,2,4)} + A^{(1,2,4)}Z(I - AA^{(1,2,4)}) : Z \in \mathbb{C}^{m\times m}\} .$$

3. Characterization of $A\{2\}$, $A\{1,2\}$ and other subsets of $A\{2\}$.

Since

$$XAX = X \tag{1.2}$$

involves X nonlinearly, a characterization of $A\{2\}$ is not obtained by merely applying Theorem 1. However, such a characterization can be reached by using a full–rank factorization of X. The rank of X will play an important role, and it will be convenient to let $A\{i,j,\dots,k\}_s$ denote the subset of $A\{i,j,\dots,k\}$ consisting of matrices of rank s.

We remark that the sets $A\{2\}_0$, $A\{2,3\}_0$, $A\{2,4\}_0$ and $A\{2,3,4\}_0$ are identical and contain a single element. For $A \in \mathbb{C}^{m\times n}$ this sole element is the $n \times m$ matrix of zeros. Having thus disposed of the case of $s = 0$, we shall consider only positive s in the remainder of this section.

The following theorem has been stated by G. W. Stewart [**1400**], who attributes it to R. E. Funderlic.

Theorem 5. Let $A \in \mathbb{C}_r^{m\times n}$ and $0 < s \le r$. Then

$$A\{2\}_s = \{YZ : Y \in \mathbb{C}^{n\times s},\ Z \in \mathbb{C}^{s\times m},\ ZAY = I_s\} . \tag{18}$$

Proof. Let

$$X = YZ , \tag{19}$$

where the conditions on Y and Z in RHS(18) are satisfied. Then Y and Z are of rank s, and X is of rank s by Ex. 1.7. Moreover,

$$XAX = YZAYZ = YZ = X .$$

On the other hand, let $X \in A\{2\}_s$, and let (19) be a full–rank factorization. Then $Y \in \mathbb{C}_s^{n\times s}$, $Z \in \mathbb{C}_s^{s\times m}$ and

$$YZAYZ = YZ . \tag{20}$$

Moreover, if $Y^{(1)}$ and $Z^{(1)}$ are any $\{1\}$–inverses, then by Lemma 1.2

$$Y^{(1)}Y = ZZ^{(1)} = I_s .$$

Thus, multiplying (20) on the left by $Y^{(1)}$ and on the right by $Z^{(1)}$ gives

$$ZAY = I_s .$$

$\square$

Corollary 5. Let $A \in \mathbb{C}_r^{m\times n}$. Then

$$A\{1,2\} = \{YZ : Y \in \mathbb{C}^{n\times r},\ Z \in \mathbb{C}^{r\times m},\ ZAY = I_r\} .$$

Proof. By Theorem 1.2,

$$A\{1,2\} = A\{2\}_r \ .$$

□

The relation $ZAY = I_s$ of (18) implies that $Z \in (AY)\{1,2,4\}$. This remark suggests the approach to the characterization of $A\{2,3\}$ on which the following theorem is based.

Theorem 6. Let $A \in \mathbb{C}_r^{m\times n}$ and $0 < s \leq r$. Then

$$A\{2,3\}_s = \{Y(AY)^\dagger : AY \in \mathbb{C}_s^{m\times s}\} \ .$$

Proof. Let $X = Y(AY)^\dagger$, where$AY \in \mathbb{C}_s^{m\times s}$. Then we have

$$AX = AY(AY)^\dagger \ . \tag{21}$$

The right member is Hermitian by (1.3), and

$$XAX = Y(AY)^\dagger AY(AY)^\dagger = Y(AY)^\dagger = X \ .$$

Thus, $X \in A\{2,3\}$. Finally, since $X \in A\{2\}$, $A \in X\{1\}$, and (21) and Lemma 1.1(f) give

$$s = \operatorname{rank} AY = \operatorname{rank} AX = \operatorname{rank} X \ .$$

On the other hand, let $X \in A\{2,3\}_s$. Then AX is Hermitian and idempotent, and is of rank s by Lemma 1.1(f), since $A \in X\{1\}$. By Ex. 1.18

$$(AX)^\dagger = AX \ ,$$

and so

$$X(AX)^\dagger = XAX = X \ .$$

Thus X is of the form described in the theorem. □

The following theorem is proved in an analogous fashion.

Theorem 7. Let $A \in \mathbb{C}_r^{m\times n}$ and $0 < s \leq r$. Then

$$A\{2,4\}_s = \{(YA)^\dagger Y : YA \in \mathbb{C}_s^{s\times m}\} \ .$$

Exercises and examples.

Ex. 14. Could Theorem 6 be sharpened by replacing $(AY)^\dagger$ by $(AY)^{(i,j,k)}$ for some i,j,k? (Which properties are actually used in the proof?) Note that AY is of full column rank; what bearing, if nay, does this have on the answer to the question?

Ex. 15. Show that if $A \in \mathbb{C}_r^{m\times n}$,

$$\begin{aligned} A\{1,2,3\} &= \{Y(AY)^\dagger : AY \in \mathbb{C}_r^{m\times r}\} \\ A\{1,2,4\} &= \{(YA)^\dagger Y : YA \in \mathbb{C}_r^{r\times m}\} \ . \end{aligned}$$

(Compare these results with Exs. 12 and 13.)

Ex. 16. The characterization of $A\{2,3,4\}$ is more difficult, and will be postponed until later in this chapter. Show, however, that if $\operatorname{rank} A = 1$, $A\{2,3,4\}$ contains exactly two elements, $A^\dagger$ and O.

4. Idempotent matrices and projectors

A comparison of Eq. (1) of the Introduction with Lemma 1.1(f) suggests that the role played by the unit matrix in connection with the ordinary inverse of a nonsingular matrix is, in a sense, assumed by idempotent matrices in relation to generalized inverses. As the properties of idempotent matrices are likely to be treated in a cursory fashion in an introductory course in linear algebra, some of them are listed in the following lemma.

Lemma 1. Let $E \in \mathbb{C}^{n\times n}$ be idempotent. Then:
(a) E^* and $I - E$ are idempotent.
(b) The eigenvalues of E are 0 and 1. The multiplicity of the eigenvalue 1 is $\operatorname{rank} E$.
(c) $\operatorname{rank} E = \operatorname{trace} E$.
(d) $E(I - E) = (I - E)E = O$.
(e) $E\mathbf{x} = \mathbf{x}$ if and only if $\mathbf{x} \in R(E)$.
(f) $E \in E\{1, 2\}$.
(g) $N(E) = R(I - E)$.

Proof. Parts (a) to (f) are immediate consequences of the definition of idempotency: (c) follows from (b) and the fact that the trace of any square matrix is the sum of its eigenvalues counting multiplicities; (g) is obtained by applying Corollary 2 to the equation $E\mathbf{x} = \mathbf{0}$. □

Lemma 2. (Langenhop [**910**]). Let a square matrix have the full–rank factorization

$$E = FG\ .$$

Then E is idempotent if and only if $GF = I$.

Proof. If $GF = I$, then clearly

$$(FG)^2 = FGFG = FG\ . \tag{22}$$

On the other hand, since F is of full column rank and G is of full row rank,

$$F^{(1)}F = GG^{(1)} = I$$

by Lemma 1.2. Thus if (22) holds, multiplication on the left by $F^{(1)}$ and on the right by $G^{(1)}$ gives $GF = I$. □

Let $P_{L,M}$ denote the transformation that carries any $\mathbf{x} \in \mathbb{C}^n$ into its projection on L along M, see § 0.1.3. It is easily verified that this transformation is linear (see Ex. 0.26). We shall call the transformation $P_{L,M}$ the *projector on L along M*.

It is well known (see, e.g., Halmos [**645**]) that every linear transformation from one finite–dimensional vector space to another can be represented by a matrix, which is uniquely determined by the linear transformation and by the choice of bases for the spaces involved. Except where otherwise specified, the basis for any finite–dimensional vector space, used in this book, is the standard basis of unit vectors. Having thus fixed the bases, there is a one–to–one correspondence between $\mathbb{C}^{m\times n}$, the *$m \times n$ complex matrices*, and $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$, the *space of linear transformations mapping $\mathbb{C}^n$ into $\mathbb{C}^m$*. This correspondence permits using the same symbol, say A, to denote both the linear transformation $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ and its matrix representation $A \in \mathbb{C}^{m\times n}$. Thus the matrix–vector equation

$$A\mathbf{x} = \mathbf{y} \quad (A \in \mathbb{C}^{,\times \mathbf{x}} \in \mathbb{C}^n, \mathbf{y} \in \mathbb{C}^m)$$

cab equally be regarded as a statement that the linear transformation A maps $\mathbf{x}$ into $\mathbf{y}$. The notion of a matrix as representing a linear transformation has appeared before, see § 0.2.5 (p. 12) and Exs. 0.13–23, and it will be utilized in an important way in Chapter 6.

In particular, linear transformations mapping $\mathbb{C}^n$ into itself are represented by the square matrices of order n. Specializing further, the next theorem establishes a one–to–one correspondence between the idempotent matrices of order n and the projectors $P_{L,M}$ where $L \oplus M = \mathbb{C}^n$. Moreover, for any two complementary subspaces L and M, a method for computing $P_{L,M}$ is given by (27) below.

Theorem 8. For every idempotent matrix $E \in \mathbb{C}^{n\times n}$, $R(E)$ and $N(E)$ are complementary subspaces with

$$E = P_{R(E),N(E)} \ . \tag{23}$$

Conversely, if L and M are complementary subspaces, there is a unique idempotent $P_{L,M}$ such that $R(P_{L,M}) = L$, $N(P_{L,M}) = M$.

Proof. Let E be idempotent of order n. Then it follows from Lemma 1(e) and 1(g) and from the equation

$$\mathbf{x} = E\mathbf{x} + (I - E)\mathbf{x} \tag{24}$$

the $\mathbb{C}^n$ is the sum of $R(E)$ and $N(E)$. Moreover, $R(E) \cap N(E) = \{\mathbf{0}\}$, since

$$E\mathbf{x} = (I - E)\mathbf{y} \implies E\mathbf{x} = E^2\mathbf{x} = E(I - E)\mathbf{y} = \mathbf{0} \ ,$$

by Lemma 1(d). Thus, $R(E)$ and $N(E)$ are complementary, and (24) shows that, for every $\mathbf{x}$, $E\mathbf{x}$ is the projection of $\mathbf{x}$ on $R(E)$ along $N(E)$. This establishes (23).

On the other hand let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots , \mathbf{x}_l\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots , \mathbf{y}_m\}$ be any two bases for L and M, respectively. Then, $P_{L,M}$ if it exists, is uniquely determined by

$$\begin{cases} P_{L,M}\,\mathbf{x}_i = \mathbf{x}_i & (i \in \overline{1,l}) \\ P_{L,M}\,\mathbf{y}_i = \mathbf{0} & (i \in \overline{1,m}) \end{cases} \ . \tag{25}$$

Let $X = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_l]$ denote the matrix whose columns are the vectors $\mathbf{x}_i$. Similarly, let $Y = [\mathbf{y}_1\ \mathbf{y}_2\ \cdots\ \mathbf{y}_m]$. Then (25) is equivalent to

$$P_{L,M}\,[X\ Y] = [X\ O] \tag{26}$$

Since $[X\ Y]$ is nonsingular, the unique solution of (26), and therefore of (25), is

$$P_{L,M} = [X\ O][X\ Y]^{-1} \ . \tag{27}$$

Since (25) implies

$$P_{L,M}\,[X\ O] = [X\ O] \ ,$$

$P_{L,M}$ as given by (27) is clearly idempotent. $\square$

The relation between the direct sum (1) and the projector[1] $P_{L,M}$ is given in the following.

Corollary 6. Let L and M be complementary subspaces of $\mathbb{C}^n$. Then, for every $\mathbf{x} \in \mathbb{C}^n$, the unique decomposition (0.2) is given by

$$P_{L,M}\,\mathbf{x} = \mathbf{y} \ , \ (I - P_{L,M})\,\mathbf{x} = \mathbf{z} \ .$$

---

[1] Our use of the term "projector" to denote either the linear transformation $P_{L,M}$ or its idempotent matrix representation is not standard in the literature. Many writers have used "projection" in the same sense. The latter usage, however, seems to us to lead to undesirable ambiguity, since "projection" also describes the image $P_{L,M}\,\mathbf{x}$ of the vector $\mathbf{x}$ under the transformation $P_{L,M}$. The use of "projection" in the sense of "image" is clearly much older (e.g., in elementary geometry) than its use in the sense of "transformation". "Projector" describes more accurately than "projection" what is meant here, and has been used in this sense by Afriat [**6**], de Boor [**193**], Bourbaki ([**213**, Ch. I, Def. 6, p. 16],[**214**, Ch. VIII, Section 1]), Greville [**578**], Przeworska–Rolewicz and Rolewicz [**1209**], Schwerdtfeger [**1326**] and Ward, Boullion and Lewis [**1536**]. Still other writers use "projector" to designate the orthogonal projector to be discussed in Section 6. This is true of Householder [**753**], Yosida [**1623**], Kantorovich and Akilov [**817**], and numerous other Russian writers. We are indebted to de Boor for several of the preceding references.

If $A^{(1)} \in A\{1\}$, we know from Lemma 1.1(f) that both $AA^{(1)}$ and $A^{(1)}A$ are idempotent, and therefore are projectors. It is of interest to find out what we can say about the subspaces associated with these projectors. In fact, we already know from Ex. 1.9 that

$$R(AA^{(1)} = R(A) \ , \ N(A^{(1)}A) = N(A) \ , \ R((A^{(1)}A)^*) = R(A^*) \ . \tag{28}$$

The following is an immediate consequence of these results.

Corollary 7. If A and X are $\{1,2\}$–inverses of each other, AX is the projector on $R(A)$ along $N(X)$, and XA is projector on $R(X)$ along $N(A)$.

An important application of projectors is to the class of diagonable matrices. (The reader will recall that a square matrix is called *diagonable* if it is similar to a diagonal matrix.) It is easily verified that a matrix $A \in \mathbb{C}^{n\times n}$ is diagonable if and only if it has n linearly independent eigenvectors. The latter fact will be used in the proof of the following theorem, which expresses an arbitrary diagonable matrix as a linear combination of projectors.

Theorem 9. (Spectral Theorem for Diagonable Matrices). Let $A \in \mathbb{C}^{n\times n}$ with k distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. Then A is diagonable if and only if there exist projectors $E_1, E_2, \dots, E_k$ such that

$$E_iE_j = O \ , \ \text{if } i \neq j \ , \tag{29}$$

$$I_n = \sum_{i=1}^{k} E_i \ , \tag{30}$$

$$A = \sum_{i=1}^{k} \lambda_i E_i \ . \tag{31}$$

Proof. *If*: For $i \in \overline{1,k}$, let $r_i = \operatorname{rank} E_i$ and let $X_i \in \mathbb{C}^{n\times r_i}$ be a matrix whose columns are a basis for $R(E_i)$. Let

$$X = [X_1 \ X_2 \ \cdots \ X_k] \ .$$

Then, by Lemma 1(c), the number of columns of X is

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} \operatorname{trace} E_i = \operatorname{trace} \sum_{i=1}^{k} E_i = \operatorname{trace} I_n = n \ ,$$

by (30). Thus X is square of order n. By the definition of X_i, there exists for each i a Y_i such that

$$E_i = X_iY_i \ .$$

Let

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \ .$$

Then

$$\begin{aligned} XY &= \sum_{i=1}^{k} \lambda_i E_i X_i = \\ &= XD \ , \end{aligned} \tag{32}$$

where

$$D = \operatorname{diag}(\lambda_1 I_1, \lambda_2 I_2, \dots, \lambda_k I_k) \ , \tag{33}$$

I_i being used to denote the unit matrix of order r_i. Since X is nonsingular, it follows from (32) that A and D are similar.
Only if: If A is diagonable,

$$AX = XD\,, \tag{34}$$

where X is nonsingular, and D can be represented in the form (33). Let X be partitioned by columns into $X_1, X_2, \dots, X_k$ in conformity with the diagonal blocks of D, and for $i = 1, 2, \dots, k$, let

$$E_i = [O \cdots O\ X_i\ O \cdots O\,]X^{-1}\,.$$

In other words, $E_i = \widetilde{X_i}X^{-1}$, where $\widetilde{X_i}$ denotes the matrix obtained from X by replacing all its columns except the columns of X_i by columns of zeros. It is then easily verified that E_i is idempotent, and that (29) and (30) hold. Finally,

$$\sum_{i=1}^{k} \lambda_i\, E_i = [\lambda_1 X_1\ \lambda_2 X_2\ \ \cdots\ \ \lambda_k X_k\,]X^{-1} = XDX^{-1} = A\,,$$

by (34). □

The idempotent matrices $\{E_i : i \in \overline{1,k}\}$ (shown in Ex. 24 below to be uniquely determined by the diagonable matrix A) are called its *principal idempotents*. Relation (31) is called the *spectral decomposition* of A. Further properties of this decomposition are studied in Exs. 24–26.

Note that $R(E_i)$ is the *eigenspace* of A (space spanned by the eigenvectors) associated with the eigenvalue λ_i, while because of (29), $N(E_i)$ is the direct sum of the eigenspaces associated with all eigenvalues of A other than λ_i.

Exercises and examples.

Ex. 17. Show that $I_n\{2\}$ consists of all idempotent matrices of order n.

Ex. 18. If E is idempotent, $X \in E\{2\}$ and $R(X) \subset R(E)$, show that X is idempotent.

Ex. 19. Let $E \in \mathbb{C}_r^{n\times n}$. Then E is idempotent if and only if its Jordan canonical form can be written as

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

Ex. 20. Show that $P_{L,M}\,A = A$ if and only if $R(A) \subset L$ and $AP_{L,M} = A$ if and only if $N(A) \supset M$.

Ex. 21. $AB(AB)^{(1)}A = A$ if and only if $\operatorname{rank} AB = \operatorname{rank} A$, and $B(AB)^{(1)}AB = B$ if and only if $\operatorname{rank} AB = \operatorname{rank} B$. (*Hint*: Use Exs. 20, 1.9 and 1.10.)

Ex. 22. A matrix $A \in \mathbb{C}^{n\times n}$ is diagonable if and only if it has n linearly independent eigenvectors.

Proof. Diagonability of A is equivalent to the existence of a nonsingular matrix X such that $X^{-1}AX = D$, which in turn is equivalent to $AX = XD$. But the latter equation expresses the fact that each column of X is an eigenvector of A, and X is nonsingular if and only if its columns are linearly independent. □

Ex. 23. Show that $I - P_{L,M} = P_{M,L}$.

Ex. 24. *principal idempotents.* Let $A \in \mathbb{C}^{n\times n}$ be a diagonable matrix with k distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. Then the idempotents $E_1, E_2, \dots, E_k$ satisfying (29)–(31) are uniquely determined by A.

PROOF. Let $\{F_i : i \in \overline{1,k}\}$ be any idempotent matrices satisfying

$$F_i F_j = O \ , \ \text{if } i \neq j \ , \tag{29$*$}$$

$$I_n = \sum_{i=1}^{k} F_i \ , \tag{30$*$}$$

$$A = \sum_{i=1}^{k} \lambda_i F_i \ . \tag{31$*$}$$

From (29) and (31) it follows that

$$E_i A = A E_i = \lambda_i E_i \quad (i \in \overline{1,k}) \ . \tag{35}$$

Similarly from (29∗) and (31∗)

$$F_i A = A F_i = \lambda_i F_i \quad (i \in \overline{1,k}) \tag{35$*$}$$

so that

$$E_i(AF_j) = \lambda_j E_i F_j$$

and

$$(E_i A) F_j = \lambda_i E_i F_j \ ,$$

proving that

$$E_i F_j = O \quad \text{if } i \neq j \ . \tag{36}$$

The uniqueness of $\{E_i : i \in \overline{1,k}\}$ now follows:

$$\begin{aligned} E_i &= E_i \sum_{j=1}^{k} F_j \quad \text{by (30$*$)} \\ &= E_i F_i \ , \quad \text{by (36)} \\ &= \left(\sum_{j=1}^{k} E_j \right) F_i \ , \quad \text{by (36)} \\ &= F_i \ , \quad \text{by (30)} \ . \end{aligned}$$

□

EX. 25. Let $A \in \mathbb{C}^{n \times n}$ be a diagonable matrix with k distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then the principal idempotents of A are given by

$$E_i = \frac{p_i(A)}{p_i(\lambda_i)} \quad (i \in \overline{1,k}) \ , \tag{37}$$

where

$$p_i(\lambda) = \prod_{\substack{j = 1 \\ j \neq i}}^{k} (\lambda - \lambda_j) \ . \tag{38}$$

PROOF. Let G_i $(i \in \overline{1,k})$ denote RHS(37) and let $E_1, E_2, \dots, E_k$ be the principal idempotents of A. For any $i, j \in \overline{1,k}$

$$\begin{aligned} G_i E_j &= \frac{1}{p_i(\lambda_i)} \prod_{\substack{h=1 \\ h \neq i}}^{k} (A - \lambda_h I) E_j \\ &= \frac{1}{p_i(\lambda_i)} \prod_{\substack{h=1 \\ h \neq i}}^{k} (\lambda_j - \lambda_h I) E_j \ , \text{ by (35)} \\ &= \begin{cases} O & \text{if } i \neq j \\ E_i & \text{if } i = j \end{cases} \end{aligned}$$

Therefore, $G_i = G_i \sum_{j=1}^{k} E_j = E_i \quad (i \in \overline{1,k})$. □

EX. 26. Let be a diagonable matrix with k distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ and principal idempotents $E_1, E_2, \dots, E_k$. Then:
(a) If $f(\lambda)$ is any polynomial,

$$f(A) = \sum_{i=1}^{k} f(\lambda_i) E_i \ .$$

(b) Any matrix commutes with A if and only if it commutes with every E_i $(i \in \overline{1,k})$.

PROOF. (a) Follows from (29), (30) and (31).
(b) Follows from (31) and (37) which express A as a linear combination of the $\{E_i : i \in \overline{1,k}\}$ and each E_i as a polynomial in A. □

EX. 27. Prove the following analog of Theorem 5 for $\{1\}$–inverses: Let $A \in \mathbb{C}_r^{m \times n}$ with $r < s \leq \min(m, n)$. Then

$$A\{1\}_s = \left\{ YZ : Y \in \mathbb{C}_s^{n \times s}, Z \in \mathbb{C}_s^{s \times m}, ZAY = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \right\} . \tag{39}$$

PROOF. Let $X = YZ$, where the conditions on Y and Z in RHS(39) are satisfied. Then $\operatorname{rank} X = s$ by Ex. 1.7. Let

$$Y = [Y_1 \ Y_2] \ , \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

where Y_1 denotes the first r columns of Y and Z_1 the first r rows of Z. Then (39) gives

$$Z_1 A Y_1 = I_r \ , \ Z_1 A Y_2 = O \ . \tag{40}$$

Let $X_1 = Y_1 Z_1$. Then it follows from the first equation (40) that $X_1 \in A\{2\}$. Since by Ex. 1.7, $\operatorname{rank} X_1 = r = \operatorname{rank} A$, $X_1 \in A\{1\}$ by Theorem 1.2. Thus

$$\begin{aligned} AXA &= AX_1AXA = AY_1(Z_1AY)ZA = AY_1[I_r \ O] \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} A \\ &= AY_1Z_1A = AX_1A = A \ . \end{aligned}$$

On the other hand, let $X \in A\{1\}_s$, and let $X = UV$ be a full–rank factorization. Then $U \in \mathbb{C}_s^{n \times s}$, $V \in \mathbb{C}_s^{s \times m}$, and

$$VAUVAU = VAU$$

and so VAU is idempotent, and is of rank r by Ex. 1.7. Thus, by Ex. 19, there is a nonsingular T such that

$$TVVAUT^{-1} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} .$$

If we now take

$$Y = UT^{-1} \ , \ Z = TV \ ,$$

then

$$Y \in \mathbb{C}_s^{n\times s}, Z \in \mathbb{C}_s^{s\times m} \ ,$$

$$ZAY = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \ , \text{ and } YZ = UV = X \ .$$

□

5. Generalized inverses with prescribed range and null space

Let $A \in \mathbb{C}^{m\times n}$ and let $A^{(1)}$ be an arbitrary element of $A\{1\}$. Let $R(A) = L$ and $N(A) = M$. By Lemma 1.1(f), $AA^{(1)}$ and $A^{(1)}A$ are idempotent. By (28) and Theorem 8,

$$AA^{(1)} = P_{L,S} \quad A^{(1)}A = P_{T,M} \ ,$$

where S is some subspace of $\mathbb{C}^m$ complementary to L, and T is some subspace of $\mathbb{C}^n$ complementary to M.

If we choose arbitrary subspaces S and T complementary to L and M, respectively, does there exist a $\{1\}$–inverse $A^{(1)}$ such that $N(AA^{(1)}) = S$ and $R(A^{(1)}A) = T$? The following theorem (parts of which have appeared previously in work of Robinson [**1280**], Langenhop [**910**], and Milne [**1052**]) answers the question in the affirmative.

Theorem 10. Let $A \in \mathbb{C}^{\times}, R(A) = L \ , \ N(A) = M \ , \ L \oplus S = \mathbb{C}^m$, and $M \oplus T = \mathbb{C}^n$. Then:
(a) X is a $\{1\}$–inverse of A such that $N(AX) = S$ and $R(XA) = T$ if and only if

$$AX = P_{L,S} \ , \quad XA = P_{T,M} \ . \tag{41}$$

(b) The general solution of (41) is

$$X = P_{T,M}A^{(1)}P_{L,S} + (I_n - A^{(1)}A)Y(I_m - AA^{(1)}) \ , \tag{42}$$

where $A^{(1)}$ is a fixed (but arbitrary) element of $A\{1\}$ and Y is an arbitrary element of $\mathbb{C}^{n\times m}$.
(c) $A^{(1,2)}_{T,S} = P_{T,M}A^{(1)}P_{L,S}$ is the unique $\{1,2\}$–inverse of A having range T and null space S.

Proof.
(a) The "if" part of the statement follows at once from Theorem 8 and Lemma 1(e), the "only if" part from Lemma 1.1(f), (28) and Theorem 8.
(b) By repeated use of Ex. 20, along with (28), we can easily verify that (41) is satisfied by $X = P_{T,M}A^{(1)}P_{L,S}$. The result then follows from Ex. 5.
(c) Since $P_{T,M}A^{(1)}P_{L,S}$ is a $\{1\}$–inverse of A, its rank is at least r by Lemma 1.1(d), while its rank does not exceed r, since rank $P_{L,S} = r$ by (41) and Lemma 1.1(f). Thus it has the same rank as A, and is therefore a $\{1,2\}$–inverse, by Theorem 1.2. It follows from parts (a) and (b) that it has the required range and null space.

On the other hand, a $\{1,2\}$–inverse of A having range T and null space S satisfies (41) and also

$$XAX = X \ . \tag{1.2}$$

By Ex. 1.24, these three equations have at most one common solution. □

Corollary 8. Under the hypotheses of Theorem 10, let $A^{(1)}_{T,S}$ be some $\{1\}$–inverse of A such that $R(A^{(1)}_{T,S}A) = T$, $N(AA^{(1)}_{T,S}) = S$, and let $A\{1\}_{T,S}$ denote the class of such $\{1\}$–inverses of A. Then

$$A\{1\}_{T,S} = \{A^{(1)}_{T,S} + (I_n - A^{(1)}_{T,S}A)Y(I_m - AA^{(1)}_{T,S}) : Y \in \mathbb{C}^{n\times m}\} \tag{43}$$

For a subspace L of $\mathbb{C}^m$, a complementary subspace of particular interest is the orthogonal complement, denoted by $L^\perp$, which consists of all vectors in $\mathbb{C}^m$ orthogonal to L. If in Theorem 10 we take $S = L^\perp$ and $T = M^\perp$, the class of $\{1\}$–inverses given by (43) is the class of $\{1, 3, 4\}$–inverses, and $A^{(1,2)}_{T,S} = A^\dagger$.

The formulas in Theorem 10 generally are not convenient for computational purposes. When this is the case, the following theorem (which extends results due to Urquhart [**1480**]) may be resorted to.

Theorem 11. Let $A \in \mathbb{C}^{\times}, U \in \mathbb{C}^{n\times p}$, $V \in \mathbb{C}^{q\times m}$, and

$$X = U(VAU)^{(1)}V ,$$

where $(VAU)^{(1)}$ is a fixed, but arbitrary element of $(VAU)\{1\}$. Then:
(a) $X \in A\{1\}$ if and only if $\operatorname{rank} VAU = r$.
(b) $X \in A\{2\}$ and $R(X) = R(U)$ if and only if $\operatorname{rank} VAU = \operatorname{rank} U$.
(c) $X \in A\{2\}$ and $N(X) = N(V)$ if and only if $\operatorname{rank} VAU = \operatorname{rank} V$.
(d) $X = A^{(1,2)}_{R(U),N(V)}$ if and only if $\operatorname{rank} U = \operatorname{rank} V = \operatorname{rank} VAU = r$.

Proof.
Proof of (a). *If*: We have $\operatorname{rank} AU = r$, since

$$r = \operatorname{rank} VAU \le \operatorname{rank} AU \le \operatorname{rank} A = r .$$

Therefore, by Ex. 1.10, $R(AU) = R(A)$, and so $A = AUY$ for some Y. Thus by Ex. 21,

$$AXA = AU(VAU)^{(1)}VAUY = AUY = A .$$

Only if: Since $X \in A\{1\}$,

$$A = AXAXA = AU(VAU)^{(1)}VAU(VAU)^{(1)}VA ,$$

and therefore $\operatorname{rank} VAU = \operatorname{rank} A = r$.
Proof of (b). *If*: By Ex. 21,

$$XAU = U(VAU)^{(1)}VAU = U ,$$

from which it follows that $XAX = X$, and also $\operatorname{rank} X = \operatorname{rank} U$. By Ex. 1.10, $R(X) = R(U)$.
Only if: Since $X \in A\{2\}$,

$$X = XAX = U(VAU)^{(1)}VAU(VAU)^{(1)}V .$$

Therefore

$$\operatorname{rank} X \le \operatorname{rank} VAU \le \operatorname{rank} U = \operatorname{rank} X .$$

Proof of (c). Similar to (b).
Proof of (d). Follows from (a), (b) and (c). □

Note that if we require only a $\{1\}$–inverse X such that $R(X) \subset R(U)$ and $N(X) \supset N(V)$, part (a) of the theorem is sufficient.

Theorem 11 can be used to prove the following modified analog of Theorem 10(c) for all $\{2\}$–inverses, and not merely $\{1, 2\}$–inverses.

THEOREM 12. Let $A \in \mathbb{C}_r^{m\times n}$, let T be a subspace of $\mathbb{C}^n$ of dimension $s \leq r$, and let S be a subspace of $\mathbb{C}^m$ of dimension $m - s$. Then, A has a $\{2\}$–inverse X such that $R(X) = T$ and $N(X) = S$ if and only if

$$AT \oplus S = \mathbb{C}^m \ , \tag{44}$$

in which case X is unique.

PROOF. *If*: Let the columns of $U \in \mathbb{C}_s^{n\times s}$ be a basis for T, and let the columns of $V^* \in \mathbb{C}_s^{m\times s}$ be a basis for $S^\perp$. Then the columns of AU span AT. Since it follows from (44) that $\dim AT = s$,

$$\operatorname{rank} AU = s \ . \tag{45}$$

A further consequence of (44) is

$$AT \cap S = \{\mathbf{0}\} \ . \tag{46}$$

Moreover, the $s \times s$ matrix VAU is nonsingular (i.e., of rank s) because

$$\begin{aligned} VAU\mathbf{y} = \mathbf{0} &\implies AU\mathbf{y} \perp S^\perp \implies AU\mathbf{y} \in S \\ &\implies AU\mathbf{y} = \mathbf{0} \quad \text{(by (46)} \\ &\implies \mathbf{y} = \mathbf{0} \quad \text{(by (45)} \ . \end{aligned}$$

Therefore, by Theorem 11,

$$X = U(VAU)^{-1}V$$

is a $\{2\}$–inverse of A having range T and null space S (see also Stewart [**1400**]).
Only if: Since $A \in X\{1\}$, AX is idempotent by Lemma 1.1(f). Moreover, $AT = R(AX)$ and $S = N(X) = N(AX)$ by (28). Thus (44) follows from Theorem 8.
Proof of uniqueness: Let X_1, X_2 be $\{2\}$–inverses of A having range T and null space S. By Lemma 1.1(f) and (28), X_1A is a projector with range T and AX_2 is a projector with null space S. Thus, by Ex. 20,

$$X_2 = (X_1A)X_2 = X_1(AX_2) = X_1 \ .$$

□

COROLLARY 9. Let $A \in \mathbb{C}_r^{m\times n}$, let T be a subspace of $\mathbb{C}^n$ of dimension r, and let S be a subspace of $\mathbb{C}^m$ of dimension $m - r$. Then, the following three statements are equivalent:
(a) $AT \oplus S = \mathbb{C}^m$.
(b) $R(A) \oplus S = \mathbb{C}^m$ and $N(A) \oplus T = \mathbb{C}^n$.
(c) There exists an $X \in A\{1, 2\}$ such that $R(X) = T$ and $N(X) = S$. □

The set of $\{2\}$–inverses of A with range T and null space S is denoted $A\{2\}_{S,T}$.

Exercises.

EX. 28. Show that $A_{T,S}^{(1,2)}$ is the unique matrix X satisfying the three equations

$$AX = P_{L,S} \ , \quad XA = P_{T,M} \ , \quad XP_{L,S} = X \ .$$

(For the Moore–Penrose inverse this was shown by Petryshyn [**1183**]. Compare Ex. 1.24.)

EX. 29. For any given matrix A, $A^\dagger$ is the unique matrix $X \in A\{1, 2\}$ such that $R(X) = R(A^*)$ and $N(X) = N(A^*)$.

EX. 30. Derive the formula of Mitra [**1058**] and Zlobec [**1652**]

$$A^\dagger = A^*YA^* \ ,$$

where Y is an arbitrary element of $(A^*AA^*)\{1\}$.

Ex. 31. Derive the formula of Decell [**391**]

$$A^{\dagger} = A^* X A^* Y A^* \ ,$$

where X and Y are any $\{1\}$–inverses of AA^* and A^*A, respectively.

Ex. 32. Penrose [**1177**] showed that the Moore–Penrose inverse of a product of two Hermitian idempotent matrices is idempotent. Prove this, using Zlobec's formula (Ex. 30).

Ex. 33. Let A be the matrix of Ex. 0.46, and let S be the subspace spanned by $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and let T be the subspace spanned by the columns of $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$. Calculate $A^{(1,2)}_{T,S}$.

Ex. 34. If E is idempotent and the columns of F and G^* are bases for $R(E)$ and $R(E^*)$, respectively, show that $E = F(GF)^{-1}G$.

Ex. 35. If A is square and $A = FG$ is a full–rank factorization, show that A has a $\{1,2\}$–inverse X with $R(X) = R(A)$ and $N(X) = N(A)$ if and only if GF is nonsingular, in which case $X = F(GF)^{-2}G$ (Cline [**352**]).

6. Orthogonal projections and orthogonal projectors

Given a vector $\mathbf{x} \in \mathbb{C}^n$ and a subspace L of $\mathbb{C}^n$, there is in L a unique vector $\mathbf{u}_{\mathbf{x}}$ that is "closest" to $\mathbf{x}$ in the sense that the "distance" $\|\mathbf{x} - \mathbf{u}\|$ is smaller for $\mathbf{u} = \mathbf{u}_{\mathbf{x}}$ than for any other $\mathbf{u} \in L$. Here, $\|\mathbf{v}\|$ denotes the *Euclidean norm* of the vector $\mathbf{v}$,

$$\|\mathbf{v}\| = +\sqrt{(\mathbf{v},\mathbf{v})} = +\sqrt{\mathbf{v}^*\mathbf{v}} = +\sqrt{\sum_{j=1}^{n} |v_j|^2} \ ,$$

where $(\mathbf{v},\mathbf{w})$ denotes the *standard inner product*, defined for $\mathbf{v},\mathbf{w} \in \mathbb{C}^n$ by

$$(\mathbf{v},\mathbf{w}) = \mathbf{w}^*\mathbf{v} = \sum_{j=1}^{n} \overline{w}_j v_j \ .$$

Not surprisingly, the vector $\mathbf{u}_{\mathbf{x}}$ that is "closest" to $\mathbf{x}$ of all vectors in L is uniquely characterized (see Ex. 37) by the fact that $\mathbf{x} - \mathbf{u}_{\mathbf{x}}$ is orthogonal to $\mathbf{u}_{\mathbf{x}}$, which we shall denote by

$$\mathbf{x} - \mathbf{u}_{\mathbf{x}} \perp \mathbf{u}_{\mathbf{x}} \ .$$

We shall therefore call the "closest" vector $\mathbf{u}_{\mathbf{x}}$ the *orthogonal projection* of $\mathbf{x}$ on L. The transformation that carries each $\mathbf{x} \in \mathbb{C}^n$ into its orthogonal projection on L we shall denote by P_L and shall call the *orthogonal projector* on L (sometimes abbreviated to "o.p. on L"). Comparison with the earlier definition of the projector on L along M (see Section 4) shows that the orthogonal projector on L is the same as the projector on L along $L^{\perp}$. (As previously noted, some writers call the orthogonal projector on L simply the projector on L.)

Being a particular case of the more general projector, the orthogonal projector is representable by a square matrix, which, in this case, is not only idempotent but also Hermitian.

In order to prove this, we shall need the relation

$$N(A) = R(A^*)^{\perp} \ , \tag{47}$$

which, in fact, arises frequently in the study of generalized inverses. Two proofs of (47) are given in Ex. 38. The first, using inner products, is immediately generalizable to transformations on Hilbert space, which will be discussed in Chapter 8. The second proof shows that, in the more restricted context of finite matrices, (47) is a consequence of the equation $A\mathbf{x} = \mathbf{0}$, which defines $N(A)$.

Let L and M be complementary orthogonal subspaces of $\mathbb{C}^n$, and consider the matrix $P^*_{L,M}$. By Lemma 1(a), it is idempotent and therefore a projector, by Theorem 8. By the use of (47) and its dual

$$N(A^*) = R(A)^\perp \tag{48}$$

(obtained by replacing A by A^* in (47)), it is readily found that

$$R(P^*_{L,M}) = M^\perp \ , \quad N(P^*_{L,M}) = L^\perp \ .$$

Thus, by Theorem 8,

$$P^*_{L,M} = P_{M^\perp,L^\perp} \ , \tag{49}$$

from which the next lemma follows easily.

Lemma 3. Let $\mathbb{C}^n = L \oplus M$. Then $M = L^\perp$ if and only if $P_{L,M}$ is Hermitian.

Just as there is a one–to–one correspondence between projectors and idempotent matrices, Lemma 3 shows that there is a one–to–one correspondence between orthogonal projectors and Hermitian idempotents. Matrices of the latter class have many striking properties, some of which are noted in the remainder of this section (including the exercises).

For any subspace L for which a basis is available, it is easy to construct the matrix P_L. The basis must first be orthonormalized (e.g., by Gram–Schmidt orthogonalization). Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\}$ be an orthonormal basis for L. Then

$$P_L = \sum_{j=1}^{l} \mathbf{x}_j \mathbf{x}_j^* \ . \tag{50}$$

The reader should verify that RHS(50) is the orthogonal projector on L, and that (27) reduces to (50) if $M = L^\perp$ and the basis is orthonormal.

In the preceding section diagonable matrices were studied in relation to projectors. The same relations will now be shown to hold between normal matrices (a subclass of diagonable matrices) and orthogonal projectors. This constitutes the *spectral theory for normal matrices.* We recall that a square matrix A is called *normal* if it commutes with its conjugate transpose

$$AA^* = A^*A \ .$$

It is well known that every normal matrix is diagonable. A normal matrix A also has the property (see Ex. 41) that the eigenvalues of A^* are the conjugates of those of A, and every eigenvector of A associated with the eigenvalue λ is also an eigenvector of A^* associated with the eigenvalue $\bar{\lambda}$.

The following spectral theorem relates normal matrices to orthogonal projectors, in the same way that diagonable matrices and projectors are related in Theorem 9.

Theorem 13. (Spectral Theorem for Normal Matrices). Let $A \in \mathbb{C}^{n\times n}$ with k distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. Then A is normal if and only if there exist orthogonal projectors $E_1, E_2, \dots, E_k$ such that

$$E_i E_j = O \ , \text{ if } i \neq j \ , \tag{51}$$

$$I_n = \sum_{i=1}^{k} E_i \ , \tag{52}$$

$$A = \sum_{i=1}^{k} \lambda_i E_i \ . \tag{53}$$

PROOF. *If*: Let A be given by (53) where the principal idempotents are Hermitian. Then

$$\begin{aligned} AA^* &= \left(\sum_{i=1}^{k} \lambda_i E_i\right)\left(\sum_{j=1}^{k} \bar{\lambda}_j E_j\right) \\ &= \sum_{i=1}^{k} |\lambda_i|^2 E_i = A^*A \,. \end{aligned}$$

Only if: Since A is normal, it is diagonable; let $E_1, E_2, \dots, E_k$ be its principal idempotents. We must show that they are Hermitian. By Ex. 41, $R(E_i)$, the eigenspace of A associated with the eigenvalue λ_i is the same as the eigenspace of A^* associated with $\bar{\lambda}_i$. Because of (51), the null spaces of corresponding principal idempotents of A and A^* are also the same (for a given $i = h$, $N(E_h)$ is the direct sum of the eigenspaces $R(E_i)$ for all $i \neq h$, i.e.,

$$N(E_h) = \sum_{\substack{i=1\\ i\neq h}}^{k} \oplus R(E_i) \quad (h \in \overline{1,k}))\,.$$

Therefore, A and A^* have the same principal idempotents, by Theorem 8. Consequently,

$$A^* = \sum_{i=1}^{k} \bar{\lambda}_i E_i \,,$$

by Theorem 9. But taking conjugate transposes in (53) gives

$$A^* = \sum_{i=1}^{k} \bar{\lambda}_i E_i^* \,,$$

and it is easily seen that the idempotents E_i^* satisfy (51) and (52). Since the spectral decomposition is unique by Ex. 24, we must have

$$E_i = E_i^* \,,\ i \in \overline{1,k} \,.$$

□

Exercises and examples.

Ex. 36. *Orthogonal subspaces* or the *Pythagorean theorem.* Let Y and Z be subspaces of $\mathbb{C}^n$. Then $Y \perp Z$ if and only if

$$\|\mathbf{y}+\mathbf{z}\|^2 = \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 \,,\ \text{for all } \mathbf{y}\in Y,\ \mathbf{z}\in Z \,. \tag{54}$$

PROOF. *If*: Let $\mathbf{y} \in Y$, $\mathbf{z} \in Z$. Then (54) implies that

$$\begin{aligned} (\mathbf{y},\mathbf{y}) + (\mathbf{z},\mathbf{z}) = \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 &= \|\mathbf{y}+\mathbf{z}\|^2 \\ &= (\mathbf{y}+\mathbf{z},\mathbf{y}+\mathbf{z}) = (\mathbf{y},\mathbf{y}) + (\mathbf{z},\mathbf{z}) + (\mathbf{y},\mathbf{z}) + (\mathbf{z},\mathbf{y}) \,, \end{aligned}$$

and therefore

$$(\mathbf{y},\mathbf{z}) + (\mathbf{z},\mathbf{y}) = 0 \,. \tag{55}$$

Now, since Z is a subspace, $i\mathbf{z} \in Z$, and replacing $\mathbf{z}$ by $i\mathbf{z}$ in (55) gives

$$0 = (\mathbf{y}, i\mathbf{z}) + (i\mathbf{z},\mathbf{y}) = i(\mathbf{y},\mathbf{z}) - i(\mathbf{z},\mathbf{y}) \,. \tag{56}$$

(Here we have used the fact that $(\alpha\mathbf{v},\mathbf{w}) = \alpha(\mathbf{v},\mathbf{w})$ and $(\mathbf{v},\beta\mathbf{w}) = \bar{\beta}(\mathbf{v},\mathbf{w})$.) It follows from (56) that

$$(\mathbf{y},\mathbf{y}) - (\mathbf{z},\mathbf{z}) = 0 \,,$$

which, in conjunction with (55) gives

$$(\mathbf{y}, \mathbf{y}) = (\mathbf{z}, \mathbf{z}) = 0 ,$$

i.e., $\mathbf{y} \perp \mathbf{z}$.

Only if: Let $Y \perp Z$. Then, for arbitrary $\mathbf{y} \in Y$, $\mathbf{z} \in Z$

$$\begin{aligned} \|\mathbf{y} + \mathbf{z}\|^2 &= (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z}) \\ &= (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z})| , \text{ since } (\mathbf{y}, \mathbf{y}) - (\mathbf{z}, \mathbf{z}) = 0 , \\ &= \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 . \end{aligned}$$

□

Ex. 37. *Orthogonal projections.* Let L be a subspace of $\mathbb{C}^n$. Then, for every $\mathbf{x} \in \mathbb{C}^n$ there is a unique vector $\mathbf{u_x}$ in L such that for all $\mathbf{u} \in L$ different from $\mathbf{u_x}$

$$\|\mathbf{x} - \mathbf{u_x}\| < \|\mathbf{x} - \mathbf{u}\| .$$

Among the vectors $\mathbf{u} \in L$, $\mathbf{u_x}$ is uniquely characterized by the fact that

$$\mathbf{x} - \mathbf{u_x} \perp \mathbf{u_x} .$$

Proof. Let $\mathbf{x} \in \mathbb{C}^n$. Since L and $L^\perp$ are complementary subspaces, there exist uniquely determined vectors $\mathbf{x}_1 \in L$, $\mathbf{x}_2 \in L^\perp$ such that

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 . \tag{57}$$

Therefore for arbitrary $\mathbf{u} \in L$,

$$\begin{aligned} \|\mathbf{x} - \mathbf{u}\|^2 &= \|\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{u}\|^2 \\ &= \|\mathbf{x}_1 - \mathbf{u}\|^2 + \|\mathbf{x}_2\|^2 , \end{aligned} \tag{58}$$

by Ex. 36, since $\mathbf{x}_1 - \mathbf{u} \in L$, $\mathbf{x}_2 \in L^\perp$. Consequently, there is a unique $\mathbf{u} \in L$, namely $\mathbf{u_x} = \mathbf{x}_1$, for which (58) is smallest.

By the uniqueness of the decomposition (57), $\mathbf{u_x} = \mathbf{x}_1$ is the only vector $\mathbf{u} \in L$ satisfying

$$\mathbf{x} - \mathbf{u} \perp \mathbf{u} .$$

□

Ex. 38. $N(A) = R(A^*)^\perp$.

Because of the importance of this relation, we give two proofs, one in terms of inner products, and the other based on matrix multiplication.

First proof. Let $A \in \mathbb{C}^{m\times n}$, and recall that for all $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{y} \in \mathbb{C}^m$

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y}) . \tag{59}$$

Let $\mathbf{x} \in N(A)$. Then LHS(59) vanishes for all $\mathbf{y} \in \mathbb{C}^m$. From (59) 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(59) vanishes for all $\mathbf{y} \in \mathbb{C}^m$. Then (59) 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 of the original relation. □

First 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 $R(A^*)$. □

Ex. 39. Let $\mathbf{x} \in \mathbb{C}^n$ and let L be an arbitrary subspace of $\mathbb{C}^n$. Then

$$\|P_L \mathbf{x}\| \leq \|\mathbf{x}\| , \tag{60}$$

with equality if and only if $\mathbf{x} \in L$. See also Ex. 53.

Proof. We have

$$\mathbf{x} = P_L\mathbf{x} + (I - P_L)\mathbf{x} = P_L\mathbf{x} + P_{L^\perp}\mathbf{x} ,$$

by Ex. 23. Then by Ex. 36,

$$\|\mathbf{x}\|^2 = \|P_L\mathbf{x}\|^2 + \|P_{L^\perp}\mathbf{x}\|^2 ,$$

from which (60) follows.

Equality holds in (60) if and only if $P_{L^\perp}\mathbf{x} = \mathbf{0}$, which is equivalent to $\mathbf{x} \in L$. □

Ex. 40. Let A be a square singular matrix, let $\{\mathbf{u}_1, \mathbf{u}_2, \dots , \mathbf{u}_n\}$ and $\{\mathbf{x}_1, \mathbf{x}_2, \dots , \mathbf{x}_n\}$ be orthonormal bases of $N(A^*)$ and $N(A)$, respectively, and let $\{\alpha_1, \alpha_2, \dots , \alpha_n\}$ be nonzero scalars. Then the matrix

$$A_0 = A + \sum_{i=1}^{n} \alpha_i \, \mathbf{u}_i \, \mathbf{x}_i^*$$

is nonsingular, and its inverse is

$$A_0^{-1} = A^\dagger + \sum_{i=1}^{n} \frac{1}{\alpha_i} \mathbf{x}_i \, \mathbf{u}_i^* \, .$$

Proof. Let X denote the expression given for A_0^{-1}. Then, from $\mathbf{x}_i^*\mathbf{x}_j = \delta_{ij} \quad (i, j \in \overline{1,n})$, it follows that

$$\begin{aligned} A_0X &= AA^\dagger + \sum_{i=1}^{n} \mathbf{x}_i \, \mathbf{x}_i^* \\ &= AA^\dagger + P_{N(A^*)} \quad \text{(by (50))} \\ &= AA^\dagger + (I_n - AA^\dagger) \quad \text{(by Lemma 1(g))} \\ &= I_n \, . \end{aligned}$$

Therefore, A_0 is nonsingular, and $X = A_0^{-1}$. □

Ex. 41. If A is normal, $A\mathbf{x} = \lambda\mathbf{x}$ if and only if $A^*\mathbf{x} = \bar{\lambda}\mathbf{x}$.

Ex. 42. If L is a subspace of $\mathbb{C}^n$ and the columns of F are a basis for L, show that

$$P_L = FF^\dagger = F(F^*F)^{-1}F^* \, .$$

(This may be simpler computationally than orthonormalizing the basis and using (50).)

Ex. 43. Let L be a subspace of $\mathbb{C}^n$. Then

$$P_{L^\perp} = I_n - P_L \, .$$

(See Ex. 23).)

Ex. 44. Let $A \in \mathbb{C}^{m\times n}$, $X \in \mathbb{C}^{n\times m}$. Then $X \in A\{2\}$ if and only if it is of the form

$$X = (EAF)^\dagger$$

where E and F are suitable Hermitian idempotents (Greville [**586**]).

PROOF. *If*: By Ex. 29,

$$R((EAF)^\dagger) \subset R(F) \ , \quad N((EAF)^\dagger) \supset N(E) \ .$$

Therefore, by Ex. 20,

$$X = (EAF)^\dagger = F(EAF)^\dagger = (EAF)^\dagger E; .$$

Consequently,

$$XAX = (EAF)^\dagger EAF(EAF)^\dagger = (EAF)^\dagger = X \ .$$

Only if: By Theorem 10(c) and Ex. 29,

$$X^\dagger = P_{R(X^*)} A P_{R(X)} \ ,$$

and, therefore, by Ex. 1.16,

$$X = \left(P_{R(X^*)} A P_{R(X)}\right)^\dagger \ . \tag{61}$$

□

Remark. Equation (61) states that if $X \in A\{2\}$, then X is the Moore–Penrose inverse of a modification of A obtained by projecting its columns on $R(X^*)$ and its rows on $R(X)$.

EX. 45. It follows from Exs. 28 and 1.24 that, for arbitrary A, $A^\dagger$ is the unique matrix X satisfying

$$AX = P_{R(A)} \ , \quad XA = P_{R(A^*)} \ , \quad XAX = X \ .$$

EX. 46. By means of Exs. 45 and 20, derive (61) directly from $XAX = X$ without using Theorem 10(c).

EX. 47. Prove the following amplification of Penrose's result stated in Ex. 32: A square matrix E is idempotent if and only if it can be expressed in the form

$$E = (FG)^\dagger$$

where F and G are Hermitian idempotents. (*Hint*: Use Ex. 17.)

In particular, derive the formula (Greville [**586**])

$$P_{L,M} = (P_{M^\perp} P_L)^\dagger = ((I - P_M)P_L)^\dagger \ . \tag{62}$$

EX. 48. Let S and T be subspaces of $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, such that

$$AT \oplus S = \mathbb{C}^m \ ,$$

and let $A^{(2)}_{T,S}$ denote the unique $\{2\}$–inverse of A having range T and null space S (see Theorem 12). Then

$$A^{(2)}_{T,S} = (P_{S^\perp} A P_T)^\dagger \ .$$

EX. 49. Show that $P_L + P_M$ is an o.p. if and only if $L \perp M$, in which case

$$P_L + P_M = P_{L+M} \ .$$

EX. 50. Show that $P_L P_M$ is an o.p. if and only if P_L and P_M commute, in which case

$$P_L P_M = P_{L\cap M} \ .$$

EX. 51. Show that $L = L \cap M \oplus L \cap M^\perp$ if and only if P_L and P_M commute.

Ex. 52. For a Hermitian matrix H we denote by $H \geq O$ the fact that H is *non–negative definite*, i.e., $(H\mathbf{x}, \mathbf{x}) \geq 0$ for all $\mathbf{x}$. For any two Hermitian matrices G, H,

$$G \geq H \text{ enotes } G - H \geq O \ .$$

The relation $\geq$ is a partial order on the set of Hermitian matrices.

Let P_L and P_M be orthogonal projectors on the subspaces L and M of $\mathbb{C}^n$, respectively. Then the following statements are equivalent:

(a) $P_L - P_L$ is an o.p.
(b) $P_L \geq P_M$.
(c) $\|P_L\mathbf{x}\| \geq \|P_M\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$.
(d) $M \subset L$.
(e) $P_L P_M = P_M$.
(f) $P_M P_L = P_M$.

Ex. 53. Let $P \in \mathbb{C}^{n\times n}$ be a projector. Then P is an orthogonal projector if and only if

$$\|P\mathbf{x}\| \leq \|\mathbf{x}\| \text{ for all } \mathbf{x} \in \mathbb{C}^n \ . \tag{63}$$

PROOF. P is an o.p. if and only if $I - P$ is an o.p. By the equivalence of statements (a) and (c) in Ex. 52, $I - P$ is an o.p. if and only if (63) holds. □

Note that for any non–Hermitian idempotent P (i.e., for any projector P which is not an orthogonal projector) there is by this exercise a vector $\mathbf{x}$ whose length is increased when multiplied by P, i.e., $\|P\mathbf{x}\| > \|\mathbf{x}\|$. For $P = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ such a vector is $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Ex. 54. Let $P \in \mathbb{C}^{n\times n}$. Then P is an o.p. if and only if

$$P = P^* P \ .$$

Ex. 55. It may be asked to what extent the results of Exs. 49–51 carry over to general projectors. this question is explored in this and the two following exercises. Let

$$\mathbb{C}^n = L \oplus M = Q \oplus S \ .$$

Then show that $P_{L,M} + P_{Q,S}$ is a projector if and only if $M \supset Q$ and $S \supset L$, in which case

$$P_{L,M} + P_{Q,S} = P_{L+Q,M\cap S} \ .$$

SOLUTION. Let $P_1 = P_{L,M}$, $P_2 = P_{Q,S}$. Then

$$(P_1 + P_2)^2 = P_1 + P_2 + P_1P_2 + P_2P_1 \ .$$

Therefore, $P_1 + P_2$ is a projector if and only if

$$P_1P_2 + P_2P_1 = O \ . \tag{64}$$

Now if $M \supset Q$ and $S \supset L$, each term of LHS(64) is O.

On the other hand, if (64) holds, multiplication by P_1 on the left and on the right, respectively, gives

$$P_1P_2 + P_1P_2P_1 = O = P_1P_2P_1 + P_2P_1 \ .$$

Subtraction then yields

$$P_1P_2 - P_2P_1 = O \ , \tag{65}$$

and (64) and (65) together imply

$$P_1P_2 = P_2P_1 = O \ ,$$

from which it follows by Lemma 1(e) that $M \supset Q$ and $S \supset L$. It is then fairly easy to show that

$$P_1 + P_2 = P_{L+Q,M\cap S} \ .$$

Ex. 56. With L, M, Q, S as in Ex. 55 show that if $P_{L,M}$ and $P_{Q,S}$ commute, then

$$P_{L,M} P_{Q,S} = P_{Q,S} P_{L,M} = P_{L \cap Q, M+S} \ . \tag{66}$$

Ex. 57. If only one of the products in (66) is equal to the projector on the right, it does not necessarily follow that the other productis the same. Instead we have the following result: With L, M, Q, S as in Ex. 55, $P_{L,M} P_{Q,S} = P_{L \cap Q, M+S}$ if and only if $Q = L \cap Q \oplus M \cap Q$. Similarly, $P_{Q,S} P_{L,M} = P_{L \cap Q, M+S}$ if and only if $L = L \cap Q \oplus L \cap S$.

Proof. Since $L \cap M = \{\mathbf{0}\}$, $(L \cap Q) \cap (M \cap Q) = \{\mathbf{0}\}$. Therefore $L \cap Q + M \cap Q = L \cap Q \oplus M \cap Q$. Since $M + S \supset M + Q$ and $L + S \supset L \cap Q$, Ex. 55 gives

$$P_{L \cap Q, M+S} + P_{M \cap Q, L+S} = P_{T,U} \ ,$$

where $T = L \cap Q \oplus M \cap Q$, $U = (L + S) \cap (M + S)$. Clearly $Q \supset T$ and $U \supset S$. Multiplying on the left by $P_{L,M}$ gives

$$P_{L,M} P_{T,U} = P_{L \cap Q, M+S} \ . \tag{67}$$

Thus, if $T = Q$, we have $U = S$, and

$$P_{L,M} P_{Q,S} = P_{L \cap Q, M+S} \ . \tag{68}$$

On the other hand, if (68) holds, (67) and (68) give

$$P_{Q,S} = P_{T,U} + H \ , \tag{69}$$

where $P_{L,M} H = O$. This implies $R(H) \subset M$. Also, since $T \subset Q$, (69) implies $R(H) \subset Q$, and therefore $R(H) \subset M \cap Q$. Consequently, $R(H) \subset T$ and therefore (69) gives $P_{T,U} P_{Q,S} = P_{Q,S}$. This implies $\operatorname{rank} P_{Q,S} \leq \operatorname{rank} P_{T,U}$. Since $Q \supset T$ it follows that $T = Q$. This proves the first statement, and the proof of the second statement is similar. □

Ex. 58. The characterization of $A\{2, 3, 4\}$ was postponed until o.p.'s had been studied. This will now be dealt with in three stages in this exercise and Exs. 59 and 61. If E is Hermitian idempotent show that $X \in E\{2, 3, 4\}$ if and only if X is Hermitian idempotent and $R(X) \subset R(E)$.

Proof. *If*: Since $R(X) \subset R(E)$, $EX = X$ be Lemma 1(e), and taking conjugate transposes gives $XE = X$. Since X is Hermitian, EX and XE are Hermitian. Finally, $XEX = X^2 = X$, since X is idempotent. Thus, $X \in E\{2, 3, 4\}$.
Only if: Let $X \in E\{2, 3, 4\}$. Then $X = XEX = EX^*X$. Therefore $R(X) \subset R(E)$. Then $EX = X$ by Lemma 1(e). But EX is Hermitian idempotent, since $X \in E\{2, 3\}$. Therefore X is Hermitian idempotent. □

Ex. 59. Let H be Hermitian non–negative definite, with spectral decomposition as in (31) with o.p.'s as its principal idempotents. Thus,

$$H = \sum_{i=1}^{k} \lambda_i \, E_i \ . \tag{70}$$

Then $X \in H\{2, 3, 4\}$ if and only if

$$X = \sum_{i=1}^{k} \lambda_i^{\dagger} \, F_i \ , \tag{71}$$

where, for each i, $F_i \in E_i\{2, 3, 4\}$.

Proof. *If*: Since E_i is Hermitian idempotent, $R(F_i) \subset R(E_i)$ by Ex. 58. Therefore (29) gives

$$E_i F_j = F_j E_i = O \quad (i \neq j) \ , \tag{72}$$

and by Lemma 1(e)

$$E_i F_i = F_i E_i = F_i \quad (i \in \overline{1, k}) \ .$$

Consequently,

$$HX = \sum_{\substack{i=1 \\ \lambda_i \neq 0}}^{k} F_i = XH \ .$$

Since each F_i is Hermitian by Ex. 58, $HX = XH$ is Hermitian. Now,

$$F_i F_j = F_i E_j F_j = O \quad (i \neq j) \ ,$$

by (72), and therefore

$$XHX = \sum_{i=1}^{k} \lambda_i^{\dagger} F_i^2 = X$$

by (71), since each F_i is idempotent.

Only if: Let $X \in H\{2,3,4\}$. Then, by (30)

$$X = IXI = \sum_{i=1}^{k} \sum_{j=1}^{k} E_i X E_j \ . \tag{73}$$

Now, (70) gives

$$HX = \sum_{i=1}^{k} \lambda_i \, E_i \, X = \sum_{i=1}^{k} \lambda_i \, X^* \, E_i \ , \tag{74}$$

since $HX = X^* H$. Similarly,

$$XH = \sum_{i=1}^{k} \lambda_i \, X \, E_i = \sum_{i=1}^{k} \lambda_i \, E_i \, X^* \ . \tag{75}$$

Multiplying by E_s on the left and by E_t on the right in both (74) and (75) and making use of (29) and the idempotency of E_s and E_t gives

$$\lambda_s E_s X E_t = \lambda_t E_s X^* E_t \tag{76}$$

$$\lambda_t E_s X E_t = \lambda_s E_s X^* E_t \ , \quad (s, t \in \overline{1,k}) \ . \tag{77}$$

Adding and subtracting (76) and (77) gives

$$(\lambda_s + \lambda_t) E_s X E_t = (\lambda_s + \lambda_t) E_s X^* E_t \ , \tag{78}$$

$$(\lambda_s - \lambda_t) E_s X E_t = -(\lambda_s - \lambda_t) E_s X^* E_t \ . \tag{79}$$

The λ_i are distinct, and are also non–negative because H is Hermitian non–negative definite. Thus, if $s \neq t$, neither of the quantities $\lambda_s + \lambda_t$ and $\lambda_s - \lambda_t$ vanishes. Therefore, (78) and (79) give

$$E_s X E_t = E_s X^* E_t = -E_s X E_t = O \quad (s \neq t) \ . \tag{80}$$

Consequently, (73) reduces to

$$X = \sum_{i=1}^{k} E_i X E_i \ . \tag{81}$$

Now, (70) gives

$$X = XHX = \sum_{i=1}^{k} \lambda_i X E_i X \ ,$$

and therefore by (80)

$$E_sXE_s = \lambda_s E_sXE_sXE_s = \lambda_s(E_sXE_s)^2 \ , \tag{82}$$

from which it follows that $E_sXE_s = O$ if $\lambda_s = 0$. Now, take

$$F_i = \lambda_i E_i X E_i \quad (i \in \overline{1,k}) \ . \tag{83}$$

Then (81) becomes (71), and we have only to show that $F_i \in E_i\{2,3,4\}$. This is trivially true for that i, if any, such that $\lambda_i = 0$. For other i, we deduce from (76) that it is idempotent. Finally, (83) gives $R(F_i) \subset R(E_i)$, and the desired conclusion follows from Ex. 58. □

Ex. 60. Prove the following corollary of Ex. 59. If H is Hermitian non–negative definite and $X \in H\{2,3,4\}$, then X is Hermitian non–negative definite, and every nonzero eigenvalue of X is the reciprocal of an eigenvalue of H.

Ex. 61. For every $A \in \mathbb{C}^{m\times n}$

$$A\{2,3,4\} = \{YA^* : Y \in (A^*A)\{2,3,4\}\} \ .$$

Ex. 62. $A\{2,3,4\}$ is a finite set if and only if the nonzero eigenvalues of A^*A are distinct (i.e., each eigenspace associated with a nonzero eigenvalue of A^*A is of dimension one). If this is the case and if there are k such eigenvalues, $A\{2,3,4\}$ contains exactly 2^k elements.

Ex. 63. Show that the matrix

$$A = \frac{1}{10}\begin{bmatrix} 9-3i & 12-4i & 10-10i \\ 3-3i & 4-4i & 0 \\ 6+6i & 8+8i & 0 \\ 6 & 8 & 0 \end{bmatrix}$$

has exactly four $\{2,3,4\}$–inverses, namely,

$$\begin{aligned}
X_1 = A^\dagger &= \frac{1}{70}\begin{bmatrix} 0 & 6+6i & 12-12i & 12 \\ 0 & 8+8i & 16-16i & 16 \\ 35+35i & -5-15i & -30+10i & -20-10i \end{bmatrix} , \\
X_2 &= \frac{1}{60}\begin{bmatrix} -9-3i & 3+3i & 6-6i & 6 \\ -12-4i & 4+4i & 8-8i & 8 \\ 25+25i & -5-15i & -30+10i & -20-10i \end{bmatrix} , \\
X_3 &= \frac{1}{420}\begin{bmatrix} 63+21i & 15+15i & 30-30i & 30 \\ 84+28i & 20+20i & 40-40i & 40 \\ 35+35i & 5+15i & 30-10i & 20+10i \end{bmatrix} , \\
X_4 &= O \ .
\end{aligned}$$

7. Efficient characterization of classes of generalized inverses

In the preceding sections, characterizations of certain classes of generalized inverses of a given matrix have been given. Most of these characterizations involve one or more matrices with arbitrary elements. In general, the number of such arbitrary elements far exceeds the actual number of degrees of freedom available.

For example, in Section 1 we obtained the characterization

$$A\{1\} = \{A^{(1)} + Z - A^{(1)}AZAA^{(1)} : Z \in \mathbb{C}^{n\times m}\} \ . \tag{4}$$

Now, as Z ranges over the entire class $\mathbb{C}^{n\times m}$, every $\{1\}$–inverse of A will be obtained repeatedly an infinite number of times unless A is a matrix of zeros. In fact, the expression in RHS(4) is unchanged if Z is replaced by $Z + A^{(1)}AWAA^{(1)}$, where W is an arbitrary element of $\mathbb{C}^{n\times m}$. We shall now see how in some cases this redundancy in the number of arbitrary parameters can be eliminated. The cases of particular interest are $A\{1\}$ because of its role in the solution of linear systems, $A\{1,2\}$ because of the symmetry inherent in the relation

$$X \in A\{1,2\} \iff A \in X\{1,2\} ,$$

and $A\{1,3\}$ and $A\{1,4\}$ because of their minimization properties, which will be studied in the next chapter.

As in (4), let $A^{(1)}$ be a fixed, but arbitrary element of $A\{1\}$, where $A \in \mathbb{C}_r^{m\times n}$. Also, let $F \in \mathbb{C}_{n-r}^{n\times(n-r)}$, $K^* \in \mathbb{C}_{m-r}^{m\times(m-r)}$, $B \in \mathbb{C}_r^{n\times r}$ be given matrices whose columns are bases for $N(A)$, $N(A^*)$ and $RA^{(1)}A)$, respectively. We shall show that the general solution of

$$AXA = A \tag{1.1}$$

is

$$X = A^{(1)} + FY + BZK , \tag{84}$$

where $Y \in \mathbb{C}^{(n-r)\times m}$ and $Z \in \mathbb{C}^{r\times(m-r)}$ are arbitrary.

Clearly $AF = O$ and $KA = O$. Therefore RHS(84) satisfies (1.1). Since $R(I_n - A^{(1)}A) = N(A)$ and $R((I_m - AA^{(1)})^*) = N(A^*)$ by (28) and Lemma 1(g), there exist uniquely defined matrices G, H, D such that

$$FG = I_n - A^{(1)}A , \quad HK = I_m - AA^{(1)} , \quad BD = A^{(1)}A . \tag{85}$$

Since these products are idempotent, we have, by Lemma 2,

$$GF = DB = I_n , \quad KH = I_m . \tag{86}$$

Moreover, it is easily verified that

$$GB = O , \quad DF = O . \tag{87}$$

Using (86) and (87), we obtain easily from (84)

$$Y = G(X - A^{(1)}) , \quad Z = D(X - A^{(1)})H . \tag{88}$$

Now, let X be an arbitrary element of $A\{1\}$. Upon substituting in (84) the expression (88) for Y and Z, it is found that (84) is satisfied. We have shown, therefore, that (84) does indeed give the general solution of (1.1).

We recall that $A^{(1)}, F, G, H, K, B, D$ are fixed matrices. Therefore, not only does (84) give X uniquely in terms of Y and Z, but also (88) gives Y and Z uniquely in terms of X. Therefore, different choices of Y and Z in (84) must yield different $\{1\}$–inverses X. Thus, the characterization (84) is completely efficient, and contains the smallest possible number of arbitrary parameters.

It is interesting to compare the number of arbitrary elements in the characterizations (4) and (84). In (4) this is mn, the number of elements of Z. In (84) it is $mn - r^2$, the total number of elements in Y and Z. Clearly, (84) contains fewer arbitrary elements, except in the trivial case $r = 0$, as previously noted.

The case of $A\{1,3\}$ is easier. If, as before, the columns of F are a basis for $N(A)$, it is readily seen that (17) can be written in the alternative form

$$A\{1,3\} = \{A^{(1,3)} + FY : Y \in \mathbb{C}^{(n-r)\times m}\} . \tag{89}$$

This is easily shown to be an efficient characterization. Here the number of arbitrary parameters is $m(m - r)$. Evidently this is less than the number in the efficient characterization (84) of $A\{1\}$, unless $r = m$, in which case every $\{1\}$–inverse is a $\{1,3\}$–inverse, since $AA^1 = I_m$ by Lemma 1.2.

Similarly, if the columns of K^* are a basis for $N(A^*)$

$$A\{1,4\} = \{A^{(1,4)} + YK : Y \in \mathbb{C}^{n\times(m-r)}\} , \tag{90}$$

where $A^{(1,4)}$ is fixed, but arbitrary element of $A\{1,4\}$.

Efficient characterization of $A\{1,2\}$ is somewhat more difficult. Let $A^{(1,2)}$ be a fixed, but arbitrary element of $A\{1,2\}$, and let

$$A^{(1,2)} = Y_0 Z_0$$

be a full–rank factorization. As before, let the columns of F and K^* form bases for the null spaces of A and A^*, respectively. Then we shall show that

$$A\{1,2\} = \{(Y_0 + FU)(Z_0 + VK) : U \in \mathbb{C}^{(n-r)\times r},\ V \in \mathbb{C}^{r\times(m-r)}\}\ . \tag{91}$$

Indeed, it is easily seen that (1.1) and (1.2) are satisfied if X is taken as the product expression in RHS(91). Moreover, if

$$FG = I_n - A^{(1,2)}A\ , \quad HK = I_m - AA^{(1,2)}\ ,$$

it can be shown that

$$U = GXAY_0\ , \quad V = Z_0AXH\ . \tag{92}$$

It is found that the product in RHS(91) reduces to X if the expressions in (92) are substituted for U and V.

Relation (91) contains $r(m+n-2r)$ arbitrary parameters. This is less than the number in the efficient characterization (84) of $A\{1\}$ by $(m-r)(n-r)$, which vanishes only if A is of full (row or column) rank, in which case every $\{1\}$–inverse is a $\{1,2\}$–inverse.

Exercises.

Ex. 64. In (85) obtain explicit formulas for G, H, and D in terms of $A, A^{(1)}, F, K, B$, and $\{1\}$–inverses of the latter three matrices.

Ex. 65. Consider the problem of obtaining all]1]–inverses of the matrix A of Ex. 0.46. Note that the parametric representation of Ex. 1.6 does not give all $\{1\}$–inverses. (In this connection see Ex. 1.11.) Obtain in two ways parametric representation that do in fact give all $\{1\}$–inverses: first by (4) and then by (84). Note that a very simple $\{1\}$–inverse (in fact, a $\{1,2\}$–inverse) is obtained by taking all the arbitrary parameters equal to zero in the representation of Ex. 1.6. Verify that possible choices of F and K are

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1+2i & 0 \\ 0 & -2 & 0 & i \\ 0 & 0 & -2 & -1-i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad K = \begin{bmatrix} 3i & 1 & 3 \end{bmatrix}\ .$$

Compare the number of arbitrary parameters in the two representations.

Ex. 66. Under the hypotheses of Theorem 10, let F and K^* be matrices whose columns are bases for $N(A)$ and $N(A^*)$, respectively. Then, (42) can be written in the alternative form

$$X = A^{(1,2)}_{T,S} + FZK\ , \tag{93}$$

where Z is an arbitrary element of $\mathbb{C}^{(n-r)\times(m-r)}$. Moreover,

$$\operatorname{rank} X = r + \operatorname{rank} Z\ . \tag{94}$$

Proof. Clearly the right member of (93) satisfies (41). On the other hand, substituting in (42) the first two equations (85) gives (93) with $Z = GYH$.

Moreover, (93) and Theorem 10(c) give

$$XP_{L,S} = A^{(1,2)}_{T,S}\ ,$$

and therefore

$$X(I_m - P_{L,S}) = FZK \ .$$

Consequently, $R(X)$ contains the range of each of the two terms of RHS(93). Furthermore, the intersection of the latter two ranges is $\{\mathbf{0}\}$, since $R(F) = N(A) = M$, which is a subspace complementary to $T = R(A^{(1,2)}_{T,S})$. Therefore, $R(X)$ is the direct sum of the two ranges mentioned, and, by statement (c) of Ex. 0.1, $\operatorname{rank} X$ is the sum of the ranks of the two terms in RHS(93).

Now, the first term is a $\{1,2\}$–inverse of A, and its rank is therefore r by Theorem 1.2, while the rank of the second term is $\operatorname{rank} Z$ by Ex. 1.7. This establishes (94). □

Ex. 67. Exercise 66 gives

$$A\{1\}_{T,S} = \{A^{(1,2)}_{T,S} + FZK \, : \, Z \in \mathbb{C}^{(n-r)\times(m-r)} \ ,$$

where $A\{1\}_{T,S}$ is defined in Corollary 8. Show that this characterization is efficient.

Ex. 68. Show that if $A \in \mathbb{C}^{m\times n}$, $A\{1\}_{T,S}$ contains matrices of all ranks from r to $\min\{m,n\}$.

Ex. 69. Let $A = ST$ be a full–rank factorization of $A \in \mathbb{C}^{m\times n}_r$, let Y_0 and Z_0 be particular $\{1\}$–inverses of T and S, respectively, and let F and K be defined as in Ex. 66. Then, show that:

$$\begin{aligned}
S\{1\} &= \{Z_0 + VK \, : \, V \in \mathbb{C}^{r\times(m-r)}\} \ , \\
T\{1\} &= \{Y_0 + FU \, : \, U \in \mathbb{C}^{(n-r)\times r}\} \ , \\
AA\{1\} &= SS\{1\} = \{S(Z_0 + VK) \, : \, V \in \mathbb{C}^{r\times(m-r)}\} \ , \\
A\{1\}A &= T\{1\}T = \{(Y_0 + FU)T \, : \, U \in \mathbb{C}^{(n-r)\times r}\} \ , \\
A\{1\} &= \{Y_0Z_0 + Y_0VK + FUZ_0 + FWK \, : \, U \in \mathbb{C}^{(n-r)\times r}, V \in \mathbb{C}^{r\times(m-r)}, W \in \mathbb{C}^{(n-r)\times(m-r)}\} \ , \\
&= A\{1,2\} + \{FXK \, : \, X \in \mathbb{C}^{(n-r)\times(m-r)}\} \ .
\end{aligned}$$

Show that all the preceding characterizations are efficient.

Ex. 70. For the matrix A of Exs. 65 and 0.46, obtain all the characterizations of Ex. 69. *Hint*: Use the full–rank factorization of A given at the end of Section 1.7 and take

$$Z_0 = \begin{bmatrix} -\frac{1}{2}i & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \end{bmatrix} \ .$$

8. Restricted generalized inverses

In a linear equation

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

with given $a \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$, the points $\mathbf{x}$ are sometimes constrained to lie in a given subspace S of $\mathbb{C}^n$, resulting in a "constrained" linear equation

$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad \mathbf{x} \in S \ . \tag{95}$$

In principle, this situation presents no difficulty since (95) is equivalent to the following, "unconstrained" but larger, linear system

$$\begin{bmatrix} A \\ P_{S^\perp} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \ , \qquad \text{where } P_{S^\perp} = I - P_S \ .$$

Another approach to the solution of (95) that does not increase the size of the problem is to interpret A as representing an element of $\mathcal{L}(S, \mathbb{C}^m)$, the space of linear transformations from S to $\mathbb{C}^m$, instead of an element of $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$, see, e.g. Sections 4 and 6.1. This interpretaion calls for the following definitions.

Let $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$, and let S be a subspace of $\mathbb{C}^n$. The *restriction* of A to S, denoted by $A_{[S]}$, is a linear transformation from S to $\mathbb{C}^m$ defined by

$$A_{[S]}\mathbf{x} = A\mathbf{x} \,, \quad \mathbf{x} \in S \,. \tag{96}$$

Conversely, let $B \in \mathcal{L}(S, \mathbb{C}^m)$. The *extension* of B to $\mathbb{C}^n$, denoted by $\operatorname{ext} B$, is the linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$ defined by

$$(\operatorname{ext} B)\mathbf{x} = \begin{cases} B\mathbf{x} & \text{if } \mathbf{x} \in S, \\ \mathbf{0} & \text{if } \mathbf{x} \in S^\perp. \end{cases} \tag{97}$$

Restricting an $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ to S and then extending to $\mathbb{C}^n$ results in $\operatorname{ext}(A_{[S]}) \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ given by

$$\operatorname{ext}(A_{[S]})\mathbf{x} = \begin{cases} A\mathbf{x} & \text{if } \mathbf{x} \in S, \\ \mathbf{0} & \text{if } \mathbf{x} \in S^\perp. \end{cases} \tag{98}$$

From (98) it should be clear that if $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ is represented by the matrix $A \in \mathbb{C}^{m \times n}$, then $\operatorname{ext}(A_{[S]})$ is represented by AP_S. The following lemma is then obvious.

LEMMA 4. Let $A \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let S be a subspace of $\mathbb{C}^n$. The system

$$A\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} \in S \tag{95}$$

is consistent if and only if the system

$$AP_S\mathbf{z} = \mathbf{b} \tag{99}$$

is consistent, in which case $\mathbf{x}$ is a solution of (95) if and only if

$$\mathbf{x} = P_S\mathbf{z} \,,$$

where $\mathbf{z}$ is a solution of (99). □

From Lemma 4 and Corollary 2 it follows that the general solution of (95) is

$$\mathbf{x} = P_S(AP_S)^{(1)}\mathbf{b} + P_S(I - (AP_S)^{(1)}AP_S)\mathbf{y} \,, \tag{100}$$
$$\text{for arbitrary } (AP_S)^{(1)} \in (AP_S)\{1\} \text{ and } \mathbf{y} \in \mathbb{C}^n \,.$$

We are thus led to study generalized inverses of $\operatorname{ext}(A_{[S]}) = AP_S$, and from (100) it appears that $P_S(AP_S)^{(1)}$, rather than $A^{(1)}$, plays the role of a $\{1\}$–inverse in solving the linear system (95); hence the following definition.

DEFINITION 1. Let $A \in \mathbb{C}^{m \times n}$ and let S be a subspace of $\mathbb{C}^n$. A matrix $X \in \mathbb{C}^{n \times m}$ is an *S–restricted $\{i, j, \dots, k\}$–inverse of* A if

$$X = P_S(AP_S)^{(i,j,\dots,k)} \tag{101}$$

for any $(AP_S)^{(i,j,\dots,k)} \in (AP_S)\{i, j, \dots, k\}$.

The role that S–restricted generalized inverses play in constrained problems is completely analogous to the role played by the corresponding generalized inverse in the unconstrained situation. Thus, for example, the following result is the constrained analog of Corollary 2.

COROLLARY 10. Let $A \in \mathbb{C}^{m \times n}$ and let S be a subspace of $\mathbb{C}^n$. Then the equation

$$A\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} \in S \tag{95}$$

is consistent if and only if

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

where X is any S–restricted $\{1\}$–inverse of A. If consistent, the general solution of (95) is

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

with X as above, and arbitrary $\mathbf{y} \in S$. □

Exercises.

EX. 71. Let I be the identity transformation in $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$ and let S be a subspace of $\mathbb{C}^n$. Show that

$$\text{ext}\,(I_{[S]}) = P_S\ .$$

EX. 72. Let $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$. Show that $A_{[R(A^*)]}$, the restriction of A to $R(A^*)$, is a one–to–one mapping of $R(A^*)$ onto $R(A)$.

SOLUTION. We show first that $A_{[R(A^*)]}$ is one–to–one on $R(A^*)$. Clearly it suffices to show that A is one–to–one on $R(A^*)$. Let $\mathbf{u}, \mathbf{v} \in R(A^*)$ and suppose that $A\mathbf{u} = A\mathbf{v}$, i.e., $\mathbf{u}$ and $\mathbf{v}$ are mapped to the same point. Then $A(\mathbf{u} - \mathbf{v}) = \mathbf{0}$, i.e.,

$$\mathbf{u} - \mathbf{v} \in N(A)\ .$$

But we also have

$$\mathbf{u} - \mathbf{v} \in R(A^*)\ ,$$

since $\mathbf{u}$ and $\mathbf{v}$ are in $R(A^*)$. Therefore

$$\mathbf{u} - \mathbf{v} \in N(A) \cap R(A^*)$$

and by (47), $\mathbf{u} = \mathbf{v}$, proving the A is one–to–one on $R(A^*)$.

We show next that $A_{[R(A^*)]}$ is a mapping onto $R(A)$, i.e., that

$$R(A_{[R(A^*)]}) = R(A)\ .$$

This follows since for any $\mathbf{x} \in \mathbb{C}^n$

$$A\mathbf{x} = AA^\dagger A\mathbf{x} = AP_{R(A^*)}\mathbf{x} = A_{[R(A^*)]}\mathbf{x}\ ,$$

EX. 73. Let $A \in \mathbb{C}^{m\times n}$. Show that

$$\text{ext}\,(A_{[R(A^*)]}) = A\ . \tag{102}$$

EX. 74. From Ex. 72 it follows that the linear transformation

$$A_{[R(A^*)]} \in \mathcal{L}(R(A^*), R(A))$$

has an inverse

$$(A_{[R(A^*)]})^{-1} \in \mathcal{L}(R(A), R(A^*))$$

Show that this inverse is the restriction of $A^\dagger$ to $R(A)$, namely

$$(A^\dagger)_{[R(A)]} = (A_{[R(A^*)]})^{-1}\ . \tag{103}$$

SOLUTION. From Exs. 72,29, and 45 it follows that, for any $\mathbf{y} \in R(A)$, $A^\dagger\mathbf{y}$ is the unique element of $R(A^*)$ satisfying

$$A\mathbf{x} = \mathbf{y}\ .$$

Therefore

$$A^\dagger\mathbf{y} = (A_{[R(A^*)]})^{-1}\mathbf{y} \quad \text{for all } \mathbf{y} \in R(A)\ .$$

EX. 75. Show that the extension of $(A_{[R(A^*)]})^{-1}$ to $\mathbb{C}^m$ is the Moore–Penrose inverse of A,

$$\text{ext}\,\left((A_{[R(A^*)]})^{-1}\right) = A^\dagger\ . \tag{104}$$

Compare with (102).

Ex. 76. Let each of the following two linear equations be consistent

$$A_1\mathbf{x} = \mathbf{b}_1\ , \tag{105a}$$
$$A_2\mathbf{x} = \mathbf{b}_2\ . \tag{105b}$$

Show that (105a) and (105b) have a common solution if and only if the linear equation

$$A_2 P_{N(A_1)}\mathbf{y} = \mathbf{b}_2 - A_2 A_1^{(1)}\mathbf{b}_1$$

is consistent, in which case the general common solution of (105a) and (105b) is

$$\mathbf{x} = A_1^{(1)}\mathbf{b}_1 + P_{N(A_1)}(A_2 P_{N(A_1)})^{(1)}(\mathbf{b}_2 - A_2 A_1^{(1)}\mathbf{b}_1) + N(A_1)\cap N(A_2)$$

or equivalently

$$\mathbf{x} = A_2^{(1)}\mathbf{b}_2 + P_{N(A_2)}(A_1 P_{N(A_2)})^{(1)}(\mathbf{b}_1 - A_1 A_2^{(1)}\mathbf{b}_2) + N(A_1)\cap N(A_2)$$

Hint. Substitute the general solution of (105a)

$$\mathbf{x} = A_1^{(1)}\mathbf{b}_1 + P_{N(A_1)}\mathbf{y}\ , \quad \mathbf{y} \text{ arbitrary}\ ,$$

in (105b).

Ex. 77. Exercise 76 illustrates the need for $P_{N(A_1)}(A_2 P_{N(A_1)})^{(1)}$, an $N(A_1)$–restricted $\{1\}$–inverse of A_2. Other applications call for other, similarly restricted, generalized inverses. The $N(A_1)$–restricted $\{1,2,3,4\}$–inverse of A_2 was studied for certain Hilbert space operators, by Minamide and ([**1054**] and [**1055**]), who characterized it as the unique solution X of the five equations

$$A_1 X = O\ ,$$
$$A_2 X A_2 = A_2 \text{ on } N(A_1)\ ,$$
$$X A_2 X = X\ ,$$
$$(A_2 X)^* = A_2 X\ ,$$

and

$$P_{N(A_1)}(XA_2)^* = XA_2 \text{ on } N(A_1)\ .$$

Show that $P_{N(A_1)}(A_2 P_{N(A_1)})^\dagger$ is the unique solution of these five equations.

9. The Bott–Duffin inverse

Consider the constrained system

$$A\mathbf{x} + \mathbf{y} = \mathbf{b}\ ,\ x \in L\ ,\ \mathbf{y} \in L^\perp\ , \tag{106}$$

with given $A \in \mathbb{C}^{n\times n}$, $\mathbf{b} \in \mathbb{C}^n$, and a subspace L of $\mathbb{C}^n$. Such systems arise in electrical network theory; see, e.g., Bott and Duffin [**202**] and Section 12 below. As in Section 8 we conclude that the consistency of (106) is equivalent to the consistency of the following system:

$$(AP_L + P_{L^\perp})\,\mathbf{z} = \mathbf{b} \tag{107}$$

and that $\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}$ is a solution of (106) if and only if

$$\mathbf{x} = P_L\mathbf{z}, \quad \mathbf{y} = P_{L^\perp}\mathbf{z} = \mathbf{b} - AP_L\mathbf{z}\ , \tag{108}$$

where $\mathbf{z}$ is a solution of (107).

If the matrix $(AP_L + P_{L^\perp})$ is nonsingular, then (106) is consistent for all $\mathbf{b} \in \mathbb{C}^m$ and the solution

$$\mathbf{x} = P_L(AP_L + P_{L^\perp})^{-1}\mathbf{b}\ , \quad \mathbf{y} = \mathbf{b} - A\mathbf{x}$$

is unique. The transformation

$$P_L(AP_L + P_{L^\perp})^{-1}$$

was introduced and studied by Bott and Duffin [**202**], who called it the *constrained inverse* of A. Since it exists only when $(AP_L + P_{L^\perp})$ is nonsingular, one may be tempted to introduce generalized inverses of this form, namely

$$P_L(AP_L + P_{L^\perp})^{(i,j,\dots,k)} \quad (1 \le i, j, \dots, k \le 4) ,$$

which do exist for all A and L. This section, however, is restricted to the Bott–Duffin inverse.

Definition 2. Let $A \in \mathbb{C}^{n\times n}$ and let L be a subspace of $\mathbb{C}^n$. If $(AP_L + P_{L^\perp})$ is nonsingular, the *Bott–Duffin inverse of A with respect to L*, denoted by $A^{(-1)}_{(L)}$, is defined by

$$A^{(-1)}_{(L)} = P_L(AP_L + P_{L^\perp})^{-1} . \tag{109}$$

Some properties of $A^{(-1)}_{(L)}$ are collected in

Theorem 14. (Bott and Duffin [**202**]). Let $(AP_L + P_{L^\perp})$ be nonsingular. Then:
(a) The equation

$$A\mathbf{x} + \mathbf{y} = \mathbf{b} \ , \ x \in L \ , \ \mathbf{y} \in L^\perp \tag{106}$$

has for every $\mathbf{b}$, the unique solution

$$\mathbf{x} = A^{(-1)}_{(L)}\mathbf{b} \ , \tag{110a}$$

$$\mathbf{y} = (I - AA^{(-1)}_{(L)})\mathbf{b} \ . \tag{110b}$$

(b) A, P_L, and $A^{(-1)}_{(L)}$ satisfy

$$P_L = A^{(-1)}_{(L)}AP_L = P_LAA^{(-1)}_{(L)} \ , \tag{111a}$$

$$A^{(-1)}_{(L)} = P_LA^{(-1)}_{(L)} = A^{(-1)}_{(L)}P_L \ . \tag{111b}$$

Proof. (a) This follows from the equivalence of (106) and (107)–(108).
(b) From (109), $P_LA^{(-1)}_{(L)} = A^{(-1)}_{(L)}$, Postmultiplying $A^{(-1)}_{(L)}(AP_L + P_{L^\perp}) = P_L$ by P_L gives $A^{(-1)}_{(L)}AP_L = P_L$. Therefore $A^{(-1)}_{(L)}P_{L^\perp} = O$ and $A^{(-1)}_{(L)}P_L = A^{(-1)}_{(L)}$. Multiplying (110b) by P_L gives $(P_L - P_LAA^{(-1)}_{(L)})\mathbf{b} = \mathbf{0}$ for all $\mathbf{b}$, thud $P_L = P_LAA^{(-1)}_{(L)}$). $\square$

From these results it follows that the Bott–Duffin inverse $A^{(-1)}_{(L)}$, whenever it exists, is the $\{1,2\}$–inverse of (P_LAP_L) having range L and null space $L^\perp$.

Corollary 11. If $AP_L + P_{L^\perp}$ is nonsingular, then
(a) $A^{(-1)}_{(L)} = (AP_L)^{(1,2)}_{L,L^\perp} = (P_LA)^{(1,2)}_{L,L^\perp} = (P_LAP_L)^{(1,2)}_{L,L^\perp}$,
(b) $(A^{(-1)}_{(L)})^{(-1)}_{(L)} = P_LAP_L$.

Proof. (a) From (111a), $\dim L = \operatorname{rank} P_L \le \operatorname{rank} A^{(-1)}_{(L)}$. Similarly from (111b), $\operatorname{rank} A^{(-1)}_{(L)} \le \dim L$, $R(A^{(-1)}_{(L)}) \subset R(P_L) = L$ and $N(A^{(-1)}_{(L)}) \supset N(P_L) = L^\perp$. Therefore

$$\operatorname{rank} A^{(-1)}_{(L)} = \dim L \tag{112}$$

and

$$R(A^{(-1)}_{(L)}) = L, \quad N(A^{(-1)}_{(L)}) = L^\perp \ . \tag{113}$$

Now $A^{(-1)}_{(L)}$ is a $\{1,2\}$–inverse of AP_L:

$$AP_LA^{(-1)}_{(L)}AP_L = AP_L \quad \text{by (111a ,}$$

and

$$A_{(L)}^{(-1)} A P_L A_{(L)}^{(-1)} = A_{(L)}^{(-1)} \quad \text{by (111a) and (111b)} . \tag{114}$$

That $A_{(L)}^{(-1)}$ is a $\{1,2\}$–inverse of $P_L A$ and of $P_L A P_L$ is similarly proved.
(b) We show first that $(A_{(L)}^{(-1)})_{(L)}^{(-1)}$ is defined, i.e., that $(A_{(L)}^{(-1)} P_L + P_{L^\perp})$ is nonsingular. From (111b), $A_{(L)}^{(-1)} P_L + P_{L^\perp} = A_{(L)}^{(-1)} + P_{L^\perp}$, which is a nonsingular matrix since its columns span $L + L^\perp = \mathbb{C}^n$, by (113). Now $P_L A P_L$ is a $\{1,2\}$–inverse of $A_{(L)}^{(-1)}$, by (a), and therefore by Theorem 1.2 and (112),

$$\operatorname{rank} P_L A P_L = \operatorname{rank} A_{(L)}^{(-1)} = \dim L .$$

This result. together with

$$R(P_L A P_L) \subset R(P_L) = L, \quad N(P_L A P_L) \supset N(P_L) = L^\perp ,$$

shows that

$$R(P_L A P_L) = L, \quad N(P_L A P_L) = L^\perp ,$$

proving that

$$\begin{aligned} P_L A P_L &= (A_{(L)}^{(-1)})_{L,L^\perp}^{(1,2)} \\ &= (A_{(L)}^{(-1)})_{(L)}^{(-1)} . \end{aligned}$$

□

Exercises.

Ex. 78. Show that the following statements are equivalent, for any $A \in \mathbb{C}^{n\times n}$ and a subspace $L \subset \mathbb{C}^n$.
(a) $AP_L + P_{L^\perp}$ is nonsingular.
(b) $\mathbb{C}^n = AL \oplus L^\perp$, i.e., $AL = \{A\mathbf{x} : \mathbf{x} \in L\}$ and $L^\perp$ are complementary subspaces of $\mathbb{C}^n$.
(c) $\mathbb{C}^n = P_L R(A) \oplus L^\perp$.
(d) $\mathbb{C}^n = P_L AL \oplus L^\perp$.
(e) $\operatorname{rank} P_L A P_L = \dim L$.
Thus, each of the above conditions is necessary and sufficient for the existence of $A_{(L)}^{(-1)}$, the Bott–Duffin inverse of A with respect to L.

Ex. 79. *A converse to Corollary 11.* If any one of the following three $\{1,2\}$–inverses exist

$$(AP_L)_{L,L^\perp}^{(1,2)} , \quad (P_L A)_{L,L^\perp}^{(1,2)} , \quad (P_L A P_L)_{L,L^\perp}^{(1,2)} ,$$

then all three exist, $AP_L + P_{L^\perp}$ is nonsingular, and

$$(AP_L)_{L,L^\perp}^{(1,2)} = (P_L A)_{L,L^\perp}^{(1,2)} = (P_L A P_L)_{L,L^\perp}^{(1,2)} = A_{(L)}^{(-1)} .$$

Hint. Condition (b) in Ex. 78 is equivalent to the existence of $(AP_L)_{L,L^\perp}^{(1,2)}$.

Ex. 80. Let K be a matrix whose columns form a basis for L. Then $A_{(L)}^{(-1)}$ exists if and only if K^*AK is nonsingular, in which case

$$A_{(L)}^{(-1)} = K(K^*AK)^{-1}K^* \quad \text{(Bott and Duffin [\textbf{202}])} .$$

PROOF. Follows from Corollary 11 and Theorem 11(d). □

Ex. 81. If A is Hermitian and and $A_{(L)}^{(-1)}$ exists, then $A_{(L)}^{(-1)}$ is Hermitian.

Ex. 82. Using the notation

$$\begin{aligned} A &= [a_{ij}] \quad (i,j \in \overline{1,n}) \\ A^{(-1)}_{(L)} &= [t_{ij}] \quad (i,j \in \overline{1,n}) \\ d_{A,L} &= \det(AP_L + L^{\perp}) \,, & (115) \\ \psi_{A,L} &= \log d_{A,L} & (116) \end{aligned}$$

show that

(a) $\dfrac{\partial \psi_{A,L}}{\partial a_{ij}} = t_{ji} \quad (i,j \in \overline{1,n})$

(b) $\dfrac{\partial t_{kl}}{\partial a_{ij}} = t_{ki}t_{jl} \quad (i,j,k,l \in \overline{1,n})$. (Bott and Duffin, [**202**, Theorem 3]).

Bott and Duffin called $d_{A,L}$ the *discriminant* of A, and $\psi_{A,L}$ the *potential* of $A^{(-1)}_{(L)}$.

Ex. 83. Let $A \in \mathbb{C}^{n\times n}$ be nonsingular, and let L be a subspace of $\mathbb{C}^n$. Then $A^{(-1)}_{(L)}$ exists if and only if $A^{(-1)}_{(L^{\perp})}$ exists.

Hint. Use $A^{-1}P_{L^{\perp}} + P_L = A^{-1}(AP_L + P_{L^{\perp}})$ to show that $(A^{-1}P_{L^{\perp}} + P_L)^{-1} = (AP_L + P_{L^{\perp}})^{-1}A$.

Ex. 84. Let $A \in \mathbb{C}^{n\times n}$ be nonsingular, let L be a subspace of $\mathbb{C}^n$, let $d_{A,L}$ and $\psi_{A,L}$ be given by (107) and (116), respectively, and similarly define

$$\begin{aligned} d_{A^{-1},L^{\perp}} &= \det(A^{-1}P_{L^{\perp}} + P_L) \\ \psi_{A^{-1},L^{\perp}} &= \log d_{A^{-1},L^{\perp}} \,. \end{aligned}$$

Then

(a) $d_{A^{-1},L^{\perp}} = \dfrac{d_{A,L}}{\det A}$

(b) $(A^{-1})^{(-1)}_{(L^{\perp})} = A - AA^{(-1)}_{(L)}A$. (Bott and Duffin, [**202**, Theorem 4]).

Ex. 85. If $\Re\langle A\mathbf{u},\mathbf{u}\rangle > 0$ for every nonzero vector $\mathbf{u}$, then $d_{A,L} \neq 0$, $\Re\langle A^{(-1)}_{(L)}\mathbf{u},\mathbf{u}\rangle \geq 0$ for every vector $\mathbf{u}$ and $\Re(t_{ii}) \geq 0$, where $A^{(-1)}_{(L)} = [t_{ij}]$. (Bott and Duffin, [**202**, Theorem 6]).

Ex. 86. Let $A, B \in \mathbb{C}^{n\times n}$ and let L be a subspace of $\mathbb{C}^n$ such that both $A^{(-1)}_{(L)}$ and $B^{(-1)}_{(L)}$ exist. Then

$$B^{(-1)}_{(L)}A^{(-1)}_{(L)} = (AP_LB)^{-1}_{(L)} \,.$$

10. An application of $\{1\}$–inverses in interval linear programming

For two vectors $\mathbf{u} = (u_i), \mathbf{v} = (v_i) \in \mathbb{R}^m$ let

$$\mathbf{u} \leq \mathbf{v}$$

denote the fact that $u_i \leq v_i$ for $i = 1,\dots,m$. A linear programming problem of the form

$$\text{maximize}\,\{\mathbf{c}^T\mathbf{x} :\ \mathbf{a} \leq A\mathbf{x} \leq \mathbf{b}\}\,, \tag{117}$$

with given $\mathbf{a},\mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$; $A \in \mathbb{R}^{m\times n}$, is called an *interval linear program* (also a *linear program with two–sided constraints*) and denoted by $IP(\mathbf{a},\mathbf{b},\mathbf{c},A)$ or simply by IP. Any linear programming problem with bounded constraint set can be written as an IP, see e.g. Robers and Ben-Israel [**1277**].

In this section, which is based on the work of Ben-Israel and Charnes [**128**], the optimal solutions of (117) are obtained by using $\{1\}$–inverses of A, in the special case where A is of full row rank. More general cases were studied by Zlobec and Ben-Israel [**1654**], [**1655**] (see also Exs. 87 and 88), and an iterative method for solving the general IP appears in Robers and Ben-Israel [**1277**]. Applications of interval programming are given in Ben-Israel, Charnes and Robers [**129**], and Robers and Ben-Israel [**1276**]. References for other applications of generalized inverses in linear programming are Pyle [**1224**] and Cline and Pyle [**359**].

The IP (117) is called *consistent* (also *feasible*) if the set

$$F = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \le A\mathbf{x} \le \mathbf{b}\} \tag{118}$$

is nonempty, in which case the elements of F are called the feasible solutions of $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$. A consistent $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is called *bounded* if

$$\max\{\mathbf{c}^T\mathbf{x} : x \in F\}$$

is finite, in which case the optimal solutions of $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ are its feasible solutions $\mathbf{x}_0$ which satisfy

$$\mathbf{c}^T\mathbf{x}_0 = \max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in F\}\ .$$

Boundedness is equivalent to $\mathbf{c} \in R(A^T)$ as the following lemma shows.

Lemma 5. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$; $A \in \mathbb{R}^{m\times n}$ be such that $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is consistent. Then $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is bounded if and only if

$$\mathbf{c} \in N(A)^{\perp}\ . \tag{119}$$

Proof. From (118), $F = F + N(A)$. Therefore

$$\begin{aligned}\max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in F\} &= \max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in F + N(A)\}\\ &= \max\{(P_{R(A^T)}\mathbf{c} + P_{N(A)}\mathbf{c})^T\mathbf{x} : \mathbf{x} \in F + N(A)\}\ , \quad \text{by (47)}\\ &= \max\{\mathbf{c}^T P_{R(A^T)}\mathbf{x} : \mathbf{x} \in F\} + \max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in N(A)\}\ ,\end{aligned}$$

where the first term

$$\max\{\mathbf{c}^T P_{R(A^T)}\mathbf{x} : \mathbf{x} \in F\} = \max\{\mathbf{c}^T A^{\dagger}A\mathbf{x} : \mathbf{a} \le A\mathbf{x} \le \mathbf{b}\}$$

is finite, and the second term

$$\max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in N(A)\}$$

is finite if and only if $\mathbf{c} \in N(A)^{\perp}$. □

We introduce now a function $\eta : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$, defined for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^m$ by

$$\eta(\mathbf{u}, \mathbf{v}, \mathbf{w}) = [\eta_i] \quad (i \in \overline{1, m})$$

where

$$\eta_i = \begin{cases} u_i & \text{if } w_i < 0,\\ v_i & \text{if } w_i > 0,\\ \lambda_i u_i + (1-\lambda_i)v_i & \text{where } 0 \le \lambda_i \le 1\ ,\ \text{if } w_i = 0\end{cases} \tag{120}$$

A component of $\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is equal to the corresponding component of $\mathbf{u}$ or $\mathbf{v}$, if the corresponding component of $\mathbf{w}$ is negative or positive, respectively. If a component of $\mathbf{w}$ is zero, then the corresponding component of $\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is the closed interval with the corresponding components of $\mathbf{u}$ and $\mathbf{v}$ as endpoints. Thus η maps points in $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ into sets in $\mathbb{R}^m$, and any statement below about $\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is meant for all values of $\eta(\mathbf{u}, \mathbf{v}, \mathbf{w})$, unless otherwise specified.

The next result gives all the optimal solutions of $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ with A of full row rank.

Theorem 15. (Ben-Israel and Charnes [**128**]). Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$; $A \in \mathbb{R}^{m\times n}$ be such that $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is consistent and bounded, and let $A^{(1)}$ be any {1}–inverse of A. Then the general optimal solution of $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is

$$\mathbf{x} = A^{(1)}\eta(\mathbf{a}, \mathbf{b}, A^{(1)T}\mathbf{c}) + \mathbf{y}\ , \quad \mathbf{y} \in N(A)\ . \tag{121}$$

PROOF. From $A \in \mathbb{R}_m^{m\times n}$ it follows that $R(A) = \mathbb{R}^m$, so that any $\mathbf{u} \in \mathbb{R}^m$ can be written as

$$\mathbf{u} = A\mathbf{x} \tag{122}$$

where

$$\mathbf{x} = A^{(1)}\mathbf{u} + \mathbf{y} , \quad \mathbf{y} \in N(A) , \quad \text{by Corollary 2.} \tag{123}$$

Substituting (122) and (123) in (117), we get, by using (119), the equivalent IP

$$\max\{\mathbf{c}^T A^{(1)}\mathbf{u} : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$$

whose general optimal solution is, by the definition (120) of η,

$$\mathbf{u} = \eta(\mathbf{a}, \mathbf{b}, A^{(1)T}\mathbf{c})$$

which gives (121) by using (123). □

Exercises.

EX. 87. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$; $A \in \mathbb{R}^{m\times n}$ be such that $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ is consistent and bounded. Let $A^{(1)} \in A\{1\}$ and let $\mathbf{z}_0 \in N(A^T)$ satisfy

$$\mathbf{z}^T \boldsymbol{\eta}_0 \le 0$$

for some $\boldsymbol{\eta}_0 \in \eta(\mathbf{a}, \mathbf{b}, (A^{(1)}P_{R(A)})^T\mathbf{c} + \mathbf{z}_0)$. Then

$$\mathbf{x}_0 = A^{(1)}P_{R(A)}\boldsymbol{\eta}_0 + \mathbf{y} , \quad \mathbf{y} \in N(A)$$

is an optimal solution of $IP(\mathbf{a}, \mathbf{b}, \mathbf{c}, A)$ if and only if it is a feasible solution (Zlobec and Ben-Israel [**1655**]).

EX. 88. Let $\mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$; $A \in \mathbb{R}^{m\times n}$ and let $\mathbf{u} \in \mathbb{R}^n$ be a positive vector such that the problem

$$\min\{\mathbf{c}^T\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{0} \le \mathbf{x} \le \mathbf{u}\} \tag{124}$$

is consistent. Let $\mathbf{z}_0 \in R(A^T)$ satisfy

$$\mathbf{z}^T \boldsymbol{\eta}_0 \le \mathbf{z}^T A^\dagger \mathbf{b}$$

for some $\boldsymbol{\eta}_0 \in \eta(\mathbf{0}, \mathbf{u}, P_{N(A)}\mathbf{c} + \mathbf{z}_0)$. Then

$$\mathbf{x}_0 = A^\dagger \mathbf{b} + P_{N(A)}\boldsymbol{\eta}_0$$

is an optimal solution of (124) if and only if it is a feasible solution (Zlobec and Ben-Israel [**1655**]).

11. A $\{1, 2\}$–inverse for the integral solution of linear equations

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*.

Let $A \in \mathbb{Z}^{m\times n}, \mathbf{b} \in \mathbb{Z}^m$ and let the linear equation

$$A\mathbf{x} = \mathbf{b} \tag{5}$$

be consistent. In many applications one has to determine if (5) has integral solutions, in which case one has to find some or all of them. If A is nonsingular and its inverse is also integral, then (5) has the unique integral solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any integral $\mathbf{b}$. 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 [**996**, p. 42].

In this section, which is based on the work of Hurt and Waid [**760**], we study the integral solution of (5) for any $A \in \mathbb{Z}^{m\times n}$ and $\mathbf{b} \in \mathbb{Z}^m$. Using the Smith normal form of A (Theorem 16 below), a $\{1,2\}$–inverse is found (Corollary 12) which can be used to determine the existence of integral solutions, and to list all of them if they exist (Corollaries 13 and 14).

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

$$PAQ = S \; . \tag{125}$$

Theorem 16. 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 [**996**, pp. 42–44].

Proof. The proof given in Marcus and Minc [**996**, 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

$$\text{interchanging two rows [columns]} \tag{126}$$

$$\text{subtracting an integer multiple of one row [column] from another row [column]} \tag{127}$$

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 (125) is the product of all the elementary row [column] operators, in the right order. □

Using the Smith normal form, a $\{1,2\}$–inverse with special integral properties can now be given.

Corollary 12. (Hurt and Waid [**760**]). Let $A \in \mathbb{Z}^{m\times n}$. Then there is an $n \times m$ matrix X satisfying

$$AXA = A \; , \tag{1.1}$$

$$XAX = X \; , \tag{1.2}$$

$$AX \in \mathbb{Z}^{m\times m}, \quad XA \in \mathbb{Z}^{n\times n} \; . \tag{128}$$

Proof. Let

$$PAQ = S \tag{125}$$

be the Smith normal form of A, and let

$$\widehat{A} = QS^{\dagger}P \; . \tag{129}$$

Then

$$PAQ = S = SS^{\dagger}S = PAQS^{\dagger}PAQ = PA\widehat{A}AQ \; ,$$

proving $A = A\widehat{A}A$. $\widehat{A}A\widehat{A} = \widehat{A}$ is similarly proved. The integrality of $A\widehat{A}$ and $\widehat{A}A$ follows from that of $PA\widehat{A} = SS^{\dagger}P$ and $\widehat{A}AQ = QS^{\dagger}S$, respectively. □

In the rest of this section we denote by $\widehat{A}, \widehat{B}$ the $\{1,2\}$–inverses of A, B as given in Corollary 12.

FIGURE 1. An example of a network

COROLLARY 13. (Hurt and Waid [**760**]). Let A, B, D be integral matrices, and let the matrix equation

$$AXB = D \tag{1}$$

be consistent. Then (1) has an integral solution if and only if the matrix

$$\widehat{A}D\widehat{B}$$

is integral, in which case the general integral solution of (1) is

$$X = \widehat{A}D\widehat{B} + Y - \widehat{A}AYB\widehat{B}\widehat{B}\ , \quad Y \in \mathbb{Z}^{n\times m}\ .$$

PROOF. Follows from Corollary 12 and Theorem 1. □

COROLLARY 14. (Hurt and Waid [**760**]). Let A and $\mathbf{b}$ be integral, and let the vector equation

$$A\mathbf{x} = \mathbf{b} \tag{5}$$

be consistent. Then (5) has an integral solution if and only if the vector

$$\widehat{A}\mathbf{b}$$

is integral, in which case the general integral solution of (5) is

$$\mathbf{x} = \widehat{A}\mathbf{b} + (I - \widehat{A}A)\mathbf{y}\ , \quad \mathbf{y} \in \mathbb{Z}^n\ .$$

Exercises.

EX. 89. 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 (126)–(127).

Hint. Use Ex. 1.3.

EX. 90. Describe in detail the algorithm mentioned in the proof of Theorem 16.

EX. 91. Use the results of Sections 10 and 11 to find the integral optimal solutions of the interval program

$$\max\{\mathbf{c}^T\mathbf{x} :\ \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and A are integral.

EX. 92. If $\mathbb{Z}$ is the *ring of polynomials with real coefficients*, or the *ring of polynomials with complex coefficients*, the results of this section hold; see, e.g., Marcus and Minc [**996**, p. 40]. Interpret Corollaries 12 and 14 in these two cases.

12. An application of the Bott–Duffin inverse to electrical networks

In this section which is based on Bott and Duffin [**202**], we keep the discussion of electrical networks at the minimum sufficient to illustrate the application of the Bott–Duffin inverse studied in Section 9. The reader is referred to the original work of Bott and Duffin for further information.

An *electrical network* is described topologically in terms of its *graph* consisting of *nodes* (also *vertices*, *junctions*, etc.) and *branches* (also *edges*), and electrically in terms of its (branch) *currents* and *voltages*.

Let the *graph* consist of m elements called *nodes* denoted by $n_i\ ,\ i \in \overline{1,m}$ (which, in the present limited discussion, can be represented by m points in the plane), and n ordered pairs of nodes called *branches* denoted by $b_j\ ,\ j \in \overline{1,n}$ (represented here by directed segments joining the paired nodes). For example, the network represented by Fig. 1 has four nodes n_1, n_2, n_3 and n_4, and six branches $b_1 = \{n_1, n_2\}$, $b_2 = \{n_2, n_3\}$, $b_3 = \{n_2, n_4\}$, $b_4 = \{n_3, n_1\}$, $b_5 = \{n_3, n_4\}$ and $b_6 = \{n_4, n_1\}$.

A graph with m nodes and n branches can be represented by an $m \times n$ matrix, called the *(node–branch) incidence matrix*, denoted by $M = [m_{ij}]$ and defined as follows:

(i) The ith row of M corresponds to the node n_i, $i \in \overline{1,m}$.

(ii) The jth column of M corresponds to the branch b_j, $j \in \overline{1,n}$.

(iii) If $b_j = \{n_k, n_l\}$, then

$$m_{ij} = \begin{cases} 1 & i = k, \\ -1 & i = l, \\ 0 & i \neq k, l. \end{cases}$$

For example, the incidence matrix of the graph of Fig. 1 is

$$M = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{bmatrix}$$

Two nodes n_k and n_l (or the corresponding rows of M) are called *directly connected* if either $\{n_k, n_l\}$ or $\{n_l, n_k\}$ is a branch, i.e. if there is a column in M having its nonzero entries in rows k and l. Two nodes n_k and n_l (or the corresponding rows of M) are called *connected* if there is a sequence of nodes

$$\{n_k, n_p, \dots, n_q, n_l\}$$

in which every two adjacent nodes are directly connected. Finally, a graph (or its incidence matrix) is called *connected* if every two nodes are connected.

In this section we consider only *direct current* (DC) *networks*, referring the reader to Bott and Duffin [**202**] and to Ex. 94 below, for alternating current (AC) networks. A DC network is described electrically in terms of two real valued functions, the *current* and the *potential*, defined on the sets of branches and nodes respectively.

For $j = 1, \dots, m$, the *current in branch* b_j, denoted by y_j, is the current (measured in amperes) flowing in b_j. The sign of y_j is *positive* if it flows in the direction of b_j, and is *negative* if it flows in the opposite direction.

For $i = 1, \dots, m$, the *potential* at node n_i, denoted by p_i, is the voltage difference (measured in volts) between n_i and some reference point, which can be taken as one of the nodes. A related function which is more often used, is the *voltage*, defined on the set of branches. For $j = 1, \dots, n$, the *voltage across branch* $b_j = \{n_k, n_l\}$, denoted by x_j, is defined as the potential difference

$$x_k = p_k = p_l \ .$$

From the definition of the incidence matrix M it is clear that the vector of branch voltages $\mathbf{x} = [x_j]$ and the vector of node potentials $\mathbf{p} = [p_i]$ are related by

$$\mathbf{x} = M^T \mathbf{p} \ . \tag{130}$$

The currents and voltages are assumed to satisfy Kirchhoff laws. The *Kirchhoff current law* is a conservation theorem for the currents (or electrical charges), stating that for each node, the net current entering the node is zero, i.e., the sum of incoming currents equals the sum of outgoing currents. From the definition of the incidence matrix M it follows that the Kirchhoff current law can be written as

$$M\mathbf{y} = \mathbf{0} \ . \tag{131}$$

The *Kirchhoff voltage law* states that the potential function is single valued. This statement usually assumes the equivalent form that the sum of the branch voltages directed around any closed circuit is zero.

From (130), (131,) and (47), it follows that the Kirchhoff current and voltage laws define two complementary orthogonal subspaces:

$N(M)$, the currents satisfying Kirchhoff current law;

$R(M^T)$, the voltages satisfying Kirchhoff voltage law.

Each branch b_j, $j \in \overline{1,n}$, of the network will be regarded as having a *series voltage generator* of v_j volts and a *parallel current generator* of w_j ampers. These are related to the branch currents and voltages by *Ohm's law*

$$a_j(x_j - v_j) + (y_j - w_j) = 0 \,, \quad j \in \overline{1,n} \,, \tag{132}$$

where $a_j > 0$ is the conductivity of the branch b_j, measured in mhos[2].

Thus the branch currents $\mathbf{y}$ and voltages $\mathbf{x}$ are found by solving the following constrained system:

$$A\mathbf{x} + \mathbf{y} = A\mathbf{v} + \mathbf{w} \,, \quad \mathbf{x} \in R(M^T) \,,\ \mathbf{y} \in N(M) \,, \tag{133}$$

where $A = [\text{diag} a_j]$ is the diagonal matrix of branch conductivities, $\mathbf{v}$ and $\mathbf{w}$ are are the given vectors of generated voltages and currents, respectively, and M is the incidence matrix. It can be shown that the Bott–Duffin inverse of A with respect to $R(M^T)$, $A^{(-1)}_{(R(M^T))}$, exists; see, e.g., Ex. 93 below. Therefore, by theorem 14, the unique solution of (133) is

$$\begin{aligned} \mathbf{x} &= A^{(-1)}_{(R(M^T))}(A\mathbf{v} + \mathbf{w}) \,, && (134) \\ \mathbf{y} &= (I - AA^{(-1)}_{(R(M^T))})(A\mathbf{v} + \mathbf{w}) \,. && (135) \end{aligned}$$

The physical significance of the matrix $A^{(-1)}_{(R(M^T))}$ should be clear from (134). The (i,j)th entry of $A^{(-1)}_{(R(M^T))}$ is the voltage across branch b_i as a result of inserting a current source of one ampere in branch b_j; $i, j \in \overline{1,n}$. Because of this property, $A^{(-1)}_{(R(M^T))}$ is called the *transfer matrix* of the network.

Since the conductivity matrix A is nonsingular, the network equations (133) can be rewritten as

$$A^{-1}\mathbf{y} + \mathbf{x} = A^{-1}\mathbf{w} + \mathbf{v} \,, \quad \mathbf{y} \in N(M) \,,\ \mathbf{x} \in R(M^T) \,. \tag{136}$$

By Exs. 93 and 83, the unique solution of (136) is

$$\begin{aligned} \mathbf{y} &= (A^{-1})^{(-1)}_{(N(M))}(A^{-1}\mathbf{w} + \mathbf{v}) \,, && (137) \\ \mathbf{x} &= (I - A^{-1}(A^{-1})^{(-1)}_{(N(M))})(A^{-1}\mathbf{w} + \mathbf{v}A^{-1}\mathbf{w} + \mathbf{v}) \,. && (138) \end{aligned}$$

The matrix $(A^{-1})^{(-1)}_{(N(M))}$ is called the *dual transfer matrix*, its (i,j)th entry being the current in branch b_i as a result of inserting a one–volt generator parallel to branch b_j. Comparing the corresponding equations in (134)–(135 and in (137)–(138, we prove that the transfer matrices $A^{(-1)}_{(R(M^T))}$ and $(A^{-1})^{(-1)}_{(N(M))}$ satisfy

$$A^{-1}(A^{-1})^{(-1)}_{(N(M))} + A^{(-1)}_{(R(M^T))}A = I \,, \tag{139}$$

which can also be proved directly from Ex. 84(b).

The correspondence between results like (134)–(135 and (137)–(138 is called *electrical duality*; see, e.g., the discussion in Bott and Duffin [**202**], Duffin [**435**], and Sharpe and Styan [**1343**], [**1344**], [**1345**], for further results on duality and on applications of generalized inverses in electrical networks.

[2] *mho*, the unit of conductance, is the reciprocal of *ohm*, the unit of resistance.

Exercises.

Ex. 93. Let $A \in \mathbb{C}^{n\times n}$ be such that $\langle A\mathbf{x}, \mathbf{x}\rangle \neq 0$ for every nonzero vector $\mathbf{x}$ in L, a subspace of $\mathbb{C}^n$. Then $A^{(-1)}_{(L)}$ exists, i.e., $(AP_L + P_{L^\perp})$ is nonsingular.

Proof. If $A\mathbf{x}+\mathbf{y} = \mathbf{0}$ for some $\mathbf{x} \in L$ and $\mathbf{y} \in L^\perp$, then $A\mathbf{x} \in L^\perp$ and therefore $\langle A\mathbf{x}, \mathbf{x}\rangle = 0$. □

See also Exs. 85 and 78(b) above.

Ex. 94. In AC networks without mutual coupling, equations (132) still hold for the branches, by using complex, instead of real, constants and variables. The complex a_j is then the admittance of branch b_j. AC networks with mutual coupling due to transformers, are still represented by (133), where the *admittance matrix* A is symmetric, its off–diagonal elements giving the mutual couplings; see, e..g., Bott and Duffin [**202**].

Ex. 95. *Incidence matrix.* Let M be a connected $m \times n$ incidence matrix. Then for any $M^{(1,3)} \in M\{1,3\}$,

$$I - MM^{(1,3)} = \frac{1}{m}\,\mathbf{e}\mathbf{e}^T ,$$

where $\mathbf{e}\mathbf{e}^T$ is the $m \times m$ matrix whose elements are all 1. See also Ijiri [**766**].

Proof. From $(I - MM^{(1,3)})M = O$ it follows for any two directly connected nodes n_i and n_j (i.e., for any column of M having its $+1$ and -1 in rows i and j), that the ith and jth columns of $I - MM^{(1,3)}$ are identical. Since M is connected, all columns of $I - MM^{(1,3)}$ are identical. Since $I - MM^{(1,3)}$ is symmetric, all its rows are also identical. Therefore, all elements of $I - MM^{(1,3)}$ are equal, say

$$I - MM^{(1,3)} = \alpha\,\mathbf{e}\mathbf{e}^T ,$$

for some real α. Now $I - MM^{(1,3)}$ is idempotent, proving that $\alpha = 1/m$. □

Ex. 96. Let M be a connected $m \times n$ incidence matrix. Then $\operatorname{rank} M = m - 1$.

Proof.

$$\begin{aligned} P_{N(M^T)} &= I - P_{R(M)} , && \text{by (48)} ,\\ &= I - MM^{(1,3)} , && \text{by Ex. 1.9 and Lemma 3} ,\\ &= \frac{1}{m}\,\mathbf{e}\mathbf{e}^T , && \text{by Ex. 95} , \end{aligned}$$

proving that $\dim N(M^T) = \operatorname{rank} P_{N(M^T)} = 1$, and therefore

$$\operatorname{rank} M = \dim R(M) = m - \dim N(M^T) = m - 1 .$$

□

Ex. 97. *Trees.* Let a connected network consist of m nodes and n branches, and let M be its incidence matrix. A *tree* is defined as consisting of the m nodes, and any $m-1$ branches which correspond to linearly independent columns of M. Show that:

(a) A tree is a connected network which contains no closed circuit.

(b) Any column of M not among the $m-1$ columns corresponding to a given tree, can be expressed uniquely as a linear combination of those $m-1$ columns, using only the coefficients $0, +1$, and -1.

(c) Any branch not in a given tree, lies in a unique closed circuit whose other branches, or the branches obtained from them by reversing their directions, belong to the tree.

EX. 98. Let $A = [\operatorname{diag} a_j]$, $a_j \neq 0$, $j \in \overline{1,n}$, and let M be a connected $m \times n$ incidence matrix. Show that the discriminant (see Ex. 82)

$$d_{A,R(M^T)} = \det\left(AP_{R(M^T)} + P_{N(M)}\right)$$

is the sum, over all trees $\{b_{j_1}, b_{j_2}, \ldots, b_{j_{m-1}}\}$ in the network, of the products

$$a_{j_1} a_{j_2} \cdots a_{j_{m-1}} \quad \text{(Bott and Duffin [\textbf{202}])}\ .$$

Suggested further reading

Section 1. Bjerhammar [**174**], Hearon [**709**], Jones [**788**], Morris and Odell [**1096**], Sheffield [**1347**].
Section 4. Afriat [**6**], Chipman and Rao [**331**], Graybill and Marsaglia [**571**], Greville [**586**], Wedderburn [**1538**].
Section 5. Ward, Boullion and Lewis [**1536**].
Section 6. Afriat [**6**], Anderson and Duffin [**26**], Ben–Israel [**112**], Chipman and Rao [**331**], Glazman and Ljubich [**542**], Greville [**586**], Petryshyn [**1183**], Stewart [**1400**].
Section 9. Rao and Mitra [**1251**].
Section 10. For applications of generalized inverses in mathematical programming see also Beltrami [**103**], Ben–Israel ([**111**], [**115**], [**118**]), Ben–Israel and Kirby [**133**], Charnes and Cooper [**302**], Charnes, Cooper and Thompson [**303**], Charnes and Kirby [**305**], Kirby [**846**], Nelson, Lewis and Boullion [**1128**], Rosen ([**1303**], [**1304**]), Zlobec [**1651**].
Section 11. Bowman and Burdet [**216**], and Charnes and Granot [**304**].

CHAPTER 3

Minimal Properties of Generalized Inverses

1. Least–squares solutions of inconsistent linear systems

For given $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$, the linear system

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

is *consistent*, i.e., has a solution for $\mathbf{x}$, if and only if $\mathbf{b} \in R(A)$. Otherwise, the *residual* vector

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} \tag{2}$$

is nonzero for all $\mathbf{x} \in \mathbb{C}^n$, and it may desired to find an *approximate solution* of (1), by which is meant a vector $\mathbf{x}$ making the residual vector (2) "closest" to zero in some sense, i.e., minimizing some norm of (2). An approximate solution that is often used, especially in statistical applications, is the *least–squares solution* of (1), defined as a vector $\mathbf{x}$ minimizing the Euclidean norm of the residual vector, i.e., minimizing the sum of squares of moduli of the residuals

$$\sum_{i=1}^{m} |r_i|^2 = \sum_{i=1}^{m} \left| b_i - \sum_{j=1}^{n} a_{ij} x_j \right|^2 = \|\mathbf{b} - A\mathbf{x}\|^2 \,. \tag{3}$$

In this section the Euclidean vector norm – see, e.g., Ex. 0.8 – is denoted simply by $\|\ \|$.

The following theorem shows that $\|A\mathbf{x} - \mathbf{b}\|$ is minimized by choosing $\mathbf{x} = X\mathbf{b}$, where $X \in A\{1,3\}$, thus establishing a relation between the $\{1,3\}$–inverses and the least–squares solutions of $A\mathbf{x} = \mathbf{b}$, characterizing each of these two concepts in terms of the other.

Theorem 1. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$. Then $\|A\mathbf{x} - \mathbf{b}\|$ is smallest when $\mathbf{x} = A^{(1,3)}\mathbf{b}$, where $A^{(1,3)} \in A\{1,3\}$. Conversely, if $X \in \mathbb{C}^{n\times m}$ has the property that, for all $\mathbf{b}$, $\|A\mathbf{x} - \mathbf{b}\|$ is smallest when $\mathbf{x} = X\mathbf{b}$, then $X \in A\{1,3\}$.

Proof. From (2.47)

$$\begin{aligned} \mathbf{b} &= (P_{R(A)} + P_{R(A)^\perp})\mathbf{b} \,. && (4)\\ \therefore\ \mathbf{b} - A\mathbf{x} &= (P_{R(A)}\mathbf{b} - A\mathbf{x}) + P_{N(A^*)}\mathbf{b} \,. \\ \therefore\ \|A\mathbf{x} - \mathbf{b}\|^2 &= \|A\mathbf{x} - P_{R(A)}\mathbf{b}\|^2 + \|P_{N(A^*)}\mathbf{b}\|^2 \,, \text{ by Ex. 0.36 .} && (5) \end{aligned}$$

Evidently, (5) assumes its minimum value if and only if

$$A\mathbf{x} = P_{R(A)}\mathbf{b} \,, \tag{6}$$

which holds if $\mathbf{x} = A^{(1,3)}\mathbf{b}$ for any $A^{(1,3)} \in A\{1,3\}$, since by Theorem 2.8, (2.28), and Lemma 2.3

$$AA^{(1,3)} = P_{R(A)} \,. \tag{7}$$

Conversely, if X is such that for all $\mathbf{b}$, $\|A\mathbf{x} - \mathbf{b}\|$ is smallest when $\mathbf{x} = X\mathbf{b}$, (6) gives $AX\mathbf{b} = P_{R(A)}\mathbf{b}$ for all $\mathbf{b}$, and therefore

$$AX = P_{R(A)} \,.$$

Thus, by Theorem 2.3, $X \in A\{1,3\}$. ∎

COROLLARY 1. A vector $\mathbf{x}$ is a least–squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if

$$A\mathbf{x} = P_{R(A)}\mathbf{b} = AA^{(1,3)}\mathbf{b} \, .$$

Thus, the general least–squares solution is

$$\mathbf{x} = A^{(1,3)}\mathbf{b} + (I_n - A^{(1,3)}A)\mathbf{y} \, , \tag{8}$$

with $A^{(1,3)} \in A\{1,3\}$ and arbitrary $\mathbf{y} \in \mathbb{C}^n$.

It will be noted that the least–squares solution is unique only when A is of full column rank (the most frequent case in statistical applications). Otherwise, (8) is an infinite set of such solutions.

Exercises, examples and supplementary notes.

EX. 1. *Normal equation.* Show that a vector $\mathbf{x}$ is a least–squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x}$ is a solution of

$$A^*A\mathbf{x} = A^*\mathbf{b} \, , \tag{9}$$

often called the *normal equation* of $A\mathbf{x} = \mathbf{b}$.

SOLUTION. It follows from (4) and (6) that $\mathbf{x}$ is a least–squares solution if and only if

$$A\mathbf{x} - \mathbf{b} \in N(A^*) \, ,$$

which is (9).

ALTERNATIVE SOLUTION. A necessary condition for the vector $\mathbf{x}^0$ to be a least–squares solution of $A\mathbf{x} = \mathbf{b}$ is that the partial derivatives $\partial f / \partial x_j$ of the function

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right)^* \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right) \tag{10}$$

vanish at $\mathbf{x}^0$, i.e., that $\nabla f(\mathbf{x}^0) = \mathbf{0}$, where

$$\nabla f(\mathbf{x}^0) = \left(\frac{\partial f}{\partial x_j}(\mathbf{x}^0) \right) \, ,$$

is the *gradient* of f at $\mathbf{x}^0$. Now it can be shown that the gradient of (8) at $\mathbf{x}^0$ is

$$\nabla f(\mathbf{x}^0) = 2A^*(A\mathbf{x} - \mathbf{b}) \, ,$$

proving the necessity of (9). The sufficiency follows from the identity

$$\begin{aligned}(A\mathbf{x} - \mathbf{b})^*(A\mathbf{x} - \mathbf{b}) &- (A\mathbf{x}^0 - \mathbf{b})^*(A\mathbf{x}^0 - \mathbf{b}) \\ &= (\mathbf{x} - x^0)^*A^*A(\mathbf{x} - x^0) + 2\Re\{(\mathbf{x} - x^0)^*A^*(\mathbf{x} - x^0)\} \, ,\end{aligned}$$

which holds for all $\mathbf{x}, \mathbf{x}^0 \in \mathbb{C}^n$.

EX. 2. For any $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$, the normal equation (9) is consistent.

EX. 3. *Ill–conditioning.* The linear equation $A\mathbf{x} = \mathbf{b}$, and the matrix A are said to be *ill–conditioned* (or badly conditioned) if the solutions are very sensitive to small changes in the data;, see, e.g. [**1145**, Chapter 8] and [**1596**]. The use of the normal equations (9) in finding least–squares solutions is limited by the fact that the matrix A^*A is ill–conditioned and very sensitive to roundoff errors, see, e.g., Taussky [**1434**] and Ex. 6.7. Methods for computing least–squares solutions which take account of this difficulty have been studied by several authors. We mention in particular Björck ([**178**], [**177**] and [**179**]), Björck and Golub [**182**], Businger and Golub [**246**] and [**247**], Golub and Wilkinson [**559**], and Noble [**1145**]. Three such methods are mentioned in Exs. 6, 10 and 11 below. These methods can be used, with slight modifications, to compute the generalized inverse. The reader who is not interested in numerical methods may skip Exs. 4 through 11.

EX. 4. The following example illustrates the ill–conditioning of the normal equation. Let

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \text{ and let the elements of } A^T A = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix}$$

be computed using double–precision and then rounded to single–precision with t binary digits. If $|\epsilon| < \sqrt{2^{-t}}$ then the rounded $A^T A$ is

$$fl(A^T A) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (\textit{fl} \text{ denotes } \textit{floaating point})$$

which is of rank 1, whereas A is of rank 2. Thus for any $\mathbf{b} \in \mathbb{R}^3$, the computed normal equation

$$fl(A^T A)\mathbf{x} = fl(A^T \mathbf{b})$$

may be inconsistent, or may have solutions which are not least–squares solutions of $A\mathbf{x} = \mathbf{b}$.

EX. 5. *Factorization methods.* Let A be factorized as

$$A = FG \tag{11}$$

where G is of full row–rank. Show that the normal equation (9) is equivalent to

$$F^* A\mathbf{x} = F^* \mathbf{b}\ . \tag{12}$$

The factorization (11) is useful if the system (12) is not ill–conditioned, or at leats not worse–conditioned than the system (1). Two such factorizations are given in Exs. 6 and 10 below.

EX. 6. *QR factorization.* Let $A \in \mathbb{C}_n^{m\times n}$ (where full column–rank is assumed form convenience; the modifications required for the general case are the subject of Ex. 9). Then the *QR factorization* of A is

$$A = QR = \widetilde{Q}\widetilde{R} \tag{13}$$

where $Q \in \mathbb{C}^{m\times n}$ is unitary (i.e. $Q^*Q = I$), $R = \begin{bmatrix} \widetilde{R} \\ O \end{bmatrix}$ where $\widetilde{R}$ is an $n \times n$ upper triangular matrix, and $\widetilde{Q}$ consists of the first n columns of Q. The columns of the unitary matrix Q form an orthonormal basis for $\mathbb{C}^m$, and it is clear from (13) that the columns of $\widetilde{Q}$ (aand the upper triangular matrix $\widetilde{R}$) maay be obtained by orthogonalizing the columns of A. (It also follows from (13) that each column of $\widetilde{Q}$ and each row of $\widetilde{R}$ is determined uniquely up to a scalar factor of modulus one.)

The two principal ways of computing the QR factorization are:

(1) Using a Gram–Schmidt type of orthogonalization; see, e.g., Rice [**1270**] and Björck [**178**] where a detailed error analysis is given for least–squares solutions.

(2) Using Householder transformations; see, e.g., Wilkinson [**1595**], Parlett [**1161**], and Golub [**550**].

These two ways are compared in [**550**].

Given the QR–factorization (13), it follows from Ex. 5 that the normal equation (9) is equivalent to

$$\widetilde{Q}^* A\mathbf{x} = \widetilde{Q}^* \mathbf{b}$$

or to

$$\widetilde{R}\mathbf{x} = \widetilde{Q}^* \mathbf{b}\ , \quad \text{since } \widetilde{Q}^*\widetilde{Q} = I_n\ . \tag{14}$$

Now $\widetilde{R}$ is upper triangular, and thus (14) is solved by backward substitution.

Ex. 7. Using the notation of Ex. 6, let

$$Q^*\mathbf{b} = \mathbf{c} = [c_i]\ ,\ \ i \in \overline{1,m}\ .$$

Show that the minimum value of $\|A\mathbf{x} - \mathbf{b}\|^2$ is $\sum_{i=n+1}^{m} |c_i|^2$.
(Hint: $\|A\mathbf{x} - \mathbf{b}\|^2 = \|Q^*(A\mathbf{x} - \mathbf{b})\|^2$ since Q is unitary.)

Ex. 8. Show that the $\widetilde{Q}\widetilde{R}$–factorization for the matrix of Ex. 4, is

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \approx \mathit{fl}(\widetilde{Q})\mathit{fl}(\widetilde{R}) = \begin{bmatrix} 1 & \frac{\epsilon}{\sqrt{2}} \\ \epsilon & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \epsilon\sqrt{2} \end{bmatrix}\ .$$

Use this to compute the least–squares solution of

$$\begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \epsilon \\ 2\epsilon \end{bmatrix}\ .$$

Answer. The (rounded) least–squares solution obtained by using (14) with the rounded matrices $\mathit{fl}(\widetilde{Q})$ and $\mathit{fl}(\widetilde{R})$ is

$$x_1 = 0,\ \ x_2 = 1\ .$$

The exact least–squares solution is

$$x_1 = \frac{\epsilon^2}{2 + \epsilon^2},\ \ x_2 = \frac{2(1 + \epsilon^2)}{2 + \epsilon^2}\ .$$

Ex. 9. Modify the results of Exs. 6 and 7 for the case $A \in \mathbb{C}_r^{m\times n}$, $r < n$.

Ex. 10. *Noble's method.* Let again $A \in \mathbb{C}_n^{m\times n}$ and assume that A is partitioned as

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad \text{where } A_1 \in \mathbb{C}_n^{n\times n}.$$

Then A may be factorized as

$$A = \begin{bmatrix} I \\ S \end{bmatrix} A_1 \quad \text{where } S = A_2 A_1^{-1} \in \mathbb{C}^{(m-n)\times n}\ . \tag{15}$$

Let now $\mathbf{b} \in \mathbb{C}^m$ be partitioned as $\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$, $\mathbf{b}_1 \in \mathbb{C}^n$. Then by Ex. 5, the normal equation reduces to

$$(I + S^*S)A_1\mathbf{x} = \mathbf{b}_1 + S^*\mathbf{b}_2 \tag{16}$$

(which reduces further to $A_1\mathbf{x} = \mathbf{b}_1$ if and only if $A\mathbf{x} = \mathbf{b}$ is consistent).

The matrix S can be obtained by applying Gauss–Jordan elimination to the matrix

$$\begin{bmatrix} A_1 & \mathbf{b}_1 & I \\ A_2 & \mathbf{b}_2 & O \end{bmatrix}$$

transforming it into

$$\begin{bmatrix} I & A_1^{-1}\mathbf{b}_1 & A_1^{-1} \\ O & \mathbf{b}_2 - S\mathbf{b}_1 & -S \end{bmatrix}$$

from which S can be read. (See Noble [**1145**, pp. 262–265].)

Ex. 11. *Iterative refinement of solutions.* Let $\mathbf{x}^{(0)}$ be an approximate solution of the consistent equation $A\mathbf{x} = \mathbf{b}$, and let $\widehat{\mathbf{x}}$ be an exact solution. Then the error $\delta\mathbf{x} = \widehat{\mathbf{x}} - \mathbf{x}^{(0)}$ satisfies

$$\begin{aligned} A\delta\mathbf{x} &= A\widehat{\mathbf{x}} - A\mathbf{x}^{(0)} \\ &= \mathbf{b} - A\mathbf{x}^{(0)} \\ &= \mathbf{r}^{(0)}, \text{ the } \textit{residual} \text{ corresponding to } \mathbf{x}^{(0)}\ . \end{aligned}$$

This suggests the following *iterative refinement of solutions*, due to Wilkinson [**1595**] (see also Moler [**1079**]):

The initial approximation: $\mathbf{x}^{(0)}$, given.
The kth residual: $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$.
The kth correction, $\delta\mathbf{x}^{(k)}$, is obtaiend by solving $A\delta\mathbf{x}^{(k)} = \mathbf{r}^{(k)}$.
The $(k+1)$st approximation: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}$.
Double precision is used in computing the residuals, but not elsewhere,
The iteration is stopped if $\|\delta\mathbf{x}^{(k+1)}\|/\|\delta\mathbf{x}^{(k)}\|$ falls below a prescribed number.
If the sequence $\{\mathbf{x}^{(k)} : k = 0, 1, \dots\}$ converges, it converges to a solution of $A\mathbf{x} = \mathbf{b}$.

The use of this method to solve linear equations which are equivalent to the normal equation, such as (14) or (16), has been successful in finding, or improving, least–squares solutions. The reader is referred to Golub and Wilkinson [**559**], Björck [**177**] and [**179**], and Björck and Golub [**182**].

Ex. 12. Show that the vector $\mathbf{x}$ is a least–squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if there is a vector $\mathbf{r}$ such that the vector $\begin{bmatrix}\mathbf{r}\\ \mathbf{x}\end{bmatrix}$ is a solution of

$$\begin{bmatrix} I & A \\ A^* & O \end{bmatrix}\begin{bmatrix}\mathbf{r}\\ \mathbf{x}\end{bmatrix} = \begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix} \tag{17}$$

Ex. 13. Let $A \in \mathbb{C}^{m\times n}$ and let $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k \in \mathbb{C}^m$. Show that a vector $\mathbf{x}$ minimizes

$$\sum_{i=1}^{k} \|A\mathbf{x} - \mathbf{b}_i\|^2$$

if and only if $\mathbf{x}$ is a least–squares solution of

$$A\mathbf{x} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{b}_i\ .$$

Ex. 14. Let $A_i \in \mathbb{C}^{m\times n}$, $\mathbf{b}_i \in \mathbb{C}^m$ $(i = 1, \dots, k)$. Show that a vector $\mathbf{x}$ minimizes

$$\sum_{i=1}^{k} \|A_i\mathbf{x} - \mathbf{b}_i\|^2 \tag{18}$$

if and only if $\mathbf{x}$ is a solution of

$$\left(\sum_{i=1}^{k} A_i^* A_i\right)\mathbf{x} = \sum_{i=1}^{k} A_i^*\mathbf{b}_i\ . \tag{19}$$

Solution. $\mathbf{x}$ minimizes (18) if and only if $\mathbf{x}$ is a least–squares solution of the system

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix}\mathbf{x} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_k \end{bmatrix}\ ,$$

whose normal equation is (19).

Ex. 15. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let α^2 be a positive real number. Show that the function

$$\|A\mathbf{x} - \mathbf{b}\|^2 + \alpha^2\|\mathbf{x}\|^2 \tag{20}$$

has a unique minimizer $\mathbf{x}_{\alpha^2}$ given by

$$\mathbf{x}_{\alpha^2} = (A^*A + \alpha^2 I)^{-1}A^*\mathbf{b} \tag{21}$$

whose norm $\|\mathbf{x}_{\alpha^2}\|$ is a monotone decreasing function of α^2.

Solution. (20) is a special case of (18) with $k = 2, A_1 = A, A_2 = \alpha I, \mathbf{b}_1 = \mathbf{b}$ and $\mathbf{b}_2 = \mathbf{0}$. Substituting these values in (19) we get

$$(A^*A + \alpha^2 I)^{-1}A^*\mathbf{x} = A^*\mathbf{b}\ ,$$

which has the unique solution (21), since $(A^*A + \alpha^2 I)$ is nonsingular.

Using (2.47) or Lemma 1 below, it is possible to write $\mathbf{b}$ (uniquely) as

$$\mathbf{b} = A\mathbf{u} + \mathbf{v}\ , \quad \mathbf{u} \in R(A^*)\ , \quad \mathbf{v} \in N(A^*)\ . \tag{22}$$

Substituting this in (21) gives

$$\mathbf{x}_{\alpha^2} = (A^*A + \alpha^2 I)^{-1}A^*A\mathbf{u}\ . \tag{23}$$

Now let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be an orthonormal basis of $R(A^*)$ consisting of eigenvectors of A^*A corresponding to nonzero eigenvalues, say

$$A^*A\mathbf{u}_j = \lambda_j\mathbf{u}_j\ , \quad \lambda_j > 0 \quad j \in \overline{1, r}\ .$$

If $\mathbf{u} = \sum_{j=1}^r \beta_j\mathbf{u}_j$ is the representation of $\mathbf{u}$ in terms of the above basis, then (23) gives

$$\mathbf{x}_{\alpha^2} = \sum_{j=1}^r \frac{\lambda_j\beta_j}{\lambda_j + \alpha^2}\mathbf{u}_j$$

whose norm squared is

$$\|\mathbf{x}_{\alpha^2}\|^2 = \sum_{j=1}^r \left(\frac{\lambda_j}{\lambda_j + \alpha^2}\right)^2 |\beta_j|^2\ ,$$

a monotone decreasing function of α^2. □

Problems of minimizing expressions like (20) in infinite–dimensional spaces and subject to linear constraints arise often in control theory. The reader is referred to [**1197**], especially to Section 4.4 and pp. 353–354 where additional references are given.

Ex. 16. *Constrained least–squares solutions.* A vector $\mathbf{x}$ is said to be a *constrained least–square solution* if $\mathbf{x}$ is a solution of the constrained minimization problem: Minimize $\|A\mathbf{x} - \mathbf{b}\|$ subject to the given constraints. Let $A_1 \in \mathbb{C}^{m_1\times n}$, $\mathbf{b}_1 \in \mathbb{C}^{m_1}$, $A_2 \in \mathbb{C}^{m_2\times n}$, $\mathbf{b}_2 \in \mathbb{C}^{m_2}$. Characterize the solutions of the problem:

$$\text{Minimize } \|A_1\mathbf{x} - \mathbf{b}_1\|^2 \text{ subject to } A_2\mathbf{x} = \mathbf{b}_2\ . \tag{24}$$

Solution. The general solution of $A_2\mathbf{x} = \mathbf{b}_2$ is

$$\mathbf{x} = A_2^{(1)}\mathbf{b}_2 + (I - A_2^{(1)}A_2)\mathbf{y}\ , \tag{25}$$

where $A_2^{(1)} \in A_2\{1\}$ and $\mathbf{y}$ ranges over $\mathbb{C}^n$. Substituting (25) in $A_1\mathbf{x} = \mathbf{b}_1$ gives the equation

$$A_1(I - A_2^{(1)}A_2)\mathbf{y} = \mathbf{b}_1 - A_1A_2^{(1)}\mathbf{b}_2\ . \tag{26}$$

Therefore $\mathbf{x}$ is a solution of (24) if and only if $\mathbf{x}$ is given by (25) where $\mathbf{y}$ is a least–squares solution of (26). □

Ex. 17. Show that a vector $\mathbf{x} \in \mathbb{C}^n$ is a solution of (24) if and only if there is a vector $\mathbf{y} \in \mathbb{C}^{m_2}$ such that the vector $\begin{bmatrix} \mathbf{x} \\ \mathbf{x} \end{bmatrix}$ is a solution of

$$\begin{bmatrix} A_1^* A_1 & A_2^* \\ A_2 & O \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} A_1^* \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} . \tag{27}$$

Compare this with Ex. 1. Similarly, find a characterization analogous to that given in Ex. 12. See also Björck and Golub [**182**].

Ex. 18. Let $A \in \mathbb{C}_r^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let $0 < p < \sqrt{r}\|\mathbf{u}\|$ where $\mathbf{u}$ is given by (22). Show that the problem

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\| \text{ subject to } \|\mathbf{x}\| = p \tag{28}$$

has the unique solution

$$\mathbf{x} = (A^*A + \alpha^2 I)^{-1} A^* \mathbf{b}$$

where α is (uniquely) determined by

$$\|(A^*A + \alpha^2 I)^{-1} A^* \mathbf{b}\| = p\ .$$

Hint. Use Ex. 15.
See also Forsythe and Golub [**505**, Section 7], and Forsythe [**504**].

2. Solutions of minimum norm

When the system (1) has a multiplicity of solutions for $\mathbf{x}$, there is a unique solution of minimum norm. This follows from Ex. 2.72, restated here as,

LEMMA 1. Let $A \in \mathbb{C}^{m\times n}$. Then A is a one–to–one mapping of $R(A^*)$ onto $R(A)$. □

COROLLARY 2. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in R(A)$. Then there is a unique solution of

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

given as the unique solution of (1) which lies in $R(A^*)$.

PROOF. By Lemma 1, Eq. (1) has a unique solution $\mathbf{x}_0$ in $R(A^*)$. Now the general solution is given as

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{y}, \quad \mathbf{y} \in N(A)\ ,$$

and by Ex. 2.36

$$\|\mathbf{x}\|^2 = \|\mathbf{x}_0\|^2 + \|\mathbf{y}\|^2$$

proving that $\|\mathbf{x}\| > \|\mathbf{x}_0\|$ unless $\mathbf{x} = \mathbf{x}_0$. □

The following theorem relates minimum–norm solutions of $A\mathbf{x} = \mathbf{b}$ and $\{1,4\}$–inverses of A, characterizing each of these two concepts in terms of the other.

THEOREM 2. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$. If $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$, the unique solution for which $\|\mathbf{x}\|$ is smallest is given by

$$\mathbf{x} = A^{(1,4)}\mathbf{b}\ ,$$

where $A^{(1,4)} \in A\{1,4\}$. Converesely, if $X \in \mathbb{C}^{n\times m}$ is such that, whenever $A\mathbf{x} = \mathbf{b}$ has a solution, $\mathbf{x} = X\mathbf{b}$ is the solution of minimum norm, then $X \in A\{1,4\}$.

PROOF. If $A\mathbf{x} = \mathbf{b}$ is consistent, then for any $A^{(1,4)} \in A\{1,4\}$, $\mathbf{x} = A^{(1,4)}\mathbf{b}$ is a solution (by Corollary 2.2), lies in $R(A^*)$ (by Ex. 1.9) and thus, by Lemma 1, is the unique solution in $R(A^*)$, and thus the unique minimum–norm solution by Corollary 2.

Conversely, let X be such that, for all $\mathbf{b} \in R(A)$, $\mathbf{x} = X\mathbf{b}$ is the solution of $A\mathbf{x} = \mathbf{b}$ of minimum norm. Setting $\mathbf{b}$ equal to each column of A, in turn, we conclude that

$$XA = A^{(1,4)}A$$

and $X \in A\{1,4\}$ by Theorem 2.4. □

The unique minimum–norm least–squares solution of $A\mathbf{x} = \mathbf{b}$, and the generalized inverse $A^\dagger$ of A, are related as follows.

COROLLARY 3. (Penrose [**1178**]). Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$. Then, among the least–squares solutions of $A\mathbf{x} = \mathbf{b}$, $A^\dagger\mathbf{b}$ is the one of minimum–norm. Conversely, if $X \in \mathbb{C}^{n\times m}$ has the property that, for all $\mathbf{b}$, $X\mathbf{b}$ is the minimum–norm least–squares solution of $A\mathbf{x} = \mathbf{b}$, then $X = A^\dagger$.

PROOF. By Corollary 1, the least–squares solutions of $A\mathbf{x} = \mathbf{b}$ coincide with the solutions of

$$A\mathbf{x} = AA^{(1,3)}\mathbf{b}\ . \tag{6}$$

Thus the minimum–norm least–squares solution of $A\mathbf{x} = \mathbf{b}$ is the minimum–norm solution of (6). But by Theorem 2, the latter is

$$\begin{aligned}\mathbf{x} &= A^{(1,4)}AA^{(1,3)}\mathbf{b}\\ &= A^\dagger\mathbf{b}\end{aligned}$$

by Theorem 1.4.

A matrix X having the properties stated in the last sentence of the theorem must satisfy $X\mathbf{b} = A^\dagger\mathbf{b}$ for all $\mathbf{b} \in \mathbb{C}^m$, and therefore $X = A^\dagger$. □

The *minimum–norm least–squares solution*, $\mathbf{x}_0 = A^\dagger\mathbf{b}$ (also called the *approximate solution*; e.g., Penrose [**1178**]) of $A\mathbf{x} = \mathbf{b}$, can thus be characterized by the following two inequalities:

$$\|A\mathbf{x}_0 - \mathbf{b}\| \le \|A\mathbf{x} - \mathbf{b}\| \quad \text{for all } \mathbf{x} \tag{29}$$

and

$$\|\mathbf{x}_0\| < \|\mathbf{x}\| \tag{30}$$

for any $\mathbf{x} \neq \mathbf{x}_0$ which gives equality in (29).

Exercises, examples and supplementary notes.

EX. 19. Let A be given as in Ex. 0.46 and let

$$\mathbf{b} = \begin{bmatrix} -i \\ 1 \\ 1 \end{bmatrix}\ .$$

Show that the general least–squares solution of $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = \frac{1}{19}\begin{bmatrix} 0 \\ 1 \\ 0 \\ -4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & -1+2i & \frac{1}{2}i \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1-i \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}$$

where $y_1, y_2, \dots, y_6$ are arbitrary, while the residual vector for the least–squares solution is

$$\frac{1}{19}\begin{bmatrix} 2i \\ 12 \\ -2 \end{bmatrix}$$

Ex. 20. In Ex. 19 show that the minimum–norm least–squares solution is

$$\mathbf{x} = \frac{1}{874}\begin{bmatrix} 0 \\ 26-36i \\ 13-18i \\ -55-9i \\ -12-2i \\ -46+59i \end{bmatrix}$$

Ex. 21. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and $\mathbf{a} \in \mathbb{C}^n$. Show that if $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$, then the unique solution for which $\|\mathbf{x} - \mathbf{a}\|$ is smallest is given by

$$\begin{aligned} \mathbf{x} &= A^{(1,4)}\mathbf{b} + (I - A^{(1,4)}A)\mathbf{a} \\ &= A^{(1,4)}\mathbf{b} + P_{N(A)}\mathbf{a} \, . \end{aligned}$$

Ex. 22. (den Broeder and Charnes [**238**], Ben–Israel [**117**]). Show that for any $A \in \mathbb{C}^{m\times n}$, as $\lambda \to 0$ through any neighborhood of 0 in $\mathbb{C}$, the following limit exists and

$$\lim_{\lambda\to 0}(A^*A + \lambda I)^{-1}A^* = A^\dagger \tag{31}$$

Solution. We must show that

$$\lim_{\lambda\to 0}(A^*A + \lambda I)^{-1}A^*\mathbf{y} = A^\dagger\mathbf{y} \tag{32}$$

for all $\mathbf{y} \in \mathbb{C}^m$. Since $N(A^*) = N(A^\dagger)$, by Ex. 2.29, (32) holds trivially for $\mathbf{y} \in N(A^*)$. Therefore it suffices to prove (32) for $\mathbf{y} \in N(A^*)^\perp = R(A)$. By Lemma 1, for any $\mathbf{y} \in R(A)$ there is a unique $\mathbf{x} \in R(A^*)$ such that $\mathbf{y} = A\mathbf{x}$. Proving (32) thus amounts to proving for all $\mathbf{x} \in R(A^*)$

$$\begin{aligned} \lim_{\lambda\to 0}(A^*A + \lambda I)^{-1}A^*A\mathbf{x} &= A^\dagger A\mathbf{x} \qquad (33) \\ &= \mathbf{x} \, , \quad \text{since } A^\dagger A = P_{R(A^*)} \, . \end{aligned}$$

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ be a basis for $R(A^*)$ consisting of eigenvectors of A^*A, say

$$A^*A\mathbf{u}_j = \lambda_j\mathbf{u}_j \quad (\lambda_j > 0, \, j \in \overline{1,r}) \, .$$

Writing $\mathbf{x} \in R(A^*)$ in terms of this basis

$$\mathbf{x} = \sum_{j=1}^{r} \xi_j\mathbf{u}_j \, ,$$

we verify that for all $\lambda \neq -\lambda_1, -\lambda_2, \dots, -\lambda_r$

$$(A^*A + \lambda I)^{-1}A^*A\mathbf{x} = \sum_{j=1}^{r} \frac{\lambda_j\xi_j}{\lambda_j + \lambda}\mathbf{u}_j \, ,$$

which tends, as $\lambda \to 0$, to $\sum_{j=1}^{r} \xi_j\mathbf{u}_j = \mathbf{x}$.

Alternative solution. Following the last solution up to (33), it suffices to show that

$$\lim_{\lambda\to 0}(A^*A + \lambda I_n)^{-1}A^*A = A^\dagger A = P_{R(A^*)} \, .$$

Now let $A^*A = FF^*$, $F \in \mathbb{C}_r^{n\times r}$ be a full–rank factorization. Then

$$(A^*A + \lambda I_n)^{-1}A^*A = (FF^* + \lambda I_n)^{-1}FF^*$$

for any λ for which the inverses exist. We now use the identity

$$(FF^* + \lambda I_n)^{-1}FF^* = F(F^*F + \lambda I_r)^{-1}F^*$$

and note that F^*F is nonsingular so that $\lim_{\lambda\to 0}(F^*F + \lambda I_r)^{-1} = (F^*F)^{-1}$. Collecting these facts we conclude that

$$\begin{aligned}\lim_{\lambda\to 0}(A^*A + \lambda I_n)^{-1}A^*A &= F(F^*F)^{-1}F^* \\ &= FF^\dagger \\ &= P_{R(A^*)}\end{aligned}$$

since the columns of F are a basis for $R(A^*A) = R(A^*)$.

Still another proof is given in Ex. 4.41.

Ex. 23. Use Exs. 15 and 22 to conclude that the solutions $\{\mathbf{x}_{\alpha^2}\}$ of the minimization problems:

$$\text{minimize}\,\{\|A\mathbf{x} - \mathbf{b}\|^2 + \alpha^2\|\mathbf{x}\|^2\}$$

converge to $A^\dagger\mathbf{b}$ as $\alpha \to 0$. Explain this result in view of Corollary 3.

Ex. 24. For a given $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$ and a positive real number p, solve the problem

$$\text{minimize}\,\|A\mathbf{x} - \mathbf{b}\| \text{ subject to } \|\mathbf{x}\| \le p\,. \tag{34}$$

Solution. If

$$\|A^\dagger\mathbf{b}\| \le p \tag{35}$$

then $\mathbf{x} = A^\dagger\mathbf{b}$ is a solution of (34), and is the unique solution if and only if (35) is an equality.

If (35) does not hold, then (34) has the unique solution given in Ex. 18.
(See also Balakrishnan [**62**, Theorem 2.3].)

Ex. 25. *Matrix spaces.* For any $A, B \in \mathbb{C}^{m\times n}$ define

$$R(A,B) = \{Y = AXB \in \mathbb{C}^{m\times n} : X \in \mathbb{C}^{n\times m}\} \tag{36}$$

and

$$N(A,B) = \{X \in \mathbb{C}^{n\times m} : AXB = O\} \tag{37}$$

which we shall call the *range* and *null space* of (A,B), respectively. Let $\mathbb{C}^{m\times n}$ be endowed with the inner product

$$\langle X, Y\rangle = \text{trace}\,Y^*X = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}\,\overline{y_{ij}}\,, \tag{38}$$

for $X = [x_{ij}]$, $Y = [y_{ij}] \in \mathbb{C}^{m\times n}$. Then for every $A, B \in \mathbb{C}^{m\times n}$ the sets $R(A,B)$ and $N(A^*, B^*)$ are complementary orthogonal subspaces of $\mathbb{C}^{m\times n}$.

Solution. As in Ex. 2.2 we use the one–to–one correspondence

$$v_{n(i-1)+j} = x_{ij} \quad (i \in \overline{1,m},\ j \in \overline{1,n}) \tag{39}$$

between the matrices $X = [x_{ij}] \in \mathbb{C}^{m\times n}$ and the vectors $\mathbf{v} = \text{vec}(X) = [v_k] \in \mathbb{C}^{mn}$. The correspondence (39) is a nonsingular linear transformation mapping $\mathbb{C}^{m\times n}$ onto C^{mn}. Linear subspaces of $\mathbb{C}^{m\times n}$ and C^{mn} thus correspond under (39).

It follows from (39) that the inner product (38) is equal to the standard inner product of the corresponding vectors vec(X) and vec(Y). Thus $\langle X, Y\rangle = \langle \text{vec}(X), \text{vec}(Y)\rangle = \text{vec}(Y)^*\text{vec}(X)$. Also, from (2.10) we deduce that under (39), $R(A,B)$ and $N(A^*, B^*)$ correspond to $R(A\otimes B^T)$ and $N(A^*\otimes B^{*T})$, respectively. By (2.8), the latter is the same as $N((A\otimes B^T)^*)$, which by (2.48) is the orthogonal complement of $R(A\otimes B^T)$ in $\mathbb{C}^{mn}$. Therefore, $R(A,B)$ and $N(A^*, B^*)$ are orthogonal complements in $\mathbb{C}^{m\times n}$.

Ex. 26. *Characterization of* $\{1,3\}$*–,* $\{1,4\}$*–, and* $\}1,2,3,4\|$*–inverses.* Let the norm used in $\mathbb{C}^{m\times n}$ be

$$\|X\| = \sqrt{\mathrm{trace} X^*X} \; ; \tag{40}$$

see, e.g., (0.24), which is the Euclidean norm of the vector $\mathrm{vec}(X)$. Show that for every $A \in \mathbb{C}^{m\times n}$:

(a) $X \in A\{1,3\}$ if and only if X is a least–squares solution of

$$AX = I_m \; , \tag{41}$$

i.e., minimizing $\|AX - I\|$ in the norm (40).

(b) $X \in A\{1,4\}$ if and only if X is a least–squares solution of

$$XA = I_n \; . \tag{42}$$

(c) $A^\dagger$ is the minimum–norm least–squares solution of both (41) and (42).

Solution. These results are based on the fact that the norm $\|X\|$ defined by (40) is merely the Euclidean norm of the corresponding vector $\mathrm{vec}(X)$.

(a) Writing the equation (41) as

$$(A \otimes I)\mathrm{vec}(X) = \mathrm{vec}(I) \; , \tag{43}$$

it follows from Corollary 1 that the general least–squares solution of (43) is

$$\mathrm{vec}(X) = (A \otimes I)^{(1,3)}\mathrm{vec}(I) + (I - (A \otimes I)^{(1,3)}(A \otimes I))\mathbf{y} \; , \tag{44}$$

where $\mathbf{y}$ is an arbitrary element of $\mathbb{C}^{mn}$. From (2.8) and (2.9) it follows that for every $A^{(1,3)} \in A\{1,3\}$, $(A^{(1,3)} \otimes I)$ is a $\{1,3\}$–inverse of $(A \otimes I)$. Therefore the general least–squares solution of (41) is the matrix corresponding to (44), namely

$$X = A^{(1,3)} + (I - A^{(1,3)}A)Y \; , \quad Y \in \mathbb{C}^{n\times m} \; ,$$

which is the general $\{1,3\}$–inverse of A by Corollary 2.3.

(b) Taking the conjugate transpose of (42), we get

$$A^*X^* = I_n \; .$$

The set of least–squares solutions of the last equation is by (a)

$$A^*\{1,3\} \; ,$$

which coincides with $A\{1,4\}$.

(c) This is left to the reader.

Ex. 27. Let A, B, D be complex matrices having dimensions consistent with the matrix equation

$$AXB = D \; .$$

Show that the minimum–norm least–squares solution of the last equation is

$$X = A^\dagger D B^\dagger \quad \text{(Penrose [\textbf{1178}])} \; .$$

Ex. 28. Let $A \in \mathbb{C}^{m\times n}$ and let X be a $\{1\}$–inverse of A; i.e., let X satisfy

$$AXA = A \; . \tag{1.1}$$

Then the following are equivalent:

(a) $X = A^\dagger$,

(b) $X \in R(A^*, A^*)$,

(c) X is the minimum–norm solution of (1.1). (Ben-Israel [**114**]).

PROOF. The general solution of (1.1) is by Theorem 2.1

$$\begin{aligned} X &= A^\dagger A A^\dagger + Y - A^\dagger A Y A A^\dagger \ , \quad Y \in \mathbb{C}^{n\times m} \\ &= A^\dagger + Y - A^\dagger A Y A A^\dagger \ . \end{aligned} \tag{45}$$

Now it is easy to verify that

$$A^\dagger \in R(A^*, A^*) \ , \quad Y - A^\dagger A Y A A^\dagger \in N(A, A) \ ,$$

and using the norm (40) it follows from Ex. 25 that X of (45) satisfies

$$\|X\|^2 = \|A^\dagger\|^2 + \|Y - A^\dagger A Y A A^\dagger\|^2 \ ,$$

and the equivalence of (a), (b) and (c) is obvious. □

EX. 29. *Restricted generalized inverses.* Let the matrix $A \in \mathbb{C}^{m\times n}$ and the subspace $S \subset \mathbb{C}^n$ be given. Then for any $\mathbf{b} \in \mathbb{C}^m$, the point $X\mathbf{b} \in S$ minimizes $\|A\mathbf{x} - \mathbf{b}\|$ in S if and only if $X = P_S(AP_S)^{(1,3)}$ is any S–restricted $\{1.3\}$–inverse of A.

PROOF. Follows from Section 2.8 and Theorem 1. □

EX. 30. Let A, S be as in Ex. 29. Then for any $\mathbf{b} \in \mathbb{C}^m$ for which the system

$$A\mathbf{x} = \mathbf{b} \ , \quad \mathbf{x} \in S \tag{2.95}$$

is consistent, $X\mathbf{b}$ is the minimum norm solution of (2.95) if and only if $X = P_S(AP_S)^{(1,4)}$ is any S–restricted $\{1.4\}$–inverse of A.

PROOF. Follows from Section 2.8 and Theorem 2. □

EX. 31. Let A, S be as above. Then for any $\mathbf{b} \in \mathbb{C}^m$, $X\mathbf{b}$ is the minimum–norm least–squares solution of (2.95) if and only if $X = P_S(AP_S)^\dagger$, the S–restricted Moore–Penrose inverse of A (Minamide and Nakamura [**1054**]).

3. Weighted generalized inverses

It may be desired to give different weights to the different squared residuals of the linear system $A\mathbf{x} = \mathbf{b}$. This is a more general problem than the one solved by the $\{1, 3\}$–inverse. A still further generalization which, however, presents no greater mathematical difficulty, is the minimizing of a given positive definite quadratic form in the residuals, or, in other words, the minimizing of

$$\|A\mathbf{x} - \mathbf{b}\|_W^2 = (A\mathbf{x} - \mathbf{b})^* W (A\mathbf{x} - \mathbf{b}) \ , \tag{46}$$

where W is a given positive definite matrix, see Ex. 0.4.

When A is not of full column rank, this problem does not have a unique solution for $\mathbf{x}$, and we may choose from the class of "generalized least–squares solutions" the one for which

$$\|\mathbf{x}\|_U^2 = \mathbf{x}^* U \mathbf{x} \tag{47}$$

is smallest, where U is a second positive definite matrix. If $A \in \mathbb{C}^{m\times n}$, W is of order m and U is of order n.

Since every inner product in $\mathbb{C}^n$ can be represented as $\mathbf{x}^* U \mathbf{y}$ for some positive definite matrix U (see Ex. 0.4), it follows that the problem of minimizing (46), and the problem of minimizing (47) among all the minimizers of (46), differ from the problems treated in Sections 1 and 2 only in the different choices of inner products and their associated norms in $\mathbb{C}^m$ and $\mathbb{C}^n$. These seemingly more general problems can be reduced by a simple transformation to the "unweighted" problems considered in Sections 1 and 2. Every positive definite matrix H has a unique positive definite *square root*: that is a positive definite K such that $K^2 = H$ (see, e.g., Ex. 32 and Ex. 6.26 below). Let us denote this K by $H^{1/2}$, and its inverse by $H^{-1/2}$.

We shall now introduce the transformations

$$\widetilde{A} = W^{1/2}AU^{-1/2},\ \widetilde{\mathbf{x}} = U^{1/2}\mathbf{x}\ ,\ \widetilde{\mathbf{b}} = W^{1/2}\mathbf{b}\ , \tag{48}$$

and it is easily verified that

$$\|A\mathbf{x} - \mathbf{b}\|_W = \|\widetilde{A}\widetilde{\mathbf{x}} - \widetilde{\mathbf{b}}\| \tag{49}$$

and

$$\|\mathbf{x}\|_u = \|\widetilde{\mathbf{x}}\|\ , \tag{50}$$

expressing the norms $\|\ \|_W$ and $\|\ \|_U$ in terms of the Euclidean norms of the transformed vectors. This observation leads to the following two theorems.

Theorem 3. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let $W \in \mathbb{C}^{m\times m}$ be positive definite. Then $\|A\mathbf{x} - \mathbf{b}\|_W$ is smallest when $\mathbf{x} = X\mathbf{b}$, where X satisfies

$$AXA = A\ , \quad (WAX)^* = WAX\ . \tag{51}$$

Conversely, if $X \in \mathbb{C}^{n\times m}$ has the property that, for all $\mathbf{b}$, $\|A\mathbf{x} - \mathbf{b}\|_W$ is smallest when $\mathbf{x} = X\mathbf{b}$, then X satisfies (51).

Proof. In view of (49), it follows from Theorem 1 that $\|A\mathbf{x} - \mathbf{b}\|_W$ is smallest when $\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}}$, where Y satisfies

$$\widetilde{A}Y\widetilde{A} = \widetilde{A}\ , \quad (\widetilde{A}Y)^* = \widetilde{A}Y\ , \tag{52}$$

and also if $Y \in \mathbb{C}^{n\times m}$ has the property that, for all $\widetilde{\mathbf{b}}$, $\|\widetilde{A}\widetilde{\mathbf{x}} - \widetilde{\mathbf{b}}\|$ is smallest when $\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}}$, then Y satisfies (52).

Now let

$$X = U^{-1/2}YW^{1/2} \tag{53}$$

so that

$$Y = U^{1/2}XW^{-1/2}\ . \tag{54}$$

Then it is easily verified by means of (48) and (54) that

$$\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}} \Longleftrightarrow \mathbf{x} = X\mathbf{b}\ , \tag{55}$$

$$\widetilde{A}Y\widetilde{A} = \widetilde{A} \Longleftrightarrow AXA = A\ , \tag{56}$$

$$(\widetilde{A}Y)^* = \widetilde{A}Y \Longleftrightarrow (WAX)^* = WAX\ .$$

□

See also Ex. 34.

Theorem 4. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let $U \in \mathbb{C}^{n\times n}$ be positive definite. If $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$, the unique solution for which $\|\mathbf{x}\|_U$ is smallest is given by

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

where X satisfies

$$AXA = A\ , \quad (UXA)^* = UXA\ . \tag{57}$$

Conversely, if $X \in \mathbb{C}^{n\times m}$ is such that, whenever $A\mathbf{x} = \mathbf{b}$ has a solution, $\mathbf{x} = X\mathbf{b}$ is the solution for which $\|\mathbf{x}\|_U$ is smallest, then X satisfies (57).

PROOF. In view of (48),

$$A\mathbf{x} = \mathbf{b} \iff \widetilde{A}\widetilde{\mathbf{x}} = \widetilde{\mathbf{b}} \,.$$

Then it follows from (50) and Theorem 2 that, if $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$, the unique solution for which $\|\mathbf{x}\|_U$ is smallest is given by $\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}}$, where Y satisfies

$$\widetilde{A}Y\widetilde{A} = \widetilde{A} \,, \quad (Y\widetilde{A})^* = Y\widetilde{A} \,, \tag{58}$$

and furthermore if $Y \in \mathbb{C}^{n\times m}$ has the property that, whenever $A\mathbf{x} = \mathbf{b}$ has a solution, $\|\mathbf{x}\|_U$ is smallest when $\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}}$, then Y satisfies (58).

As in the proof of Theorem 3, let X be given by (53), so that (54) holds. Then we have, in addition tgo (55) and (56),

$$(Y\widetilde{A})^* = Y\widetilde{A} \iff (UXA)^* = UXA \,.$$

□

See also Ex. 36.

From Theorems 3 and 4 and Corollary 3, we can easily deduce:

COROLLARY 4. Let $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, and let $W \in \mathbb{C}^{m\times m}$ and $U \in \mathbb{C}^{n\times n}$ be positive definite. Then, there is a unique matrix

$$X = A^{(1,2)}_{(W,U)} \in A\{1,2\}$$

satisfying

$$(WAX)^* = WAX \,, \quad (UXA)^* = UXA \,. \tag{59}$$

Moreover, $\|A\mathbf{x}-\mathbf{b}\|_W$ assumes its minimum value for $\mathbf{x} = X\mathbf{b}$, and in the set of vectors $\mathbf{x}$ for which this minimum value is assumed, $\mathbf{x} = X\mathbf{b}$ is the one for which $\|\mathbf{x}\|_U$ is smallest.

If $Y \in \mathbb{C}^{n\times m}$ has the property that, for all $\mathbf{b}$, $\mathbf{x} = Y\mathbf{b}$ is the vector of $\mathbb{C}^m$ for which $\|\mathbf{x}\|_U$ is smallest among those for which $\|A\mathbf{x} - \mathbf{b}\|_W$ assumes its minimum value, then $Y = A^{(1,2)}_{(W,U)}$. □

See also Exs, 37–44.

Exercises.

EX. 32. *Square root.* Let H be Hermitian positive definite with the spectral decomposition

$$H = \sum_{i=1}^{k} \lambda_i E_i \,. \tag{2.70}$$

Then

$$H^{1/2} = \sum_{i=1}^{k} \lambda_i^{1/2} E_i \,.$$

EX. 33. *Cholesky factorization.* Let H be Hermitian positive definite. Then it can be factorized as

$$H = R_H^* R_H \,, \tag{60}$$

where R_H is an upper triangular matrix. (60) is called the *Cholesky factorization* of H; see, e.g., Wilkinson [**1595**].

Show that the results of Section 3 can be derived by using the Cholesky factorization

$$U = R_U^* R_U \quad \text{and} \quad W = R_W^* R_W \tag{61}$$

of U and W, respectively, instead of their square–root factorizations.

Hint: Instead of (48) use

$$\widetilde{A} = R_W A R_U^{-1} \,, \; \widetilde{\mathbf{x}} = R_U \mathbf{x} \,, \; \widetilde{\mathbf{b}} = R_W \mathbf{b} \,.$$

EX. 34. Let A, $\mathbf{b}$ and W be as in Theorem 3. Show that a vector $\mathbf{x} \in \mathbb{C}^n$ minimizes $\|A\mathbf{x} - \mathbf{b}\|_W$ if and only if $\mathbf{x}$ is a solution of

$$A^*WA\mathbf{x} = A^*W\mathbf{b} ,$$

and compare with Ex. 1.

EX. 35. Let $A_1 \in \mathbb{C}^{m_1 \times n}$, $\mathbf{b}_1 \in \mathbb{C}^{m_1}$, $A_2 \in \mathbb{C}^{m_2 \times n}$, $\mathbf{b}_2 \in R(A_2)$, and let $W \in \mathbb{C}^{m_1 \times m_1}$ be positive definite. Consider the problem

$$\text{minimize } \|A_1\mathbf{x} - \mathbf{b}_1\|_W \text{ subject to } A_2\mathbf{x} = \mathbf{b}_2 . \tag{62}$$

Show that a vector $\mathbf{x} \in \mathbb{C}^n$ is a minimizer of (62) if and only if there is a vector $\mathbf{y} \in \mathbb{C}^{m_2}$ such that the vector $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ is a solution of

$$\begin{bmatrix} A_1^*WA_1 & A_2^* \\ A_2 & O \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} A_1^*W\mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} .$$

Compare with Ex. 17.

EX. 36. Let $A \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in R(A)$, and let $U \in \mathbb{C}^{n \times n}$ be positive definite. Show that the problem

$$\text{minimize } \|\mathbf{x}\|_U \text{ subject to } A\mathbf{x} = \mathbf{b} \tag{63}$$

has the unique minimizer

$$\mathbf{x} = U^{-1}A^*(AU^{-1}A^*)^{(1)}\mathbf{b}$$

and the minimum value

$$\mathbf{b}^*(AU^{-1}A^*)^{(1)}\mathbf{b}$$

where $(AU^{-1}A^*)^{(1)}$ is any $\{1\}$–inverse of $AU^{-1}A^*$ (Rao [**1241**, p. 49]).

Outline of solution. (63) is equivalent to the problem

$$\text{minimize } \|\widetilde{\mathbf{x}}\| \text{ subject to } \widetilde{A}\widetilde{\mathbf{x}} = \widetilde{\mathbf{b}}$$

where $\widetilde{\mathbf{x}} = U^{1/2}\widetilde{\mathbf{b}}$, $\widetilde{A} = AU^{-1/2}$, $\widetilde{\mathbf{b}} = \mathbf{b}$. The unique minimizer of the last problem is, by Theorem 2,

$$\widetilde{\mathbf{x}} = Y\widetilde{\mathbf{b}} \ , \text{ for any } Y \in \widetilde{A}\{1,4\} .$$

Therefore the unique minimizer of (63) is

$$\mathbf{x} = U^{-1/2}X\mathbf{b} \ , \text{ for any } X \in (AU^{-1/2})\{1,4\} .$$

Complete the proof by choosing

$$X = U^{-1/2}A^*(AU^{-1}A^*)^{(1)}$$

which by Theorem 1.3 is a $\{1,2,4\}$–inverse of $AU^{-1/2}$.

EX. 37. *The weighted inverse* $A^{(1,2)}_{(W,U)}$. Chipman [**327**] first called attention to the unique $\{1,2\}$–inverse given by Corollary 4. However, instead of the second equation of (59) he used

$$(XAV)^* = XAV .$$

Show that these two relations are equivalent. How are U and V related?

EX. 38. Use Theorems 3 and 4 to show that

$$A^{(1,2)}_{(W,U)} = U^{-1/2}(W^{1/2}AU^{-1/2})^{\dagger}W^{1/2} ,$$

or equivalently, using (61),

$$A^{(1,2)}_{(W,U)} = R_U^{-1}(R_W A R_U^{-1})^{\dagger}R_W .$$

Ex. 39. Use Exs. 34 and 36 to show that

$$A^{(1,2)}_{(W,U)} = U^{-1}A^*WA(A^*WAU^{-1}A^*WA)^{(1)}A^*W .$$

Ex. 40. For a given A and an arbitrary $X \in A\{1,2\}$, do there exist positive definite matrices W and U such that $X = A^{(1,2)}_{(W,U)}$? Show that this question reduces to the following simpler one. Given an idempotent E, is there a positive definite V, such that VE is Hermitian? Show that such a V is given by

$$V = E^*HE + (I - E^*)K(I - E) ,$$

where H and K are arbitrary positive definite matrices. (This slightly generalizes a result of Ward, Boullion and Lewis [**1536**], who took $H = K = I$.)

SOLUTION. Since H and K are positive definite, $\mathbf{x}^*V\mathbf{x} = 0$ only if both the equations

$$E\mathbf{x} = \mathbf{0} \ , \ \ (I - E)\mathbf{x} = \mathbf{0} \tag{64}$$

hold. But addition of the two equations (64) gives $\mathbf{x} = \mathbf{0}$. Therefore V is positive definite. Moreover

$$VE = E^*HE$$

is clearly Hermitian,

Ex. 41. As a particular illustration, let

$$E = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

and show that V can be taken as any matrix of the form

$$V = \begin{bmatrix} a & a \\ a & b \end{bmatrix} \tag{65}$$

where $b > a > 0$. Show that (65) can be written in the form

$$V = aE^*E + c(I - E^*)(I - E) ,$$

where a and c are arbitrary positive scalars.

Ex. 42. Use Ex. 40 to prove that if X is an arbitrary $\{1,2\}$–inverse of A, there exist positive definite W and U such that $X = A^{(1,2)}_{(W,U)}$. (Ward, Boullion and Lewis [**1536**]).

Ex. 43. Show that

$$A^{(1,2)}_{(W,U)} = A^{(1,2)}_{T,S}$$

(see Theorem 2.10(c)), where the subspaces T, S and the positive definite matrices W, U are related by

$$T = U^{-1}N(A)^{\perp} \tag{66}$$

and

$$S = W^{-1}R(A) \tag{67}$$

or equivalently, by

$$U = P^*_{N(A),T}U_1P_{N(A),T} + P^*_{T,N(A)}U_2P_{T,N(A)} \tag{68}$$

and

$$W = P^*_{R(A),S}W_1P_{R(A),S} + P^*_{S,R(A)}W_2P_{S,R(A)} \tag{69}$$

where U_1, U_2, W_1, and W_2 are arbitrary positive definite matrices of appropriate dimensions.

SOLUTION. From (59), we have

$$XA = U^{-1}A^*X^*U \ ,$$

and therefore

$$R(X) = R(XA) = U^{-1}R(A^*) = U^{-1}N(A)^{\perp}$$

by Corollary 2.7 and (2.47). Also,

$$AX = W^{-1}XAW \ ,$$

and therefore

$$N(X) = N(AX) = N(A^*W) = W^{-1}N(A^*) = W^{-1}R(A)^{\perp}$$

by Corollary 2.7 and (2.48). Finally, from Exs. 40 and 2.23 it follows that the general positive definite matrix U mapping T onto $N(A)^{\perp}$ is given by (68). Equation (69) is similarly proved.

Ex. 44. Let $A = FG$ be a full–rank factorization. Use Ex. 43 and Theorem 2.11(d) to show that

$$A^{(1,2)}_{(W,U)} = U^{-1}G^*(F^*WAU^{-1}G^*)^{-1}F^*W \ .$$

Compare with Ex. 38.

4. Essentially strictly convex norms and the associated projectors and generalized inverses*

[1]In the previous sections various generalized inverses were characterized and studied in terms of their minimization properties with respect to the class of ellipsoidal (or weighted Euclidean) norms

$$\|\mathbf{x}\|_U = (\mathbf{x}^*U\mathbf{x})^{1/2} \ , \tag{47}$$

where U is positive definite.

Given any two ellipsoidal norms $\| \ \|_W$ and $\| \ \|_U$ on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, (defined by (47) and two given positive definite matrices $W \in \mathbb{C}^{m\times m}$ and $U \in \mathbb{C}^{n\times n}$), it was shown in Corollary 4 that every $A \in \mathbb{C}^{m\times n}$ has a unique $\{1,2\}$–inverse $A^{(1,2)}_{(W,U)}$ with the following minimization property:

For any $\mathbf{b} \in \mathbb{C}^m$, the vector $A^{(1,2)}_{(W,U)}\mathbf{b}$ satisfies

$$\|AA^{(1,2)}_{(W,U)}\mathbf{b} - \mathbf{b}\|_W \leq \|A\mathbf{x} - \mathbf{b}\|_W \ , \quad \text{for all } \mathbf{x} \in \mathbb{C}^n \ , \tag{70}$$

and

$$\|A^{(1,2)}_{(W,U)}\mathbf{b}\|_U < \|\mathbf{x}\|_U \tag{71}$$

for any $A^{(1,2)}_{(W,U)}\mathbf{b} \neq \mathbf{x} \in \mathbb{C}^n$ which gives equality in (70). In particular, for $W = I_m$ and $U = I_n$ the inverse mentioned above is the Moore–Penrose inverse

$$A^{(1,2)}_{(I_m,I_n)} = A^{\dagger} \quad \text{for every } A \in \mathbb{C}^{m\times n} \ .$$

In this section, which is based on Erdelsky [**466**], Newman and Odell [**1139**] and Holmes [**743**], similar minimizations are attempted for norms in the more general class of essentially strictly convex norms. The resulting projectors and generalized inverses are, in general, not even linear transformations, but they still retain many useful properties that justify their study.

In this section we denote by $\alpha, \beta, \phi, \ldots$ various vector norms on finite–dimensional spaces; see, e.g., Ex. 0.6.

[1]This section requires familiarity with the basic properties of convex functions and convex sets in finite–dimensional spaces; see, e.g., Rockafellar [**1295**].

Let ϕ be a norm on $\mathbb{C}^n$ and let L be a subspace of $\mathbb{C}^n$. Then for any point $\mathbf{x} \in \mathbb{C}^n$ there is a point $\mathbf{y} \in L$ which is "closest" to $\mathbf{x}$ in the norm ϕ, i.e., a point $\mathbf{y} \in L$ satisfying

$$\phi(\mathbf{y} - \mathbf{x}) = \inf\{\phi(\mathbf{l} - \mathbf{x}) : \mathbf{l} \in L\} \; ; \tag{72}$$

see Ex. 45 below. Generally, the closest point is not unique; see, e.g., Ex. 46. However, Lemma 1 below guarantees the uniqueness of closest points, for the special class of essentially strictly convex norms.

From the definition of a vector norm (see § 0.1.5), it is obvious that every norm ϕ on $\mathbb{C}^n$ is a *convex function*, i.e., for every $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ and $0 \leq \lambda \leq 1$,

$$\phi(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \;\leq\; \lambda\phi(\mathbf{x}) + (1-\lambda)\phi(\mathbf{y}) \; .$$

A function $\phi : \mathbb{C}^n \to \mathbb{R}$ is called *strictly convex* if for all $\mathbf{x} \neq \mathbf{y} \in \mathbb{C}^n$ and $0 < \lambda < 1$,

$$\phi(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \;<\; \lambda\phi(\mathbf{x}) + (1-\lambda)\phi(\mathbf{y}) \; . \tag{73}$$

If $\phi : \mathbb{C}^n \to \mathbb{R}$ is a norm, then (73) is clearly violated for $\mathbf{y} = \mu\mathbf{x}$, $\mu \geq 0$. Thus a norm ϕ on $\mathbb{C}^n$ is not strictly convex. Following Holmes [**743**], a norm ϕ on $\mathbb{C}^n$ is called *essentially strictly convex* (abbreviated *e.s.c*) if ϕ satisfies (73) for all $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \notin \{\mu\mathbf{x} : \mu \geq 0\}$. Equivalently, a norm ϕ on $\mathbb{C}^n$ is e.s.c. if

$$\left.\begin{array}{c} \mathbf{x} \neq \mathbf{y} \in \mathbb{C}^n, \; \phi(\mathbf{x}) = \phi(\mathbf{y}) \\ 0 < \lambda < 1 \end{array}\right\} \implies \phi(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \;<\; \lambda\phi(\mathbf{x}) + (1-\lambda)\phi(\mathbf{y}) \; . \tag{74}$$

The following lemma is a special case of a result in Clarkson [**347**].

Lemma 2. Let ϕ be any e.s.c. norm on $\mathbb{C}^n$. Then for any subspace $L \subset \mathbb{C}^n$ and any point $\mathbf{x} \in \mathbb{C}^n$, there is a unique point $\mathbf{y} \in L$ closest to $\mathbf{x}$, i.e.,

$$\phi(\mathbf{y} - \mathbf{x}) = \inf\{\phi(\mathbf{l} - \mathbf{x}) : \mathbf{l} \in L\} \; . \tag{72}$$

Proof. If $\mathbf{y}_1$, $\mathbf{y}_2 \in L$ satisfy (72) and $\mathbf{y}_1 \neq \mathbf{y}_2$, then for any $0 < \lambda < 1$

$$\phi(\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2 - \mathbf{x}) \;<\; \phi(\mathbf{y}_1 - \mathbf{x}) \; , \quad \text{by (74)} \; ,$$

showing that the point $\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2$, which is in L, is closer to $\mathbf{x}$ than $\mathbf{y}_1$, a contradiction. $\square$

Definition 1. Let ϕ be an e.s.c. norm on $\mathbb{C}^n$ and let L be a subspace of $\mathbb{C}^n$. Then the *ϕ–metric projector on* L, denoted by $P_{L,\phi}$ is the mapping $P_{L,\phi} : \mathbb{C}^n \to L$ assigning to each point in $\mathbb{C}^n$ its (unique) closest point in L, i.e.

$$P_{L,\phi}(\mathbf{x}) \in L$$

and

$$\phi(P_{L,\phi}(\mathbf{x}) - \mathbf{x}) \leq \phi(\mathbf{l} - \mathbf{x}), \quad \text{for all } \mathbf{x} \in \mathbb{C}^n \; , \mathbf{l} \in L \; . \tag{75}$$

If ϕ is a general norm, then the projector $P_{L,\phi}$ defined as above is a point–to–set mapping, since the closest point $P_{L,\phi}(\mathbf{x})$ need not be unique for all $\mathbf{x} \in \mathbb{C}^n$ and $L \subset \mathbb{C}^n$. An excellent survey of metric projectors in normed linear spaces, is given in Holmes [**743**, Section 32]; see also Exs. 65–73 below.

Some properties of $P_{L,\phi}$ in the e.s.c. case are collected in the following theorem, a special case of results by Aronszajn and Smith, and Hirschfeld; see also Singer [**1365**, p. 140, Theorem 6.1].

Theorem 5. Let ϕ be an e.s.c. norm on $\mathbb{C}^n$. Then for any subspace L of $\mathbb{C}^n$ and every point $\mathbf{x} \in \mathbb{C}^n$:

(a) $P_{L,\phi}(\mathbf{x}) = \mathbf{x}$ if and only if $\mathbf{x} \in L$,
(b) $P^2_{L,\phi}(\mathbf{x}) = P_{L,\phi}(\mathbf{x})$,
(c) $P_{L,\phi}(\lambda\mathbf{x}) = \lambda P_{L,\phi}(\mathbf{x})$ for all $\lambda \in \mathbb{C}$,
(d) $P_{L,\phi}(\mathbf{x} + \mathbf{y}) = P_{L,\phi}(\mathbf{x}) + \mathbf{y}$ for all $\mathbf{y} \in L$,
(e) $P_{L,\phi}(\mathbf{x} - P_{L,\phi}(\mathbf{x})) = \mathbf{0}$,

(f) $|\phi(\mathbf{x} - P_{L,\phi}(\mathbf{x})) - \phi(\mathbf{y} - P_{L,\phi}(\mathbf{y}))| \le \phi(\mathbf{x} - \mathbf{y})$ for all $\mathbf{y} \in \mathbb{C}^n$,
(g) $\phi(\mathbf{x} - P_{L,\phi}(\mathbf{x})) \le \phi(\mathbf{x})$,
(h) $\phi(P_{L,\phi}(\mathbf{x})) \le 2\phi(\mathbf{x})$,
(i) $P_{L,\phi}$ is continuous on $\mathbb{C}^n$.

Proof. (a) Follows from (72) and (75) since the infimum in (72) is zero if and only if $\mathbf{x} \in L$.
(b)

$$\begin{aligned} P^2_{L,\phi}(\mathbf{x}) &= P_{L,\phi}(P_{L,\phi}(\mathbf{x})) \\ &= P_{L,\phi}(\mathbf{x}) \quad \text{by (a), since } P_{L,\phi}(\mathbf{x}) \in L\ . \end{aligned}$$

(c) For any $\mathbf{z} \in L$ and $\lambda \ne 0$

$$\begin{aligned} \phi(\lambda\mathbf{x} - \mathbf{z}) &= \phi\left(\lambda\mathbf{x} - \lambda\frac{\mathbf{z}}{\lambda}\right) \\ &= |\lambda|\phi\left(\mathbf{x} - \frac{\mathbf{z}}{\lambda}\right) \\ &\ge |\lambda|\phi\left(\mathbf{x} - P_{L,\phi}(\mathbf{x})\right) \quad \text{by (75)} \\ &= \phi(\lambda\mathbf{x} - \lambda P_{L,\phi}(\mathbf{x}))\ , \end{aligned}$$

which proves (c) for $\lambda \ne 0$. For $\lambda = 0$, (c) is obvious.
(d) From (75) it follows that for all $\mathbf{z} \in L$

$$\phi(P_{L,\phi}(\mathbf{x}) + \mathbf{y} - (\mathbf{x} + \mathbf{y})) \le \phi(\mathbf{z} + \mathbf{y} - (\mathbf{x} + \mathbf{y}))\ ,$$

proving (d).
(e) Follows from (d).
(f) For all $\mathbf{x}$, $\mathbf{y} \in \mathbb{C}^n$

$$\phi(\mathbf{x} - P_{L,\phi}(\mathbf{x})) \le \phi(\mathbf{x} - P_{L,\phi}(\mathbf{y})) \le \phi(\mathbf{x} - \mathbf{y}) + \phi(\mathbf{y} - P_{L,\phi}(\mathbf{y}))$$

and thus

$$\phi(\mathbf{x} - P_{L,\phi}(\mathbf{x})) - \phi(\mathbf{y} - P_{L,\phi}(\mathbf{y})) \le \phi(\mathbf{x} - \mathbf{y})\ ,$$

from which (f) follows by interchanging $\mathbf{x}$ and $\mathbf{y}$.

(g) follows from (f) by taking $\mathbf{y} = \mathbf{0}$.

(h)

$$\begin{aligned} \phi(P_{L,\phi}(\mathbf{x})) &\le \phi(P_{L,\phi}(\mathbf{x}) - \mathbf{x}) + \phi(\mathbf{x}) \\ &\le 2\,\phi(\mathbf{x}) \quad \text{by (g)} \end{aligned}$$

(i) Let $\{\mathbf{x}_k\} \subset \mathbb{C}^n$ be a sequence converging to $\mathbf{x}$

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}\ .$$

Then the sequence $\{P_{L,\phi}(\mathbf{x}_k)\}$ is bounded, by (h), and hence contains a convergent subsequence, also denoted by $\{P_{L,\phi}(\mathbf{x}_k)\}$. Let

$$\lim_{k\to\infty} P_{L,\phi}(\mathbf{x}_k) = \mathbf{y}\ .$$

Then

$$\phi(P_{L,\phi}(\mathbf{x}_k) - \mathbf{x}_k) \leq \phi(P_{L,\phi}(\mathbf{x}) - \mathbf{x}_k)$$

for $k = 1, 2, \dots$ and in the limit,

$$\phi(\mathbf{y} - \mathbf{x}) \leq \phi(P_{L,\phi}(\mathbf{x}) - \mathbf{x})$$

proving the $\mathbf{y} = P_{L,\phi}(\mathbf{x})$. □

The function $P_{L,\phi}$ is homogeneous by Theorem 5(c), but in general it is not additive; i.e., it does not necessarily satisfy

$$P_{L,\phi}(\mathbf{x} + \mathbf{y}) = P_{L,\phi}(\mathbf{x}) + P_{L,\phi}(\mathbf{y}) \ , \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \ .$$

Thus, in general, $P_{L,\phi}$ is not a linear transformation. The following three corollaries deal with cases where $P_{L,\phi}$ is linear.

For any $\mathbf{l} \in L$ we define the *inverse image of* $\mathbf{l}$ *under* $P_{L,\phi}$, denoted by $P_{L,\phi}^{-1}(\mathbf{l})$, as

$$P_{L,\phi}^{-1}(\mathbf{l}) = \{\mathbf{x} \in \mathbb{C}^n : \ P_{L,\phi}(\mathbf{x}) = \mathbf{l}\} \ .$$

We recall that a *linear manifold* (also *affine set*, *flat*, *linear variety*) in $\mathbb{C}^n$ is a set of the form

$$\mathbf{x} + L = \{\mathbf{x} + \mathbf{l} : \ \mathbf{l} \in L\} \ ,$$

where $\mathbf{x}$ and L are a given point and subspace, respectively, in $\mathbb{C}^n$.

The following result is a special case of Theorem 6.4 in Singer [**1365**].

Corollary 5. Let ϕ be an e.s.c. norm on $\mathbb{C}^n$ and let L be a subspace of $\mathbb{C}^n$. Then the following statements are equivalent.

(a) $P_{L,\phi}$ is additive.
(b) $P_{L,\phi}^{-1}(\mathbf{0})$ is a linear subspace.
(c) $P_{L,\phi}^{-1}(\mathbf{l})$ is a linear manifold for any $\mathbf{l} \in L$.

Proof. First we show that

$$P_{L,\phi}^{-1}(\mathbf{0}) = \{\mathbf{x} - P_{L,\phi}(\mathbf{x}) : \ \mathbf{x} \in \mathbb{C}^n\} \ . \tag{76}$$

From Theorem 5(f) it follows that

$$P_{L,\phi}^{-1}(\mathbf{0}) \supset \{\mathbf{x} - P_{L,\phi}(\mathbf{x}) : \ \mathbf{x} \in \mathbb{C}^n\} \ .$$

The reverse containment follows by writing each $\mathbf{x} \in P_{L,\phi}^{-1}(\mathbf{0})$ as

$$\mathbf{x} = \mathbf{x} - P_{L,\phi}(\mathbf{x}) \ .$$

The equivalence of (a) and (b) is obvious from (76). The equivalence of (b) and (c) follows from

$$P_{L,\phi}^{-1}(\mathbf{l}) = \mathbf{l} + P_{L,\phi}^{-1}(\mathbf{0}), \quad \text{for all } \mathbf{l} \in L \ , \tag{77}$$

which is a result of Theorem 5(d) and (e). □

Corollary 6. Let L be a hyperplane of $\mathbb{C}^n$, i.e., an $(n-1)$–dimensional subspace of $\mathbb{C}^n$. Then $P_{L,\phi}$ is additive for any e.s.c. norm ϕ on $\mathbb{C}^n$.

Proof. Let $\mathbf{u}$ be a vector not contained in L. Then any $\mathbf{x} \in \mathbb{C}^n$ is uniquely represented as

$$\mathbf{x} = \lambda\mathbf{u} + \mathbf{l}, \quad \text{where } \lambda \in \mathbb{C}, \mathbf{l} \in L \ ,$$

Therefore, by (76),

$$\begin{aligned} P_{L,\phi}^{-1}(\mathbf{0}) &= \{\lambda\mathbf{u} + (\mathbf{l} - P_{L,\phi}(\lambda\mathbf{u} + \mathbf{l})) : \ \lambda \in \mathbb{C}, \mathbf{l} \in L\} \\ &= \{\lambda\mathbf{u} + P_{L,\phi}(-\lambda\mathbf{u}) : \ \lambda \in \mathbb{C}\} \ , \quad \text{by Theorem 5(d)} \\ &= \{\lambda(\mathbf{u} - P_{L,\phi}(\mathbf{u})) : \ \lambda \in \mathbb{C}\} \ , \quad \text{by Theorem 5(c)} \end{aligned}$$

is a line, proving that $P_{L,\phi}$ is additive, by Corollary 5. □

COROLLARY 7. (Erdelsky [**466**]). Let ϕ be an e.s.c. norm on $\mathbb{C}^n$ and let r be an integer, $1 \leq r < n$. If $P_{L,\phi}$ is additive for all r–dimensional subspaces of $\mathbb{C}^n$, then it is additive for all subspaces of higher dimension.

PROOF. Let L be a subspace with $\dim L > r$, and assume that $P_{L,\phi}$ is not additive. Then by Corollary 5, $P_{L,\phi}^{-1}(\mathbf{0})$ is not a subspace, i.e., there exist $\mathbf{x}_1$, $\mathbf{x}_2 \in P_{L,\phi}^{-1}(\mathbf{0})$ such that $P_{L,\phi}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{y} \neq \mathbf{0}$. Let now M be an r–dimensional subspace of L which contains $\mathbf{y}$. Then $\mathbf{x}_1$, $\mathbf{x}_2 \in P_{M,\phi}^{-1}(\mathbf{0})$, but $P_{M,\phi}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{y} \neq \mathbf{0}$, a contradiction of the hypothesis that $P_{M,\phi}$ is additive. □

See also Exs. 68–71 for additional results on the linearity of the projectors $P_{L,\phi}$.

Following Boullion and Odell [**209**, pp. 43–44] we define generalized inverses associated with pairs of e.s.c. norms as follows.

DEFINITION 2. Let α and β be e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively. For any $A \in \mathbb{C}^{m\times n}$ we define the *generalized inverse associated with* α *and* β, (also called the α–β *generalized inverse*, see, e.g, Boullion and Odell [**209**, p. 44]), denoted by $A^{(-1)}_{\alpha,\beta}$, as

$$A^{(-1)}_{\alpha,\beta} = (I - P_{N(A),\beta})A^{(1)}P_{R(A),\alpha} \ , \tag{78}$$

where $A^{(1)}$ is any $\{1\}$–inverse of A.

RHS(78) means that the three transformations

$$P_{R(A),\alpha} : \mathbb{C}^m \to R(A) \ ,$$
$$A^{(1)} : \mathbb{C}^m \to \mathbb{C}^n \ ,$$

and

$$(I - P_{N(A),\beta}) : \mathbb{C}^n \to P^{-1}_{N(a),\beta}(\mathbf{0}),$$

see, e.g., (76), are performed in this order. We show now that $A^{(-1)}_{\alpha,\beta}$ is a single–valued transformation which does not depend on the particular $\{1\}$–inverse used in its definition. For any $\mathbf{y} \in \mathbb{C}^m$, the set

$$\{A^{(1)}P_{R(A),\alpha}(\mathbf{y}) : \ A^{(1)} \in A\{1\}\}$$

obtained as $A^{(1)}$ ranges over $A\{1\}$, is, by Theorem 1.2, the set of solutions of the linear equation

$$A\mathbf{x} = P_{R(A),\alpha}(\mathbf{y}) \ ,$$

a set which can be written as

$$A^{\dagger}P_{R(A),\alpha}(\mathbf{y}) + \{\mathbf{z} : \ \mathbf{z} \in N(A)\} \ .$$

Now, for any $\mathbf{z} \in N(A)$, it follows from Theorem 5(a) and (d) that

$$(I - P_{N(A),\beta})(A^{\dagger}P_{R(A),\alpha}(\mathbf{y}) + \mathbf{z}) = (I - P_{N(A),\beta})A^{\dagger}P_{R(A),\alpha}(\mathbf{y})$$

proving that

$$A^{(-1)}_{\alpha,\beta}(\mathbf{y}) = (I - P_{N(A),\beta})A^{\dagger}P_{R(A),\alpha}(\mathbf{y}), \quad \text{for all } \mathbf{y} \in \mathbb{C}^n \ , \tag{79}$$

independently of the $\{1\}$–inverse $A^{(1)}$ used in the definition (78).

If the norms α and β are Euclidean, then $P_{R(A),\alpha}$ and $P_{N(A),\beta}$ reduce to the orthogonal projectors $P_{R(A)}$ and $P_{N(A)}$, respectively, and $A^{(-1)}_{\alpha,\beta}$ is, by (79), just the Moore–Penrose inverse $A^{\dagger}$; see also Exs. 66–69 and 73 below. Thus many properties of $A^{\dagger}$ are specializations of the corresponding properties of $A^{(-1)}_{\alpha,\beta}$, some of which are collected in the following theorem. In particular, the minimization properties of $A^{\dagger}$ are special cases of statements (i) and (j) below.

THEOREM 6. (Erdelsky [**466**], Newman and Odell [**1139**]). Let α and β be e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively. Then, for any $A \in \mathbb{C}^{m\times n}$:

(a) $A^{(-1)}_{\alpha,\beta} : \mathbb{C}^m \to \mathbb{C}^n$ is a homogeneous transformation.

(b) $A^{(-1)}_{\alpha,\beta}$ is additive (hence linear) if $P_{R(A),\alpha}$ and $P_{N(A),\beta}$ are additive.

(c) $N(A^{(-1)}_{\alpha,\beta}) = P^{-1}_{R(A),\alpha}(\mathbf{0})$,

(d) $R(A^{(-1)}_{\alpha,\beta}) = P^{-1}_{N(A),\beta}(\mathbf{0})$,

where, as in the case of linear transformations, we denote

$$\begin{aligned} N(A^{(-1)}_{\alpha,\beta}) &= \{\mathbf{y} \in \mathbb{C}^m : A^{(-1)}_{\alpha,\beta}(\mathbf{y}) = \mathbf{0}\}, \\ R(A^{(-1)}_{\alpha,\beta}) &= \{A^{(-1)}_{\alpha,\beta}(\mathbf{y}) : \mathbf{y} \in \mathbb{C}^m\}. \end{aligned}$$

(e) $AA^{(-1)}_{\alpha,\beta} = P_{R(A),\alpha}$.

(f) $A^{(-1)}_{\alpha,\beta}A = I - P_{N(A),\beta}$.

(g) $AA^{(-1)}_{\alpha,\beta}A = A$.

(h) $A^{(-1)}_{\alpha,\beta}AAA^{(-1)}_{\alpha,\beta} = A^{(-1)}_{\alpha,\beta}$.

(i) For any $\mathbf{b} \in \mathbb{C}^m$, an α*–approximate solution* of

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

is defined as any vector $\mathbf{x} \in \mathbb{C}^n$ minimizing $\alpha(A\mathbf{x} - \mathbf{b})$. Then $\mathbf{x}$ is an an α–approximate solution of (1) if and only if

$$A\mathbf{x} = AA^{(-1)}_{\alpha,\beta}(\mathbf{b}). \tag{80}$$

(j) For any $\mathbf{b} \in \mathbb{C}^m$, the equation

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

has a unique α–approximate solution of minimal β–norm, given by $A^{(-1)}_{\alpha,\beta}(\mathbf{b})$; that is, for every $\mathbf{b} \in \mathbb{C}^m$,

$$\alpha(AA^{(-1)}_{\alpha,\beta}(\mathbf{b}) - \mathbf{b}) \leq \alpha(A\mathbf{x} - \mathbf{b}\ , \quad \text{for all } \mathbf{x} \in \mathbb{C}^n\ , \tag{81}$$

and

$$\beta(A^{(-1)}_{\alpha,\beta}(\mathbf{b})) \leq \beta(\mathbf{x}) \tag{82}$$

for any $\mathbf{x} \neq A^{(-1)}_{\alpha,\beta}(\mathbf{b})$ with equality in (81).

PROOF. (a) Follows from the definition and Theorem 5(c).
(b) Obvious from definition (78).
(c) From (78) it is obvious that

$$N(A^{(-1)}_{\alpha,\beta}) \supset P^{-1}_{R(A),\alpha}(\mathbf{0})\ .$$

Conversely, if $\mathbf{y} \neq P^{-1}_{R(A),\alpha}(\mathbf{0})$; i.e., if $P^{-1}_{R(A),\alpha}(\mathbf{y}) \neq \mathbf{0}$, then $A^\dagger P^{-1}_{R(A),\alpha}(\mathbf{y}) \neq \mathbf{0}$ since $(A^\dagger)_{[R(A)]}$ is nonsingular (see Ex. 2.74), and consequently

$$(I - P_{N(A),\beta})A^\dagger P^{-1}_{R(A),\alpha}(\mathbf{y}) \neq \mathbf{0}\ , \quad \text{by Theorem 5(a).}$$

(d) From (76) and the definition (78) it is obvious that

$$R(A^{(-1)}_{\alpha,\beta}) \subset P^{-1}_{N(A),\beta}(\mathbf{0})\ .$$

Conversely, let $\mathbf{x} \in P^{-1}_{N(A),\beta}(\mathbf{0})$. Then, by (76),

$$\begin{aligned}
\mathbf{x} &= (I - P_{N(A),\beta})\mathbf{z}\ , \quad \text{for some } \mathbf{z} \in \mathbb{C}^n \\
&= (I - P_{N(A),\beta})P_{R(A^*)}\mathbf{z}\ , \quad \text{by Theorem 5(d)} \\
&= (I - P_{N(A),\beta})A^{\dagger}A\mathbf{z} \\
&= (I - P_{N(A),\beta})A^{\dagger}P_{R(A),\alpha}(A\mathbf{z}) \\
&= A^{(-1)}_{\alpha,\beta}(A\mathbf{z})\ . \qquad (83)
\end{aligned}$$

(e) Obvious from (79).
(f) For any $\mathbf{z} \in \mathbb{C}^n$ it follows from (83) that

$$(I - P_{N(A),\beta})\mathbf{z} = A^{(-1)}_{\alpha,\beta}(A\mathbf{z})\ .$$

(g) Obvious from (e) and Theorem 5(a).
(h) Obvious from (f) and (d).
(i) A vector $\mathbf{x} \in \mathbb{C}^n$ is an α–approximate solution of (1) if and only if

$$\alpha(A\mathbf{x} - \mathbf{b}) \leq \alpha(\mathbf{y} - \mathbf{b})\ , \quad \text{for all } \mathbf{y} \in R(A)\ ,$$

or equivalently

$$\begin{aligned}
A\mathbf{x} &= P_{R(A),\alpha}(\mathbf{b})\ , \quad \text{by (75)} \\
&= AA^{(-1)}_{\alpha,\beta}(\mathbf{b})\ , \quad \text{by (e)}\ .
\end{aligned}$$

(j) From (80) it follows that $\mathbf{x}$ is an α–approximate solution of (1) if and only if

$$\begin{aligned}
\mathbf{x} &= A^{\dagger}AA^{(-1)}_{\alpha,\beta}(\mathbf{b}) + \mathbf{z}\ , \quad \mathbf{z} \in N(A) \qquad (84) \\
&= A^{\dagger}P_{R(A),\alpha}(\mathbf{b}) + \mathbf{z}\ , \quad \mathbf{z} \in N(A)\ , \quad \text{by (e)}\ .
\end{aligned}$$

Now, by Lemma 2 and Definition 1, the β–norm of

$$A^{\dagger}P_{R(A),\alpha}(\mathbf{b}) + \mathbf{z}\ , \quad \mathbf{z} \in N(A)\ ,$$

is minimized uniquely at

$$\mathbf{z} = -P_{N(A),\beta}A^{\dagger}P_{R(A),\alpha}(\mathbf{b})\ ,$$

which substituted in (84) gives

$$\begin{aligned}
\mathbf{x} &= (I - P_{N(A),\beta})A^{\dagger}P_{R(A),\alpha}(\mathbf{b}) \\
&= A^{(-1)}_{\alpha,\beta}(\mathbf{b})\ .
\end{aligned}$$

□

See Exs. 73–76 for additional results on the generalized inverse $A^{(-1)}_{\alpha,\beta}$.

Exercises and examples.

Ex. 45. *Closest points.* Let ϕ be anorm on $\mathbb{C}^n$ and let L be a nonempty closed set in $\mathbb{C}^n$. Then, for any $\mathbf{x} \in \mathbb{C}^n$, the infimum

$$\inf\{\phi(\mathbf{l} - \mathbf{x} : \mathbf{l} \in L\}$$

is attained at some point $\mathbf{y} \in L$ called ϕ–closest to $\mathbf{x}$ in L.

PROOF. Let $\mathbf{z} \in L$. Then the set

$$K = L \cap \{\mathbf{l} \in \mathbb{C}^n : \phi(\mathbf{l} - \mathbf{x}) \le \phi(\mathbf{z} - \mathbf{x})\}$$

is closed (being the intersection of two closed sets) and bounded, hence compact. The continuous function $\phi(\mathbf{l} - \mathbf{x})$ attains its minimum at some $\mathbf{l} \in K$, but by definition of K,

$$\inf\{\phi(\mathbf{l} - \mathbf{x} : \mathbf{l} \in K\} = \inf\{\phi(\mathbf{l} - \mathbf{x} : \mathbf{l} \in L\} .$$

□

Ex. 46. Let ϕ be the ℓ_1 norm on $\mathbb{R}^2$,

$$\phi(\mathbf{x}) = \phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = |x_1| + |x_2|$$

see, e.g., Ex. 0.8, and let $L = \{\mathbf{x} \in \mathbb{R}^2 : x_1 + x_2 = 1\}$. Then the set of ϕ–closest points in L to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is $\left\{\begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} : -1 \le \alpha \le 1\right\}$.

Ex. 47. Let $\| \ \|$ be the Euclidean norm on $\mathbb{C}^n$, let $S \subset \mathbb{C}^n$ be a convex set and let $\mathbf{x}, \mathbf{y}$ be two points in $\mathbb{C}^n$: $\mathbf{x} \notin S$ and $\mathbf{y} \in S$. Then the following statements are equivalent:

(a) $\mathbf{y}$ is $\| \ \|$–closest to $\mathbf{x}$ in S.

(b) $\mathbf{s} \in S \implies \Re\langle \mathbf{y} - \mathbf{x}, \mathbf{s} - \mathbf{y}\rangle \ge 0$.

PROOF. (adapted from Goldstein [**545**, p. 99]).

(a) $\implies$ (b) For any $0 \le \lambda \le 1$ and $\mathbf{s} \in S$,

$$\mathbf{y} + \lambda(\mathbf{s} - \mathbf{y}) \in S .$$

Now

$$\begin{aligned} 0 &\le \|\mathbf{x} - \mathbf{y} - \lambda(\mathbf{s} - \mathbf{y})\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 \\ &= 2\lambda\Re\langle \mathbf{y} - \mathbf{x}, \mathbf{s} - \mathbf{y}\rangle + \lambda^2\|\mathbf{s} - \mathbf{x}\|^2 \\ &< 0 \text{ if } \quad \Re\langle \mathbf{y} - \mathbf{x}, \mathbf{s} - \mathbf{y}\rangle < 0 \text{ and } 0 < \lambda < -\frac{2\Re\langle \mathbf{y} - \mathbf{x}, \mathbf{s} - \mathbf{y}\rangle}{\|\mathbf{s} - \mathbf{y}\|^2} , \end{aligned}$$

a contradiction to (a).

(b) $\implies$ (a) For any $\mathbf{s} \in S$,

$$\begin{aligned} \|\mathbf{x} - \mathbf{s}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 &= \|\mathbf{s}\|^2 - 2\Re\langle \mathbf{s}, \mathbf{x}\rangle + 2\Re\langle \mathbf{y}, \mathbf{x}\rangle - \|\mathbf{y}\|^2 \\ &= \|\mathbf{s} - \mathbf{y}\|^2 + 2\Re\langle \mathbf{y} - \mathbf{x}, \mathbf{s} - \mathbf{y}\rangle \\ &\ge 0 \quad \text{if (b)} . \end{aligned}$$

□

Ex. 48. *A hyperplane separation theorem.* Let S be a nonempty closed convex set in $\mathbb{C}^n$, $\mathbf{x}$ a point not in S. Then there is a real hyperplane

$$\{\mathbf{z} \in \mathbb{C}^n : \Re\langle \mathbf{u}, \mathbf{z}\rangle = \alpha\} \quad \text{for some } \mathbf{0} \ne \mathbf{u} \in \mathbb{C}^n , \ \alpha \in \mathbb{R}$$

which separates S and $\mathbf{x}$, in the sense that

$$\mathbf{s} \in S \implies \Re\langle \mathbf{u}, \mathbf{s}\rangle \ge \alpha ,$$

and

$$\Re\langle \mathbf{u}, \mathbf{x}\rangle < \alpha .$$

PROOF. Let $\mathbf{x}_S$ be the $\| \; \|$–closest point to $\mathbf{x}$ in S, where $\| \; \|$ is the Euclidean norm, The point $\mathbf{x}_S$ is unique, by the same proof as in Lemma 2, since $\| \; \|$ is e.s.c. Then, for any $\mathbf{s} \in S$,

$$\begin{aligned} \Re\langle \mathbf{x}_S - \mathbf{x}, \mathbf{s} \rangle &\geq \Re\langle \mathbf{x}_S - \mathbf{x}, \mathbf{x}_S \rangle \ , \quad \text{by Ex. 47} \ , \\ &> \Re\langle \mathbf{x}_S - \mathbf{x}, \mathbf{x} \rangle \ , \end{aligned}$$

since

$$\Re\langle \mathbf{x}_S - \mathbf{x}, \mathbf{x}_S - \mathbf{x} \rangle = \|\mathbf{x}_S - \mathbf{x}\|^2 > 0 \ ,$$

The proof is completed by choosing

$$\mathbf{u} = \mathbf{x}_S - \mathbf{x} \ , \quad \alpha = \Re\langle \mathbf{x}_S - \mathbf{x}, \mathbf{x}_S \rangle \ .$$

□

EX. 49. *Gauge functions and their duals.* A function $\phi : \mathbb{C}^n \to \mathbb{R}$ is called a *gauge function* (also a *Minkowski functional*) if

(G1) ϕ is continuous, and for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$,
(G2) $\phi(\mathbf{x}) \geq 0$ and $\phi(\mathbf{x}) = 0$ only if $\mathbf{x} = \mathbf{0}$,
(G3) $\phi(\alpha\mathbf{x}) = \alpha\phi(\mathbf{x})$ for all $\alpha \geq 0$, and
(G4) $\phi(\mathbf{x} + \mathbf{y}) \leq \phi(\mathbf{x}) + \phi(\mathbf{y})$.

A gauge function $\phi : \mathbb{C}^n \to \mathbb{R}$ is called *symmetric* if for all $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{C}^n$,

(G5) $\phi(\mathbf{x}) = \phi(x_1, x_2, \ldots, x_n) = \phi(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$

for every permutation $\{\pi(1), \pi(2), \ldots, \pi(n)\}$ of $\{1, 2, \ldots, n\}$, and

(G6) $\phi(\mathbf{x}) = \phi(x_1, x_2, \ldots, x_n) = \phi(\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_n x_n)$

for every scalar sequence $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ satisfying

$$\begin{cases} |\lambda_i| = 1 & \text{if } \phi : \mathbb{C}^n \to \mathbb{R} \ , \\ \lambda_i = \pm 1 & \text{if } \phi : \mathbb{R}^n \to \mathbb{R} \ , \end{cases} \quad i \in \overline{1, n} \ .$$

Let $\phi : \mathbb{C}^n \to \mathbb{R}$ satisfy (G1)–(G3). The *dual function*[2] of ϕ is the function $\phi_D : \mathbb{C}^n \to \mathbb{R}$ defined by

$$\phi_D(\mathbf{y}) = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\Re\langle \mathbf{y}, \mathbf{x} \rangle}{\phi(\mathbf{x})} \ . \tag{85}$$

Then:
(a) The supremum in (85) is attained, and

$$\phi_D(\mathbf{y}) = \max_{\mathbf{x} \in S_i} \frac{\Re\langle \mathbf{y}, \mathbf{x} \rangle}{\phi(\mathbf{x})} \ , \quad i = 1 \text{ or } \phi \ , \tag{86}$$

where

$$S_1 = \{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| = 1\} \tag{87}$$

and

$$S_\phi = \{\mathbf{x} \in \mathbb{C}^n : \phi(\mathbf{x}) = 1\} \ . \tag{88}$$

(b) ϕ_D is a gauge function.
(c) ϕ_D satisfies (G5) [(G6)] if ϕ does.
(d) If ϕ is a gauge function (i.e., if ϕ also satisfies (G4)), then ϕ is the conjugate of ϕ_D (Bonnesen and Fenchel [**192**], von Neumann [**1506**]).

---

[2]Originally, ϕ_D was called the *conjugate* of ϕ by Bonnesen and Fenchel [**192**] and von Neumann [**1506**]. However, in the modern convexity literature, the word *conjugate function* has a different meaning, see, e.g., Rockafellar [**1295**].

PROOF. (a). From (G3) it follows that the constraint $\mathbf{x} \neq \mathbf{0}$ in (85) can be replaced by $\mathbf{x} \in S_1$, or alternatively, by $\mathbf{x} \in S_\phi$. The supremum is attained since S_1 is compact.

(b),(c). The continuity of ϕ_D follows from (G1), (86) and the compactness of S_1. It is easy to show that ϕ shares with ϕ_D each of the properties (G2), (G3), (G5), and (G6), while (G4) holds for ϕ_D, by definition (85), without requiring that it hold for ϕ.

(d). From (85) it follows that

$$\Re\langle \mathbf{y}, \mathbf{x}\rangle \leq \phi(\mathbf{x})\phi_D(\mathbf{y}) \;, \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \;, \tag{89}$$

and hence

$$\phi(\mathbf{x}) \geq \sup_{\mathbf{y} \neq \mathbf{0}} \frac{\Re\langle \mathbf{y}, \mathbf{x}\rangle}{\phi_D(\mathbf{y})} \;. \tag{90}$$

To show equality in (90) we note that the set

$$B = \{\mathbf{z} : \phi(\mathbf{z}) \leq 1\}$$

is a closed convex set in $\mathbb{C}^n$, an easy consequence of the definition of a gauge function. From the hyperplane separation theorem (see, e.g., Ex. 48 above) we conclude:

$$\begin{gathered}\text{If a point } \mathbf{x} \text{ is contained in every closed half–space } \{\mathbf{z} : \Re\langle \mathbf{u}, \mathbf{z}\rangle \leq 1\} \\ \text{which contains } B, \text{ then } \mathbf{x} \in B, \text{ i.e., } \phi(\mathbf{x}) \leq 1 \;.\end{gathered} \tag{91}$$

From (86) and (88) it follows that

$$B \subset \{\mathbf{z} : \Re\langle \mathbf{y}, \mathbf{z}\rangle \leq 1\}$$

is equivalent to

$$\phi_D(\mathbf{y}) \leq 1 \;.$$

Statement (91) is thus equivalent to

$$\{\phi_D(\mathbf{y}) \leq 1 \implies \Re\langle \mathbf{y}, \mathbf{x}\rangle \leq 1\} \implies \phi(\mathbf{x}) \leq 1$$

which proves equality in (90). □

Ex. 50. *Convex bodies and gauge functions.* A *convex body* in $\mathbb{C}^n$ is a closed bounded convex set with nonempty interior.

Let $B \subset \mathbb{C}^n$ be a convex body and let $\mathbf{0} \in \operatorname{int} B$ where $\operatorname{int} B$ denotes the *interior* of B. The *gauge function* (or *Minkowki functional*) *of* B is the function $\phi^B : \mathbb{C}^n \to \mathbb{R}$ defined by

$$\phi^B(\mathbf{x}) = \inf\{\lambda > 0 : \mathbf{x} \in \lambda B\} \;. \tag{92}$$

Then:
(a) ϕ^B is a gauge function, i.e., it satisfies (G1)–(G4) of Ex. 49.
(b) $B = \{\mathbf{x} \in \mathbb{C}^n : \phi^B(\mathbf{x}) \leq 1\}$.
(c) $\operatorname{int} B = \{\mathbf{x} \in \mathbb{C}^n : \phi^B(\mathbf{x}) < 1\}$.

Conversely, if $\phi : \mathbb{C}^n \to \mathbb{R}$ is any gauge function, then ϕ is the gauge function ϕ^B of a convex body B defined by

$$B = \{\mathbf{x} \in \mathbb{C}^n : \phi^B(\mathbf{x}) \leq 1\} \;, \tag{93}$$

which has $\mathbf{0}$ as an interior point.

Thus (92) and (93) establish a one–to–one correspondence between all gauge functions $\phi : \mathbb{C}^n \to \mathbb{R}$ and all convex bodies $B \subset \mathbb{C}^n$ with $\mathbf{0} \in \operatorname{int} B$.

Ex. 51. A set $B \in \mathbb{C}^n$ is called *equilibrated* if

$$\mathbf{x} \in B, \; |\lambda| \leq 1 \implies \lambda\mathbf{x} \in B \;.$$

Clearly, $\mathbf{0}$ is an interior point of any equilibrated convex body.

Let B be a convex body, $\mathbf{0} \in \operatorname{int} B$. Then B is equilibrated if and only if its gauge function ϕ^B satisfies

$$\phi^B(\lambda \mathbf{x}) = |\lambda| \phi^B(\mathbf{x}) \quad \text{for all } \lambda \in \mathbb{C},\, \mathbf{x} \in \mathbb{C}^n \,. \tag{94}$$

EX. 52. *Vector norms.* From the definition of a *vector norm* (§ 0.1.5) and a gauge function (Ex. 49) it follows that a function $\phi : \mathbb{C}^n \to \mathbb{R}$ is a norm if and only if ϕ is a gauge function satisfying (94).

Thus (92) and (93) establish a one–to–one correspondence between all norms $\phi : \mathbb{C}^n \to \mathbb{R}$ and all equilibrated convex bodies $B \in \mathbb{C}^n$ (Householder [**753**, Chapter 2]).

EX. 53. If a norm $\phi : \mathbb{C}^n \to \mathbb{R}$ is *unitarily invariant* (i.e., if $\phi(U\mathbf{x}) = \phi(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{C}^n$ and any unitary matrix $U \in \mathbb{C}^{n\times n}$) then ϕ is a symmetric gauge function (see Ex. 49). Is the converse true?

EX. 54. *Dual norms.* The *dual* (also *polar*) of a nonempty set $B \subset \mathbb{C}^n$ is the set B_D defined by

$$B_D = \{\mathbf{y} \in \mathbb{C}^n : \mathbf{x} \in B \implies \Re\langle \mathbf{y}, \mathbf{x}\rangle \le 1\} \,. \tag{95}$$

Let $B \subset \mathbb{C}^n$ be an equilibrated convex body. Then

(a) B_D is an equilibrated convex body.

(b) $(B_D)_D = B$, i.e., B is the dual of its dual.

(c) Let ϕ^B be the norm corresponding to B via (92). Then the dual of ϕ^B, computed by (85),

$$\phi^B_D(\mathbf{y}) = \sup_{\mathbf{x}\neq \mathbf{0}} \frac{\Re\langle \mathbf{y}, \mathbf{x}\rangle}{\phi^B(\mathbf{x})} \,, \tag{96}$$

is the norm corresponding to B_D. The norm ϕ^B_D, defined by (96), is called the *dual* of ϕ^B.

(d) $\left(\phi^B_D\right)_D = \phi^B$, i.e., ϕ^B is the dual of its dual. Such pairs $\{\phi^B, \phi^B_D\}$ are called *dual norms* (Householder, [**753**, Chapter 2]).

EX. 55. *ℓ_p–norms.* If ϕ is an ℓ_p–norm, $p \ge 1$, (see Exs. 0.7–8), then its dual is an ℓ_q–norm where q is determined by

$$\frac{1}{p} + \frac{1}{q} = 1 \,.$$

In particular, the ℓ_1 and ℓ_∞ norms are dual, while the Euclidean norm (the ℓ_2–norm) is self–dual.

EX. 56. *The generalized Cauchy–Schwartz inequality.* Let $\{\phi, \phi_D\}$ be dual norms on $\mathbb{C}^n$. Then

$$\Re\langle \mathbf{y}, \mathbf{x}\rangle \le \phi(\mathbf{x})\phi_D(\mathbf{y}) \,, \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \,, \tag{89}$$

and for any $\mathbf{x} \neq \mathbf{0}\,[\mathbf{y} \neq \mathbf{0}]$ there exists a $\mathbf{y} \neq \mathbf{0}\,[\mathbf{x} \neq \mathbf{0}]$ giving equality in (89). Such pairs $\{\mathbf{x}, \mathbf{y}\}$ are called *dual vectors* (with respect to the norm ϕ).

If ϕ is the Euclidean norm, then (89) reduces to the classical Cauchy–Schwartz inequality (0.4), (Householder [**753**]).

EX. 57. *A Tchebycheff solution of $A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{C}^{(n+1)\times n}n$.* A *Tchebycheff approximate solution* of the system

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

is, by the definition in Theorem 6(i), a vector $\mathbf{x}$ minimizing the Tchebycheff norm

$$\|\mathbf{r}\|_\infty = \max_{i=1,\dots,m} \{|r_i|\}$$

of the residual vector

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} \,. \tag{2}$$

Let $A \in \mathbb{C}^{(n+1)\times n}_n$ and $\mathbf{b} \in \mathbb{C}^{n+1}$ be such that (1) is inconsistent. Then (1) has a unique Tchebycheff approximate solution given by

$$\mathbf{x} = A^\dagger(\mathbf{b} + \mathbf{r}) \,, \tag{97}$$

where the residual $\mathbf{r} = [r_i]$ is

$$r_i = \frac{\sum_{j=1}^{n+1} |(P_{N(A^*)}\mathbf{b})_j|^2}{\sum_{j=1}^{n+1} |(P_{N(A^*)}\mathbf{b})_j|} \frac{(P_{N(A^*)}\mathbf{b})_i}{|(P_{N(A^*)}\mathbf{b})_i|} , \quad i \in \overline{1, n+1} . \tag{98}$$

(The real case appeared in Cheney [**324**, p. 41] and Meicler [**1015**].)

Proof. From

$$\mathbf{r}(\mathbf{x}) - \mathbf{b} = -A\mathbf{x} \in R(A)$$

it follows that any residual $\mathbf{r}$ satisfies

$$P_{N(A^*)}\mathbf{r} = P_{N(A^*)}\mathbf{b}$$

or equivalently

$$\langle P_{N(A^*)}\mathbf{b}, \mathbf{r}\rangle = \langle \mathbf{b}, P_{N(A^*)}\mathbf{b}\rangle , \tag{99}$$

since $\dim N(A^*) = 1$ and $\mathbf{b} \notin R(A)$. (Equation (99) represents the hyperplane of residuals; see, e.g., Cheney [**324**, Lemma, p. 40]). A routine computation now shows, that among all vectors $\mathbf{r}$ satisfying (99) there is a unique vector of minimum Tchebycheff norm given by (98), from which (97) follows since $N(A) = \{\mathbf{0}\}$. □

Ex. 58. Let $A \in \mathbb{C}_n^{(n+1)\times n}$ and $\mathbf{b} \in \mathbb{C}^{n+1}$ be such that (1) is inconsistent. Then, for any norm ϕ on $\mathbb{C}^n$, a ϕ–approximate solution of (1) is given by

$$\mathbf{x} = A^\dagger(\mathbf{b} + \mathbf{r}) ,$$

where the residual $\mathbf{r}$ is a dual vector of $P_{N(A^*)}\mathbf{b}$ with respect to the norm ϕ, and the error of approximation is

$$\phi(\mathbf{r}) = \frac{\langle \mathbf{b}, P_{N(A^*)}\mathbf{b}\rangle}{\phi_D(P_{N(A^*)}\mathbf{b})} .$$

Proof. Follows from (99) and Ex. 56. □

Ex. 59. Let $\{\phi, \phi_D\}$ be dual norms with unit balls $B = \{\mathbf{x} : \phi(\mathbf{x}) \le 1\}$ and $B_D = \{\mathbf{y} : \phi_D(\mathbf{y}) \le 1\}$, respectively, and let $\{\mathbf{x}_0, \mathbf{y}_0\}$ be dual vectors of norm one, i.e., $\phi(\mathbf{x}_0) = 1$, $\phi_D(\mathbf{y}) = 1$ and

$$\Re\langle \mathbf{x}_0, \mathbf{y}_0\rangle = \phi(\mathbf{x}_0)\phi_D(\mathbf{y}_0) .$$

Then

(a) The hyperplane

$$H = \{\mathbf{x} : \Re\langle \mathbf{y}_0, \mathbf{x}\rangle = \phi(\mathbf{x}_0)\phi_D(\mathbf{y}_0)\}$$

supports B at $\mathbf{x}_0$, that is, $\mathbf{x}_0 \in H$ and B lies on one side of H, i.e.,

$$\mathbf{x} \in B \implies \Re\langle \mathbf{y}_0, \mathbf{x}\rangle \le \Re\langle \mathbf{y}_0, \mathbf{x}_0\rangle = \phi(\mathbf{x}_0)\phi_D(\mathbf{y}_0) .$$

(b) The hyperplane

$$\{\mathbf{y} : \Re\langle \mathbf{x}_0, \mathbf{y}\rangle = \phi(\mathbf{x}_0)\phi_D(\mathbf{y}_0)\}$$

supports B_D at $\mathbf{y}_0$.

Proof. Follows from (89). □

Ex. 60. A closed convex set B is called *rotund* if its boundary contains no line segments, or equivalently, if each one of its boundary points is an extreme point.

A closed convex set is called *smooth* if it has, at each boundary point, a unique supporting hyperplane.

Show that an equilibrated convex body B is rotund if and only if its dual set B_D is smooth.

PROOF. *If.* if B is not rotund then its boundary contains two points $\mathbf{x}_0 \neq \mathbf{x}_1$ and the line segment $\{\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_0 : 0 \leq \lambda \leq 1\}$ joining them; that is

$$\phi(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_0) = 1, \quad 0 \leq \lambda \leq 1 ,$$

where ϕ is the gauge function of B.

For any $0 < \lambda < 1$ let $\mathbf{y}_\lambda$ be a dual vector of $\mathbf{x}_\lambda = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_0$ with $\phi_D(\mathbf{y}_\lambda) = 1$. Then

$$\Re\langle\mathbf{x}_\lambda, \mathbf{y}_\lambda\rangle = 1$$

and, by (89)

$$\Re\langle\mathbf{x}_0, \mathbf{y}_\lambda\rangle = \Re\langle\mathbf{x}_1, \mathbf{y}_\lambda\rangle = 1 ,$$

showing that $\mathbf{y}_\lambda$ is a dual vector of both $\mathbf{x}_0$ and $\mathbf{x}_1$, and by Ex. 59(b), both hyperplanes

$$\{\mathbf{y} : \Re\langle\mathbf{x}_\lambda, \mathbf{y}\rangle = 1\}, \quad \lambda = 0, 1$$

support B_D at $\mathbf{y}_\lambda$.

Only if. Follows by reversing the above steps. ☐

For additional results and references on rotundity see the survey of Cudia [**368**].

Ex. 61. Let ϕ be a norm on $\mathbb{C}^n$ and let B be its unit ball,

$$B = \{\mathbf{x} : \phi(\mathbf{x}) \leq 1\} .$$

Then

(a) ϕ is e.s.c. if and only if B is rotund.

(b) ϕ is Gateaux differentiable; that is the limit

$$\phi'(\mathbf{x};\mathbf{y}) = \lim_{t\to 0} \frac{\phi(\mathbf{x}+t\mathbf{y}) - \phi(\mathbf{x})}{t}$$

exists for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, if and only if B is smooth.

Ex. 62. Give an example of dual norms $\{\phi, \phi_D\}$ such that ϕ is e.s.c. but ϕ_D is not.

SOLUTION. Let

$$B = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_1 \geq \frac{1}{2}(x_2+1)^2 - 1,\ x_2 \geq \frac{1}{2}(x_1+1)^2 - 1 \right\} .$$

Then B is an equilibrated convex body. B is rotund but not smooth (the points $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ are "corners" of B), so, by Ex. 58, the dual set B_D is not rotund. Hence, by Ex. 61(a), the gauge function ϕ^B is an e.s.c. norm, but its dual ϕ^B_D is not.

Ex. 63. *Norms of homogeneous transformations* (Bauer [**98**], Householder [**753**]). Let α and β be norms on $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. Let $A : \mathbb{C}^n \to \mathbb{C}^m$ be a continuous transformation that is homogeneous; that is

$$A(\lambda\mathbf{x}) = \lambda A(\mathbf{x}) , \quad \text{for all } \lambda \in \mathbb{C},\ \mathbf{x} \in \mathbb{C}^n .$$

The *norm* (also *least upper bound*) *of* A *corresponding to* $\{\alpha, \beta\}$, denoted by $\|A\|_{\alpha,\beta}$ (also by $\text{lub}_{\alpha,\beta}(A)$) is defined as

$$\|A\|_{\alpha,\beta} = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\beta(A\mathbf{x})}{\alpha(\mathbf{x})} = \max_{\alpha(\mathbf{x})=1} \beta(A\mathbf{x}) , \tag{100}$$

since A is continuous and homogeneous. Then for any A, A_1, A_2 as above:

(a) $\|A\|_{\alpha,\beta} \geq 0$ with equality if and only if A is the zero transformation.

(b) $\|\lambda A\|_{\alpha,\beta} = |\lambda| \, \|A\|_{\alpha,\beta}$ for all $\lambda \in \mathbb{C}$.

(c) $\|A_1 + A_2\|_{\alpha,\beta} \leq \|A_1\|_{\alpha,\beta} + \|A_2\|_{\alpha,\beta}$.

(d) If B_α, B_β are the unit balls of α, β, respectively, then

$$\|A\|_{\alpha,\beta} = \inf\{\lambda > 0 : \, AB_\alpha \subset \lambda B_\beta\} .$$

(e) If $A_1 : \mathbb{C}^n \to \mathbb{C}^m$ and $A_2 : \mathbb{C}^m \to \mathbb{C}^p$ are continuous homogeneous transformations, and if α, β, and γ are norms on $\mathbb{C}^n, \mathbb{C}^m$, and $\mathbb{C}^p$, respectively, then

$$\|A_2 A_1\|_{\alpha,\gamma} \leq \|A_1\|_{\alpha,\beta} \, \|A_2\|_{\beta,\gamma} .$$

(f) If $A : \mathbb{C}^n \to \mathbb{C}^m$ is a linear transformation, and if $\alpha = \beta$, i.e., the same norm is used in $\mathbb{C}^n$ and $\mathbb{C}^m$, then definition (100) reduces to that given in Ex. 0.28.

Ex. 64. Let α and β be norms on $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. Then for any $A \in \mathbb{C}^{m\times n}$

$$\|A\|_{\alpha,\beta} = \|A^*\|_{\beta_D,\alpha_D} . \tag{101}$$

Proof. From (89) and (100) it follows that for all $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{y} \in \mathbb{C}^m$

$$\Re\langle A\mathbf{x}, \mathbf{y}\rangle \;\leq\; \beta(A\mathbf{x})\, \beta_D(\mathbf{y}) \;\leq\; \|A\|_{\alpha,\beta}\, \alpha(\mathbf{x})\, \beta_D(\mathbf{y}) ,$$

with equality for at least one pair $\mathbf{x} \neq \mathbf{0}$, $\mathbf{y} \neq \mathbf{0}$. The dual inequalities

$$\Re\langle \mathbf{x}, A^*\mathbf{y}\rangle \;\leq\; \alpha(\mathbf{x})\, \alpha_D(A^*\mathbf{y}) \;\leq\; \|A^*\|_{\beta_D,\alpha_D}\, \beta_D(\mathbf{y})\, \alpha(\mathbf{x})$$

then show that

$$\|A\|_{\alpha,\beta} \;\leq\; \|A^*\|_{\beta_D,\alpha_D} ,$$

from which (101) follows by reversing the roles of A and A^* and by using Ex. 54(d). □

Ex. 65. *Projective bounds* (Erdelsky [**466**]). Let α be an e.s.c. norm on $\mathbb{C}^n$. The *projective bound* of α, denoted by $Q(\alpha)$, is defined as

$$Q(\alpha) = \sup_{L} \|P_{L,\alpha}\|_{\alpha,\alpha} , \tag{102}$$

where the supremum is taken over all subspaces L with domension $1 \leq \dim L \leq n-1$. (The α–metric projector $P_{L,\alpha}$ is continuous and homogeneous, by Theorem 5(c) and (i), allowing the use of (100) to define $\|P_{L,\alpha}\|_{\alpha,\alpha}$). Then

(a) The supremum in (102) is finite and is attained for a k–dimensional subspace, for each $k = 1, 2, \ldots, n-1$.

(b) The projective bound satisfies

$$1 \leq Q(\alpha) < 2 \tag{103}$$

and the upper limit is approached arbitrarily closely by e.s.c. norms.

Proof. (a) It can be shown that the $n-1$ sets of real numbers

$$S_k = \{\alpha(P_{L,\alpha}(\mathbf{x})) : \, \alpha(\mathbf{x}) = 1 \,, \; L \text{ is } k\text{–dimensional}\} \,, \; k = 1, 2, \ldots, n-1,$$

are identical, bounded, and contain the supremum $Q(\alpha)$.

(b) From Theorem 5(a) and (h) it follows that

$$1 \leq Q(\alpha) \leq 2 \ .$$

Let $\mathbf{x}$ be such that $\|P_{L,\alpha}\|_{\alpha,\alpha} = \alpha(P_{L,\alpha}(\mathbf{x}))$ and $\alpha(\mathbf{x}) = 1$. Then $P_{L,\alpha}(\mathbf{x}) \neq \mathbf{0}$ and consequently

$$1 = \alpha(\mathbf{x}) = \alpha(\mathbf{0} - \mathbf{x}) > \alpha(P_{L,\alpha}(\mathbf{x}) - \mathbf{x})$$

and

$$\|P_{L,\alpha}\|_{\alpha,\alpha} = \alpha(P_{L,\alpha}(\mathbf{x})) \leq \alpha(P_{L,\alpha}(\mathbf{x}) - \mathbf{x}) + \alpha(\mathbf{x}) < 2 \ ,$$

proving (102). Let $\{B_k\}$ be a sequence of rotund equilibrated convex bodies in $\mathbb{R}^2$ satisfying

$$B_{k+1} \subset B_k \ , \quad k = 1, 2, \ldots$$

and "converging" to

$$B = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : \ |x_1| \leq 1, \ |x_2| \leq 1 \right\} \ .$$

Then the corresponding norms $\{\phi^{B_k}\}$ are e.s.c., by Ex. 61(a), and "approximate" ϕ^B, which is the ℓ_∞–norm on $\mathbb{R}^2$,

$$\phi^B \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \max\{|x_1|, \ |x_2|\} \ .$$

Finally, by (92)

$$\phi^{B_k} \leq \phi^{B_{k+1}}, \quad k = 1, 2, \ldots$$

and

$$\sup_k Q(\phi^{B_k}) = 2 \ .$$

□

Ex. 66. *Projective norms* (Erdelsky [**466**]). An e.s.c. norm α on $\mathbb{C}^n$ for which the projective bound

$$Q(\alpha) = 1$$

is called a *projective norm.* All ellipsoidal norms

$$\|\mathbf{x}\|_U = (\mathbf{x}^* U \mathbf{x})^{1/2} \ , \quad U \text{ positive definite} \tag{47}$$

are projective.

Conversely, for spaces of dimension ≥ 3, all projective norms are ellipsoidal, both in the real case (Kakutani [**799**]) and in the complex case (Bohnenblust [**191**]). An example of a nonellipsoidal projective norm on $\mathbb{R}^2$ is

$$\alpha \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{cases} (|x_1|^p + |x_2|^p)^{1/p} & \text{if } x_1 x_2 \geq 0 \\ (|x_1|^q + |x_2|^q)^{1/q} & \text{if } x_1 x_2 < 0 \end{cases}$$

where $(1/p) + (1/q) = 1 \ , \ 1 < p \neq 2$.

Ex. 67. (Erdelsky [**466**]). If α is a projective norm, L is a subspace for which the α–metric projector $P_{L,\alpha}$ is linear, and N denotes

$$N = P_{L,\alpha}^{-1}(\mathbf{0}) \ , \tag{104}$$

then

$$L = P_{N,\alpha}^{-1}(\mathbf{0}) \ . \tag{105}$$

PROOF. $L \subset P_{N,\alpha}^{-1}(\mathbf{0})$. If $\mathbf{x} \in L$ and $\mathbf{y} \in N$ then

$$P_{L,\alpha}^{-1}(\mathbf{x}+\mathbf{y}) = \mathbf{x} \ ,$$

by Theorem 5(a) and consequently,

$$\begin{aligned} \alpha(\mathbf{x}) &\leq \|P_{L,\alpha}\|_{\alpha,\alpha}\, \alpha(\mathbf{x}+\mathbf{y}) \\ &\leq Q(\alpha)\, \alpha(\mathbf{x}+\mathbf{y}) \\ &= \alpha(\mathbf{x}+\mathbf{y}) \end{aligned}$$

for all $\mathbf{y} \in N$, proving that $P_{N,\alpha}(\mathbf{x}) = \mathbf{0}$.

$P_{N,\alpha}^{-1}(\mathbf{0}) \subset L$. If $\mathbf{x} \in P_{N,\alpha}^{-1}(\mathbf{0})$, then, by (76), it can be written as

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 \ , \quad \mathbf{x}_1 \in L \ , \ \mathbf{x}_2 \in N \ .$$

Therefore,

$$\begin{aligned} \mathbf{0} = P_{N,\alpha}(\mathbf{x}) &= P_{N,\alpha}(\mathbf{x}_1) + \mathbf{x}_2 \ , \quad \text{by Theorem 5(d)} \\ &= \mathbf{x}_2 \ , \quad \text{since } L \subset P_{N,\alpha}^{-1}(\mathbf{0}) \ , \end{aligned}$$

proving that

$$\mathbf{x} = \mathbf{x}_1 \in L \ .$$

□

Projective norms and the linearity of metric projectors.

The following four exercises probe the relations between the linearity of the α–metric projector $P_{L,\alpha}$ and the projectivity of the norm α. Exercise 68 shows that

$$\alpha \text{ projective} \implies P_{L,\alpha} \text{ linear for all } L \ ,$$

and a partial converse is proved in Ex. 70.

Ex. 68. (Erdelsky [**466**]). If α is a projective norm on $\mathbb{C}^n$, then $P_{L,\alpha}$ is linear for all subspaces L of $\mathbb{C}^n$.

PROOF. By Corollary 7 it suffices to prove linearity of $P_{L,\alpha}$ for all one–dimensional subspaces L.

Let $\dim L = 1$, $\mathbf{l} \in L$, $\alpha(\mathbf{l}) = 1$, and let $\mathbf{l}+N$ be a supporting hyperplane of $B_\alpha = \{\mathbf{x} : \alpha(\mathbf{x}) \leq 1\}$ at $\mathbf{l}$. Since

$$\alpha(\mathbf{l}) \leq \alpha(\mathbf{x}) \ , \quad \text{for all } \mathbf{x} \in \mathbf{l} + N \ ,$$

it follows from Definition 1 that

$$P_{N,\alpha}(\mathbf{l}) = \mathbf{0}$$

and hence

$$L \subset P_{N,\alpha}^{-1}(\mathbf{0}) \ .$$

Now $P_{N,\alpha}$ is linear by Corollary 6, since $\dim N = n-1$, which also shows that $P_{N,\alpha}^{-1}(\mathbf{0})$ is a 1–dimensional subspace, by (76), and hence

$$L = P_{N,\alpha}^{-1}(\mathbf{0}) \ .$$

From Ex. 67 it follows then that

$$N = P_{L,\alpha}^{-1}(\mathbf{0}) \ ,$$

and the linearity of $P_{L,\alpha}$ is established by Corollary 5(b). □

Ex. 69. (Erdelsky [**466**]). If α is an e.s.c. norm on $\mathbb{C}^n$, L is a subspace for which $P_{L,\alpha}$ is linear, and N denotes

$$N = P_{L,\alpha}^{-1}(\mathbf{0}) \ , \tag{104}$$

then

$$L = P_{N,\alpha}^{-1}(\mathbf{0}) \tag{105}$$

if, and only if,

$$P_{L,\alpha} + P_{N,\alpha} = 1 \ .$$

Proof. Follows from (76). □

Ex. 70. (Erdelsky [**466**]). Let α be an e.s.c. norm on $\mathbb{C}^n$ and let $1 \leq k \leq n-1$ be an integer such that, for every k–dimensional subspace L of $\mathbb{C}^n$:

$$P_{L,\alpha} \text{ is linear}$$

and

$$L = P_{N,\alpha}^{-1}(\mathbf{0}) \ , \tag{105}$$

where N is given by (104). Then α is projective.

Proof. Let α be nonprojective; i.e., let $Q(\alpha) > 1$. Then there is a k–dimensional subspace L and two points $\mathbf{x}, \mathbf{y}$ in $\mathbb{C}^n$ such that

$$\mathbf{y} = P_{L,\alpha}(\mathbf{x}) \tag{106}$$

and

$$\alpha(\mathbf{y}) = \|P_{L,\alpha}\|_{\alpha,\alpha}\, \alpha(\mathbf{x}) = Q(\alpha)\, \alpha(\mathbf{x}) > \alpha(\mathbf{x}) \ . \tag{107}$$

Let $N = P_{L,\alpha}^{-1}(\mathbf{0})$. Then

$$\mathbf{0} \neq \mathbf{y} - \mathbf{x} \in N \ , \quad \text{by (107),(106) and (76)} \tag{108}$$

and

$$\alpha(\mathbf{x}) = \alpha(\mathbf{y} - (\mathbf{y} - \mathbf{x})) < \alpha(\mathbf{y}) \ . \tag{109}$$

Now

$$\begin{aligned} \mathbf{y} &= P_{L,\alpha}(\mathbf{y}) + P_{N,\alpha}(\mathbf{y}) \ , \quad \text{by (105) and Ex. 69} \\ &= \mathbf{y} + P_{N,\alpha}(\mathbf{y}) \ , \quad \text{by (106) and Theorem 5(a)} \ , \end{aligned} \tag{110}$$

proving that

$$P_{N,\alpha}(\mathbf{y}) = \mathbf{0} \ , \tag{111}$$

which, by (108) and (75), contradicts (109). □

Ex. 71. (Newman and Odell [**1139**]). Let ϕ_p be the ℓ_p–norm, $1 < p < \infty$, on $\mathbb{C}^n$. The P_{L,ϕ_p} is linear for every subspace L if and only if $p = 2$.

Ex. 72. (Erdelsky [**466**]). *Essentially strictly convex norms.* Let α be an e.s.c. norm on $\mathbb{C}^n$, $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ and L a subspace of $\mathbb{C}^n$. Then

$$\mathbf{x} \in P_{L,\alpha}^{-1}(\mathbf{0})$$

if, and only if, there is a dual $\mathbf{y}$ of $\mathbf{x}$ with respect to α (i.e., a vector $\mathbf{y} \neq \mathbf{0}$ satisfying $\langle \mathbf{y}, \mathbf{x} \rangle = \alpha(\mathbf{x})\, \alpha_D(\mathbf{y})$), such that $\mathbf{y} \in L^\perp$.

Ex. 73. (Erdelsky [**466**]). If α and α_D are both e.s.c. norms on $\mathbb{C}^n$, L is a subspace of $\mathbb{C}^n$ for which $P_{L,\alpha}$ is linear, and $N = P_{L,\alpha}^{-1}(\mathbf{0})$, then

(a) $L^\perp = P_{N^\perp,\alpha_D}^{-1}(\mathbf{0})$,

(b) $P_{N^\perp,\alpha_D} = (P_{L,\alpha})^*$.

Proof. (a) Since both α and α_D are e.s.c., it follows from Exs. 61(a), 60, and 59 that every $\mathbf{0} \neq \mathbf{x}$ has a dual $\mathbf{0} \neq \mathbf{y}$ with respect to α, and $\mathbf{x}$ is a dual of $\mathbf{y}$. Now

$$\mathbf{y} \in P_{N^\perp,\alpha_D}^{-1}(\mathbf{0}) \Longleftrightarrow \mathbf{x} \in N^{\perp\perp} = N ,$$

by Ex. 72, which also show that

$$\mathbf{x} \in N \Longleftrightarrow \mathbf{y} \in L^\perp ,$$

proving (a).

(b) By (a) and Corollary 5(b), $P_{N^\perp,\alpha_D}$ is linear. Let $\mathbf{x}$ and $\mathbf{y}$ be arbitrary vectors, written as

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 , \quad \mathbf{x}_1 \in L ,\ \mathbf{x}_2 \in N , \quad \text{by (76)}$$

and

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 , \quad \mathbf{y}_1 \in N^\perp ,\ \mathbf{y}_2 \in L^\perp , \quad \text{by (a) and (76)} .$$

Then

$$\langle P_{L,\alpha}(\mathbf{x}), \mathbf{y}\rangle = \langle \mathbf{x}_1, \mathbf{y}_1\rangle = \langle \mathbf{x}, P_{N^\perp,\alpha_D}(\mathbf{y})\rangle .$$

□

Ex. 74. (Erdelsky [**466**]). *Dual norms.* Let α and α_D be dual norms on $\mathbb{C}^n$. Then:

(a) If α and α_D are both e.s.c., then $Q(\alpha) = Q(\alpha_D)$.

(b) If α is projective, then α_D is e.s.c.

(c) If α is projective, then so is α_D.

α–β Generalized Inverses

Ex. 75. (Erdelsky [**466**]). Let α and β be Let α and β be e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, and let $A \in \mathbb{C}^{m\times n}$.

If $B \in \mathbb{C}^{n\times m}$ satisfies

$$AB = P_{R(A),\alpha} , \tag{112}$$

$$BA = I - P_{N(A),\beta} , \tag{113}$$

$$\operatorname{rank} B = \operatorname{rank} A , \tag{114}$$

then

$$B = A_{\alpha,\beta}^{(-1)} .$$

Thus, if the α–β generalized inverse of A is linear, it can be defined by (112)–(114).

Ex. 76. (Erdelsky [**466**]). Let α and β be e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively. Then

$$(A_{\alpha,\beta}^{(-1)})_{\beta,\alpha}^{(-1)} = A , \quad \text{for all } A \in \mathbb{C}^{m\times n} \tag{115}$$

if and only if α and β are projective norms.

Proof. If α and β are projective, then $A_{\alpha,\beta}^{(-1)}$ is linear for any A, by Theorem 6(b) and Ex. 68. Let $\widehat{R} = R(A_{\alpha,\beta}^{(-1)})$ and $\widehat{N} = N(A_{\alpha,\beta}^{(-1)})$. Then by Exs. 67, 68, 72 and Theorem 6(c),(d),(e),(g), and (h),

$$\begin{aligned} A^{(-1)}_{\alpha,\beta}\, A &= I - P_{N(A),\beta} = P_{\widehat{R},\beta}\ , \\ A\, A^{(-1)}_{\alpha,\beta} &= P_{R(A),\alpha} = I - P_{\widehat{N},\alpha}\ , \\ \operatorname{rank} A^{(-1)}_{\alpha,\beta} &= \operatorname{rank} A\ , \end{aligned}$$

and (115) follows from Ex. 75.

Only if. If (115) holds for all $A \in \mathbb{C}^{m\times n}$ then

$$\begin{aligned} I - P_{N(A),\beta} = A^{(-1)}_{\alpha,\beta}\, A = A^{(-1)}_{\alpha,\beta} \left(A^{(-1)}_{\alpha,\beta}\right)^{(-1)}_{\beta,\alpha} = P_{\widehat{R},\beta}\ , \\ P_{R(A),\alpha} = A\, A^{(-1)}_{\alpha,\beta} = \left(A^{(-1)}_{\alpha,\beta}\right)^{(-1)}_{\beta,\alpha} A^{(-1)}_{\alpha,\beta} = I - P_{\widehat{N},\alpha}\ , \end{aligned}$$

and α and β are projective by Ex. 70. □

EX. 77. (Erdelsky [**466**]). If α and β are e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, then

$$(A^{(-1)}_{\alpha,\beta})^* = (A^*)^{(-1)}_{\beta_D,\alpha_D}\ , \quad \text{for all } A \in \mathbb{C}^{m\times n}\ . \tag{116}$$

PROOF. From Theorem 6(d) and (f) and Exs. 67, 68, and 69

$$\begin{aligned} A\, A^{(-1)}_{\alpha,\beta} &= P_{R(A),\alpha} = I - P_{N,\alpha}\ , & N &= P^{-1}_{R(A),\alpha}(\mathbf{0})\ , \\ A^{(-1)}_{\alpha,\beta}\, A &= I - P_{N(A),\beta} = P_{M,\beta}\ , & M &= P^{-1}_{N(A),\beta}(\mathbf{0})\ , \end{aligned}$$

and

$$\begin{aligned} R(A) &= P^{-1}_{N,\alpha}(\mathbf{0})\ , \\ N(A) &= P^{-1}_{M,\beta}(\mathbf{0})\ . \end{aligned}$$

Since α_D and β_D are e.s..c. norms, by Ex. 74(b), it follows from Ex. 73(b) that

$$\begin{aligned} A\, A^{(-1)}_{\alpha,\beta} &= I - \left(P_{R(A)^\perp,\alpha_D}\right)^* = I - \left(P_{N(A^*),\alpha_D}\right)^*\ , \\ A^{(-1)}_{\alpha,\beta}\, A &= \left(P_{N(A)^\perp,\beta_D}\right)^* = \left(P_{R(A^*),\beta_D}\right)^*\ , \end{aligned}$$

and hence

$$\begin{aligned} (A^{(-1)}_{\alpha,\beta})^*\, A^* &= I - P_{N(A^*),\alpha_D}\ , \\ A^*\,(A^{(-1)}_{\alpha,\beta})^* &= P_{R(A^*),\beta_D}\ , \end{aligned}$$

from which (116) follows by using Ex. 75. □

EX. 78. (Erdelsky [**466**]). If α and β are e.s.c. norms on $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, then for any $O \neq A \in \mathbb{C}^{m\times n}$

$$\begin{aligned} \frac{1}{\|A^{(-1)}_{\alpha,\beta}\|_{\beta,\alpha}} &\leq \inf\{\|X\|_{\alpha,\beta} :\ X \in \mathbb{C}^{m\times n}\ ,\ \operatorname{rank}(A+X) < \operatorname{rank} A\} \\ &\leq \frac{q}{\|A^{(-1)}_{\alpha,\beta}\|_{\beta,\alpha}}\ , \end{aligned} \tag{117}$$

where

$$q = 1 \quad \text{if } \operatorname{rank} A = m\ ,$$

and

$$q = Q(\alpha) \quad \text{otherwise}\ .$$

In particular, if α is projective,

$$\frac{1}{\|A^{(-1)}_{\alpha,\beta}\|_{\beta,\alpha}} = \inf\left\{\|X\|_{\alpha,\beta} : X \in \mathbb{C}^{m\times n}\ ,\ \operatorname{rank}(A+X) < \operatorname{rank} A\right\} . \tag{118}$$

A special case of (118) is given in Ex. 6.15 below.

5. An extremal property of the Bott–Duffin inverse with application to electrical networks

An important extremal property of the Bott–Duffin inverse, studied in Sections 2.9 and 2.12, is stated in the following theorem.

THEOREM 7. (Bott and Duffin [**202**]). Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, and let L be a subspace of $\mathbb{C}^n$ such that $A^{(-1)}_{(L)}$ exists[3]. Then for any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, the quadratic function

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{v})^* A(\mathbf{x}-\mathbf{v}) - \mathbf{w}^*\mathbf{x} \tag{119}$$

has a unique stationary value in L, when

$$\mathbf{x} = A^{(-1)}_{(L)}(A\mathbf{v}+\mathbf{w}) . \tag{120}$$

Conversely, if the Hermitian matrix A and the subspace L are such that for any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, the quadratic function (119) has a stationary value in L, then $A^{(-1)}_{(L)}$ exists and the stationary point is unique for any $\mathbf{v}$, $\mathbf{w}$ and given by (120).

PROOF. A *stationary point* of q in L is a point $\mathbf{x} \in L$ at which the gradient

$$\nabla q(\mathbf{x}) = \left[\frac{\partial}{\partial x_j} q(\mathbf{x})\right] \quad (j \in \overline{1,n})$$

is orthogonal to L, i.e., $\nabla q(\mathbf{x}) \in L^\perp$. The value of q at a stationary point is called a *stationary value* of q.

Differentiating (119) we see that the sought stationary point $\mathbf{x} \in L$ satisfies

$$\nabla q(\mathbf{x}) = A(\mathbf{x}-\mathbf{v}) - \mathbf{w} \in L^\perp ,$$

and by taking $\mathbf{y} = -\nabla q(\mathbf{x})$ we conclude that $\mathbf{x}$ is a stationary point of q in L if and only if $\mathbf{x}$ is a solution of

$$A\mathbf{x} + \mathbf{y} = A\mathbf{v} + \mathbf{w}\ , \quad \mathbf{x} \in L, \mathbf{y} \in L^\perp . \tag{121}$$

Thus the existence of a stationary value of q for any $\mathbf{v}$, $\mathbf{w}$ is equivalent to the consistency of (121) for any $\mathbf{v}$, $\mathbf{w}$, i.e., to the existence of $A^{(-1)}_{(L)}$, in whish case (120) is the unique stationary point in L. $\square$

COROLLARY 8. Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite and let L be a subspace of $\mathbb{C}^n$. Then for any $\mathbf{v}$, $\mathbf{w} \in \mathbb{C}^n$ the function

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{v})^* A(\mathbf{x}-\mathbf{v}) - \mathbf{w}^*\mathbf{x} \tag{119}$$

has a unique minimum in L, when

$$\mathbf{x} = A^{(-1)}_{(L)}(A\mathbf{v}+\mathbf{w}) . \tag{120}$$

PROOF. Follows from Theorem 7, since $A^{(-1)}_{(L)}$ exists, by Ex. 2.93, and the stationary value of q is actually a minimum since A is positive definite. $\square$

[3]See Ex. 2.78 for conditions equivalent to the existence of $A^{(-1)}_{(L)}$.

We return now to the direct current electrical network of Section 2.12, consisting of m nodes $\{n_i : i \in \overline{1,m}\}$ and n branches $\{b_j : j \in \overline{1,n}\}$, with

$a_j > 0$, the *conductance* of b_j,
$A = [\operatorname{diag} a_j]$, the *conductance matrix*,
x_j, the *voltage across* b_j,
y_j, the *current in* b_j,
v_j, the *voltage generated by the sources in series with* b_j,
w_j, the *current generated by the sources in parallel with* b_j, and
M, the (node–branch) *incidence matrix*.

We recall that the branch voltages $\mathbf{x}$ and currents $\mathbf{y}$ are uniquely determined by the following three physical laws:

$$A\mathbf{x} + \mathbf{y} = A\mathbf{v} + \mathbf{w} \qquad (\textit{Ohm's law}), \tag{122}$$

$$\mathbf{y} \in N(M) \qquad (\textit{Kirchhoff's current law}), \tag{123}$$

$$\mathbf{x} \in R(M^T) \qquad (\textit{Kirchhoff's voltage law}), \tag{124}$$

and that $\mathbf{x}$, $\mathbf{y}$ are related by

$$\mathbf{x} = A^{(-1)}_{(R(M^T))}(A\mathbf{v} + \mathbf{w}) \,, \tag{2.134}$$

$$\mathbf{y} = (I - AA^{(-1)}_{(R(M^T))})(A\mathbf{v} + \mathbf{w}) \,, \tag{2.135}$$

or dually, by (2.138) and (2.137).

A classical variational principle of Kelvin ([**1449**]) and Maxwell ([**1006**, pp. 903–908]), states that the voltages $\mathbf{x}$ and the currents $\mathbf{y}$ are such that the rate of energy dissipation is minimized. This variational principle is given in the following corollary.

Corollary 9. Let A, M, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{v}$, $\mathbf{w}$ be as above. Then

(a) The vector $\mathbf{x}_0$ of branch voltages is the unique minimizer of

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{v})^* A(\mathbf{x} - \mathbf{v}) - \mathbf{w}^*\mathbf{x} \tag{119}$$

in $R(M^T)$, and the vector $\mathbf{y}_0$ of branch currents is

$$\mathbf{y}_0 = -\nabla q(\mathbf{x}_0) = -A(\mathbf{x}_0 - \mathbf{v}) + \mathbf{w} \in R(M^T)^\perp = N(M) \,. \tag{125}$$

(b) The vector $\mathbf{y}_0$ is the unique minimizer of

$$p(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{w})^* A^{-1}(\mathbf{y} - \mathbf{w}) - \mathbf{v}^*\mathbf{y} \tag{126}$$

in $N(M)$, and the vector $\mathbf{x}_0$ is

$$\mathbf{x}_0 = -\nabla p(\mathbf{y}_0) = -A^{-1}(\mathbf{y}_0 - \mathbf{w}) + \mathbf{v} \in N(M)^\perp = R(M^T) \,. \tag{127}$$

Proof. Since the conductance matrix A is positive definite, it follows by comparing (120) and (2.134) that $\mathbf{x}_0$ is the unique minimizer of (119) in $R(M^T)$, and the argument used in the proof of Theorem 7 shows that $\mathbf{y}_0 = -\nabla q(\mathbf{x}_0)$ as given in (125). Part (b) follows from the dual derivation (2.137) and (2.138) of $\mathbf{y}_0$ and $\mathbf{x}_0$, respectively, as solutions of the dual network equations (2.136). □

Corollary 9 shows that the voltage $\mathbf{x}$ is uniquely determined by the function (119) to be minimized subject to Kirchhoff's voltage law (124). Kirchhoff's current law (123) and Ohm's law (122) are then consequences of (125).

Dually, the current $\mathbf{y}$ is uniquely determined by the function (126) to be minimized subject to Kirchhoff's current law (123), and the other two laws (122) and (124) then follow from (127).

Further references on the extremal properties of the network functions and solutions are Dennis [**394**], Stern [**1397**] and [**1398**], and Guillemin [**629**]. Corollary 9 is a special case of the Duality Theory of Convex Programming; see, e.g., Rockafellar [**1295**].

Exercises.

EX. 79. Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite, and let the subspace $L \subset \mathbb{C}^n$ and the vector $\mathbf{w} \in \mathbb{C}^n$ be given. Then the quadratic function

$$\frac{1}{2}\,\mathbf{x}^* A\mathbf{x} - \mathbf{w}^*\mathbf{x} \tag{128}$$

has a minimum in L if and only if the system

$$A\mathbf{x} - \mathbf{w} \in L^{\perp}\,, \quad \mathbf{x} \in L \tag{129}$$

is consistent, in which case the solutions $\mathbf{x}$ of (129) are the minimizers of (128) in L.

EX. 80. Show that the consistency of (129) is equivalent to the condition

$$\mathbf{x} \in L,\, A\mathbf{x} = \mathbf{0} \implies \mathbf{w}^*\mathbf{x} = 0\,,$$

which is obviously equivalent to the boundedness from below of (128) in L, hence to the existence of a minimizer in L.

EX. 81. Show that $A_{(L)}^{(-1)}$ exists if and only if the system (129) has a unique solution for any $\mathbf{w} \in \mathbb{C}^n$, in which case this solution is

$$\mathbf{x} = A_{(L)}^{(-1)}\,\mathbf{w}\,.$$

EX. 82. Give the general solution of (129) in case it is consistent but $A_{(L)}^{(-1)}$ does not exist.

Suggested further reading

Section 1. Desoer and Whalen [**396**], Erdélyi and Ben-Israel [**477**], Leringe and Wedin [**930**], Osborne [**1154**], Peters and Wilkinson [**1182**], and the references on applications to statistics given at the end of the Introduction.

For various applications in control theory and in system theory, see Balakrishnan [**62**], Barnett [**87**], Ho and Kalman [**735**], Kalman ([**807**], [**808**], [**809**], [**810**]), Kalman, Ho and Narendra [**812**], Kishi [**859**], Kuo and Mazda [**895**], Minamide and Nakamura ([**1054**], [**1055**]), Porter ([**1197**], [**1198**]), Porter and Williams ([**1200**], [**1199**]), Wahba and Nashed [**1512**], and Zadeh and Desoer [**1625**].

Section 2. Erdélyi and Ben-Israel [**477**], Osborne [**1155**], Rosen [**1305**].

CHAPTER 4

Spectral Generalized Inverses

1. Introduction

In this chapter we shall study generalized inverses having some of the spectral properties (i.e., properties relating to eigenvalues and eigenvectors) of the inverse of a nonsingular matrix. Only square matrices are considered, since only they have eigenvalues and eigenvectors.

The four Penrose equations of Chapter 1,

$$\begin{aligned} AXA &= A\,, & (1)\\ XAX &= X\,, & (2)\\ (AX)^* &= AX\,, & (3)\\ (XA)^* &= XA\,, & (4) \end{aligned}$$

will now be supplemented further by the following equations applicable only to square matrices

$$A^kXA = A^k\,, \tag{1^k}$$

$$AX = XA\,, \tag{5}$$

$$A^kX = XA^k\,, \tag{5^k}$$

$$AX^k = X^kA\,, \tag{6^k}$$

In these equations k is a given positive integer. For example, we shall have occasion to refer to a $\{1^k, 2, 5\}$*–inverse* of A.

2. Spectral properties of a nonsingular matrix

If A is nonsingular it is easy to see that every eigenvector of A associated with the eigenvalue λ is also an eigenvector of A^{-1} associated with the eigenvalue λ^{-1}. (A nonsingular matrix does not have 0 as an eigenvalue.)

A matrix $A \in \mathbb{C}^{n\times n}$ that is not diagonable does not have n linearly independent eigenvectors (see Ex. 2.22). However, it does have n linearly independent principal vectors. Following Wilkinson [**1595**], we define a *principal vector* of A of *grade* p associated with the eigenvalue λ as a vector $\mathbf{x}$ such that

$$(A-\lambda I)^p\mathbf{x} = \mathbf{0}\,, \quad (A-\lambda I)^{p-1}\mathbf{x} \neq \mathbf{0}\,. \tag{7}$$

Here p is some positive integer.

Evidently principal vectors are a generalization of eigenvectors. In fact, an eigenvector is a principal vector of grade 1. We shall find it convenient to abbreviate "principal vector of grade p associated with the eigenvalue λ" to "λ–vector of A of grade p".

It is not difficult to show (see Ex. 3) that, if A is nonsingular, a vector $\mathbf{x}$ is a λ^{-1}–vector of A^{-1} of grade p if and only if it is a λ–vector of A of grade p. In the remainder of this chapter, we shall explore the extent to which singular matrices have generalized inverses with comparable spectral properties.

Exercises.

Ex. 1. A square matrix A is diagonable if and only if all its principal vectors are eigenvectors.

Ex. 2. For a given eigenvalue λ, the maximal grade of the λ–vectors of A is the multiplicity of λ as root of the minimal polynomial of A.

Ex. 3. If A is nonsingular, $\mathbf{x}$ is a λ^{-1}–vector of A^{-1} of grade p if and only if it is a λ–vector of A of grade p. [*Hint*: Show that $A^{-p}(A-\lambda I)^p = (-\lambda)^p(A^{-1}-\lambda^{-1}I)^p$. Using this and the analogous relation obtained by replacing A by A^{-1}, show that $(A-\lambda I)^r\mathbf{x} = \mathbf{0}$ if and only if $(A^{-1}-\lambda^{-1}I)^r\mathbf{x} = \mathbf{0}$ for $r = 0, 1, \dots$]

Ex. 4. If A is nonsingular and diagonable, A^{-1} is the *only* matrix related to A by the property stated in Ex. 3.

Ex. 5. If A is nonsingular and not diagonable, there are matrices other than A^{-1} having the spectral relationship to A described in Ex. 3. For example, consider

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} , \quad X = \begin{bmatrix} \lambda^{-1} & c \\ 0 & \lambda^{-1} \end{bmatrix} \quad (\lambda, c \neq 0) .$$

Show that for $p = 1, 2$, $\mathbf{x}$ is a λ^{-1}–vector of X of grade p if and only if it is a λ–vector of A of grade p. (Note that $X = A^{-1}$ for $c = -\lambda^{-2}$.)

3. Spectral inverse of a diagonable matrix

In investigating the existence of generalized inverses of a singular square matrix, we shall begin with diagonable matrices, because they are the easiest to deal with. Evidently some extension must be made of the spectral property enjoyed by nonsingular matrices, because a singular matrix has 0 as one of its eigenvalues. Given a diagonable matrix $A \in \mathbb{C}^{n\times n}$, let us seek a matrix X such that every eigenvector of A associated with the eigenvalue λ (for every λ in the spectrum of A) is also an eigenvector of X associated with the eigenvalue $\lambda^\dagger$, where $\lambda^\dagger$ is defined in (1.8).

Since A has n linearly independent eigenvectors, there is a nonsingular matrix P, having such a set of eigenvectors as columns, such that

$$AP = PJ \tag{8}$$

where

$$J = \mathrm{diag}(\lambda_1, \lambda_2, \dots , \lambda_n)$$

is a Jordan form of A. We shall need the diagonal matrix obtained from J by replacing each diagonal element λ_i by $\lambda_i^\dagger$. By Ex. 1.20, this is, in fact, the Moore–Penrose inverse of J; that is,

$$J^\dagger = \mathrm{diag}(\lambda_1^\dagger, \lambda_2^\dagger, \dots , \lambda_n^\dagger) .$$

Because of the spectral requirement imposed on X, we must have

$$XP = PJ^\dagger . \tag{9}$$

Solving (8) and (9) for A and X givs

$$X = PJP^{-1} , \quad X = PJ^\dagger P^{-1} . \tag{10}$$

Since J and $J^\dagger$ are both diagonal, they commute with each other. As a result, it follows from (10) that $X \in A\{1, 2, 5\}$.

We do not wish to limit our consideration to diagonable matrices. We began with them because they are easier to work with. The result just obtained suggests that we should examine the existence and properties (especially spectral properties) of $\{1, 2, 5\}$–inverses for square matrices in general.

4. The group inverse

It follows from (5) and from Corollary 2.7 that a $\{1,2,5\}$–inverse of A, if it exists, is a $\{1,2\}$–inverse X such that $R(X) = R(A)$ and $N(X) = N(A)$. By Theorem 2.10, there is at most one such inverse.

This unique $\{1,2,5\}$–inverse is called the *group inverse* of A, and is denoted by $A^{\#}$. The name “group inverse” was given by I. Erdélyi [**469**], because the positive and negative powers of a given matrix A (the latter being interpreted as powers of $A^{\#}$), together with the projector $AA^{\#}$ as the unit element, constitute an Abelian group, see Ex. 13. Both he and Englefield [**465**] (who called it the “commuting reciprocal inverse”) drew attention to the spectral properties of the group inverse. As we shall see later, however, the group inverse is a particular case of the *Drazin inverse* [**423**], or $\{1^k,2,5\}$–inverse, which predates [**469**] and [**465**].

The group inverse is not restricted to diagonable matrices; however, it does not exist for all square matrices. By Section 2.5 and Theorem 2.10, such an inverse exists if and only if $R(A)$ and $N(A)$ are complementary subspaces. We show in Theorem 1 that this is equivalent to

$$R(A) = R(A^2)\ .$$

In this connection, the following definition is useful.

Definition 1. The smallest positive integer k for which

$$\operatorname{rank} A^k = \operatorname{rank} A^{k+1}\ , \tag{11}$$

holds, is called the *index*[1] of A.

The index will be studied in Section 6 below. For now we state.

Theorem 1. A square matrix A has a group inverse if and only if its index is 1, or, in other words, if and only if

$$\operatorname{rank} A = \operatorname{rank} A^2\ . \tag{12}$$

When the group inverse exists, it is unique.

Proof. Let $A \in \mathbb{C}^{n\times n}$. If A is nonsingular, $R(A) = \mathbb{C}^n$ and $N(A) = \{\mathbf{0}\}$. Thus $R(A)$ and $N(A)$ are trivially complementary. Since a nonsingular matrix has index 1, it remains to prove the theorem for singular matrices. Now, for any positive integer k,

$$\dim R(A^k) + \dim N(A^k) = \operatorname{rank} A^k + \operatorname{nullity} A^k = n\ .$$

It therefore follows from statement (c) of Ex. 0.1 that $R(A^k)$ and $N(A^k)$ are complementary if and only if

$$R(A^k) \cap N(A^k) = \{\mathbf{0}\}\ . \tag{13}$$

Since, for any positive integer k,

$$R(A^{k+1}) \subset R(A^k)\ ,$$

and

$$N(A^k) \subset N(A^{k+1})\ ,$$

it follows that (13) is equivalent to

$$\dim R(A^k) = \dim R(A^{k+1})\ . \tag{14}$$

The statement of the theorem is the special case $k = 1$. $\square$

[1]Some writers (e.g., MacDuffee [**986**]) define the index as the degree of the minimal polynomial.

An alternative proof of uniqueness is as follows. Let $X, Y \in A\{1,2,5\}$, $E = AX = XA$, and $F = AY = YA$. Then $E = F$ since

$$E = AX = AYAX = FE \,,$$
$$F = YA = YAXA = FE \,.$$

Therefore,

$$X = EX = FX = YE = YF = Y \,.$$

The following theorem gives an equivalent condition for the existence of $A^{\#}$ that is often more convenient in numerical work, and also an explicit formula for $A^{\#}$.

Theorem 2. (Cline [**352**]). Let a square matrix A have the full–rank factorization

$$A = FG \,. \tag{15}$$

Then A has a group inverse if and only if GF is nonsingular, in which case

$$A^{\#} = F(GF)^{-2}G \,. \tag{16}$$

Proof. Let $r = \operatorname{rank} A$. Then $GF \in \mathbb{C}^{r\times r}$. Now

$$A^2 = FGFG \,,$$

and so

$$\operatorname{rank} A^2 = \operatorname{rank} GF$$

by Ex. 1.7. Therefore (12) holds if and only if GF is nonsingular, and the first part of the theorem is established. It is easily verified that (1), (2), and (5) hold with A given by (15) and X by RHS(16). Formula (16) then follows from the uniqueness of the group inverse. □

For an important class of matrices, the group inverse and the Moore–Penrose inverse are the same. We shall call a square matrix A *range–Hermitian* (such a matrix is also called an *EP$_r$* or *EP matrix*, e.g., Schwerdtfeger [**1326**], Pearl [**1168**] and other writers) if

$$R(A^*) = R(A) \,, \tag{17}$$

or, equivalently, if

$$N(A^*) = N(A) \,, \tag{18}$$

the equivalence follows from (2.47).

Using the notation of Theorem 2.10, the preceding discussion shows that

$$A^{\#} = A^{(1,2)}_{R(A),N(A)} \,,$$

while Ex. 2.29 establishes that

$$A^{\dagger} = A^{(1,2)}_{R(A^*),N(A^*)} \,.$$

The two inverses are equal, therefore, if and only if $R(A) = R(A^*)$ and $N(A) = N(A^*)$. But this is true if and only if A is range–Hermitian. Thus we have proved:

Theorem 3. $A^{\#} = A^{\dagger}$ if and only if A is range–Hermitian. □

The approach of (10) can be extended from diagonable matrices to all square matrices of index 1. To do this we shall need the following lemma.

Lemma 1. Let J be a square matrix in Jordan form. Then J is range–Hermitian if and only if it has index 1.

PROOF. *Only if*: Follows from Ex. 7.

If: If J is nonsingular, $\operatorname{rank} J = \operatorname{rank} J^2$ and J is range–Hermitian by Ex. 15. If J has only 0 as an eigenvalue, it is nilpotent. In this case, it follows easily from the structure of the Jordan form that $\operatorname{rank} J^2 < \operatorname{rank} J$ unless J is the null matrix O, in which case it is trivially range–Hermitian.

If J has both zero and nonzero eigenvalues, it can be partitioned in the form

$$J = \begin{bmatrix} J_1 & O \\ O & J_2 \end{bmatrix} ,$$

where J_1 is nonsingular and has as eigenvalues the nonzero eigenvalues of J, while J_2 is nilpotent. By the same reasoning employed in the preceding paragraph, $\operatorname{rank} J = \operatorname{rank} J^2$ implies $J_2 = O$. It then follows from Ex. 15 that J is range–Hermitian. □

THEOREM 4. (Erdélyi [**469**]). Let A have index 1 and let

$$A = PJP^{-1} ,$$

where P is nonsingular and J is a Jordan normal form of A. Then

$$A^{\#} = PJ^{\dagger}P^{-1} . \tag{19}$$

PROOF. It is easily verified that relations (1), (2), (5), and (12) are similarity invariants. Therefore

$$J^{\#} = P^{-1}A^{\#}P \tag{20}$$

and also $\operatorname{rank} J = \operatorname{rank} J^2$. It then follows from Lemma 1 and Theorem 3 that

$$J^{\#} = J^{\dagger} , \tag{21}$$

and (19) follows from (20) and (21). □

Exercises.

Ex. 6. Let $A \in \mathbb{C}^{n\times n}$. If for some positive integer k,

$$R(A^{k+1}) = R(A^k) , \tag{22}$$

then, for all integers $\ell > k$,

$$R(A^{\ell+1}) = R(A^{\ell}) .$$

[*Hint*: $R(A^{k+1}) = AR(A^k)$ and $R(A^{\ell}) = A^{\ell-k}R(A^k)$.]

Ex. 7. Every range–Hermitian matrix has index 1.

PROOF. If A is range–Hermitian, then by (2.47), $N(A) = R(A)^{\perp}$. Thus $R(A)$ and $N(A)$ are complementary subspaces. □

Ex. 8. If A is nonsingular, $A^{\#} = A^{-1}$.

Ex. 9. $A^{\#\#} = A$.

Ex. 10. $A^{*\#} = A^{\#*}$.

Ex. 11. $A^{T\#} = A^{\#T}$.

Ex. 12. $(A^{\ell})^{\#} = (A^{\#})^{\ell}$ for every positive integer ℓ.

Ex. 13. Let A have index 1 and denote $(A^{\#})^j$ by A^{-j} for $j = 1, 2, \dots$ Also denote $AA^{\#}$ by A^0. Then show that

$$A^{\ell}A^m = A^{\ell+m}$$

for all integers ℓ and m. (Thus, the "powers" of A, positive, negative and zero, constitute an Abelian group under matrix multiplication.)

Ex. 14. Show that

$$A^{\#} = A(A^3)^{(1)}A\,, \tag{23}$$

where $(A^3)^{(1)}$ is an arbitrary element of $A^3\{1\}$.

Ex. 15. Show that a nonsingular matrix is range–Hermitian.

Ex. 16. Show that a normal matrix is range–Hermitian. [*Hint*: Use Corollary 1.2.]

Remark. It follows from Exs. 7 and 16 that

$$\begin{aligned}\{\text{Hermitian matrices}\} &\subset \{\text{normal matrices}\}\\ &\subset \{\text{range–Hermitian matrices}\} \subset \{\text{matrices of index 1}\}\,.\end{aligned}$$

Ex. 17. A square matrix A is range–Hermitian if and only if A commutes with $A^\dagger$.

Ex. 18. (Katz [**824**]). A square matrix A is range–Hermitian if and only if there is a matrix Y such that $A^* = YA$.

Ex. 19. (Katz and Pearl [**826**]). A matrix in $\mathbb{C}^{n\times n}$ is range–Hermitian if and only if it is similar to a matrix of the form

$$\begin{bmatrix} A & O \\ O & O \end{bmatrix}\,,$$

where A is nonsingular.

Proof. See Lemma 1. □

5. Spectral properties of the group inverse

Even when A is not diagonable, the group inverse has spectral properties comparable to those of the inverse of a nonsingular matrix. However, in this case, $A^{\#}$ is not the only matrix having such properties. This has already been illustrated in the case of a nonsingular matrix (see Ex. 5).

We note that if a square matrix A has index 1, its 0–vectors are all of grade 1, i.e., null vectors of A. This follows from the fact that (12) implies $N(A^2) = N(A)$ by Ex. 1.10.

The following two lemmas are needed in order to establish the spectral properties of the group inverse. The second is stated in greater generality than is required for the immediate purpose because it will be used in connection with spectral generalized inverses other than the group inverse.

Lemma 2. Let $\mathbf{x}$ be a λ–vector of A with $\lambda \neq 0$. Then $\mathbf{x} \in R(A^\ell)$ where ℓ is an arbitrary positive integer.

Proof. We have

$$(A - \lambda I)^p \mathbf{x} = \mathbf{0}$$

for some positive integer p. Expanding the left member by the binomial theorem, transposing the last term, and dividing by its coefficient $(-\lambda)^{p-1} \neq 0$ gives

$$\mathbf{x} = c_1 A\mathbf{x} + c_2 A^2\mathbf{x} + \cdots + c_p A^p\mathbf{x}\,, \tag{24}$$

where

$$c_i = (-1)^{i-1}\lambda^{-i}\binom{p}{i}\,.$$

Successive multiplication of (24) by A gives

$$\begin{aligned} A\mathbf{x} &= c_1 A^2\mathbf{x} + c_2 A^3\mathbf{x} + \cdots + c_p A^{p+1}\mathbf{x}\,,\\ A^2\mathbf{x} &= c_1 A^3\mathbf{x} + c_4 A^4\mathbf{x} + \cdots + c_p A^{p+2}\mathbf{x}\,,\\ \cdots &= \cdots\cdots\cdots\cdots \\ A^{\ell-1}\mathbf{x} &= c_1 A^{\ell}\mathbf{x} + c_4 A^{\ell+1}\mathbf{x} + \cdots + c_p A^{p+\ell-1}\mathbf{x}\,, \end{aligned} \tag{25}$$

Successive substitution of equations (25) in RHS(24) gives eventually

$$\mathbf{x} = A^{\ell} q(A)\mathbf{x} ,$$

where q is some polynomial. $\square$

Lemma 3. Let A be a square matrix and let

$$XA^{\ell+1} = A^{\ell} \tag{26}$$

for some positive integer ℓ. Then every λ–vector of A of grade p for $\lambda \neq 0$ is a λ^{-1}–vector of X of grade p.

Proof. The proof will be by induction on the grade p. Let $\lambda \neq 0$ and $A\mathbf{x} = \lambda\mathbf{x}$. Then $A^{\ell+1}\mathbf{x} = \lambda^{\ell+1}\mathbf{x}$, and therefore $\mathbf{x} = \lambda^{-\ell-1}A^{\ell+1}\mathbf{x}$. Accordingly,

$$X\mathbf{x} = \lambda^{-\ell-1}XA^{\ell+1}\mathbf{x} = \lambda^{-1}\mathbf{x} .$$

proving the lemma for $p = 1$.

Suppose the lemma is true for $p = 1, 2, \dots , r$, and let $\mathbf{x}$ be a λ–vector of A of grade $r+1$. Then, by Lemma 2,

$$\mathbf{x} = A^{\ell}\mathbf{y}$$

for some $\mathbf{y}$. Thus

$$\begin{aligned}(X - \lambda^{-1}I)\,\mathbf{x} &= (X - \lambda^{-1}I)A^{\ell}\,\mathbf{y} = X(A^{\ell} - \lambda^{-1}A^{\ell+1})\,\mathbf{y}\\ &= X(I - \lambda^{-1}A)A^{\ell}\,\mathbf{y} = -\lambda^{-1}X(A - \lambda I)\,\mathbf{x} .\end{aligned}$$

By the induction hypothesis, $(A - \lambda I)\,\mathbf{x}$ is a λ^{-1}–vector of X of grade r. Consequently

$$\begin{aligned}(X - \lambda^{-1}I)^{r}(A - \lambda I)\,\mathbf{x} &= \mathbf{0} ,\\ \mathbf{z} = (X - \lambda^{-1}I)^{r-1}(A - \lambda I)\,\mathbf{x} &\neq \mathbf{0}\\ X\mathbf{z} &= \lambda^{-1}\mathbf{z} .\end{aligned}$$

Therefore

$$\begin{aligned}(X - \lambda^{-1}I)^{r+1}\,\mathbf{x} &= -\lambda^{-1}X(X - \lambda^{-1}I)^{r}(A - \lambda I)\,\mathbf{x} = \mathbf{0} ,\\ (X - \lambda^{-1}I)^{r}\,\mathbf{x} &= -\lambda^{-1}X\,\mathbf{z} = -\lambda^{-2}\mathbf{z} \neq \mathbf{0} .\end{aligned}$$

This completes the induction. $\square$

The following theorem shows that for every matrix A of index 1, the group inverse is the only matrix in $A\{1\}$ or $A\{2\}$ having spectral properties comparable to those of the inverse of a nonsingular matrix. For convenience, we introduce:

Definition 2. X is an *S–inverse* of A (or A and X *S–inverses* of each other) if they share the property that, for every $\lambda \in \mathbb{C}$ and every vector $\mathbf{x}$, $\mathbf{x}$ is a λ–vector of A of grade p if and only if it is a $\lambda^{\dagger}$–vector of X of grade p.

Theorem 5. Let $A \in \mathbb{C}^{n\times n}$ have index 1. Then $A^{\#}$ is the unique S–inverse of A in $A\{1\} \cup A\{2\}$. If A is diagonable, $A^{\#}$ is the only S–inverse of A.

Proof. First we shall show that $A^{\#}$ is an S–inverse of A. Since $X = A^{\#}$ satisfies (26) with $\ell = 1$, it follows from Lemma 3 that $A^{\#}$ satisfies the "if" part of the definition of S–inverse for $\lambda \neq 0$. Replacing A by $A^{\#}$ establishes the "only if" part for $\lambda \neq 0$, since $A^{\#\#} = A$ (see Ex. 9).

Since both A and $A^{\#}$ have index 1, all their 0–vectors are null vectors as pointed out in the second paragraph of this session. Thus, in order to prove that $A^{\#}$ satisfies the definition of S–inverse for $\lambda = 0$, we need only show that $N(A) = N(A^{\#})$. But this follows from the commutativity of A and $A^{\#}$) and Ex. 1.10.

Let $r = \operatorname{rank} A$ and consider the equation

$$AP = PJ$$

where P is nonsingular and J is a Jordan form of A. The columns of P are λ–vectors of A. Since A has index 1, those columns which are not null vectors are associated with nonzero eigenvalues, and are therefore in $R(A)$ by Lemma 2. Since there are r of them and they are linearly independent, they span $R(A)$. But, by hypothesis, these columns are also λ^{-1}–vectors of X and therefore in $R(X)$. Since $\operatorname{rank} X = r$, these r vectors span $R(X)$, and so $R(X) = R(A)$. Thus X is a $\{1,2\}$–inverse of A such that $R(X) = R(A)$ and $N(X) = N(A)$. But $A^{\#}$ is the only such matrix, and so $X = A^{\#}$.

It was shown in Section 3 that if A is diagonable, an S–inverse of A must be a $\{1,2,5\}$–inverse. Since $A^{\#}$ is the only such inverse, this completes the proof. □

6. The Drazin inverse

We have seen that the group inverse does not exist for all square matrices, but only those of index 1. However, we shall show in this section that every square matrix has a unique $\{1^k, 2, 5\}$–inverse, where k is its index. This inverse is called the *Drazin inverse*, because it was first studied by Drazin [**423**] (though in the more general context of rings and semigroups without specific reference to matrices). The spectral properties of the Drazin inverse of a square matrix have been studied by Cline [**352**] and Greville [**583**]; not all of them will be mentioned here.

It is readily seen that the set of three equations (1^k), (2) and (5) is equivalent to the set

$$AX = XA\ , \tag{5}$$

$$A^{k+1}X = A^k\ , \tag{27}$$

$$AX^2 = X\ . \tag{28}$$

It is evident also that if (27) holds for some positive integer k, then it holds for every integer $\ell > k$. It follows also from (27) that

$$\operatorname{rank} A^k = \operatorname{rank} A^{k+1}\ . \tag{11}$$

Therefore, a solution X for (27) (and, consequently, of the set (5), (27), (28)) exists only if (11) holds. We shall show presently that if (11) holds, there is a u nique X (the Drazin inverse of A) satisfying (5), (27), and (28).

The next lemma collects properties of the *matrix index* (see Definition 1) that are needed below.

Lemma 4. Let $A \in \mathbb{C}^{n\times n}$ have index k. Then:
(a) All matrices $\{A^\ell : \ell \geq k\}$ have the same rank, the same range and the same null space.
(b) Their transposes $\{(A^\ell)^T : \ell \geq k\}$ all have the same rank, the same range and the same null space.
(c) Their conjugate transposes $\{(A^\ell)^* : \ell \geq k\}$ all have the same rank, the same range and the same null space.
(d) Moreover, for no ℓ less than k do A^ℓ and a higher power of A (or their transposes or conjugate transposes) have the same range or the same null space.

Proof. It may be well to point out first that (11) necessarily holds for *some* positive integer k (see Ex. 20).
(a) It follows from (11) and Ex. 1.10 that

$$R(A^{k+1}) = R(A^k)\ . \tag{22}$$

Therefore (27) holds for some X, and multiplication on the left by $A^{\ell-k}$ gives

$$A^\ell = A^{\ell+1}X \quad (\ell \geq k)\ . \tag{29}$$

It follows from (29) that all the matrices $\{A^\ell : \ell \geq k\}$ have the same range and the same rank. From Ex. 1.10 and the fact that A^k and A^ℓ have the same rank, it follows that they have the same

null space. (See Ex. 6 for an alternative proof of $R(A^\ell) = R(A^{\ell+1}$ for all $\ell \geq k$.).
(b) and (c). The statements about the transposes and conjugate transposes are obtained by applying (a) to A^T and A^* and noting that $(A^\ell)^T = (A^T)^\ell$ and $(A^\ell)^* = (A^*)^\ell$.
(d) If an equality of ranges of the kind ruled out in part (d) should occur, there must be some $\ell < k$ such that A^ℓ or its transpose or conjugate transpose have the same range as the corresponding matrix with exponent $\ell + 1$. But this would imply $\operatorname{rank} A^\ell = \operatorname{rank} A^{\ell+1}$, and k would not be the index of A. Similarly, equality of null spaces would imply that A^ℓ and $A^{\ell+1}$ have the same nullity, and therefore the same rank. □

Theorem 6. Let $A \in \mathbb{C}^{n\times n}$. Then the following statements are equivalent:
(a) The index of A is k.
(b) The smallest positive exponent for which (27) holds is k.
(c) If A is singular and $m(\lambda)$ is its minimal polynomial, k is the multiplicity of $\lambda = 0$ as a zero of $m(\lambda)$.
(d) If A is singular, k is the maximal grade of 0–vectors of A.

Proof. (a) $\Longleftrightarrow$ (b). Clearly (29) implies

$$\operatorname{rank} A^{\ell+1} = \operatorname{rank} A^\ell \ , \tag{30}$$

and by Ex. 1.10, (30) implies

$$R(A^{\ell+1}) = R(A^\ell) \ ,$$

so that (29) holds. Thus (30) and (29) are equivalent, proving (a)
(b) $\Longleftrightarrow$ (c). Let

$$m(\lambda) = \lambda^\ell p(\lambda)$$

where $p(0) \neq 0$. Let k be defined by (b), and we must now show that $k = \ell$. We have

$$p(A)A^\ell = O \ .$$

If $\ell > k$, then

$$O = p(A)A^\ell X = p(A)A^{\ell-1} \ ,$$

where $\lambda^{\ell-1}p(\lambda)$ is of lower degree than $m(\lambda)$, contrary to the definition of the minimal polynomial.
Since $p(0) \neq 0$, we can write[2]

$$m(\lambda) = c\lambda^\ell(1 - \lambda q(\lambda)) \ , \tag{31}$$

where $c \neq 0$ and q is a polynomial. It follows that

$$A^{\ell+1}q(A) = A^\ell \ . \tag{32}$$

If $\ell < k$, this would contradict (b).
(a) $\Longleftrightarrow$ (d). Let A have index k and the h be the maximal grade of the 0–vectors of A. We must whow that $h - k$. The definition of h implies that $N(A^\ell) = N(A^h)$ for all $\ell \geq h$, but $N(A^{h-1})$ is a proper subsace of $A(A^h)$. It follows from Lemma 4 that $h = k$. □

The following lemma will be used in proving the existence of a unique $\{1^k, 2, , 5\}$–inverse of a square matrix of index k.

Lemma 5. If Y is a $\{1^\ell, 5\}$–inverse of s quare matrix A, then

$$X = A^\ell Y^{\ell+1}$$

is a $\{1^\ell, 2, 5\}$–inverse.

---

[2]For this device we are indebted to M. R. Hestenes (see [**126**, p. 687, footnote 56]).

PROOF. We have

$$A^{\ell+1}Y = A^{\ell}, \quad AY = YA .$$

Clearly X satisfies (5). We have then

$$A^{\ell}XA = A^{2\ell+1}Y^{\ell+1} = A^{2\ell}Y^{\ell} = A^{2\ell-1}Y^{\ell-1} = \cdots = A^{\ell} ,$$

and

$$XAX = A^{2\ell+1}Y^{2\ell+2} = A^{2\ell}Y^{2\ell+1} = \cdots = A^{\ell}Y^{\ell+1} = X .$$

□

THEOREM 7. Let $A \in \mathbb{C}^{n\times n}$ have index k. Then A has a unique $\{1^k, 2, 5\}$–inverse, which is expressible as a polynomial in A, and is also the unique $\{1^{\ell}, 2, 5\}$–inverse for every $\ell \geq k$.

PROOF. The matrix $q(A)$ of (32) is a $\{1^k, 5\}$–inverse of A. Therefore, by Lemma 5,

$$X = A^k (q(A))^{k+1} \tag{33}$$

is a $\{1^k, 2, 5\}$–inverse. This proves the existence of such an inverse.

A matrix X that satisfies (27) clearly satisfies (29) for all $\ell \geq k$. Therefore, a $\{1^k, 2, , 5\}$–inverse of A is a $\{1^{\ell}, 2, 5\}$–inverse for all $\ell \geq k$.

Uniqueness will be proved by adapting the proof of uniqueness of the group inverse given in th remark following Theorem 1. Let $X, Y \in A\{1^{\ell}, 2, 5\}$, $E = AX = XA$, and $F = AY = YA$. Note that E and F are idempotent. Then $E = F$, since

$$E = AX = A^{\ell}X^{\ell} = AYA^{\ell}X^{\ell} = FAX = FE ,$$
$$F = YA = Y^{\ell}A^{\ell} = Y^{\ell}A^{\ell}XA = YAE = FE .$$

The proof is then completed exactly as in the case of the group inverse. □

This unique $\{1^k, 2, 5\}$–inverse is the *Drazin inverse*, and we shall denote it by A^D. The group inverse is the particular case of the Drazin inverse for matrices of index 1.

COROLLARY 1. (Englefield [**465**]). Let $A \in \mathbb{C}^{n\times n}$. Then there is a $\{1, 2\}$–inverse of A expressible as a polynomial in A if and only if A has index 1, in which case the only such inverse is the group inverse, which is given by

$$A^{\#} = A(q(A))^2 , \tag{34}$$

where q is defined by (31).

PROOF. *Only if*: A $\{1, 2\}$–inverse of A that is a polynomial in A necessarily commutes with A, and is therefore a $\{1, 2, 5\}$–inverse. The group inverse $A^{\#}$ is the only such inverse, and A has a group inverse if and only if its index is 1.

If: If A has index 1, it has a group inverse, which is a $\{1, 2\}$–inverse, and in this case coincides with the Drazin inverse. It is therefore expressible as a polynomial in A by Theorem 7. Formula (34) is merely the specia;ization of (33) for $k = 1$. □

COROLLARY 2. (Pearl [**1170**]). Let $A \in \mathbb{C}^{n\times n}$. Then $A^{\dagger}$ is expressible as polynomial in A if and only if A is range–Hermitian.

Exercises.

Ex. 20. Let $A \in \mathbb{C}^{n\times n}$. Show that (11) holds for some k between 1 and n, inclusive.

PROOF. Since $n \geq \text{rank}\,(A^k) \geq \text{rank}\,(A^{k+1}) \geq 0$, eventually $\text{rank}\,A^k = \text{tank}\,A^{k+1}$ for some $k \in \overline{1,n}$. □

Ex. 21. $(A^*)^D = (A^D)^*$.

Ex. 22. $(A^T)^D = (A^D)^T$.

Ex. 23. $(A^\ell)^D = (A^D)^\ell$ for $\ell = 1, 2, \ldots$.

Ex. 24. If A has index k, A^ℓ has index 1 and $(A^\ell)^\# = (A^D)^\ell$ for $\ell \geq k$.

Ex. 25. $(A^D)^D = A$ if and only if A has index 1 (Drazin).

Ex. 26. A^D has index 1, and $(A^D)^\# = A^2A^D$.

Ex. 27. $((A^D)^D)^D = A^D$ (Drazin).

Ex. 28. If A has index k, $R(A^D) = R(A^\ell)$ and $N(A^D) = N(A^\ell)$ for all $\ell \geq k$.

Ex. 29. $R(A^D)$ is the subspace spanned by all the λ–vectors of A for all nonzero eigenvalues λ, and $N(A^D)$ is the subspace spanned by all the 0–vectors of A, and these are complementary subspaces.

Ex. 30. $AA^D = A^DA$ is idempotent and is the projector on $R(A^D)$ along $N(A^D)$. Alternatively, if A has index k, it is the projector on $R(A^\ell)$ along $N(A^\ell)$ for all $\ell \geq k$.

Ex. 31. If A and X are S–inverses of each other, they have the same index.

Ex. 32. $A^D(A^D)^\# = AA^D$.

Ex. 33. Let $A \in \mathbb{C}^{n\times n}$ have index k. Then, for all $\ell \geq k$,

$$A^D = A^\ell(q(A))^{\ell+1} \, ,$$

where q is defined by (31).

Ex. 34. If A is nilpotent, $A^D = O$.

Ex. 35. If $\ell > m > 0$, $A^m(A^D)^\ell = (A^D)^{\ell-m}$.

Ex. 36. If $m > 0$ and $\ell - m \geq k$, $A^\ell(A^D)^m = A^{\ell-m}$.

Ex. 37. Let A have index k, and define as follows a set of matrices B_j where j ranges over all the integers:

$$B_j = \begin{cases} A^j & \text{for } j \geq k, \\ A^k(A^D)^{k-j} & \text{for } 0 \leq j < k, \\ (A^D)^{-j} & \text{for } j < 0. \end{cases}$$

Is the set of matrices $\{B_j\}$ an Abelian group under matrix multiplication with unit element B_0 and multiplication rule $B_\ell B_m = B_{\ell+m}$? Is there an equivalent, but easier way of defining the matrices B_j?

Ex. 38. If A has index k and $\ell \geq k$, show that

$$A^D = A^\ell(A^{2\ell+1})^{(1)}A^\ell \, , \tag{35}$$

where $(A^{2\ell+1})^{(1)}$ is an arbitrary element of $A^{2\ell+1}\{1\}$ (Greville [**583**]). Note that (23) is a particular case of (35).

Ex. 39. Let $A \in \mathbb{C}^{n\times n}$. Then A has index 1 if and only if the limit

$$\lim_{\lambda\to 0} (\lambda I_n + A)^{-1}A$$

exists, in which case

$$\lim_{\lambda\to 0} (\lambda I_n + A)^{-1}A = AA^\# \quad (\text{ Ben–Israel [}\mathbf{117}\text{]}) \, .$$

Remark. Here $\lambda \to 0$ means $\lambda \to 0$ through any neighborhood of 0 in $\mathbb{C}$ which excludes the nonzero eigenvalues of A.

PROOF. Let rank $A = r$ and let $A = FG$ be a full–rank factorization. Then the identity

$$(\lambda I_n + A)^{-1}A = F(\lambda I_r + GF)^{-1}G$$

holds whenever the inverse in question exists. Therefore the existence of $\lim_{\lambda \to 0}(\lambda I_n + A)^{-1}A$ is equivalent to the existence of $\lim_{\lambda \to 0}(\lambda I_r + GF)^{-1}$ which, in turn, is equivalent to the nonsingularity of GF. The proof is completed by using Theorems 1 and 2. ☐

EX. 40. Let $A \in \mathbb{C}^{n\times n}$. Then A is range–Hermitian if and only if

$$\lim_{\lambda \to 0}(\lambda I_n + A)^{-1}P_{R(A)} = A^{\dagger} \ .$$

PROOF. Follows from Ex. 39 and Theorem 3. ☐

EX. 41. Let $O \neq A \in \mathbb{C}^{m\times n}$. Then

$$\lim_{\lambda \to 0}(\lambda I_n + A^*A)^{-1}A^* = A^{\dagger} \quad \text{(den Broeder and Charnes [\textbf{238}])} \ . \tag{3.31}$$

PROOF.

$$\begin{aligned} \lim_{\lambda \to 0}(\lambda I_n + A^*A)^{-1}A^* &= \lim_{\lambda \to 0}(\lambda I_n + A^*A)^{-1}P_{R(A^*A)}A^* \\ &\qquad \text{(since } R(A^*) = R(A^*A)) \\ &= (A^*A)^{\dagger}A^* \quad \text{(by Ex. 40 since } A^*A \text{ is range–Hermitian)} \\ &= A^{\dagger} \quad \text{(by Ex. 1.16(d))} \ . \end{aligned}$$

☐

7. Spectral properties of the Drazin inverse

The spectral properties of the Drazin inverse are the same as those of the group inverse with regard to nonzero eigenvalues and the associated eigenvectors, but weaker for 0–vectors. The necessity for such weakening is apparent from the following theorem.

THEOREM 8. Let $A \in \mathbb{C}^{n\times n}$ and let $X \in A\{1\} \cup A\{2\}$ be an S–inverse of A. Then both A and X have index 1.

PROOF. First, let $X \in A\{1\}$, and suppose that $\mathbf{x}$ is a 0–vector of A of grade 2. Then, $A\mathbf{x}$ is a null–vector of A. Since X is an S–inverse of A, $A\mathbf{x}$ is also a null–vector of X. Thus,

$$\mathbf{0} = XA\mathbf{x} = AXA\mathbf{x} = A\mathbf{x} \ ,$$

which contradicts the assumption that $\mathbf{x}$ is a 0–vector of A of grade 2. Hence, A has no 0–vectors of grade 2, and therefore has index 1, by Theorem 6(d). By Ex. 31, X has also index 1.

If $X \in A\{2\}$, we reverse the roles of A and X. ☐

Accordingly, we relax the definition of the S–inverse (Definition 2, p. 131) as follows.

DEFINITION 3. X is an S'*–inverse* of A if, for all $\lambda \neq 0$, a vector $\mathbf{x}$ is a λ^{-1}–vector of X of grade p if and only if it is a λ–vector of A of grade p, and $\mathbf{x}$ is a 0–vector of X if and only if it is a 0–vector of A (without regard to grade).

THEOREM 9. For every square matrix A, A and A^D are S'–inverses of each other.

PROOF. Since A^D satisfies

$$A^D A^{k+1} = A^k \ , \quad A(A^D)^2 = A^D \ ,$$

the part of Definition 3 relating to nonzero eigenvalues follows from Lemma 3. Since A^D has index 1 by Ex. 26, all its 0–vectors are null vectors. Thus the part of Definition 3 relating to 0–vectors follows from Ex. 29. □

8. Index 1–nilpotent decomposition of a square matrix

The following theorem plays an important role in the study of spectral generalized inverses of matrices of index greater than 1. It is implicit in Wederburn's [**1538**] results on idempotent and nilpotent parts, but is not stated by him in this form.

THEOREM 10. A square matrix A has a unique decomposition

$$A = B + N \ , \tag{36}$$

such that B has index 1, N is nilpotent, and

$$NB = BN = O \ . \tag{37}$$

Moreover,

$$B = (A^D)^{\#} \ . \tag{38}$$

PROOF. Suppose A has a decomposition (36) such that B has index 1, N is nilpotent and (37) holds. We shall first show that this implies (38), and therefore the decomposition is unique if it exists.

Since

$$B^{\#} = B(B^{\#})^2 = (B^{\#})^2 B \ ,$$

we have

$$B^{\#} N = N B^{\#} = O \ .$$

Consequently,

$$AB^{\#} = BB^{\#} = B^{\#} A \ . \tag{39}$$

Moreover,

$$A(B^{\#})^2 = B(B^{\#})^2 = B^{\#} \ . \tag{40}$$

Because of (37), we have

$$A^\ell = (B + N)^\ell = B^\ell + N^\ell \quad (\ell = 1, 2, \dots) \ . \tag{41}$$

If ℓ is sufficiently large so that $N^\ell = O$,

$$A^\ell = B^\ell \ ,$$

and for such ℓ,

$$A^{\ell+1} B^{\#} = B^{\ell+1} B^{\#} = B^\ell \ . \tag{42}$$

It follows from (39),(40), and (42) that $X = B^{\#}$ satisfies (5),(27), and (28), and therefore

$$B^{\#} = A^D \ ,$$

which is equivalen to (38).

It remains to show that this decomposition has the required properties. Clearly B has index 1. By taking

$$N = A - (A^D)^\# \tag{43}$$

and noting that

$$(A^D)^\# = A^2 A^D$$

by Ex. 26, it is easily verified that (37) holds. Therefore (41) follows, and if k is the index of A,

$$A^k = B^k + N^k = A^{2k}(A^D)^k + N^k = A^k + N^k \ ,$$

and therefore $N^k = O$. □

We shall call the matrix N given by (43) the *nilpotent part* of A and shall denote it by $A^{(N)}$.

Theorem 11. Let $A \in \mathbb{C}^{n\times n}$. Then A and X are S'–inverses of each other if

$$X^D = (A^D)^\# \ . \tag{44}$$

Moreover, if $X \in A\{1\} \cup A\{2\}$, it is an S'–inverse of A only if (44) holds.

Proof. If (44), A and X have the same range and the same null space, and consequently the projectors XX^D and $AA^D = A^D(A^D)^\#$ are equal. Thus, if ℓ is the maximum of the indices of A and X,

$$XA^{\ell+1} = X(A^D)^\# A^D A^{\ell+1} = XX^D A^\ell = A^\ell \tag{45}$$

by Ex. 30. By interchanging the roles of A and X we obtain also

$$AX^{\ell+1} = X^\ell \ . \tag{46}$$

From (45) and (46), Lemma 3, Ex. 29 and the fact that A^D and X^D have the same null space, we deduce that A and X are S'–inverses of each other.

On the other hand, let A and X be S'–inverses of each other, and let $X \in A\{1\}$. Then, by Ex. 29,

$$N(A^D) = N(X^D) \ ,$$

and so,

$$(A^D)^\# X^{(N)} = (X^D)^\# A^{(N)} = O \ .$$

Similarly, since

$$R(A^D) = R(X^D) \ ,$$

(2.42) gives

$$N(A^{D*}) = N(X^{D*}) \ ,$$

and therefore

$$X^{(N)}(A^D)^\# = (X^D)^\# A^{(N)} = O \ .$$

Consequently

$$A = AXA = (A^D)^\#(X^D)^\#(A^D)^\# + A^{(N)}X^{(N)}A^{(N)} \ ,$$

and therefore

$$A^D = A^D A A^D = AA^D(X^D)^\# AA^D = (X^D)^\# \ , \tag{47}$$

since AA^D is the projector on the range of $(X^D)^\#$ along its null space. But (47) is equivalent to (44).

If $X \in A\{2\}$, we reverse the roles of A and X. □

Referring back to the proof of Theorem 5, we note that if A has index 1, a matrix X that is an S–inverse of A and also either a $\{1\}$–inverse or a $\{2\}$–inverse, is automatically a $\{1,2\}$–inverse. However, a similar remark does not apply when the index of A is greater than 1 and X is an S'–inverse of A. This is because $A^{(N)}$ is no longer a null matrix (as it is when A has index 1) and its properties must be taken into account. (For details see Ex. 49.)

9. Quasi–commuting inverses

Erdélyi [**470**] calls A and X *quasi–commuting inverses* of each other if they are $\{1,2,5^k,6^k\}$–inverses of each other for some positive integer k. He noted that such pairs of matrices the spectrum of X is obtained by replacing each eigenvalue λ of A by $\lambda^\dagger$. The following theorem shows that quasi–commuting inverses have much more extensive spectral properties.

Theorem 12. If A and X are quasi–commuting inverses, they are S'–inverses.

Proof. If A and X are $\{1,2,5^\ell,6^\ell\}$–inverses of each other, then

$$XA^{\ell+1} = A^\ell XA = A^\ell ,$$

and similarly,

$$AX^{\ell+1} = X^\ell . \tag{46}$$

In view of Lemma 3 and Ex. 29, all that remains in order to prove that A and X are S'–inverses of each other is to show that A^D and X^D have the same null space. Now,

$$\begin{aligned} A^D\mathbf{x} = \mathbf{0} \quad &\Longrightarrow \quad \mathbf{0} = A^{\ell+1}A^D\mathbf{x} = A^\ell\mathbf{x} \\ &\Longrightarrow \quad \mathbf{0} = X^{2\ell}A^\ell\mathbf{x} = A^\ell X^{2\ell}\mathbf{x} = X^\ell\mathbf{x} \quad \text{(by (46))} \\ &\Longrightarrow \quad \mathbf{0} = (X^D)^{\ell+1}X^\ell\mathbf{x} = X^D\mathbf{x} . \end{aligned}$$

Since the roles of A and X are symmetrical, the reverse implication follows by interchanging them. □

Corollary 3. A and X are quasi–commuting inverses of each other if and only if (44) holds and $A^{(N)}$ and $X^{(N)}$ are $\{1,2\}$–inverses of each other.

Proof. *If*: A and X are $\{1,2\}$–inverses of each other by Ex. 46. Choose ℓ sufficiently large so that $(A^{(N)})^\ell = O$. Then

$$\begin{aligned} XA^\ell &= ((X^D)^\# + X^{(N)})((A^D)^\#)^\ell \\ &= ((X^D)^\# + X^{(N)})((X^D)^\#)^\ell = ((X^D)^\#)^{\ell-1} = A^\ell X . \end{aligned}$$

By interchanging A and X, it follows also that A commutes with X^ℓ.

Only if: By Theorem 12, A and X are S'–inverses of each other. Then, by Theorem 11, (44) holds, and by Ex. 49, $A^{(N)}$ and $X^{(N)}$ are $\{1,2\}$–inverses of each other. □

10. Other spectral generalized inverses

Greville [**583**] calls X a *strong spectral inverse* if equations (10) are satisfied. Although this is not quite obvious, the relationship is a reciprocal one, and they can be called syrong spectral inverses of each other. If A has index 1, Theorem 4 shows that $A^\#$ is the only strong spectral inverse. Greville has shown that strong spectral inverses are quasi–commuting, but, for a matrix A with index greater than 1, the set of strong spectral inverses is a proper subset of the set of quasi–commuting inverses. Strong spectral inverses have some remarkable and, in some respects, complicated properties, and there are a number of open questions concerning them. As these properties relate to matrices of index greater than 1, which are not for most purposes a very important class, they will not be discussed further here. The interested reader may consult Greville [**583**].

Cline [**352**] has pointed out that a square matrix A of index 1 has a $\{1,2,3\}$–inverse whose range is $R(A)$. This is, therefore, a "least–squares" inverse and also has spectral properties (see Exs. 49 and 50). Greville [**584**] has extended this notion to square matrices of arbitrary index, but his extension raises some questions that have not been answered (see the conclusion of [**584**]).

Exercises.

Ex. 42. If A has index 1, $A^{(N)} = O$.

Ex. 43. If A is nilpotent, $\operatorname{rank} A^{\ell+1} < \operatorname{rank} A^{\ell}$ unless $A^{\ell} = O$,

Ex. 44. If A is nilpotent, the smallest positive integer ℓ such that $A^{\ell} = O$ is called the *index of nilpotency* of A. Show that this is the same as the index of A (see Definition 1).

Ex. 45. A and $A^{(N)}$ have the same index.

Ex. 46. $\operatorname{rank} A = \operatorname{rank} A^{D} + \operatorname{rank} A^{(N)}$.

Ex. 47. $A^{D}A^{(N)} = A^{(N)}A^{D} = O$.

Ex. 48. Every 0–vector of A of grade p is a 0–vector of $A^{(N)}$ of grade p.

Ex. 49. Let A and X satisfy (44). Then $X \in A\{1\}$ if and only if $A^{(N)} \in A^{(N)}\{1\}$. Similar statements with $\{1\}$ replaced by $\{2\}$ and by $\{1,2\}$ are also true.

Ex. 50. If A has index 1, show that $X = A^{\#}AA^{\dagger} \in A\{1,2,3\}$ (Cline). Show that this X has the properties of an S–inverse of A with respect to nonzero eigenvalues (but, in general, not with respect to 0–vectors). What is the condition on A that this X be an S–inverse of A?

Ex. 51. For square A with arbitrary index, Greville has suggested as an extension of Cline's inverse

$$X = A^{D}AA^{\dagger} + A^{(1)}A^{(N)}A^{\dagger} \ ,$$

where $A^{(1)}$ is an arbitrary element of $A\{1\}$. Show that $X \in A\{1,2,3\}$ and has some spectral properties. Describe its spectral properties precisely.

Ex. 52. Can a matrix A of index greater than 1 have an S–inverse? It can if we are willing to accept an "inverse" that is neither a $\{1\}$–inverse nor a $\{2\}$–inverse. Let

$$A^{(S)} = A^{D} + A^{(N)} \ .$$

Show that $A^{(S)}$ is an S–inverse of A and that $X = A^{(S)}$ is the unique solution of the four equations

$$\begin{aligned} AX &= XA \ , & A^{\ell+1}X &= A^{\ell} \ , \\ AX^{\ell+1} &= X^{\ell} \ , & A - X &= A^{\ell}X^{\ell}(A - X) \end{aligned}$$

for every positive integer ℓ not less than the index of A. Show also that $A^{(S)} = A^{\#}$ if A has index 1 and $(A^{(S)})^{(S)} = A$. In your opinion, can $A^{(S)}$ properly be called a generalized inverse of A?

Ex. 53. Let F be a square matrix ofindex 1, and let G be such that $R(FG) \subset R(G)$. Then,

$$R(FG) = R(F) \cap R(G) \ .$$

Proof. Evidently, $R(FG) \subset R(F)$ and therefore

$$R(FG) \subset R(F) \cap R(G) \ .$$

Now let $\mathbf{x} \in R(F) \cap R(G)$, and we must show that $\mathbf{x} \in R(FG)$. Since F has index 1, it has a group inverse $F^{\#}$, which, by Corollary 1, can be expressed as a polynomial in F, say $p(F)$. We have

$$\mathbf{x} = F\mathbf{y} = G\mathbf{z}$$

for some $\mathbf{y}, \mathbf{z}$, and therefore

$$\mathbf{x} = FF^{\#}\mathbf{x} = FF^{\#}G\mathbf{z} = Fp(F)G\mathbf{z} \ .$$

Since $R(FG) \subset R(G)$,

$$FG = GH$$

for some H, and, consequently,

$$F^\ell G = GH^\ell$$

for every non–negative integer ℓ. Thus

$$\mathbf{x} = Fp(F)G\mathbf{z} = FGp(H)\mathbf{z} \subset R(FG) . \qquad \square$$

(This is a slight extension of a result of Arghiriade [**36**].)

Ex. 54. *The "reverse–order" property for the Moore–Penrose inverse.* For some pairs of matrices A, B the relation

$$(AB)^\dagger = B^\dagger A^\dagger \tag{48}$$

holds, and for others it does not. There does not seem to be a simple criterion for distinguishing the cases in which (48) holds. The following result is due to Greville [**582**].

For matrices A, B such that AB exists,

$$(AB)^\dagger = B^\dagger A^\dagger \tag{48}$$

if and only if

$$R(A^*AB) \subset R(B) \quad \text{and} \quad R(BB^*A^*) \subset R(A^*) . \tag{49}$$

Proof. *If*: We have

$$BB^\dagger A^*AB = A^*AB \tag{50}$$

and

$$A^\dagger ABB^*A^* = BB^*A^* . \tag{51}$$

Taking conjugate transposes of both sides of (50) gives

$$B^*A^*ABB^\dagger = B^*A^*A , \tag{52}$$

and then multiplying on the right by $A^\dagger$ and on the left by $(AB)^{*\dagger}$ yields

$$ABB^\dagger A^\dagger = AB(AB)^\dagger . \tag{53}$$

Multiplying (51) on the left by $B^\dagger$ and on the right by $(AB)^{*\dagger}$ gives

$$B^\dagger A^\dagger AB = (AB)^\dagger AB . \tag{54}$$

It follows from (53) and (54) that $B^\dagger A^\dagger \in (AB)\{1, 3, 4\}$.

Finally, the equations,

$$B^*A^* = B^*BB^\dagger A^\dagger AA^* , \quad B^\dagger A^\dagger = B^\dagger B^{*\dagger}B^*A^*A^{*\dagger}A^\dagger$$

show that

$$\operatorname{rank} B^\dagger A^\dagger = \operatorname{rank} B^*A^* = \operatorname{rank} AB ,$$

and therefore $B^\dagger A^\dagger \in (AB)\{2\}$ by Theorem 1.2, and so (48) holds.

Only if: We have

$$B^*A^* = B^\dagger A^\dagger ABB^*A^* ,$$

and multiplying on the left by ABB^*B gives

$$ABB^*(I - A^\dagger A)BB^*A^* = O .$$

Since the left member is Hermitian and $I - A^\dagger A$ is idempotent, it follows that

$$(I - A^\dagger A)BB^*A^* = O ,$$

which is equivalent to (51). In an analogous manner, (50) is obtained. □

EX. 55. (Arghiriade [**36**].) For matrices A, B such that AB exists, (48) holds if and only if A^*ABB^* is range–Hermitian.

PROOF. We shall show that the condition that A^*ABB^* be range–Hermitian is equivalent to (49), and the result will then follow from Ex. 54. Let C denote A^*ABB^*, and observe that

$$R(A^*AB) = R(C)\ , \qquad R(BB^*A^*) = R(C^*)$$

because

$$CB^{*\dagger} = A^*AB\ , \qquad C^*A^\dagger = BB^*A^*\ .$$

Therefore it is sufficient to prove that $R(C) = R(C^*)$ if and only if $R(C) \subset R(B)$ and $R(C^*) \subset R(A^*)$.

If: A^*A and BB^* are Hermitian, and therefore of index 1 by Ex. 7. Since $R(BB^*) = R(B)$ by Corollary 1.2, it follows from Ex. 53 with $F = A^*A$, $G = BB^*$ that

$$R(C) = R(A^*) \cap R(B)\ .$$

Reversing the assignments of F and G gives

$$R(C^*) = R(A^*) \cap R(B)\ .$$

Thus $R(C) = R(C^*)$.

Only if: Obvious. □

EX. 56. (Cline [**352**].) If ℓ is any integer not less than the index of A,

$$(A^D)^\dagger = (A^\ell)^\dagger A^{2\ell+1}(A^\ell)^\dagger\ .$$

[*Hint:* Use Ex. 2.48, noting that $\mathbb{R}(A^D) = R(A^\ell)$ and $N(A^D) = N(A^\ell)$.]

EX. 57. If the mtrices A, E in $\mathbb{C}^{m\times n}$ satisfy

$$R(E) \subset R(A)\ , \tag{55}$$

$$R(E^*) \subset R(A^*)\ , \tag{56}$$

and

$$\|A^\dagger E\| < 1 \tag{57}$$

for any multiplicative matrix norm (see p. 13), then

$$(A+E)^\dagger = (I + A^\dagger E)^{-1}A^\dagger\ . \tag{58}$$

PROOF. The matrix $B = I + A^\dagger E$ is nonsingular by (57) and Exs. 0.35 and 0.41. Since

$$\begin{aligned} A + E &= A + AA^\dagger E\ , \quad \text{by (55)} \\ &= A(I + A^\dagger E)\ , \end{aligned}$$

it suffices to show that the matrices A and $B = I + A^\dagger$ have the "reverse order" property (48)

$$(A(I + A^\dagger E))^\dagger = (I + A^\dagger E)^{-1}A^\dagger\ ,$$

which by Ex. 54 is equivalent to

$$R(A^*AB) \subset R(B) \tag{59}$$

and

$$R(BB^*A^*) \subset R(A^*)\ . \tag{60}$$

Now (59) holds since B is nonsingular, and (60) follows from

$$\begin{aligned} R(BB^*A^*) &= R((I + A^\dagger E)(I + A^\dagger E)^* A^*) \\ &= R(A^* + E^* A^{\dagger *} A^* + A^\dagger E(I + A^\dagger E)^* A^*) \\ &\subset R(A^*) \ , \quad \text{by (56)} \ . \end{aligned}$$

□

Ex. 58. *Error bounds for generalized inverses* (Ben–Israel [**110**]). Let A, E satisfy (55), (56), and (57). Then,

$$\|(A+E)^\dagger - A^\dagger\| \leq \frac{\|A^\dagger E\| \|A^\dagger\|}{1 - \|A^\dagger E\|} \ . \tag{61}$$

If (55) and (56) hold, but (57) is replaced by

$$\|A^\dagger\| \|E\| < 1 \ , \tag{62}$$

then

$$\|(A+E)^\dagger - A^\dagger\| \leq \frac{\|A^\dagger\|^2 \|E\|}{1 - \|A^\dagger E\|} \ . \tag{63}$$

Proof. From Ex. 57 it follows that

$$\begin{aligned} (A+E)^\dagger - A^\dagger &= (I + A^\dagger E)^{-1} A^\dagger - A^\dagger \\ &= \sum_{k=0}^{\infty} (-1)^k (A^\dagger E)^k A^\dagger - A^\dagger \ , \quad \text{by (57) and Ex. 0.41,} \\ &= \sum_{k=1}^{\infty} (-1)^k (A^\dagger E)^k A^\dagger \end{aligned}$$

and hence,

$$\begin{aligned} \|(A+E)^\dagger - A^\dagger\| &\leq \sum_{k=1}^{\infty} \|(A^\dagger E)\|^k \|A^\dagger\| \\ &= \frac{\|A^\dagger E\| \|A^\dagger\|}{1 - \|A^\dagger E\|} \ , \quad \text{by (57)} \ . \end{aligned}$$

The condition (62) [which is stronger than (57)] then implies (63). □

For further results see Stewart [**1399**], Wedin [**1540**], [**1541**], Pereyra [**1180**], Golub and Pereyra [**556**] and Moore and Nashed [**1090**].

Suggested further reading

Section 4. For range–Hermitian matrices see Arghiriade [**36**], Katz [**824**], Katz and Pearl [**826**], Pearl ([**1168**], [**1169**], [**1170**]). For matrices of index 1 see Ben–Israel [**117**]. For the group inverse see Robert [**1278**].
Section 10. Poole and Boullion [**1195**], Ward, Boullion and Lewis [**1537**], and Scroggs and Odell [**1331**].

CHAPTER 5

Generalized Inverses of Partitioned Matrices

1. Introduction

In this chapter we study linear equations and matrices in partitioned form. For example, in computing a (generalized or ordinary) inverse of a matrix $A \in \mathbb{C}^{m\times n}$, the size or difficulty of the problem may be reduced if A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} .$$

The typical result here is the sought inverse expressed in terms of the submatrices A_{ij}.

Partitioning by columns and by rows is used in Section 2 to solve linear equations, and to compute generalized inverses and related items.

Intersections of linear manifolds are studied in Section 3, and used in Section 4 to obtain common solutions of pairs of linear equations and to invert matrices partitioned by rows.

Greville's method for computing $A^\dagger$ for $A \in \mathbb{C}^{m\times n}$, $n \geq 2$, is based on partitioning A as

$$A = \begin{bmatrix} A_{n-1} & \mathbf{a}_n \end{bmatrix}$$

where $\mathbf{a}_n$ is the nth column of A. $A^\dagger$ is then expressed in terms of $\mathbf{a}_n$ and $A_{n-1}^\dagger$, which is computed in the same way, using the partition

$$A_{n-1} = \begin{bmatrix} A_{n-2} & \mathbf{a}_{n-1} \end{bmatrix} , \quad \text{etc.}$$

Greville's method and some of its consequences are studied in Section 5.

Bordered matrices, the subject of Section 6, are matrices of the form

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix}$$

where $A \in \mathbb{C}^{m\times n}$ is given and U and V are chosen so that the resulting bordered matrix is nonsingular. Moreover,

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix}^{-1} = \begin{bmatrix} A^\dagger & V^{*\dagger} \\ U^\dagger & O \end{bmatrix}$$

expressing generalized inverses in terms of an ordinary matrix.

2. Partitioned matrices and linear equations

Consider the linear equation

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

with given matrix A and vector $\mathbf{b}$, in the following three cases.

Case 1. $A \in \mathbb{C}_r^{r\times n}$, i.e. A is of full row rank. Let the columns of A be rearranged, if necessary, so that the first r columns are linearly independent. A rearrangement of columns may be interpreted as postmultiplication by a suitable permutation matrix; thus,

$$AQ = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} Q^T , \tag{2}$$

where Q is an $n \times n$ permutation matrix (hence $Q^{-1} = Q^T$) and A_1 consists of r linearly independent columns, so that $A_1 \in \mathbb{C}_r^{r\times r}$, i.e., A_1 is nonsingular.

The matrix A_2 is in $\mathbb{C}^{r\times(n-r)}$ and if $n = r$, this matrix and other items indexed by the subscript 2 are to be interpreted as absent.

Corresponding to (2), let the vector $\mathbf{x}$ be partitioned

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} , \quad \mathbf{x}_1 \in \mathbb{C}^r . \tag{3}$$

Using (2) and (3) we rewrite (1) as

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} Q^T \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{b} \tag{4}$$

easily shown to be satisfied by the vector

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = Q \begin{bmatrix} A_1^{-1}\mathbf{b} \\ O \end{bmatrix} , \tag{5}$$

which is thus a particular solution of (1).

The general solution of (1) is obtained by adding to (5) the general element of $N(A)$, i.e., the general solution of

$$A\mathbf{x} = \mathbf{0} . \tag{6}$$

In (2) the columns of A_2 are linear combinations of the columns of A_1, say,

$$A_2 = A_1 T \quad \text{or} \quad T = A_1^{-1} A_2 \in \mathbb{C}^{r\times(n-r)} , \tag{7}$$

where the matrix T is called the *multiplier* corresponding to the partition (2), a name suggested by T being the "ratio" of the last $n - r$ columns of AQ to its first r columns.

Using (2), (3), and (7) permits writing (6) as

$$A_1 \begin{bmatrix} I_r & T \end{bmatrix} Q^T \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{0} , \tag{8}$$

whose general solution is clearly

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} \mathbf{y} , \tag{9}$$

where $\mathbf{y} \in \mathbb{C}^{n-r}$ is arbitrary.

Adding (5) and (9) we obtain the general solution of (1):

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = Q \begin{bmatrix} A_1^{-1}\mathbf{b} \\ O \end{bmatrix} + Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} \mathbf{y} , \quad \mathbf{y} \text{ arbitrary} . \tag{10}$$

Thus an advantage of partitioning A as in (2). is that it permits solving (1) by working with matrices smaller or more convenient than A. We also note that the null space of A is completely determined by the multiplier T and the permutation matrix Q, indeed (9) shows that the columns of the $n \times (n - r)$ matrix

$$Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} \tag{11}$$

form a basis for $N(A)$.

Case 2. $A \in \mathbb{C}_r^{m\times r}$**, i.e.** A **is of full column rank.** Unlike Case 1, here the linear equation (1) may be inconsistent. If, however, (1) is consistent, then it has a unique solution. Partitioning the rows of A is useful for both checking the consistency of (1) and for computing its solution, if consistent.

Let the rows of A be rearranged, if necessary, so that the first r rows are linearly independent. This is written, analogously to (2), as

$$PA = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad \text{or} \quad A = P^T \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} , \tag{12}$$

where P is an $m \times m$ permutation matrix, and $A_1 \in \mathbb{C}_r^{r\times r}$.

If $m = r$, the matrix A_2 and other items with the subscript 2 are to be interpreted as absent.

In (12) the rows of A_2 are linear combinations of the rows of A_1, say,

$$A_2 = SA_1 \quad \text{or} \quad S = A_2 A_1^{-1} \in \mathbb{C}^{(m-r)\times r} , \tag{13}$$

where again S is called the *multiplier* corresponding to the partition (12), giving the "ratio" of the last $(m-r)$ rows of PA to its first r rows.

Corresponding to (12) let the permutation matrix P be partitioned as

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} , \quad P_1 \in \mathbb{C}^{r\times m} . \tag{14}$$

Equation (1) can now be written, using (12), (13), and (14), as

$$\begin{bmatrix} I_r \\ S \end{bmatrix} A_1 \mathbf{x} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \mathbf{b} , \tag{15}$$

from which the conclusions below easily follow:

(a) Equation (1) is consistent if and only if

$$P_2 \mathbf{b} = SP_1 \mathbf{b} \tag{16}$$

i.e., the "ratio" of the last $m-r$ components of the vector $P\mathbf{b}$ to its first r components is the multiplier S of (13).

(b) If (16) holds, then the unique solution of (1) is

$$\mathbf{x} = A_1^{-1} P_1 \mathbf{b} . \tag{17}$$

From (a) we note that the range of A is completely determined by the multiplier S and the permutation matrix P. Indeed, the columns of the $m \times r$ matrix

$$P^T \begin{bmatrix} I_r \\ S \end{bmatrix} \tag{18}$$

form a basis for $R(A)$.

Case 3. $A \in \mathbb{C}_r^{m\times n}$**, with** $r \leq mnn$**.** This *general case* has some of the characteristics of both cases 1 and 2, as here we partition both the columns and rows of A.

Since rank $A = r$, A has at least one nonsingular $r \times r$ submatrix A_{11}, which by a rearrangement of rows and columns can be brought to the top left corner of A, say

$$PAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \tag{19}$$

where A and Q are permutation matrices, and $A_{11} \in |Cmnrrrr$.

By analogy with (2) and (12) we may have to interpret some of these submatrices as absent, e.g., A_{12} and A_{22} are absent if $n = r$.

By analogy with (7) and (13) there are *multipliers* $T \in \mathbb{C}^{r\times(n-r)}$ and $S \in \mathbb{C}^{(m-r)\times r}$, satisfying

$$\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} T \quad \text{and} \quad \begin{bmatrix} A_{21} & A_{22} \end{bmatrix} = S \begin{bmatrix} A_{11} & A_{12} \end{bmatrix} . \tag{20}$$

These multipliers are given by

$$T = A_{11}^{-1} A_{12} \quad \text{and} \quad S = A_{21} A_{11}^{-1} . \tag{21}$$

Combining (19) and (20) results in the following *partition* of $A \in \mathbb{C}_r^{m\times n}$

$$\begin{aligned} A &= P^T \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} Q^T \\ &= P^T \begin{bmatrix} I_r \\ S \end{bmatrix} A_{11} \begin{bmatrix} I_r & T \end{bmatrix} Q^T , \end{aligned} \tag{22}$$

where $A_{11} \in \mathbb{C}_r^{r\times r}$, P and Q are permutation matrices, and S and T are given by (21).

As in cases 1 and 2 we conclude that the multipliers S and T, and the permutation matrices P and Q, carry all the information about the range and null space of A.

Lemma 1. Let $A \in \mathbb{C}_r^{m\times n}$ be partitioned as in (22). Then

(a) The columns of the $n \times (n-r)$ matrix

$$Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} \tag{11}$$

form a basis for $N(A)$.

(b) The columns of the $m \times r$ matrix

$$P^T \begin{bmatrix} I_r \\ S \end{bmatrix} \tag{18}$$

form a basis for $R(A)$. □

Returning to the linear equation (1), it may be partitioned by using (22) and (14), in analogy with (4) and (15), as follows:

$$\begin{bmatrix} I_r \\ S \end{bmatrix} A_{11} \begin{bmatrix} I_r & T \end{bmatrix} Q^T \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \mathbf{b} . \tag{23}$$

The following theorem summarizes the situation, and includes the results of cases 1 and 2 as special cases.

Theorem 1. Let $A \in \mathbb{C}_r^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$ be given, and let the linear equation

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

be partitioned as in (23). Then

(a) Equation (1) is consistent if and only if[1]

$$P_2 \mathbf{b} = S P_1 \mathbf{b} \tag{16}$$

(b) If (16) holds, the general solution of (1) is

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = Q \begin{bmatrix} A_{11}^{-1} P_1 \mathbf{b} \\ O \end{bmatrix} + Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} \mathbf{y} , \tag{24}$$

where $\mathbf{y} \in \mathbb{C}^{n-r}$ is arbitrary. □

The partition (22) is useful also for computing generalized inverses. We collect some of these results in the following.

---

[1]By convention, (16) is satisfied if $m = r$, in which case P_2, and S are interpreted as absent.

Theorem 2. Let $A \in \mathbb{C}_r^{m\times n}$ be partitioned as in (22). Then

(a) A $\{1,2\}$–inverse of A is

$$A^{(1,2)} = Q \begin{bmatrix} A_{11}^{-1} & O \\ O & O \end{bmatrix} P \quad \text{(Rao [\textbf{1241}])} . \tag{25}$$

(b) A $\{1,2,3\}$–inverse of A is

$$A^{(1,2,3)} = Q \begin{bmatrix} A_{11}^{-1} \\ O \end{bmatrix} (I_r + S^*S)^{-1} \begin{bmatrix} I_r & S^* \end{bmatrix} P \quad \text{(Meyer and Painter [\textbf{1032}])} . \tag{26}$$

(c) A $\{1,2,4\}$–inverse of A is

$$A^{(1,2,4)} = Q \begin{bmatrix} I_r \\ T^* \end{bmatrix} (I_r + TT^*)^{-1} \begin{bmatrix} A_{11}^{-1} & O \end{bmatrix} P . \tag{27}$$

(d) The Moore–Penrose inverse of A is

$$A^{\dagger} = Q \begin{bmatrix} I_r \\ T^* \end{bmatrix} (I_r + TT^*)^{-1} A_{11}^{-1} (I_r + S^*S)^{-1} \begin{bmatrix} I_r & S^* \end{bmatrix} P \quad \text{(Noble [\textbf{1144}])} . \tag{28}$$

Proof. The partition (22) is a full–rank factorization of A (see Lemma 1.4),

$$A = FG , \quad F \in \mathbb{C}_r^{m\times r} , \quad G \in \mathbb{C}_r^{r\times n} \tag{29}$$

with

$$F = P^T \begin{bmatrix} I_r \\ S \end{bmatrix} A_{11} , \quad G = \begin{bmatrix} I_r & T \end{bmatrix} Q^T \tag{30}$$

or alternatively

$$F = P^T \begin{bmatrix} I_r \\ S \end{bmatrix} , \quad G = A_{11} \begin{bmatrix} I_r & T \end{bmatrix} Q^T . \tag{31}$$

The theorem now follows from Ex. 1.25 and Ex. 1.15 by using (29) with either (30) or (31). □

Exercises.

Ex. 1. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \quad A_{11} \text{ nonsingular} . \tag{32}$$

Then

$$\operatorname{rank} A = \operatorname{rank} A_{11} \tag{33}$$

if and only if

$$A_{22} = A_{21} A_{11}^{-1} A_{12} \quad \text{(Brand [\textbf{233}])} . \tag{34}$$

Ex. 2. Let A, A_{11} satisfy (32) and (33). Then the general solution of

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

is given by

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} -A_{11}^{-1} A_{12} \mathbf{x}_2 \\ \mathbf{x}_2 \end{bmatrix} , \quad \mathbf{x}_2 \text{ arbitrary} .$$

Ex. 3. Let A, A_{11} satisfy (32) and (33). Then the linear equation

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{35}$$

is consistent if and only if

$$A_{21} A_{11}^{-1} \mathbf{b}_1 = \mathbf{b}_2$$

in which case the general solution of (35) is given by

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11}^{-1}\mathbf{b}_1 - A_{11}^{-1} A_{12}\mathbf{x}_2 \\ \mathbf{x}_2 \end{bmatrix} , \quad \mathbf{x}_2 \text{ arbitrary} .$$

Ex. 4. Let A, A_{11} satisfy (32) and (33). Then

$$A^\dagger = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}^* T_{11}^* \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^* ,$$

where

$$T_{11} = \left(\begin{bmatrix} A_{11} & A_{12} \end{bmatrix} A^* \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \right)^{-1} \quad \text{(Zlobec [\textbf{1652}])} .$$

Ex. 5. Let $A \in \mathbb{C}_r^{n\times n}$, $r < n$, be partitioned by

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_r \\ S \end{bmatrix} A_{11} \begin{bmatrix} I_r & T \end{bmatrix} , \quad A_{11} \in \mathbb{C}_r^{r\times r} . \tag{36}$$

Then the group inverse $A^\#$ exists if and only if $I_r + ST$ is nonsingular, in which case

$$A^\# = \begin{bmatrix} I_r \\ S \end{bmatrix} ((I_r + TS)A_{11}(I_r + TS))^{-1} \begin{bmatrix} I_r & T \end{bmatrix} \quad \text{(Robert [\textbf{1278}])} . \tag{37}$$

Ex. 6. Let $A \in \mathbb{C}_r^{n\times n}$ be partitioned as in (36). Then A is range–Hermitian if and only if $S = T^*$.

Ex. 7. Let $A \in \mathbb{C}_r^{m\times n}$ be partitioned as in (22). Then the following orthogonal projectors are given in terms of the multipliers S, T and the permutation matrices P, Q as:

(a) $P_{R(A)} = P^T \begin{bmatrix} I_r \\ S \end{bmatrix} (I_r + S^*S)^{-1} \begin{bmatrix} I_r & S^* \end{bmatrix} P$,

(b) $P_{R(A^*)} = Q \begin{bmatrix} I_r \\ T^* \end{bmatrix} (I_r + TT^*)^{-1} \begin{bmatrix} I_r & T \end{bmatrix} Q^T$,

(c) $P_{N(A)} = Q \begin{bmatrix} -T \\ I_{n-r} \end{bmatrix} (I_{n-r} + T^*T)^{-1} \begin{bmatrix} -T^* & I_{n-r} \end{bmatrix} Q^T$,

(d) $P_{N(A^*)} = P^T \begin{bmatrix} -S \\ I_{m-r} \end{bmatrix} (I_{m-r} + SS^*)^{-1} \begin{bmatrix} -S & I_{m-r} \end{bmatrix} P$.

remark. (a) and (d) are alternative computations since

$$P_{R(A)} + P_{N(A^*)} = I_m .$$

The computation (a) requires inverting the $r \times r$ positive definite matrix $I_r + S^*S$, while in (d) the dimension of the positive definite matrix to be inverted is $(m - r) \times (m - r)$. Accordingly (a) may be preferred if $r < m - r$,

Similarly (b) and (c) are alternative computations since

$$P_{R(A^*)} + P_{N(A)} = I_n$$

with (b) preferred if $r < n - r$.

EX. 8. (Albert [**12**]) For a Hermitian matrix H we denote by
$H \geq O$ the fact that H is positive semidefinite,
$H > O$ that H is positive definite. Let

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{22} \end{bmatrix} ,$$

where H_{11} and H_{22} are Hermitian. Then:

(a) $H \geq O$ if and only if

$$H_{11} \geq O \ , \quad H_{11}H_{11}^{\dagger}H_{12} = H_{12} \quad \text{and} \quad H_{22} - H_{12}^*H_{11}^{\dagger}H_{12} \geq O$$

(b) $H > O$ if and only if

$$H_{11} > O \ , \quad H_{11} - H_{12}H_{22}^{\dagger}H_{12}^* > O \quad \text{and} \quad H_{22} - H_{12}^*H_{11}^{-1}H_{12} > O$$

EX. 9. (Rohde [**1297**]) Let

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{22} \end{bmatrix} ,$$

be Hermitian positive semidefinite, and denote

$$H^{(\alpha)} = \begin{bmatrix} H_{11}^{(\alpha)} + H_{11}^{(\alpha)}H_{12}G^{(\alpha)}H_{12}^*H_{11}^{(\alpha)} & -H_{11}^{(\alpha)}H_{12}G^{(\alpha)} \\ -G^{(\alpha)}H_{12}^*H_{11}^{(\alpha)} & G^{(\alpha)} \end{bmatrix} , \tag{38}$$

where

$$G = H_{22} - H_{12}^*H_{11}^{(\alpha)}H_{12}$$

and α is an integer, or a set of integers, to be specified below. Then:

(a) The relation (38) is an identity for $\alpha = 1$ and $\alpha = \{1, 2\}$. This means that RHS(38) is an $\{\alpha\}$–inverse of H if in it one substitutes the $\{\alpha\}$–inverses of H_{11} and G as indicated.

(b) If H_{22} is nonsingular and $\operatorname{rank} H = \operatorname{rank} H_{11} + \operatorname{rank} H_{22}$, then (38) is an identity with $\alpha = \{1, 2, 3\}$ and $\alpha = \{1, 2, 3, 4\}$.

3. Intersection of manifolds

For any vector $\mathbf{f} \in \mathbb{C}^n$ and a subspace L of $\mathbb{C}^n$, the set

$$\mathbf{f} + L = \{\mathbf{f} + \boldsymbol{\ell} : \boldsymbol{\ell} \in L\} \tag{39}$$

is called a (*linear*) *manifold* (also *affine set*). The vector $\mathbf{f}$ in (39) is not unique, indeed

$$\mathbf{f} + L = (\mathbf{f} + \boldsymbol{\ell}) + L \quad \text{for any } \boldsymbol{\ell} \in L \ .$$

This nonuniqueness suggests singling out the representation

$$(\mathbf{f} - P_L\mathbf{f}) + L = P_{L^\perp}\mathbf{f} + L \tag{40}$$

of the manifold (39) and calling it the *orthogonal representation* of $\mathbf{f} + L$. We note that $P_{L^\perp}\mathbf{f}$ is the unique vector of least Euclidean norm in the manifold (39).

In this section we study the intersection of two manifolds

$$\{\mathbf{f} + L\} \cap \{\mathbf{g} + M\} \tag{41}$$

for given vectors $\mathbf{f}$ and $\mathbf{g}$, and given subspaces L and M in $\mathbb{C}^n$. The results are needed in Section 4 below where the common solutions of pairs of linear equations are studied. Let such a pair be

$$A\mathbf{x} = \mathbf{a} \tag{42a}$$

and

$$B\mathbf{x} = \mathbf{b} \tag{42b}$$

where A and B are given matrices with n columns, and $\mathbf{a}$ and $\mathbf{b}$ are given vectors. Assuming (42a) and (42b) to be consistent, their solutions are the manifolds

$$A^{\dagger}\mathbf{a} + N(A) \tag{43a}$$

and

$$B^{\dagger}\mathbf{b} + N(B) , \tag{43b}$$

respectively. If the intersection of these manifolds

$$\{A^{\dagger}\mathbf{a} + N(A)\} \cap \{B^{\dagger}\mathbf{b} + N(B)\} \tag{44}$$

is nonempty, then it is the set of common solutions of (42a)–(42b). This is the main reason for our interest in intersections of manifolds, whose study here includes conditions for the intersection (41) to be nonempty, in which case its properties and representations are given.

Since linear subspaces are manifolds, this special case is considered first.

Lemma 2. Let L and M be subspaces of $\mathbb{C}^n$, with P_L and P_M the corresponding orthogonal projectors. Then

$$\begin{aligned} P_{L+M} &= (P_L + P_M)(P_L + P_M)^{\dagger} \\ &= (P_L + P_M)^{\dagger}(P_L + P_M) . \end{aligned} \tag{45}$$

Proof. Clearly $L + M = R(\begin{bmatrix} P_L & P_M \end{bmatrix})$. Therefore,

$$\begin{aligned} P_{L+M} &= \begin{bmatrix} P_L & P_M \end{bmatrix} \begin{bmatrix} P_L & P_M \end{bmatrix}^{\dagger} \\ &= \begin{bmatrix} P_L & P_M \end{bmatrix} \begin{bmatrix} P_L \\ P_M \end{bmatrix} \begin{bmatrix} P_L & P_M \end{bmatrix}^{\dagger} \quad \text{(by Ex. 10)} \\ &= (P_L + P_M)(P_L + P_M)^{\dagger} , \quad \text{since } P_L \text{ and } P_M \text{ are idempotent} \\ &= (P_L + P_M)^{\dagger}(P_L + P_M) \end{aligned}$$

since a Hermitian matrix commutes with its Moore–Penrose inverse. □

The interesection of any two subspaces L and M in $\mathbb{C}^n$ is a subspace $L \cap M$ in $\mathbb{C}^n$, nonempty since $\mathbf{0} \in L \cap M$. The orthogonal projector $P_{L\cap M}$ is given in terms of P_L and P_M in the following.

Theorem 3. (Anderson and Duffin [**26**]). Let L, M, P_L, and P_M be as in Lemma 2. Then

$$\begin{aligned} P_{L\cap M} &= 2P_L(P_L + P_M)^{\dagger}P_M \\ &= 2P_M(P_L + P_M)^{\dagger}P_L \end{aligned} \tag{46}$$

Proof. Since $M \subset L + M$, it follows that

$$P_{L+M}P_M = P_M = P_M P_{L+M} , \tag{47}$$

and by using (45)

$$(P_L + P_M)(P_L + P_M)^{\dagger}P_M = P_M = P_M(P_L + P_M)^{\dagger}(P_L + P_M) . \tag{48}$$

Subtracting $P_M(P_L + P_M)^{\dagger}P_M$ from the first and last expressions in (48) gives

$$P_L(P_L + P_M)^{\dagger}P_M = P_M(P_L + P_M)^{\dagger}P_L . \tag{49}$$

Now, let

$$H = 2P_L(P_L + P_M)^{\dagger}P_M = 2P_M(P_L + P_M)^{\dagger}P_L .$$

Evidently, $R(H) \subset L \cap M$, and therefore

$$\begin{aligned} H = P_{L\cap M} H &= P_{L\cap M}\big(P_L(P_L+P_M)^\dagger P_M + P_M(P_L+P_M)^\dagger P_L\big) \\ &= P_{L\cap M}(P_L+P_M)^\dagger(P_L+P_M) \\ &= P_{L\cap M}P_{L+M} \quad \text{(by Lemma 2)} \\ &= P_{L\cap M}\ , \end{aligned}$$

since $L \cap M \subset L + M$. □

Other expressions for $L \cap M$ are given in the following theorem.

Theorem 4. (Lent [**924**]). Let L and M be subspaces of $\mathbb{C}^n$. Then

(a) $L \cap M = \begin{bmatrix} P_L & O \end{bmatrix} N(\begin{bmatrix} P_L & -P_M \end{bmatrix}) = \begin{bmatrix} O & P_M \end{bmatrix} N(\begin{bmatrix} P_L & -P_M \end{bmatrix})$

(b) $ = N(P_{L^\perp} + P_{M^\perp})$

(c) $ = N(I - P_L P_M) = N(I - P_M P_L)$.

Proof. (a) $\mathbf{x} \in L \cap M$ if and only if

$$\mathbf{x} = P_L\mathbf{y} = P_M\mathbf{z} \quad \text{for some } \mathbf{y}, \mathbf{z} \in \mathbb{C}^n\ ,$$

which is equivalent to

$$\mathbf{x} = \begin{bmatrix} P_L & O \end{bmatrix}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} O & P_M \end{bmatrix}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}\ , \text{ where } \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \in N(\begin{bmatrix} P_L & -P_M \end{bmatrix})\ .$$

(b) Let $\mathbf{x} \in L \cap M$. Then $P_{L^\perp}\mathbf{x} = P_{M^\perp}\mathbf{x} = \mathbf{0}$, proving that $\mathbf{x} \in N(P_{L^\perp} + P_{M^\perp})$. Conversely, let $\mathbf{x} \in N(P_{L^\perp} + P_{M^\perp})$, i.e.,

$$(I - P_L)\mathbf{x} + (I - P_M)\mathbf{x} = \mathbf{0}$$

or

$$2\mathbf{x} = P_L\mathbf{x} + P_M\mathbf{x}$$

and therefore

$$2\|\mathbf{x}\| \le \|P_L\mathbf{x}\| + \|P_M\mathbf{x}\|\ ,$$

by the triangle inequality for norms. But by Ex. 2.39,

$$\|P_L\mathbf{x}\| \le \|\mathbf{x}\|\ , \quad \|P_M\mathbf{x}\| \le \|\mathbf{x}\|\ .$$

Therefore,

$$\|P_L\mathbf{x}\| = \|\mathbf{x}\| = \|P_M\mathbf{x}\|$$

and so, by Ex. 2.39,

$$P_L\mathbf{x} = \mathbf{x} = P_M\mathbf{x}\ ,$$

proving $\mathbf{x} \in L \cap M$.

(c) Let $\mathbf{x} \in L \cap M$. Then $\mathbf{x} = P_L\mathbf{x} = P_M\mathbf{x} = P_L P_M\mathbf{x}$, and therefore $\mathbf{x} \in N(I - P_L P_M)$. Conversely, let $\mathbf{x} \in N(I - P_L P_M)$ and therefore,

$$\mathbf{x} = P_L P_M \mathbf{x} \in L\ . \tag{50}$$

Also,

$$\begin{aligned}\|P_M\mathbf{x}\|^2 + \|P_{M^\perp}\mathbf{x}\|^2 &= \|\mathbf{x}\|^2 \\ &= \|P_L P_M \mathbf{x}\|^2 \\ &\le \|P_M\mathbf{x}\|^2 \ , \quad \text{by Ex. 2.39.}\end{aligned}$$

Therefore,

$$P_{M^\perp}\mathbf{x} = \mathbf{0} \ , \quad \text{i.e., } \mathbf{x} \in M$$

and by (50),

$$\mathbf{x} \in L \cap M \ .$$

The remaining equality in (c) is proved similarly. □

The intersection of manifolds, which if nonempty is itself a manifold, can now be determined.

Theorem 5. (Ben–Israel [**112**], Lent [**924**]). Let $\mathbf{f}$ and $\mathbf{g}$ be vectors in $\mathbb{C}^n$ and let L and M be subspaces of $\mathbb{C}^n$. Then the intersection of manifolds

$$\{\mathbf{f} + L\} \cap \{\mathbf{g} + M\} \tag{41}$$

is nonempty if and only if

$$\mathbf{g} - \mathbf{f} \in L + M \ , \tag{51}$$

in which case

$$\begin{aligned}
&\text{(a)}\quad \{\mathbf{f}+L\}\cap\{\mathbf{g}+M\} &&= \mathbf{f} + P_L(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f}) + L\cap M \\
&\text{(a}'\text{)} &&= \mathbf{g} - P_M(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f}) + L\cap M \\
&\text{(b)} &&= \mathbf{f} + (P_{L^\perp}+P_{M^\perp})^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M \\
&\text{(b}'\text{)} &&= \mathbf{g} - (P_{L^\perp}+P_{M^\perp})^\dagger P_{L^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M \\
&\text{(c)} &&= \mathbf{f} + (I - P_M P_L)^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M \\
&\text{(c}'\text{)} &&= \mathbf{g} - (I - P_L P_M)^\dagger P_{L^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M.
\end{aligned}$$

Proof. $\{\mathbf{f}+L\}\cap\{\mathbf{g}+M\}$ is nonempty if and only if

$$\mathbf{f} + \boldsymbol{\ell} = \mathbf{g} + \mathbf{m} \ , \quad \text{for some } \boldsymbol{\ell} \in L \ , \ \mathbf{m} \in M \ ,$$

which is equivalent to

$$\mathbf{g} - \mathbf{f} = \boldsymbol{\ell} - \mathbf{m} \in L + M \ .$$

We now prove (a), (b), and (c). The primed statements (a′), (b′), and (c′) are proved similarly to their unprimed counterparts.

(a) The points $\mathbf{x} \in \{\mathbf{f}+L\}\cap\{\mathbf{g}+M\}$ are characterized by

$$\mathbf{x} = \mathbf{f} + P_L\mathbf{u} = \mathbf{g} + P_M\mathbf{v} \ , \quad \text{for some } \mathbf{u},\mathbf{v} \in \mathbb{C}^n \ . \tag{52}$$

Thus

$$\begin{bmatrix} P_L & -P_M \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{g} - \mathbf{f} \ . \tag{53}$$

The linear equation (53) is consistent, since (41) is nonempty, and therefore the general solution of (53) is

$$\begin{aligned}\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} &= \begin{bmatrix} P_L & -P_M \end{bmatrix}^\dagger (\mathbf{g}-\mathbf{f}) + N(\begin{bmatrix} P_L & -P_M \end{bmatrix}) \\ &= \begin{bmatrix} P_L \\ -P_M \end{bmatrix} (P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f}) + N(\begin{bmatrix} P_L & -P_M \end{bmatrix}) \quad \text{by Ex. 10.}\end{aligned} \tag{54}$$

Substituting (54) in (52) gives

$$\begin{aligned}\mathbf{x} &= \mathbf{f} + \begin{bmatrix} P_L & O \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \\ &= \mathbf{f} + P_L(P_L + P_M)^{\dagger}(\mathbf{g} - \mathbf{f}) + L \cap M\end{aligned}$$

by Theorem 4(a).

(b) Writing (52) as

$$P_L\mathbf{u} - P_M\mathbf{v} = \mathbf{g} - \mathbf{f}$$

and multiplying by $P_{M^\perp}$ gives

$$P_{M^\perp}P_L\mathbf{u} = P_{M^\perp}(\mathbf{g} - \mathbf{f}) \ , \tag{55}$$

which implies

$$(P_{L^\perp} + P_{M^\perp})P_L\mathbf{u} = P_{M^\perp}(\mathbf{g} - \mathbf{f}) \ . \tag{56}$$

The general solution of (56) is

$$\begin{aligned}P_L\mathbf{u} &= (P_{L^\perp} + P_{M^\perp})^{\dagger}P_{M^\perp}(\mathbf{g} - \mathbf{f}) + N(P_{L^\perp} + P_{M^\perp}) \\ &= (P_{L^\perp} + P_{M^\perp})^{\dagger}P_{M^\perp}(\mathbf{g} - \mathbf{f}) + L \cap M \ ,\end{aligned}$$

by Theorem 4(b), which when substituted in (52) proves (b).

(c) Equation (55) can be written as

$$(I - P_MP_L)P_L\mathbf{u} = P_{M^\perp}(\mathbf{g} - \mathbf{f})$$

whose general solution is

$$\begin{aligned}P_L\mathbf{u} &= (I - P_MP_L)^{\dagger}P_{M^\perp}(\mathbf{g} - \mathbf{f}) + N(I - P_MP_L) \\ &= (I - P_MP_L)^{\dagger}P_{M^\perp}(\mathbf{g} - \mathbf{f}) + L \cap M \ ,\end{aligned}$$

by Theorem 4(c), which when substituted in (52) proves (c). □

Theorem 5 verifies that the intersection (41), if nonempty, is itself a manifold. We note, in passing, that parts (a) and (a′) of Theorem 5 give the same representation of (41); i.e., if (51) holds, then

$$\mathbf{f} + P_L(P_L + P_M)^{\dagger}(\mathbf{g} - \mathbf{f}) = \mathbf{g} - P_M(P_L + P_M)^{\dagger}(\mathbf{g} - \mathbf{f}) \ . \tag{57}$$

Indeed, (51) implies that

$$\begin{aligned}\mathbf{g} - \mathbf{f} &= P_{L+M}(\mathbf{g} - \mathbf{f}) \\ &= (P_L + P_M)(P_L + P_M)^{\dagger}(\mathbf{g} - \mathbf{f}) \ ,\end{aligned}$$

which gives (57) by rearrangement of terms.

It will now be proved that parts (a), (a′), (b), and (b′) of Theorem 5 give orthogonal representations of

$$\{\mathbf{f} + L\} \cap \{\mathbf{g} + M\} \tag{41}$$

if the representations $\{\mathbf{f} + L\}$ and $\{\mathbf{g} + M\}$ are orthogonal, i.e., if

$$\mathbf{f} \in L^\perp \ , \quad \mathbf{g} \in M^\perp \ . \tag{58}$$

Corollary 1. Let L and M be subspaces of $\mathbb{C}^n$, and let

$$\mathbf{f} \in L^\perp \ , \quad \mathbf{g} \in M^\perp \ . \tag{58}$$

If (41) is nonempty, then each of the four representations given below is orthogonal.

$$
\begin{aligned}
&\text{(a)} \quad \{\mathbf{f}+L\}\cap\{\mathbf{g}+M\} &&= \mathbf{f} + P_L(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f}) + L\cap M\\
&\text{(a}'\text{)} &&= \mathbf{g} - P_M(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f}) + L\cap M\\
&\text{(b)} &&= \mathbf{f} + (P_{L^\perp}+P_{M^\perp})^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M\\
&\text{(b}'\text{)} &&= \mathbf{g} - (P_{L^\perp}+P_{M^\perp})^\dagger P_{L^\perp}(\mathbf{g}-\mathbf{f}) + L\cap M\ .
\end{aligned}
$$

PROOF. Each of the above representations is of the form

$$\{\mathbf{f}+L\}\cap\{\mathbf{g}+M\} = \mathbf{v} + L\cap M\ , \tag{59}$$

which is an orthogonal representation if and only if

$$P_{L\cap M}\mathbf{v} = \mathbf{0}\ . \tag{60}$$

In the proof we use the facts

$$P_{L\cap M} = P_L P_{L\cap M} = P_{L\cap M}P_L = P_M P_{L\cap M} = P_{L\cap M}P_M\ , \tag{61}$$

which hold since $L\cap M$ is contained in both L and M.

(a) Here $\mathbf{v} = \mathbf{f} + P_L(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f})$. The matrix P_L+P_M is Hermitian, and therefore $(P_L+P_M)^\dagger$ is a polynomial in powers of P_L+P_M, by Theorem 4.7. From (61) it follows therefore that

$$P_{L\cap M}(P_L+P_M)^\dagger = (P_L+P_M)^\dagger P_{L\cap M} \tag{62}$$

and (60) follows from

$$
\begin{aligned}
P_{L\cap M}\mathbf{v} &= P_{L\cap M}\mathbf{f} + P_{L\cap M}P_L(P_L+P_M)^\dagger(\mathbf{g}-\mathbf{f})\\
&= P_{L\cap M}\mathbf{f} + (P_L+P_M)^\dagger P_{L\cap M}(\mathbf{g}-\mathbf{f}) \quad \text{(by (61) and (62))}\\
&= \mathbf{0}\ , \quad \text{by (58)}\ .
\end{aligned}
$$

(a$'$) follows from (57) and (a).

(b) Here $\mathbf{v} = \mathbf{f} + (P_{L^\perp}+P_{M^\perp})^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f})$. The matrix $P_{L^\perp}+P_{M^\perp}$ is Hermitian, and therefore $(P_{L^\perp}+P_{M^\perp})^\dagger$ is a polynomial in $P_{L^\perp}+P_{M^\perp}$, which implies that

$$P_{L\cap M}(P_{L^\perp}+P_{M^\perp})^\dagger = O\ . \tag{63}$$

Finally, (60) follows from

$$
\begin{aligned}
P_{L\cap M}\mathbf{v} &= P_{L\cap M}\mathbf{f} + P_{L\cap M}(P_{L^\perp}+P_{M^\perp})^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f})\\
&= \mathbf{0}\ , \quad \text{by (58) and (58)}\ .
\end{aligned}
$$

(b$'$) If (58) holds, then

$$
\begin{aligned}
\mathbf{g}-\mathbf{f} &= P_{L^\perp+M^\perp}(\mathbf{g}-\mathbf{f})\\
&= (P_{L^\perp}+P_{M^\perp})^\dagger(P_{L^\perp}+P_{M^\perp})(\mathbf{g}-\mathbf{f})\ ,
\end{aligned}
$$

by Lemma 2, and therefore

$$\mathbf{f} + (P_{L^\perp}+P_{M^\perp})^\dagger P_{M^\perp}(\mathbf{g}-\mathbf{f}) = \mathbf{g} - (P_{L^\perp}+P_{M^\perp})^\dagger P_{L^\perp}(\mathbf{g}-\mathbf{f})\ ,$$

which proves (b$'$) identical to (b), if (58) is satisfied. □

Finally, we characterize subspaces L and M for which the intersection (41) is always nonempty.

COROLLARY 2. Let L and M be subspaces of $\mathbb{C}^n$. Then the intersection

$$\{\mathbf{f}+L\}\cap\{\mathbf{g}+M\} \tag{41}$$

is nonempty for all $\mathbf{f},\mathbf{g}\in\mathbb{C}^n$ if and only if

$$L^\perp\cap M^\perp = \{\mathbf{0}\}\ . \tag{64}$$

PROOF. The intersection (41) is by Theorem 5 nonempty for all $\mathbf{f}, \mathbf{g} \in \mathbb{C}^n$, if and only if

$$L + M = \mathbb{C}^n ,$$

which is equivalent to

$$\begin{aligned} \{\mathbf{0}\} &= (L+M)^{\perp} \\ &= L^{\perp} \cap M^{\perp} , \end{aligned}$$

by Ex. 11(b). □

Exercises and examples.

EX. 10. Let P_L and P_M be $n \times n$ orthogonal projectors. Then

$$\begin{bmatrix} P_L & \pm P_M \end{bmatrix}^{\dagger} = \begin{bmatrix} P_L \\ \pm P_M \end{bmatrix} (P_L + P_M)^{\dagger} \tag{65}$$

PROOF. Use $A^{\dagger} = A^*(AA^*)^{\dagger}$ with $A = \begin{bmatrix} P_L & \pm P_M \end{bmatrix}$, and the fact that P_L and P_M are Hermitian idempotents. □

EX. 11. Let L and M be subspaces of $\mathbb{C}^n$. Then:
(a) $(L \cap M)^{\perp} = L^{\perp} + M^{\perp}$
(b) $(L^{\perp} \cap M^{\perp})^{\perp} = L + M$.

PROOF. (a) Evidently $L^{\perp} \subset (L \cap M)^{\perp}$ and $M^{\perp} \subset (L \cap M)^{\perp}$; hence

$$L^{\perp} + M^{\perp} \subset (L \cap M)^{\perp} .$$

Conversely, from $L^{\perp} \subset L^{\perp} + M^{\perp}$ it follows that

$$(L^{\perp} + M^{\perp})^{\perp} \subset L^{\perp\perp} = L .$$

Similarly $(L^{\perp} + M^{\perp})^{\perp} \subset M$, hence

$$L^{\perp} + M^{\perp})^{\perp} \subset L \cap M$$

and by taking orthogonal complements

$$(L \cap M)^{\perp} \subset L^{\perp} + M^{\perp} .$$

(b) Follows from (a) by replacing L and M by $L^{\perp}$ and $M^{\perp}$, respectively. □

EX. 12. (von Neumann [**1507**]). Let $L_1, L_2, \dots , L_k$ be any k linear subspaces of $\mathbb{C}^n$, $k \geq 2$, and let

$$Q = P_{L_k} P_{L_{k-1}} \cdots P_{L_2} P_{L_1} P_{L_2} \cdots P_{L_{k-1}} P_{L_k} . \tag{66}$$

Then the orthogonal projector on $\bigcap_{i=1}^{k} L_i$ is $\lim_{m \to \infty} Q^m$.

EX. 13. (Pyle [**1223**]). The matrix Q of (66) is Hermitian, so let its spectral decomposition be given by

$$Q = \sum_{i=1}^{q} \lambda_i E_i$$

where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$$

are the distinct eigenvalues of Q, and

$$E_1, E_2, \dots , E_q$$

are the corresponding orthogonal projectors satisfying

$$E_1 + E_2 + \cdots + E_q = I$$

and

$$E_iE_j = O \quad \text{if } i \neq j \ ,$$

Then

$$1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \geq 0$$

and

$$\bigcap_{i=1}^{k} L_i \neq \{\mathbf{0}\} \text{ if and only if } \lambda_1 = 1 \ ,$$

in which case the orthogonal projector on $\bigcap_{i=1}^{k} L_i$ is E_1.

Ex. 14. *A closed–form expression.* Using the notation of Ex. 13, the orthogonal projector on $\bigcap_{i=1}^{k} L_i$ is

$$Q^{\nu} + \left[(Q^{\nu+1} - Q^{\nu})^{\dagger} - (Q^{\nu} - Q^{\nu-1})^{\dagger}\right]^{\dagger} \ , \quad \text{for } \nu = 2, 3, \dots \tag{67}$$

If λ_q, the smallest eigenvalue of Q, is positive then (67) holds also for $\nu = 1$, in which case Q^0 is taken as I. (Pyle [**1223**]).

4. Common solutions of linear equations and generalized inverses of partitioned matrices

Consider the pair of linear equations

$$A\mathbf{x} = \mathbf{a} \tag{42a}$$

$$B\mathbf{x} = \mathbf{b} \tag{42b}$$

with given vectors $\mathbf{a}, \mathbf{b}$ and matrices A, B having n columns.

Assuming (42a) and (42b) to be consistent, we study here their common solutions, if any, expressing them in terms of the solutions of (42a) and (42b).

The common solutions of (42a) and (42b) are the solutions of the partitioned linear equation

$$\begin{bmatrix} A \\ B \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \ , \tag{68}$$

which is often the starting point, the partitioning into (42a) and (42b) being used to reduce the size or difficulty of the problem.

The solutions of (42a) and (42b) constitute the manifolds

$$A^{\dagger}\mathbf{a} + N(A) \tag{43a}$$

and

and

$$B^{\dagger}\mathbf{b} + N(B) \ , \tag{43b}$$

respectively. Thus the intersection

$$\{A^{\dagger}\mathbf{a} + N(A)\} \cap \{B^{\dagger}\mathbf{b} + N(B)\} \tag{44}$$

is the set of solutions of (68), and (68) is consistent if and only if (44) is nonempty.

The results of Section 3 are applicable to determining the intersection (44). In particular, Theorem 5 yields the following

Corollary 3. Let A and B be matrices with n columns, and let $\mathbf{a}$ and $\mathbf{b}$ be vectors such that each of the equations (42a) and (42b) is consistent. Then (42a) and (42b) have common solutions if and only if

$$B^\dagger \mathbf{b} - A^\dagger \mathbf{a} \in N(A) + N(B) \tag{69}$$

in which case the set of common solutions is the manifold

$$\text{(a)} \qquad A^\dagger \mathbf{a} + P_{N(A)}(P_{N(A)} + P_{N(B)})^\dagger (B^\dagger \mathbf{b} - A^\dagger \mathbf{a}) + N(A) \cap N(B)$$

$$\text{(a}'\text{)} \qquad = B^\dagger \mathbf{b} - P_{N(B)}(P_{N(A)} + P_{N(B)})^\dagger (B^\dagger \mathbf{b} - A^\dagger \mathbf{a}) + N(A) \cap N(B)$$

$$\text{(b)} \qquad = (A^\dagger A + B^\dagger B)^\dagger (A^\dagger \mathbf{a} + B^\dagger \mathbf{b}) + N(A) \cap N(B).$$

Proof. Follows from Theorem 5 by substituting

$$\mathbf{f} = A^\dagger \mathbf{a}, \quad L = N(A), \quad \mathbf{g} = B^\dagger \mathbf{b}, \quad M = N(B)\ . \tag{70}$$

Thus (69), (a), and (a′) follow directly from (51), (a), and (a′) of Theorem 5, respectively, by using (70).

That (b) follows from Theorem 5(b) or 5(b′) is proved as follows. Substituting (70) in Theorem 5(b) gives

$$\begin{aligned}\{A^\dagger \mathbf{a} + N(A)\} &\cap \{B^\dagger \mathbf{b} + N(B)\} \\ &= A^\dagger \mathbf{a} + (A^\dagger A + B^\dagger B)^\dagger B^\dagger B (B^\dagger \mathbf{b} - A^\dagger \mathbf{a}) + N(A) \cap N(B) \\ &= (A^\dagger - (A^\dagger A + B^\dagger B)^\dagger B^\dagger B A^\dagger)\mathbf{a} + (A^\dagger A + B^\dagger B)^\dagger B^\dagger \mathbf{b} \\ &\quad + N(A) \cap N(B)\ ,\end{aligned} \tag{71}$$

since $P_{N(X)^\perp} = P_{R(X^*)} = X^\dagger X$ for $X = A, B$.

Now $R(A^\dagger) = R(A^*) \subset R(A^*) + R(B^*)$ and therefore

$$A^\dagger = (A^\dagger A + B^\dagger B)^\dagger (A^\dagger A + B^\dagger B) A^\dagger$$

by Lemma 2, from which it follows that

$$A^\dagger - (A^\dagger A + B^\dagger B)^\dagger B^\dagger B A^\dagger = (A^\dagger A + B^\dagger B)^\dagger A^\dagger\ ,$$

which when substituted in (71) gives (b). □

Since each of the parts (a), (a′), and (b) of Corollary 3 gives the solutions of the partitioned equation (68), these expressions can be used to obtain the generalized inverses of partitioned matrices.

Theorem 6. (Ben–Israel [**112**], Katz [**825**], Mihalyffy [**1048**]). Let A and B be matrices with n columns. Then each of the following expressions is a $\{1, 2, 4\}$–inverse of the partitioned matrix $\begin{bmatrix} A \\ B \end{bmatrix}$:

$$\text{(a)} \qquad X = \begin{bmatrix} A^\dagger & O \end{bmatrix} + P_{N(A)}(P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \tag{72}$$

$$\text{(a}'\text{)} \qquad Y = \begin{bmatrix} O & B^\dagger \end{bmatrix} - P_{N(B)}(P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \tag{73}$$

$$\text{(b)} \qquad Z = (A^\dagger A + B^\dagger B)^\dagger \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix}\ . \tag{74}$$

Moreover, if

$$R(A^*) \cap R(B^*) = \{\mathbf{0}\}\ , \tag{75}$$

then each of the expressions (72), (73), (74) is the Moore–Penrose inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$.

PROOF. From Corollary 3 it follows that whenever

$$\begin{bmatrix} A \\ B \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} , \tag{68}$$

is consistent, then $X \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$, $Y \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$, and $Z \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$ are among its solutions. Also the representations (43a) and (43b) are orthogonal, and therefore, by Corollary 1, the representations (a), (a′) and (b) of Corollary 3 are also orthogonal. Thus $X \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$, $Y \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$, and $Z \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$ are all perpendicular to

$$N(A) \cap N(B) = N \begin{bmatrix} A \\ B \end{bmatrix} .$$

By Theorem 3.2, it follows therefore that X, Y, and Z are $\{1,4\}$–inverses of $\begin{bmatrix} A \\ B \end{bmatrix}$.

We show now that X, Y, and Z are $\{2\}$–inverses of $\begin{bmatrix} A \\ B \end{bmatrix}$.

(a) From (72) we get

$$X \begin{bmatrix} A \\ B \end{bmatrix} = A^\dagger A + P_{N(A)}(P_{N(A)} + P_{N(B)})^\dagger(-A^\dagger A + B^\dagger B) .$$

But

$$-A^\dagger A + B^\dagger B) = P_{N(A)} - P_{N(B)} = (P_{N(A)} + P_{N(B)}) - 2P_{N(B)} .$$

Therefore, by Lemma 2 and Theorem 3,

$$\begin{aligned} X \begin{bmatrix} A \\ B \end{bmatrix} &= A^\dagger A + P_{N(A)} P_{N(A)+N(B)} - P_{N(A)\cap N(B)} \\ &= A^\dagger A + P_{N(A)} - P_{N(A)\cap N(B)} \quad (\text{since } N(A) \subset N(A) + N(B)) \\ &= I_n - P_{N(A)\cap N(B)} \quad (\text{since } P_{N(A)} = I - A^\dagger A) . \end{aligned} \tag{76}$$

Since $R(H^\dagger) = R(H^*) = N(H)^\perp$ for $H = A, B$,

$$P_{N(A)\cap N(B)} A^\dagger = O , \quad P_{N(A)\cap N(B)} B^\dagger = O , \tag{77}$$

and therefore (76) gives

$$X \begin{bmatrix} A \\ B \end{bmatrix} X = X - P_{N(A)\cap N(B)} P_{N(A)} (P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} .$$

Since

$$P_{N(A)\cap N(B)} = P_{N(A)\cap N(B)} P_{N(A)} = P_{N(A)\cap N(B)} P_{N(B)} ,$$

$$\begin{aligned} X \begin{bmatrix} A \\ B \end{bmatrix} X &= X - \frac{1}{2} P_{N(A)\cap N(B)} (P_{N(A)} + P_{N(B)})(P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \\ &= X - \frac{1}{2} P_{N(A)\cap N(B)} P_{N(A)+N(B)} \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \quad (\text{by Lemma 2}) \\ &= X - \frac{1}{2} P_{N(A)\cap N(B)} \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \ (\text{since } N(A)\cap N(B) \subset N(A) + N(B)) \\ &= X \quad (\text{by (77)}) . \end{aligned}$$

(a′) That Y given by (73) is a $\{2\}$–inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$ is similarly proved.

(b) The proof that Z given by (74) is a $\{2\}$–inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$ is easy since

$$Z \begin{bmatrix} A \\ B \end{bmatrix} = (A^\dagger A + B^\dagger B)^\dagger (A^\dagger A + B^\dagger B)$$

and therefore

$$\begin{aligned} Z \begin{bmatrix} A \\ B \end{bmatrix} Z &= (A^\dagger A + B^\dagger B)^\dagger (A^\dagger A + B^\dagger B)(A^\dagger A + B^\dagger B)^\dagger \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix} \\ &= (A^\dagger A + B^\dagger B)^\dagger \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix} \\ &= Z \ . \end{aligned}$$

Finally, we show that (75) implies that X, Y, and Z given by (72), (73), and (74) respectively, are $\{3\}$–inverses of $\begin{bmatrix} A \\ B \end{bmatrix}$. Indeed (75) is equivalent to

$$N(A) + N(B) = \mathbb{C}^n \ , \tag{78}$$

since $N(A) + N(B) = \{R(A^*) \cap R(B^*)\}^\perp$ by Ex. 11(b).

(a) From (72) it follows that

$$BX = \begin{bmatrix} BA^\dagger & O \end{bmatrix} + BP_{N(A)}(P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \ . \tag{79}$$

But

$$(P_{N(A)} + P_{N(B)})(P_{N(A)} + P_{N(B)})^\dagger = I_n \tag{80}$$

by (78) and Lemma 2. Therefore

$$P_{N(A)}(P_{N(A)} + P_{N(B)})^\dagger = B(P_{N(A)} + P_{N(B)} - P_{N(A)})(P_{N(A)} + P_{N(B)})^\dagger = B \ ,$$

and so (79) becomes

$$BX = \begin{bmatrix} O & BB^\dagger \end{bmatrix} \ .$$

Consequently,

$$\begin{bmatrix} A \\ B \end{bmatrix} X = \begin{bmatrix} AA^\dagger & O \\ O & BB^\dagger \end{bmatrix} \ ,$$

which proves that X is a $\{3\}$–inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$.

(a′) That Y given by (73) is a $\{3\}$–inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$ whenever (75) holds is similarly proved, or, alternatively, (72) and (73) give

$$\begin{aligned} Y - X &= \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} - (P_{N(A)} + P_{N(B)})(P_{N(A)} + P_{N(B)})^\dagger \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \\ &= O \ , \quad \text{by (80)} \ . \end{aligned}$$

(b) Finally we show that Z is the Moore–Penrose inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$ when (75) holds. By Ex. 2.29, the Moore–Penrose inverse of any matrix H is the only $\{1,2\}$–inverse U such that $R(U) = R(H^*)$ and $N(U) = N(H^*)$. Thus, $H^\dagger$ is also the unique matrix $U \in H\{1,2,4\}$ such that $N(H^*) \subset N(U)$. Now, Z has already been shown to be a $\{1,2,4\}$–inverse of $\begin{bmatrix} A \\ B \end{bmatrix}$, and it therefore suffices to prove that

$$N(\begin{bmatrix} A^* & B^* \end{bmatrix}) \subset N(Z) \ . \tag{81}$$

Let $\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} \in N(\begin{bmatrix}A^* & B^*\end{bmatrix})$. Then

$$A^*\mathbf{u} + B^*\mathbf{v} = \mathbf{0}\ ,$$

and therefore

$$A^*\mathbf{u} = -B^*\mathbf{v} = \mathbf{0}\ , \tag{82}$$

since, by (75), the only vector common to $R(A^*)$ and $R(B^*)$ is the zero vector. Since $N(H^\dagger) = N(H^*)$ for any H, (82) gives

$$A^\dagger\mathbf{u} = B^\dagger\mathbf{v} = \mathbf{0}\ ,$$

and therefore by (74), $Z\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} = \mathbf{0}$. Thus (81) is established, and the proof is complete. $\square$

If a matrix is partitioned by columns instead of by rows, then Theorem 6 may still be used. Indeed,

$$\begin{bmatrix}A & B\end{bmatrix} = \begin{bmatrix}A^*\\ B^*\end{bmatrix}^* \tag{83}$$

permits using Theorem 6 to obtain generalized inverses of $\begin{bmatrix}A^*\\ B^*\end{bmatrix}$, which is partitioned by rows, and then translating the results to the matrix $\begin{bmatrix}A & B\end{bmatrix}$, partitioned by columns.

In working with the conjugate transposes of a matrix, we note that

$$\begin{aligned} X \in A\{i\} &\Longleftrightarrow X^* \in A^*\{i\}\ , \quad (i = 1, 2)\ ,\\ X \in A\{3\} &\Longleftrightarrow X^* \in A^*\{4\}\ ,\\ X \in A\{4\} &\Longleftrightarrow X^* \in A^*\{3\}\ . \end{aligned} \tag{84}$$

Applying Theorem 6 to $\begin{bmatrix}A^*\\ B^*\end{bmatrix}$ as in (83), and using (84), we obtain the following.

Corollary 4. Let A and B be matrices with n rows. Then each of the following expressions is a $\{1, 2, 3\}$–inverse of the partitioned matrix $\begin{bmatrix}A & B\end{bmatrix}$:

(a) $$X = \begin{bmatrix}A^\dagger\\ O\end{bmatrix} + \begin{bmatrix}-A^\dagger\\ B^\dagger\end{bmatrix}(P_{N(A^*)} + P_{N(B^*)})^\dagger P_{N(A^*)}\ , \tag{85}$$

(a′) $$Y = \begin{bmatrix}O\\ B^\dagger\end{bmatrix} - \begin{bmatrix}-A^\dagger\\ B^\dagger\end{bmatrix}(P_{N(A^*)} + P_{N(B^*)})^\dagger P_{N(B^*)}\ , \tag{86}$$

(b) $$Z = \begin{bmatrix}A^\dagger\\ B^\dagger\end{bmatrix}(AA^\dagger + BB^\dagger)^\dagger\ . \tag{87}$$

Moreover, if

$$R(A) \cap R(B) = \{\mathbf{0}\}\ , \tag{88}$$

then each of the expressions (85), (86), (87) is the Moore–Penrose inverse of $\begin{bmatrix}A & B\end{bmatrix}$. $\square$

Other and more general results on Moore–Penrose inverses of partitioned matrices were given in Cline [**350**]. However, these results are too formidable for reproduction here.

Exercises and examples.

Ex. 15. Let the partitioned matrix $\begin{bmatrix} A \\ B \end{bmatrix}$ be nonsingular. Then

$$\begin{aligned}
&\text{(a)} & \begin{bmatrix} A \\ B \end{bmatrix}^{-1} &= \begin{bmatrix} A^\dagger & O \end{bmatrix} + P_{N(A)}(P_{N(A)} + P_{N(B)})^{-1} \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \\
&\text{(a}'\text{)} & &= \begin{bmatrix} O & B^\dagger \end{bmatrix} - P_{N(B)}(P_{N(A)} + P_{N(B)})^{-1} \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \\
&\text{(b)} & &= (A^\dagger A + B^\dagger B)^{-1} \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix} .
\end{aligned}$$

PROOF. Follows from Theorem 6. Indeed the nonsingularity of $\begin{bmatrix} A \\ B \end{bmatrix}$ guarantees that (75) is satisfied, and also that the matrices $P_{N(A)}+P_{N(B)}$ and $A^\dagger A+B^\dagger B = P_{R(A^*)}+P_{R(B^*)}$ are nonsingular. □

Ex. 16. Let $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \end{bmatrix}$. Then $\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is nonsingular. We calculate now its inverse using Ex. 15(b).

Here

$$A^\dagger = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} , \qquad A^\dagger A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ,$$

$$B^\dagger = \frac{1}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} , \qquad B^\dagger B = \frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} ,$$

$$A^\dagger A + B^\dagger B = \frac{1}{10}\begin{bmatrix} 7 & 9 \\ 9 & 13 \end{bmatrix} , \qquad (A^\dagger A + B^\dagger B)^{-1} = \begin{bmatrix} 13 & -9 \\ -9 & 7 \end{bmatrix} ,$$

and finally,

$$\begin{aligned}
\begin{bmatrix} A \\ B \end{bmatrix}^{-1} &= (A^\dagger A + B^\dagger B)^{-1} \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix} \\
&= \begin{bmatrix} 13 & -9 \\ -9 & 7 \end{bmatrix} \frac{1}{10} \begin{bmatrix} 7 & 9 \\ 9 & 13 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} .
\end{aligned}$$

Ex. 17. *Series expansion.* Let the partitioned matrix $\begin{bmatrix} A \\ B \end{bmatrix}$ be nonsingular. Then

$$A^\dagger A + B^\dagger B = I + K , \tag{89}$$

where K is Hermitian and

$$\|K\| < 1 . \tag{90}$$

From (89) and (90) it follows that

$$(A^\dagger A + B^\dagger B)^{-1} = \sum_{j=0}^{\infty} (-1)^j K^j , \tag{91}$$

Substituting (91) in Ex. 15(b) gives

$$\begin{bmatrix} A \\ B \end{bmatrix}^{-1} = \sum_{j=0}^{\infty} (-1)^j K^j \begin{bmatrix} A^\dagger & B^\dagger \end{bmatrix} . \tag{92}$$

Similarly,

$$P_{N(A)} + P_{N(B)} = I - A^\dagger A + I - B^\dagger B$$
$$= I - K \ , \quad \text{with } K \text{ as in (89)}$$

and therefore

$$(P_{N(A)} + P_{N(B)})^{-1} = \sum_{j=0}^{\infty} K^j \ . \tag{93}$$

Substituting (93) in Ex. 15(a) gives

$$\begin{bmatrix} A \\ B \end{bmatrix}^{-1} = \begin{bmatrix} A^\dagger & O \end{bmatrix} + (I - A^\dagger A) \sum_{j=0}^{\infty} K^j \begin{bmatrix} -A^\dagger & B^\dagger \end{bmatrix} \ . \tag{94}$$

Ex. 18. Let the partitioned matrix $\begin{bmatrix} A \\ B \end{bmatrix}$ be nonsingular. Then the solution of

$$\begin{bmatrix} A \\ B \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \ , \tag{68}$$

for any given $\mathbf{a}$ and $\mathbf{b}$ is

$$\mathbf{x} = \sum_{j=0}^{\infty} (-1)^j K^j (A^\dagger \mathbf{a} + B^\dagger \mathbf{b}) \tag{95}$$

$$= A^\dagger \mathbf{a} + (I - A^\dagger A) \sum_{j=0}^{\infty} K^j (B^\dagger \mathbf{b} - A^\dagger \mathbf{a}) \ , \tag{96}$$

with K given by (89).

Proof. Use (92) and (94). □

Remark. If the nonsingular matrix $\begin{bmatrix} A \\ B \end{bmatrix}$ is ill–conditioned, then slow convergence may be expected in (91) and (93), and hence in (92) and (94). Even then the convergence of (95) or (96) may be reasonable for certain vectors $\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$. Thus for example, if $\|B^\dagger \mathbf{b} - A^\dagger \mathbf{a}\|$ is sufficiently small, then (96) may be reasonably approximated by its first few terms.

Ex. 19. *Common solutions for n matrix equations.* For each $i \in \overline{1,n}$ let the matrices $A_i \in \mathbb{C}^{p\times q}$, $B_i \in \mathbb{C}^{p\times r}$ be given, and consider the n matrix equations

$$A_i X = B_i \quad i \in \overline{1,n} \ . \tag{97}$$

For $k \in \overline{1,n}$ define recursively

$$\begin{aligned} C_k &= A_k F_{k-1} \ , & D_k &= B_k - A_k E_{k-1} \ , \\ E_k &= E_{k-1} + F_{k-1} C_k^\dagger D_k \text{ and} & F_k &= F_{k-1}(I - C_k^\dagger C_k) \ , \end{aligned} \tag{98}$$

where

$$E_0 = O_{q\times r} \ , \qquad F_0 = I_q \ .$$

Then the n matrix equations (97) have a common solution if and only if

$$C_i C_i^\dagger D_i = D_i \quad i \in \overline{1,n} \ , \tag{99}$$

in which case the general common solution of (97) is

$$X = E_n + F_n Z \;, \tag{100}$$

where $Z \in \mathbb{C}^{q\times r}$ is arbitrary (Morris and Odell [**1096**]).

Ex. 20. (Morris and Odell [**1096**]). For $i \in \overline{1,n}$ let $A_i \in \mathbb{C}^{1\times q}$, and let C_i be defined by (98) for $i \in \overline{1,n}$. Let the vectors $\{A_1, A_2, \dots, A_k\}$ be linearly independent. Then the vectors $\{A_1, A_2, \dots, A_{k+1}\}$ are linearly independent if and only if $C_{k+1} = O$.

Ex. 21. (Morris and Odell [**1096**]). For $i \in \overline{1,n}$ let A_i, C_i be as in Ex. 20. For any $k \le n$ the vectors $\{C_1, C_2, \dots, C_k\}$ are orthogonal and span the subspace spanned by $\{A_1, A_2, \dots, A_k\}$.

5. Greville's method and related results

Greville's method for computing the Moore–Penrose inverse $A^\dagger$ of a matrix $A \in \mathbb{C}^{m\times n}$ is a finite iterative method. The main variant of this method, described in Theorem 7 below, uses n iterations. At the kth iteration ($k = 1, 2, \dots, n$) it computes $A_k^\dagger$, where A_k is the submatrix of A consisteing of its first k columns.

First we need some notation. For $k = 2, \dots, n$ the matrix A_k is partitioned as

$$A_k = \begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix} \tag{101}$$

where $\mathbf{a}_k$ is the kth column of A. For $k = 2, \dots, n$ let the vectors $\mathbf{d}_k$ and $\mathbf{c}_k$ be defined by

$$\mathbf{d}_k = A_{k-1}^\dagger \mathbf{a}_k \tag{102}$$

$$\begin{aligned} \mathbf{c}_k &= \mathbf{a}_k - A_{k-1}\mathbf{d}_k \qquad (103)\\ &= \mathbf{a}_k - A_{k-1}A_{k-1}^\dagger \mathbf{a}_k \\ &= \mathbf{a}_k - P_{R(A_{k-1})}\mathbf{a}_k \\ &= P_{N(A_{k-1}^*)}\mathbf{a}_k \;. \end{aligned}$$

Theorem 7. (Greville [**580**]). Let $A \in \mathbb{C}^{m\times n}$. Using the above notation, the Moore–Penrose inverse of A_k ($k = 2, \dots, n$) is

$$\begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix}^\dagger = \begin{bmatrix} A_{k-1}^\dagger - \mathbf{d}_k \mathbf{b}_k^* \\ \mathbf{b}_k^* \end{bmatrix} \;, \tag{104}$$

where

$$\mathbf{b}_k^* = \mathbf{c}_k^\dagger \quad \text{if } \mathbf{c}_k \ne \mathbf{0} \;, \tag{105}$$

$$\mathbf{b}_k^* = (1 + \mathbf{d}_k^*\mathbf{d}_k)^{-1}\mathbf{d}_k^* A_{k-1}^\dagger \quad \text{if } \mathbf{c}_k = \mathbf{0} \;. \tag{106}$$

Proof. Let $A_k^\dagger = \begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix}^\dagger$ be partitioned as

$$A_k^\dagger = \begin{bmatrix} B_k \\ \mathbf{b}_k^* \end{bmatrix} \tag{107}$$

where $\mathbf{b}_k^*$ is the kth row of $A_k^\dagger$. Multiplying (101) and (107) gives

$$A_k A_k^\dagger = A_{k-1}B_k + \mathbf{a}_k \mathbf{b}_k^* \;. \tag{108}$$

Now by (101), Ex. 2.29 and Corollary 2.7

$$N(A_{k-1}^\dagger) = N(A_{k-1}^*) \supset N(A_k^*) = N(A_k^\dagger) = N(A_k A_k^\dagger) \;,$$

and it follows from Ex. 2.20 that

$$A_{k-1}^\dagger A_k A_k^\dagger = A_{k-1}^\dagger \;. \tag{109}$$

Moreover, since

$$R(A_k^\dagger) = R(A_k^*)$$

by Ex. 2.29, it follows from (101), (107), and Corollary 2.7 that

$$R(B_k) \subset R(A_{k-1}^*) = R(A_{k-1}^\dagger) = R(A_{k-1}^\dagger A_{k-1}) \ ,$$

and therefore

$$A_{k-1}^\dagger A_{k-1} B_k = B_k \tag{110}$$

by Ex. 2.20. It follows from (109) and (110) that premultiplication of (108) by $A_{k-1}^\dagger$ gives

$$\begin{aligned} A_{k-1}^\dagger &= B_k + A_{k-1}^\dagger \mathbf{a}_k \mathbf{b}_k^* \\ &= B_k + \mathbf{d}_k \mathbf{b}_k^* \ , \end{aligned} \tag{111}$$

by (102). Thus we may write

$$\begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix}^\dagger = \begin{bmatrix} A_{k-1}^\dagger - \mathbf{d}_k \mathbf{b}_k^* \\ \mathbf{b}_k^* \end{bmatrix} \ , \tag{104}$$

with $\mathbf{b}_k^*$ still to be determined. We distinguish two cases according as $\mathbf{a}_k$ is or is not in $R(A_{k-1})$, i.e., according as $\mathbf{c}_k$ is or is not $\mathbf{0}$.

Case I ($\mathbf{c}_k \neq \mathbf{0}$)

By using (111), (108) becomes

$$\begin{aligned} A_k A_k^\dagger &= A_{k-1} A_{k-1}^\dagger + (\mathbf{a}_k - A_{k-1}\mathbf{d}_k)\mathbf{b}_k^* \\ &= A_{k-1} A_{k-1}^\dagger + \mathbf{c}_k \mathbf{b}_k^* \end{aligned} \tag{112}$$

by (103). Since $A_k A_k^\dagger$ is Hermitian, it follows from (112) that $\mathbf{c}_k \mathbf{b}_k^*$ is Hermitian, and therefore

$$\mathbf{b}_k^* = \delta \mathbf{c}_k^* \ , \tag{113}$$

where δ is some real number. From (101) and (103) we obtain

$$A_k = A_k A_k^\dagger A_k = \begin{bmatrix} A_{k-1} + \mathbf{c}_k \mathbf{b}_k^* A_{k-1} & \mathbf{a}_k - \mathbf{c}_k + (\mathbf{b}_k^* \mathbf{a}_k)\mathbf{c}_k \end{bmatrix} \ ,$$

and comparison with (101) shows that

$$\mathbf{b}_k^* \mathbf{a}_k = 1 \ , \tag{114}$$

since $\mathbf{c}_k \neq \mathbf{0}$. Now, by (103),

$$\mathbf{c}_k = P\mathbf{a}_k \ ,$$

where P denotes the orthogonal projector on $N(A_{k-1}^*)$. Therefore, (113) and (114) give

$$\begin{aligned} 1 &= \mathbf{b}_k^* \mathbf{a}_k = \delta \mathbf{c}_k^* \mathbf{a}_k = \delta \mathbf{a}_k^* P \mathbf{a}_k \\ &= \delta \mathbf{a}_k^* P^2 \mathbf{a}_k = \delta \mathbf{c}_k^* \mathbf{c}_k \ , \end{aligned} \tag{115}$$

since P is idempotent. By (113), (115), aand Ex. 1.17(a)

$$\mathbf{b}_k^* = \delta \mathbf{c}_k^* = \mathbf{c}_k^\dagger \ .$$

Case II ($\mathbf{c}_k = \mathbf{0}$)

Here $R(A_k) = R(A_{k-1})$, and so, by (107) and (2.48),

$$\begin{aligned} N(\mathbf{b}_k^*) \supset N(A_k^\dagger) = N(A_k^*) &= N(A_{k-1}^*) = N(A_{k-1}^\dagger) \\ &= N(A_{k-1} A_{k-1}^\dagger) \ . \end{aligned}$$

Therefore, by Ex. 2.20,

$$\mathbf{b}_k^* A_{k-1} A_{k-1}^\dagger = \mathbf{b}_k^* \,. \tag{116}$$

Now, (101) and (104) give

$$A_k^\dagger A_k = \begin{bmatrix} A_{k-1}^\dagger - \mathbf{d}_k \mathbf{b}_k^* A_{k-1} & (1-\alpha)\mathbf{d}_k \\ \mathbf{b}_k^* A_{k-1} & \alpha \end{bmatrix} , \tag{117}$$

where

$$\alpha = \mathbf{b}_k^* \mathbf{a}_k \tag{118}$$

is a scalar (real, in fact, since it is a diagonal element of a Hermitian matrix). Since (117) is Hermitian we have

$$\mathbf{b}_k^* A_{k-1} = (1-\alpha)\mathbf{d}_k^* \,.$$

Thus, by (116),

$$\mathbf{b}_k^* = \mathbf{b}_k^* A_{k-1} A_{k-1}^\dagger = (1-\alpha)\mathbf{d}_k^* A_{k-1}^\dagger \,. \tag{119}$$

Substitution of (119) in (118) gives

$$\alpha = (1-\alpha)\mathbf{d}_k^* \mathbf{d}_k \,, \tag{120}$$

by (102). Adding $1-\alpha$ to both sides of (120) gives

$$(1-\alpha)(1+\mathbf{d}_k^* \mathbf{d}_k) = 1$$

and substitution for $1-\alpha$ in (119) gives (106). □

Greville's method as described above, thus computes $A^\dagger$ recursively in terms of $A_k^\dagger$ ($k = 1, 2, \dots, n$). This method was adapted by Greville [**580**] for the computation of $A^\dagger \mathbf{y}$, for any $\mathbf{y} \in \mathbb{C}^m$, without computing $A^\dagger$. This is done as follows:

Let

$$\widetilde{A} = \begin{bmatrix} A & \mathbf{y} \end{bmatrix} \,. \tag{121}$$

Then (104) gives

$$A_k^\dagger \widetilde{A} = \begin{bmatrix} A_{k-1}^\dagger \widetilde{A} - \mathbf{d}_k \mathbf{b}_k^* \widetilde{A} \\ \mathbf{b}_k^* \widetilde{A} \end{bmatrix} \,. \tag{122}$$

By (102) it follows that $\mathbf{d}_k$ is the kth column of $A_{k-1}^\dagger \widetilde{A}$ for $k = 2, \dots, n$. Therefore only the vector $\mathbf{b}_k^* \widetilde{A}$ is needed to get $A_k^\dagger \widetilde{A}$ from $A_{k-1}^\dagger \widetilde{A}$ by (122).

If $\mathbf{c}_k = \mathbf{0}$, then (106) gives $\mathbf{b}_k^* \widetilde{A}$ as

$$\mathbf{b}_k^* \widetilde{A} = (1 + \mathbf{d}_k^* \mathbf{d}_k)^{-1} \mathbf{d}_k^* A_{k-1}^\dagger \widetilde{A} \qquad (\mathbf{c}_k = \mathbf{0}) \,. \tag{123}$$

If $\mathbf{c}_k \neq \mathbf{0}$, then from (105)

$$\mathbf{b}_k^* \widetilde{A} = (\mathbf{c}_k^* \mathbf{c}_k)^{-1} \mathbf{c}_k^* \widetilde{A} \qquad (\mathbf{c}_k \neq \mathbf{0}) \,. \tag{124}$$

The computation of (124) is simplified by noting that the kth element of the vector $\mathbf{c}_k^* \widetilde{A}$ is $\mathbf{c}_k^* \mathbf{a}_k$ ($k = 1, 2, \dots, n$). Premultiplying (103) by $\mathbf{c}_k^*$ we obtain

$$\mathbf{c}_k^* \mathbf{c}_k = \mathbf{c}_k^* \mathbf{a}_k \,, \tag{125}$$

since $\mathbf{c}_k^* A_{k-1} = \mathbf{0}$ by (103). Thus the vector (124) may be computed by computing $\mathbf{c}_k^* \widetilde{A}$ and normalizing it by dividing by its kth element. In the Greville method as described above, the matrix to be inverted is modified at each iteration by adjoining an additional column. This is the natural approach to some applications. Consider, for example, the *least–squares polynomial approximation problem* where a real function $y(t)$ is to be approximated by polynomials $\sum_{j=0}^{k} x_j t^j$. In the discrete version of this problem, the function $y(t)$ is represented by the m–dimensional vector

$$\mathbf{y} = \big[y_i\big] = \big[y(t_i)\big] \quad (i = 1, \dots, m) , \tag{126}$$

whose ith component is the function y evaluated at $t = t_i$, where the points $t_1, t_2, \dots, t_m$ are given. Similarly, the polynomial t^j $(j = 0, 1, \dots)$ is represented by the m–dimensional vector

$$\mathbf{a}_{j+1} = \big[a_{i,j+1}\big] = \big[(t_i)^j\big] \quad (i = 1, \dots, m) . \tag{127}$$

The problem is, therefore, for a given approximation error $\epsilon > 0$ to find an integer $k = k(\epsilon)$ and a vector $\mathbf{x} \in \mathbb{R}^{k-1}$ such that

$$\|A_{k-1}\mathbf{x} - \mathbf{y}\| \le \epsilon , \tag{128}$$

where $\mathbf{y}$ is given by (126) and $A_{k-1} \in \mathbb{R}^{m\times(k-1)}$ is the matrix

$$A_{k-1} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_{k-1} \end{bmatrix} \tag{129}$$

for $\mathbf{a}_j$ given by (127). For any k, the Euclidean norm $\|A_{k-1}\mathbf{x} - \mathbf{y}\|$ is minimized by

$$\mathbf{x} = A_{k-1}^{\dagger}\mathbf{y} . \tag{130}$$

If for a given k, the vector (130) does not satisfy (128), i.e., if

$$\|A_{k-1}A_{k-1}^{\dagger}\mathbf{y} - \mathbf{y}\| > \epsilon , \tag{131}$$

then we try achieving (128) with the matrix

$$A_k = \begin{bmatrix} A_{k-1} & \mathbf{a}_k \end{bmatrix} , \tag{101}$$

where, in effect, the degree of the approximating polynomial has been increased from $k-2$ to $k-1$. Greville's method described above computes $A_k^{\dagger}\mathbf{y}$ in terms of $A_{k-1}^{\dagger}\mathbf{y}$, and is thus the natural method for solving the above polynomial approximation problem and similar problems in approximation and regression.

There are applications on the other hand which call for modifying the matrix to be inverted by adjoining additional rows. Consider, for example, the problem of solving (or approximating the solution of) the following linear equation:

$$\sum_{j=1}^{n} A_{ij}x_j = y_i \quad (i = 1, \dots, k-1) , \tag{132}$$

where n is fixed and the data $\{a_{ij}, y_i : i = 1, \dots, k-1,\ j = 1, \dots, n\}$ are the result of some experiment or observation repeated $k-1$ times, with the row

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} & y_i \end{bmatrix} \quad (i = 1, \dots, k-1)$$

the result of the ith experiment.

Let $\widehat{x}_{k-1}$ be the least–squares solution of (132), i.e.,

$$\widehat{x}_{k-1} = A_{(k-1)}^{\dagger}\mathbf{y}_{(k-1)} , \tag{133}$$

where $A_{(k-1)} = \big[a_{ij}\big]$ and $\mathbf{y}_{(k-1)} = \big[y_i\big]$, $i = 1, \dots, k-1$; $j = 1, \dots, n$. If the results of an additional experiment or observation become available after (132) is solved, then it is necessary to update the solution (133) in light of the additional information. This explains the need for the variant of Greville's method described in Corollaries 5 and 6 below, for which some notation is needed.

Let n be fixed and let $A_{(k)} \in \mathbb{C}^{k\times n}$ be partitioned as

$$A_{(k)} = \begin{bmatrix} A_{(k-1)} \\ \mathbf{a}_k^* \end{bmatrix} , \quad \mathbf{a}_k^* \in \mathbb{C}^{1\times n} . \tag{134}$$

Also, in analogy with (102) and (103), let

$$\mathbf{d}_k^* = \mathbf{a}_k^* A_{(k-1)}^\dagger , \tag{135}$$

$$\mathbf{c}_k^* = \mathbf{a}_k^* - \mathbf{d}_k^* A_{(k-1)} . \tag{136}$$

COROLLARY 5. (Kishi [**859**]). Using the above notation

$$A_{(k)}^\dagger = \begin{bmatrix} A_{(k-1)}^\dagger - \mathbf{b}_k \mathbf{d}_k^* & \mathbf{b}_k \end{bmatrix} , \tag{137}$$

where

$$\mathbf{b}_k = \mathbf{c}_k^{*\dagger} , \quad \text{if } \mathbf{c}_k^* \neq \mathbf{0} \tag{138}$$

$$\mathbf{b}_k = (1 + \mathbf{d}_k^* \mathbf{d}_k)^{-1} A_{(k-1)}^\dagger \mathbf{d}_k , \quad \text{if } \mathbf{c}_k^* = \mathbf{0} . \tag{139}$$

PROOF. Follows by applying Theorem 7 to the conjugate transpose of the matrix (134). □

In some applications it is necessary to compute

$$\widehat{x}_k = A_{(k)}^\dagger \mathbf{y}_{(k)} \quad \text{for given } \mathbf{y}_{(k)} \in \mathbb{C}^k$$

but $A_{(k)}^\dagger$ is not needed. Then $\widehat{x}_k$ may be obtained from $\widehat{x}_{k-1}$ very simply, as follows.

COROLLARY 6. (Albert and Sittler [**16**]). Let the vector $\mathbf{y}_{(k)} \in \mathbb{C}^k$ be partitioned as

$$\mathbf{y}_{(k)} = \begin{bmatrix} \mathbf{y}_{(k-1)} \\ y_k \end{bmatrix} , \quad y_k \in \mathbb{C} , \tag{140}$$

and let

$$\widehat{x}_k = A_{(k)}^\dagger \mathbf{y}_{(k)} , \quad \widehat{x}_{k-1} = A_{(k-1)}^\dagger \mathbf{y}_{(k-1)} \tag{141}$$

using the notation (134). Then

$$\widehat{x}_k = \widehat{x}_{k-1} + (y_k - \mathbf{a}_k^* \widehat{x}_{k-1}) \mathbf{b}_k , \tag{142}$$

with $\mathbf{b}_k$ given by (138) or (139).

PROOF. Follows directly from Corollary 5. □

Exercises and examples.

EX. 22. *A converse of Theorem 7.* Let the matrix $A_{k-1} \in \mathbb{C}^{m\times(k-1)}$ be obtained from $A_k \in \mathbb{C}^{m\times k}$ by deleting its kth column $\mathbf{a}_k$. If A_k is of full column rank

$$\begin{bmatrix} A_{(k-1)}^\dagger \\ \mathbf{0}^T \end{bmatrix} = A_k^\dagger - \frac{A_k^\dagger \mathbf{b}_k \mathbf{b}_k^*}{\mathbf{b}_k^* \mathbf{b}_k} , \tag{143}$$

where $\mathbf{b}_k^*$ is the last row of $A_k^\dagger$ (Fletcher [**496**]).

EX. 23. Let

$$A_2 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & -1 \end{bmatrix} .$$

Then

$$A_1^\dagger = \mathbf{a}_1^\dagger = \begin{bmatrix}1\\2\\0\end{bmatrix}^\dagger = \frac{1}{5}\begin{bmatrix}1 & 2 & 0\end{bmatrix},$$

$$\mathbf{d}_2 = A_1^\dagger \mathbf{a}_2 = \frac{1}{5}\begin{bmatrix}1 & 2 & 0\end{bmatrix}\begin{bmatrix}0\\1\\-1\end{bmatrix} = \frac{2}{5},$$

$$\mathbf{c}_2 = \mathbf{a}_2 - A_1\mathbf{d}_2 = \begin{bmatrix}0\\1\\-1\end{bmatrix} - \begin{bmatrix}1\\2\\0\end{bmatrix}\frac{2}{5} = \begin{bmatrix}-\frac{2}{5}\\ \frac{1}{5}\\ -1\end{bmatrix},$$

and by (105)

$$\mathbf{b}_2^* = \mathbf{c}_2^\dagger = \frac{1}{6}\begin{bmatrix}-2 & 1 & -5\end{bmatrix}.$$

$A_2^\dagger$ is now computed by (104) as

$$\begin{aligned} A_2^\dagger &= \frac{1}{5}\begin{bmatrix}1 & 2 & 0\\0 & 0 & 0\end{bmatrix} + \begin{bmatrix}-\frac{2}{5}\\1\end{bmatrix}\begin{bmatrix}-\frac{2}{6} & \frac{1}{6} & -\frac{5}{6}\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ -\frac{1}{3} & \frac{1}{6} & -\frac{5}{6}\end{bmatrix}. \end{aligned}$$

Let now $\mathbf{a}_2^\dagger$ be computed by (143), i.e., by deleting $\mathbf{a}_1$ from A_2. Interchanging columns of A_2 and rows of $A_2^\dagger$ we obtain

$$A_2^\dagger \mathbf{b}_2 = \begin{bmatrix}-\frac{1}{3} & \frac{1}{6} & -\frac{5}{6}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}\frac{1}{3}\\ \frac{1}{3}\\ \frac{1}{3}\end{bmatrix} = \begin{bmatrix}-\frac{1}{3}\\ \frac{1}{3}\end{bmatrix}$$

and

$$\mathbf{b}_2^* \mathbf{b}_2 = \begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}\frac{1}{3}\\ \frac{1}{3}\\ \frac{1}{3}\end{bmatrix} = \frac{1}{3},$$

and finally from (143)

$$\begin{aligned} \begin{bmatrix}A_1^\dagger\\ \mathbf{0}^T\end{bmatrix} &= \begin{bmatrix}-\frac{1}{3} & \frac{1}{6} & -\frac{5}{6}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix} - \begin{bmatrix}-1\\1\end{bmatrix}\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\\ &= \begin{bmatrix}-1\\1\end{bmatrix}\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\\ &= \begin{bmatrix}0 & \frac{1}{2} & -\frac{1}{2}\\ 0 & 0 & 0\end{bmatrix}, \end{aligned}$$

or

$$\mathbf{a}_2^\dagger = \begin{bmatrix}0\\1\\-1\end{bmatrix}^\dagger = \frac{1}{2}\begin{bmatrix}0\\1\\-1\end{bmatrix} \quad \text{(Fletcher [\textbf{496}])}.$$

6. Generalized inverses of bordered matrices

Partitioning was shown above to permit working with submatrices smaller in size and better behaved (e.g., nonsingular) than the original matrix. In this section a nonsingular matrix is obtained from the original matrix by adjoining to it certain matrices. Thus from a given matrix $A \in \mathbb{C}^{m\times n}$ we obtain the matrix

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix} , \tag{144}$$

which, under certain conditions on U and V^*, is nonsingular, and from its inverse $A^\dagger$ can be read off. These ideas find applications in differential equations (Reid [**1261**]) and eigenvalue computation (Blattner [**185**]).

The following theorem is based on the results of Blattner [**185**].

THEOREM 8. Let $A \in \mathbb{C}_r^{m\times n}$ and let the matrices U and V satisfy

(a) $U \in \mathbb{C}_{(m-r)}^{m\times(m-r)}$ and the columns of U are a basis for $N(A^*)$.

(b) $V \in \mathbb{C}_{(n-r)}^{n\times(n-r)}$ and the columns of U are a basis for $N(A)$.

Then the matrix

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix} , \tag{144}$$

is nonsingular and its inverse is

$$\begin{bmatrix} A^\dagger & V^{*\dagger} \\ U^\dagger & O \end{bmatrix} . \tag{145}$$

PROOF. Premultiplying (145) by (144) gives

$$\begin{bmatrix} AA^\dagger + UU^\dagger & AV^{*\dagger} \\ V^*A^\dagger & V^*V^{*\dagger} \end{bmatrix} . \tag{146}$$

Now, $R(U) = N(A^*) = R(A)^\perp$ by assumption (a) and (2.47), and therefore

$$AA^\dagger + UU^\dagger = I_n \tag{147}$$

by Ex. 2.43. Moreover,

$$V^*A^\dagger = V^*A^\dagger AA^\dagger = V^*A^*A^{\dagger *}A^\dagger = (AV)^*A^{\dagger *}A^\dagger = O , \tag{148}$$

by (1.2), (1.4), and assumption (b), while

$$\begin{aligned} AV^{*\dagger} &= AV^{\dagger *} = A(V^\dagger VV^\dagger)^* = A(V^\dagger V^{\dagger *}V^*)^* \\ &= AVV^\dagger V^{\dagger *} = O , \end{aligned} \tag{149}$$

by (1.2), Ex. 1.16(b), (1.3), and assumption (b). Finally, V^* is of full row rank by assumption (b), and therefore

$$V^*V^{*\dagger} = I_{n-r} , \tag{150}$$

by Lemma 1.2(b). By (147)–(150), (146) reduces to I_{m+n-r}, and therefore (144) is nonsingular and (145) is its inverse. □

The next two corollaries apply Theorem 8 for the solution of linear equations.

COROLLARY 7. Let A, U, V be as in Theorem 8, let $\mathbf{b} \in \mathbb{C}^n$, and consider the linear equation

$$A\mathbf{x} = \mathbf{b} . \tag{1}$$

Then the solution $\mathbf{x}, \mathbf{y}$ of

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} . \tag{151}$$

satisfies

$$\mathbf{x} = A^{\dagger}\mathbf{b} \ , \quad \text{the minimal–norm least squares solution of (1)} \ ,$$
$$U\mathbf{y} = P_{N(A^*)}\mathbf{b} \ , \quad \text{the residual of (1)} \ .$$

COROLLARY 8. (Cramer's rule, Ben–Israel [**120**], Verghese [**1502**]). Let $A, U, V, \mathbf{b}$ be as in Corollary 7. Then the minimal–norm least–squares solution $\mathbf{x} = [x_j]$ of (1) is given by

$$x_j = \frac{\det \begin{bmatrix} A[j \leftarrow \mathbf{b}] & U \\ V^*[j \leftarrow \mathbf{0}] & O \end{bmatrix}}{\det \begin{bmatrix} A & U \\ V^* & O \end{bmatrix}} \ , \quad j \in \overline{1,n} \ . \tag{152}$$

PROOF. Apply the proof of Cramer's rule, Ex. 0.49, to (151). □

Exercises.

EX. 24. *A special case of Theorem 8.* Let $A \in \mathbb{C}_r^{m\times n}$ and let the matrices $U \in \mathbb{C}^{m\times(m-r)}$ and $V \in \mathbb{C}^{n\times(n-r)}$ satisfy

$$AV = O \ , \quad V^*V = I_{n-r} \ , \quad A^*U = O \ , \quad \text{and} \quad U^*U = I_{m-r} \ . \tag{153}$$

Then the matrix

$$\begin{bmatrix} A & U \\ V^* & O \end{bmatrix} \ , \tag{144}$$

is nonsingular and its inverse is

$$\begin{bmatrix} A^{\dagger} & V \\ U^* & O \end{bmatrix} \quad \text{(Reid [\textbf{1261}])} \ . \tag{154}$$

EX. 25. Let A, U, and V be as in Ex. 24, and let

$$\alpha = \min\{\|A\mathbf{x}\| : \mathbf{x} \in R(A^*), \ \|\mathbf{x}\| = 1\} \ , \tag{155}$$
$$\beta = \max\{\|A^{\dagger}\mathbf{y}\| : \mathbf{y} \in \mathbb{C}^n, \ \|\mathbf{y}\| = 1\} \ . \tag{156}$$

Then

$$\alpha\beta = 1 \quad \text{(Reid [\textbf{1261}])} \ . \tag{157}$$

PROOF. If $\mathbf{y} \in \mathbb{C}^n$, $\|\mathbf{y}\| = 1$, then

$$\mathbf{z} = A^{\dagger}\mathbf{y}$$

is the solution of

$$A\mathbf{z} = (I_m - UU^*)\mathbf{y} \ , \quad V^*\mathbf{z} = \mathbf{0} \ ,$$

Therefore,

$$\begin{aligned} \alpha\|A^{\dagger}\mathbf{y}\| &= \alpha\|\mathbf{z}\| \le \|A\mathbf{z}\| \quad \text{(by (155))} \\ &= \|(I_m - UU^*)\mathbf{y}\| \\ &\le \|\mathbf{y}\| \quad \text{(by Ex. 2.39 since } I_m - UU^* \text{ is an orthogonal projector)} \\ &= 1 \ , \end{aligned}$$

Therefore $\alpha\beta \le 1$. On the other hand, let

$$\mathbf{x} \in R(A^*) \ , \quad \|x\| = 1 \ ;$$

then

$$\mathbf{x} = A^{\dagger}A\mathbf{x} \ ,$$

so that

$$1 = \|\mathbf{x}\| = \|A^{\dagger}A\mathbf{x}\| \leq \beta\|A\mathbf{x}\| \ ,$$

proving that $\alpha\beta \geq 1$, and completing the proof. □

See also Exs. 6.5 and 6.7.

Ex. 26. *A generalization of Theorem 8.* Let

$$A = \begin{bmatrix} B & C \\ D & O \end{bmatrix}$$

be nonsingular of order n, where B is $m \times p$, $0 < m < n$ and $0 < p < n$. Then A^{-1} is of the form

$$A^{-1} = \begin{bmatrix} E & F \\ G & O \end{bmatrix} \ , \tag{158}$$

where E is $p \times m$, if and only if B is of rank $m + p - n$, in which case

$$E = B^{(1,2)}_{N(D),R(C)} \ , \quad F = D^{(1,2)}_{N(B),\{\mathbf{0}\}} \ , \quad G = C^{(1,2)}_{R(I_{n-p}),R(B)} \ . \tag{159}$$

Proof. We first observe that since A is nonsingular, C is of full column rank $n-p$, for otherwise the columns of A would not be linearly independent. Similarly, D is of full row rank $n - m$. Since C is $m \times (n - p)$, it follows that $n - p \leq m$, or, in other words,

$$m + p \geq n \ .$$

If. Since A is nonsingular, the $m \times n$ matrix $\begin{bmatrix} B & C \end{bmatrix}$ is of full row rank m, and therefore of column rank m. Therefore, a basis for $\mathbb{C}^m$ can be chosen from among its columns. Moreover, this basis can be chosen so that it includes all $n - p$ columns of C, and the remaining $m + p - n$ basis elements are columns of B. Since B is of rank $m + p - n$, the latter columns span $R(B)$. Therefore $R(B) \cap R(C) = \{\mathbf{0}\}$, and consequently $R(B)$ and $R(C)$ are complementary subspaces. Similarly, we can shos that $R(B^*)$ and $R(D^*)$ are complementary subspaces of $\mathbb{C}^p$, and therefrore their orthogonal complements $N(B)$ and $N(D)$ aare complementary spaces.

The results of the preceding paragraph guarantee the existence of all $\{1, 2\}$–inverses in the right member of Eqs. 159). if X now denotes RHS(158) with E, F, G given by (159), as easy computation shows that $AX = I_n$.

Only if. It was shown in the "if" part of the proof that rank B is at least $m + p - n$. If A^{-1} is of the form (158) we must have

$$BF = O \ . \tag{160}$$

Since A^{-1} is nonsingular, it follows from (158) that F is of full column rank $n - m$. Thus, (160) exhibits $n - m$ independent linear relations among the columns of B. Therefore the rank of B is at most $p - (n - m) = m + p - n$. This completes the proof. □

Suggested further reading

Section 2. Ben–Israel [**116**], Burns, Carlson, Haynsworth and Markham [**245**], and Carlson, Haynsworth and Markham [**293**].
Section 3. Afriat [**6**].
Section 4. Hartwig [**668**], and Harwood, Lovass–Nagy and Powers [**698**].
Section 5. Meyer [**1024**].
Section 6. Further references on bordered matrices are Blattner [**185**], Reid [**1261**], Hearon [**705**], and Germain–Bonne [**539**].
Further extensions of Cramer's rule are Cimmino [**343**], Wang ([**1519**], [**1522**]), and Werner [**1575**].

CHAPTER 6

A Spectral Theory for Rectangular Matrices

1. Introduction

Linear transformations in $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ and their matrix representations (see § 0.2.5, p. 12) are studied in this chapter, resulting in the simplest (diagonal) representations of linear transformations.

The main result, Theorem 2 (a restatement of the Autonne–Eckart–Young Theorem) states that for any $A \in \mathbb{C}_r^{m\times n}$ with *singular values*[1] $\alpha(A) = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ ordered by

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0 \,, \tag{1}$$

and for any scalars $d(A) = \{d_1, d_2, \ldots, d_r\}$ satisfying

$$|d_i| = \alpha_i \,, \quad i \in \overline{1, r} \tag{2}$$

there exist two unitary matrices $U \in U^{m\times m}$ (the set of $m \times m$ unitary matrices) and $V \in U^{n\times n}$ such that the $m \times n$ matrix

$$D = U^* A V = \begin{bmatrix} d_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & d_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix} \tag{3}$$

is diagonal. Thus any $m \times n$ complex matrix is unitarily equivalent to a diagonal matrix

$$A = U D V^* \,. \tag{4}$$

The corresponding statement for linear transformations is that for any linear transformations $A : \mathbb{C}^n \to \mathbb{C}^m$ with $\dim R(A) = r$, and for any set of scalars $d(A) = \{d_1, d_2, \ldots, d_r\}$ satisfying (2), there exist two orthogonal bases $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ and $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, such that the corresponding matrix representation $A_{\{\mathcal{U},\mathcal{V}\}}$ is diagonal,

$$A_{\{\mathcal{U},\mathcal{V}\}} = \begin{bmatrix} d_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & d_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix} ,$$

i.e.,

$$\begin{cases} A\mathbf{v}_j = d_j \mathbf{u}_j \,, & j = 1, \ldots, r \,, \\ A\mathbf{v}_j = \mathbf{0} \,, & j = r+1, \ldots, n \,. \end{cases} \tag{5}$$

If $d(A) = \alpha(A)$, i.e., if the scalars $\{d_1, \ldots, d_r\}$ in (2) are chosen as the singular values of A,

$$d_i = \alpha_i \,, \quad i \in \overline{1, r} \,, \tag{6}$$

[1] See Exs. 1–11 below.

then (4)

$$A = UDV^* \ , \quad D = \left[\begin{array}{ccc|c} \alpha_1 & & & \\ & \ddots & & O \\ & & \alpha_r & \\ \hline & O & & O \end{array}\right] \tag{7}$$

is called the *singular value decomposition* (abbreviated *SVD*) of A. In the general case, we will call (4) a *UDV^*–decomposition* of A.

While (6) is the most common choice, there are cases where other choices seem more natural. Thus if $A \in \mathbb{C}_r^{n\times n}$ is normal, then the choice $d(A) = \lambda(A)\backslash\{0\}$, i.e., choosing the scalars $\{d_1, \ldots , d_r\}$ to be the nonzero eigenvalues of A, guarantees that $U = V$ in (3) and (4), giving the spectral theorem for normal matrices[2] as a special case of (4).

The UDV^*–decomposition studied in Section 2 is the basis for a generalized spectral theory for rectangular matrices; this theory generalizes and extends the classical spectral theory for normal matrices (Theorem 2.13), replacing orthogonal projectors and eigenvalues by partial isometries and scalars $d(A)$ satisfying (2), respectively. This generalized spectral theory, essentially due to Penrose [**1177**], Lanczos [**906**], Hestenes ([**723**], [**724**], [**725**], [**726**]), and Hawkins and Ben–Israel [**700**], is developed in Section 4, following the discussion of partial isometries in Section 3.

Exercises and examples.

Ex. 1. *Singular values.* Let $A \in \mathbb{C}_r^{m\times n}$ and let $\lambda_j(A^*A)$, $j \in \overline{1,n}$, denote the eigenvalues of A^*A ordered by

$$\lambda_1(A^*A) \geq \lambda_2(A^*A) \geq \cdots \geq \lambda_r(A^*A) > \lambda_{r+1}(A^*A) = \cdots = \lambda_n(A^*A) = 0 \ . \tag{8}$$

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

$$\alpha_j(A) = +\sqrt{\lambda_j(A^*A)} \ , \quad j \in \overline{1,r} \ . \tag{9}$$

The *set of singular values* of A is denoted by $\alpha(A)$. Ordering the eigenvalues of AA^* as in (8), it follows from

$$\lambda_j(AA^*) = \lambda_j(A^*A) \ , \quad j = 1, \ldots , \min\{m, n\}$$

that the singular values can be defined equivalently by

$$\alpha_j(A) = +\sqrt{\lambda_j(AA^*)} \ , \quad j \in \overline{1,r} \ . \tag{10}$$

Ex. 2. A and A^* have the same singular values.

Ex. 3. Unitarily equivalent matrices have the same singular values.

Proof. Let $A \in \mathbb{C}^{m\times n}$, and let $U \in U^{m\times m}$ and $V \in U^{n\times n}$ be any two unitary matrices. Then the matrix

$$(UAV)(UAV)^* = UAVV^*A^*U^* = UAA^*U^*$$

is similar to AA^*, and thus has the same eigenvalues. Therefore the matrices UAV and A have the same singular values. $\square$

Ex. 4. (Lanczos [**906**]). Let $A \in \mathbb{C}_r^{m\times n}$. Then the matrix $\begin{bmatrix} O & A \\ A^* & O \end{bmatrix}$ has $2r$ nonzero eigenvalues given by $\pm\,\alpha_j(A)$, $j \in \overline{1,r}$.

[2]Theorem 2.13; see also Ex. 0.16(a) and Ex. 25 below.

Ex. 5. *An extremal characterization of singular values.* Let $A \in \mathbb{C}_r^{m\times n}$. Then

$$\alpha_k(A) = \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1,\, \mathbf{x} \perp \mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}\ , \quad k = 1, \dots, r\ , \tag{11}$$

where

$\|\ \|$ denotes the Euclidean norm,

$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}\}$ is an orthonormal set of vectors in $\mathbb{C}^n$, defined recursively by

$\|A\mathbf{x}_1\| = \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1\}$

$\|A\mathbf{x}_j\| = \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1,\, \mathbf{x} \perp \mathbf{x}_1, \dots, \mathbf{x}_{j-1}\}\ , \quad j = 2, \dots, k-1\ ,$

and RHS(11) is the (attained) supremum of $\|A\mathbf{x}\|$ over all vectors $\mathbf{x} \in \mathbb{C}^n$ with norm one, which are perpendicular to $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}$.

Proof. Follows from the corresponding extremal characterization of the eigenvalues of A^*A (see, e.g., Marcus and Minc [**996**, p. 114]),

$$\begin{aligned}\lambda_k(A^*A) &= \max\{\langle \mathbf{x}, A^*A\mathbf{x}\rangle : \|\mathbf{x}\| = 1,\, \mathbf{x} \perp \mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}\\ &= \langle \mathbf{x}_k, A^*A\mathbf{x}_k\rangle\ , \quad k = 1, \dots, n\end{aligned}$$

since $\langle \mathbf{x}, A^*A\mathbf{x}\rangle = \langle A\mathbf{x}, A\mathbf{x}\rangle = \|A\mathbf{x}\|^2$. Here the vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ are an orthonormal set of eigenvectors of A^*A,

$$A^*A\mathbf{x}_k = \lambda_k(A^*A)\mathbf{x}_k\ , \quad k = 1, \dots, n\ .$$

□

The singular values can be characterized equivalently as

$$\begin{aligned}\alpha_k(A) &= \max\{\|A^*\mathbf{y}\| : \|\mathbf{y}\| = 1,\, \mathbf{y} \perp \mathbf{y}_1, \dots, \mathbf{y}_{k-1}\}\ ,\\ &= \|A^*\mathbf{y}_k\|\end{aligned}$$

where the vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_r\}$ are an orthonormal set of eigenvectors of AA^*, corresponding to its positive eigenvalues

$$AA^*\mathbf{y}_k = \lambda_k(AA^*)\mathbf{y}_k\ , \quad k \in \overline{1, r}\ .$$

We can interpret this extremal characterization as follows: let the columns of A be $\mathbf{a}_j$, $j = 1, \dots, n$. Then

$$\|A^*\mathbf{y}_k\|^2 = \sum_{j=1}^{n} |\langle \mathbf{a}_j, \mathbf{y}\rangle|^2$$

Thus $\mathbf{y}_1$ is a normalized vector maximizing the sum of squares of mduli of its inner products with the columns of A, the maximum value being $\alpha_1^2(A)$, etc.

Ex. 6. If $A \in \mathbb{C}_r^{n\times n}$ is normal and its eigenvalues are ordered by

$$|\lambda_1(A)| \geq |\lambda_2(A)| \geq \cdots \geq |\lambda_r(A)| > |\lambda_{r+1}(A)| = \cdots = |\lambda_n(A)| = 0$$

then the singular values of A are

$$\alpha_j(A) = |\lambda_j(A)|\ , \quad j \in \overline{1, r}\ .$$

Hint. Use Ex. 5 and the spectral theorem for normal matrices, Theorem 2.13.

Ex. 7. Let $A \in \mathbb{C}_r^{m\times n}$, and let the singular values of $A^\dagger$ be ordered by

$$\alpha_1(A^\dagger) \geq \alpha_2(A^\dagger) \geq \cdots \geq \alpha_r(A^\dagger)\ .$$

Then

$$\alpha_j(A^\dagger) = \frac{1}{\alpha_{r-j+1}(A)}\ , \quad j \in \overline{1, r}\ . \tag{12}$$

PROOF.

$$\begin{aligned}\alpha_j^2(A^\dagger) &= \lambda_j(A^{\dagger *}A^\dagger)\ , \quad \text{by definition (9)}\\ &= \lambda_j((AA^*)^\dagger)\ , \quad \text{since } A^{\dagger *}A^\dagger = A^{*\dagger}A^\dagger = (AA^*)^\dagger\\ &= \frac{1}{\lambda_{r-j+1}(AA^*)}\\ &= \frac{1}{\alpha_{r-j+1}^2(A)}\ , \quad \text{by definition (10)}\ .\end{aligned}$$

□

Ex. 8. Let $\| \ \|$ be the matrix norm

$$\|A\| = (\text{trace} A^*A)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \tag{13}$$

defined on $\mathbb{C}^{m\times n}$, see, e.g., Ex. 0.27. Then for any $A \in \mathbb{C}_r^{m\times n}$,

$$\|A\|^2 = \sum_{j=1}^{r}\alpha_j^2(A)\ . \tag{14}$$

PROOF. Follows from $\text{trace} A^*A = \sum_{j=1}^{r}\lambda_j(A^*A)$. □

See also Ex. 51 below.

Ex. 9. Let $\| \ \|_2$ be the *spectral norm*, defined on $\mathbb{C}^{m\times n}$ by

$$\begin{aligned}\|A\|_2 &= \max\{\sqrt{\lambda} :\ \lambda \text{ an eigenvalue of } A^*A\}\\ &= \alpha_1(A)\ ;\end{aligned} \tag{15}$$

see, e.g., Ex. 0.32. Then for any $A \in \mathbb{C}_r^{m\times n}$, $r \geq 1$,

$$\|A\|_2\|A^\dagger\|_2 = \frac{\alpha_1(A)}{\alpha_r(A)}\ . \tag{16}$$

PROOF. Follows from Ex. 7 and definition (15). □

Ex. 10. *A condition number.* Let A be an $n\times n$ nonsingular matrix, and consider the equation

$$A\mathbf{x} = \mathbf{b} \tag{17}$$

for $\mathbf{b} \in \mathbb{C}^n$. The sensitivity of the solution of (17) to changes in the right–hand side $\mathbf{b}$, is indicated by the *condition number* of A, defined for ant multiplicative matrix norm $\| \ \|$ by

$$\text{cond}(A) = \|A\|\|A^{-1}\|\ . \tag{18}$$

Indeed, changing $\mathbf{b}$ to $(\mathbf{b}+\delta\mathbf{b})$ results in a change of the solution $\mathbf{x} = A^{-1}\mathbf{b}$ to $\mathbf{x}+\delta\mathbf{x}$, with

$$\delta\mathbf{x} = A^{-1}\delta\mathbf{b}\ . \tag{19}$$

For any consistent pair of vector and matrix norms (see Exs. 0.28–0.30), it follows from (17) that

$$\|\mathbf{b}\| \leq \|A\|\|\mathbf{x}\|\ . \tag{20}$$

Similarly, from (19)

$$\|\delta\mathbf{x}\| \leq \|A^{-1}\|\|\delta\mathbf{b}\|\ . \tag{21}$$

From (20) and (21) we get the following bound:

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A\|\|A^{-1}\|\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} = \text{cond}(A)\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} \tag{22}$$

relating the change of the solution to the change in data and the condition number (18).

The *spectral condition number* corresponding to the spectral norm (15) is, by (16)

$$\text{cond}(A) = \frac{\alpha_1(A)}{\alpha_n(A)} . \tag{23}$$

Prove that for this condition number

$$\text{cond}(A^*A) = (\text{cond}(A))^2 ,$$

showing that A^*A is worse conditioned than A, if $\text{cond}(A) > 1$ (Taussky [**1434**]).

Ex. 11. *Weyl's inequalities.* Let $A \in \mathbb{C}_r^{m\times n}$ have eigen values $\lambda_1, \dots , \lambda_n$ ordered by

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq \lambda_n|$$

and singular values

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r .$$

Then

$$\sum_{j=1}^{k} |\lambda_j| \geq \sum_{j=1}^{k} \alpha_j , \tag{24}$$

$$\prod_{j=1}^{k} |\lambda_j| \geq \prod_{j=1}^{k} \alpha_j , \tag{25}$$

for $k = 1, \dots , r$ (Weyl [**1590**], Marcus and Minc [**996**, pp. 115–116]).

A Historical note

TEXT (p. 242)

2. The UDV^* decomposition

The UDV^* decomposition studied here is a variation of the singular value decomposition proved by Beltrami, Jordan, and Sylvester for square real matrices (see, e.g., MacDuffee [**986**, p. 78]), by Autonne [**49**] for square complex matrices, and by Eckart and Young [**451**] for rectangular matrices. Our approach follows that of Eckart and Young [**451**]. First we require the following theorem.

Theorem 1. Let $O \neq A \in \mathbb{C}_r^{m\times n}$, let $\alpha(A)$, the singular values of A, be

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0 , \tag{1}$$

and let $d(A) = \{d_1, \dots , d_r\}$ be any complex scalars satisfying

$$|d_i| = \alpha_i , \quad i \in \overline{1, r} . \tag{2}$$

Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots , \mathbf{u}_r\}$ be an orthonormal set of eigenvectors of AA^* corresponding to its nonzero eigenvalues:

$$AA^*\mathbf{u}_i = \alpha_i^2 \mathbf{u}_i , \quad i \in \overline{1, r} \tag{26}$$

$$\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \delta_{ij} , \quad i, j \in \overline{1, r} . \tag{27}$$

Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots , \mathbf{v}_r\}$ be defined by

$$\mathbf{v}_i = \frac{1}{d_i} A^*\mathbf{u}_i , \quad i \in \overline{1, r} \tag{28}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal set of eigenvectors of A^*A corresponding to its nonzero eigenvalues

$$A^*A\mathbf{v}_i = \alpha_i^2 \mathbf{v}_i \ , \quad i \in \overline{1,r} \tag{29}$$

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij} \ , \quad i, j \in \overline{1,r} \ . \tag{30}$$

Furthermore

$$\mathbf{u}_i = \frac{1}{d_i} A\mathbf{v}_i \ , \quad i \in \overline{1,r} \ . \tag{31}$$

Dually, let the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ satisfy (29) and (30) and let the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ be defined by (31). Then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ satisfy (26), (27) and (28).

Proof. Let $\{\mathbf{v}_i : i \in \overline{1,r}\}$ be given by (28). Then

$$\begin{aligned} A^*A\mathbf{v}_i &= \frac{1}{\overline{d_i}} A^*AA^*\mathbf{u}_i \\ &= d_i A^*\mathbf{u}_i \ , \quad \text{by (26) and (2)} \\ &= \alpha_i^2 \mathbf{v}_i \ , \quad \text{by (28) and (2)} \end{aligned}$$

and

$$\begin{aligned} \langle \mathbf{v}_i, \mathbf{v}_i \rangle &= \frac{1}{\overline{d_i} d_j} \langle A^*\mathbf{u}_i, A^*\mathbf{u}_j \rangle \\ &= \frac{1}{\overline{d_i} d_j} \langle AA^*\mathbf{u}_i, \mathbf{u}_j \rangle \\ &= \frac{d_i}{d_j} \langle \mathbf{u}_i, \mathbf{u}_j \rangle \ , \quad \text{by (26) and (2)} \\ &= \delta_{ij} \ , \quad \text{by (27)} \ . \end{aligned}$$

Equations (31) follow from (28) and (26). The dual statement follows by interchanging A and A^*. $\square$

An easy consequence of Theorem 1 is the following.

Theorem 2. (Autonne [**49**], Eckart and Young [**451**]). Let $O \neq A \in \mathbb{C}_r^{m\times n}$, and let $d(A) = \{d_1, \ldots, d_r\}$ be complex scalars satisfying

$$|d_i| = \alpha_i \ , \quad i \in \overline{1,r} \ , \tag{2}$$

where

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0 \tag{1}$$

are the singular values of A.

Then there exist unitary matrices $U \in U^{m\times m}$ and $V \in U^{n\times n}$ such that the matrix

$$D = U^*AV = \begin{bmatrix} d_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & d_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix} \tag{3}$$

is diagonal.

Proof. For the given $A \in \mathbb{C}_r^{m\times n}$ we construct two such matrices U and V as follows.

Let the vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ in $\mathbb{C}^m$ satisfy (26) and (27), and thus form an orthonormal basis of $R(AA^*) = R(A)$; see, e.g., Corollary 1.2. Let $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ be an orthonormal basis of $R(A)^\perp = N(A^*)$. Then the set $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ is an orthonormal basis of $\mathbb{C}^m$ satisfying (26) and

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

The matrix U defined by

$$U = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m\end{bmatrix} \tag{33}$$

is thus an $m \times m$ unitary matrix.

Let now the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ in $\mathbb{C}^n$ be defined by (28). Then these vectors satisfy (29) and (30), and thus form an orthonormal basis of $R(A^*A) = R(A^*)$. Let $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ be an orthonormal basis of $R(A^*)^\perp = N(A)$. Then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{C}^n$ satisfying (29) and

$$A\mathbf{v}_i = \mathbf{0}\,, \quad i \in \overline{r+1, n}\,. \tag{34}$$

The matrix V defined by

$$V = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_r & \mathbf{v}_{r+1} & \dots & \mathbf{v}_n\end{bmatrix} \tag{35}$$

is thus an $n \times n$ unitary matrix.

With U and V as given above, the matrix

$$D = U^*AV = [d_{ij}]\,, \quad i \in \overline{1, m},\, j \in \overline{1, n}$$

satisfies

$$d_{ij} = \mathbf{u}_i^* A\mathbf{v}_j = 0 \quad \text{if } i > r \text{ or } j > r\,, \quad \text{by (32) and (34)}\,,$$

and for $i, j = 1, \dots, r$

$$\begin{aligned} d_{ij} &= \mathbf{u}_i^* A\mathbf{v}_j \\ &= \frac{1}{d_j}\mathbf{u}_i^* AA^*\mathbf{u}_j\,, \quad \text{by (28)} \\ &= d_j\,\mathbf{u}_i^*\mathbf{u}_j\,, \quad \text{by (26) and (2)} \\ &= d_j\,\delta_{ij}\,, \quad \text{by (27)}\,, \end{aligned}$$

completing the proof. $\square$

A corresponding decomposition of $A^\dagger$ is given in

Corollary 1. (Penrose [**1177**]). Let A, D, U, and V be as in Theorem 2. Then

$$A^\dagger = VD^\dagger U^* \tag{36}$$

where

$$D^\dagger = \left[\begin{array}{ccc|c} \frac{1}{d_1} & & & \\ & \ddots & & O \\ & & \frac{1}{d_r} & \\ \hline & O & & O \end{array}\right]. \tag{37}$$

PROOF. Equation (36) follows from (4) and Ex. 1.21. The form (37) for $D^\dagger$ is obvious since

$$D = \left[\begin{array}{ccc:c} d_1 & & & \\ & \ddots & & O \\ & & d_r & \\ \hdashline & O & & O \end{array}\right] .$$

□

Exercises and examples.

Ex. 12. Let $A \in \mathbb{C}_r^{m\times n}$, let $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ satisfy (26) and (27), and let $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ be given by (28). Then

$$A = \sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^* . \tag{38}$$

PROOF. The vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ form an orthonormal basis for $R(A^*)$. Therefore,

$$\sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^* \mathbf{x} = \mathbf{0} \quad \text{for all } \mathbf{x} \in R(A^*)^\perp = N(A) ,$$

and for any $j = 1, \dots, r$

$$\begin{aligned} \sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^* \mathbf{v}_j &= d_j \mathbf{u}_j , \quad \text{by (30)} \\ &= A\mathbf{v}_j , \quad \text{by (31)} \end{aligned}$$

proving that for all $\mathbf{x} \in \mathbb{C}^n$

$$\sum_{i=1}^{r} d_i \mathbf{u}_i \mathbf{v}_i^* \mathbf{x} = A\mathbf{x} .$$

□

Ex. 13. *Best matrix approximations of given rank.* For a given $A \in \mathbb{C}_r^{m\times n}$ and an integer k, $1 \le k \le r$, a *best rank-k approximation* of A is a matrix $A_{(k)} \in \mathbb{C}_k^{m\times n}$ satisfying

$$\|A - A_{(k)}\| = \inf_{X \in \mathbb{C}_k^{m\times n}} \|A - X\| , \tag{39}$$

where $\|\ \|$ is the matrix norm (13).

For the matrices D, U, and V of Theorem 2, let $D_{(k)}$, $U_{(k)}$, and $V_{(k)}$ denote their submatrices defined by

$$D_{(k)} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_k \end{bmatrix} \in \mathbb{C}^{k\times k} , \quad U_{(k)} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix} \in \mathbb{C}^{m\times k} , \quad V_{(k)} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix} \in \mathbb{C}^{n\times k} . \tag{40}$$

Then a best rank-k approximation of A is

$$A_{(k)} = U_{(k)} D_{(k)} V_{(k)}^* , \tag{41}$$

which is unique if, and only if, the k th and the $(k+1)$ st singular values of A are distinct:

$$\alpha_k \ne \alpha_{k+1} . \tag{42}$$

The approximation error of $A_{(k)}$ is

$$\|A - A_{(k)}\| = \left(\sum_{i=k+1}^{r} \alpha_i^2\right)^{1/2} \quad \text{(Eckart and Young [\textbf{450}])} . \tag{43}$$

PROOF. Using Ex. 0.34 and (3) we have, for any $X \in \mathbb{C}^{m\times n}$,

$$\|A - X\|^2 = \|U^*(A - X)V\|^2 = \|D - Y\|^2 = f(Y) , \quad \text{say} , \tag{44}$$

where

$$Y = U^* X V = [y_{ij}] . \tag{45}$$

Let L be any subspace with $\dim L \leq k$, and let P_L denote the orthogonal projector on L. Then the matrix $Y = P_L D$ minimizes $f(Y)$ among all matrices Y with $R(Y) \subset L$, and the corresponding minimum value is

$$\begin{aligned} \|D - P_L D\|^2 = \|QD\|^2 &= \operatorname{trace} D^* Q^* Q D \\ &= \operatorname{trace} D^* Q D = \sum_{i=1}^{m} \alpha_i^2 q_{ii} \end{aligned} \tag{46}$$

where $Q = I - P_L = [q_{ij}]$ is the orthogonal projector on $L^\perp$. Now

$$\begin{aligned} \inf_{X \in \mathbb{C}_k^{m\times n}} \|A - X\|^2 &= \inf_{Y \in \mathbb{C}_k^{m\times n}} \|D - Y\|^2 \\ &= \inf \{\|D - P_L D\|^2 : \text{over all subspaces } L \text{ with } \dim L \leq k\} \\ &= \inf \left\{ \sum_{i=1}^{m} \alpha_i^2 q_{ii} : Q = [q_{ij}] = P_{L^\perp}, \dim L \leq k \right\} \end{aligned} \tag{47}$$

and since $0 \leq q_{ii} \leq 1$ (why?), $\sum_{i=1}^{m} q_{ii} = m - \dim L$, it follows that the minimizing

$$Q = \begin{bmatrix} O & O \\ O & I_{m-k} \end{bmatrix} \text{ is unique if and only if } \alpha_k \neq \alpha_{k+1} ,$$

and the minimizing Y is accordingly

$$Y = P_L D = \begin{bmatrix} I_k & O \\ O & O \end{bmatrix} D$$

or

$$y_{ij} = \begin{cases} d_i & , \text{ if } 1 \leq i = j \leq k \\ 0 & , \text{ otherwise} . \end{cases} \tag{48}$$

The remaining statements are easily proved. $\square$

See also Householder and Young [**755**], Golub and Kahan [**553**]; Gaches, Rigal, and Rousset de Pina [**526**], and Franck [**514**].

EX. 14. Let $A \in \mathbb{C}_r^{m\times n}$. Then, using the notation of Ex. 13,

$$A = U_{(r)} D_{(r)} V_{(r)}^* = A_{(r)} , \tag{49}$$

$$A^\dagger = V_{(r)} D_{(r)}^{-1} U_{(r)}^* = A_{(r)}^\dagger . \tag{50}$$

Ex. 15. Let $O \neq A \in \mathbb{C}_r^{m\times n}$ have singular values

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$$

and let $M_{r-1} = \bigcup_{k=0}^{r-1} \mathbb{C}_k^{m\times n}$ be the set of $m \times n$ matrices of rank $\leq r-1$. Then the distance, using either norm (13) or the spectral norm (15), of A from M_{r-1} is

$$\inf_{X\in M_{r-1}} \|A - X\| = \alpha_r \ . \tag{51}$$

Two easy consequences of (51) are:

(a) Let A be as above, and let $B \in \mathbb{C}^{m\times n}$ satisfy

$$\|V\| < \alpha_r \ ;$$

then

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

(b) For any $0 \leq k \leq \min\{m, n\}$, the $m \times n$ matrices of rank $\leq k$ form a closed set in $\mathbb{C}^{m\times n}$.

In particular, the $n \times n$ singular matrices form a closed set in $\mathbb{C}^{n\times n}$. For any nonsingular $A \in \mathbb{C}^{n\times n}$th singular values

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n > 0$$

the smallest singular value α_n is a measure of the nonsingularity of A.

Ex. 16. *A minimal rank matrix approximation.* Let $A \in \mathbb{C}^{m\times n}$ and let $\epsilon > 0$. Find a matrix $B \in \mathbb{C}^{m\times n}$ of minimal rank, satisfying

$$\|A - B\| \leq \epsilon$$

for the norm (13).

Solution. Using the notation of Ex. 13,

$$B = A_{(k)} \ ,$$

where k is determined by

$$\left(\sum_{i=k}^{r} \alpha_i(A)^2\right)^{1/2} > \epsilon \ , \ \left(\sum_{i=k+1}^{r} \alpha_i(A)^2\right)^{1/2} \leq \epsilon \quad \text{(Golub [\textbf{549}])} \ .$$

Ex. 17. *A unitary matrix approximation.* Let $U^{n\times n}$ denote the set of $n \times n$ unitary matrices. Let $A \in \mathbb{C}_r^{n\times n}$ with a singular–value decomposition

$$A = UDV^* \ , \quad D == \left[\begin{array}{ccc|c|c} \alpha_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \alpha_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{array}\right] .$$

Then

$$\inf_{W\in U^{n\times n}} \|A - W\| = \|D - I\| = \sqrt{\sum_{i=1}^{r}(1-\alpha_i)^2 + n - r}$$

is attained for

$$W = UV^* \quad \text{(Fan and Hoffman [\textbf{481}], Mirsky [\textbf{1056}], Golub [\textbf{549}])} \ .$$

Ex. 18. The following generalization of Ex. 17 arises in factor analysis; see, e.g., Green [**573**] and Schönemann [**1318**].

For given $A, B \in \mathbb{C}^{m\times n}$, find a $W \in U^{n\times n}$ such that

$$\|A - BW\| \le \|A - BX\| \quad \text{for any } X \in U^{n\times n} \ .$$

SOLUTION. $W = UV^*$ where $B^*A = UDV^*$ is a singular–value decomposition of B^*A.

Ex. 19. Let $A_{(k)}$ be a best rank–k approximation of $A \in \mathbb{C}_r^{m\times n}$ (as given by Ex. 13). Then $A^*_{(k)}$, $A_{(k)}A^*_{(k)}$, and $A^*_{(k)}A_{(k)}$ are best rank–k approximations of A, AA^*, and A^*A, respectively. If A is normal, then $A^j_{(k)}$ is a best rank–k approximation of A^j for all $j = 1, 2\ldots$ (Householder and Young [**755**]).

Ex. 20. *Real matrices.* If $A \in \mathbb{R}_r^{m\times n}$, then the unitary matrices U and V in the singular value decomposition (7) can also be taken to be real, hence orthogonal.

Ex. 21. *Simultaneous diagonalization.* Let $A_1, A_2 \in \mathbb{C}^{m\times n}$. Then the following are equivalent:

(a) There exist two unitary matrices U, V such that both

$$D_1 = U^*A_1V \ ,$$
$$D_2 = U^*A_2V$$

are diagonal real matrices (in which case one of them, say D_1, can be assumed to be non–negative).

(b) $A_1A_2^*$ and $A_2^*A_1$ are both Hermitian (Eckart and Young [**451**]).

Ex. 22. Let $A_1, A_2 \in \mathbb{C}^{n\times n}$ be Hermitian matrices. Then the following are equivalent:

(a) There is a unitary matrix U such that both

$$D_1 = U^*A_1U \ ,$$
$$D_2 = U^*A_2U$$

are diagonal real matrices.

(b) A_1A_2 and A_2A_1 are both Hermitian.

(c) $A_1A_2 = A_2A_1$.

Ex. 23. Let $A_1, A_2 \in \mathbb{C}^{m\times n}$. Then the following are equivalent:

(a) There exist two unitary matrices U, V such that both

$$D_1 = U^*A_1V \ ,$$
$$D_2 = U^*A_2V$$

are diagonal matrices.

(b) There is a polynomial f such that

$$A_1A_2^* = f(A_2A_1^*)$$
$$A_2^*A_1 = f(A_1^*A_2) \quad \text{(Williamson [\textbf{1599}])} \ .$$

Ex. 24. *Normal matrices.* If $O \ne A \in \mathbb{C}_r^{n\times n}$ is normal and its nonzero eigenvalues are ordered by

$$|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_r| > 0 \ ,$$

then the scalars $d(A) = \{d_1, \ldots, d_r\}$ in (2) can be chosen as the corresponding eigenvalues

$$d_i = \lambda_i \ , \quad i \in \overline{1, r} \ . \tag{52}$$

This choice reduces both (28) and (31) to

$$\mathbf{u}_i = \mathbf{v}_i \ , \quad i \in \overline{1, r} \ . \tag{53}$$

PROOF. The first claim follows from Ex. 6.

Using Exs. 0.16 it can be shown that all four matrices A, A^*, AA^*, and A^*A have common eigenvectors. Therefore, the vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ of (26) and (27) are also eigenvectors of A^*, and (28) reduces to (53). $\square$

Ex. 25. *Normal matrices.* If $O \neq A \in \mathbb{C}_r^{n\times n}$ is normal, and the scalars $d(A)$ are chosen by (52), then the UDV^*–decomposition (4) of A reduces to the statement that A is unitarily similar to a diagonal matrix

$$A = UDU^* \ ; \quad \text{see Ex. 0.16(a)} \ .$$

3. Partial isometries and the polar decomposition theorem

A linear transformation $U : \mathbb{C}^n \to \mathbb{C}^m$ is called a *partial isometry* (sometimes also a *subunitary transformation*) if it is norm preserving on the orthogonal complement of its null space, i.e., if

$$\|U\mathbf{x}\| = \|\mathbf{x}\| \quad \text{for all } \mathbf{x} \in N(U)^\perp = R(U^*) \ , \tag{54}$$

or equivalently, if it is distance preserving

$$\|U\mathbf{x} - U\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in N(U)^\perp \ .$$

Except where otherwise indicated, the norms used here are the Euclidean vector norm and the corresponding spectral norm for matrices, see Ex. 0.32.

Partial isometries in Hilbert spaces were studied extensively by von Neumann [**1507**], Halmos [**646**], Halmos and McLaughlin [**647**], Erdelyi [**474**], and others. Most of the results given here are special cases for the finite dimensional space $\mathbb{C}^n$.

A nonsingular partial isometry is called an *isometry* (or a *unitary transformation*). Thus a linear transformation $U : \mathbb{C}^n \to \mathbb{C}^n$ is an isometry if $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$.

We recall that $U \in \mathbb{C}^{n\times n}$ is a unitary matrix if and only if $U^* = U^{-1}$. Analogous characterizations of partial isometries are collected in the following theorem, drawn from Halmos [**646**], Hestenes [**723**] and Erdélyi [**467**].

Theorem 3. Let $U \in \mathbb{C}^{m\times n}$. Then the following eight statements are equivalent.

- (a) U is a partial isometry.
- (a*) U^* is a partial isometry.
- (b) U^*U is an orthogonal projector.
- (b*) UU^* is an orthogonal projector.
- (c) $UU^*U = U$.
- (c*) $U^*UU^* = U^*$.
- (d) $U^* = U^\dagger$.
- (d*) $U^\dagger$ is a partial isometry.

Proof. We prove (a) $\Longleftrightarrow$ (b), (a) $\Longleftrightarrow$ (e), and (b) $\Longleftrightarrow$ (c) $\Longleftrightarrow$ (d). The obvious equivalence (c) $\Longleftrightarrow$ (c*) then takes care of the dual statements (a*) and (b*).

(a) $\Longrightarrow$ (b). Since $R(U^*U) = R(U^*)$, (b) can be rewritten as

$$U^*U = P_{R(U^*)} \ . \tag{55}$$

From Ex. 0.16(b) it follows for any Hermitian $H \in \mathbb{C}^{n\times n}$ that

$$\langle H\mathbf{x}, \mathbf{x}\rangle = 0 \ , \quad \text{for all } \mathbf{x} \in \mathbb{C}^n \ , \tag{56}$$

implies $H = O$. Consider now the matrix

$$H = P_{R(U^*)} - U^*U \ .$$

Clearly,

$$\langle H\mathbf{x}, \mathbf{x}\rangle = 0 \quad \text{for all } \mathbf{x} \in R(U^*)^\perp = N(U) \ ,$$

while for $\mathbf{x} \in R(U^*)$

$$\begin{aligned} \langle P_{R(U^*)}\mathbf{x}, \mathbf{x}\rangle &= \langle \mathbf{x}, \mathbf{x}\rangle \\ &= \langle U\mathbf{x}, U\mathbf{x}\rangle \quad \text{by (a)} \\ &= \langle U^*U\mathbf{x}, \mathbf{x}\rangle \ . \end{aligned}$$

Thus (a) implies that the Hermitian matrix $H = P_{R(U^*)} - U^*U$ satisfies (56), which in turn implies (55).

(b) $\Longrightarrow$ (a). This follows from

$$\begin{aligned} \langle U\mathbf{x}, U\mathbf{x}\rangle &= \langle U^*U\mathbf{x}, \mathbf{x}\rangle \\ &= \langle P_{R(U^*)}\mathbf{x}, \mathbf{x}\rangle \quad \text{by (55)} \\ &= \langle \mathbf{x}, \mathbf{x}\rangle \quad \text{if } \mathbf{x} \in R(U^*) \ . \end{aligned}$$

(a) $\Longleftrightarrow$ (e). Since

$$\mathbf{y} = U\mathbf{x} \ , \quad \mathbf{x} \in R(U^*)$$

is equivalent to

$$\mathbf{x} = U^\dagger \mathbf{y} \ , \quad \mathbf{y} \in R(U) \ ,$$

it follows that

$$\langle U\mathbf{x}, U\mathbf{x}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle \quad \text{for all } \mathbf{x} \in R(U^*)$$

is equivalent to

$$\langle \mathbf{y}, \mathbf{y}\rangle = \langle U^\dagger\mathbf{y}, U^\dagger\mathbf{y}\rangle \quad \text{for all } \mathbf{y} \in R(U) = N(U^\dagger)^\perp \ .$$

(b) $\Longleftrightarrow$ (c) $\Longleftrightarrow$ (d). The obvious equivalence (c) $\Longleftrightarrow$ (c*) states that $U^* \in U\{1\}$ if, and only if, $U^* \in U\{2\}$. Since U^* is (always) a $\{3,4\}$–inverse of U, it follows that U^* is a $\{1\}$–inverse of U if, and only if, $U^* = U^\dagger$. □

Returning to the UDV^*–decomposition of Section 2, we identify some useful partial isometries in the following theorem.

Theorem 4. (Hestenes [**723**]). Let $O \neq A \in \mathbb{C}_r^{m\times n}$, and let

$$A = UDV^* \ , \tag{4}$$

where the unitary matrices $U \in U^{m\times m}$, $V \in U^{n\times n}$ and the diagonal matrix $D \in \mathbb{C}^{m\times n}$ are given as in Theorem 2. Let $U_{(r)}$, $D_{(r)}$, and $V_{(r)}$ be defined by (40). Then

(a) The matrices $U_{(r)}$, $V_{(r)}$ are partial isometries with

$$U_{(r)}U_{(r)}^* = P_{R(A)} \ , \quad U_{(r)}^*U_{(r)} = I_r \ , \tag{57}$$

$$V_{(r)}V_{(r)}^* = P_{R(A^*)} \ , \quad V_{(r)}^*V_{(r)} = I_r \ . \tag{58}$$

(b) The matrix

$$E = U_{(r)}V_{(r)}^* \tag{59}$$

is a partial isometry with

$$EE^* = P_{R(A)} \ , \quad E^*E = P_{R(A^*)} \ . \tag{60}$$

PROOF. (a) That $U_{(r)}$, $V_{(r)}$ are partial isometries is obvious from their definitions and the unitarity of U and V (see, e.g., Ex. 28). Now

$$U_{(r)}^* U_{(r)} = I_r \ ,$$

by Definition (40), since U is unitary, and

$$\begin{aligned} P_{R(A^*)} &= A^\dagger A = A_{(r)}^\dagger A_{(r)} \ , \quad \text{by Ex. 13} \\ &= V_{(r)} D_{(r)}^{-1} U_{(r)}^* U_{(r)} D_{(r)} V_{(r)}^* \ , \quad \text{by (49) and (50)} \\ &= V_{(r)} V_{(r)}^* \ , \end{aligned}$$

with the remaining statements in (a) similarly proved.

(b) using (57) and (58), it can be verified that

$$E^\dagger = V_{(r)} U_{(r)}^* = E^* \ ,$$

from which (60) follows easily. □

The partial isometry E thus maps $R(A^*)$ isometrically onto $R(A)$. Since A also maps $R(A^*)$ onto $R(A)$, we should expect A to be a "multiple" of E. This is the essence of the following theorem, proved by Autonne [**49**] and Williamson [**1599**] for square matrices, by Penrose [**1177**] for rectangular matrices, and by Murray and von Neumann [**1103**] for linear operators in Hilbert spaces.

THEOREM 5. (The polar decomposition theorem). Let $O \neq A \in \mathbb{C}_r^{m\times n}$. Then A can be written as

$$A = GE = EH \ , \tag{61}$$

where $E \in \mathbb{C}^{m\times n}$ is a partial isometry and $G \in \mathbb{C}^{m\times m}$, $H \in \mathbb{C}^{n\times n}$ are Hermitian and positive semi–definite.

The matrices E, G, and H are uniquely determined by

$$R(E) = R(G) \ , \tag{62}$$

$$R(E^*) = R(H) \ , \tag{63}$$

in which case

$$G^2 = AA^* \ , \tag{64}$$

$$H^2 = A^*A \ , \tag{65}$$

and E is given by

$$E = U_{(r)} V_{(r)}^* \ . \tag{59}$$

PROOF. Let

$$A = UDV^* \ , \quad D = \begin{bmatrix} \alpha_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \alpha_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix} \tag{7}$$

be the singular–value decomposition of A. For any k, $r \leq k \leq \min\{m,n\}$, we use (40) to define the three matrices

$$D_{(k)} = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_r \end{bmatrix} \in \mathbb{C}^{k\times k} \ , \quad U_{(k)} = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_k \end{bmatrix} \in \mathbb{C}^{m\times k} \ , \quad V_{(k)} = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_k \end{bmatrix} \in \mathbb{C}^{n\times k} \ .$$

Then (7) can be rewritten as

$$\begin{aligned} A &= U_{(k)} D_{(k)} V_{(k)}^* \\ &= (U_{(k)} D_{(k)} U_{(k)}^*)(U_{(k)} V_{(k)}^*) \ , \quad \text{since } U_{(k)}^* U_{(k)} = I_k \\ &= (U_{(k)} V_{(k)}^*)(V_{(k)} D_{(k)} V_{(k)}^*) \ , \quad \text{since } V_{(k)}^* V_{(k)} = I_k \ , \end{aligned}$$

which proves (61) with the partial isometry

$$E = U_{(k)} V_{(k)}^* \tag{66}$$

and the positive semi–definite matrices

$$G = U_{(k)} D_{(k)} U_{(k)}^* \ , \quad H = V_{(k)} D_{(k)} V_{(k)}^* \ . \tag{67}$$

This also shows E to be nonunique if $r < \min\{m, n\}$, in which case G and H are also nonunique, for (67) can then be replaced by

$$\begin{aligned} G &= U_{(k)} D_{(k)} U_{(k)}^* + \mathbf{u}_{k+1} \mathbf{u}_{k+1}^* \ , \\ H &= V_{(k)} D_{(k)} V_{(k)}^* + \mathbf{v}_{k+1} \mathbf{v}_{k+1}^* \ , \end{aligned}$$

which satisfies (61) for the E given in (66).

Let now E and G satisfy (62). Then from (61)

$$AA^* = GEE^*G = GEE^\dagger G = GP_{R(E)}G = G^2 \ ,$$

which proves (64) and the uniqueness of G; see also Ex. 26 below. The uniqueness of E follows from

$$E = EE^\dagger E = GG^\dagger E = G^\dagger GE = G^\dagger A \ . \tag{68}$$

Similarly (63) implies (65) and the uniqueness of H, E.

Finally from

$$\begin{aligned} G^2 &= AA^* \\ &= U_{(r)} D_{(r)} V_{(r)}^* V_{(r)} D_{(r)} U_{(r)}^* \ , \quad \text{by (49)} \\ &= U_{(r)} D_{(r)}^2 U_{(r)}^* \end{aligned}$$

we conclude that

$$G = U_{(r)} D_{(r)} U_{(r)}^*$$

and consequently

$$G^\dagger = U_{(r)} D_{(r)}^{-1} U_{(r)}^* \ . \tag{69}$$

Therefore,

$$\begin{aligned} E &= G^\dagger A \ , \quad \text{by (68)} \\ &= U_{(r)} D_{(r)}^{-1} U_{(r)}^* U_{(r)} D_{(r)} V_{(r)}^* \ , \quad \text{by (69) and (49)} \\ &= U_{(r)} V_{(r)}^* \ , \quad \text{proving (59)} \ . \end{aligned}$$

□

If, in the proof of Theorem 5, one uses a general UDV^*–decomposition of A instead of the singular–value decomposition, then the matrices G and H defined by (67) are merely normal matrices, and need not be Hermitian. Hence, the following corollary.

COROLLARY 2. Let $O \neq A \in \mathbb{C}_r^{m\times n}$. Then, for any choice of the scalars $d(A)$ in (2), there exist a partial isometry $E \in \mathbb{C}^{m\times n}$ and two normal matrices $G \in \mathbb{C}^{m\times m}$, $H \in \mathbb{C}^{n\times n}$, satisfying (61). The matrices E, G, and H are uniquely determined by (62) and (63), in which case

$$GG^* = AA^* \ , \tag{70}$$

$$H^*H = A^*A \ , \tag{71}$$

and E is given by (59). □

Theorem 5 is the matrix analog of the polar decomposition of a complex number

$$z = x + iy \ , \quad x, y \text{ real}$$

as

$$z = |z|e^{i\theta} \ , \tag{72}$$

where

$$|z| = (z\bar{z})^{1/2} = (x^2 + y^2)^{1/2}$$

and

$$\theta = \arctan \frac{y}{x} \ .$$

Indeed, the complex scalar z in (72) corresponds to the matrix A in (61), while $\bar{z}$, $|z|$. and $e^{i\theta}$ correspond to A^*, G (or H) and E, respectively. This analogy is natural since $|z| = (z\bar{z})^{1/2}$ corresponds to the square roots $G = (AA^*)^{1/2}$ or $H = (A^*A)^{1/2}$, while the scalar $e^{i\theta}$ satisfies

$$|ze^{i\theta}| = |z| \quad \text{for all } z \in \mathbb{C} \ ,$$

which justifies its comparison to the partial isometry E; see also Exs. 44 and 48.

Exercises and examples.

EX. 26. *Square roots.* Let $A \in \mathbb{C}_r^{n\times n}$ be Hermitian positive semi–definite. Then there exists a unique Hermitian positive semi–definite matrix $B \in \mathbb{C}_r^{n\times n}$ satisfying

$$B^2 = A \ ; \tag{73}$$

B is called the *square root* of A, denoted by $A^{1/2}$.

PROOF. Writing A as

$$A = UDU^* \ , \quad U \text{ unitary} \ , \quad D = \begin{bmatrix} \lambda_1 & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \lambda_r & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix}$$

we see that

$$B = UD^{1/2}U^* \ , \quad D^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & & & \vdots & \\ & \ddots & & \vdots & O \\ & & \lambda_r^{1/2} & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & O & & \vdots & O \end{bmatrix}$$

is a Hermitian positive semi–definite matrix satisfying (73). To prove uniqueness, assume that B is a Hermitian matrix satisfying (73). Then, since B and $A = B^2$ commute, it follows from Ex. 22 that

$$B = U\widetilde{D}U^*$$

where $\widetilde{D}$ is diagonal and real, by Ex. 0.16(b), hence

$$\widetilde{D} = D^{1/2} \ , \quad \text{by (73).}$$

$\square$

Ex. 27. *Linearity of isometries.* Let X, Y be real normed vector spaces and let $f : X \to Y$ be isometric, i.e.,

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|_Y = \|\mathbf{x}_1 - \mathbf{x}_2\|_X \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in X$$

where $\| \ \|_X$ and $\| \ \|_Y$ are the norms in X and Y, respectively. If $f(\mathbf{0}) = \mathbf{0}$ then f is a linear transformation (Mazur and Ulam). For extensions and references see Dunford and Schwartz [**441**, p. 91] and Vogt [**1503**].

Ex. 28. *Partial isometries.* If the $n \times n$ matrix U is unitary, and $U_{(k)}$ is any $n \times k$ submatrix of U, then $U_{(k)}$ is a partial isometry. Conversely, if $W \in \mathbb{C}_k^{n\times k}$ is a partial isometry, the there is an $n \times (n-k)$ partial isometry V such that the matrix $U = \begin{bmatrix} W & V \end{bmatrix}$ is unitary.

Ex. 29. Any matrix unitarily equivalent to a partial isometry is a partial isometry.

PROOF. Let $A = UBV^*$, $U \in U^{m\times m}$, $V \in U^{n\times n}$. Then

$$\begin{aligned} A^\dagger &= VB^\dagger U^* \ , && \text{by Ex. 1.21} \\ &= VB^*U^* \ , && \text{if } B \text{ is a partial isometry} \\ &= A^* \ . \end{aligned}$$

$\square$

Ex. 30. Let $A \in \mathbb{C}_r^{m\times n}$ be a partial isometry with singular values $\alpha(A) = \{\alpha_i : i \in \overline{1,r}\}$. Then

$$\alpha_i = 1 \ , \quad i = 1, \dots, r \ .$$

Consequently, in any UDV^*–decomposition of a partial isometry, teh diagonal factor

$$D = \left[\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & O \\ & & d_r & \\ \hline & O & & O \end{array}\right]$$

has $|d_i| = 1, \ i = 1, \dots, r$.

Ex. 31. A linear transformation $E : \mathbb{C}^n \to \mathbb{C}^m$ with $\dim R(E) = r$ is a partial isometry if, and only if, there are two orthonormal bases $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ and $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ of $R(E^*)$ and $R(E)$, respectively, such that

$$\mathbf{u}_i = E\mathbf{v}_i \ , \quad i = 1, \dots, r \ .$$

Ex. 32. *Contractions.* A matrix $A \in \mathbb{C}^{m\times n}$ is called a *contraction* if

$$\|A\mathbf{x}\| \le \|\mathbf{x}\| \quad \text{for all } \mathbf{x} \in \mathbb{C}^n \ . \tag{74}$$

For any $A \in \mathbb{C}^{\times n}$ the following statements are equivalent:

(a) A is a contraction.
(b) A^* is a contraction.

(c) For any subspace L of $\mathbb{C}^m$ containing $R(A)$, the matrix $P_L - AA^*$ is positive semi–definite.

PROOF. (a) $\Longleftrightarrow$ (b). By Exs. 0.28 and 0.32, (a) is equivalent to

$$\|A\|_2 \le 1 \ ,$$

but

$$\|A\|_2 = \|A^*\|_2 \quad \text{by (15) and Ex. 2.}$$

(b) $\Longleftrightarrow$ (c). By definition (74), the statement (b) is equivalent to

$$\begin{aligned} 0 &\le \langle \mathbf{x}, \mathbf{x}\rangle - \langle A^*\mathbf{x}, A^*\mathbf{x}\rangle \\ &= \langle (I - AA^*)\mathbf{x}, \mathbf{x}\rangle \quad \text{for all } \mathbf{x} \in \mathbb{C}^m \ , \end{aligned}$$

which in turn is equivalent to (c). □

Ex. 33. Let $A \in \mathbb{C}^{m\times n}$ be a contraction and let L be any subspace of $\mathbb{C}^m$ containing $R(A)$. Then the $(m+n)\times(m+n)$ matrix $M(A)$ defined by

$$M(A) = \begin{bmatrix} A & \vdots & \sqrt{P_L - AA^*} \\ \cdots & \cdots & \cdots\cdots\cdots \\ O & \vdots & O \end{bmatrix}$$

is a partial isometry (Halmos and McLaughlin [**647**], Halmos [**646**]).

PROOF. The square root $\sqrt{P_L - AA^*}$ exists and is unique by Exs. 32(c) and 26. The proof the follows by verifying that

$$M(A)M(A)^*M(A) = M(A) \ .$$

□

Ex. 34. *Eigenvalues of partial isometries.* Let U be an $n\times n$ partial isometry and let λ be an eigenvalue of U corresponding to the eigenvector $\mathbf{x}$. Then

$$|\lambda| = \frac{\|P_{R(U^*)}\mathbf{x}\|}{\|\mathbf{x}\|} \ ;$$

hence

$$|\lambda| \le 1 \quad \text{(Erdélyi [\textbf{467}]).}$$

PROOF. From $U\mathbf{x} = \lambda\mathbf{x}$ we conclude

$$|\lambda|\|\mathbf{x}\| = \|U\mathbf{x}\| = \|UP_{R(U^*)}\mathbf{x}\| = \|P_{R(U^*)}\mathbf{x}\| \ .$$

□

Ex. 35. The partial isometry

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$$

has the following eigensystem:

$$\lambda = 0\,, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \in N(U)\,,$$

$$\lambda = 1\,, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \in R(U^*)\,,$$

$$\lambda = \sqrt{3}/2\,, \quad \mathbf{x} = \begin{bmatrix} 0 \\ \sqrt{3}/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{3}/2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1/2 \end{bmatrix}\,, \quad \begin{bmatrix} 0 \\ \sqrt{3}/2 \\ 0 \end{bmatrix} \in R(U^*)\,, \ \begin{bmatrix} 0 \\ 0 \\ 1/2 \end{bmatrix} \in N(U)\,.$$

Ex. 36. *Normal partial isometries.* Let U be an $n \times n$ partial isometry. Then U is normal if and only if it is range–Hermitian.

PROOF. Since any normal matrix is range–Hermitian, only the "if" part need proof. Let U be range–Hermitian, i.e., let $R(U) = R(U^*)$. Then $UU^* = U^*U$, by Theorem 3. □

Ex. 37. Let U be an $n \times n$ partial isometry. If U is normal, then its eigenvalues have absolute values 0 or 1.

PROOF. For any nonzero eigenvalue λ of a normal partial isometry U, it follows from $U\mathbf{x} = \lambda\mathbf{x}$ that $\mathbf{x} \in R(U) = R(U^*)$, and therefore

$$|\lambda|\|\mathbf{x}\| = \|U\mathbf{x}\| = \|\mathbf{x}\|\,.$$

□

Ex. 38. The converse of Ex. 37 is false. Consider, for example, the partial isometry

$$U = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\,.$$

Ex. 39. Let $E \in \mathbb{C}^{n\times n}$ be a contraction. Then E is a normal partial isometry if, and only if, the eigenvalues of E have absolute values 0 or 1 and rank E = rank E^2 (Erdelyi [**471**, Lemma 2]).

Ex. 40. A matrix $E \in \mathbb{C}^{n\times n}$ is a normal partial isometry if, and only if,

$$E = U \begin{bmatrix} W & \vdots & O \\ \cdots & \cdots & \cdots \\ O & \vdots & O \end{bmatrix} U^*\,,$$

where U and W are unitary matrices (Erdelyi [**471**]).

Ex. 41. *Polar decompositions.* Let $A \in \mathbb{C}^{n\times n}$, and let

$$A = GE\,, \tag{61}$$

where G is positive semi–definite and E is a partial isometry satisfying

$$R(E) = R(G)\,. \tag{62}$$

Then A is normal if, and only if

$$GE = EG\,,$$

in which case E is a normal partial isometry (Hearon [**707**, Theorem 1], Halmos [**646**, Problem 108]).

Ex. 42. Let $A \in \mathbb{C}^{n\times n}$ have the polar decompositions (61) and (62). Then A is a partial isometry if and only if G is an orthogonal projector.

PROOF. *If.* Let

$$G = G^* = G^2 \ . \tag{75}$$

Then

$$\begin{aligned} AA^* &= GEE^*G \ , \quad \text{by (61)} \\ &= G^2 \ , \quad \text{since } EE^* = P_{R(G)} \text{ by Theorem 3(b}^*\text{) and (62)} \\ &= G \ , \quad \text{by (75)} \ , \end{aligned}$$

proving that A is a partial isometry by Theorem 3(b*).

Only if. Let A be a partial isometry and let $A = GE$ be its unique polar decomposition determined by (62). Then

$$AA^* = G^2$$

is a Hermitian idempotent, by Theorem 3(b*), and hence its square root is also idempotent. □

EX. 43. Let $A \in \mathbb{C}^{n\times n}$ have the polar decompositions (61) satisfying (62) and (63). Then α is a singular value of A if, and only if,

$$A\mathbf{x} = \alpha E\mathbf{x} \ , \quad \text{for some } \mathbf{0} \neq \mathbf{x} \in R(E^*) \tag{76}$$

or equivalently, if and only if

$$A^*\mathbf{y} = \alpha E^*\mathbf{y} \ , \quad \text{for some } \mathbf{0} \neq \mathbf{y} \in R(E) \quad \text{(Hestenes [}\mathbf{723}\text{])} \ . \tag{77}$$

PROOF. From (61) it follows that (76) is equivalent to

$$G(E\mathbf{x}) = \alpha(E\mathbf{x}) \ ,$$

which, by (64), is equivalent to

$$AA^*(E\mathbf{x}) = \alpha^2(E\mathbf{x}) \ .$$

The equivalence of (77) is similarly proved. □

EX. 44. Let z be any complex number with the polar decomposition

$$z = |z|e^{i\theta} \ . \tag{76}$$

Then, for any real α, the following inequalities are obvious:

$$|z - e^{i\theta}| \leq |z - e^{i\alpha}| \leq |z - e^{i\theta}| \ .$$

Fan and Hoffman [**481**] established the followinganalogous matrix inequalities:

Let $A \in \mathbb{C}^{n\times n}$ be decomposed as

$$A = UH \ ,$$

where U is unitary and H is positive semi–definite. Then for any unitary $W \in U^{n\times n}$, the inequalities

$$\|A - U\| \leq \|A - W\| \leq \|A + U\|$$

hold for every unitarily invariant norm.

Give the analogous inequalities for the polar decomposition of rectangular matrices given in Theorem 5.

EX. 45. *Generalized Cayley transforms.* Let L be a subspace of $\mathbb{C}^n$. Then the equations

$$U = (P_L + iH)(P_L - iH)^\dagger \ , \tag{78}$$

$$H = i(P_L - U)(P_L + U)^\dagger \ , \tag{79}$$

establish a one–to–one correspondence between all Hermitian matrices H with

$$R(H) \subset L \tag{80}$$

and all normal partial isometries U with

$$R(U) = L \tag{81}$$

whose spectrum excludes -1 (Ben–Israel [**109**], Pearl [**1166**], [**1167**] and Nanda [**1110**]).

PROOF. Note that

$$(P_L \pm iH) \quad \text{and} \quad (P_L + U)$$

map L onto itself for Hermitian H satisfying (80) and normal partial isometries satisfying (81), whose spectrum excludes -1. Since on L, $(P_L \pm iH)$ and $(P_L \pm U)$ reduce to $(I \pm iH)$ and $(I \pm U)$, respectively, the proof follows from the classical theorem; see, e.g., Gantmacher [**533**, Vol. I, p. 279]. $\square$

EX. 46. Let H be a given Hermitian matrix. Let L_1 and L_2 be two subspaces containing $R(H)$, and let U_1 and U_2 be the normal partial isometries defined, respectively, by (78). If $L_1 \subset L_2$ then $U_1 = U_2 P_{L_1}$, i.e., U_1 is the restriction of U_2 to L_1. Thus the "minimal" normal partial isometry corresponding to a given Hermitian matrix H is

$$U = (P_{R(H)} + iH)(P_{R(H)} - iH)^\dagger \ .$$

EX. 47. A well known inequality of Fan and Hoffman [**481**, Theorem 3] is extended to the singular case as follows.

If H_1, H_2 are Hermitian with $R(H_1) = R(H_2)$ and if

$$U_k = (P_{R(H_k)} + iH_k)(P_{R(H_k)} - iH_k)^\dagger \ , \quad k = 1, 2 \ ,$$

then

$$\|U_1 - U_2\| \le 2\|H_1 - H_2\|$$

for every unitaril invariant norm (Ben–Israel [**109**]).

Trace inequalities.

EX. 48. Let z be a complex scalar. Then, for any real α, the following inequality is obvious:

$$|z| \ge \Re\{z e^{i\alpha}\} \ .$$

An analogous matrix inequality can be stated as follows:

Let $H \in \mathbb{C}^{n\times n}$ be Hermitian positive semi–definite. Then

$$\operatorname{trace} H \ge \Re\{\operatorname{trace}(HW)\} \ , \quad \text{for all } W \in U^{n\times n} \ .$$

where $U^{n\times n}$ is the class of $n \times n$ unitary matrices.

PROOF. Suppose there is a $W_0 \in U^{n\times n}$ with

$$\operatorname{trace} H < \Re\{\operatorname{trace}(HW_0)\} \ . \tag{82}$$

Let

$$H = UDU^* \quad \text{with } U \in U^{n\times n}$$

and

$$D = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix} \ ,$$

where $\{\alpha_1, \cdots, \alpha_n\}$ are the eigenvalues of H. Then

$$\begin{aligned}\sum \alpha_i &= \mathrm{trace} H < \Re\{\mathrm{trace}(UDU^* W_0)\}\ , \quad \text{by (82)} \\ &= \Re\{\mathrm{trace} A\}\ , \quad \text{where } A = UDV^*\ ,\ V^* = U^* W_0 \\ &= \Re\{\sum \lambda_i\}\ , \quad \text{where } \{\lambda_1, \dots, \lambda_n\} \text{ are the eigenvalues of } A\ .\end{aligned} \tag{83}$$

But $AA^* = UDV^*VDU^* = UD^2U^*$, proving that the nonzero $\{\alpha_i\}$ are the singular values of A. Thus (83) implies that

$$\sum \alpha_i < \sum |\lambda_i|\ ,$$

a contradiction of Weyl's inequality (24). □

EX. 49. Let $A \in \mathbb{C}_r^{m\times n}$ be given, and let $W_\ell^{m\times n}$ denote the class of all partial isometries in $\mathbb{C}_\ell^{m\times n}$, where $\ell = \min\{m, n\}$. Then

$$\sup_{W \in W_\ell^{m\times n}} \Re\{\mathrm{trace}(AW)\}$$

is attained for some $W_0 \in W_\ell^{m\times n}$. Moreover, AW_0 is Hermitian positive semi–definite, and

$$\sup_{W \in W_\ell^{m\times n}} \Re\{\mathrm{trace}(AW)\} = \mathrm{trace}(AW_0) = \sum_{i=1}^{r} \alpha_i\ , \tag{84}$$

where $\{\alpha_1, \dots, \alpha_r\}$ are the singular values of A. (For $m = n$, and unitary W, this result is due to von Neumann [**1506**].)

PROOF. Without a loss of generality, assume that $m \le n$. Let

$$A = GE \tag{61}$$

be a polar decomposition, where the partial isometry E is taken to be of full rank (using (66) with $k = m$), so $E \in W_m^{m\times n}$. The, for any $W \in W_m^{m\times n}$,

$$\begin{aligned}\mathrm{trace}(AW) &= \mathrm{trace}(GEW) \\ &= \mathrm{trace}\left(\begin{bmatrix} G & \vdots & O \\ \cdots & \cdots & \cdots \\ O & \vdots & O \end{bmatrix} \begin{bmatrix} E \\ \cdots \\ E^\perp \end{bmatrix} \begin{bmatrix} W & \vdots & W^\perp \end{bmatrix} \right)\ ,\end{aligned} \tag{85}$$

where the submatrices $E^\perp$ and $W^\perp$ are chosen so as to make

$$\begin{bmatrix} E \\ \cdots \\ E^\perp \end{bmatrix} \text{ and } \begin{bmatrix} W & \vdots & W^\perp \end{bmatrix}$$

unitary matrices; see, e.g., Ex. 28. Since

$$\begin{bmatrix} G & \vdots & O \\ \cdots & \cdots & \cdots \\ O & \vdots & O \end{bmatrix} \text{ is positive semi–definite, and } \begin{bmatrix} E \\ \cdots \\ E^\perp \end{bmatrix} \begin{bmatrix} W & \vdots & W^\perp \end{bmatrix}$$

is unitary, it follows from Ex. 48 and (85), that

$$\sup_{W \in W_m^{n\times m}} \Re\{\mathrm{trace}(AW)\}$$

is attained for $W_0 \in W_m^{n\times m}$ satisfying

$$AW_0 = G\ ,$$

and (84) follows from (67). □

Ex. 50. Let $A \in \mathbb{C}_r^{m\times n}$ and $B \in \mathbb{C}_s^{n\times m}$ have singular values

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$$

and

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_s > 0$$

respectively. Then

$$\sup_{X\in U^{n\times n},\, W\in U^{m\times m}} \Re\{\text{trace}(AXBW)\}$$

is attained for some $X_0 \in U^{n\times n}, W_0 \in U^{m\times m}$, and is given by

$$\text{trace}\,(AX_0BW_0) = \sum_{i=1}^{\min\{r,s\}} \alpha_i\beta_i \; .$$

This result was proved by von Neumann [**1506**, Theorem 1] for the case $m = n$. The general case is proved by "squaring the matrices A and B, i.e., adjoining sero rows and columns to make them square.

Gauge functions and singular values.
The following two exercises relate gauge functions (Ex. 3.49) to matrix norms and inequalities. The unitarily invariant matrix norms are characterized in Ex. 51 as symmetric gauge functions of the singular values. For square matrices these results were proved by von Neumann [**1506**] and Mirsky [**1056**].

Ex. 51. *Unitarily invariant matrix norms.* We use here the notation of Ex. 3.49.

Let the functions $\|\;\|_\phi : \mathbb{C}^{m\times n} \to \mathbb{R}$ and $\widehat{\phi} : \mathbb{C}^{mn} \to \mathbb{R}$ be defined, for any function $\phi : \mathbb{R}^\ell \to \mathbb{R}$, $\ell = \min\{m, n\}$, as follows: For any $A = [a_{ij}] \in \mathbb{C}^{m\times n}$ with singular values

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0 \; ,$$

$\|A\|_\phi$ and $\widehat{\phi}(a_{11}, \ldots , a_{mn})$ are defined as

$$\|A\|_\phi = \widehat{\phi}(a_{11}, \ldots , a_{mn}) = \phi(\alpha_1, \ldots , \alpha_r, 0, \ldots , 0) \; . \tag{86}$$

Then:

(a) If $\phi : \mathbb{R}^\ell \to \mathbb{R}$ satisfies conditions (G1)–(G3) of Ex. 3.49, so does $\widehat{\phi} : \mathbb{C}^{mn} \to \mathbb{R}$.

(b) $\|UAV\|_\phi = \|A\|_\phi$ for all $A \in \mathbb{C}^{m\times n}$, $U \in U^{m\times m}$, $V \in U^{n\times n}$.

(c) Let $\phi : \mathbb{R}^\ell \to \mathbb{R}$ satisfies conditions (G1)–(G3) of Ex. 3.49, and let $\phi_D : \mathbb{R}^\ell \to \mathbb{R}$ be its dual, defined by (3.85). Then, for any $A \in \mathbb{C}^{m\times n}$, the following supremum is attained, and

$$\sup_{X\in\mathbb{C}^{n\times m},\, \|X\|_\phi=1} \Re\{\text{trace}(AX)\} = \|A\|_{\phi_D} \; . \tag{87}$$

(d) If $\phi : \mathbb{R}^\ell \to \mathbb{R}$ is a symmetric gauge function, then $\widehat{\phi} : \mathbb{C}^{mn} \to \mathbb{R}$ is a gauge function, and $\|\;\|_\phi : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a unitarily invariant norm.

(e) If $\|\;\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a unitarily invariant norm, then there is a symmetric gauge function $\phi : \mathbb{R}^\ell \to \mathbb{R}$ such that $\|\;\| = \|\;\|_\phi$.

Proof. (a) Follows from definition (86).

(b) Obvious by Ex. 3.

(c) For the given $A \in \mathbb{C}^{m\times n}$

$$\begin{aligned}
\sup_{X\in\mathbb{C}^{n\times m},\,\|X\|_\phi=1} &\Re\{\mathrm{trace}(AX)\} = \\
&= \sup_{X\in\mathbb{C}^{n\times m},\,\|X\|_\phi=1} \Re\{\mathrm{trace}(AUXV) : U \in U^{n\times n}, V \in U^{m\times m}\}\ , \quad \text{by (b)} \\
&= \sup_{\phi(\xi_1,\dots,\xi_\ell)=1} \sum_i \alpha_i \xi_i\ , \quad \text{by Ex. 50} \\
&= \phi_D(\alpha_1,\dots,\alpha_r)\ , \quad \text{by (3.86) and (3.88)} \\
&= \|A\|_{\phi_D}\ , \quad \text{by (86)}\ ,
\end{aligned}$$

where

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$$

and

$$\xi_1 \geq \xi_2 \geq \cdots \geq \xi_\ell > 0$$

are the singular values of A and X, respectively.

(d) Let ϕ_D be the dual of ϕ, and let $\widehat{(\phi_D)} : \mathbb{C}^{mn} \to \mathbb{R}$ be defined by (86) as

$$\widehat{(\phi_D)}(a_{11},\dots,a_{mn}) = \|A\|_{\phi_D}\ , \quad \text{for } A = [a_{ij}]\ .$$

Then

$$\begin{aligned}
\widehat{(\phi_D)}(a_{11},\dots,a_{mn}) &= \|A^*\|_{\phi_D}\ , \quad \text{by Ex. 2} \\
&= \sup_{X=[x_{ij}]\in\mathbb{C}^{m\times n},\widehat{\phi}(x_{11},\dots,x_{mn})=1} \Re\{\mathrm{trace}(A^*X)\}\ , \quad \text{by (87)} \\
&= \sup_{\widehat{\phi}(x_{11},\dots,x_{mn})=1} \sum_{i,j} \overline{a_{ij}}\, x_{ij}
\end{aligned}$$

proving that $\widehat{(\phi_D)} : \mathbb{C}^{mn} \to \mathbb{R}$ is the dual of $\widehat{\phi} : \mathbb{C}^{mn} \to \mathbb{R}$, by using (3.86) and (3.88). Since ϕ is the dual ϕ_D (by Ex. 3.49(d)), it follows that $\widehat{\phi}$ is the dual of $\widehat{(\phi_D)}$ and, by Ex. 3.49(d), $\widehat{\phi} : \mathbb{C}^{mn} \to \mathbb{R}$ is a gauge function, That $\|\ \|_\phi$ is a unitarily invariant norm follows then from (b) and Ex. 3.53.

(e) Let $\|\ \| : \mathbb{C}^{m\times n} \to \mathbb{R}$ be a unitarily invariant matrix norm, and define $\phi : \mathbb{R}^\ell \to \mathbb{R}$ by

$$\phi(\mathbf{x}) = \phi(x_1, x_2, \dots, x_\ell) = \|[\mathrm{diag}|x_i|]\|\ ,$$

where

$$[\mathrm{diag}|x_i|] = \begin{bmatrix} |x_1| & & \\ & \ddots & \\ & & |x_\ell| \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}\ .$$

Then ϕ is a symmetric gauge function and $\|\ \| = \|\ \|_\phi$. □

Ex. 52. *Inequalities for singular values.* Let $A, B \in \mathbb{C}^{m\times n}$ and let

$$\alpha_1 \geq \cdots \geq \alpha_r > 0$$

and

$$\beta_1 \geq \cdots \geq \beta_s > 0$$

be the singular values of A and B, respectively. Then for any symmetric gauge function $\phi : \mathbb{R}^\ell \to \mathbb{R}$, $\ell = \min\{m, n\}$, the singular values

$$\gamma_1 \geq \cdots \geq \gamma_t > 0$$

of $A + B$ satisfy

$$\phi(\gamma_1, \dots, \gamma_t, 0, \dots, 0) \le \phi(\alpha_1, \dots, \alpha_r, 0, \dots, 0) + \phi(\beta_1, \dots, \beta_s, 0, \dots, 0) \tag{88}$$

(von Neumann [**1506**]).

Proof. The inequality (88) follows from (86) and Ex. 51(d), since

$$\|A + B\|_\phi \le \|A\|_\phi + \|B\|_\phi \ .$$

□

4. A spectral theory for rectangular matrices

The following theorem, due to Penrose [**1177**], is a generalization to rectangular matrices of the classical spectral theorem for normal matrices (Theorem 2.13).

Theorem 6. (Spectral theorem for rectangular matrices). Let $O \neq A \in \mathbb{C}_r^{m\times n}$, and let $d(A) = \{d_1, \dots, d_r\}$ be complex scalars satisfying

$$|d_i| = \alpha_i \ , \quad i = 1, \dots, r \tag{2}$$

where

$$\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_r > 0 \qquad \text{(refeq:6-2)}$$

are the singular values, $\alpha(A)$, of A.

Then there exist r partial isometries $\{E_i : i = 1, \dots, r\}$ in $\mathbb{C}_1^{m\times n}$ satisfying

$$E_i E_j^* = O \ , \quad E_i^* E_j = O \ , \quad 1 \le i \neq j \le r \tag{89}$$

$$E_i E^* A = A E^* E_i \ , \quad i = 1, \dots, r \tag{90}$$

where

$$E = \sum_{i=1}^{r} E_i \tag{91}$$

is the partial isometry given by (59), and

$$A = \sum_{i=1}^{r} d_i E_i \tag{92}$$

Furthermore, for each $i = 1, \dots, r$, the partial isometry $(\bar{d}_i/|d_i|)E_i$ is unique if the corresponding singular value is simple, i.e., if $\alpha_i < \alpha_{i-1}$ and $\alpha_i > \alpha_{i+1}$ for $2 \le i \le r$ and $1 \le i \le r-1$, respectively.

Proof. Let the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ satisfy (26) and (27), let vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ be defined by (28), and let

$$E_i = \mathbf{u}_i \mathbf{v}_i^* \ , \quad i = 1, \dots, r \ . \tag{93}$$

The E_i is a partial isometry by Theorem 3(c), since $E_i E_i^* E_i = E_i$ by (27) and (30), from which (89) also follows. The statement on uniqueness follows from (93), (2), (26), (27), and (28). The result (92) was proved in Ex. 12, which also shows the matris E of (59) to be given by (91). Finally, (90) follows from (91), (92). and (89). □

As shown by the proof of Theorem 6, the spectral representation (92) of A is just a way of rewriting its UDV^*–decomposition. The following spectral representation of $A^\dagger$ similarly follows from Corollary 1.

Corollary 3. Let A, d_i, and E_i, $i = 1, \dots, r$, be as in Theorem 6. Then

$$A^\dagger = \sum_{i=1}^{r} \frac{1}{d_i} E_i \ . \tag{94}$$

□

If $A \in \mathbb{C}_r^{n\times n}$ is a normal matrix with nonzero eigenvalues $\{\lambda_i : i = 1, \dots , r\}$ ordered by

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_r| \,,$$

then, by Ex. 24, the choice

$$d_i = \lambda_i \,, \quad i \in \overline{1,r} \tag{52}$$

guarantees that

$$\mathbf{u}_i = \mathbf{v}_i \,, \quad i \in \overline{1,r} \tag{53}$$

and consequently, the partial isometries E_i of (93) are orthogonal projectors

$$P_i = \mathbf{u}_i\mathbf{u}_i^* \,, \quad i \in \overline{1,r} \tag{95}$$

and (92) reduces to

$$A = \sum_{i=1}^{r} \lambda_i P_i \,, \tag{96}$$

giving the spectral theorem for normal matrices as a special case of Theorem 5.

The classical spectral theory for square matrices (see, e.g., Dunford and Schwartz [**441**, pp. 556–565] makes extensive use of matrix functions $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$, induced by scalar functions $f : \mathbb{C} \to \mathbb{C}$, according to the definition given in Ex. 53. Similarly, the spectral theory for rectangular matrices given here uses matrix functions $f : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ which correspond to scalar functions $f : \mathbb{C} \to \mathbb{C}$, according to the following.

Definition 1. Let $f : \mathbb{C} \to \mathbb{C}$ be any scalar function. Let $A \in \mathbb{C}_r^{m\times n}$ have a spectral representation

$$A = \sum_{i=1}^{r} d_i E_i \tag{92}$$

as in Theorem 6. Then the *matrix function* $f : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ *corresponding to* $f : \mathbb{C} \to \mathbb{C}$ is defined at A by

$$f(A) = \sum_{i=1}^{r} f(d_i) E_i \,. \tag{97}$$

Note that the value of $f(A)$ defined by (97) depends on the particular choice of the scalars $d(A)$ in (2). In particular, for a normal matrix $A \in \mathbb{C}^{n\times n}$, the choice of $d(A)$ by (52) reduces (97) to the classical definition – see (114) below – in the case that $f(0) = 0$ or that A is nonsingular.

Let

$$A = U_{(r)} D_{(r)} V_{(r)}^* \,, \quad D_{(r)} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_r \end{bmatrix} \tag{49}$$

be a UDV^*–decomposition of a given $A \in \mathbb{C}_r^{m\times n}$. Then Definition 1 gives $f(A)$ as

$$f(A) = U_{(r)} f(D_{(r)}) V_{(r)}^* \,, \quad f(D_{(r)}) = \begin{bmatrix} f(d_1) & & \\ & \ddots & \\ & & f(d_r) \end{bmatrix} \tag{98}$$

An easy consequence of Theorem 6 and Definition 1 is the following:

Theorem 7. Let $f, g, h : \mathbb{C} \to \mathbb{C}$ be scalar functions and let $f, g, h : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the corresponding matrix functions defined by Definition 1.

Let $A \in \mathbb{C}_r^{m\times n}$ have a UDV^*–decomposition

$$A = U_{(r)} D_{(r)} V_{(r)}^* \tag{49}$$

and let the partial isometry E be given by

$$E = U_{(r)} V_{(r)}^* \ . \tag{59}$$

Then

(a) If $f(z) = g(z) + h(z)$, then $f(A) = g(A) + h(A)$.
(b) If $f(z) = g(z)h(z)$, then $f(A) = g(A)E^* h(A)$.
(c) If $f(z) = g(h(z))$, then $f(A) = g(h(A))$.

Proof. Parts (a) and (c) are obvious by Definition 97).

(b) If $f(z) = g(z)h(z)$, then

$$\begin{aligned} g(A)E^*h(A) &= \left(\sum_{i=1}^{r} g(d_i)E_i\right)\left(\sum_{j=1}^{r} E_j^*\right)\left(\sum_{k=1}^{r} h(d_k)E_i\right) , \\ &\qquad \text{by (97) and (91)} , \\ &= \sum_{i=1}^{r} g(d_i)h(d_i)E_i \ , \quad \text{by (89) and Theorem 3(c)} , \\ &= \sum_{i=1}^{r} f(d_i)E_i = f(A) \ . \end{aligned}$$

$\square$

For matrix functions defined as above, an analog of Cauchy's integral theorem is given in Corollary 4 below. First we require

Lemma 1. Let $A \in \mathbb{C}_r^{m\times n}$ be represented by

$$A = \sum_{i=1}^{r} d_i E_i \tag{92}$$

Let $\{\widehat{d_j} : j = 1, \ldots , q\}$ be the set of distinct $\{d_i : i = 1, \ldots , r\}$ and let

$$\widehat{E_j} = \sum_{i} \{E_i : d_i = \widehat{d_j}\} \ , \quad j = 1, \ldots , q \tag{99}$$

For each $j \in \overline{1, q}$ let Γ_j be a contour (i.e., a closed rectifiable Jordan curve, positiveky oriented ib the customary way) surrounding $\widehat{d_j}$ but no other $\widehat{d_k}$.
Then:

(a) For each $j \in \overline{1, q}$, $\widehat{E_j}$ is a partial isometry and

$$\widehat{E_j}^* = \frac{1}{2\pi i} \int_{\Gamma_i} (zE - A)^\dagger dz \ . \tag{100}$$

(b) If $f : \mathbb{C} \to \mathbb{C}$ is analytic in a domain containing the set surrounded by

$$\Gamma = \bigcup_{j=1}^{q} \Gamma_j \ ,$$

then

$$\sum_{j=1}^{r} f(d_j)E_j^* = \frac{1}{2\pi i}\int_\Gamma f(z)(zE-A)^\dagger dz\ ; \tag{101}$$

in particular,

$$A^\dagger = \frac{1}{2\pi i}\int_\Gamma \frac{1}{z}(zE-A)^\dagger dz\ ; \tag{102}$$

PROOF. (a) From (89) and Theorem 3 it follows that $\widehat{E_j}$ and $\widehat{E_j}^*$ are partial isometries for each $j = 1, \ldots, q$. Also, from (91), (92), and Corollary 3

$$(zE-A)^\dagger = \sum_{k=1}^{r} \frac{1}{z-d_k}\, E_k^*\ , \tag{103}$$

hence

$$\begin{aligned}
\frac{1}{2\pi i}\int_{\Gamma_j}(zE-A)^\dagger dz &= \sum_{k=1}^{r}\left(\frac{1}{2\pi i}\int_{\Gamma_j}\frac{dz}{z-d_k}\right)E_k^* \\
&= \sum_{\{d_k=\widehat{d}_j\}} E_k^* \\
&\quad \text{by the assumptions on } \Gamma_j \text{ and Cauchy's integral theorem} \\
&= \widehat{E_j}^*\ , \quad \text{by (99)}\ .
\end{aligned}$$

(b) Similarly we calculate

$$\begin{aligned}
\frac{1}{2\pi i}\int_\Gamma f(z)(zE-A)^\dagger dz &= \sum_{j=1}^{q}\sum_{k=1}^{r}\left(\frac{1}{2\pi i}\int_{\Gamma_j}\frac{f(z)}{z-d_k}\right)E_k^* \\
&= \sum_{j=1}^{q} f(\widehat{d}_j)\widehat{E_j}^* \\
&= \sum_{j=1}^{r} f(d_j)E_j^*\ , \quad \text{proving (101)}\ .
\end{aligned}$$

Finally, (102) follows from (101) and Corollary 3. $\square$

COROLLARY 4. Let A, E, Γ, and f be as in Lemma 1. Then

$$f(A) = E\left(\frac{1}{2\pi i}\int_\Gamma f(z)(zE-A)^\dagger dz\right)E\ . \tag{104}$$

PROOF. Using (91) and (101) we calculate

$$\begin{aligned}
E\left(\frac{1}{2\pi i}\int_\Gamma f(z)(zE-A)^\dagger dz\right)E &= \left(\sum_{i=1}^{r} E_i\right)\left(\sum_{j=1}^{r} f(d_j)E_j^*\right)\left(\sum_{k=1}^{r} E_k\right) \\
&= \sum_{j=1}^{r} f(d_j)E_j\ , \quad \text{by (89) and Theorem 3(c)}\ , \\
&= f(A)\ .
\end{aligned}$$

$\square$

The *generalized resolvent* of a matrix $A \in \mathbb{C}^{m\times n}$ is the function $R(z, A) : \mathbb{C} \to \mathbb{C}^{n\times m}$ given by

$$R(z,A) = (zE - A)^{\dagger} \ , \tag{105}$$

where the partial isometry E is given as in Theorem 6. This definition is suggested by the classical definition of the *resolvent* of a square matrix as

$$R(z,A) = (zI - A)^{-1} \ , \quad \text{for all } z \notin \lambda(A) \ .$$

In analogy to the classical case – see, e.g., Dunford and Schwartz [**441**, p. 568] – we state the following identity, known as the (*first*) *resolvent equation*.

Lemma 2. Let $A \in \mathbb{C}_r^{m\times n}$ and let $d(A)$ and E be as in THeorem 6. Then

$$R(\lambda, A) - R(\mu, A) = (\mu - \lambda) R(\lambda, A) R(\mu, A) \tag{106}$$

for any scalars $\lambda, \mu \notin d(A)$.

Proof.

$$\begin{aligned}
R(\lambda, A) - R(\mu, A) &= (\lambda E - A)^{\dagger} - (\mu E - A)^{\dagger} \ , \quad \text{by (105)} \\
&= \sum_{k=1}^{r} \left(\frac{1}{\lambda - d_k} - \frac{1}{\mu - d_k} \right) E_k^* \ , \quad \text{by (103)} \\
&= \sum_{k=1}^{r} \left(\frac{\mu - \lambda}{(\lambda - d_k)(\mu - d_k)} \right) E_k^* \\
&= (\mu - \lambda) \left(\sum_{k=1}^{r} \frac{1}{\lambda - d_k} E_k^* \right) E \left(\sum_{\ell=1}^{r} \frac{1}{\mu - d_\ell} E_\ell^* \right) \ , \\
&\quad \text{by (89), (91) and Theorem 3(c)} \ , \\
&= (\mu - \lambda) R(\lambda, A) R(\mu, A) \ , \quad \text{by (103)} \ .
\end{aligned}$$

□

The resolvent equation, (106), is used in the following lemma, based on Lancaster [**902**, p. 552].

Lemma 3. Let $A \in \mathbb{C}^{m\times n}$, let $d(A)$ an dE be given as in Theorem 6, and let the scalar functions $f, g : \mathbb{C} \to \mathbb{C}$ be analytic in a domain D containing $d(A)$. If Γ is a contour surrounding $d(A)$ and lying in the interior of D, then

$$\begin{aligned}
&\left(\frac{1}{2\pi i} \int_{\Gamma} f(\lambda) R(\lambda, A) d\lambda \right) E \left(\frac{1}{2\pi i} \int_{\Gamma} g(\lambda) R(\lambda, A) d\lambda \right) \\
&\qquad\qquad = \ \frac{1}{2\pi i} \int_{\Gamma} f(\lambda) g(\lambda) R(\lambda, A) d\lambda
\end{aligned} \tag{107}$$

Proof. Let Γ_1 be a contour surrounding Γ and still lying in the interior of D. Then

$$\frac{1}{2\pi i} \int_{\Gamma} g(\lambda) R(\lambda, A) d\lambda = \frac{1}{2\pi i} \int_{\Gamma_1} g(\mu) R(\mu, A) d\mu \ ,$$

which when substituted in LHS(107) gives

$$\begin{aligned}\left(\frac{1}{2\pi i}\int_{\Gamma} f(\lambda)R(\lambda,A)d\lambda\right)E\left(\frac{1}{2\pi i}\int_{\Gamma_1} g(\mu)R(\mu,A)d\mu\right)& \\ &= -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma} f(\lambda)g(\mu)R(\lambda,A)ER(\mu,A)d\lambda d\mu \\ &= \frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma} f(\lambda)g(\mu)\frac{R(\lambda,A)-R(\mu,A)}{\lambda-\mu}d\lambda d\mu\ , \quad \text{by (106)} \\ &= \frac{1}{4\pi^2}\int_{\Gamma} f(\lambda)R(\lambda,A)\left(\int_{\Gamma_1}\frac{g(\mu)}{\lambda-\mu}d\mu\right)d\lambda \\ &\quad -\frac{1}{4\pi^2}\int_{\Gamma_1}\left(\int_{\Gamma}\frac{f(\lambda)}{\lambda-\mu}d\lambda\right)g(\mu)R(\mu,A)d\mu \\ &= \frac{1}{2\pi i}\int_{\Gamma} f(\lambda)g(\lambda)R(\lambda,A)d\lambda\ , \quad \text{since } \int_{\Gamma_1}\frac{g(\mu)}{\lambda-\mu}d\mu = -2\pi i g(\lambda) \\ &\quad \text{and} \qquad \int_{\Gamma}\frac{f(\lambda)}{\lambda-\mu}d\lambda = 0\ ,\end{aligned}$$

by our assumptions on Γ, Γ_1. □

We illustrate now the application of the above concepts to the solution of the matrix equation

$$AXB = D \tag{108}$$

studied in Theorem 2.1. Here the matrices $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{k\times \ell}$, and $D \in \mathbb{C}^{m\times \ell}$ are given, and, in addition, the matrices A and B have spectral representations , given by Theorem 6 as follows:

$$A = \sum_{i=1}^{p} d_i^A E_i^A\ , \qquad E^A = \sum_{i=1}^{p} E_i^A\ , \quad p = \operatorname{rank} A \tag{109}$$

and

$$B = \sum_{i=1}^{q} d_i^B E_i^B\ , \qquad E^B = \sum_{i=1}^{q} E_i^B\ , \quad q = \operatorname{rank} B\ . \tag{110}$$

Theorem 8. Let A, B, D be as above, and let Γ_1 and Γ_2 be contours surrounding $d(A) = \{d_1^A, \ldots, d_p^A\}$ and $d(B) = \{d_1^B, \ldots, d_q^B\}$, respectively. If (108) is consistent, then it has the following solution:

$$X = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\frac{R(\lambda,A)DR(\mu,B)}{\lambda\mu}\,d\mu\,d\lambda\ . \tag{111}$$

Proof. From (104) it follows that

$$A = E^A\left(\frac{1}{2\pi i}\int_{\Gamma_1}\lambda R(\lambda,A)d\lambda\right)E^A$$

and

$$B = E^B\left(\frac{1}{2\pi i}\int_{\Gamma_2}\mu R(\mu,B)d\mu\right)E^B\ .$$

Therefore,

$$
\begin{aligned}
AXB &= E^A \left[\frac{1}{2\pi i} \int_{\Gamma_1} \lambda R(\lambda, A) d\lambda \right] \\
&\quad \times E^A \left[\frac{1}{2\pi i} \int_{\Gamma_1} \frac{R(\lambda, A)}{\lambda} D \left(\frac{1}{2\pi i} \int_{\Gamma_2} \frac{R(\mu, B)}{\mu} d\mu \right) d\lambda \right] \\
&\quad \times E^B \left[\frac{1}{2\pi i} \int_{\Gamma_2} \mu R(\mu, B) d\mu \right] E^B \\
&= E^A \left[\frac{1}{2\pi i} \int_{\Gamma_1} R(\lambda, A) d\lambda \right] D \left[\frac{1}{2\pi i} \int_{\Gamma_2} R(\mu, B) d\mu \right] E^B \ , \\
&\qquad \text{by a double application of Lemma 3} \\
&= E^A (E^A)^* D (E^B)^* E^B \ , \quad \text{by (101) with } f \equiv 1 \\
&= P_{R(A)} D P_{R(B^*)} \ , \quad \text{by (60)} \\
&= AA^\dagger D B^\dagger B \\
&= D \quad \text{if and only if (108) is consistent, by Theorem 2.1 .}
\end{aligned}
$$

Alternatively, it follows from (102) that X in (111) is $X = A^\dagger D B^\dagger$, a solution of (108) if it is consistent. □

For additional results along these lines see Lancaster [**902**] and Wimmer and Ziebur [**1608**].

Exercises and examples.

Ex. 53. *Matrix functions: The classical definition.* For any $A \in \mathbb{C}^{n\times n}$ with spectrum $\sigma(A)$, let $F(A)$ denote the class of all functions $f : \mathbb{C} \to \mathbb{C}$ which are analytic in some open set containing $\sigma(A)$. For any scalar function which is analytic in some open set, the *corresponding matrix function* $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is defined, at those $A \in \mathbb{C}^{n\times n}$ such that $f \in F(A)$, by

$$f(A) = p(A) \ , \tag{112}$$

where $p(A)$ is any polynomial such that, for each $\lambda \in \sigma(A)$,

$$p^{(i)}(\lambda) = f^{(i)}(\lambda) \ , \quad i = 0, 1, \dots , \nu(\lambda) - 1 \tag{113}$$

where $\nu(\lambda)$ is the index (see Definition 4.1) of the matrix $A - \lambda I$, also called the *index of the eigenvalue* λ.

For other definitions of matrix functions, and their relations to the one given here, see Rinehart [**1274**]. Additional results and references on matrix functions are Dunford and Schwartz [**441**, pp. 556–565], Gantmacher [**533**], Frame [**508**], and Lancaster [**902**].

Ex. 54. If $A \in \mathbb{C}^{n\times n}$ is normal with a spectral representation

$$A = \sum_{i=1}^{r} \lambda_i P_i \tag{96}$$

then, for any $f \in F(A)$, definition (112) gives

$$f(A) = \sum_{i=1}^{r} f(\lambda_i) P_i + f(0) P_{N(A)} \ , \tag{114}$$

since the eigenvalues of a normal matrix have index one.

Ex. 55. *Generalized powers* The matrix function $f : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ corresponding to the scalar function

$$f(z) = z^k \ , \quad k \text{ any integer ,}$$

is denoted by

$$f(A) = A^{<k>}$$

and called the *generalized kth power* of $A \in \mathbb{C}^{m\times n}$. Definition 1 shows that

$$A^{<k>} = \sum_{i=1}^{r} d_i^k E_i \ , \ \text{by (97)} \tag{115}$$

or equivalently

$$A^{<k>} = U_{(r)} D_{(r)}^k V_{(r)}^* \ , \ \text{by (98)} \ . \tag{116}$$

The generalized powers of A satisfy

$$A^{<k>} = \begin{cases} E \ , & k = 0 \ , \\ A^{<k-1>} E^* A \ , & k \geq 1 \ , \ \text{in particular } A^{<1>} = A \ , \\ A^{<k+1>} E^* A^{<-1>} \ , & k \leq -1 \ . \end{cases} \tag{117}$$

Ex. 56. If in Theorem 6 the scalars $d(A)$ are chosen as the singular values of A, i.e., if $d(A) = \alpha(A)$, then for any integer k

$$A^{*<k>} = A^{<k>*} \ , \tag{118}$$

$$A^{<2k+1>} = A(A^*A)^k = (AA^*)^k A \ , \tag{119}$$

in particular

$$A^{<-1>} = A^{*\dagger} \ . \tag{120}$$

Ex. 57. If $A \in \mathbb{C}_r^{n\times n}$ is normal, and if the scalars $d(A)$ are chosen as the eigenvalues of A, i.e., if $d(A) = \sigma(A)$, then

$$A^{<k>} = \begin{cases} A^k \ , & k \geq 1 \ , \\ P_{R(A)} \ , & k = 0 \ , \\ (A^\dagger)^k \ , & k \leq -1 \ . \end{cases} \tag{121}$$

Ex. 58. *Ternary powers.* From (119) follows the definition of a polynomial in ternary powers of $A \in \mathbb{C}^{m\times n}$, as a polynomial

$$\sum_k p_k A^{<2k+1>} = \sum_k p_k (AA^*)^k A \ .$$

Such polynomials were studied by Hestenes [**726**] in the more general context of ternary algebras.

In (124) below, we express $A^\dagger$ as a polynomial in ternary powers of A^*. First we require the following.

Ex. 59. Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, and let a vanishing polynomial of A, i.e., a polynomial $m(\lambda)$ satisfying $m(A) = O$, be given in the form

$$m(\lambda) = c\lambda^\ell (1 - \lambda q(\lambda)) \tag{4.31}$$

where $c \neq 0$, $\ell \geq 0$, and the leading coefficient of q is 1.

Then

$$A^\dagger = q(A) + q(O)[Aq(A) - I] \ , \tag{122}$$

and in particular

$$A^{-1} = q(A) \ , \quad \text{if } A \text{ is nonsingular (Albert [\textbf{13}, p. 75]) .}$$

Proof. From (4.31) it follows that

$$A^\ell = A^{\ell+1}q(A)$$

and since A is Hermitian

$$\begin{aligned} A^\dagger &= (A^\dagger)^{\ell+1}A^\ell = AA^\dagger q(A) \\ &= AA^\dagger[q(A)-q(O)] + AA^\dagger q(O) \\ &= q(A) - q(O) + AA^\dagger q(O) \end{aligned} \tag{123}$$

since $q(A)-q(O)$ contains only positive powers of A. Postmultiplying (123) by A gives

$$\begin{aligned} A^\dagger A &= [q(A)-q(O)]A + Aq(O) \\ &= q(A)A = Aq(A)\ , \end{aligned}$$

which when substituted in (123), gives (122). □

Alternatively, (122) can be shown to follow from the results of Section 4.6, since here $A^D = A^\dagger$.

Ex. 60. Let $A \in \mathbb{C}^{m\times n}$ and let

$$m(\lambda) = c\lambda^\ell(1-\lambda q(\lambda)) \tag{4.31}$$

be a vanishing polynomial of A^*A, as in Ex. 59. Then

$$A^\dagger = q(A^*A)A^* \tag{124}$$

(Penrose [**1177**], Hestenes [**726**], Ben-Israel and Charnes [**126**]).

Proof. From (122) it follows that

$$(A^*A)^\dagger = q(A^*A) + q(O)[A^*Aq(A^*A) - I]\ ,$$

so, by Ex. 1.16(d),

$$A^\dagger = (A^*A)^\dagger A^* = q(A^*A)A^*\ .$$

□

A computational method based on (124) is given in Decell [**390**] and in Albert [**13**].

Ex. 61. *Partial isometries.* Let $W \in \mathbb{C}^{m\times n}$. Then W is a partial isometry if and only if

$$W = e^{iA}$$

for some $A \in \mathbb{C}^{m\times n}$.

Proof. Follows from (98) and Exs. 29–30. □

Ex. 62. Let $U \in \mathbb{C}^{n\times n}$. Then U is a unitary matrix if and only if

$$U = e^{iH} + P_{N(H)} \tag{125}$$

for some Hermitian matrix $H \in \mathbb{C}^{n\times n}$. Note that the exponential in (125) is defined according to Definition 1. For the classical definition given in Ex. 53, Eq. (125) should be replaced by

$$U = e^{iH}\ . \tag{125'}$$

Ex. 63. *Polar decompositions.* Let $A \in \mathbb{C}_r^{m\times n}$ and let

$$A = GE = EH \tag{61}$$

be a polar decomposition of A, given as in Corollary 2. Then for any function f, Definition 1 gives

$$f(A) = f(G)E = Ef(H)\ , \tag{126}$$

in particular

$$A^{<k>} = G^kE = EH^k\ , \quad \text{for any integer } k\ . \tag{127}$$

Suggested further reading

Section 2. Businger and Golub [**247**], Golub and Kahan [**553**], Golub and Reinsch [**557**], Good [**562**], Hartwig [**671**], Hestenes [**723**], Lanczos [**906**], and Wedin [**1540**].
Section 3. Erdélyi ([**467**], [**473**], [**471**], [**474**], [**472**]), Erdélyi and Miller [**478**], Halmos and Wallen [**648**], Hearon ([**707**], [**706**]), Hestenes ([**723**], [**724**], [**725**], [**726**]) and Poole and Boullion [**1195**].

CHAPTER 7

Computational Aspects of Generalized Inverses

1. Introduction

There are three principal situations in which it is required to obtain numerically a generalized inverse of a given matrix: (i) the case in which any $\{1\}$–inverse will suffice, (ii) the cases in which any $\{1,3\}$–inverse (or sometimes any $\{1,4\}$–inverse) will do, and (iii) the case in which a $\{2\}$–inverse having a specified range and null space is required.

The inverse desired in case (iii) is, in the majority of cases, the Moore–Penrose inverse, which is the unique $\{2\}$–inverse of the given matrix A having the same range and null space as A^*. The Drazin inverse can also be fitted into this pattern, being the unique $\{2\}$–inverse of A having the same range and null space as A^ℓ, where ℓ is any integer not less than the index of A. When $\ell = 1$, this is the group inverse.

Generalized inverses are closely associated with linear equations, orthonormalization, least squares solutions, singular values, and various matrix factorizations. These topics have been studied extensively, and many excellent references are available in the numerical analysis literature. for this reason we can keep this chapter brief, restricting our effort to listing some computational methods for generalized inversion, and discussing the mathematics behind these methods. No error analysis is attempted.

Iterative methods for generalized inversion are discussed in Section 5. The remaining sections deal with direct methods.

2. Computation of unrestricted $\{1\}$–inverses and $\{1,2\}$–inverses

Let A be a given matrix for which a $\{1\}$–inverse is desired, when any $\{1\}$–inverse will suffice. If it should happen that A is of such a structure, or has risen in such a manner, that a nonsingular submatrix of maximal order is known, we can write

$$PAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \tag{5.19}$$

where A_{11} is nonsingular and P and Q are permutation matrices used to bring the nonsingular submatrix into the upper left position. (If A is of full (column or row) rank, some of the submatrices in (5.19) will be absent.) Since rank A is the order of A_{11}, this implies that

$$A_{22} = A_{21}A_{11}^{-1}A_{12} \quad (\text{ Brand } [\mathbf{233}]) \tag{5.34}$$

and a $\{1,2\}$–inverse of A is

$$A^{(1,2)} = Q \begin{bmatrix} A_{11}^{-1} & O \\ O & O \end{bmatrix} P \quad (\text{ C. R. Rao } [\mathbf{1241}]) . \tag{5.25}$$

In the more usual case in which a nonsingular submatrix of maximal order is not known, and likewise, rank A is not known, perhaps the simplest method is that of Section 1.2, using Gaussian elimination to bring A to Hermite normal form. (It should be noted, however, that we are employing the “looser” definition of the Hermite normal form, Definition 0.1 (p. 22), and not the strict definition used in some texts, e.g., Marcus and Minc [**996**].) Thus if

$$EAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} \tag{0.46}$$

(with modifications in the case that A is of full rank), where E is nonsingular and P is a permutation matrix, then

$$A^{(1)} = P \begin{bmatrix} I_r & K \\ O & L \end{bmatrix} E \tag{1}$$

is a $\{1\}$–inverse of A for arbitrary L. Of course, the simplest choice is $L = O$, which gives the $\{1,2\}$–inverse

$$A^{(1,2)} = P \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} E \, .$$

On the other hand, when A is square, a nonsingular $\{1\}$–inverse may sometimes be desired. This is obtained by taking L in (1) to be nonsingular. The simplest choice for L is a unit matrix, which gives

$$A^{(1)} = PE \, .$$

If the calculations are performed on a computer, then, as in the nonsingular case, the accuracy may depend on the choice of pivots used in the Gaussian elimination. (For a discussion of pivoting see, e.g., Pennington [**1176**]; for a simple illustration, see Ex. 2 below.)

Exercises.

Ex. 1. Show that (1) gives a $\{1,2\}$–inverse of A if and only if $L = O$.

Ex. 2. Consider the two nonsingular matrices

$$A = \begin{bmatrix} \epsilon & 1 \\ 0 & 1 \end{bmatrix} , \quad B = \begin{bmatrix} \epsilon & 1 \\ 1 & 1 \end{bmatrix} ,$$

where ϵ is a small, positive number. Compare the various ways (i.e., choices of pivots) of transforming A and B to their Hermite normal forms. The objective is a numerically stable process, which here means to avoid, or to postpone, division by ϵ.

3. Computation of unrestricted $\{1,3\}$–inverses

Let $A \in \mathbb{C}_r^{m\times n}$ and let

$$A = FG \tag{2}$$

be a full–rank factorization. Then, by Ex. 1.25(b),

$$X = G^{(1)} F^{\dagger} \, , \tag{3}$$

where $G^{(1)}$ is an arbitrary element of $G\{1\}$, is a $\{1,2,3\}$–inverse of A. If the factorization (2) has been obtained from the Hermite normal form of A by the procedure described in Section 1.7, then

$$F = AP_1 \, , \tag{4}$$

where P_1 denotes the first r columns of the permutation matrix P. Moreover, we may take $G^{(1)} = P_1$, and (3) gives

$$X = P_1 F^{\dagger} \, . \tag{5}$$

Since F is of full column rank,

$$F^{\dagger} = (F^* F)^{-1} F^* \tag{6}$$

by (1.24). Thus (4), (6), and (5), in that order, give a $\{1,2,3\}$–inverse of A.

Note that (4) shows that F is a submatrix of A consisting of r linearly independent columns. In fact, the only purpose served by the computation of the Hermite normal form is in the selection of the r columns. Thus, the method is equally valid if r linearly independent columns have been determined in some other manner; see, e.g., Exs. 4–5 below.

Observe also that (5) shows that each of the r rows of $F^\dagger$ is a row of X (in general, not the corresponding row), while the remaining $n-r$ rows of X are rows of zeros. Thus, in the language of linear programming, X is a "basic" $\{1,2,3\}$–inverse of A.

Exercises.

Ex. 3. Use (4), (6), and (5) to obtain a $\{1\}$–inverse of

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} .$$

Ex. 4. *Gram–Schmidt orthogonalization.* Given a nonzero $A \in \mathbb{C}^{m\times n}$, a full column rank submatrix can be found by the *Gram–Schmidt orthogonalization process* (abbreviated GSO) as follows.

Applying GSO (without normalization) to the columns $[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ of A gives an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ of $R(A)$, where

$$\mathbf{v}_1 = \mathbf{a}_{c_1} \quad \text{if } \mathbf{a}_{c_1} \neq \mathbf{0} = \mathbf{a}_j \text{ for } 1 \leq j < c_1 \tag{7a}$$

$$\mathbf{x}_j = \mathbf{a}_j - \sum_{\ell=1}^{k-1} \frac{\langle \mathbf{a}_j, \mathbf{v}_\ell \rangle}{\|\mathbf{v}_\ell\|^2} , \quad j = c_{k-1}+1, c_{k-1}+2, \dots, c_k \tag{7b}$$

and

$$\mathbf{v}_k = \mathbf{x}_{c_k} \quad \text{if } \mathbf{x}_{c_k} \neq \mathbf{0} = \mathbf{x}_j \text{ for } c_{k-1}+1 \leq j < c_k \; , \; k = 2, \dots, r \; . \tag{7c}$$

Then

$$F = [\mathbf{a}_{c_1}, \mathbf{a}_{c_2}, \dots, \mathbf{a}_{c_r}] \; , \quad P_1 = [\mathbf{e}_{c_1}, \mathbf{e}_{c_2}, \dots, \mathbf{e}_{c_r}] \tag{8}$$

are two matrices satisfying (4).

Ex. 5. If F is given by (8) and (7), then

$$F^\dagger = \begin{bmatrix} \mathbf{a}_{c_1}^*/\|\mathbf{a}_{c_1}\|^2 \\ \mathbf{a}_{c_2}^*/\|\mathbf{a}_{c_2}\|^2 \\ \cdots\cdots \\ \mathbf{a}_{c_r}^*/\|\mathbf{a}_{c_r}\|^2 \end{bmatrix} \tag{9}$$

Thus the GSO gives everything needed in (5) to compute a $\{1,2,3\}$–inverse. See also Ex. 9 below.

Ex. 6. Use (7), (8), (9), and (5) to calculate a $\{1,2,3\}$–inverse of the matrix given in Ex. 3.

4. Computation of $\{2\}$–inverses with prescribed range and null space

Let $A \in \mathbb{C}_r^{m\times n}$, let $A\{2\}_{S,T}$ contain an nonzero matrix X, and let U and V be such that $R(U) = R(X)$, $N(V) = N(X)$, and the product VAU us defined. Then, by Theorems 2.11 and 2.12, rank U = rank V = rank VAU, and

$$X = U(VAU)^{(1)}V \; , \tag{10}$$

where $(VAU)^{(1)}$ is an arbitrary element of $(VAU)\{1\}$. This is the basic formula for the case considered in this section. Zlobec's formula

$$A^\dagger = A^*(A^*AA^*)^{(1)}A^* \tag{11}$$

(see Ex. 2.30) and Greville's formula

$$A^D = A^\ell (A^{2\ell+1})^{(1)} A^\ell \tag{4.35}$$

where ℓ is a positive integer not less than the index of A, are particular cases. Formula (10) has the advantage that it does not require inversion of any nonsingular matrix. Aside from matrix multiplication, only the determination of a $\{1\}$–inverse of VAU is needed, and this can be obtained by the method of Section 1.2.

It should be noted, however, that when ill–conditioning of A is a problem, this is accentuated by forming products like A^*AA^* or $A^{2\ell+1}$, and in such cases, other methods are preferable.

In the case of the Moore–Penrose inverse, Noble's formula

$$A^\dagger = Q \begin{bmatrix} I_r \\ T^* \end{bmatrix} (I_r + TT^*)^{-1} A_{11}^{-1} (I_r + S^*S)^{-1} \begin{bmatrix} I_r & S^* \end{bmatrix} P \tag{5.28}$$

is available, if a maximal nonsingular (and well–conditioned) submatrix A_{11} is known, where the permutation matrices P and Q and the "multipliers" S and T are defined by

$$\begin{aligned} A &= P^T \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} Q^T \\ &= P^T \begin{bmatrix} I_r \\ S \end{bmatrix} A_{11} \begin{bmatrix} I_r & T \end{bmatrix} Q^T \ ; \ \text{see Ex. 7.} \end{aligned} \tag{1.19}$$

Otherwise, it is probably best to use the method of Section 1.7 to obtain a full–rank factorization

$$A = FG \ . \tag{1.19}$$

Then, the Moore–Penrose inverse is

$$A^\dagger = G^*(F^*AG^*)^{-1}F^* \ , \tag{1.22}$$

while the group inverse is

$$A^\# = F(GF)^{-2}G \ , \tag{4.16}$$

whenever GF is nonsingular.

In the computation of A^D when the index of A exceeds 1, it is not easy to avoid raising A to a power. When ill–conditioning of A is serious, perhaps the best method is the sequential procedure of Cline [**352**], which involves full–rank factorization of matrices of successively smaller order, until a nonsingular matrix is reached. Thus, we take

$$A = B_1 G_1 \ , \tag{12}$$

$$G_i B_i = B_{i+1} G_{i+1} \quad (i = 1, 2, \dots , k-1) \ , \tag{13}$$

where k is the index of A. Then

$$A^D = B_1 B_2 \cdots B_k (G_k B_k)^{-k-1} G_k G_{k-1} \cdots G_1 \ . \tag{14}$$

Exercises.

EX. 7. *Noble's method.* Let the nonzero matrix $A \in \mathbb{C}_r^{m\times n}$ be transformed to a column–permuted Hermite normal form

$$PEAQ = \begin{bmatrix} I_r & \vdots & T \\ \cdots & \cdots & \cdots \\ O & \vdots & O \end{bmatrix} = (PEP^T)(PAQ) \tag{15}$$

where P and Q are permutation matrices and E is a product of elementary row matrices of types (i) and (ii) (see Section 1.2),

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

which does not involve permutation of rows.

Then E can be chosen so that

$$PE\begin{bmatrix}A & \vdots & I_m\end{bmatrix}\begin{bmatrix}Q & \vdots & O\\ \cdots & \cdots & \cdots\\ O & \vdots & P^T\end{bmatrix} = \begin{bmatrix}I_r & \vdots & T & \vdots & A_{11}^{-1} & \vdots & O\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ O & \vdots & O & \vdots & -S & \vdots & I_{m-r}\end{bmatrix} \tag{16}$$

giving all the matrices P, Q, T, S, and A_{11}^{-1} which appear in (5.28). Note that after the left–hand portion of RHS(16) has been brought to the form (15), still further row operations may be needed to bring the right–hand portion to the required form (Noble [**1144**]).

Ex. 8. *Singular value decomposition.* Let

$$A = U_{(r)} D_{(r)} V_{(r)}^* \tag{6.49}$$

be a singular value decomposition of $A \in \mathbb{C}^{m\times n}$. Then

$$\begin{aligned} A^\dagger &= V_{(r)} D_{(r)}^{-1} U_{(r)}^* \\ &= V_{(r)} (U_{(r)}^* A V_{(r)})^{-1} U_{(r)}^* \end{aligned} \tag{6.50}$$

is shown to be a special case of (10) by taking

$$U = V_{(r)} , \quad V = U_{(r)}^* .$$

A method for computing the Moore–Penrose inverse, based on (6.50), has been developed by Golub and Kahan [**553**]. See also Businger and Golub [**246**], [**247**] and Golub and Reinsch [**557**].

Ex. 9. *Gram–Schmidt orthogonalization.* The GSO of Exs. 4–5 can be modified to compute the Moore–Penrose inverse. This method is due to Rust and Burrus and Schneeberger [**1312**]; see also Albert [**13**, Chapter V].

Ex. 10. For the matrix A of Ex. 3, calculate $A^\dagger$ by:
(a) Zlobec's formula (11).
(b) Noble's formula (5.28).
(c) MacDuffee's formula (1.22).
(d) Greville's method, Section 5.5.

Ex. 11. For the matrix A of Ex. 3, calculate $A^\#$ by:
(a) Formula (4.23).
(b) Cline's formula (4.16).

Ex. 12. Show that, for a matrix A of index k that is not nilpotent, with B_i and G_i defined by (12) and (13), $G_k B_k$ is nonsingular. [*Hints*: Express A^k and A^{k+1} in terms of B_i and G_i ($i = 1, 2, \dots, k$), and let r_k denote the number of columns of B_k, which is also the number of rows of G_k. Show that $\operatorname{rank} A^k = r_k$, while $\operatorname{rank} A^{k+1} = \operatorname{rank} G_k B_k$. Therefore $\operatorname{rank} A^{k+1} = \operatorname{rank} A^k$ implies that $G_k B_k$ is nonsingular.]

Ex. 13. Use Theorem 4.7 to verify (14).

5. Iterative methods for computing $A^\dagger$

An iterative method for computing $A^\dagger$ is a set of instructions for generating a sequence $\{X_k : k = 1, 2, \dots\}$ converging to $A^\dagger$. The instructions specify how to select the initial approximation X_0, how to proceed from X_k to X_{k+1} for each k, and when to stop, having obtained a reasonable approximation of $A^\dagger$.

The rate of convergence of such an iterative method is determined in terms of the corresponding sequence of *residuals* $\{R_k : k = 0, 1, \dots\}$

$$R_k = P_{R(A)} - AX_k , \quad k = 0, 1, \dots \tag{17}$$

which converges to O as $X_k \to A^\dagger$. An iterative method is said to be a *p* $\underline{th}$*–order method*, for some $p > 1$, if there is a positive constant c such that

$$\|R_{k+1}\| \leq c\|R_k\|^p \ , \quad k = 0, 1, \dots \tag{18}$$

for any multiplicative matrix norm; see, e.g., Ex. 0.27.

In analogy with the nonsingular case – see, e.g., Householder [**753**, pp. 94–95] – we consider iterative methods of the type

$$X_{k+1} = X_k + C_k R_k \ , \quad k = 0, 1, \dots \ , \tag{19}$$

where $\{C_k \ : \ k = 0, 1, \dots\}$ is a suitable sequence, and X_0 is the initial approximation (to be specified).

One objection to (19) as an iterative method for computing $A^\dagger$ is that (19) requires at each iteration the residual R_k, for which one needs the projection $P_{R(A)}$, whose computation is a task comparable to computing $A^\dagger$. This difficulty will be overcome here by choosing the sequence $\{C_k\}$ in (19) to satisfy

$$C_k = C_k P_{R(A)} \ , \quad k = 0, 1, \dots \tag{20}$$

For such a choice we have

$$\begin{aligned} C_k R_k &= C_k \left(P_{R(A)} - AX_k\right) \ , \quad \text{by (17)} \\ &= C_k \left(I - AX_k\right) \ , \quad \text{by (20)} \ , \end{aligned} \tag{21}$$

and (19) can therefore be rewritten as

$$X_{k+1} = X_k + C_k T_k \ , \quad k = 0, 1, \dots \tag{22}$$

where

$$T_k = I - AX_k \ , \quad k = 0, 1, \dots \tag{23}$$

The iterative method (19), or (22), is suitable for the case where A is an $m \times n$ matrix with $m \leq n$, for then R_k and T_k are $m \times m$ matrices. However, if $m > n$ the following dual version of (19) is preferable to it

$$X'_{k+1} = X'_k + R'_k C'_k \ , \quad k = 0, 1, \dots \ , \tag{19$'$}$$

where

$$R'_k = P_{R(A^*)} - X'_k A \tag{17$'$}$$

and $\{C'_k : k = 0, 1, \dots\}$ is a suitable sequence, satisfying

$$C'_k = P_{R(A^*)} C'_k \ , \quad k = 0, 1, \dots \tag{20$'$}$$

a condition which allows rewriting (19$'$) as

$$X'_{k+1} = X'_k + T'_k C'_k \ , \quad k = 0, 1, \dots \tag{22$'$}$$

where

$$T'_k = I - X'_k A \ , \quad k = 0, 1, \dots \tag{23$'$}$$

Indeed, if $m > n$ then (22$'$) is preferable to (22), for the former method uses the $n \times n$ matrix T'_k while the latter uses T_k, which is an $m \times m$ matrix.

Since all the results and proofs pertaining to the iterative method (19) or (22) hold true, with obvious modifications, fo r the dual method (19′) or (22′), we will, for the same of convenience, restrict the discussion to the case

$$m \leq n \,, \tag{24}$$

leaving to the reader the details of the complementary case.

A *first–order iterative method for computing* $A^\dagger$, of type (22), is presented in the following.

Theorem 1. Let $O \neq A \in \mathbb{C}^{m\times n}$ and let the initial approximation X_0 and its residual R_0 satisfy

$$X_0 \in R(A^*, A^*) \tag{25}$$

(i.e. $X_0 = A^*BA^*$ for some $B \in \mathbb{C}^{m\times n}$, see Ex. 3.25, p. 96), and

$$\rho(R_0) < 1 \tag{26}$$

respectively. Then the sequence

$$\begin{aligned} X_{k+1} &= X_k + X_0T_k \\ &= X_k + X_0(I - AX_k) \,, \quad k = 0, 1, \dots \end{aligned} \tag{27}$$

converges to $A^\dagger$ as $k \to \infty$, and the corresponding sequence of residuals satisfies

$$\|R_{k+1}\| \leq \|R_0\|\|R_k\| \,, \quad k = 0, 1, \dots \tag{28}$$

for any multiplicative matrix norm.

Proof. The sequence (27) is obtained from (22) by choosing

$$C_k = X_0 \,, \quad k = 0, 1, \dots \tag{29}$$

a choice which, by (25), satisfies (20), and allows rewriting (27) as

$$\begin{aligned} X_{k+1} &= X_k + X_0R_k \\ &= X_k + X_0(P_{R(A)} - AX_k) \,, \quad k = 0, 1, \dots \end{aligned} \tag{30}$$

From (30) we compute the residual

$$\begin{aligned} R_{k+1} &= P_{R(A)} - AX_{k+1} \\ &= P_{R(A)} - AX_k - AX_0R_k \\ &= R_k - AX_0R_k \\ &= P_{R(A)}R_k - AX_0R_k \,, \quad \text{by (17)} \\ &= R_0R_k \,, \quad k = 0, 1, \dots \\ &= R_0^{k+2} \,, \quad \text{by repeating the argument.} \end{aligned} \tag{31}$$

For any multiplicative matrix norm, it follows from (31) that

$$\|R_{k+1}\| \leq \|R_0\|\|R_k\| \,. \tag{28}$$

From

$$R_{k+1} = R_0^{k+2} \,, \quad k = 0, 1, \dots \tag{31}$$

it also follows, by using (26) and Ex. 0.38, that the sequence of residuals converges to the zero matrix:

$$P_{R(A)} - AX_k \to O \quad \text{as} \quad k \to \infty \,. \tag{32}$$

We will prove now that the sequence (27) converges. Rewriting the sequence (27) as

$$X_{k+1} = X_k + X_0 R_k \ , \tag{30}$$

it follows from (31) that

$$\begin{aligned} X_{k+1} &= X_k + X_0 R_0^{k+1} \\ &= X_{k_1} + X_0 R_0^k + X_0 R_0^{k+1} \\ &= X_0 \left(I + R_0 + R_0^2 + \cdots + R_0^{k+1}\right) , \quad k = 0, 1, \ldots \end{aligned} \tag{33}$$

which, by (26) and Exs. 0.38–0.39, converges to a limit X_∞.

Finally we will show that $X_\infty = A^\dagger$. From (32) it follows that

$$AX_\infty = P_{R(A)} \ ,$$

and in particular, that X_∞ is a $\{1\}$–inverse of A. From (25) and (27) it is obvious that all X_k lie in $R(A^*, A^*)$, and therefore

$$X_\infty \in R(A^*, A^*) \ ,$$

proving that $X_\infty = A^\dagger$, since $A^\dagger$ is the unique $\{1\}$–inverse which lies in $R(A^*, A^*)$; see Ex. 3.28. □

For any integer $p \geq 2$, a *p th–order iterative method for computing* $A^\dagger$, of type (22), is described in the following.

Theorem 2. Let $O \neq A \in \mathbb{C}^{m\times n}$ and let the initial approximation X_0 and its residual R_0 satisfy (25) and (26), respectively. Then for any integer $p \geq 2$, the sequence

$$\begin{aligned} X_{k+1} &= X_k \left(I + T_k + T_k^2 + \cdots + T_k^{p-1}\right) \\ &= X_k \left(I + (I - AX_k) + (I - AX_k)^2 + \cdots + (I - AX_k)^{p-1}\right) , \ k = 0, 1, \ldots \end{aligned} \tag{34}$$

converges to $A^\dagger$ as $k \to \infty$, and the corresponding sequence of residuals satisfies

$$\|R_{k+1}\| \leq \|R_k\|^p \ , \quad k = 0, 1, \ldots \tag{35}$$

Proof. The sequence (34) is obtained from (22) by choosing

$$C_k = X_k \left(I + T_k + T_k^2 + \cdots + T_k^{p-2}\right) . \tag{36}$$

From (25) and (34) it is obvious that all the X_k lie in $R(A^*, A^*)$, and therefore the sequence $\{C_k\}$, given by (36), satisfies (20), proving that the sequence (34) can be rewritten in tyhe form (19)

$$X_{k+1} = X_k \left(I + R_k + R_k^2 + \cdots + R_k^{p-1}\right) , \ k = 0, 1, \ldots \tag{37}$$

From (37) we compute

$$\begin{aligned} R_{k+1} &= P_{R(A)} - AX_{k+1} \\ &= P_{R(A)} - AX_k \left(I + R_k + R_k^2 + \cdots + R_k^{p-1}\right) \\ &= R_k - AX_k \left(R_k + R_k^2 + \cdots + R_k^{p-1}\right) , \end{aligned} \tag{38}$$

Now for any $j = 1, \ldots, p-1$

$$\begin{aligned} R_k^j - AX_k R_k^j &= P_{R(A)} R_k^j - AX_k R_k^j \\ &= R_k R_k^j = R_k^{j+1} \ , \end{aligned}$$

and therefore, the last line in (38) collapses to

$$R_{k+1} = R_k^p \ , \tag{39}$$

which implies (35). The remainder of the proof, namely, that the sequence (37) converges to $A^\dagger$, can be given analogously to the proof of Theorem 1. $\square$

The iterative methods (27) and (34) are related by the following:

THEOREM 3. Let $O \neq A \in \mathbb{C}^{m\times n}$ and let the sequence $\{X_k : k = 0, 1, \dots\}$ be constructed as in Theorem 1. Let p be any integer ≥ 2, and let a sequence $\{\widetilde{X}_j : j = 0, 1, \dots\}$ be constructed as in Theorem 2 with the same initial approximation X_0 as the first sequence

$$\widetilde{X}_0 = X_0 \ ,$$
$$\widetilde{X}_{j+1} = \widetilde{X}_j \, (I + \widetilde{T}_j + \widetilde{T}_j^2 + \cdots + \widetilde{T}_j^{p-1}) \ , \ j = 0, 1, \dots \ , \tag{34}$$

where

$$\widetilde{T}_j = I - A\widetilde{X}_k \ , \ j = 0, 1, \dots \tag{23}$$

Then

$$\widetilde{X}_j = X_{p^j - 1} \ , \ j = 0, 1, \dots \tag{40}$$

PROOF. We use induction on j to prove (40), which obviously holds for $j = 0$. Assuming

$$\widetilde{X}_j = X_{p^j - 1} \ , \tag{40}$$

we will show that

$$\widetilde{X}_{j+1} = X_{p^{j+1} - 1} \ .$$

From

$$X_k = X_0 \, (I + R_0 + R_0^2 + \cdots + R_0^k) \tag{33}$$

and (40), it follows that

$$\widetilde{X}_j = X_0 \, (I + R_0 + R_0^2 + \cdots + R_0^{p^j - 1}) \ . \tag{41}$$

Rewriting (34) as

$$\widetilde{X}_{j+1} = \widetilde{X}_j \, (I + \widetilde{R}_j + \widetilde{R}_j^2 + \cdots + \widetilde{R}_j^{p-1}) \ , \tag{37}$$

it follows from

$$\begin{aligned} \widetilde{R}_j &= P_{R(A)} - A\widetilde{X}_j \\ &= P_{R(A)} - AX_{p^j - 1} \ , \ \text{by (40)} \ , \\ &= R_{p^j - 1} \\ &= R_0^{p^j} \ , \quad \text{by (31)} \ , \end{aligned}$$

that

$$\begin{aligned} \widetilde{X}_{j+1} &= \widetilde{X}_j \, (I + R_0^{p^j} + R_0^{2p^j} + \cdots + R_0^{(p-1)p^j}) \\ &= X_0 \, (I + R_0 + R_0^2 + \cdots + R_0^{p^j - 1})(I + R_0^{p^j} + R_0^{2p^j} + \cdots + R_0^{(p-1)p^j}) \ , \ \text{by (41)} \ , \\ &= X_0 \, (I + R_0 + R_0^2 + \cdots + R_0^{p^{j+1} - 1}) \\ &= X_{p^{j+1} - 1} \ , \quad \text{by (33)} \ . \end{aligned}$$

$\square$

Theorem 3 shows that an approximation $\widetilde{X}_j$ obtained by the p th–order method (34) in j iterations, will require $p^j - 1$ iterations of the 1 st–order method (27), both methods using the same initial approximation. For any two iterative methods of different orders, the higher–order method will, in general, require fewer iterations but more computations per iteration. A discussion of the optimal order p for methods of type (34) is given in Ex. 20.

Exercises and examples.

Ex. 14. The condition

$$X_0 \in R(A^*, A^*) \tag{25}$$

is necessary for the convergence of the iterative methods (27) and (34): let

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \epsilon \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \ \epsilon \neq 0 ,$$

and let

$$X_0 = A + B ,$$

Then

$$R_0 = P_{R(A)} - AX_0 = O$$

and in particular (26) holds, but

$$X_0 \notin R(A^*, A^*)$$

and both sequences (27) and (34) reduce to

$$X_k = X_0 , \quad k = 0, 1, \dots ,$$

without converging to $A^\dagger$.

Ex. 15. Let $O \neq A \in \mathbb{C}^{m\times n}$, and let X_0 and $R_0 = P_{R(A)} - AX_0$ satisfy

$$X_0 \in R(A^*, A^*) , \tag{25}$$

$$\rho(R_0) < 1 . \tag{26}$$

Then

$$A^\dagger = X_0 (I - R_o)^{-1} . \tag{42}$$

PROOF. The proof of Theorem 1 shows $A^\dagger$ to be the limit of

$$X_k = X_0 \left(I + R_0 + R_o^2 + \cdots + R_0^k\right) \tag{33}$$

as $k \to \infty$. But the sequence (33) converges, by Ex. 0.39, to RHS(42). □

The special case $X_0 = \beta A^*$

A frequent choice of the initial approximation X_0, in the iterative methods (27) and (34), is

$$X_0 = \beta A^* \tag{43}$$

for a suitable real scalar β. This special case is treated in the following three exercises.

Ex. 16. Let $O \neq A \in \mathbb{C}_r^{m\times n}$, let β be a real scalar, and let

$$R_0 = P_{R(A)} - \beta AA^* \,,$$
$$T_0 = I - \beta AA^* \,.$$

Then the following are equivalent.

(a) The scalar β satisfies

$$0 < \beta < \frac{2}{\lambda_1(AA^*)} \,, \tag{44}$$

where

$$\lambda_1(AA^*) \geq \lambda_2(AA^*) \geq \cdots \geq \lambda_r(AA^*) > 0$$

are the nonzero eigenvalues of AA^*.

(b) $\rho(R_0) < 1$.

(c) $\rho(T_0) \leq 1$ and $\lambda = -1$ is not an eigenvalue of T_0.

Proof. The nonzero eigenvalues of R_0 and T_0 are among

$$\{1 - \beta\lambda_i(AA^*) : i = 1, \dots, r\}$$

and

$$\{1 - \beta\lambda_i(AA^*) : i = 1, \dots, m\}$$

respectively. The equivalence of (a), (b), and (c) then follows from the observation that (44) is equivalent to

$$|1 - \beta\lambda_i(AA^*)| < 1 \,, \quad i = 1, \dots, r \,.$$

□

Ex. 17. Let $O \neq A \in \mathbb{C}_r^{m\times n}$, and let the real scalar β satisfy

$$0 < \beta < \frac{2}{\lambda_1(AA^*)} \,. \tag{44}$$

Then:

(a) The sequence

$$X_0 = \beta A^* \,, \quad X_{k+1} = X_k\,(I - \beta AA^*) + \beta A^* \,, \ k = 0, 1, \dots \tag{45}$$

or equivalently

$$X_k = \beta \sum_{j=0}^{k} A^*\,(I - \beta AA^*)^j \,, \ k = 0, 1, \dots \tag{46}$$

is a first–order method for computing $A^\dagger$.

(b) The corresponding residuals $R_k = P_{R(A)} - AX_k$ are given by

$$R_k = (P_{R(A)} - \beta AA^*)^{k+1} \,, \ k = 0, 1, \dots \tag{47}$$

(c) For any k, the spectral norm of R_k, $\|R_k\|_2$, is minimized by choosing

$$\beta = \frac{2}{\lambda_1(AA^*) + \lambda_r(AA^*)} \,, \tag{48}$$

in which case the minimal $\|R_k\|_2$ is

$$\|R_k\|_2 = \left(\frac{\lambda_1(AA^*) - \lambda_r(AA^*)}{\lambda_1(AA^*) + \lambda_r(AA^*)}\right)^{k+1} \,, \ k = 0, 1 \dots \tag{49}$$

PROOF. (a) Substituting (43) in (27) results in (45) or equivalently in (46).
(b) Follows from (46).
(c) R_k is Hermitian and therefore

$$\begin{aligned} \|R_k\|_2 &= \rho(R_k)\ , \text{ by Ex. 0.38} \\ &= \rho(R_0^{k+1})\ , \text{ by (31)} \\ &= \rho^{k+1}(R_0)\ , \text{ by Ex. 0.37}\ . \end{aligned}$$

Thus, $\|R_k\|_2$ is minimized by the same β that minimizes $\rho(R_0)$. Since the nonzero eigenvalues of $R_0 = P_{R(A)} - \beta AA^*$ are

$$1 - \beta\lambda_i(AA^*)\ ,\ i = 1, \dots, r$$

it is clear that β minimizes

$$\rho(R_0) = \max\{|1 - \beta\lambda_i(AA^*)| : i = 1, \dots, r\}$$

if and only if

$$-(1 - \beta\lambda_1(AA^*)) = 1 - \beta\lambda_r(AA^*)\ , \tag{50}$$

which is (48). Finally (49) is obtained by substituting (48) in

$$\begin{aligned} \rho(R_k) &= \max\left\{|1 - \beta\lambda_i(AA^*)|^{k+1} : i = 1, \dots, r\right\}\ , \text{ by (47)} \\ &= |1 - \beta\lambda_r(AA^*)|^{k+1}\ , \text{ for } \beta \text{ satisfying (50)}\ . \end{aligned}$$

□

EX. 18. Let A, β be as in Ex. 17. Then for any integer $p \geq 2$, the sequence

$$X_{k+1} = X_k\left(I + T_k + T_k^2 + \cdots + T_k^{p-1}\right) \tag{34}$$

with

$$X_0 = \beta A^* \tag{43}$$

is a p th–order method for computing $PA^\dagger$. The corresponding residuals are

$$R_k = (P_{R(A)} - \beta AA^*)^{p^{k+1}}$$

and their spectral norms are minimized by β of (48). The iterative methods of Exs. 17 and 18 were studied by Ben-Israel and Cohen [**130**], Petryshyn [**1183**], and Zlobec [**1650**].

EX. 19. *A second–order iterative method.* An important special case of Theorem 2 is the case $p = 2$, resulting in the following second–order iterative method for computing $A^\dagger$. Let $O \neq A \in \mathbb{C}^{m\times n}$ and let the initial approximation X_0 and its residual R_0 satisfy (25) and (26), respectively. Then the sequence

$$X_{k+1} = X_k\left(2I - AX_k\right)\ , \quad k = 0, 1, \dots \tag{51}$$

converges to $A^\dagger$ as $k \to \infty$, and the corresponding sequence of residuals satisfies

$$\|R_{k+1}\| \leq \|R_k\|^2\ , \quad k = 0, 1, \dots \quad \square \tag{52}$$

The iterative method (51) is a generalization of the well–known method of Schulz [**1323**] for the iterative inversion of a nonsingular matrix, see, e.g., Householder [**753**, p. 95]. The method (51) was studied by Ben–Israel [**108**], Ben–Israel and Cohen [**130**], Petryshyn [**1183**], and Zlobec [**1650**]. A detailed error analysis of (51) is given in Söderstörm and Stewart [**1376**].

EX. 20. *Discussion of the optimum order p.* As in Theorem 3 we denote by $\{X_k\}$ and $\{\widetilde{X}_k\}$ the sequences generated by the 1 st–order method (27) and by the p th–order method (34), respectively, using the same initial approximation $X_0 = \widetilde{X}_0$. Taking the sequence $\{X_k\}$ as the standard for comparing different orders p in (34), we use (40) to conclude that for each $k = 0, 1, \dots$, the smallest integer $\widetilde{k}$ such that the iterate $\widetilde{X}_{\widetilde{k}}$ is beyond X_k is the smallest integer $\widetilde{k}$ satisfying

$$p^{\widetilde{k}} - 1 \geq k$$

and therefore

$$\widetilde{k} = \langle \ln(k+1)/\ln p \rangle \ , \tag{53}$$

where for any real α, $\langle\alpha\rangle$ is the smallest integer $\geq \alpha$.

In assessing the computational effort per iteration, we assume that the amount of computational effort required to add or subtract an identity matrix is negligible compared to the effort to perform a matrix multiplication. Assuming (24) and hence the usage of the methods (27) and (34), rather than their duals based on (22′), we define a unit of computational effort as the effort required to multiply two $m \times m$ matrices. Accordingly, premultiplying an $n \times m$ matrix by an $m \times n$ matrix requires n/m units, as does the premultiplication of an $m \times m$ matrix by an $n \times m$ matrix. The iteration

$$\begin{aligned} X_{k+1} &= X_k\,(I + T_k + T_k^2 + \cdots + T_k^{p-1}) \\ &= X_k\,(I + T_k(I + \cdots + T_k(I + T_k)\cdots)) \end{aligned} \tag{34}$$

thus requires:

n/m units of effort to compute T_k,
$p-2$ units of effort to compute $T_k(I + \cdots + T_k(I + T_k)\cdots))$,
n/m units of effort to multiply $X_k\,(I + \cdots + T_k^{p-1})$, adding to

$$p - 2 + 2\frac{n}{m} \tag{54}$$

units of effort.

The figure (54) can be improved for certain p. For example, the iteration (34) can be written for $p = 2^q$, $q = 1, 2, \dots$, as

$$\begin{aligned} X_{k+1} &= X_k \prod_{j=1}^{2^{q-1}} (I + T_k^j) \\ &= X_k\,(I + T_k)(I + T_k^2)(I + T_k^4)\cdots(I + T_k^{2^{q-1}}) \end{aligned} \tag{55}$$

requiring only

$$2(q-1) + 2\frac{n}{m} \tag{56}$$

units of effort, improving on (54) for all $q \geq 3$; see also Lonseth [**969**].

In comparing the first–order iterative method (27) and the second–order method (51) (obtained from (34) for $p = 2$) one sees that both methods require $2(n/m)$ units of effort per iteration. Therefore, by Theorem 3, the second–order method (51) is superior to the first–order method (27).

For a given integer $k = 1, 2, \dots$ we define the *optimal order p* as as the order of the iterative method (34) which, starting with an initial approximation X_0, minimizes the computational effort required to obtain, or go beyond, the approximation X_k, obtained by the first–order method (27) in k iterations.

Combining (53), (54), and (55) it follows that for a given k, the optimal p is the integer p minimizing

$$\left(p - 2 + 2\frac{n}{m}\right) \langle \frac{\ln(k+1)}{\ln p} \rangle \ , \quad p = 2, 3, \dots \ , \ p \neq 2^q \ , \ q = 1, 2, \dots \tag{57}$$

or

$$\left(2q - 2 + 2\frac{n}{m}\right) \langle \frac{\ln(k+1)}{q \ln 2} \rangle \ , \quad p = 2^q \ , \ q = 1, 2, \dots \tag{58}$$

Lower bounds for (57) and (58) are

$$\ln(k+1) \frac{p - 2 + 2(n/m)}{\ln p} \ , \quad p = 2, 3, \dots \ , \ p \neq 2^q \ , \ q = 1, 2, \dots \tag{57$'$}$$

and

$$\ln(k+1) \frac{2q - 2 + 2(n/m)}{q \ln 2} \ , \quad p = 2^q \ , \ q = 1, 2, \dots \tag{58$'$}$$

respectively, suggesting the following definition which is independent of k. The *approximate optimum order* p is the integer $p \geq 2$ minimizing

$$f(p) = \begin{cases} \dfrac{p - 2 + 2(n/m)}{\ln p} & p \neq 2^q \ , \ p = 1, 2, \dots \\ \dfrac{2(q - 1 + (n/m)}{q \ln 2} & p = 2^q \ , \ q = 1, 2, \dots \end{cases} \tag{59}$$

The approximate optimum order p depends on the ratio n/m.

Ex. 21. *Iterative methods for computing projections.* Since $AA^\dagger = P_{R(A)}$, it follows that for any sequence $\{X_k\}$, the sequence $\{Y_k = AX_k\}$ satisfies

$$Y_k \to P_{R(A)} \quad \text{if} \quad X_k \to A^{|dag} \ .$$

Thus, for any iterative method for computing $A^\dagger$ defined by a sequence of successive approximations $\{X_k\}$, there is an associated iterative method for computing $P_{R(A)}$ defined by the sequence $\{Y_k = AX_k\}$. Similarly, an iterative method for computing $P_{R(A^*)}$ is given by the sequence $\{Y'_k = X_k A\}$ since $A^\dagger A = P_{R(A^*)}$.

The residuals R_k, $k = 0, 1, \dots$, of the sequence $\{Y_k\}$ will still be defined by (1), or equivalently

$$R_k = P_{R(A)} - Y_k \ , \quad k = 0, 1, \dots \tag{60}$$

Therefore, the iterative method $\{Y_k = AX_k\}$ for computing $P_{R(A)}$ is of the same order as the iterative method $\{X_k\}$ for computing $A^\dagger$.

In particular, a $p\,\underline{\text{th}}$–order iterative method for computing $P_{R(A)}$, based on Theorem 2, is given as follows.

Let $O \neq A \in \mathbb{C}^{m \times n}$ and let the initial approximation Y_0 and its residual R_0 satisfy

$$Y_0 \in R(A, A^*) \tag{61}$$

$$\rho(R_0) < 1 \ , \tag{26}$$

respectively. Then for any integer $p \geq 2$, the sequence

$$Y_{k+1} = Y_k \left(I + T_k + T_k^2 + \cdots + T_k^{p-1}\right) \tag{62}$$

with

$$T_k = I - Y_k \ , \quad k = 0, 1, \dots$$

converges to $P_{R(A)}$ as $k \to \infty$, and the corresponding sequence of residuals (60) satisfies

$$\|R_{k+1}\| \leq \|R_k\|^p \ , \quad k = 0, 1, \dots \ , \tag{35}$$

for any multiplicative matrix norm.

Ex. 22. *A monotone property of (62).* Let $O \neq A \in \mathbb{C}^{m\times n}$, let p be an even positive integer, and let the sequence $\{Y_k\}$ be given by (62), (61), (26), and the additional condition that Y_0 be Hermitian. Then the sequence $\{\text{trace}\, Y_k : \ k = 1, 2, \dots\}$ is monotone increasing and converges to rank A.

Proof. From (60), (31), and Theorem 3 it follows that

$$Y_k = P_{R(A)} - R_0^{p^k} \ . \tag{63}$$

From the fact that the trace of a matrix equals the sum of its eigenvalues, it follows that

$$\text{trace}\, P_{R(A)} = \dim\, R(A) = \text{rank}\, A$$

and

$$\begin{aligned}\text{trace}\, R_0^{p^k} &= \sum_{i=1}^{m} \lambda_i(R_0^{p^k}) \\ &= \sum_{i=1}^{m} \lambda_i^{p^k}(R_0)\end{aligned}$$

which is a monotone decreasing sequence converging to zero, since p is even, R_0 is Hermitian (by (60) and the assumption that Y_0 is Hermitian), and therefore its eigenvalues $\lambda_i(R_0)$, which by (26) have moduli less than 1, are real. The proof is completed by noting that, by (63),

$$\begin{aligned}\text{trace}\, Y_k &= \text{trace}\, P_{R(A)} - \text{trace}\, R_0^{p^k} \\ &= \text{rank}\, A - \sum_{i=1}^{m} \lambda_i^{p^k}(R_0) \ .\end{aligned}$$

$\square$

Ex. 23. *A lower bound on rank A.* Let $O \neq A \in \mathbb{C}^{m\times n}$ and let the sequence $\{Y_k : \ k = 0, 1, \dots\}$ be as in Ex. 22. Then

$$\text{rank}\, A \geq \langle \text{trace}\, Y_k \rangle \ , \quad k = 1, 2, \dots \tag{64}$$

where $\langle \alpha \rangle$ is the smallest integer $\geq \alpha$.

Ex. 24. *Iterative methods for computing matrix products involving generalized inverses.* In some applications one has to compute a matrix product $A^\dagger B$ or $BA^\dagger$, where $A \in \mathbb{C}^{m\times n}$ is given, and B is a given matrix or vector. The iterative methods for computing $A^\dagger$ given above can be adapted for computing such products.

Consider, for example, the iterative method

$$X_{k+1} = X_k \left(I + T_k + T_k^2 + \cdots + T_k^{p-1}\right) \ , \quad k = 0, 1, \dots \tag{34}$$

where p is an integer ≥ 2,

$$T_k = I - AX_k \ , \quad k = 0, 1, \dots \tag{23}$$

and the initial approximation X_0 satisfies (25) and (26). A corresponding iterative method for computing $BA^\dagger$, for a given $B \in \mathbb{C}^{q\times n}$, is given as follows.

Let $X_0 \in \mathbb{C}^{n\times m}$ satisfy (25) and (26) and let the sequence $\{Z_k : \ k = 0, 1, \dots\}$ be given by

$$Z_0 = BX_0 \ , \tag{65}$$

$$Z_{k+1} = Z_k M_k \ , \quad k = 0, 1, \dots \tag{66}$$

where

$$M_k = I + T_k + T_k^2 + \cdots + T_k^{p-1}\ , \quad k = 0, 1, \dots \tag{67}$$

$$T_{k+1} = I + M_k(T_k - I)\ , \quad k = 0, 1, \dots \tag{68}$$

and

$$T_0 = I - AX_0\ .$$

Then the sequence $\{Z_k\}$ converges to $BA^\dagger$ as $k \to \infty$ (Garnett, Ben–Israel and Yau [**534**]).

Suggested further reading

Section 4. Albert [**13**], Businger and Golub [**247**], Decell [**389**], Germain–Bonne [**539**], Golub and Reinsch [**557**], Graybill, Meyer and Painter [**572**], Ijiri [**766**], Kublanovskaya [**889**], Noble ([**1144**], [**1146**]), Pereyra and Rosen [**1181**], Peters and Wilkinson [**1182**], Pyle [**1222**], Shinozaki, Sibuya and Tanabe [**1351**], Stallings and Boullion [**1388**], Tewarson ([**1439**], [**1444**], [**1445**]), Urquhart [**1480**], and Willner [**1601**].

Section 5. Kammerer and Nashed ([**814**], [**813**], [**815**], [**816**]), Nashed ([**1114**], [**1115**]), Showalter [**1353**], Showalter and Ben–Israel [**1354**], Whitney and Meany [**1591**], and Zlobec ([**1650**], [**1653**]).

CHAPTER 8

Generalized Inverses of Linear Operators between Hilbert Spaces

1. Introduction

The observation that generalized inverses are like prose ("Good Heavens! For more than forty years I have been speaking prose without knowing it" – Molière, *Le Bourgois Gentilhomme*) is nowhere truer than in the literature of linear operators. In fact, generalized inverses of integral and differential operators were studied by Fredholm, Hilbert, Schmoidt, Bounitzky, Hurwitz, and others, before E. H. Moore introduced generalized inverses in an algebraic setting; see, e.g., the historic survey in Reid [**1263**].

This chapter is a brief and biased introduction to generalized inverses of linear operators between Hilbert spaces, with special emphasis on the similarities to the finite–dimensional case. Thus the spectral theory of such operators is omitted.

Following the preliminaries in Section 2, generalized inverses are introduced in Section 3. Applications to integral and differential operators are sampled in Exs. 18–37. The minimization properties of generalized inverses are studied in Section 4. Integral and series representations of generalized inverses, and iterative methods for their computation are given in Section 5.

This chapter requires familiarity with the basic concepts of linear functional analysis, in particular, the theory of linear operators in Hilbert space.

2. Hilbert spaces and operators: Preliminaries and notation

In this section we have collected, for convenience, some preliminary results, which can be found, in the form stated here or in a more general form, in the standard texts on functional analysis; see, e.g., Taylor [**1436**] and Yosida [**1623**].

(A) Our Hilbert spaces will be denoted by $\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$, etc. In each space, the *inner product* of two vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\langle x, y\rangle$ and the *norm* is denoted by $\|\ \|$. The *closure* of a subset L of $\mathcal{H}$ will be denoted by $\overline{L}$ and its *orthogonal complement* by $L^\perp$. $L^\perp$ is a closed subspace of $\mathcal{H}$, and

$$L^\perp = \overline{L}^\perp .$$

The *sum*, $M + N$, of two subsets $M, N \subset \mathcal{H}$ is

$$M + N = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in M, \mathbf{y} \in N\} .$$

If M, N are subspaces of $\mathcal{H}$ and $M \cap N = \{\mathbf{0}\}$, then $M + N$ is called the *direct sum* of M and N, and denoted by $M \oplus N$. If in addition $M \subset N^\perp$ we denote their sum by $M \overset{\perp}{\oplus} N$ and call it the *orthogonal direct sum* of M and N. Even if the subspaces M, N are closed, their sum $M + N$ need not be closed; see, e.g., Ex. 1. An orthogonal direct sum of two closed subspaces is closed. Conversely, if L, M are closed subspaces of $\mathcal{H}$ and $M \subset L$, then

$$L = M \overset{\perp}{\oplus} (L \cap M^\perp) . \tag{1}$$

If (1) holds for two subspaces $M \subset L$, we say that L is *decomposable with respect to* M. See Exs. 5–6.

(B) The *(Cartesian) product* of $\mathcal{H}_1, \mathcal{H}_2$ will be denoted by

$$\mathcal{H}_{1,2} = \mathcal{H}_1 \times \mathcal{H}_2 = \{\{\mathbf{x}, \mathbf{y}\} : \mathbf{x} \in \mathcal{H}_1, \mathbf{y} \in \mathcal{H}_2\}$$

where $\{\mathbf{x}, \mathbf{y}\}$ is an ordered pair. $\mathcal{H}_{1,2}$ is a Hilbert space with inner product

$$\langle \{\mathbf{x}_1, \mathbf{y}_1\}, \{\mathbf{x}_2, \mathbf{y}_2\} \rangle = \langle \mathbf{x}_1, \mathbf{y}_1 \rangle + \langle \mathbf{x}_2, \mathbf{y}_2 \rangle \ .$$

Let $J_i : \mathcal{H}_i \to \mathcal{H}_{1,2}$, $i = 1, 2$ be defined by

$$J_1 \mathbf{x} = \{\mathbf{x}, \mathbf{0}\} \quad \text{for all } \mathbf{x} \in \mathcal{H}_1$$

and

$$J_2 \mathbf{y} = \{\mathbf{0}, \mathbf{y}\} \quad \text{for all } \mathbf{y} \in \mathcal{H}_2 \ .$$

The transformations J_1 and J_2 are isometric isomorphisms, mapping $\mathcal{H}_1$ and $\mathcal{H}_2$ onto

$$\mathcal{H}_{1,0} = J_1 \mathcal{H}_1 = \mathcal{H}_1 \times \{\mathbf{0}\}$$

and

$$\mathcal{H}_{0,2} = J_2 \mathcal{H}_2 = \{\mathbf{0}\} \times \mathcal{H}_2 \ ,$$

respectively. Here $\{\mathbf{0}\}$ is an appropriate zero space.

(C) Let $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ denote the *class of linear operators from $\mathcal{H}_1$ to $\mathcal{H}_2$*. In what follows we will use *operator* to mean a linear operator. For any $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ we denote the *domain* of T by $D(T)$, the *domain* of T by $R(T)$, the *domain* of T by $N(T)$, and the *domain* of T by $C(T)$, where

$$C(T) = D(T) \cap N(T)^{\perp} \ . \tag{2}$$

The *graph*, $G(T)$, of a $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is

$$G(T) = \{\{\mathbf{x}, T\mathbf{x}\} : \mathbf{x} \in D(T)\} \ .$$

Clearly, $G(T)$ is a subspace of $\mathcal{H}_{1,2}$, and $G(T) \cap \mathcal{H}_{0,2} = \{\mathbf{0}, \mathbf{0}\}$. Conversely, if G is a subspace of $\mathcal{H}_{1,2}$ and $G(T) \cap \mathcal{H}_{0,2} = \{\mathbf{0}, \mathbf{0}\}$, then G is the graph of a unique $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, defined for any point $\mathbf{x}$ in its domain

$$D(T) = J_1^{-1} P_{\mathcal{H}_{1,0}} G(T)$$

by

$$T\mathbf{x} = \mathbf{y} \ ,$$

where $\mathbf{y}$ is the unique vector in $\mathcal{H}_2$ such that $\{\mathbf{x}, \mathbf{y}\} \in G$, and $P_{\mathcal{H}_{1,0}}$ is the orthogonal projector: $\mathcal{H}_{1,2} \to \mathcal{A}_{1,0}$, see (L) below.

Similarly, for any $T \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ the of T, $G^{-1}(T)$, is defined by

$$G^{-1}(T) = \{\{T\mathbf{y}, \mathbf{y}\} : \mathbf{y} \in D(T)\} \ .$$

A subspace G in $\mathcal{H}_{1,2}$ is an inverse graph of some $T \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ if and only if $G \cap \mathcal{H}_{1,0} = \{\mathbf{0}, \mathbf{0}\}$, in which case T is uniquely determined by G (von Neumann [**1507**]).

(D) An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is called *closed* if $G(T)$ is a closed subspace of $\mathcal{H}_{1,2}$. Equivalently, T is closed if

$$\mathbf{x}_n \in D(T),\ \mathbf{x}_n \to \mathbf{x}_0,\ T\mathbf{x}_n \to \mathbf{y}_0 \implies \mathbf{x}_0 \in D(T) \text{ and } T\mathbf{x}_0 = \mathbf{y}_0$$

where $\to$ denotes strong convergence. A closed operator has a closed null space. The subclass of *closed operators* in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is denoted by $\mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$.

(E) An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is called *bounded* if its norm $\|T\|$ is finite, where

$$\|T\| = \sup_{\mathbf{0} \neq \mathbf{x} \in \mathcal{H}_1} \frac{\|T\mathbf{x}\|}{\|\mathbf{x}\|} \ .$$

The subclass of *bounded operators* in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is denoted by $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. If $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, then it may be assumed, without loss of generality, that $D(T)$ is closed or even that $D(T) = \mathcal{H}_1$. A bounded $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is closed if and only if $D(T)$ is closed. Thus we may write $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2) \subset \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$.

Conversely, a closed $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ is bounded if $D(T) = \mathcal{H}_1$. This statement is the *closed graph theorem.*

(F) Let $T_1, T_2 \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ with $D(T_1) \subset D(T_2)$. If $T_2\mathbf{x} = T_1\mathbf{x}$ for all $\mathbf{x} \in D(T_1)$, then T_2 is called an *extension* of T_1 and T_1 is called a *restriction* of T_2. These relations are denoted by

$$T_1 \subset T_2$$

or by

$$T_1 = (T_2)_{[D(T_1)]} \ .$$

Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and let the restriction of T to $C(T)$ be denoted by T_0

$$T_0 = T_{[C(T)]} \ .$$

Then

$$G(T_0) = \{\{\mathbf{x}, T\mathbf{x}\} : \mathbf{x} \in C(T)\}$$

satisfies

$$G(T_0) \cap \mathcal{H}_{1,0} = \{\mathbf{0}, \mathbf{0}\}$$

and hence is the inverse graph of an operator $S \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ with

$$D(S) = R(T_0) \ .$$

Clearly,

$$ST\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \in C(T) \ ,$$

and

$$TS\mathbf{y} = \mathbf{y} \quad \text{for all } \mathbf{y} \in R(T_0) \ .$$

Thus, if T_0 is considered as an operator in $\mathcal{L}(\overline{C(T)}, \overline{R(T_0)})$, then T_0 is invertible in its domain. The inverse T_0^{-1} is closed if and only if T_0 is closed. For $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, both $C(T)$ aand T_0 nay be trivial; see, e.g., Exs. 2 and 4.

(G) An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is called *dense* (or *densely defined*) if $\overline{D(T)} = \mathcal{H}_1$. Since any $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ can be considered to be an element of $T \in \mathcal{L}(\overline{D(T)}, \mathcal{H}_2)$, any operator can be assumed to be dense without loss of generality.

For any $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, the condition $\overline{D(T)} = \mathcal{H}_1$ is equivalent to

$$G(T)^{\perp} \cap \mathcal{H}_{1,0} = \{\mathbf{0}, \mathbf{0}\} \ ,$$

where

$$G(T)^{\perp} = \{\{\mathbf{y}, \mathbf{z}\} : \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{z}, T\mathbf{x} \rangle = 0 \text{ for all } \mathbf{x} \in D(T)\} \subset \mathcal{H}_{1,2} \ .$$

Thus for any dense $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, $G(T)^{\perp}$ is the inverse graph of a unique operator in $\mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$. This operator is $-T^*$, where T^*, the *adjoint* of T, satisfies

$$\langle T^*\mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{y}, T\mathbf{x} \rangle \quad \text{for all } \mathbf{x} \in D(T) \ .$$

(H) For any dense $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$,

$$\overline{N(T)} = R(T^*)^{\perp} \ , \quad N(T^*) = R(T)^{\perp} \ . \tag{3}$$

In particular, $T\,[T^*]$ has a dense range if and only if $T^*\,[T]$ is one–to–one.

(I) Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ be dense.

If both T and T^* have inverses, then $(T^{-1})^* = (T^*)^{-1}$.

T has a bounded inverse if and only if $R(T^*) = \mathcal{H}_1$.

T^* has a bounded inverse if $R(T) = \mathcal{H}_2$. The converse holds if T is closed.

T^* has a bounded inverse and $R(T^*) = \mathcal{H}_1$ if and only if T has a bounded inverse and $\overline{R(T)} = \mathcal{H}_1$ (Taylor [**1436**], Goldberg [**543**]).

(J) An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is called *closable* (or *preclosed*) if T has a closed extension. Equivalently, T is closable if

$$\overline{G(T)} \cap \mathcal{H}_{0,2} = \{\mathbf{0}, \mathbf{0}\} ,$$

in which case $\overline{G(T)}$ is the graph of an operator $\overline{T}$, called the *closure* of T. $\overline{T}$ is the minimal closed extension of T.

Since $G(T)^{\perp\perp} = \overline{G(T)}$ it follows that for a dense T, T^{**} is defined only if T is closable, in which case

$$T \subset T^{**} = \overline{T}$$

and

$$T = T^{**}$$

if and only if T is closed.

(K) A dense operator $T \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ is called *symmetric* if

$$T \subset T^*$$

and *self–adjoint* if

$$T = T^* ,$$

in which case it is called *non–negative*, and denoted by $T \geq O$, if

$$\langle T\mathbf{x}, \mathbf{x}\rangle \geq 0 \quad \text{for all } \mathbf{x} \in D(T) .$$

If $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ is dense, then T^*T and TT^* are non–negative, and $I + TT^*$ and $I + T^*T$ have bounded inverses (von Neumann [**1504**]).

(L) An operator $T \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ is an *orthogonal projector* if

$$P = P^* = P^2 ,$$

in which case $R(P)$ is closed and

$$\mathcal{H} = R(P) \overset{\perp}{\oplus} N(P) .$$

Conversely, if L is a closed subspace of $\mathcal{H}$, then there is a unique orthogonal projector P_L such that

$$L = R(P_L) \quad \text{and} \quad L^\perp = N(P_L) .$$

(M) An operator $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ is called *normally solvable* if $R(T)$ is closed, which, by (3), is equivalent to the following condition: The equation

$$T\mathbf{x} = \mathbf{y}$$

is consistent if and only if $\mathbf{y}$ is orthogonal to any solution $\mathbf{u}$ of

$$T^*\mathbf{u} = \mathbf{0} .$$

This condition accounts for the name "normally solvable".

For any $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$, the following statements are equivalent:

(a) T us normally solvable.

(b) The restriction $T_0 = T_{[C(T)]}$ has a bounded inverse.

(c) The non–negative number

$$\gamma(T) = \inf \left\{ \frac{\|T\mathbf{x}\|}{\|\mathbf{x}\|} : \mathbf{0} \neq \mathbf{x} \in C(T) \right\} \tag{4}$$

is positive (Hestenes [**725**, Theorem 3.3]).

2.1. Exercises and examples.

Ex. 1. *A nonclosed sum of closed subspaces.* Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, and let

$$D = J_1 D(T) = \{\{\mathbf{x}, \mathbf{0}\} : \mathbf{x} \in D(T)\} .$$

Without loss of generality we assume that $D(T)$ is closed. Then D is closed. Also $G(T)$ is closed since T is bounded. But

$$G(T) + D$$

is nonclosed if $R(T)$ is nonclosed, since

$$\{\mathbf{x}, \mathbf{y}\} \in G(T) + D \iff \mathbf{y} \in R(T) \quad \text{(Halmos [\textbf{646}, p. 26]).}$$

Ex. 2. *Unbounded linear functionals.* Let T be an unbounded linear functional on $\mathcal{H}$. Then $N(T)$ is dense in $\mathcal{H}$, and consequently $N(T)^\perp = \{\mathbf{0}\}$, $C(T) = \{\mathbf{0}\}$.

An example of such a functional on $L^2[0, \infty]$ is

$$Tx = \int_0^\infty tx(t)dt .$$

To show that $N(T)$ is dense, let $x_0 \in L^2[0, \infty]$ with $Tx_0 = \alpha$. Then a sequance $\{x_n\} \subset N(T)$ converging to x_0 is

$$x_n(t) = \begin{cases} x_0(t) & \text{if } t < 1 \text{ or } t > n+1 \\ x_0(t) - \frac{\alpha}{nt} & \text{if } 1 \le t \le n+1 \end{cases}$$

Indeed,

$$\|x_n - x_0\|^2 = \int_1^{n+1} \frac{\alpha^2}{(nt)^2}\, dt = \frac{\alpha^2}{n(n+1)} \to 0 .$$

Ex. 3. Let D be a dense subspace of $\mathcal{H}$, and let F be a closed subspace such that $F^\perp$ is finite dimensional. Then

$$\overline{D \cap F} = F \quad \text{(Erdelyi and Ben–Israel [\textbf{477}, Lemma 5.1]).}$$

Ex. 4. *An operator with trivial carrier.* Let D be any proper dense subspace of $\mathcal{H}$ and choose $\mathbf{x} \notin D$. Let $F = [x]^\perp$, where $[\mathbf{x}]$ is the line generated by $\mathbf{x}$. Then $\overline{D \cap F} = F$, by Ex. 3. However, $D \not\subset F$, so we can choose a subspace $A \neq \{\mathbf{0}\}$ in D such that

$$D = A \oplus (D \cap F) .$$

Define $T \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ by

$$D(T) = D$$

and

$$T(\mathbf{y} + \mathbf{z}) = \mathbf{y} \quad \text{if } \mathbf{y} \in A,\ \mathbf{z} \in D \cap F .$$

Then

$$\begin{aligned} N(T) &= D \cap F , \\ \overline{N(T)} &= \overline{D \cap F} = F , \\ N(T)^\perp &= F^\perp = [\mathbf{x}] , \\ C(T) &= D(T) \cap N(T)^\perp = \{\mathbf{0}\} . \end{aligned}$$

EX. 5. Let L, M be subspaces of $\mathcal{H}$ and let $M \subset L$. Then

$$L = M \overset{\perp}{\oplus} (L \cap M^{\perp}) \tag{1}$$

if and only if

$$P_{\overline{M}}\mathbf{x} \in M \quad \text{for all } \mathbf{x} \in L\,.$$

In particular, a space is decomposable with respect to any closed subspace (Arghiriade [**40**]).

EX. 6. Let L, M, N be subspaces of $\mathcal{H}$ such that

$$L = M \overset{\perp}{\oplus} N\,.$$

Then

$$M = L \cap N^{\perp}\,, \quad N = L \cap M^{\perp}\,.$$

Thus an orthogonal direct sum is decomposable with respect to each summand.

EX. 7. *A bounded operator with nonclosed range.* Let ℓ^2 denote the Hilbert space of square summable sequences and let $T \in \mathcal{B}(\ell^2, \ell^2)$ be defined, for some $0 < k < 1$, by

$$T(\alpha_0,\, \alpha_1, \alpha_2,\, \dots\,,\, \alpha_n,\, \dots) = (\alpha_0,\, k\alpha_1,\, k^2\alpha_2,\, \dots\,,\, k^n\alpha_n,\, \dots)\,.$$

Consider the sequence

$$\mathbf{x}_n = \left(1,\, \frac{1}{2k},\, \frac{1}{3k^2},\, \dots\,,\, \frac{1}{nk^{n-1}},\, 0,\, 0,\, \dots\right)\,,$$

and the vector

$$\mathbf{y} = \lim_{n\to\infty} T\mathbf{x}_n = \left(1,\, \frac{1}{2},\, \frac{1}{3},\, \dots\,,\, \frac{1}{n},\, \dots\right)\,.$$

Then,

$$\mathbf{y} \in \overline{R(T)}\,,\ \mathbf{y} \notin R(T)\,.$$

EX. 8. *Linear integral operators.* Let $L^2 = L^2[a,b]$, the Lebesgue square integrable functions on the finite interval $[a,b]$. Let $K(s,t)$ be an L^2–*kernel* on $a \leq s, t, \leq b$, meaning that the Lebesgue integral

$$\int_a^b \int_a^b |K(s,t)|^2 ds dt$$

exists and is finite; see, e.g., Smithies [**1375**, Section 1.6].

Consider the two operators $T_1, T_2 \in \mathcal{B}(L^2, L^2)$ defined by

$$(T_1\mathbf{x})(s) = \int_a^b K(s,t)\mathbf{x}(t)dt\,, \quad a \leq s \leq b\,,$$

$$(T_2\mathbf{x})(s) = \mathbf{x}(s) - \int_a^b K(s,t)\mathbf{x}(t)dt\,, \quad a \leq s \leq b\,,$$

called *Fredholm integral operators of the first kind* and *the second kind*, respectively. Then

(a) $R(T_2)$ is closed.

(b) $R(T_1)$ is nonclosed unless it is finite dimensional.

More generally, if $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is completely continuous then $R(T)$ is nonclosed unless it is finite dimensional (Kammerer and Nashed [**815**, Proposition 2.5]).

EX. 9. Let $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$. Then T is normally solvable if and only if T^* is. Also, T is normally solvable if and only if RR^* or T^*T is.

3. Generalized inverses of linear operators between Hilbert spaces

A natural definition of generalized inverses in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is the following one due to Tseng [**1464**].

DEFINITION 1. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. Then an operator $T^q \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is a *Tseng generalized inverse* (abbreviated g.i.) of T if

$$R(T) \subset D(T^g) \tag{5}$$

$$R(T^g) \subset D(T) \tag{6}$$

$$T^g T\mathbf{x} = P_{\overline{R(T^g)}}\mathbf{x} \quad \text{for all } \mathbf{x} \in D(T) \tag{7}$$

$$TT^g\mathbf{y} = P_{\overline{R(T)}}\mathbf{y} \quad \text{for all } \mathbf{y} \in D(T^g)\ . \tag{8}$$

This definition is symmetric in T and T^g, thus T is a g.i. of T^g.

An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ may have a unique g.i., or infinitely many g.i.'s or it may have none. We will show in Theorem 1 that T is a g.i. if and only if its domain is decomposable with respect to its null space,

$$\begin{aligned} D(T) &= N(T) \overset{\perp}{\oplus} (D(T) \cap N(T)^{\perp}) \\ &= N(T) \overset{\perp}{\oplus} C(T)\ . \end{aligned} \tag{9}$$

By Ex. 5, this condition is satisfied if $N(T)$ is closed. Thus it holds for all closed operators, and in particular for bounded operators. If T has g.i.'s, then it has a maximal g.i., some of whose properties are collected in Theorem 2. For bounded operators with closed range, the maximal g.i. coincides with the Moore–Penrose inverse, and will likewise be denoted by $T^{\dagger}$. See Theorem 3.

For operators T without g.i.'s, the maximal g.i. $T^{\dagger}$ can be "approximated" in several ways, with the objective of retaining as many of its useful properties as possible. One such approach, due to Erdélyi [**475**] is described in Definition 3 and Theorem 4.

Some properties of g.i.'s, when they exist, are given in the following three lemmas, due to Arghiriade [**40**], which are needed later.

LEMMA 1. If $T^g \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is a g.i. of $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, then $D(T)$ is decomposable with respect to $R(T^g)$.

PROOF. Follows from Ex. 5 since, for any $\mathbf{x} \in D(T)$

$$P_{\overline{R(T^g)}}\mathbf{x} = T^g T\mathbf{x}\ , \quad \text{by (7)}\ .$$

□

LEMMA 2. If $T^g \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is a g.i. of $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, then T is a one–to–one mapping of $R(T^g)$ onto $R(T)$.

PROOF. Let $\mathbf{y} \in R(T)$. Then

$$\mathbf{y} = P_{\overline{R(T)}}\mathbf{y} = TT^g\mathbf{y}\ , \quad \text{by (8)}\ ,$$

proving that $T(R(T^g)) = R(T)$.

Now we prove that T is one–to–one on $R(T^g)$. Let $\mathbf{x}_1, \mathbf{x}_2 \in R(T^g)$ satisfy

$$T\mathbf{x}_1 = T\mathbf{x}_2\ .$$

Then

$$\mathbf{x}_1 = P_{\overline{R(T^g)}}\mathbf{x}_1 = T^g T\mathbf{x}_1 = T^g T\mathbf{x}_2 = P_{\overline{R(T^g)}}\mathbf{x}_2 = \mathbf{x}_2\ .$$

□

Lemma 3. If $T^g \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is a g.i. of $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, then:

$$N(T) = D(T) \cap R(T^g)^\perp \tag{10}$$

and

$$C(T) = R(T^g)\ . \tag{11}$$

Proof. Let $\mathbf{x} \in D(T)$. Then, by Lemma 1,

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2\ , \quad \mathbf{x}_1 \in R(T^g)\ , \quad \mathbf{x}_2 \in D(T) \cap R(T^g)^\perp\ , \quad \mathbf{x}_1 \perp \mathbf{x}_2\ . \tag{12}$$

Now

$$\mathbf{x}_1 = P_{\overline{R(T^g)}}\mathbf{x} = T^g T(\mathbf{x}_1 + \mathbf{x}_2) = T^g T \mathbf{x}_1$$

and therefore

$$T^g T \mathbf{x}_2 = \mathbf{0}\ ,$$

which, by Lemma 2 with T and T^g interchanged, implies that

$$T\mathbf{x}_2 = \mathbf{0}\ , \tag{13}$$

hence

$$D(T) \cap R(T^g)^\perp \subset N(T)\ .$$

Conversely, let $\mathbf{x} \in N(T)$ be decomposed as in (12). Then

$$\begin{aligned} \mathbf{0} = T\mathbf{x} &= T(\mathbf{x}_1 + \mathbf{x}_2) \\ &= T\mathbf{x}_1\ , \quad \text{by (13)}\ , \end{aligned}$$

which, by Lemma 2, implies that $\mathbf{x}_1 = \mathbf{0}$ and therefore

$$N(T) \subset D(T) \cap R(T^g)^\perp\ ,$$

completing the proof of (10).

Now

$$\begin{aligned} D(T) &= R(T^g) \overset{\perp}{\oplus} (D(T) \cap R(T^g)^\perp)\ , \quad \text{by Lemma 1}\ , \\ &= R(T^g) \overset{\perp}{\oplus} N(T)\ , \end{aligned}$$

which, by Ex. 6, implies that

$$R(T^g) = D(T) \cap N(T)^\perp\ ,$$

proving (11). □

The existence of g.i.'s is settled in the following theorem announced, without proof, by Tseng [**1464**]. Our proof follows that of Arghiriade [**40**].

Theorem 1. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. Then T has a g.i. if and only if

$$D(T) = N(T) \overset{\perp}{\oplus} C(T)\ , \tag{9}$$

in which case, for any subspace $L \subset R(T)^\perp$, there is a g.i. T^g_L of T, with

$$D(T^g_L) = R(T) \overset{\perp}{\oplus} L \tag{14}$$

and

$$N(T^g_L) = L\ . \tag{15}$$

PROOF. If T has a g.i., then (9) follows from Lemmas 1 and 3.

Conversely, suppose that (9) holds. Then

$$R(T) = T(D(T)) = T(C(T)) = R(T_0) \ , \tag{16}$$

where $T_0 = T_{[C(T)]}$ is the restriction of T to $C(T)$. The inverse T_0^{-1} exists, by Section 2(F), and satisfies

$$R(T_0^{-1}) = C(T)$$

and, by (16,)

$$D(T_0^{-1}) = R(T) \ .$$

For any subspace $L \subset R(T)^\perp$, consider the extension T_L^g of T_0^{-1} with domain

$$D(T_L^g) = R(T) \stackrel{\perp}{\oplus} L \tag{14}$$

and null space

$$N(T_L^g) = L \ . \tag{15}$$

From its definition, it follows that T_L^g satisfies

$$D(T_L^g) \supset R(T)$$

and

$$R(T_L^g)R(T_0^{-1}) = C(T) \subset D(T) \ . \tag{17}$$

For any $\mathbf{x} \in D(T)$

$$\begin{aligned} T_L^g T\mathbf{x} &= T_L^g T P_{\overline{C(T)}}\mathbf{x} \ , \quad \text{by (9)} \\ &= T_0^{-1} T_0 P_{\overline{C(T)}}\mathbf{x} \ , \quad \text{by Ex. 5} \\ &= P_{\overline{R(T_L^g)}}\mathbf{x} \ , \quad \text{by (17)} \ . \end{aligned}$$

Finally, any $\mathbf{y} \in D(T_L^g)$ can be written, by (14), as

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 \ , \quad \mathbf{y}_1 \in R(T) \ , \quad \mathbf{y}_2 \in L \ , \quad \mathbf{y}_1 \perp \mathbf{y}_2 \ ,$$

and therefore

$$\begin{aligned} TT_L^g\mathbf{y} &= TT_L^g\mathbf{y}_1 \ , \quad \text{by (15)} \\ &= T_0 T_0^{-1}\mathbf{y}_1 \\ &= \mathbf{y}_1 \\ &= P_{\overline{R(T)}}\mathbf{y} \ . \end{aligned}$$

Thus T_L^g is a g.i. of T. □

The g.i. T_L^g is uniquely determined by its domain (14) and null space (15); see Ex. 10.

The maximal choice of the subspace L in (14) and (15) is $L = R(T)^\perp$. For this choice we have the following

DEFINITION 2. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ satisfy (9). Then the *maximal g.i.* of T, denoted by $T^\dagger$, is the g.i. of T with domain

$$D(T^\dagger) = R(T) \stackrel{\perp}{\oplus} R(T)^\perp \tag{18}$$

and null space

$$N(T^\dagger) = R(T)^\perp \ . \tag{19}$$

By Ex. 10, the g.i. $T^\dagger$ so defined is unique. It is maximal in the sense that any other g.i. of T is a restriction of $T^\dagger$.

Moreover, $T^\dagger$ is dense, by (18), and has a closed null space, by (19). Choosing L as a dense subspace of $R(T)^\perp$ shows that an operator T may have infinitely many dense g.i.'s T^g_L. Also, T may have infinitely many g.i.'s T^g_L with closed null space, each obtained by choosing L as a closed subspace of $R(T)^\perp$. However, $T^\dagger$ is the unique dense g.i. with closed null space; see Ex. 11.

For closed operators, the maximal g.i. can be alternatively defined, by means of the following construction due to Hestenes [**725**], see also Landesman [**908**].

Let $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$ be dense. Since $N(T)$ is closed, it follows, from Ex. 5, that

$$D(T) = N(T) \overset{\perp}{\oplus} C(T) \,, \tag{9}$$

and therefore

$$G(T) = N \overset{\perp}{\oplus} C \,, \tag{20}$$

where, using the notation of Section 2(B), (C), and (F),

$$N = J_1 N(T) = G(T) \cap \mathcal{H}_{1,0} \tag{21}$$

$$C = \{\{\mathbf{x}, T\mathbf{x}\} : \mathbf{x} \in C(T)\} \,, \tag{22}$$

Similarly, since T^* is closed, it follows from Section 2(G), that

$$G(T)^\perp = N^* \overset{\perp}{\oplus} C^* \tag{23}$$

with

$$N^* = J_2 N(T^*) = G(T)^\perp \cap \mathcal{H}_{0,2} \,, \tag{24}$$

$$C^* = \{\{-T^*\mathbf{y}, \mathbf{y}\} : \mathbf{y} \in C(T^*)\} \,. \tag{25}$$

Now

$$\begin{aligned} \mathcal{H}_{1,2} &= G(T) \overset{\perp}{\oplus} G(T)^\perp \,, \quad \text{since } T \text{ is closed} \\ &= (N \overset{\perp}{\oplus} C) \overset{\perp}{\oplus} (N^* \overset{\perp}{\oplus} C^*) \,, \quad \text{by (20) and (23)} \\ &= (C \overset{\perp}{\oplus} N^*) \overset{\perp}{\oplus} (C^* \overset{\perp}{\oplus} N) \\ &= G^\dagger \overset{\perp}{\oplus} G^{\dagger *} \,, \end{aligned} \tag{26}$$

where

$$G^\dagger = C \overset{\perp}{\oplus} N^* \,, \tag{27}$$

$$G^{\dagger *} = C^* \overset{\perp}{\oplus} N \,. \tag{28}$$

Since

$$G^\dagger \cap \mathcal{H}_{1,0} = \{\mathbf{0}, \mathbf{0}\} \,, \quad \text{by Section 2(F)} \,,$$

it follows that $G^\dagger$ is the inverse graph of an operator $T^\dagger \in \mathcal{C}(\mathcal{H}_2, \mathcal{H}_1)$, with domain

$$\begin{aligned} J_2^{-1} P_{\mathcal{H}_{0,2}} G^\dagger &= T(C(T)) \overset{\perp}{\oplus} N(T^*) \\ &= R(T) \overset{\perp}{\oplus} R(T)^\perp \,, \quad \text{by (16) and (3),} \end{aligned}$$

and null space

$$J_2^{-1}M^* = N(T^*) = R(T)^\perp$$

and such that

$$T^\dagger T\mathbf{x} = P_{\overline{C(T)}}\,\mathbf{x}\ , \quad \text{for any } \mathbf{x} \in N(T) \overset{\perp}{\oplus} C(T)\ ,$$

and

$$TT^\dagger\mathbf{y} = P_{\overline{R(T)}}\,\mathbf{y}\ , \quad \text{for any } \mathbf{y} \in R(T) \overset{\perp}{\oplus} R(T)^\perp\ .$$

Thus $T^\dagger$ is the maximal g.i. of Definition 2.

Similarly, $G^{\dagger *}$ is the graph of the operator $-T^{*\dagger} \in \mathcal{C}(\mathcal{H}_1,\mathcal{H}_2)$, which is the maximal g.i. of $-T^*$.

This elegant construction makes obvious the properties of the maximal g.i., collected in the following:

Theorem 2. (Hestenes [**725**]). Let $T \in \mathcal{C}(\mathcal{H}_1,\mathcal{H}_2)$ be dense. Then

(a) $T^\dagger \in \mathcal{C}(\mathcal{H}_2,\mathcal{H}_1)$,

(b) $D(T^\dagger) = R(T) \overset{\perp}{\oplus} N(T^*)$, $\quad N(T^\dagger) = N(T^*)$,

(c) $R(T^\dagger) = C(T)$,

(d) $T^\dagger T\mathbf{x} = P_{\overline{R(T^\dagger)}}\mathbf{x}$ for any $\mathbf{x} \in D(T)$,

(e) $TT^\dagger\mathbf{y} = P_{\overline{R(T)}}\mathbf{y}$ for any $\mathbf{y} \in D(T^\dagger)$,

(f) $T^{\dagger\dagger} = T$,

(g) $T^{*\dagger} = T^{\dagger *}$,

(h) $N(T^{*\dagger}) = N(T)$,

(i) T^*T and $T^\dagger T^{*\dagger}$ are non–negative and

$$(T^*T)^\dagger = T^\dagger T^{*\dagger}\ , \quad N(T^*T) = N(T)\ ,$$

(j) TT^* and $T^{*\dagger}T^\dagger$ are non–negative and

$$(TT^*)^\dagger = T^{*\dagger}T^\dagger\ , \quad N(TT^*) = N(T^*)\ .$$

□

For bounded operators with closed range, various characterizations of the maximal g.i. are collected in the following:

Theorem 3. (Petryshyn [**1183**]). If $T \in \mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)$ and $R(T)$ is closed, then $T^\dagger$ is characterized as the unique solution X of the following equivalent systems:

(a) $TXT = T\ ,\ XTX = X\ ,\ (TX)^* = TX\ ,\ (XT)^* = XT$,

(b) $TX = P_{R(T)}\ ,\ N(X^*) = N(T)$,

(c) $TX = P_{R(T)}\ ,\ XT = P_{R(T^*)}\ ,\ XTX = X$,

(d) $XTT^* = T^*\ ,\ XX^*T^* = X$,

(e) $XT\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in R(T^*)$,
$X\mathbf{y} = \mathbf{0}$ for all $\mathbf{y} \in N(T^*)$,

(f) $XT = P_{R(T^*)}\ ,\ N(X) = N(T^*)$,

(g) $TX = P_{R(T)}\ ,\ XT = P_{R(X)}$. □

The notation $T^\dagger$ is justified by Theorem 3(a), which lists the four *Penrose equations* (1.1)–(1.4).

If $T \in \mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)$ does not satisfy (9), then it has no g.i., by Theorem 1. In this case one can still approximate $T^\dagger$ by an operator that has some properties of $T^\dagger$, and reduces to it if $T^\dagger$ exists. Such an approach, due to Erdélyi [**475**], is described in the following

Definition 3. Let $T \in \mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)$ and let T_r be the restriction of T defined by

$$D(T_r) = N(T) \overset{\perp}{\oplus} C(T)\ ,\ N(T_r) = N(T)\ . \tag{29}$$

The (*Erdelyi*) *g.i.* of T is defined as $T_r^\dagger$, which exists since T_r satisfies (9).

The inverse graph of $T_r^\dagger$ is

$$G^{-1}(T_r) = \{\{\mathbf{x}, T\mathbf{x}+\mathbf{z}\} : \mathbf{x} \in C(T)\ ,\ \mathbf{z} \in (T(C(T)))^\perp\}\ , \tag{30}$$

from which the following properties of $T_r^\dagger$ can be easily deduced.

THEOREM 4. (Erdélyi [**475**]). Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and let its restriction T_r be defined by (29). Then

(a) $T_r^\dagger = T^\dagger$ if $T^\dagger$ exists,

(b) $D(T_r^\dagger) = T(C(T)) \overset{\perp}{\oplus} T(C(T))^\perp$,
and in general, $\overline{R(T)} \not\subset D(T_r^\dagger)$,

(c) $R(T_r^\dagger) = C(T)$, $\overline{R(T_r^\dagger)} = N(T)^\perp$,

(d) $T_r^\dagger T\mathbf{x} = P_{\overline{R(T_r^\dagger)}}\mathbf{x}$ for all $\mathbf{x} \in D(T_r)$,

(e) $TT_r^\dagger\mathbf{y} = P_{\overline{R(T)}}\mathbf{y}$ for all $\mathbf{y} \in D(T_r^\dagger)$,

(f) $D((T_r^\dagger)_r^\dagger) = \overline{N(T)} \overset{\perp}{\oplus} C(T)$,

(g) $R((T_r^\dagger)_r^\dagger) = T(C(T))$,

(h) $N((T_r^\dagger)_r^\dagger) = \overline{N(T)}$,

(i) $T \subset (T_r^\dagger)_r^\dagger$ if (9) holds ,

(j) $T = (T_r^\dagger)_r^\dagger$ if and only if $N(T)$ is closed ,

(k) $T_r^{\dagger *} \subset (T^*)_r^\dagger$ if T is dense and closable. □

See also Ex. 15.

Exercises and examples.

EX. 10. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ have g.i.'s and let L be a subspace of $R(T)^\perp$. Then the conditions

$$D(T_L^g) = R(T) \overset{\perp}{\oplus} L \tag{14}$$

$$N(T_L^g) = L \tag{15}$$

determine a unique g.i., which is thus equal to T_L^g as constructed in the proof of Theorem 1.

PROOF. Let T^g be a g.i. of T satisfying (14) and (15), and let $\mathbf{y} \in D(T^g)$ be written as

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2\ ,\quad \mathbf{y}_1 \in R(T)\ ,\quad \mathbf{y}_2 \in L\ .$$

Then

$$\begin{aligned} T^g\mathbf{y} &= T^g\mathbf{y}_1\ ,\quad \text{by (15)}\\ &= T^gT\mathbf{x}_1\ ,\quad \text{for some } \mathbf{x}_1 \in D(T)\\ &= P_{\overline{R(T^g)}}\mathbf{x}_1\ ,\quad \text{by (7)}\\ &= P_{\overline{C(T)}}\mathbf{x}_1\ ,\quad \text{by (7)}\ . \end{aligned}$$

We claim that this determines T^g uniquely. For, suppose there is an $\mathbf{x}_2 \in D(T)$ with $\mathbf{y}_1 = T\mathbf{x}_2$. Then, as above,

$$T^g\mathbf{y} = P_{\overline{C(T)}}\mathbf{x}_2$$

and therefore

$$\begin{aligned} P_{\overline{C(T)}}\mathbf{x}_1 - P_{\overline{C(T)}}\mathbf{x}_2 &= P_{\overline{C(T)}}(\mathbf{x}_1 - \mathbf{x}_2)\\ &= \mathbf{0} \quad \text{since } \mathbf{x}_1 - \mathbf{x}_2 \in N(T)\ . \end{aligned}$$

□

EX. 11. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ have g.i.'s. Then $T^\dagger$ is the unique dense g.i. with closed null space.

PROOF. Let T^g be any dense g.i. with closed null space. Then

$$\begin{aligned} D(T^g) &= N(T^g) \overset{\perp}{\oplus} C(T^g) \ , \quad \text{by Theorem 1} \\ &= N(T^g \overset{\perp}{\oplus} R(T) \ , \quad \text{by (11)} \ , \end{aligned}$$

which, together with the assumptions $\overline{D(T^g)} = \mathcal{H}_2$ and $N(T^g) = \overline{N(T^g)}$, implies that

$$N(T^g) = R(T)^\perp \ .$$

Thus, T^g has the same domain and null space as $T^\dagger$, and therefore $T^g = T^\dagger$, by Ex. 10. □

EX. 12. Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ have a closed range $R(T)$ and let $T_1 \in \mathcal{B}(\mathcal{H}_1, R(T))$ be defined by

$$T_1\mathbf{x} = T\mathbf{x} \quad \text{for all } \mathbf{x} \in \mathcal{H}_1 \ .$$

Then
(a) T_1^* is the restriction of T^* to $R(T)$.
(b) The operator $T_1 T_1^* \in \mathcal{B}(R(T), R(T))$ is invertible.
(c) $T^\dagger = P_{R(T^*)} T_1^* (T_1 T_1^*)^{-1} P_{R(T)}$ (Kurepa [**896**]).

EX. 13. Let $T \in \mathcal{C}(\mathcal{H}_1, \mathcal{H}_2)$. Then $R(T)$ is closed if and only if $T^\dagger$ is bounded (Landesman [**908**]).

PROOF. Follows from Section 2(M). □

EX. 14. Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ have closed range. Then

$$T^\dagger = (T^*T)^\dagger T^* = T^*(TT^*)^\dagger \quad \text{(Desoer and Whalen [\textbf{396}])}.$$

EX. 15. For arbitrary $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ consider its extension $\widetilde{T}$ with

$$D(\widetilde{T}) = D(T) + \overline{N(T)} \ , \quad N(\widetilde{T}) = \overline{N(T)} \ , \quad \widetilde{T} = T \text{ on } D(T) \ , \tag{31}$$

which coincides with T if $N(T)$ is closed. Since $D(\widetilde{T})$ is decomposable with respect to $N(\widetilde{T})$, it might seem that $\widetilde{T}$ can be used to obtain $\widetilde{T}^\dagger$, a substitute for (possibly nonexisting) $T^\dagger$.

Show that $\widetilde{T}$ is not well defined by (31) if

$$D(T) \cap \overline{N(T)} \neq N(T) \quad \text{and} \quad N(\widetilde{T}) \neq D(\widetilde{T}) \ , \tag{32}$$

which is the only case of interest since otherwise $D(T)$ is decomposable with respect to $N(T)$ or $\widetilde{T}$ is identically O in its domain.

PROOF. By (32) there exist $\mathbf{x}_0$ and $\mathbf{y}$ such that

$$\mathbf{x}_0 \in D(T) \cap \overline{N(T)} \ , \quad \mathbf{x}_0 \notin N(T)$$

and

$$\mathbf{y} \in D(T) \ , \quad \mathbf{y} \notin \overline{N(T)} \ .$$

Then

$$\widetilde{T}(\mathbf{x}_0 + \mathbf{y}) = \widetilde{T}\mathbf{y} \ , \quad \text{since } \mathbf{x}_0 \in N(\widetilde{T})$$

and on the other hand

$$\begin{aligned} \widetilde{T}(\mathbf{x}_0 + \mathbf{y}) &= T(\mathbf{x}_0 + \mathbf{y}) \ , \quad \text{since } \mathbf{x}_0, \mathbf{y} \in D(T) \\ &\neq T\mathbf{y} \ , \quad \text{since } \mathbf{x}_0 \notin N(T) \ . \end{aligned}$$

□

Ex. 16. Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ have closed range. Then

$$\|T^\dagger\| = \frac{1}{\gamma(T)} ,$$

where $\gamma(T)$ is defined in (4) (Petryshyn [**1183**, Lemma 2]).

Ex. 17. Let $F \in \mathcal{B}(\mathcal{H}_3, \mathcal{H}_2)$ and $G \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_3)$ with $R(G) = \mathcal{H}_3 = R(F^*)$, and define $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ by $A = FG$, Then

$$\begin{aligned} A^\dagger &= G^*(GG^*)^{-1}(F^*F)^{-1}F^* \\ &= G^\dagger F^\dagger \quad \text{(Holmes [\textbf{743}, p. 223])}. \end{aligned}$$

Compare with Theorem 1.5 and Ex. 1.15.

Ex. 18. *Generalized inverses of linear integral operators.* In this exercise and in Exs. 19–25 below we consider the *Fredholm integral equation of the second kind*

$$x(s) - \lambda \int_a^b K(s,t)x(t)dt = y(s) , \quad a \leq s \leq b , \tag{33}$$

written for short as

$$(I - \lambda K)\mathbf{x} = \mathbf{y} ,$$

where all functions are complex, $[a,b]$ is a bounded interval, λ is a complex scalar and $K(s,t)$ is a L^2–*Kernel* on $[a,b] \times [a,b]$; see Ex. 8. Writing L^2 for $L^2[a,b]$, we need the following facts from the Fredholm thoery of integral equations; see, e.g., Smithies [**1375**]. For any λ, K as above

(a) $(I - \lambda K) \in \mathcal{B}(L^2, l^2)$,

(b) $(I - \lambda K)^* = I - \overline{\lambda} K^*$, where $K^*(s,t) = \overline{K(t,s)}$.

(c) The null spaces $N(I - \lambda K)$ and $N(I - \overline{\lambda} K^*)$ have equal finite dimensions,

$$\dim N(I - \lambda K) = \dim N(I - \overline{\lambda} K^*) = n(\lambda) , \text{ say} . \tag{34}$$

(d) A scalar λ is called a *regular value* of K if $n(\lambda) = 0$, in which case the operator $I - \lambda K$ has an inverse $(I - \lambda K)^{-1} \in \mathcal{B}(L^2, L^2)$ written as

$$(I - \lambda K)^{-1} = I + \lambda R , \tag{35}$$

where $R = R(s,t;\lambda)$ is an L^2–kernel called the resolvent of K.

(e) A scalar λ is called an *eigenvalue* of K if $n(\lambda) > 0$, in which case any nonzero $\mathbf{x} \in N(I - \lambda K)$ is called an *eigenfunction* of K corresponding to λ. For any λ and, in particular, for any eigenvalue λ, both range spaces $R(I - \lambda K)$ and $R(I - \overline{\lambda} K^*)$ are closed and, by (3),

$$R(I - \lambda K) = N(I - \overline{\lambda} K^*)^\perp , \quad R(I - \overline{\lambda} K^*) = N(I - \lambda K)^\perp . \tag{36}$$

Thus, if λ is a regular value of K then (33) has, for any $\mathbf{y} \in L^2$, a unique solution given by

$$\mathbf{x} = (I + \lambda R)\mathbf{y} ,$$

that is

$$x(s) = y(s) + \lambda \int_a^b R(s,t,\lambda)y(t)dt , \ a \leq s \leq b . \tag{37}$$

If λ is an eigenvalue of K then (33) is consistent if and only if $\mathbf{y}$ is orthogonal to every $\mathbf{u} \in N(I - \overline{\lambda} K^*)$, in which case the general solution of (33) is

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{n(\lambda)} c_i \mathbf{x}_i , \quad c_i \text{ arbitrary scalars} , \tag{38}$$

where $\mathbf{x}_0$ is a particular solution of (33) and $\{\mathbf{x}_1, \ldots , \mathbf{x}_{n(\lambda)}\}$ is a basis of $N(I - \lambda K)$.

EX. 19. *Pseudo resolvents.* Let λ be an eigenvalue of K. Following Hurwitz [**761**], an L^2–kernel $R = R(s,t,\lambda)$ is called a *pseudo resolvent* of K if for any $\mathbf{y} \in R(I - \lambda K)$, the function

$$x(s) = y(s) + \lambda \int_a^b R(s,t,\lambda) y(t) dt \tag{37}$$

is a solution of (33).

A pseudo resolvent was constructed by Hurwitz as follows.

Let λ_0 be an eigenvalue of K, and let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be orthonormal bases of $N(I - \lambda_0 K)$ and $N(I - \overline{\lambda_0} K^*)$ respectively. Then λ_0 is a regular value of the kernel

$$K_0(s,t) = K(s,t) - \frac{1}{\lambda_0} \sum_{i=1}^n u_i(s) \overline{x_i(t)} \ , \tag{39}$$

written for short as

$$K_0 = K - \frac{1}{\lambda_0} \sum_{i=1}^n \mathbf{u}_i \mathbf{x}_i^*$$

and the resolvent R_0 of K_0 is a pseudo resolvent of K, satisfying

$$\begin{aligned} (I + \lambda_0 R_0)(I - \lambda_0 K)\mathbf{x} &= \mathbf{x} \ , \quad \text{for all } \mathbf{x} \in R(I - \overline{\lambda_0} K^*) \\ (I - \lambda_0 K)(I + \lambda_0 R_0)\mathbf{y} &= \mathbf{y} \ , \quad \text{for all } \mathbf{y} \in R(I - \lambda_0 K) \\ (I + \lambda_0 R_0)\mathbf{u}_i &= \mathbf{x}_i \ , \quad i = 1, \dots, n \ . \end{aligned} \tag{40}$$

PROOF. Follows from the matrix case, Ex. 2.40. □

EX. 20. A comparison with Theorem 2.2 shows that $I + \lambda R$ is a $\{1\}$–inverse of $I - \lambda K$, if R is a pseudo resolvent of K. As with $\{1\}$–inverses, the pseudo resolvent is nonunique. Indeed, for $R_0, \mathbf{u}_i, \mathbf{x}_i$ as above, the kernel

$$R_0 + \sum_{i,j,=1}^n c_{ij} \mathbf{x}_i \mathbf{u}_j^* \tag{41}$$

is a pseudo resolvent of K for any choice of scalars c_{ij}.

The pseudo resolvent constructed by Fredholm [**515**], who called the resulting operator $I + \lambda R$ a *pseudo inverse* of $I - \lambda K$, is the first explicit application, known to us, of a generalized inverse.

The class of all pseudo resolvents of a given kernel K is characterized as follows.

Let K be an L^2–kernel, let λ_0 be an eigenvalue of K and let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be orthonormal bases of $N(I - \lambda_0 K)$ and $N(I - \overline{\lambda_0} K^*)$ respectively. An L^2–kernel R is a pseudo resolvent of K if and only if

$$R = K + \lambda_0 K R - \frac{1}{\lambda_0} \sum_{i=1}^n \boldsymbol{\beta}_i \mathbf{u}_i^* \ , \tag{42a}$$

$$R = K + \lambda_0 R K - \frac{1}{\lambda_0} \sum_{i=1}^n \mathbf{x}_i \boldsymbol{\alpha}_i \ , \tag{42b}$$

where $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i \in L^2$ satisfy

$$\langle \boldsymbol{\alpha}_i, \mathbf{x}_j \rangle = \delta_{ij} \ , \ \langle \boldsymbol{\beta}_i, \mathbf{u}_j \rangle = \delta_{ij} \ , \quad i,j = 1, \dots, n \ . \tag{43}$$

Here KT stands for the kernel $KR(s,t) = \int_a^b K(s,u) R(u,t) du$, etc.

If λ is a regular value of K then (42) reduces to

$$R = K + \lambda K R \ , \ R = K + \lambda R K \ , \tag{44}$$

which uniquely determines the resolvent $R(s,t,\lambda)$ (Hurwitz [**761**]).

Ex. 21. Let K, λ_0, $\mathbf{x}_i$, $\mathbf{u}_i$, and R_0 be as above. Then the maximal g.i. of $I - \lambda_0 K$ is

$$(I - \lambda_0 K)^\dagger = I + \lambda_0 R_0 - \sum_{i=1}^{n} \mathbf{x}_i \mathbf{u}_i^* \ , \tag{45}$$

corresponding to the pseudo resolvent

$$R = R_0 - \frac{1}{\lambda_0} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{u}_i^* \ . \tag{46}$$

Ex. 22. Let $K(s,t) = u(s)\overline{v(t)}$, where

$$\int_a^b u(s)\overline{v(t)} = 0 \ .$$

Then every scalar λ is a regular value of K.

Ex. 23. Consider the equation

$$x(s) - \lambda \int_{-1}^{1} (1 + 3st) x(t) dt = y(s) \tag{47}$$

with $K(s,t) = 1 + 3st$. The resolvent is

$$R(s,t;\lambda) = \frac{1 + 3st}{1 - 2\lambda} \ .$$

K has a single eigenvalue $\lambda = \frac{1}{2}$ and an orthonormal basis of $N(I - \frac{1}{2}K)$ is

$$\left\{ x_1(s) = \frac{1}{\sqrt{2}},\ x_2(s) = \frac{\sqrt{3}}{\sqrt{2}} \right\}$$

which, by symmetry, is also an orthonormal basis of $N(I - \frac{1}{2}K^*)$. From (39) we get

$$\begin{aligned} K_0(s,t) &= K(s,t) - \frac{1}{\lambda_0} \sum u_i(s)\,\overline{x_i(t)} \\ &= (1 + 3st) - 2\left(\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} + \frac{\sqrt{3}}{\sqrt{2}}s\frac{\sqrt{3}}{\sqrt{2}}t \right) \\ &= 0 \ , \end{aligned}$$

and the resolvent of $K_0(s,t)$ is therefore

$$R_0(s,t,;\lambda) = 0 \ .$$

If $\lambda \neq \frac{1}{2}$, then for each $y \in L^2[-1,1]$ equation (47) has a unique solution

$$x(s) = y(s) + \lambda \int_{-1}^{1} \frac{1 + 3st}{1 - 2\lambda} y(t) dt \ .$$

If $\lambda = \frac{1}{2}$, then (47) is consistent if and only if

$$\int_{-1}^{1} y(t) dt = 0 \ , \quad \int_{-1}^{1} t y(t) dt = 0 \ ,$$

in which case the general solution is

$$x(s) = y(s) + c_1 + c_2 s \ , \quad c_1, c_2 \text{ arbitrary} \ .$$

EX. 24. Let

$$K(s,t) = 1 + s + 3st\ , \quad -1 \leq s,t \leq 1\ .$$

Then $\lambda = \frac{1}{2}$ is the only eigenvalue and

$$\dim\ N(I - \frac{1}{2}K) = 1\ .$$

An orthonormal basis of $N(I - \frac{1}{2}K)$ is the single vector

$$x_1(s) = \frac{\sqrt{3}}{\sqrt{2}} s\ , \quad -1 \leq s \leq 1\ .$$

An orthonormal basis of $N(I - \frac{1}{2}K^*)$ is

$$u_1(s) = \frac{1}{\sqrt{2}}\ , \quad -1 \leq s \leq 1\ .$$

The Hurwitz kernel (39) is

$$\begin{aligned} K_0(s,t) &= (1 + s + 3st) - 2\left(\frac{1}{\sqrt{2}}\frac{\sqrt{3}}{\sqrt{2}}t\right) \\ &= 1 + s - \sqrt{3}t + 3st\ , \quad -1 \leq s,t \leq 1\ . \end{aligned}$$

Compute the resolvent R_0 of K_0, which is a pseudo resolvent of K. (*Hint*: Use the following exercise).

EX. 25. *Degenerate kernels.* A kernel $K(s,t)$ is called *degenerate* if it is a finite sum of products of L^2 functions, as follows:

$$K(s,t) = \sum_{i=1}^{m} f_i(s)\,\overline{g_i(t)}\ . \tag{48}$$

Degenerate kernels are convenient because they reduce the integral equation (33) to a finite system of linear equations. Also, any L^2–kernel can be approximated, arbitrarily close, by a degenerate kernel; see, e.g., Smithies [**1375**, p. 40], and Halmos [**646**, Problem 137].

Let $K(s,t)$ be given by (48). Then

(a) The scalar λ is an eigenvalue of (48) if and only if $1/\lambda$ is an eigenvalue of the $m \times m$ matrix

$$B = [b_{ij}]\ , \quad \text{where } b_{ij} = \int_a^b f_j(s)\,\overline{g_i(s)}ds\ .$$

(b) Any eigenfunction of K $[K^*]$ corresponding to an eigenvalue λ $[\overline{\lambda}]$ is a linear combination of the m functions $f_1, \dots, f_m$ $[g_1, \dots, g_m]$.

(c) If λ is a regular value of (48), then the resolvent at λ is

$$R(s,t,;\lambda) = \frac{\det \begin{bmatrix} 0 & \vdots & f_1(s) & \cdots & f_m(s) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\overline{g_1(t)} & \vdots & & & \\ \vdots & \vdots & & I - \lambda B & \\ -\overline{g_m(t)} & \vdots & & & \end{bmatrix}}{\det(I - \lambda B)}\ .$$

See also Kantorovich and Krylov [**818**, Chapter II].

Generalized inverses of linear differential operators.
The following 12 exercises deal with generalized inverses of closed dense operators $L \in \mathcal{C}(\mathcal{S}_1, \mathcal{S}_2)$ with $\overline{D(L)} = \mathcal{S}_1$, where:

(i) $\mathcal{S}_1, \mathcal{S}_2$ are spaces of (scalar or vector) functions which are either the Hilbert space $L^2[a,b]$ or the *space of continuous functions* $C[a,b]$, where $[a,b]$is a given finite real interval. Since $C[a,b]$ is a dense subspace of $L^2[a,b]$, a closed dense linear operator mapping $C[a,b]$ into $\mathcal{S}_2$ may be considered as a dense operator in $\mathcal{C}(L^2[a,b], \mathcal{S}_2)$.

(ii) L is defined for all $\mathbf{x}$ in its domain $D(L)$ by

$$L\mathbf{x} = \ell\mathbf{x} \ , \tag{49}$$

where ℓ is a *differential expression*, for example, in the vector case

$$\ell\mathbf{x} = A_1(t)\frac{d}{dt}\mathbf{x} + A_0(t)\mathbf{x} \ , \tag{50}$$

where $A_0(t), A_1(t)$ are $n \times n$ matrix coefficients, with suitable regularity conditions; see, e.g., Ex. 31 below.

(iii) The domain of L consists of those functions in $\mathcal{S}_1$ for which ℓ makes sense and $\ell\mathbf{x} \in \mathcal{S}_2$, and which satisfy certain conditions, such as *initial* or *boundary conditions.*

If a differential operator L is invertible and there is a kernel (function, or matrix in the vector case)

$$G(s,t) \ , \quad a \le s,t \le b \ ,$$

such that for all $\mathbf{y} \in R(L)$

$$(L^{-1}\mathbf{y})(s) = \int_a^b G(s,t)y(t)dt \ , \quad a \le s \le b \ ,$$

then $G(s,t)$ is called the *Green's function* (or *matrix*) of L. In this case, for any $\mathbf{y} \in R(L)$, the unique solution of

$$L\mathbf{x} = \mathbf{y} \tag{51}$$

is given by

$$\mathbf{x}(s) = \int_a^b G(s,t)\mathbf{y}(t)dt \ , \quad a \le s \le b \ . \tag{52}$$

If L is not invertible, but there is a kernel $G(s,t)$ such that, for any $\mathbf{y} \in R(L)$, a particular solution of (51) is given by (52), then $G(s,t)$ is called a *generalized Green's function* (or *matrix*) of L. A generalized Green's function of L is therefore a kernel of an integral operator which is a generalized inverse of L.

Generalized Green's functions were introduced by Hilbert [**734**] in 1904, and consequently studied by Myller, Westfall and Bounitzky [**212**], Elliott ([**457**], [**458**]), and Reid [**1260**]; see, e.g., the historical survey in [**1263**].

EX. 26. *Derivatives.* Let

$\mathcal{S}$ = the real space $L^2[0,\pi]$ of real valued functions,
$\mathcal{S}^1$ = the absolutely continuous functions $x(t)$, $0 \le t \le \pi$, whose derivatives x' are in $\mathcal{S}$,
$\mathcal{S}^2 = \{x \in \mathcal{S}^1 : x' \in \mathcal{S}^1\}$,

and let L be the differential operator d/dt with

$$D(L) = \{x \in \mathcal{S}^1 : x(0) = x(\pi) = 0\} \ .$$

Then

(a) $L \in \mathcal{C}(\mathcal{S},\mathcal{S})$, $\overline{D(L)} = \mathcal{S}$, $C(L) = D(L)$,

$$R(L) = \left\{ y \in \mathcal{S} : \int_0^\pi y(t)dt = 0 \right\} = \overline{R(L)} \ .$$

(b) The adjoint L^* is the operator $-d/dt$ with

$$D(L^*) = \mathcal{S}^1 \ , \quad c(L^*) = \mathcal{S}^1 \cap R(L) \ , \quad R(L^*) = \mathcal{S} \ .$$

(c) $L^*L = -\dfrac{d^2}{dt^2}$ with $D(L^*L) = \{x \in \mathcal{S}^2 : x(0) = x(\pi) = 0\}$ and $R(L^*L) = \mathcal{S}$.

(d) $LL* = -\dfrac{d^2}{dt^2}$ with $D(LL^*) = \{x \in \mathcal{S}^2 : x'(0) = x'(\pi) = 0\}$ and $R(LL^*) = R(L)$.

(e) $L^\dagger$ is defined on $D(L^\dagger = \mathcal{S}$ by

$$(L^\dagger y)(t) = \int_0^t y(s)ds - \frac{t}{\pi}\int_0^\pi y(s)ds \ , \quad 0 \le t \le \pi$$

(Hestenes [**725**, Example 1]).

Ex. 27. For L of Ex. 26, determine which of the following equations hold and interpret your results:

(a) $L^{\dagger *} = L^{*\dagger}$,

(b) $L^\dagger = (L^*L)^\dagger L^* = L^*(LL^*)^\dagger$,

(c) $L^{\dagger\dagger} = L$.

Ex. 28. *Gradients.* Let

$\mathcal{S} =$ the real space $L^2([a,b] \times [a,b])$ of real valued functions $x(t_1,t_2)$, $0 \le t_1, t_2 \le \pi$.

$\mathcal{S}^1 =$ the subclass of $\mathcal{S}$ with the properties

(i) $x(t_1,t_2)$ is absolutely continuous in $t_1[t_2]$ for almost all $t_2[t_1]$, $0 \le t_1, t_2 \le \pi$;

(ii) the partial derivatives $\partial x/\partial t_1$, $\partial x/\partial t_2$ which exist almost everywhere are in $\mathcal{S}$,

and let L be the gradient operator

$$\ell x = \begin{bmatrix} \dfrac{\partial x}{\partial t_1} \\ \dfrac{\partial x}{\partial t_2} \end{bmatrix}$$

with domain

$$D(L) = \left\{ x \in \mathcal{S}^1 : \begin{cases} x(0,t_2) = x(\pi,t_2) = 0 \text{ for almost all } t_2 \ , \\ x(t_1,0) = x(t_1,\pi) = 0 \text{ for almost all } t_1 \ , \end{cases} \ 0 \le t_1, t_2 \le \pi \right\}$$

Then:

(a) $L \in \mathcal{C}(\mathcal{S}, \mathcal{S} \times \mathcal{S})$, $\overline{D(L)} = \mathcal{S}$.

(b) The adjoint L^* is the negative of the divergence operator

$$\ell^* \mathbf{y} = \ell^* \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = -\frac{\partial y_1}{\partial t_1} - \frac{\partial y_2}{\partial t_2}$$

with

$$D(L^*) = \{\mathbf{y} \in \mathcal{S} \times \mathcal{S} : \mathbf{y} \in C^1\} \ .$$

(c) L^*L is the negative of the Laplacian operator

$$L^*L = -\left[\frac{\partial^2}{\partial t_1^2} + \frac{\partial^2}{\partial t_2^2} \right]$$

(d) The Green's function of L^*L is

$$G(s_1,s_2,t_1,t_2) = \frac{4}{\pi^2}\sum_{m,n=1}^{\infty} \frac{1}{m^2+n^2} \sin(ms_1)\sin(ns_2)\sin(mt_1)\sin(nt_2)\ ,$$
$$0 \le s_i, t_j \le \pi\ .$$

(e) If

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathcal{S}\times\mathcal{S}\ ,$$

then

$$(L^\dagger y)(t_1,t_2) = \sum_{j=1}^{2}\int_0^\pi \int_9^\pi \frac{\partial}{\partial s_j} G(s_1,s_2,t_1,t_2)\, y_j(s_1,s_2)\, ds_1\, ds_2$$

(Landesman [**908**, Section 5]).

EX. 29. *Ordinary linear differential equations with homogeneous boundary conditions.* Let
$\mathcal{S}$ = the real space $L^2[a,b]$ of real valued functions,
$C^k[a,b]$ = the real valued functions on $[a,b]$ with k derivatives and

$$x^{(k)} = \frac{d^k x}{dt^k} \in C[a,b]\ ,$$

$\mathcal{S}^k = \{x \in C^{k-1}[a,b] :\ x^{k-1}$ absolutely continuous , $x^{(k)} \in \mathcal{S}\}$
and let L be the operator

$$\ell = \sum_{i=1}^{n} a_i(t)\left(\frac{d}{dt}\right)^i\ ,\ a_i \in C^i[a,b]\ ,\ i = 0,1,\dots,n\ , \qquad (53)$$
$$a_n(t) \ne 0\ ,\ a \le t \le b\ ,$$

with domain $D(L)$ consisting of all $x \in \mathcal{S}^n$ which satisfy

$$M\widehat{x} = 0\ , \qquad (54)$$

where $M \in \mathbb{R}_m^{m\times 2n}$ is a matrix with a specific null space $N(M)$, and $\widehat{x} \in \mathbb{R}^{2n}$ is the boundary vector

$$\widehat{x}^T = [x(a), x'(a), \cdots, x^{(n-1)}(a); x(b), x'(b), \cdots, x^{(n-1)}(b)]\ .$$

Finally let $\widetilde{L}$ be the operator ℓ of (53) with $D(\widetilde{L}) = \mathcal{S}^n$. Then

(a) $L \in \mathcal{C}(\mathcal{S},\mathcal{S})\ ,\quad \overline{D(L)} = \mathcal{S}\ .$
(b) $\dim N(\widetilde{L}) = n = \dim N(\widetilde{L}^*)\ .$
(c) $N(L) \subset N(\widetilde{L})\ ,\ N(L^*) \subset N(\widetilde{L}^*)$, hence $\dim N(L) \le n$ and $\dim N(L^*) \le n$.
(d) $R(L)$ is closed.
(e) The restriction $L_0 = L_{[C(L)]}$ of L to its carrier is a one–to–one mapping of $C(L)$ onto $R(L)$;

$$L_0 \in \mathcal{C}(C(L), R(L))\ .$$

(f) $L_0^{-1} \in \mathcal{B}(R(L), C(L))\ .$
(g) $L^\dagger$, the extension of L_0^{-1} to all of $\mathcal{S}$ with $N(L^\dagger) = R(L)^\perp$ is bounded and satisfies

$$LL^\dagger y = P_{R(L)}y\ ,\quad \text{for all } y \in \mathcal{S}$$
$$L^\dagger L x = P_{N(L)^\perp}x\ ,\quad \text{for all } y \in D(L)\ .$$

For proofs of (a) and (d) see Halperin [**649**] and Schwartz [**1325**]. The proof of (e) is contained in Section 2(F), and (f) follows from the closed graph theorem (Locker [**968**]).

Ex. 30. For L as in Ex. 29, find the generalized Green's function which corresponds to $L^\dagger$, i.e., find the kernel $L^\dagger(s,t)$ such that

$$(L^\dagger y)(s) = \int_a^b L^\dagger(s,t)\, y(t)\, dt \quad \text{for all } y \in D(L^\dagger) = \mathcal{S} \, .$$

SOLUTION. A generalized Green's function of $\widetilde{L}$ is (see Coddington and Levinson [**360**, Theorem 6.4])

$$\widetilde{G}(s,t) = \begin{cases} \sum_{j=1}^{n} \dfrac{x_j(s)\det(X_j(t))}{a_n(t)\det(X(t))} \, , & a \le t \le s \le b \\ 0 \, , & a \le s \le t \le b \end{cases} \tag{55}$$

where

$\{x_1, \dots, x_k\}$ is an orthonormal basis of $N(L)$,

$\{x_1, \dots, x_k, x_{k+1}, \dots, x_n\}$ is an orthonormal basis of $N(\widetilde{L})$,

$$X(t) = \left[x_j^{(i-1)}(t) \right] \, , \quad i, j = 1, \dots, n \, ,$$

$X_j(t)$ is the matrix obtained from $X(t)$ by replacing the j th column by $[0, 0, \dots, 0, 1]^T$.

Since $R(L) \subset R(\widetilde{L})$ it follows, for any $y \in R(L)$, that the general solution of

$$Lx = y$$

is

$$x(s) = \int_a^b \widetilde{G}(s,t)\, y(t)\, dt + \sum_{i=1}^{n} c_i\, x_i(s) \, , \tag{56}$$

$$c_i \text{ arbitrary} \, .$$

Writing the particular solution $L^\dagger y$ in the form (56)

$$L^\dagger y = x_0 + \sum_{i=1}^{n} c_i\, x_i \, , \tag{57}$$

$$x_0(s) = \int_a^b \widetilde{G}(s,t)\, y(t)\, dt \, ,$$

we determine its coefficients $\{c_1, \dots, c_n\}$ as follows:

(a) The coefficients $\{c_1, \dots, c_k\}$ are determined by $L^\dagger y \in N(L)^\perp$, since, by (57),

$$\langle L^\dagger y, x_j \rangle = 0 \implies c_j = -\langle x_0, x_j \rangle \, , \quad j = 1, \dots, k \, .$$

(b) The remaining coefficients $\{c_{k+1}, \dots, c_n\}$ are determined by the boundary condition (54). Indeed, writing (57) as

$$L^\dagger y = x_0 + Xc \, , \quad c^T = [c_1, \dots, c_n] \, ,$$

it follows from (54) that

$$M\widehat{x}_0 + M\widehat{X}c = 0 \, , \quad \text{where } \widehat{X} = \begin{bmatrix} X(a) \\ X(b) \end{bmatrix} \, . \tag{58}$$

A solution of (58) is

$$c = -(M\widehat{X})^{(1)} M\widehat{x}_0 \, , \tag{59}$$

where $(M\widehat{X})^{(1)} \in \mathbb{R}^{n\times m}$ is any $\{1\}$–inverse of $M\widehat{X} \in \mathbb{R}^{m\times n}$. Now $\{x_1, \dots, x_k\} \subset D(L)$, and therefore

$$M\widehat{X} = \begin{bmatrix} O & B \end{bmatrix} \, , \quad B \in \mathbb{R}^{m\times(n-k)}_{n-k} \, .$$

Thus, we may use in (59),

$$(M\widehat{X})^{(1)} = \begin{bmatrix} O \\ B^{(1)} \end{bmatrix} , \quad \text{for any } B^{(1)} \in B\{1\} ,$$

obtaining

$$c = -\begin{bmatrix} O \\ B^{(1)} \end{bmatrix} M\widehat{x}_0 ,$$

which uniquely determines $\{c_{k+1}, \dots , c_n\}$.

Substituting these coefficients $\{c_1, \dots , c_n\}$ in (56) finally gives $L^\dagger(s,t)$ (Locker [**968**]).

Ex. 31. *The vector case.* Let $\mathcal{S}_n$ and $\mathcal{S}_n^k$ denote the spaces of n–dimensional vector functions whose components belong to $\mathcal{S}$ and $\mathcal{S}^k$, respectively, of Ex. 29. Let L be the differential operator

$$\ell\mathbf{x} = A_1(t)\frac{d\mathbf{x}}{dt} + A_0(t)\mathbf{x} , \quad a \le t \le b \tag{50}$$

where A_0, A_1 are $n \times n$ matrix functions satisfying[1]

(i) $A_0(t)$ is continuous on $[a, b]$.

(ii) $A_1(t)$ is continuously differentiable and nonsingular on $[a, b]$,

with domain $D(L)$ consisting of those vector functions $\mathbf{x} \in \mathcal{S}_n^1$ which satisfy

$$M\widehat{\mathbf{x}} = \mathbf{0} , \tag{54}$$

where $M \in \mathbb{R}_m^{m\times 2n}$ is a matrix with a specified null space $N(M)$ and $\widehat{\mathbf{x}} \in \mathbb{R}^{2n}$ is the boundary vector

$$\widehat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}(a) \\ \mathbf{x}(b) \end{bmatrix} . \tag{60}$$

Let $\widetilde{L}$ be the differential operator (50) with domain $D(\widetilde{L}) = \mathcal{S}_n^1$. Then

(a) $L \in \mathcal{C}(\mathcal{S}_n, \mathcal{S}_n)$, $\overline{D(L)} = \mathcal{S}_n$.

(b) The adjoint of L is the operator L^* defined by

$$\ell^*\mathbf{y} = -\frac{d}{dt}(A_1^*(t)\mathbf{y}) + A_0^*(t)\mathbf{y} \tag{61}$$

on its domain

$$\begin{aligned} D(L^*) &= \{\mathbf{y} \in \mathcal{S}_n^1 : \mathbf{y}^*(b)\mathbf{x}(b) - \mathbf{y}^*(a)\mathbf{x}(a) = 0 \text{ for all } \mathbf{x} \in D(L)\} \\ &= \left\{\mathbf{y} \in \mathcal{S}_n^1 : P^* \begin{bmatrix} I & O \\ O & -I \end{bmatrix} \widehat{\mathbf{y}} = \mathbf{0} \text{ for any } P \in \mathbb{R}_{2n-m}^{(2n-m)\times 2n} \text{ with} MP = O\right\} \end{aligned} \tag{62}$$

(c) $\dim N(\widetilde{L}) = n$.

(d) Let

$$k = \dim N(L) \quad \text{and} \quad k^* = \dim N(L^*) .$$

Then

$$\max\{0, n-m\} \le k \le \min\{n, 2n-m\}$$

and

$$k + m = k^* + n .$$

(e) $R(L) = N(L^*)^\perp$, $R(L^*) = N(L)^\perp$,

hence both $R(L)$ and $R(L^*)$ are closed.

(f) Let

$$X(t) = [\mathbf{x}_1(t), \dots , \mathbf{x}_n(t)]$$

[1] Much weaker regularity conditions will do; see, e.g., Reid [**1262**] and [**1264**, Chapter III].

be a *fundamental matrix* of $\widetilde{L}$, i.e., let the vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ form a basis of $N(\widetilde{L})$. Then

$$\widetilde{G}(s,t) = \frac{1}{2}\text{sign}(s-t)X(s)X(t)^{-1} \tag{63}$$

is a generalized Green's matrix of $\widetilde{L}$.

(g) Let $(M\widehat{X})^{(1)}$ be any $\{1\}$–inverse of $M\widehat{X}$ where $\widehat{X} = \begin{bmatrix} X(a) \\ X(b) \end{bmatrix}$. Then

$$G(s,t) = \frac{1}{2}X(s)\left(\text{sign}(s-t)I - (M\widehat{X})^{(1)}M\begin{bmatrix} I & O \\ O & -I \end{bmatrix}\widehat{X}\right)X(t)^{-1} \tag{64}$$

is a generalized Green's matrix of L (Reid [**1263**] and [**1264**, Chapter III]).

Proof of (g). For any $\mathbf{y} \in R(L)$, the general solution of

$$L\mathbf{x} = \mathbf{y} \tag{51}$$

is

$$\mathbf{x}(s) = \int_a^b \widetilde{G}(s,t)\mathbf{y}(t)dt + \sum_{i=1}^n c_i\mathbf{x}_i(s) \tag{56}$$

or

$$\mathbf{x} = \mathbf{x}_0 + X\mathbf{c}\ , \quad \mathbf{c}^T = [c_1, \ldots, c_n]$$

and from (54) it follows that

$$\mathbf{c} = -(M\widehat{X})^{(1)}M\widehat{\mathbf{x}}_0 \tag{59}$$

and (64) follows by substituting (59) in (56). □

Ex. 32. The differential expression

$$\ell\, x = \sum_{i=1}^n a_i(t)\,\frac{d^i x}{dt^i}\ , \quad x \text{ scalar function} \tag{53}$$

is a special case of

$$\ell\mathbf{x} = A_1(t)\frac{d\mathbf{x}}{dt} + A_0(t)\mathbf{x}\ , \quad x \text{ vector function}\ . \tag{50}$$

Ex. 33. *The class of all generalized Green's functions.* Let L be as in Ex. 31 and let $X_0(t)$ and $Y_0(t)$ be $n \times k$ and $n \times k^*$ matrix functions whose columns are bases of $N(L)$ and $N(L^*)$, respectively. Then a kernel $H(s,t)$ is a generalized Green's matrix of L if and only if

$$H(s,t) = G(s,t) + X_0(s)A^*(t) + B(s)Y_0^*(t)\ , \tag{65}$$

where $G(s,t)$ is any generalized Green's matrix of L (in particular (64)), and $A(t)$ and $B(s)$ are $n \times k$ and $n \times k^*$ matrix functions which are Lebesgue measurable and essentially bounded. (Reid [**1262**]).

Ex. 34. Let $X_0(t)$ and $Y_0(t)$ be a sin Ex. 33. If $\Theta(t)$ and $\Psi(t)$ are Lebesgue measurable and essentially bounded matrix functions such that the matrices

$$\int_a^b \Theta^*(t)\,X_0(t)\,dt\ , \quad \int_a^b Y_0^*(t)\,\Psi(t)\,dt$$

are nonsingular, then L has a unique generalized Green's function $G_{\Theta,\Psi}$ such that

$$\begin{matrix} \int_a^b \Theta^*(s)\,G(s,t) = O \\ \int_a^b G(s,t)\,\Psi(t)\,dt = O \end{matrix} \quad a \le s,t \le b\ . \tag{66}$$

Thus the generalized inverse determined by $G_{\Theta,\Psi}$ has null space spanned by the columns of Ψ and range which is the orthogonal complement of the columns of Θ. Compare with Section 2.5. (Reid [**1262**]).

Ex. 35. *Existence and properties of* $L^\dagger$. If in Ex. 34 we take

$$\Theta = X_0 \ , \quad \Psi = Y_0 \ ,$$

then we get a generalized inverse of L which has the same range and null space as L^*. This generalized inverse is the analog of the *Moore–Penrose inverse* of L and will likewise be denoted by $L^\dagger$.

Show that $L^\dagger$ satisfies the four *Penrose equations* (1.1)–(1.4) as far as can be expected.

(a) $LL^\dagger L = L$,
(b) $L^\dagger L L^\dagger = L^\dagger$,
(c) $LL^\dagger = P_{R(L)}$,
$(LL^\dagger)^* = P_{R(L)}$ on $D(L^*)$,
(d) $L^\dagger L = P_{R(L^*)}$ on $D(L)$,
$(L^\dagger L)^* = P_{R(L^*)}$ (Loud [**973**], [**974**]).

Ex. 36. *Loud's construction of* $L^\dagger$. Just as in the matrix case (see Theorem 2.10(c) and Ex. 2.29) it follows here that

$$L^\dagger = P_{R(L^*)}\, G\, P_{R(L)} \ , \tag{67}$$

where G is any generalized Green's matrix.

In computing $P_{R(L^*)}$ and $P_{R(L)}$ we use Ex. 31(e) to obtain

$$P_{R(L^*)} = I - P_{N(L)} \ , \quad P_{R(L)} = I - P_{N(L^*)} \ . \tag{68}$$

Here $P_{N(L)}$ and $P_{N(L^*)}$ are integral operators of the first kind with kernels

$$K_{N(L)} = X_0(s) \left(\int_a^b X_0^*(u) X_0(u) du \right)^{-1} X_0^*(t) \tag{69}$$

and

$$K_{N(L^*)} = Y_0(s) \left(\int_a^b Y_0^*(u) Y_0(u) du \right)^{-1} Y_0^*(t) \ , \tag{70}$$

respectively, where X_0 and Y_0 are as in Ex. 33.

Thus, for any generalized Green's matrix $G(s,t)$, $L^\dagger$ has the kernel

$$L^\dagger(s,t) = G(s,t) - \int_a^b K_{N(L)}(s,u)\, G(u,t)\, du - \int_a^b G(s,u)\, K_{N(L^*)}(u,t)\, du + \int_a^b \int_a^b K_{N(L)}(s,u)\, G(u,v)\, K_{N(L^*)}(v,t)\, du\, dv \tag{71}$$

(Loud [**974**]).

Ex. 37. Let L be the differential operator given by

$$\ell\, \mathbf{x} = \mathbf{x}' - B(t)\mathbf{x} \ , \quad 0 \le t \le 1$$

with boundary conditions

$$\mathbf{x}(0) = \mathbf{x}(1) = \mathbf{0} \ .$$

Then the adjoint L^* is given by

$$\ell^*\, \mathbf{y} = -\mathbf{y}' - B(t)^*\mathbf{y}$$

with no boundary conditions.

Let $X(t)$ be a fundamental matrix for

$$\ell\,\mathbf{x} = \mathbf{0}\ .$$

Then $X(t)^{*-1}$ is a fundamental matrix for

$$\ell^*\,\mathbf{y} = \mathbf{0}\ .$$

Now $N(L) = \{\mathbf{0}\}$ and therefore $K_{N(L)} = O$. Also, $N(L^*)$ is spanned by the columns of $X(t)^{*-1}$, so by (70)

$$K_{N(L^*)} = X(s)^{*-1}\left(\int_a^b X(u)X(u)^{*-1}du\right) X(t)^{-1}\ . \tag{72}$$

A generalized Green's matrix for L is

$$G(s,t) = \begin{cases} X(s)X(t)^{-1}\ , & 0 \le s < t \le 1 \\ O\ , & 0 \le t < s \le 1 \end{cases} \tag{73}$$

Finally, by (71),

$$L^\dagger(s,t) = G(s,t) - \int_0^1 G(s,u)\,K_{N(L^*)}(u,t)\,du\ ,$$

with G and $K_{N(L^*)}$ given by (73) and (72), respectively (Loud [**974**, pp. 201–202]).

4. Minimal properties of generalized inverses

In this section, which is based on Erdélyi and Ben–Israel [**477**], we develop certain distinguishing minimal properties of generalized inverses of operators between Hilbert spaces. The matrix case appears in Chapter 3.

Definition 4. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and consider the linear equation

$$T\mathbf{x} = \mathbf{y}\ . \tag{74}$$

If the infimum

$$\|T\mathbf{x}' - \mathbf{y}\| = \inf_{\mathbf{x}\in D(T)} \|T\mathbf{x} - \mathbf{y}\| \tag{75}$$

is attained by a vector $\mathbf{x}' \in D(T)$, then $\mathbf{x}'$ is called an *extremal solution* of (74). Among the extremal solutions there may exist a unique vector $\mathbf{x}_0$ of least norm

$$\|\mathbf{x}_0\| < \|\mathbf{x}'\|\ ,$$

for all extremal solutions $\mathbf{x}' \neq \mathbf{x}_0$. Then $\mathbf{x}_0$ is called the *least extremal solution*.

Other names for extremal solutions are *virtual solutions* (Tseng [**1467**]), and *approximate solutions*.

Example 38 shows that extremal solutions need not exist. their existence is characterized in the following theorem.

Theorem 5. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. Then

$$T\mathbf{x} = \mathbf{y} \tag{74}$$

has an extremal solution if and only if

$$P_{\overline{R(T)}}\mathbf{y} \in R(T)\ . \tag{76}$$

PROOF. For every $\mathbf{x} \in D(T)$

$$\begin{aligned}\|T\mathbf{x} - \mathbf{y}\|^2 &= \|P_{\overline{R(T)}}(T\mathbf{x} - \mathbf{y})\|^2 + \|P_{R(T)^\perp}(T\mathbf{x} - \mathbf{y})\|^2 \\ &= \|P_{\overline{R(T)}}(T\mathbf{x} - \mathbf{y})\|^2 + \|P_{R(T)^\perp}\mathbf{y}\|^2 .\end{aligned}$$

Thus

$$\|T\mathbf{x} - \mathbf{y}\| \geq \|P_{R(T)^\perp}\mathbf{y}\| , \quad \text{for all } \mathbf{x} \in D(T)$$

with equality if and only if

$$T\mathbf{x} = P_{\overline{R(T)}}\mathbf{y} . \tag{77}$$

Clearly,

$$\inf_{\mathbf{x} \in D(T)} \|T\mathbf{x} - \mathbf{y}\| = P_{\overline{R(T)}}\mathbf{y} , \tag{78}$$

which is attained if and only if (77) is satisfied for some $\mathbf{x} \in D(T)$. □

See also Ex. 45.

The existence of extremal solutions does not guarantee the existence of a least extremal solution; see, e.g., Ex. 40. Before settling this issue we require

LEMMA 4. Let $\mathbf{x}'$ and $\mathbf{x}''$ be extremal solutions of (74). Then
(a) $P_{N(T)^\perp}\mathbf{x}' = P_{N(T)^\perp}\mathbf{x}''$
(b) $P_{\overline{N(T)}}\mathbf{x}' \in N(T)$ if and only if $P_{\overline{N(T)}}\mathbf{x}'' \in N(T)$.

PROOF. (a) From (77),

$$T\mathbf{x}' = T\mathbf{x}'' = P_{\overline{R(T)}}\mathbf{y}$$

and hence

$$T(\mathbf{x}' - \mathbf{x}'') = \mathbf{0} , \tag{79}$$

proving (a).

(b) From (79),

$$\mathbf{x}' - \mathbf{x}'' = P_{\overline{N(T)}}(\mathbf{x}' - \mathbf{x}'')$$

and then

$$P_{\overline{N(T)}}\mathbf{x}' = P_{\overline{N(T)}}\mathbf{x}'' + (\mathbf{x}' - \mathbf{x}'') ,$$

proving (b). □

The existence of the least extremal solution is characterized in the following:

THEOREM 6. (Erdélyi and Ben–Israel [**477**]). Let $\mathbf{x}$ be an extremal solution of (74). There exists a least extremal solution if and only if

$$P_{\overline{N(T)}}\mathbf{x} \in N(T) \tag{80}$$

in which case, the least extremal solution is

$$\mathbf{x}_0 = P_{N(T)^\perp}\mathbf{x} . \tag{81}$$

PROOF. Let $\mathbf{x}'$ be an extremal solution of (74). Then

$$\begin{aligned}\|\mathbf{x}'\|^2 &= \|P_{\overline{N(T)}}\mathbf{x}'\|^2 + \|P_{N(T)^\perp}\mathbf{x}'\|^2 \\ &= \|P_{\overline{N(T)}}\mathbf{x}'\|^2 + \|P_{N(T)^\perp}\mathbf{x}\|^2 , \quad \text{by Lemma 4 ,}\end{aligned}$$

proving that

$$\|\mathbf{x}'\| \geq \|P_{N(T)^\perp}\mathbf{x}\|$$

with equality if and only if

$$P_{\overline{N(T)}}\mathbf{x}' = \mathbf{0} . \tag{82}$$

If. Let condition (80) be satisfied and define

$$\mathbf{x}_0 = \mathbf{x} - P_{\overline{N(T)}}\mathbf{x} .$$

Then $\mathbf{x}_0$ is an extremal solution since

$$T\mathbf{x}_0 = T\mathbf{x} .$$

Also

$$P_{\overline{N(T)}}\mathbf{x}_0 = \mathbf{0} ,$$

which, by (82), proves that $\mathbf{x}_0$ is the least extremal solution.

Only if. Let $\mathbf{x}_0$ be the least extremal solution of (74). Then, by (82),

$$\mathbf{x}_0 = P_{\overline{N(T)}}\mathbf{x}_0 + P_{N(T)^\perp}\mathbf{x}_0 = P_{N(T)^\perp}\mathbf{x} ,$$

and hence

$$\mathbf{x}_0 = \mathbf{x} - P_{\overline{N(T)}}\mathbf{x} .$$

But

$$T\mathbf{x}_0 = T\mathbf{x} ,$$

since both $\mathbf{x}_0$ and $\mathbf{x}$ are extremal solutions, and therefore

$$TP_{\overline{N(T)}} = \mathbf{0} ,$$

proving (80). □

As in the matrix case (see Corollary 3.3), here too a unique generalized inverse is characterized by the property that it gives the least extremal solution whenever it exists. We define this inverse as follows:

DEFINITION 5. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, let

$$C(T) = D(T) \cap N(T)^\perp , \tag{2}$$

$$B(T) = D(T) \cap \overline{N(T)} , \tag{83}$$

and let $A(T)$ be a subspace satisfying

$$D(T) = A(T) \oplus \left(B(T) \overset{\perp}{\oplus} C(T) \right) . \tag{84}$$

(Examples 43 and 44 below show that, in the general case, this complicated decomposition cannot be avoided.) Let

$$G_0 = \{\{\mathbf{x}, T\mathbf{x}\} : \mathbf{x} \in C(T)\} , \quad G_1 = G(T)^\perp \cap \mathcal{H}_{0,2} = J_2 R(T)^\perp .$$

The *extremal g.i.* of T, denoted by $T_e^\dagger$, is defined by its inverse graph

$$G_0 + G_1 = \{\{\mathbf{x}, T\mathbf{x} + \mathbf{z}\} : \mathbf{x} \in C(T),\, \mathbf{z} \in R(T)^\perp\}\ .$$

The following properties of $T_e^\dagger$ aare easy consequences of the above construction.

THEOREM 7. (Erdélyi and Ben–Israel [**477**]). Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. Then

(a) $D(T_e^\dagger) = T(C(T)) \overset{\perp}{\oplus} R(T)^\perp$, and in general, $R(T) \not\subset D(T_e^\dagger)$.
(b) $R(T_e^\dagger) = C(T)$.
(c) $N(T_e^\dagger) = R(T)^\perp$.
(d) $TT_e^\dagger \mathbf{y} = P_{\overline{R(T)}}\, \mathbf{y}$, for all $\mathbf{y} \in D(T_e^\dagger)$.
(e) $T_e^\dagger T\mathbf{x} = P_{\overline{R(T_e^\dagger)}}\, \mathbf{x}$, for all $\mathbf{x} \in N(T) \overset{\perp}{\oplus} C(T)$. □

See also Exs. 41–42 below.

The extremal g.i. $T_e^\dagger$ is characterized in terms of the least extremal solution, as follows:

THEOREM 8. (Erdélyi and Ben–Israel [**477**]). The least extremal solution $\mathbf{x}_0$ of (74) exists if and only if

$$\mathbf{y} \in D(T_e^\dagger)\ , \tag{85}$$

in which case

$$\mathbf{x}_0 = T_e^\dagger \mathbf{y}\ . \tag{86}$$

PROOF. Assume (85). By Theorem 7(a)

$$P_{\overline{R(T)}}\, \mathbf{y} = \mathbf{y}_0 \in T(C(T)) \subset R(T)\ ,$$

and, by Theorem 5, extremal solutions do exist. Let $\mathbf{x}_0$ be the unique vector in $C(T)$ such that

$$P_{\overline{R(T)}}\, \mathbf{y} = \mathbf{y}_0 = T\mathbf{x}_0\ .$$

Then, by Theorem 3(a), (c), and (e),

$$T_e^\dagger \mathbf{y} = T_e^\dagger \mathbf{y}_0 = T_e^\dagger T\mathbf{x}_0 = \mathbf{x}_0\ ,$$

and by Theorem 3(d),

$$\|T\mathbf{x}_0 - \mathbf{y}\| = \|TT_e^\dagger \mathbf{y} - \mathbf{y}\| = \|P_{\overline{R(T)}}\, \mathbf{y} - \mathbf{y}\| = \|P_{R(T)^\perp}\, \mathbf{y}\|\ ,$$

which, by (78), shows that $\mathbf{x}_0$ is an extremal solution. Since

$$\mathbf{x}_0 \in R(T_e^\dagger) \subset N(T)^\perp\ ,$$

it follows, from Lemma 4, that

$$\mathbf{x}_0 = P_{N(T)^\perp}\, \mathbf{x}$$

for any extremal solution $\mathbf{x}$ of (74). By Theorem 6, $\mathbf{x}_0$ is the least extremal solution.

Conversely, let $\mathbf{x}_0$ be the least extremal solution whose existence we assume. By Theorem 2, $\mathbf{x}_0 \in C(T)$, and by Theorem 3(e),

$$T_e^\dagger T\mathbf{x}_0 = \mathbf{x}_0\ .$$

Since $\mathbf{x}_0$ is an extremal solution, it follows from (77) that

$$T\mathbf{x}_0 = P_{\overline{R(T)}}\, \mathbf{y} \in T(C(T))$$

and therefore

$$\begin{aligned}\mathbf{x}_0 &= T_e^\dagger T\mathbf{x} = T_e^\dagger P_{\overline{R(T)}}\mathbf{y} \\ &= T_e^\dagger \mathbf{y} \ .\end{aligned}$$

□

If $N(T)$ is closed then $T_e^\dagger$ coincides with the maximal g.i. $T^\dagger$. Thus for closed operators, and in particular for bounded operators, $T_e^\dagger$ should be replaced by $T^\dagger$ in the statement of Theorem 8

4.1. Exercises and examples.

Ex. 38. *A linear equation without extremal solution.* Let T and $\mathbf{y}$ be as in Ex. 7. Then

$$T\mathbf{x} = \mathbf{y}$$

has no extremal solutions.

Ex. 39. it was noted in Ex. 8, that, in general, the Fredholm integral operator of the first kind has a nonclosed range. Consider the kernel

$$G(s,t) = \begin{cases} s(1-t)\ , & 0 \le s \le t \le 1 \\ t(1-s)\ , & 0 \le t \le s \le 1 \end{cases}$$

which is a generalized Green's function of the operator

$$-\frac{d^2}{dt^2}\ , \quad 0 \le t \le 1\ .$$

Let $T \in \mathcal{B}(L^2[0,1], L^2[0,1])$ be defined by

$$(Tx)(s) = \int_0^1 G(s,t)\,x(t)\,dt\ .$$

Show that there exists a $\mathbf{y} \in L^2[0,1]$ for which

$$Tx = y$$

has no extremal solution.

Ex. 40. *An equation without a least extremal solution.* Consider the unbounded functional on $L^2[0,\infty]$

$$Tx = \int_0^\infty tx(t)dt$$

discussed in Ex. 2. Then the equation

$$Tx = 1$$

is consistent, and each of the functions

$$x_n(t) = \begin{cases} \dfrac{1}{nt}\ , & 1 \le t \le n+1 \\ 0\ , & \text{otherwise} \end{cases}$$

is a solution, $n = 1, 2, \ldots$. Since

$$\|x_n\|^2 = \int_1^{n+1} \frac{1}{(nt)^2}\,dt = \frac{1}{n(n+1)} \to 0\ ,$$

there is no extremal solution of least norm.

Ex. 41. *Properties of* $(T_e^\dagger)^\dagger$. By Theorem 7(a) and (c), it follows that $D(T_e^\dagger)$ is decomposable with respect to $N(T_e^\dagger)$. Thus $T_e^\dagger$ has a maximal (Tseng) g.i., denoted by $T_e^{\dagger\dagger}$. Some of its properties are listed below.

(a) $G(T_e^{\dagger\dagger}) = \{\{\mathbf{x}+\mathbf{z}, T\mathbf{x}\} : \mathbf{x} \in C(T), \mathbf{z} \in C(T)^\perp\}$.

(b) $D(T_e^{\dagger\dagger}) = C(T) \overset{\perp}{\oplus} C(T)^\perp$.

(c) $R(T_e^{\dagger\dagger}) = T(C(T))$.

(d) $N(T_e^{\dagger\dagger}) = C(T)^\perp$.

Ex. 42. Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and let

$$D_0(T) = N(T) \overset{\perp}{\oplus} C(T) .$$

Then

(a) $D(T_e^{\dagger\dagger}) = C(T) \overset{\perp}{\oplus} \overline{N(T)} \overset{\perp}{\oplus} D_0(T)^\perp$, a refinement of Ex. 41(b).

(b) $D_0(T) \subset D(T) \cap D(T_e^{\dagger\dagger})$ and $T_{[D_0(T)]} = (T_e^{\dagger\dagger})_{[D_0(T)]}$.

(c) $T_e^{\dagger\dagger}$ is an extension of T if and only if $D(T)$ is decomposable with respect to $N(T)$, in which case $T_e^{\dagger\dagger}$ is an extension by zero to $\overline{N(T)} \overset{\perp}{\oplus} D(T)^\perp$.

Ex. 43. *An example of* $A(T) \neq \{\mathbf{0}\}$, $A(T) \subset D(T_e^{\dagger\dagger})$. Let T be the operator defined in Ex. 4. Then, by Ex. 4,

$$\begin{aligned} B(T) &= D(T) \cap \overline{N(T)} \\ &= D \cap (\overline{D \cap F}) \\ &= D \cap F \\ &= N(T) , \end{aligned}$$

and

$$C(T) = \{\mathbf{0}\} ,$$

showing that

$$A(T) \neq \{\mathbf{0}\} , \quad \text{by (84)} .$$

Thus

$$A(T) = A \quad \text{of Ex. 4,}$$

and

$$D(T_e^\dagger) = A^\perp = N(T_e^\dagger) .$$

Finally, from $C(T)^\perp = \mathcal{H}$,

$$D(T_e^{\dagger\dagger}) = \mathcal{H} \supset A$$

with

$$N(T_e^{\dagger\dagger}) = \mathcal{H} .$$

Ex. 44. *An example of* $A(T) \neq \{\mathbf{0}\}$, $A(T) \cap D(T_e^{\dagger\dagger}) = \{\mathbf{0}\}$. Let $\mathcal{H}$ be a Hilbert space and let M, N be subspaces of $\mathcal{H}$ such that

$$M \neq \overline{M} , \quad N \neq \overline{N} \subset M^\perp .$$

Choose

$$\mathbf{y} \in \overline{M} \setminus M \quad \text{and} \quad \mathbf{z} \in M^{\perp} \setminus (M^{\perp} \overset{\perp}{\oplus} (N^{\perp} \cap M^{\perp}) \; ;$$

let

$$\mathbf{x} = \mathbf{y} + \mathbf{z}$$

and

$$D = M \oplus N \oplus [\mathbf{x}]$$

where $[\mathbf{x}]$ is the line spanned by $\mathbf{x}$. Define $T \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ on $D(T) = D$ by

$$T(\mathbf{u} + \mathbf{v} + \alpha\mathbf{x}) = \mathbf{v} + \alpha\mathbf{x} \;, \quad \mathbf{u} \in M \;, \quad \mathbf{v} \in N \;, \quad \alpha\mathbf{x} \in [\mathbf{x}] \;.$$

Then

$$C(T) = N \;, \quad N(T) = M \;, \quad A(T) = [\mathbf{x}]$$

and

$$\mathbf{x} \notin D(T_e^{\dagger\dagger}) \;.$$

Ex. 45. Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. Then

$$T\mathbf{x} = \mathbf{y} \tag{74}$$

has an extremal solution if and only if there is a positive scalar β such that

$$|\langle \mathbf{y}, \mathbf{z} \rangle|^2 \leq \beta \langle \mathbf{z}, AA^*\mathbf{z} \rangle \;, \quad \text{for every } \mathbf{z} \in N(AA^*)^{\perp}$$

(Tseng [**1467**]; see also Holmes [**743**, Section 35]).

Ex. 46. Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, $S \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_3)$ be normally solvable, and let

$$T_S = T_{[N(S)]}$$

denote the restriction of T to $N(S)$. If T_S is also normally solvable , then $T_S^{\dagger}$ is called the $N(S)$–*restricted pseudoinverse of* T. It is the unique solution X of the following five equations

$$\begin{aligned} SX &= O \;, \\ XTX &= X \;, \\ (TX)^* &= TX \;, \\ TXT &= T \quad \text{on } N(S) \;, \\ P_{N(S)}(XT)^* &= XT \quad \text{on } N(S) \end{aligned} \qquad \text{(Minamide and Nakamura [\textbf{1055}]).}$$

Ex. 47. Let T, S, and $T_S^{\dagger}$ be as in Ex. 46. Then for any $\mathbf{y}_0 \in \mathcal{H}_2$ and $\mathbf{z}_0 \in R(S)$, the least extremal solution of

$$T\mathbf{x} = \mathbf{y}_0$$

subject to

$$S\mathbf{x} = \mathbf{z}_0$$

is given by

$$\mathbf{x}_0 = T_S^{\dagger}(\mathbf{y}_0 - TS^{\dagger}\mathbf{z}_0) + S^{\dagger}\mathbf{z}_0 \qquad \text{(Minamide and Nakamura [\textbf{1055}]).}$$

EX. 48. Let $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ be Hilbert spaces, let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ with $R(T) = \mathcal{H}_2$ and let $S \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_3)$. For any $\mathbf{y} \in \mathcal{H}_2$, there is a unique $\mathbf{x}_0 \in \mathcal{H}_1$ satisfying

$$T\mathbf{x} = \mathbf{y} \tag{74}$$

and which minimizes the functional

$$\|S\mathbf{x}\|^2 + \|\mathbf{x}\|^2$$

over all solutions of (74). This $\mathbf{x}_0$ is given by

$$\mathbf{x}_0 = (I + S^*S)^{-1}T^\dagger \mathbf{y}_0$$

where $\mathbf{y}_0$ is the unique vector in $\mathcal{H}_2$ satisfying

$$\mathbf{y} = T(I + S^*S)^{-1}T^\dagger \mathbf{y}_0 \qquad \text{(Porter and Williams [\textbf{1199}]).}$$

EX. 49. Let $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, T$, and S be as above. Then for any $\mathbf{y} \in \mathcal{H}_2, \mathbf{x}_1 \in \mathcal{H}_1$, and $\mathbf{y}_1 \in \mathcal{H}_2$ there is a unique $\mathbf{x}_0 \in \mathcal{H}_1$ which is a solution of

$$T\mathbf{x} = \mathbf{y} \tag{74}$$

and which minimizes

$$\|S\mathbf{x} - \mathbf{y}_1\|^2 + \|\mathbf{x} - \mathbf{x}_1\|^2$$

from among all solutions of (74). This $\mathbf{x}_0$ is given by

$$\mathbf{x}_0 = (I + S^*S)^{-1}(T^\dagger \mathbf{y}_0 + \mathbf{x}_0 + S^*\mathbf{y}_1)$$

where $\mathbf{y}_0$ is the unique vector in $\mathcal{H}_2$ satisfying

$$\mathbf{y} = T(I + S^*S)^{-1}(T^\dagger \mathbf{y}_0 + \mathbf{x}_1 + S^*\mathbf{y}_1) \qquad \text{(Porter and Williams [\textbf{1199}]).}$$

5. Series and integral representations and iterative computation of generalized inverses

Direct computational methods, in which the exact solution requires a finite number of steps (such as the elimination methods of Sections 7.2–7.4) cannot be used, in general, for the computation of generalized inverses of operators. The exceptions are operators with nice algebraic properties, such as the integral and differential operators of Exs. 18–37 with their finite–dimensional null spaces. In the general case, the only computable representations of generalized inverses involve infinite series, or integrals, approximated by suitable iterative methods. Such representations and methods are sampled in this section, based on Showalter and Ben–Israel [**1354**], where the proofs, omitted here, can be found.

To motivate the idea behind our development consider the problem of minimizing

$$f(\mathbf{x}) = \langle A\mathbf{x} - \mathbf{y}, A\mathbf{x} - \mathbf{y}\rangle \ , \tag{87}$$

where $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces.

Treating $\mathbf{x}$ as a function $\mathbf{x}(t)$, $t \geq 0$, with $\mathbf{x}(0) = \mathbf{0}$, we differentiate (87):

$$\begin{aligned} \frac{d}{dt} f(\mathbf{x}) &= 2\Re\langle A\mathbf{x} - \mathbf{y}, A\dot{\mathbf{x}}\rangle \ , \quad \dot{\mathbf{x}} = \frac{d}{dt}\mathbf{x} \\ &= 2\Re\langle A^*(A\mathbf{x} - \mathbf{y}), \dot{\mathbf{x}}\rangle \end{aligned} \tag{88}$$

and setting

$$\dot{\mathbf{x}} = -A^*(A\mathbf{x} - \mathbf{y}) \ , \tag{89}$$

it follows from (88) that

$$\frac{d}{dt} f(\mathbf{x}) = -2\|A^*(A\mathbf{x} - \mathbf{y})\|^2 < 0 \,. \tag{90}$$

This version of the steepest descent method, given in Rosenbloom [**1308**], results in $f(\mathbf{x}(t))$ being a monotone decreasing function of t, asymptotically approaching its infimum as $t \to \infty$. We expect $\mathbf{x}(t)$ to approach asymptotically $A^\dagger \mathbf{y}$, so by solving (89)

$$\mathbf{x}(t) = \int_0^t \exp\{-A^*A(t-s)\}A^*\mathbf{y}\, ds \tag{91}$$

and observing that $\mathbf{y}$ is arbitrary we get

$$A^\dagger = \lim_{t\to\infty} \exp\{-A^*A(t-s)\}A^*\, ds \tag{92}$$

which is the essence of Theorem 9.

Here as elsewhere in this section, the convergence is in the strong operator topology. Thus the limiting expression

$$A^\dagger = \lim_{t\to\infty} B(t) \quad \text{or} \quad B(t) \to A^\dagger \text{ or } t \to \infty \tag{93}$$

means that for all $\mathbf{y} \in D(A^\dagger)$

$$A^\dagger \mathbf{y} = \lim_{t\to\infty} B(t)\mathbf{y}$$

in the sense that

$$\lim_{t\to\infty} \|(A^\dagger - B(t))\mathbf{y}\| = 0 \,. \tag{94}$$

A numerical integration of (89) with suitably chosen step size similarly results in

$$A^\dagger = \sum_{k=0}^{\infty} (I - \alpha A^*A)^k \alpha A^* \,, \tag{95}$$

where

$$0 < \alpha < \frac{2}{\|A\|^2} \,, \tag{96}$$

which is the essence of Theorem 10.

In statements like (94) it is necessary do distinguish between points $\mathbf{y} \in \mathcal{H}_2$ relative to the given $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. Indeed, the three cases

$$P_{\overline{R(A)}}\mathbf{y} \in R(AA^*) \,, \quad P_{\overline{R(A)}}\mathbf{y} \in (R(A) \setminus R(AA^*)) \,, \quad P_{\overline{R(A)}}\mathbf{y} \in (\overline{R(A)} \setminus R(A))$$

have different rates of convergence in (94). Here $\mathbf{x} \in (X \setminus Y)$ means $\mathbf{x} \in X$, $\mathbf{x} \notin Y$. We abbreviate these as follows:

$$\begin{aligned} (\mathbf{y} \in \mathrm{I}) &\quad \text{means} \quad P_{\overline{R(A)}}\mathbf{y} \in R(AA^*) \,, \\ (\mathbf{y} \in \mathrm{II}) &\quad \text{means} \quad P_{\overline{R(A)}}\mathbf{y} \in (R(A) \setminus R(AA^*)) \,, \\ (\mathbf{y} \in \mathrm{III}) &\quad \text{means} \quad P_{\overline{R(A)}}\mathbf{y} \in (\overline{R(A)} \setminus R(A)) \,. \end{aligned} \tag{97}$$

We note that $A^\dagger \mathbf{y}$ is not defined for $(\mathbf{y} \in \mathrm{III})$, a case which does not exist if $R(A)$ is closed.

THEOREM 9. (Showalter and Ben–Israel [**1354**]). Let $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and define, for $t \geq 0$

$$\begin{aligned} L_1(t) &= \int_0^1 \exp\{-A^*A(t-s)\}\,ds\,, \\ L_2(t) &= \int_0^1 \exp\{-AA^*(t-s)\}\,ds\,, \\ B(t) &= L_1(t)A^* = A^*L_2(t)\,. \end{aligned} \tag{98}$$

Then:

(a) $\|(A^\dagger - B(t))\mathbf{y}\|^2 \leq \dfrac{\|A^\dagger\mathbf{y}\|^2\|(AA^*)^\dagger\mathbf{y}\|^2}{\|(AA^*)^\dagger\mathbf{y}\|^2 + 2\|A^\dagger\mathbf{y}\|^2 t}$ if ($\mathbf{y} \in$ I) and $t \geq 0$.

(b) $\|(A^\dagger - B(t))\mathbf{y}\|^2$ is a decreasing function of $t \geq 0$,
with limit zero as $t \to \infty$, if ($\mathbf{y} \in$ II).

(c) $\|(P_{\overline{R(A)}} - AB(t))\mathbf{y}\|^2 \leq \dfrac{\|\mathbf{y}\|^2\|A^\dagger\mathbf{y}\|^2}{\|A^\dagger\mathbf{y}\|^2 + 2\|\mathbf{y}\|^2 t}$ if ($\mathbf{y} \in$ I) or ($\mathbf{y} \in$ II), and $t \geq 0$.

(d) $\|(P_{\overline{R(A)}} - AB(t))\mathbf{y}\|^2$ is a decreasing function of $t \geq 0$,
with limit zero as $t \to \infty$, if ($\mathbf{y} \in$ III). □

Note that even though $A^\dagger\mathbf{y}$ is not defined for ($\mathbf{y} \in$ III), still

$$AB(t) \to P_{\overline{R(A)}} \quad \text{as } t \to \infty\,.$$

The discrete version of Theorem 9 is the following theorem.

THEOREM 10. (Showalter and Ben–Israel [**1354**]). Let $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, let c be a real number, $0 < c < 2$, and let

$$\alpha = \frac{c}{\|A\|^2}\,.$$

For ant $\mathbf{y} \in \mathcal{H}_2$ define

$$\mathbf{x} = T^\dagger\mathbf{y} \quad \text{if } (\mathbf{y} \in \text{I}) \text{ or } (\mathbf{y} \in \text{II})$$

and define the sequence

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{0}\,, \quad \mathbf{x}_0 = \mathbf{0}\,, \\ (\mathbf{y} - \mathbf{y}_{N+1}) &= (I - \alpha AA^*)\,(\mathbf{y} - \mathbf{y}_N) \quad \text{if } (\mathbf{y} \in \text{I}) \text{ or } (\mathbf{y} \in \text{II}) \text{ or } (\mathbf{y} \in \text{III}) \\ (\mathbf{x} - \mathbf{x}_{N+1}) &= (I - \alpha A^*A)\,(\mathbf{x} - \mathbf{x}_N) \quad \text{if } (\mathbf{y} \in \text{I}) \text{ or } (\mathbf{y} \in \text{II}) \\ N &= 1, 2, \ldots \end{aligned}$$

Then the sequence

$$B_N = \sum_{k=0}^{N} (I - \alpha A^*A)^k \alpha A^*\,, \quad N = 0, 1, \ldots \tag{99}$$

converges to $A^\dagger$ as follows:

(a) $\|(A^\dagger - B_N)\mathbf{y}\|^2 \leq \dfrac{\|A^\dagger\mathbf{y}\|^2\|(AA^*)^\dagger\mathbf{y}\|^2}{\|(AA^*)^\dagger\mathbf{y}\|^2 + N[(2-c)c/\|A\|^2]\|A^\dagger\mathbf{y}\|^2}$ if ($\mathbf{y} \in$ I) and $N = 1, 2, \ldots$

(b) $\|(A^\dagger - B_N)\mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{x}_N\|^2$ converges monotonically to zero if ($\mathbf{y} \in$ II).

(c) $\|(P_{\overline{R(A)} - AB_N}\mathbf{y}\|^2 \leq \dfrac{\|\mathbf{y}\|^2\|A^\dagger\mathbf{y}\|^2}{\|A^\dagger\mathbf{y}\|^2 + N[(2-c)c/\|A\|^2]\|\mathbf{y}\|^2}$ if ($\mathbf{y} \in$ I) or ($\mathbf{y} \in$ II) and $N = 1, 2, \ldots$

(d) $\|(P_{\overline{R(A)} - AB_N}\mathbf{y}\|^2 = \|\mathbf{y} - \mathbf{y}_N\|^2$ converges monotonically to zero if ($\mathbf{y} \in$ III). □

The convergence $B_N \to A^\dagger$, in the uniform operator topology, was established by Petryshyn [**1183**], restricting A to have closed range.

As in the matrix case, studied in Section 7.5, higher–order iterative methods are more efficient means of summing the series (95) than the first–order method (99). Two such methods, of order $p \geq 2$, are given in the following:

THEOREM 11. (Showalter and Ben–Israel [**1354**]). Let A, α and $\{B_N : N = 0, 1, \ldots\}$ be as in Theorem 10. Let p be an integer

$$p \geq 2$$

and define the sequence $\{C_{N,p} : N = 0, 1 \ldots\}$ and $\{D_{N,p} : N = 0, 1 \ldots\}$ as follows:

$$C_{0,p} = \alpha A^* , \quad C_{N+1,p} = C_{N,p} \sum_{k=0}^{p-1} (I - AC_{N,p})^k , \tag{100}$$

$$D_{0,p} = \alpha A^* , \quad D_{N+1,p} = D_{N,p} \sum_{k=0}^{p} \binom{p}{k} (-AD_{N,p})^{k-1} . \tag{101}$$

Then, for all $N = 0, 1, \ldots$,

$$B_{(p^{N+1}-1)} = C_{N+1,p} = D_{N+1,p} . \tag{102}$$

$\square$

Consequently $\{C_{N,p}\}$ and $\{D_{N,p}\}$ are p th–order iterative methods for computing $A^{|dag}$, with the convergence rates established in Theorem 10; e.g.,

$$\|(A^\dagger - C_{N,p})\mathbf{y}\|^2 \leq \frac{\|A^\dagger \mathbf{y}\|^2 \|(AA^*)^\dagger \mathbf{y}\|^2}{\|(AA^*)^\dagger \mathbf{y}\|^2 + (p^N - 1)[(2-c)c/\|A\|^2]\|A^\dagger \mathbf{y}\|^2} \quad \text{if } (\mathbf{y} \in \mathrm{I}) \text{ and } N = 1, 2, \ldots$$

The series (100) is somewhat simpler to use if the term $(I - AC_{N,p})^k$ can be evaluated by only $k-1$ operator multiplications, e.g. for matrices. The form (101) is preferable otherwise, e.g. for integral operators,

For other iterative methods and comprehensive bibliographies on the subject see Kammerer and Nashed ([**814**]–[**816**]) and Zlobec [**1653**].

5.1. Exercises and examples.

EX. 50. Let $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ have closed range, let $\mathbf{b} \in \mathcal{H}_2$ and let[2] $B \in R(A^*, A^*)$. Then the sequence

$$\mathbf{x}_{k+1} = \mathbf{x}_k - B(A\mathbf{x}_k - \mathbf{b}) , \quad k = 0, 1, \ldots \tag{103}$$

converges to $A^\dagger \mathbf{b}$ for all $\mathbf{x}_0 \in R(A^*)$ if

$$\rho(P_{R(A^*)} - BA) < 1$$

where $\rho(T)$ denotes the spectral radius of T; see, e.g. Taylor [**1436**, p. 262] (Zlobec [**1653**]).

The choice $B = \alpha A^*$ in (103) reduces it to the iterative method (99). Other choices of B aare given in the following exercise.

EX. 51. *Splitting methods.* Let A be as in Ex. 50, and write

$$A = M + N , \tag{104}$$

where $M \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ has closed range and $N(A) = N(M)$. Choosing

$$B = w\, M^\dagger , \quad w \neq 0$$

in (103) gives

$$\mathbf{x}_{k+1} = [(1-w)I - wM^\dagger N]\mathbf{x}_k + wM^\dagger \mathbf{b} , \quad \mathbf{x}_0 \in R(A^*) , \tag{105}$$

[2]For $S, T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ with closed ranges, $R(S, T) = \{Z : Z = SWR \text{ for some } W \in \mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)\}$.

in particular, for $w = 1$,

$$\mathbf{x}_{k+1} = -wM^{\dagger}N\,\mathbf{x}_k + M^{\dagger}\mathbf{b}\ , \quad \mathbf{x}_0 \in R(A^*)\ . \tag{106}$$

(Zlobec [**1653**], Berman and Neumann [**148**], Berman and Plemmons [**149**]).

Suggested further reading

Section 1. For alternative or more general treatments of generalized inverses of operators see F. V. Atkinson ([**45**], [**46**]), Beutler ([**157**], [**158**]), Davis and Robinson [**381**], Hamburger [**650**], Hansen and Robinson [**653**], Hestenes [**726**], Holmes [**743**], Leach [**920**], Nashed ([**1115**]–[**1119**]), Nashed and Votruba ([**1122**]–[**1124**]), Pietsch [**1186**], Porter and Williams ([**1199**], [**1200**]), Przeworska–Rolewicz and Rolewicz [**1209**], Sheffield [**1346**], Votruba [**1511**], Wyler [**1618**] and Zarantonello [**1626**].

Section 3. For integral equations see K. E. Atkinson [**47**], Courant and Hilbert [**366**], Kammerer and Nashed ([**814**]–[**815**]), Korganoff and Pavel-Parvu [**873**], Lonseth [**969**], and Rall [**1236**].

For applications to Wiener–Hopf operators see Lent [**924**].

For applications to differential operators see also Bradley ([**231**], [**232**]), Courant and Hilbert [**366**], Greub and Rheinboldt [**577**], Kallina [**806**], Locker [**967**], Tucker [**1468**], and Wyler [**1619**].

For application in bifurcation theory see Stakgold [**1387**].

Bibliography

1. K. Abdel-Malek and Harn-Jou Yeh, *On the determination of starting points for parametric surface intersections*, Computer-aided Design **29** (1997), no. 1, 21–35.
2. N. N. Abdelmalek, *On the solutions of the linear least squares problems and pseudo–inverses*, Computing **13** (1974), no. 3-4, 215–228.
3. V. M. Adukov, *Generalized inversion of block Toeplitz matrices*, Linear Algebra and its Applications **274** (1998), 85–124.
4. ______, *Generalized inversion of finite rank Hankel and Toeplitz operators with rational matrix symbols*, Linear Algebra and its Applications **290** (1999), no. 1-3, 119–134.
5. S. N. Afriat, *On the latent vectors and characteristic values of products of pairs of symmetric idempotents*, Quart. J. Math. Oxford Ser. (2) **7** (1956), 76–78.
6. ______, *Orthogonal and oblique projectors and the characteristics of pairs of vector spaces*, Proc. Cambridge Philos. Soc. **53** (1957), 800–816.
7. J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, *The theory of splines and their applications*, Academic Press, New York, 1967.
8. A. C. Aitken, *On least squares and linear combinations of observations*, Proceedings of the Royal Society of Edinburgh, Sec A **55** (1934), 42–47.
9. Y. Akatsuka and T. Matsuo, *Optimal control of linear discrete systems using the generalized inverse of a matrix*, Techn Rept. 13, Institute of Automatic Control, Nagoya Univ., Nagoya, Japan, 1965.
10. I. S. Alalouf and G. P. H. Styan, *Characterizations of estimability in the general linear model*, Ann. Statist. **7** (1979), no. 1, 194–200.
11. ______, *Estimability and testability in restricted linear models*, Math. Operationsforsch. Statist. Ser. Statist. **10** (1979), no. 2, 189–201.
12. A. Albert, *Conditions for positive and nonnegative definiteness in terms of pseudo–inverses*, SIAM J. Appl. Math. **17** (1969), 434–440.
13. ______, *Regression and the Moore–Penrose Pseudoinverse*, Academic Press, New York, 1972.
14. ______, *The Gauss–Markov theorem for regression models with possibly singular covariabes*, SIAM J. Appl. Math. **24** (1973), 182–187.
15. ______, *Statistical applications of the pseudo inverse*, In Nashed [**1116**], pp. 525–548.
16. A. Albert and R. W. Sittler, *A method for computing least squares estimators that keep up with the data*, SIAM J. Control **3** (1965), 384–417.
17. V. Aleksić and V. Rakočević, *Approximate properties of the Moore-Penrose inverse*, VIII Conference on Applied Mathematics (Tivat, 1993), Univ. Montenegro, Podgorica, 1994, pp. 1–14.
18. E. L. Allgower, K. Böhmer, A. Hoy, and V. Janovský, *Direct methods for solving singular nonlinear equations*, ZAMM Z. Angew. Math. Mech. **79** (1999), 219–231.
19. M. Altman, *A generalization of Newton's method*, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. **3** (1955), 189–193.
20. ______, *On a generalization of Newton's method*, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. **5** (1957), 789–795.
21. J. K. Amburgey, T. O. Lewis, and T. L. Boullion, *On computing generalized characteristic vectors and values for a rectangular matrix*, In Boullion and Odell [**207**], pp. 267–275.
22. A. R. Amir-Moéz, *Geometry of generalized inverses*, Math. Mag. **43** (1970), 33–36.
23. ______, *Extreme properties of linear transformations*, Polygonal Publ. House, Washington, NJ, 1990.
24. C. L. Anderson, *A geometric theory of pseudoinverses and some applications in statistics*, Master's thesis in statistics, Southern Methodist Univ., 1967.
25. W. N. Anderson, Jr., *Shorted operators*, SIAM J. Appl. Math. **20** (1971), 520–525.
26. W. N. Anderson, Jr. and R. J. Duffin, *Series and parallel addition of matrices*, SIAM J. Appl. Math. **26** (1969), 576–594, (see [**878**]).
27. W. N. Anderson, Jr. and M. Schreiber, *The infimum of two projections*, Acta Sci. Math. (Szeged) **33** (1972), 165–168.
28. W. N. Anderson, Jr. and G. E. Trapp, *Inequalities for the parallel connection of resistive n-port networks*, J. Franklin Inst. **209** (1975), no. 5, 305–313.
29. ______, *Shorted operators. II*, SIAM J. Appl. Math. **28** (1975), 60–71, (this concept first introduced by Krein [**880**]).
30. ______, *Analytic operator functions and electrical networks*, In Campbell [**267**], pp. 12–26.
31. ______, *Inverse problems for means of matrices*, SIAM J. Algebraic Discrete Methods **7** (1986), no. 2, 188–192.
32. T. Ando, *Generalized Schur complements*, Linear Algebra and its Applications **27** (1979), 173–186.

33. Mihai Anitescu, Dan I. Coroian, M. Zuhair Nashed, and Florian A. Potra, *Outer inverses and multibody system simulation*, Numer. Funct. Anal. Optim. **17** (1996), no. 7-8, 661–678.
34. P. M. Anselone and P. J. Laurent, *A general method for the construction of interpolating or smoothing spline-functions*, Numer. Math. **12** (1968), 66–82.
35. H. Anton and C. S. Duris, *On minimum norm and best approximate solutions of $Av = b$ in normed spaces*, J. Approximation Theory **16** (1976), no. 3, 245–250.
36. E. Arghiriade, *Sur les matrices qui sont permutables avec leur inverse généralisée*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **35** (1963), 244–251.
37. ______, *On the generalized inverse of a product of matrices*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **5** (1967), 37–42.
38. ______, *Remarques sur l'inverse généralisée d'un produit de matrices*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **42** (1967), 621–625.
39. ______, *Sur quelques équations fonctionnelles de matrices*, Rev. Roumaine Math. Pures Appl. **12** (1967), 1127–1133.
40. ______, *Sur l'inverse généralisée d'un operateur lineaire dans les espaces de Hilbert*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **45** (1968), 471–477.
41. E. Arghiriade and A. Dragomir, *Une nouvelle définition de l'inverse généralisée d'une matrice*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **35** (1963), 158–165.
42. ______, *Remarques sur quelques théoremes relatives a l'inverse généralisée d'un operateur lineaire dans les espaces de hilbert*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **46** (1969), 333–338.
43. I. K. Argyros, *Local convergence theorems of Newton's method for nonlinear equations using outer or generalized inverses*, Czechoslovak Math. J. **50(125)** (2000), no. 3, 603–614.
44. S. Aronowitz and B. E. Eichinger, *Petrie matrices and generalized inverses*, J. Math. Phys. **16** (1975), 1278–1283.
45. F. V. Atkinson, *The normal solvability of linear equations in normed spaces (russian)*, Mat. Sbornik N.S. **28(70)** (1951), 3–14.
46. ______, *On relatively regular operators*, Acta Sci. Math. Szeged **15** (1953), 38–56.
47. K. E. Atkinson, *The solution of non-unique linear integral equations*, Numer. Math. **10** (1967), 117–124, (see also [**1090**]).
48. L. Autonne, Bull. Soc. Math. France **30** (1902), 121–133.
49. ______, *Sur les matrices hypohermitiennes et sur les matrices unitaires*, Ann. Univ. Lyon **38** (1917), 1–77.
50. G. Backus, *Inference from inadequate and inaccurate data. I, II*, Proc. Nat. Acad. Sci. U.S.A. 65 (1970), 1–7; ibid. **65** (1970), 281–287.
51. G. Backus and F. Gilbert, *Uniqueness in the inversion of inaccurate gross Earth data*, Philos. Trans. Roy. Soc. London Ser. A **266** (1970), no. 1173, 123–192.
52. C. Badea and M. Mbekhta, *Generalized inverses and the maximal radius of regularity of a Fredholm operator*, Integral Equations Operator Theory **28** (1997), no. 2, 133–146.
53. ______, *Compressions of resolvents and maximal radius of regularity*, Trans. Amer. Math. Soc. **351** (1999), no. 7, 2949–2960.
54. J. K. Baksalary and R. Kala, *The matrix equation $AX - YB = C$*, Linear Algebra and its Applications **25** (1979), 41–43.
55. ______, *The matrix equation $AXB + CYD = E$*, Linear Algebra and its Applications **30** (1980), 141–147.
56. ______, *Two properties of a nonnegative definite matrix*, Bull. Acad. Polon. Sci. Sér. Sci. Math. **28** (1980), no. 5-6, 233–235 (1981).
57. ______, *Range invariance of certain matrix products*, Linear and Multilinear Algebra **14** (1983), no. 1, 89–96.
58. J. K. Baksalary and T. Mathew, *Rank invariance criterion and its application to the unified theory of least squares*, Linear Algebra and its Applications **127** (1990), 393–401.
59. J. K. Baksalary, P. R. Pordzik, and G. Trenkler, *A note on generalized ridge estimators*, Comm. Statist. Theory Methods **19** (1990), no. 8, 2871–2877.
60. J. K. Baksalary and F. Pukelsheim, *On the Löwner, minus, and star partial orderings of nonnegative definite matrices and their squares*, Linear Algebra and its Applications **151** (1991), 135–141.
61. J. K. Baksalary, S. Puntanen, and H. Yanai, *Canonical correlations associated with symmetric reflexive generalized inverses of the dispersion matrix*, Linear Algebra and its Applications **176** (1992), 61–74.
62. A. V. Balakrishnan, *An operator theoretic formulation of a class of control problems and a steepest descent method of solution*, J. Soc. Indust. Appl. Math. Ser. A: Control **1** (1963), 109–127.
63. K. F. Baldwin and A. E. Hoerl, *Bounds of minimum mean squared error in ridge regression*, Comm. Statist. A—Theory Methods **7** (1978), no. 13, 1209–1218.
64. J. A. Ball, M. Rakowski, and B. F. Wyman, *Coupling operators, Wedderburn-Forney spaces, and generalized inverses*, Linear Algebra and its Applications **203/204** (1994), 111–138.
65. K. S. Banerjee, *Singularity in Hotelling's weighing designs and generalized inverses*, Ann. Math. Statist. **37** (1966), 1021–1032, (erratum, *ibid* **40**(1969), 710).

66. K. S. Banerjee and W. T. Federer, *On the structure and analysis of singular fractional replicates*, Ann. Math. Statist. **39** (1968), 657–663.
67. R. B. Bapat, *Generalized inverses with proportional minors*, Linear Algebra and its Applications **211** (1994), 27–33.
68. ______, *Moore–Penrose inverse of the incidence matrix of a tree*, Linear and Multilinear Algebra **42** (1997), no. 2, 159–167.
69. ______, *Structure of a nonnegative regular matrix and its generalized inverses*, Linear Algebra and its Applications **268** (1998), 31–39.
70. ______, *Linear Algebra and Linear Models*, second ed., Hindustan Book Agency, New Delhi, 1999.
71. ______, *Linear estimation in models based on a graph*, Linear Algebra and its Applications **302/303** (1999), 223–230.
72. ______, *Resistance distance in graphs*, Mathematics Student **68** (1999), 87–98.
73. ______, *Moore-Penrose inverse of set inclusion matrices*, Linear Algebra and its Applications **318** (2000), no. 1-3, 35–44.
74. R. B. Bapat and A. Ben-Israel, *Singular values and maximum rank minors of generalized inverses*, Linear and Multilinear Algebra **40** (1995), no. 2, 153–161.
75. R. B. Bapat and R. E. Hartwig, *A master of the row space and the column space: the mathematical work of Sujit Kumar Mitra*, In Bapat et al. [**80**], (special issue of *Linear Algebra and its Applications* **211** (1994)), pp. 5–14.
76. R. B. Bapat, S. K. Jain, and S. Pati, *Weighted Moore-Penrose inverse of a Boolean matrix*, Linear Algebra and its Applications **255** (1997), 267–279.
77. R. B. Bapat, S. K. Jain, and K. M. Prasad, *Generalized power symmetric stochastic matrices*, Proc. Amer. Math. Soc. **127** (1999), no. 7, 1987–1994.
78. R. B. Bapat, S. K. Jain, and L. E. Snyder, *Nonnegative idempotent matrices and the minus partial order*, Linear Algebra and its Applications **261** (1997), 143–154.
79. R. B. Bapat and D. M. Kulkarni, *Minors of some matrices associated with a tree*, Algebra and Its Applications. Contemporary Mathematics 259 (D. V. Huynh, S. K. Jain, S. R. Lopez-Permouth Ed.), American Math Society, Providence, RI, 2000, pp. 45–66.
80. R. B. Bapat, S. K. Mitra, and R. Hartwig (eds.), *Generalized inverses. papers from the workshop on g-inverses held in calcutta, december 11–16, 1993*, New York, North-Holland Publishing Co., 1994, (special issue of *Linear Algebra and its Applications* **211** (1994)).
81. R. B. Bapat and Sukanta Pati, *Algebraic connectivity and the characteristic set of a graph*, Linear and Multilinear Algebra **45** (1998), no. 2-3, 247–273.
82. R. B. Bapat and K. M. Prasad, *Cochran's theorem and related results on matrix rank over a commutative ring*, Statistical Inference and Design of Experiments (U. J. Dixit and M. R. Satam Ed.), Narosa Publishing House, 1999, pp. 125–133.
83. R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press, Cambridge, 1997.
84. R. B. Bapat, K. P. S. Bhaskara Rao, and K. Manjunatha Prasad, *Generalized inverses over integral domains*, Linear Algebra and its Applications **140** (1990), 181–196.
85. R. B. Bapat and D. W. Robinson, *The Moore-Penrose inverse over a commutative ring*, Linear Algebra and its Applications **177** (1992), 89–103.
86. G. P. Barker and S. L. Campbell, *Internal stability of two-dimensional systems*, Linear and Multilinear Algebra **14** (1983), no. 4, 365–369.
87. S. Barnett, *Matrices in Control Theory*, Van Nostrand Reinhold, London, 1971.
88. G. Basile, *Alcune osservazioni sulla pseudoinversa di una matrice rettangolare.*, Atti Accad. Sci. Ist. Bologna Cl. Sci. Fis. Rend. (12) **6** (1968/1969), no. fasc., 1–2, 236–240.
89. T. S. Baskett and I. J. Katz, *Theorems on products of EP_r matrices*, Linear Algebra and its Applications **2** (1969), 87–103.
90. H. Bateman, *A formula for the solving function of a certain integral equation of the second kind*, Transactions of the Cambridge Philosophical Society **20** (1908), 179–187.
91. ______, *On the application of integral equations to the determination of upper and lower limits of a double integral*, Transactions of the Cambridge Philosophical Society **21** (1908), 123–128.
92. ______, *The reality of the roots of certain transcendental equations occurring in the theory of integral equations*, Transactions of the Cambridge Philosophical Society **20** (1908), 371–381.
93. ______, *On the numerical solution of linear integral equations*, Proc. Roy. Soc. London Ser. A **100** (1922), 441–449.
94. D. Batigne, *Integral generalized inverses of integral matrices*, Linear Algebra and its Applications **22** (1978), 125–134.
95. D. R. Batigne, F. J. Hall, and I. J. Katz, *Further results on integral generalized inverses of integral matrices*, Linear and Multilinear Algebra **6** (1978/79), no. 3, 233–241.
96. F. L. Bauer, *A further generalization of the Kantorovič inequality*, Numer. Math. **3** (1961), 117–119.
97. ______, *Elimination with weighted row combinations for solving linear equations and least squares problems*, Numer. Math. **7** (1965), 338–352, (republished, pp. 119–133 in [**1598**]).
98. ______, *Theory of norms*, Computer Science Dept. CS 75, Stanford University, Stanford, 1967.

99. F. L. Bauer, J. Stoer, and C. Witzgall, *Absolute and monotonic norms*, Numer. Math. **3** (1961), 257–264.
100. E. F. Beckenbach and R. Bellman, *Inequalities*, 3rd ed., Springer-Verlag, New York, 1971.
101. R. Bellman, *Introduction to Matrix Analysis*, 2nd ed., McGraw-Hill Book Co., New York, 1970.
102. E. Beltrami, *Sulle funzioni bilineari*, Giornale di Matematiche ad Uso degli Studenti Delle Universita **11** (1873), 98–106, (an English translation by D. Boley is available as University of Minnesota, Department of Computer Science, Technical Report 90–37, 1990.).
103. E. J. Beltrami, *A constructive proof of the Kuhn–Tucker multiplier rule*, J. Math. Anal. Appl. **26** (1969), 297–306.
104. A. Ben-Israel, *On direct sum decompositions of Hestenes algebras*, Israel J. Math. **2** (1964), 50–54.
105. ______, *An iterative method for computing the generalized inverse of an arbitrary matrix*, Math. Comput. **19** (1965), 452–455.
106. ______, *A modified Newton-Raphson method for the solution of systems of equations*, Israel J. Math. **3** (1965), 94–98.
107. ______, *A Newton-Raphson method for the solution of systems of equations*, J. Math. Anal. Appl. **15** (1966), 243–252.
108. ______, *A note on an iterative method for generalized inversion of matrices*, Math. Comput. **20** (1966), 439–440.
109. ______, *A note on the Cayley transform*, Notices Amer. Math. Soc. **13** (1966), 599.
110. ______, *On error bounds for generalized inverses*, SIAM J. Numer. Anal. **3** (1966), 585–592, (see also [**1399**]).
111. ______, *On iterative methods for solving nonlinear least squares problems over convex sets*, Israel J. Math. **5** (1967), 211–214.
112. ______, *On the geometry of subspaces in Euclidean n–spaces*, SIAM J. Appl. Math. **15** (1967), 1184–1198.
113. ______, *On applications of generalized inverses in nonlinear analysis*, In Boullion and Odell [**207**], pp. 183–202.
114. ______, *On decompositions of matrix spaces with applications to matrix equations*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **45** (1968), 122–128.
115. ______, *On optimal solutions of* 2*-person* 0*-sum games*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **44** (1968), 512–516.
116. ______, *A note on partitioned matrices and equations*, SIAM Rev. **11** (1969), 247–250.
117. ______, *On matrices of index zero or one*, SIAM J. Appl. Math. **17** (1969), 1118–1121, (see [**1028**], [**1309**]).
118. ______, *On Newton's method in nonlinear programming*, In Kuhn [**890**], pp. 339–352.
119. ______, *Applications of generalized inverses to programming, games, and networks*, In Nashed [**1116**], pp. 495–523.
120. ______, *A Cramer rule for least-norm solutions of consistent linear equations*, Linear Algebra and its Applications **43** (1982), 223–226, (extended in [**1502**]).
121. ______, *Generalized inverses of matrices: a perspective of the work of Penrose*, Math. Proc. Cambridge Philos. Soc. **100** (1986), no. 3, 407–425.
122. ______, *A volume associated with $m \times n$ matrices*, Linear Algebra and its Applications **167** (1992), 87–111, (this concept was introduced by Good [**563**]).
123. ______, *The change-of-variables formula using matrix volume*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 1, 300–312 (electronic).
124. ______, *An application of the matrix volume in probability*, Linear Algebra and its Applications **321** (2001), 9–25.
125. ______, *A local inverse for nonlinear mappings*, Numerical Algorithms **25** (2001), 37–46.
126. A. Ben-Israel and A. Charnes, *Contributions to the theory of generalized inverses*, J. Soc. Indust. Appl. Math. **11** (1963), 667–699.
127. ______, *Generalized inverses and the Bott-Duffin network analysis*, J. Math. Anal. Appl. **7** (1963), 428–435, (corrigendum in *J. Math. Anal. Appl.* **18**(1967), 393).
128. ______, *An explicit solution of a special class of linear programming problems*, Operations Res. **16** (1968), 1166–1175, (see [**145**], [**170**], [**1277**], [**1313**]).
129. A. Ben-Israel, A. Charnes, and P. D. Robers, *On generalized inverses and interval linear programming*, In Boullion and Odell [**207**], pp. 53–70.
130. A. Ben-Israel and D. Cohen, *On iterative computation of generalized inverses and associated projections*, SIAM J. Numer. Anal. **3** (1966), 410–419.
131. A. Ben-Israel and T. N. E. Greville, *Generalized inverses: theory and applications*, Wiley-Interscience [John Wiley & Sons], New York, 1974, (reprinted by Robert E. Krieger Publishing Co. Inc., Huntington, NY, 1980.).
132. ______, *Some topics in generalized inverses of matrices*, In Nashed [**1116**], pp. 125–147.
133. A. Ben-Israel and M. J. L. Kirby, *A characterization of equilibrium points of bimatrix games*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **46** (1969), 402–407.
134. A. Ben-Israel and S. J. Wersan, *An elimination method for computing the generalized inverse of an arbitrary complex matrix*, J. Assoc. Comput. Mach. **10** (1963), 532–537.
135. A. Ben-Tal and M. Teboulle, *A geometric property of the least squares solution of linear equations*, Linear Algebra and its Applications **139** (1990), 165–170, (see [**142**], [**406**], [**122**], [**136**], [**188**], [**502**]).

136. ———, *Addenda: "A geometric property of the least squares solution of linear equations" [Linear Algebra Appl.* **139** *(1990), 165–170]*, Linear Algebra and its Applications **180** (1993), 5.
137. M. W. Benson and P. O. Frederickson, *Fast parallel algorithms for the Moore-Penrose pseudo-inverse*, Hypercube multiprocessors 1987 (Knoxville, TN, 1986), SIAM, Philadelphia, PA, 1987, pp. 597–604.
138. ———, *Fast pseudo-inverse algorithms on hypercubes*, Multigrid methods (Copper Mountain, CO, 1987), Dekker, New York, 1988, pp. 23–33.
139. M. Benzi and C. D. Meyer, Jr., *A direct projection method for sparse linear systems*, SIAM J. Sci. Comput. **16** (1995), no. 5, 1159–1176.
140. I.S. Berezin and N.P. Zhidkov, *Computing Methods*, Pergamon Press, London, 1965.
141. L. Berg, *Über quasivertauschbare Matrixinversen*, Rostock. Math. Kolloq. (1980), no. 15, 5–10.
142. ———, *Three results in connection with inverse matrices*, Linear Algebra and its Applications **84** (1986), 63–77, (see also [**135**]).
143. P. G. Bergman, R. Penfield, R. Schiller, and H. Zatkis, *The Hamiltonian of the general theory of relativity with electromagnetic field*, Physical Review **52** (1950), 1950.
144. A. Berman, *Nonnegative matrices which are equal to their generalized inverse*, Linear Algebra and its Applications **9** (1974), 261–265.
145. ———, *Generalized interval programming*, Bull. Calcutta Math. Soc. **71** (1979), no. 3, 169–176.
146. A. Berman and S. K. Jain, *Nonnegative generalized inverses of powers of nonnegative matrices*, Proceedings of the Victoria Conference on Combinatorial Matrix Analysis (Victoria, BC, 1987), vol. 107, 1988, pp. 169–179.
147. A. Berman and M. Neumann, *Consistency and splittings*, SIAM J. Numer. Anal. **13** (1976), no. 6, 877–888.
148. ———, *Proper splittings of rectangular matrices*, SIAM J. Appl. Math. **31** (1976), no. 2, 307–312.
149. A. Berman and R. J. Plemmons, *Monotonicity and the generalized inverse*, SIAM J. Appl. Math. **22** (1972), 155–161.
150. ———, *Cones and iterative methods for best least squares solutions of linear systems*, SIAM J. Numer. Anal. **11** (1974), 145–154.
151. ———, *Inverses of nonnegative matrices*, Linear and Multilinear Algebra **2** (1974), 161–172.
152. ———, *Nonnegative Matrices in the Mathematical Sciences*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994, (revised reprint of the 1979 original).
153. L. Bernard, *A generalized inverse method for asymptotic linear programming*, Mathematical Programming **43** (1989), 71–86.
154. ———, *An efficient basis update for asymptotic linear programming*, Linear Algebra and its Applications **184** (1993), 83–102.
155. M. Bertero, C. De Mol, and E. R. Pike, *Linear inverse problems with discrete data. I. General formulation and singular system analysis*, Inverse Problems **1** (1985), no. 4, 301–330.
156. J. Bérubé, R. E. Hartwig, and G. P. H. Styan, *On canonical correlations and the degrees of nonorthogonality in the three-way layout*, Statistical sciences and data analysis (Tokyo, 1991), VSP, Utrecht, 1993, pp. 247–252.
157. F. J. Beutler, *The operator theory of the pseudo-inverse. I. Bounded operators*, J. Math. Anal. Appl. **10** (1965), 451–470.
158. ———, *The operator theory of the pseudo-inverse. II. Unbounded operators with arbitrary range*, J. Math. Anal. Appl. **10** (1965), 471–493.
159. F. J. Beutler and W. L. Root, *The operator pseudoinverse in control and systems identification*, In Nashed [**1116**], pp. 397–494.
160. J. H. Bevis, F. J. Hall, and R. E. Hartwig, *Consimilarity and the matrix equation* $A\overline{X} - XB = C$, Current trends in matrix theory (Auburn, Ala., 1986), North-Holland, New York, 1987, pp. 51–64.
161. ———, *The Drazin inverse of a semilinear transformation and its matrix representation*, Linear Algebra and its Applications **97** (1987), 229–242.
162. J. H. Bevis, F. J. Hall, and I. J. Katz, *Integer generalized inverses of incidence matrices*, Linear Algebra and its Applications **39** (1981), 247–258.
163. R. Bhatia, *Perturbation bounds for matrix eigenvalues*, Longman Scientific & Technical, Harlow, 1987.
164. ———, *Letter to the editor: "The n-dimensional Pythagorean theorem" [Linear and Multilinear Algebra* **26** *(1990), no. 1-2, 9–13; MR 90k:51031] by S.-Y. T. Lin and Y. F. Lin*, Linear and Multilinear Algebra **30** (1991), no. 1-2, 155, (see [**960**]).
165. ———, *Matrix analysis*, Springer-Verlag, New York, 1997.
166. P. Bhimasankaram, *A characterization of subclasses of generalized inverses of specified rank*, Sankhyā Ser. A **36** (1974), no. 2, 214–218.
167. ———, *On generalized inverses of a block in a partitioned matrix*, Linear Algebra and its Applications **109** (1988), 131–143.
168. ———, *Rank factorization of a matrix and its applications*, Math. Sci. **13** (1988), no. 1, 4–14, (see [**1267**]).
169. P. Bhimasankaram and Thomas Mathew, *On ordering properties of generalized inverses of nonnegative definite matrices*, Linear Algebra and its Applications **183** (1993), 131–146.
170. M. Bilodeau, *Sur une représentation explicite des solutions optimales d'un programme linéaire*, Canad. Math. Bull. **29** (1986), no. 4, 419–425.
171. Z. W. Birnbaum, *Introduction to probability and mathematical statistics*, Harper & Brothers, Publishers, New York, 1962.
172. A. Bjerhammar, *Application of calculus of matrices to method of least squares with special reference to*

geodetic calculations, Trans. Roy. Inst. Tech. Stockholm **1951** (1951), no. 49, 86 pp. (2 plates).

173. ———, *Rectangular reciprocal matrices, with special reference to geodetic calculations*, Bull. Géodésique (1951), 188–220.
174. ———, *A generalized matrix algebra*, Trans. Roy. Inst. Tech. Stockholm **1958** (1958), no. 124, 32 pp.
175. ———, *Studies with generalized matrix algebra*, Bull. Géodésique (N.S.) No. **85** (1967), 193–210.
176. ———, *Theory of errors and generalized matrix inverses*, Elsevier Scientific Publishing Co., Amsterdam, 1973.
177. Å. Björck, *Iterative refinement of linear least squares solutions I*, BIT **7** (1967), 257–278.
178. ———, *Solving linear least squares problems by Gram–Schmidt orthogonalization*, BIT **7** (1967), 1–21.
179. ———, *Iterative refinement of linear least squares solutions II*, BIT **8** (1968), 8–30.
180. ———, *A uniform numerical method for linear estimation from general gauss-markov models*, Proceedings of the First Symposium on Comutational Staistics (COMPSTAT), (G. Bruckmann, F. Ferschl and L. Schmetterer, Editors), Physica Verlag, Vienna, 1974, pp. 131–140.
181. Å. Björck and Tommy Elfving, *Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations*, BIT **19** (1979), 145–163.
182. Å. Björck and G. H. Golub, *Iterative refinement of linear least squares solutions by householder transformation*, BIT **7** (1967), 322–337.
183. ———, *Numerical methods for computing angles between linear subspaces*, Mathematics of Computation **27** (1973), 579–594.
184. J. Blatter and E. W. Cheney, *On the existence of extremal projections*, J. Approximation Theory **6** (1972), 72–79.
185. J. W. Blattner, *Bordered matrices*, J. Soc. Indust. Appl. Math. **10** (1962), 528–536.
186. ———, *On the convergence of a certain matrix iteration*, Bul. Inst. Politehn. Iaşi (N.S.) **10 (14)** (1964), no. 3-4, 43–46.
187. L. Bober and P. Chrzan, *Application of the generalized Moore-Penrose matrix inversion to the estimation of a classical econometric model under additional constraints*, Przegląd Statyst. **25** (1978), no. 3, 315–324 (1979).
188. E. Y. Bobrovnikova and S. A. Vavasis, *A norm bound for projections with complex weights*, Linear Algebra and its Applications **307** (2000), no. 1-3, 69–75, (A complex version of the bounds in [**1404**], [**1455**]).
189. P. T. Boggs, *The convergence of the Ben-Israel iteration for nonlinear least squares problems*, Math. Comp. **30** (1976), no. 135, 512–522.
190. E. Bohl and P. Lancaster, *Perturbation of spectral inverses applied to a boundary layer phenomenon arising in chemical networks*, Linear Algebra and its Applications **180** (1993), 35–59.
191. F. Bohnenblust, *A characterization of complex Hilbert spaces*, Portugal. Math. **3** (1942), 103–109.
192. T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Springer, Berlin, 1934.
193. C. de Boor, *The Method of Projections as applied to the Numerical Solution of Two Point Boundary Value Problems using Cubic Splines*, Doctoral dissertation in mathematics, University of Michigan, Ann Arbor, MI, 1966.
194. ———, *Convergence of abstract splines*, J. Approx. Theory **31** (1981), no. 1, 80–89.
195. J. C. G. Boot, *The computation of the generalized inverse of singular or rectangular matrices*, Amer. Math. Monthly **70** (1963), 302–303.
196. E. Boroş, *On the generalized inverse of an EP_r matrix*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **2** (1964), 33–38.
197. ———, *On certain properties of EP_r matrices*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **3** (1965), 77–84.
198. ———, *Das verallgemeinerte Inverse eines linearen Operators in Vektorräumen mit nicht ausgearteter Hermitescher Metrik über einem kommutativen Körper*, J. Reine Angew. Math. **252** (1972), 68–78.
199. ———, *The generalized inverse of linear operators on spaces with indefinite metric*, An. Univ. Timişoara Ser. Ştiinţ. Mat. **21** (1983), no. 1-2, 9–44.
200. E. Boroş and I. Sturz, *On quasi-inverse matrices*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **1** (1963), 59–66.
201. N. K. Bose and Sujit K. Mitra, *Generalized inverse of polynomial matrices*, IEEE Trans. Automatic Control **23** (1978), no. 3, 491–493.
202. R. Bott and R. J. Duffin, *On the algebra of networks*, Trans. Amer. Math. Soc. **74** (1953), 99–109.
203. R. H. Bouldin, *The pseudo-inverse of a product*, SIAM J. Appl. Math. **24** (1973), 489–495.
204. ———, *Selfadjoint approximants*, Indiana Univ. Math. J. **27** (1978), no. 2, 299–307.
205. ———, *Generalized inverses and factorizations*, In Campbell [**267**], pp. 233–249.
206. T. L. Boullion, *Contributions to the Theory of Pseudoinverses*, Ph.D. thesis, University of Texas, Austin, 1966.
207. T. L. Boullion and P. L. Odell (eds.), *Proceedings of the Symposium on Theory and Applications Generalized Inverses of Matrices*, Lubbock, Texas Tech. Press, 1968.
208. ———, *A note on the Scroggs-Odell pseudoinverse*, SIAM J. Appl. Math. **17** (1969), 7–10, (correction of [**1331**, Theorem 6]).
209. ———, *Generalized Inverse Matrices*, John Wiley & Sons, New York, 1971.

210. T. L. Boullion and G. D. Poole, *A characterization of the general solution of the matrix equation $AX + XB = C$*, Indust. Math. **20** (1970), 91–95.
211. Z. Boulmaarouf, M. Fernandez Miranda, and J-Ph. Labrousse, *An algorithmic approach to orthogonal projections and Moore-Penrose inverses*, Numer. Funct. Anal. Optim. **18** (1997), no. 1-2, 55–63.
212. E. Bounitzky, *Sur la fonction de Green des équations differentielles linéaires ordinaires*, J. Math. Pures Appl. **5** (1909), no. 6, 65–125.
213. N. Bourbaki, *Eléments de Mathématiques. Livre V. Espaces Vectoriels Topologiques*, Hermann & Cie, Paris, 1953.
214. ______, *Eléments de Mathématiques. Livre II. Algèbre*, Hermann & Cie, Paris, 1958.
215. H. J. Bowdler, R. S. Martin, G. Peters, and J. H. Wilkinson, *Solutions of real and complex systems of linear equations*, Numer. Math. **8** (1966), 217–239, (republished, pp. 93–110 in [**1598**]).
216. V. J. Bowman and C.-A. Burdet, *On the general solution to systems of mixed–integer linear equations*, SIAM J. Appl. Math. **26** (1974), 120–125.
217. Yu. E. Boyarintsev, *General solutions of boundary value problems for singular systems of ordinary differential equations*, Čisl. Metody Meh. Splošn. Sredy **8** (1977), no. 7, 12–21.
218. ______, *A representation of the Drazin inverse matrix*, Numerical Methods of Optimization (Applied Mathematics) (Russian), Akad. Nauk SSSR Sibirsk. Otdel. Ènerget. Inst., Irkutsk, 1978, pp. 176–179.
219. ______, *Regulyarnye i Singulyarnye Sistemy Lineinykh Obyknovennykh Differentsialnykh Uravnenii*, "Nauka" Sibirsk. Otdel., Novosibirsk, 1980.
220. ______, *Solving a pair of matrices*, Approximate Methods for Solving Operator Equations and their Applications, Akad. Nauk SSSR Sibirsk. Otdel. Ènerget. Inst., Irkutsk, 1982, pp. 35–47.
221. ______, *Representation of the solutions of a system of linear algebraic equations by means of generalized inverse matrices*, Computational Methods in Linear Algebra (Russian) (Moscow, 1982), Akad. Nauk SSSR Otdel Vychisl. Mat., Moscow, 1983, pp. 33–45.
222. ______, *The solving pair of matrices and its application*, Current problems in numerical and applied mathematics (Novosibirsk, 1981), "Nauka" Sibirsk. Otdel., Novosibirsk, 1983, pp. 52–55.
223. ______, *Degenerate systems and the index of a variable matrix*, Differential Equations and Numerical Methods, "Nauka" Sibirsk. Otdel., Novosibirsk, 1986, pp. 105–114, 277.
224. ______, *Methods of Solving Singular Systems of Ordinary Differential Equations*, John Wiley & Sons Ltd., Chichester, 1992, (translation of the 1988 Russian original).
225. ______, *A resolving transformation of unknowns in an implicit system of ordinary differential equations*, Algebrodifferential Systems and Methods for their Solution (Russian), VO "Nauka", Novosibirsk, 1993, pp. 4–19, 90.
226. Yu. E. Boyarintsev, V. A. Danilov, A. A. Loginov, and V. F. Chistyakov, *Chislennye Metody Resheniya Singulyarnykh Sistem*, "Nauka" Sibirsk. Otdel., Novosibirsk, 1989.
227. Yu. E. Boyarintsev and V. M. Korsukov, *The structure of a general continuously differentiable solution of a boundary value problem for a singular system of ordinary differential equations*, Questions in Applied Mathematics (Russian), Sibirsk. Ènerget. Inst., Akad. Nauk SSSR Sibirsk. Otdel., Irkutsk, 1977, pp. 73–93.
228. F. Brackx, R. Delanghe, and J. Van hamme, *Generalized inverses of elliptic systems of differential operators with constant coefficients and related REDUCE programs for explicit calculations*, Rend. Circ. Mat. Palermo (2) Suppl. (1987), no. 16, 21–28.
229. H. W. Braden, *R-matrices and generalized inverses*, J. Phys. A **30** (1997), no. 15, L485–L493.
230. ______, *The equations $A^T X \pm X^T A = B$*, SIAM J. Matrix Anal. Appl. **20** (1999), no. 2, 295–302 (electronic).
231. J. S. Bradley, *Adjoint quasi–differential operators of Euler type*, Pacific J. Math. **16** (1966), 213–237.
232. ______, *Generalized Green's matrices for compatible differential systems*, Michigan Math. J. **13** (1966), 97–108.
233. L. Brand, *The solution of linear algebraic equations*, Math. Gaz. **46** (1962), 203–237.
234. C. Brezinski, *Projection methods for linear systems*, J. Comput. Appl. Math. **77** (1997), no. 1-2, 35–51, (ROLLS Symposium (Leipzig, 1996)).
235. C. Brezinski, M. Morandi Cecchi, and M. Redivo Zaglia, *The reverse bordering method*, SIAM J. Matrix Anal. Appl. **15** (1994), no. 3, 922–937.
236. C. Brezinski and M. Redivo Zaglia, *Extrapolation Methods. Theory and practice, with 1 IBM-PC floppy disk (5.25 inch)*, North-Holland Publishing Co., Amsterdam, 1991.
237. P. Broadbridge and H. G. Petersen, *Use of generalized inverses in the construction of Hamiltonians for constrained dynamical systems*, Confronting the infinite (Adelaide, 1994), World Sci. Publishing, River Edge, NJ, 1995, pp. 307–318.
238. C. G. den Broeder Jr. and A. Charnes, *Contributions to the theory of generalized inverses for matrices*, Dept. of math., Purdue University, Lafayette, IN, 1957, (Reprinted as ONR Res. Memo. No. 39, Northwestern University, Evanston, IL, 1962).
239. R. C. Brown, *Generalized Green's functions and generalized inverses for linear differential systems with Stieltjes boundary conditions*, J. Differential Equations **16** (1974), 335–351.
240. R. Bru and N. Thome, *Group inverse and group involutory matrices*, Linear and Multilinear Algebra **45** (1998), no. 2-3, 207–218.

241. J. T. Bruening, *A new formula for the Moore-Penrose inverse*, Current trends in matrix theory (Auburn, Ala., 1986), North-Holland, New York, 1987, pp. 65–74.
242. R. S. Bucy, *Comments on a paper by F. E. Udwadia and R. E. Kalaba: "A new perspective on constrained motion" [Proc. Roy. Soc. London Ser. A* **439** *(1992), no. 1906, 407–410; MR 94b:70027]*, Proc. Roy. Soc. London Ser. A **444** (1994), no. 1920, 253–255, (see [**801**]).
243. Hamza Bulut and S. Aysun Bulut, *Spectral decompositions and generalized inverses in a circularization network flow problem*, J. Math. Anal. Appl. **174** (1993), 390–402.
244. W. Burmeister, *Inversionfreie Verfahren zur lösung nichtlinearer Operatorgleichungen*, Zeit. angew. Math. Mech. **52** (1972), 101–110.
245. F. Burns, D. Carlson, E. V. Haynsworth, and T. Markham, *Generalized inverse formulas using the Schur complement*, SIAM J. Appl. Math. **26** (1974), 254–259.
246. P. A. Businger and G. H. Golub, *Linear least squares by Householder transformations*, Numer. Math. **7** (1965), 269–276, (republished, pp. 111–118 in [**1598**]).
247. ______, *Algorithm 358: Singular value decomposition of a complex matrix*, Comm. ACM **12** (1969), 564–565.
248. C. A. Butler and T. D. Morley, *A note on the shorted operator*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 2, 147–155.
249. ______, *Six generalized Schur complements*, Linear Algebra and its Applications **106** (1988), 259–269.
250. Kim Ki-Hang Butler, *A Moore-Penrose inverse for Boolean relation matrices*, Combinatorial mathematics (Proc. Second Australian Conf., Univ. Melbourne, Melbourne, 1973), Springer, Berlin, 1974, pp. 18–28. Lecture Notes in Math., Vol. 403.
251. G. D. Callon and C. W. Groetsch, *The method of weighting and approximation of restricted pseudosolutions*, J. Approx. Theory **51** (1987), no. 1, 11–18.
252. S. L. Campbell, *Differentiation of the Drazin inverse*, SIAM J. Appl. Math. **30** (1976), no. 4, 703–707.
253. ______, *The Drazin inverse of an infinite matrix*, SIAM J. Appl. Math. **31** (1976), no. 3, 492–503, (see [**265**]).
254. ______, *Optimal control of autonomous linear processes with singular matrices in the quadratic cost functional*, SIAM J. Control Optimization **14** (1976), no. 6, 1092–1106.
255. ______, *Linear systems of differential equations with singular coefficients*, SIAM J. Math. Anal. **8** (1977), no. 6, 1057–1066.
256. ______, *On continuity of the Moore-Penrose and Drazin generalized inverses*, Linear Algebra and Appl. **18** (1977), no. 1, 53–57.
257. ______, *Optimal control of discrete linear processes with quadratic cost*, Internat. J. Systems Sci. **9** (1978), no. 8, 841–847.
258. ______, *Singular perturbation of autonomous linear systems. II*, J. Differential Equations **29** (1978), no. 3, 362–373.
259. ______, *Limit behavior of solutions of singular difference equations*, Linear Algebra and its Applications **23** (1979), 167–178.
260. ______, *Nonregular singular dynamic Leontief systems*, Econometrica **47** (1979), no. 6, 1565–1568.
261. ______, *On a singularly perturbed autonomous linear control problem*, IEEE Trans. Automat. Control **24** (1979), no. 1, 115–117.
262. ______, *Continuity of the Drazin inverse*, Linear and Multilinear Algebra **8** (1979/80), no. 3, 265–268.
263. ______, *Singular Systems of Differential Equations*, Pitman (Advanced Publishing Program), Boston, Mass., 1980.
264. ______, *On an assumption guaranteeing boundary layer convergence of singularly perturbed systems*, Automatica—J. IFAC **17** (1981), no. 4, 645–646.
265. ______, *The Drazin inverse of an operator*, [**267**], pp. 250–260.
266. ______, *On positive controllers and linear quadratic optimal control problems*, Internat. J. Control **36** (1982), no. 5, 885–888.
267. S. L. Campbell (ed.), *Recent Applications of Generalized Inverses*, Boston, Mass., Pitman (Advanced Publishing Program), 1982.
268. ______, *Singular Systems of Differential Equations. II*, Pitman (Advanced Publishing Program), Boston, Mass., 1982.
269. ______, *Index two linear time-varying singular systems of differential equations*, SIAM J. Algebraic Discrete Methods **4** (1983), no. 2, 237–243.
270. ______, *Control problem structure and the numerical solution of linear singular systems*, Math. Control Signals Systems **1** (1988), no. 1, 73–87.
271. ______, *Uniqueness of completions for linear time varying differential algebraic equations*, Linear Algebra and its Applications **161** (1992), 55–67.
272. ______, *Least squares completions for nonlinear differential algebraic equations*, Numer. Math. **65** (1993), no. 1, 77–94.
273. S. L. Campbell and C. D. Meyer, Jr., *Continuity properties of the Drazin pseudoinverse*, Linear Algebra and its Applications **10** (1975), 77–83.
274. ______, *EP operators and generalized inverses*, Canad. Math. Bull **18** (1975), no. 3, 327–333.
275. ______, *Weak Drazin inverses*, Linear Algebra and its Applications **20** (1978), no. 2, 167–178.
276. ______, *Generalized Inverses of Linear Transformations*, Pitman (Advanced Publishing Program), Boston, Mass., 1979, (reprinted by Dover, 1991).
277. S. L. Campbell, C. D. Meyer, Jr., and N. J. Rose, *Applications of the Drazin inverse to linear systems*

of differential equations with singular constant coefficients, SIAM J. Appl. Math. **31** (1976), no. 3, 411–425.

278. S. L. Campbell and G. D. Poole, *Computing nonnegative rank factorizations*, Linear Algebra and its Applications **35** (1981), 175–182.
279. ———, *Convergent regular splittings for nonnegative matrices*, Linear and Multilinear Algebra **10** (1981), no. 1, 63–73.
280. S. L. Campbell and M. Rakowski, *Explicit formulae for completions of linear time varying singular systems of differential equations*, Circuits Systems Signal Process. **13** (1994), no. 2-3, 185–199.
281. S. L. Campbell and N. J. Rose, *Singular perturbation of autonomous linear systems. III*, Houston J. Math. **4** (1978), no. 4, 527–539.
282. ———, *Singular perturbation of autonomous linear systems*, SIAM J. Math. Anal. **10** (1979), no. 3, 542–551.
283. ———, *A second order singular linear system arising in electric power systems analysis*, Internat. J. Systems Sci. **13** (1982), no. 1, 101–108.
284. S. L. Campbell and K. D. Yeomans, *Behavior of the nonunique terms in general DAE integrators*, Appl. Numer. Math. **28** (1998), no. 2-4, 209–226.
285. Wei Ping Cao and Ji Pu Ma, *Perturbation of* A_0^+h, Numer. Math. J. Chinese Univ. (English Ser.) **3** (1994), no. 1, 96–103.
286. ———, *The pointwise continuity of the M-P generalized inverses* A_x^+, Acta Math. Sinica **40** (1997), no. 2, 287–295.
287. ———, *Moore-Penrose generalized inverses of closed operators*, Nanjing Daxue Xuebao Shuxue Bannian Kan **16** (1999), no. 1, 75–81.
288. S. R. Caradus, *An equational approach to products of relatively regular operators*, Aequationes Math. **15** (1977), no. 1, 55–62.
289. ———, *Generalized Inverses and Operator Theory*, Queen's University, Kingston, Ont., 1978.
290. D. Carlson, *Matrix decompositions involving the Schur complement*, SIAM J. Appl. Math. **28** (1975), 577–587.
291. ———, *What are Schur complements, anyway?*, Linear Algebra and its Applications **74** (1986), 257–275.
292. ———, *Generalized inverse invariance, partial orders, and rank-minimization problems for matrices*, Current trends in matrix theory (Auburn, Ala., 1986), North-Holland, New York, 1987, pp. 81–87.
293. D. Carlson, E. V. Haynsworth, and T. Markham, *A generalization of the Schur complement by means of the Moore–Penrose inverse*, SIAM J. Appl. Math. **26** (1974), 169–175.
294. R. Caron, H. J. Greenberg, and A. Holder, *Analytic centers and repelling inequalities*, Tech. Report CCM 142, Center for Computational Mathematics, University of Colorado at Denver, 1999, (to appear in European Journal of Operations Research).
295. N. Castro González, *On the convergence of semi-iterative methods to the Drazin inverse solution of linear equations in Banach spaces*, Collect. Math. **46** (1995), no. 3, 303–314.
296. N. Castro González and J. J. Koliha, *Semi-iterative methods for the Drazin inverse solution of linear equations in Banach spaces*, Numer. Funct. Anal. Optim. **20** (1999), no. 5-6, 405–418.
297. ———, *Perturbation of the Drazin inverse for closed linear operators*, Integral Equations Operator Theory **36** (2000), no. 1, 92–106.
298. N. Castro González, J. J. Koliha, and Yimin Wei, *Perturbation of the Drazin inverse for matrices with equal eigenprojections at zero*, Linear Algebra and its Applications **312** (2000), no. 1-3, 181–189.
299. Jian Miao Cen, *Fuzzy matrix partial orderings and generalized inverses*, Fuzzy Sets and Systems **105** (1999), no. 3, 453–458.
300. N. N. Chan, *On a downdating formula for regression*, J. Statist. Plann. Inference **46** (1995), no. 3, 347–350, (see [**442**]).
301. E. Chang, *The generalized inverse and interpolation theory*, In Campbell [**267**], pp. 196–219.
302. A. Charnes and W. W. Cooper, *Structural sensitivity analysis in linear programming and an exact product form left inverse*, Naval Res. Logist. Quart. **15** (1968), 517–522.
303. A. Charnes, W. W. Cooper, and G. L. Thompson, *Constrained generalized medians and hypermedians as deterministic equivalents for two–stage linear programs under uncertainty*, Management Sci. **12** (1965), 83–112.
304. A. Charnes and F. Granot, *Existence and representation of Diophantine and mixed Diophantine solutions to linear equations and inequalities*, Center for cybernetic studies, The University of Texas, Austin, TX, 1973.
305. A. Charnes and M. J. L. Kirby, *Modular design, generalized inverses and convex programming*, Operations Res. **13** (1965), 836–847.
306. Guoliang Chen, Musheng Wei, and Yifeng Xue, *Perturbation analysis of the least squares solution in Hilbert spaces*, Linear Algebra and its Applications **244** (1996), 69–80.
307. Guoliang Chen and Yimin Wei, *Perturbation analysis for the projection of a point onto an affine set in a Hilbert space*, Chinese Ann. Math. Ser. A **19** (1998), no. 4, 405–410, (translation in *Chinese J. Contemp. Math.* **19**(1998), 245–252).
308. Guoliang Chen and Yifeng Xue, *Perturbation analysis for the operator equation* $Tx = b$ *in Banach spaces*, J. Math. Anal. Appl. **212** (1997), no. 1, 107–125.
309. ———, *The expression of the generalized inverse of the perturbed operator under Type I perturbation in Hilbert spaces*, Linear Algebra and its Applications **285** (1998), no. 1-3, 1–6.

310. Han Fu Chen, *Two kinds of linear estimators in Hilbert spaces, and their connection*, Acta Math. Sinica **25** (1982), no. 6, 671–679.
311. X. Chen, M. Z. Nashed, and L. Qi, *Convergence of Newton's method for singular smooth and nonsmooth equations using adaptive outer inverses*, SIAM J. Optim. **7** (1997), 445–462.
312. Xuzhou Chen and R. E. Hartwig, *The group inverse of a triangular matrix*, Linear Algebra and its Applications **237/238** (1996), 97–108.
313. ———, *The hyperpower iteration revisited*, Linear Algebra Appl. **233** (1996), 207–229.
314. Yong-Lin Chen, *On the weighted projector and weighted generalized inverse matrices*, Acta Math. Appl. Sinica **6** (1983), no. 3, 282–291.
315. ———, *The generalized Bott-Duffin inverse and its applications*, Linear Algebra and its Applications **134** (1990), 71–91.
316. ———, *Representations and Cramer rules for the solution of a restricted matrix equation*, Linear and Multilinear Algebra **35** (1993), no. 3-4, 339–354.
317. ———, *Finite algorithms for the* (2)*-generalized inverse* $A^{(2)}_{T,S}$, Linear and Multilinear Algebra **40** (1995), no. 1, 61–68.
318. ———, *Iterative methods for solving restricted linear equations*, Appl. Math. Comput. **86** (1997), no. 2-3, 171–184.
319. ———, *Defining equations and explicit expressions for the generalized inverse* $A^{(2)}_{T,S}$, J. Nanjing Norm. Univ. Nat. Sci. Ed. **23** (2000), no. 2, 5–8.
320. Yong-Lin Chen and Xin Chen, *Representation and approximation of the outer inverse* $A^{(2)}_{T,S}$ *of a matrix* A, Linear Algebra and its Applications **308** (2000), no. 1-3, 85–107.
321. Yonghong Chen, S. J. Kirkland, and M. Neumann, *Group generalized inverses of* M*-matrices associated with periodic and nonperiodic Jacobi matrices*, Linear and Multilinear Algebra **39** (1995), no. 4, 325–340.
322. ———, *Nonnegative alternating circulants leading to* M*-matrix group inverses*, Linear Algebra and its Applications **233** (1996), 81–97.
323. Yonghong Chen and M. Neumann, M*-matrix generalized inverses of* M*-matrices*, Linear Algebra and its Applications **256** (1997), 263–285.
324. E. W. Cheney, *Introduction to Approximation Theory*, McGraw–Hill Book Co., New York, 1966.
325. H. Chernoff, *Locally optimal designs for estimating parameters*, Ann. Math. Statist. **24** (1953), 586–602.
326. J. S. Chipman, *On least squares with insufficient observations*, J. Amer. Statist. Assoc. **54** (1964), 1078–1111.
327. ———, *Specification problems in regression analysis*, In Boullion and Odell [**207**], pp. 114–176.
328. ———, *"Proofs" and proofs of the Eckart-Young theorem*, Stochastic Processes and Functional Analysis (Riverside, CA, 1994), Dekker, New York, 1997, pp. 71–83.
329. ———, *Linear restrictions, rank reduction, and biased estimation in linear regression*, Linear Algebra and its Applications **289** (1999), no. 1-3, 55–74.
330. J. S. Chipman and M. M. Rao, *On the treatment of linear restrictions in regression analysis*, Econometrica **32** (1964), 198–209.
331. ———, *Projections, generalized inverses and quadratic forms*, J. Math. Anal. Appl. **9** (1964), 1–11.
332. H. Chitwood, *Generalized Green's matrices for linear differential systems*, SIAM J. Math. Anal. **4** (1973), 104–110.
333. Han Hyuk Cho, *Regular fuzzy matrices and fuzzy equations*, Fuzzy Sets and Systems **105** (1999), no. 3, 445–451.
334. K. K. Choong and J. Y. Kim, *A numerical strategy for computing the stability boundaries for multiloading systems by using generalized inverse and continuation method*, Engineering Structures **23** (2001), 715–724.
335. Shui-Nee Chow and Yun Qiu Shen, *Bifurcations via singular value decompositions*, Appl. Math. Comput. **28** (1988), no. 3, part I, 231–245.
336. Ole Christensen, *Frames and pseudo-inverses*, J. Math. Anal. Appl. **195** (1995), no. 2, 401–414.
337. ———, *Operators with closed range, pseudo-inverses, and perturbation of frames for a subspace*, Canad. Math. Bull. **42** (1999), no. 1, 37–45.
338. ———, *Frames, Riesz bases, and discrete Gabor/wavelet expansions*, Bull. Amer. Math. Soc. **38** (2001), no. 3, 273–291.
339. Moody T. Chu, *On a numerical treatment for the curve-tracing of the homotopy method*, Numer. Math. **42** (1983), no. 3, 323–329.
340. Moody T. Chu, R. E. Funderlic, and G. H. Golub, *On a variational formulation of the generalized singular value decomposition*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 4, 1082–1092.
341. G. Ciecierska, *A note on another method of computing the Moore-Penrose inverse of a matrix*, Demonstratio Math. **31** (1998), no. 4, 879–886.
342. G. Cimmino, *Inversione delle corrispondenze funzionali lineari ed equazioni differenziali*, Rivista Mat. Univ. Parma **1** (1950), 105–116.
343. ———, *Cramer's rule without the notion of determinant*, Atti Accad. Sci. Istit. Bologna Cl. Sci. Fis. Rend. (14) **3** (1985/86), 115–138 (1987).
344. ———, *An unusual way of solving linear systems*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **80** (1986), no. 1-2, 6–7 (1987).
345. ———, *On some identities involving spherical means*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **83** (1989), 69–72 (1990).
346. J. F. Claerbout, *Geophysical Estimation by Example*, Stanford Exploration Project, Stanford University, 2001, (on–line book).
347. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. **40** (1936), 396–414.

348. J.-J. Climent, M. Neumann, and A. Sidi, *A semi-iterative method for real spectrum singular linear systems with an arbitrary index*, J. Comput. Appl. Math. **87** (1997), no. 1, 21–38.
349. J.-J. Climent, N. Thome, and Y. Wei, *A geometrical approach on generalized inverses by Neumann–type series*, Linear Algebra and its Applications **332/334** (2001), 535–542.
350. R. E. Cline, *Representations for the generalized inverse of a partitioned matrix*, J. Soc. Indust. Appl. Math. **12** (1964), 588–600.
351. ______, *Representations for the generalized inverse of sums of matrices*, J. Soc. Indust. Appl. Math. Ser. B. Numer. Anal. **2** (1965), 99–114.
352. ______, *Inverses of rank invariant powers of a matrix*, SIAM J. Appl. Math. **5** (1968), 182–197.
353. ______, *Elements of the theory of generalized inverses for matrices. (umap modules and monographs in undergraduate mathematics and its applications project). the umap expository monograph series)*, EDC/UMAP, Newton, Mass., 1979.
354. ______, *Note on an extension of the Moore-Penrose inverse*, Linear Algebra and its Applications **40** (1981), 19–23.
355. R. E. Cline and R. E. Funderlic, *The rank of a difference of matrices and associated generalized inverses*, Linear Algebra and its Applications **24** (1979), 185–215.
356. R. E. Cline and T. N. E. Greville, *An extension of the generalized inverse of a matrix*, SIAM J. Appl. Math. **19** (1970), 682–688.
357. ______, *A Drazin inverse for rectangular matrices*, Linear Algebra and its Applications **29** (1980), 53–62.
358. R. E. Cline and R. J. Plemmons, *l_2-solutions to underdetermined linear systems*, SIAM Rev. **18** (1976), no. 1, 92–106.
359. R. E. Cline and L. D. Pyle, *The generalized inverse in linear programming. an intersecton projection method and the solution of a class of structured linear programming problems*, SIAM J. Appl. Math. **24** (1973), 338–351.
360. E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw–Hill Book Co., 1955.
361. L. Collatz, *Aufgaben monotoner Art*, Arch. Math. **3** (1952), 366–376.
362. D. Constales, *A closed formula for the Moore-Penrose generalized inverse of a complex matrix of given rank*, Acta Math. Hungar. **80** (1998), no. 1-2, 83–88.
363. Corrado Corradi, *A note on the solution of separable nonlinear least-squares problems with separable nonlinear equality constraints*, SIAM J. Numer. Anal. **18** (1981), no. 6, 1134–1138.
364. ______, *Computing methods for restricted estimation in linear models*, Statistica (Bologna) **42** (1982), no. 1, 55–68, (See [**529**]).
365. R. W. Cottle, *Manifestations of the Schur complement*, Linear Algebra and Appl. **8** (1974), 189–211.
366. R. Courant and D. Hilbert, *Methods of Mathematical Physics. Vol. I*, Interscience Publishers, New York, 1953, (First published in German 1924).
367. D. E. Crabtree and E. V. Haynsworth, *An identity for the Schur complement of a matrix*, Proc. Amer. Math. Soc. **22** (1969), 364–366.
368. D. F. Cudia, *Rotundity*, Convexity, Proc, Sympos. Pure Math. Vol. VII (V. Klee, Editor), Amer. Math. Soc., Providence, R.I., 1963, pp. 73–97.
369. C. G. Cullen and K. J. Gale, *A functional definition of the determinant*, Amer. Math. Monthly **72** (1965), 403–406.
370. B. Cvetkov, *A new method of computation in the theory of least squares*, Austral. J. Appl. Sci. **6** (1955), 274–280.
371. Hua Dai, *An algorithm for symmetric generalized inverse eigenvalue problems*, Linear Algebra and its Applications **296** (1999), no. 1-3, 79–98.
372. J. F. Dalphin and V. Lovass-Nagy, *Best least squares solutions to finite difference equations using the generalized inverse and tensor product methods*, Journal of the ACM **20** (1973), no. 2, 279–289.
373. D. F. Davidenko, *On a new method of numerical solution of systems of nonlinear equations*, Doklady Akad. Nauk SSSR (N.S.) **88** (1953), 601–602.
374. ______, *The evaluation of determinants by the method of variation of parameters*, Soviet Math. Dokl. **1** (1960), 316–319.
375. ______, *Inversion of matrices by the method of variation of parameters*, Soviet Math. Dokl. **1** (1960), 279–282.
376. ______, *The method of variation of parameters as applied to the computation of eigenvalues and eigenvectors of matrices*, Soviet Math. Dokl. **1** (1960), 364–367.
377. ______, *Pseudo-inversion and construction of generalized solutions of linear equations that arise in calculations for nuclear reactors*, Computational processes and systems, No. 6 (Russian), “Nauka”, Moscow, 1988, pp. 98–109.
378. C. Davis, *Separation of two linear subspaces*, Acta Sci. Math. Szeged **19** (1958), 172–187.
379. ______, *Completing a matrix so as to minimize the rank*, Topics in operator theory and interpolation, Birkhäuser, Basel, 1988, pp. 87–95.
380. C. Davis and W. M. Kahan, *Some new bounds on perturbation of subspaces*, Bull. Amer. Math. Soc. **75** (1969), 863–868.
381. D. L. Davis and D. W. Robinson, *Generalized inverses of morphisms*, Linear Algebra and its Applications **5** (1972), 329–338.
382. Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl. **21** (2000), no. 4, 1253–1278 (electronic).

383. ———, *On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl. **21** (2000), no. 4, 1324–1342 (electronic).
384. Alvaro R. De Pierro and Musheng Wei, *Reverse order law for reflexive generalized inverses of products of matrices*, Linear Algebra and its Applications **277** (1998), 299–311.
385. O. Y. De Vel and E. V. Krishnamurthy, *An iterative pipelined array architecture for the generalized matrix inversion*, Information Processing Letters **26** (1988), no. 5, 263–267.
386. L. de Vito, *Sugli autovalori e sulle autosoluzioni di una classe di trasformazioni hermitiane*, Rend. Sem. Mat. Univ. Padova **25** (1956), 144–175.
387. P. Dean and J. Porrill, *Pseudo-inverse control in biological systems: a learning mechanism for fixation stability*, Neural Networks **11** (1998), 1205–1218.
388. Stanley R. Deans, *The Radon transform and some of its applications (revised reprint of the 1983 original)*, Robert E. Krieger Publishing Co. Inc., Malabar, FL, 1993.
389. H. P. Decell, Jr., *An alternate form of the generalized inverse of an arbitrary complex matrix*, SIAM Rev. **7** (1965), 356–358, (see [**1652**]).
390. ———, *An application of the Cayley-Hamilton theorem to generalized matrix inversion*, SIAM Rev. **7** (1965), 526–528, (extended in [**1520**]).
391. ———, *On the derivative of the generalized inverse of a matrix*, Linear and Multilinear Algebra **1** (1973/74), 357–359.
392. H. P. Decell, Jr. and P. L. Odell, *On the fixed point probability vector of regular or ergodic transition matrices*, J. Amer. Statist. Assoc. **62** (1967), 600–602.
393. J. P. Dedieu and M. Shub, *Newton's method for overdetermined systems of equations*, Math. of Comput. **69** (2000), 1099–1115.
394. J. B. Dennis, *Mathematical Programming and Electrical Networks*, MIT Press, Cambridge, Mass., 1959.
395. B. A. Dent and A. Newhouse, *Polynomials orthogonal over discrete domains*, SIAM Rev. **1** (1959), 55–59.
396. C. A. Desoer and B. H. Whalen, *A note on pseudoinverses*, J. Soc. Indust. Appl. Math. **11** (1963), 442–447.
397. M. Desplas, *Matrice pseudo-inverse de Moore-Penrose et variables duales généralisées en programmation mathématique*, RAIRO Rech. Opér. **26** (1992), no. 4, 313–360.
398. P. Deuflhard and G. Heindl, *Affine invariant convergence theorems for Newton's method and extensions to related methods*, SIAM J. Numer. Anal. **16** (1979), 1–10.
399. P. Deuflhard and W. Sautter, *On rank-deficient pseudoinverses*, Linear Algebra and its Applications **29** (1980), 91–111.
400. E. Deutsch, *Semi–inverses, reflexive semi–inverses, and pseudo–inverses of an arbitrary linear transformation*, Linear Algebra and its Applications **4** (1971), 313–322.
401. E. Deutsch and M. Neumann, *Derivatives of the Perron root at an essentially nonnegative matrix and the group inverse of an M-matrix*, J. Math. Anal. Appl. **102** (1984), no. 1, 1–29.
402. F. Deutsch, *The angle between subspaces of a Hilbert space*, Approximation theory, wavelets and applications (Maratea, 1994), Kluwer Acad. Publ., Dordrecht, 1995, pp. 107–130.
403. M. E. Díaz Lozano, *The class of $\{1, 2, 4\}$-inverses as the solution of a linear problem*, Bol. Soc. Mat. Mexicana (3) **4** (1998), no. 2, 223–228.
404. V. D. Didenko and B. Silbermann, *Extension of C^*-algebras and Moore-Penrose stability of sequences of additive operators*, Linear Algebra and its Applications **275/276** (1998), 121–140.
405. F. Diele, L. Lopez, and R. Peluso, *The Cayley transform in the numerical solution of unitary differential systems*, Adv. Comput. Math. **8** (1998), no. 4, 317–334.
406. I. I. Dikin, *On the speed of an iterative process*, Upravlyaemye Sistemi **12** (1974), 54–60, (see [**1496**]).
407. Jiu Ding, *Perturbation results for projecting a point onto a linear manifold*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 3, 696–700 (electronic).
408. ———, *Perturbation of systems of linear algebraic equations*, Linear and Multilinear Algebra **47** (2000), no. 2, 119–127.
409. ———, *Lower and upper bounds in the perturbation of general linear algebraic equations*, Appl. Math. Lett. **14** (2001), no. 1, 49–52.
410. Jiu Ding and Liang Jiao Huang, *On the perturbation of the least squares solutions in Hilbert spaces*, Linear Algebra and its Applications **212/213** (1994), 487–500.
411. ———, *Perturbation of generalized inverses of linear operators in Hilbert spaces*, J. Math. Anal. Appl. **198** (1996), no. 2, 506–515.
412. D. S. Djordjević and P. S. Stanimirović, *Universal iterative methods for computing generalized inverses*, Acta Math. Hungar. **79** (1998), no. 3, 253–268.
413. ———, *Applications of the Groetsch theorem*, Indian J. Pure Appl. Math. **31** (2000), no. 3, 277–286.
414. Dragomir Ž. Doković, *On the generalized inverse for matrices*, Glasnik Mat.-Fiz. Astronom. Ser. II Društvo Mat. Fiz. Hrvatske **20** (1965), 51–55.
415. J. Dommanget, *L'inverse d'un cracovien rectangulaire: Son emploi dans la résolution des systèmes d'équations linéaires*, Publ. Sci. Tech. Ministère de l'Air (Paris) Notes Tech. No. **128** (1963), 11–41.
416. J. Douglas, Jr. and C. M. Pearcy, *On convergence of alternating direction procedures in the presence of singular operators*, Numer. Math. **5** (1963), 175–184.

417. A. Doust and V. E. Price, *The latent roots and vectors of a singular matrix*, Comput. J. **7** (1964), 222–227.
418. A. Dragomir, *On the generalized inverse of a matrix*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **3** (1965), 123–128.
419. A. Dragomir and P. Dragomir, *Formulas of the Bjerhammar type for the generalized inverse of an arbitrary matrix*, Proceedings of the Conference on Algebra (Romanian) (Timişoara, 1986) (Timişoara), Univ. Timişoara, 1987, pp. 33–36.
420. A. Dragomir and M. Fildan, *L'inverse généralisé d'un opérateur linéaire*, An. Univ. Timişoara Ser. Şti. Mat. **7** (1969), 55–65.
421. P. Dragomir, *On the Greville-Moore formula for calculating the generalized inverse matrix*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **1** (1963), 115–119.
422. ———, *The generalized inverse of a bilinear form*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **2** (1964), 71–76.
423. M. P. Drazin, *Pseudo inverses in associative rings and semigroups*, Amer. Math. Monthly **65** (1958), 506–514.
424. ———, *Natural structures on semigroups with involution*, Bull. Amer. Math. Soc. **84** (1978), no. 1, 139–141.
425. ———, *Differentiation of generalized inverses*, In Campbell [**267**], pp. 138–144.
426. ———, *Extremal definitions of generalized inverses*, Linear Algebra and its Applications **165** (1992), 185–196.
427. Pierre Druilhet, *Optimality of neighbour balanced designs*, J. Statist. Plann. Inference **81** (1999), no. 1, 142–152.
428. H. Drygas, *On a generalization of the Farkas theorem*, Unternehmensforschung **13** (1969), 283–290.
429. ———, *The Coordinate–Free Approach to Gauss–Markov Estimation*, Springer–Verlag, Berlin, 1970.
430. ———, *Consistency of the least squares and Gauss–Markov estimators in regression models*, Z. Wahrscheinlichkeitstheorie u. verw. Gebiete **17** (1971), 309–326.
431. ———, *Estimation and prediction for linear models in general spaces (Vorträge auf der Ersten Sommerschule über Probleme der Modellwahl und Parameterschätzung in der Regressions-Analyse, Zinnowitz, 1974)*, Math. Operationsforsch. Statist. **6** (1975), no. 2, 301–324.
432. ———, *Gauss-Markov estimation for multivariate linear models with missing observations*, Ann. Statist. **4** (1976), no. 4, 779–787.
433. ———, *On the unified theory of least squares*, Probab. Math. Statist. **5** (1985), no. 2, 177–186.
434. H. Drygas and J. Srzednicka, *A new result on Hsu's model of regression analysis*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **24** (1976), no. 12, 1133–1136.
435. R. J. Duffin, *Network models*, In Wilf and Harary [**1594**], pp. 65–91.
436. R. J. Duffin and T. D. Morley, *Inequalities induced by network connections. II. Hybrid connections*, J. Math. Anal. Appl. **67** (1979), no. 1, 215–231.
437. ———, *Inequalities induced by network connections*, In Campbell [**267**], pp. 27–49.
438. R. J. Duffin and A. C. Schaeffer, *A class of nonharmonic Fourier series*, Trans. Amer. Math. Soc. **72** (1952), 341–366.
439. R. J. Duffin and G. E. Trapp, *Hybrid addition of matrices-network theory concept*, Applicable Anal. **2** (1972/73), 241–254.
440. J. W. Duke, *A note on EP linear transformations*, Linear Algebra and its Applications **3** (1970), 379–382.
441. N. Dunford and J. T. Schwartz, *Linear Operators. Part I*, Interscience, New York, 1957.
442. T. T. Dunne and M. Stone, *Downdating the Moore-Penrose generalized inverse for cross-validation of centred least squares prediction*, J. Roy. Statist. Soc. Ser. B **55** (1993), no. 2, 369–375.
443. Arthur M. DuPré and Seymour Kass, *Distance and parallelism between flats in* $\mathbb{R}^n$, Linear Algebra and its Applications **171** (1992), 99–107.
444. C. S. Duris, *Optimal quadrature formulas using generalized inverses. I. General theory and minimum variance formulas*, Math. Comp. **25** (1971), 495–504.
445. C. S. Duris and V. P. Sreedharan, *Chebyshev and* l^1*-solutions of linear equations using least squares solutions*, SIAM J. Numer. Anal. **5** (1968), 491–505.
446. P. S. Dwyer, *Some applications of matrix derivatives in multivariate analysis*, J. Amer. Statist. Assoc **62** (1967), 607–625.
447. P. S. Dwyer and M. S. Macphail, *Symbolic matrix derivatives*, Ann. Math. Statistics **19** (1948), 517–534.
448. N. Eagambaram, *Generalized inverses with nonnegative principal minors*, Linear Algebra and its Applications **111** (1988), 293–312.
449. ———, $(i, j, \cdots, k)$*-inverses via bordered matrices*, Sankhyā Ser. A **53** (1991), no. 3, 298–308.
450. C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika **1** (1936), 211–218.
451. ———, *A principal axis transformation for non-Hermitian matrices*, Bull. Amer. Math. Soc. **45** (1939), 118–121.
452. E. Egerváry, *On a property of the projector matrices and its application to the canonical representation of matrix functions*, Acta Sci. Math. Szeged **15** (1953), 1–6.
453. M. Eiermann, I. Marek, and W. Niethammer, *On the solution of singular linear systems of algebraic equations by semi-iterative methods*, Numer. Math. **53** (1988), no. 3, 265–283.

454. E. Eitelberg and H. Hanselmann, *Comments on: "On system realization by matrix generalized inverses" (Internat. J. Control* **26** *(1977), no. 5, 745–751) by V. Lovass-Nagy, R. J. Miller and D. L. Powers*, Internat. J. Control **27** (1978), no. 4, 651–652, (see [**976**]).
455. L. Eldén, *Perturbation theory for the least squares problem with linear equality constraints*, SIAM J. Numer. Anal. **17** (1980), 338–350.
456. ______, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT **22** (1983), 487–502.
457. W. W. Elliott, *Generalized Green's functions for compatible differential systems*, Amer. J. Math. **50** (1928), 243–258.
458. ______, *Green's functions for differential systems containing a parameter*, Amer. J. Math. **51** (1929), 397–416.
459. L. Elsner and Kh. D. Ikramov, *Normal matrices: an update*, Linear Algebra and its Applications **285** (1998), no. 1-3, 291–303, (see [**602**]).
460. C. Elster, *Recovering wavefronts from difference measurements in lateral shearing interferometry*, J. Comput. Appl. Math. **110** (1999), no. 1, 177–180.
461. H. W. Engl and C. W. Groetsch, *A higher order approximation technique for restricted linear least-squares problems*, Bull. Austral. Math. Soc. **37** (1988), no. 1, 121–130.
462. H. W. Engl and R. Kress, *A singular perturbation problem for linear operators with an application to electrostatic and magnetostatic boundary and transmission problems*, Math. Methods Appl. Sci. **3** (1981), no. 2, 249–274.
463. H. W. Engl and M. Z. Nashed, *New extremal characterizations of generalized inverses of linear operators*, J. Math. Anal. Appl. **82** (1981), no. 2, 566–586.
464. H. W. Engl and A. Neubauer, *On projection methods for solving linear ill-posed problems*, Model optimization in exploration geophysics (Berlin, 1986), Vieweg, Braunschweig, 1987, pp. 73–92.
465. M. J. Englefield, *The commuting inverses of a square matrix*, Proc. Cambridge Philos. Soc. **62** (1966), 667–671.
466. P. J. Erdelsky, *Projections in a normed linear space and a generalization of the pseudo–inverse*, Doctoral dissertation in mathematics, California Inst. Tech., Pasadena, CA, 1969.
467. I. Erdélyi, *On partial isometries in finite dimensional Euclidean spaces*, SIAM J. Appl. Math. **14** (1966), 453–467.
468. ______, *On the "reversed order law" related to the generalized inverse of matrix products*, J. Assoc. Comput. Mach. **13** (1966), 439–443.
469. ______, *On the matrix equation $Ax = \lambda Bx$*, J. Math. Anal. Appl. **17** (1967), 119–132.
470. ______, *The quasi–commuting inverse for a square matrix*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **42** (1967), 626–633.
471. ______, *Normal partial isometries closed under multiplication on unitary spaces*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **43** (1968), 186–190.
472. ______, *Partial isometries and generalized inverses*, In Boullion and Odell [**207**], pp. 203–217.
473. ______, *Partial isometries closed under multiplication on Hilbert spaces*, J. Math. Anal. Appl. **22** (1968), 546–551.
474. ______, *Partial isometries defined by a spectral property on unitary spaces*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII **44** (1968), 741–747.
475. ______, *A generalized inverse for arbitrary operators between hilbert spaces*, Proceedings of the Cambridge Philosophical Society **71** (1972), 43–50.
476. ______, *Spectral decompositions for generalized inversions*, In Campbell [**267**], pp. 261–274.
477. I. Erdélyi and A. Ben-Israel, *Extremal solutions of linear equations and generalized inversion between Hilbert spaces*, J. Math. Anal. Appl. **39** (1972), 298–313.
478. I. Erdélyi and F. R. Miller, *Decomposition theorems for partial isometries*, J. Math. Anal. Appl. **30** (1970), 665–679.
479. C. A. Eschenbach, F. J. Hall, and Zhongshan Li, *Sign pattern matrices and generalized inverses*, Linear Algebra and its Applications **211** (1994), 53–66.
480. D. J. Evans, Wenyu Sun, R. J. B. de Sampaio, and J. Y. Yuan, *The restricted generalized inverses corresponding to constrained quadratic system*, Int. J. Comput. Math. **62** (1996), no. 3-4, 285–296.
481. Ky Fan and A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc. **6** (1955), 111–116.
482. R. W. Farebrother, *Further results on the mean square error of ridge regression*, J. Roy. Statist. Soc. Ser. B **38** (1976), no. 3, 248–250.
483. ______, *An historical note on recursive residuals*, J. Roy. Statist. Soc. Ser. B **40** (1978), no. 3, 373–375.
484. ______, *A class of statistical estimators related to principal components*, Linear Algebra and its Applications **289** (1999), no. 1-3, 121–126.
485. R. Featherstone and O. Khatib, *Load independence of the dynamically consistent inverse of the jacobian matrix. source*, International Journal of Robotics Research **16** (1997), 168–170.
486. C. A. Felippa, K. C. Park, and M. R. Justino Filho, *The construction of free-free flexibility matrices as generalized stiffness inverses*, Computers & Structures **68** (1998), 411–418.
487. I. S. Fenyő, *A representation of the generalized inverse in Hilbert spaces*, Rend. Sem. Mat. Fis. Milano **50** (1980), 23–29 (1982).
488. M. Ferrante and P. Vidoni, *A Gaussian-generalized inverse Gaussian finite-dimensional filter*, Stochastic Process. Appl. **84** (1999), no. 1, 165–176.

489. M. Fiedler, *Moore–Penrose involutions in the classes of Laplacians and simplices*, Linear and Multilinear Algebra **39** (1995), no. 1-2, 171–178.
490. M. Fiedler and T. L. Markham, *Some connections between the Drazin inverse, P-matrices, and the closure of inverse M-matrices*, Linear Algebra and its Applications **132** (1990), 163–172.
491. ———, *A characterization of the Moore-Penrose inverse*, Linear Algebra and its Applications **179** (1993), 129–133.
492. J. A. Fill and D. E. Fishkind, *The Moore-Penrose generalized inverse for sums of matrices*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 2, 629–635 (electronic).
493. P. A. Filmore and J. P. Williams, *On operator ranges*, Advances in Math. **7** (1971), 254–281.
494. A. G. Fisher, *On construction and properties of the generalized inverse*, SIAM J. Appl. Math. **15** (1967), 269–272.
495. R. Fletcher, *Generalized inverse methods for the best least squares solution of systems of non-linear equations*, Comput. J. **10** (1968), 392–399.
496. ———, *A technique for orthogonalization*, J. Inst. Math. Appl. **5** (1969), 162–166.
497. ———, *Generalized inverses for nonlinear equations and optimization*, pp. 75–85, In [**1226**], 1970.
498. Delia Flores de Chela, *Generalized inverses on normed vector spaces*, Linear Algebra and its Applications **26** (1979), 243–263.
499. ———, *Approximations in the l_∞ norm and the generalized inverse*, Linear Algebra and its Applications **42** (1982), 3–21.
500. ———, *Generalized inverses under the l_1-norm*, Linear Algebra and its Applications **94** (1987), 237–261.
501. N. Florsch and J. Hinderer, *Bayesian estimation of the free core nutation parameters from the analysis of precise tidal gravity data*, Physics of The Earth and Planetary Interiors **117** (2000), 21–35.
502. A. Forsgren, *On linear least-squares problems with diagonally dominant weight matrices*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 4, 763–788.
503. A. Forsgren and G. Sporre, *On weighted linear least-squares problems related to interior methods for convex quadratic programming*, SIAM J. Matrix Anal. Appl. **23** (2001), 42–56.
504. G. E. Forsythe, *The maximum and minimum of a positive definite quadratic polynomial on a sphere are convex functions of the radius*, SIAM J. Appl. Math. **19** (1970), 551–554.
505. G. E. Forsythe and G. H. Golub, *On the stationary values of a second–degree polynomial on the unit sphere*, SIAM J. Appl. Math. **13** (1965), 1050–1068.
506. D. J. Foulis, *Baer *-semigroups*, Proc. Amer. Math. Soc. **11** (1960), 648–654.
507. ———, *Relative inverses in Baer*–semigroups*, Michigan Math. J. **10** (1963), 65–84.
508. J. S. Frame, *Matrix functions and applications. I. matrix operations and generalized inverses*, IEEE Spectrum **1** (1964), 209–220.
509. ———, *Matrix functions and applications. II. Functions of a matrix*, IEEE Spectrum **1** (1964), no. 4, 102–108.
510. ———, *Matrix functions and applications. IV. Matrix functions and constituent matrices*, IEEE Spectrum **1** (1964), no. 6, 123–131.
511. ———, *Matrix functions and applications. V. Similarity reductions by rational or orthogonal matrices*, IEEE Spectrum **1** (1964), no. 7, 103–109.
512. J. S. Frame and H. E. Koenig, *Matrix functions and applications. III. Applications of matrices to systems analysis*, IEEE Spectrum **1** (1964), no. 5, 100–109.
513. C. Franchetti and E. W. Cheney, *Orthogonal projections in spaces of continuous functions*, J. Math. Anal. Appl. **63** (1978), no. 1, 253–264.
514. P. Franck, *Sur la distance minimale d'une matrice régulière donnée au lieu des matrices singulières*, Deux. Congr. Assoc. Franc. Calcul. et Trait. Inform. Paris 1961, Gauthiers–Villars, Paris, 1962, pp. 55–60.
515. I. Fredholm, *Sur une classe d'équations fonctionnelles*, Acta Math. **27** (1903), 365–390.
516. A. Friedlander, J. M. Martínez, and H. D. Scolnik, *Generalized inverses and a new stable secant type minimization algorithm*, Optimization techniques (Proc. 8th IFIP Conf., Würzburg, 1977), Part 2, Springer, Berlin, 1978, pp. 136–146. Lecture Notes in Control and Informat. Sci., Vol. 7.
517. J. D. Fulton, *Generalized inverses of matrices over a finite field*, Discrete Math. **21** (1978), no. 1, 23–29.
518. ———, *Generalized inverses of matrices over fields of characteristic two*, Linear Algebra and its Applications **28** (1979), 69–76.
519. R. E. Funderlic and C. D. Meyer, Jr., *Sensitivity of the stationary distribution vector for an ergodic Markov chain*, Linear Algebra and its Applications **76** (1986), 1–17.
520. R. Gabriel, *Extensions of generalized algebraic complement to arbitrary matrices (romanian)*, Stud. Cerc. Mat. **17** (1965), 1567–1581.
521. ———, *Das verallgemeinerte Inverse einer Matrix deren Elemente einem beliebigen Körper angehören*, J. Reine Angew. Math. **234** (1969), 107–122.
522. ———, *Das verallgemeinerte Inverse einer Matrix über einem beliebigen Körperanalytisch betrachtet*, J. Reine Angew. Math. **244** (1970), 83–93.
523. ———, *Das verallgemeinerte Inverse einer Matrix über einen beliebigen Körper—mit Skelettzerlegungen berechnet*, Rev. Roumaine Math. Pures Appl. **20** (1975), 213–225.

524. ———, *Das verallgemeinerte Inverse in Algebren*, Rev. Roumaine Math. Pures Appl. **20** (1975), 311–324, (corrigendum: Rev. Roumaine Math. Pures Appl. **20**(1975), 747).
525. R. Gabriel and R. E. Hartwig, *The Drazin inverse as a gradient*, Linear Algebra and its Applications **63** (1984), 237–252.
526. J. Gaches, J.-L. Rigal, and X. Rousset de Pina, *Distance euclidienne d'une application linéaire σ au lieu des applications de rang r donné*, C. R. Acad. Sci. Paris **260** (1965), 5672–5674.
527. A. Galántai, *The theory of Newton's method*, J. Comput. Appl. Math. **124** (2000), no. 1-2, 25–44, (Numerical analysis 2000, Vol. IV, Optimization and nonlinear equations).
528. A. Galántai and G. Varga, *A relaxation method for the computation of generalized inverses of matrices*, Közlemények—MTA Számitástechn. Automat. Kutato Int. Budapest (1976), no. 17, 57–62.
529. A. R. Gallant and T. M. Gerig, *Computations for constrained linear models*, J. Econometrics **12** (1980), no. 1, 59–84, (See [**364**]).
530. A. M. Galperin and Z. Waksman, *On pseudoinverses of operator products*, Linear Algebra and its Applications **33** (1980), 123–131.
531. ———, *Ulm's method under regular smoothness*, Numer. Funct. Anal. Optim. **19** (1998), no. 3-4, 285–307.
532. W. Gander, *Algorithms for the polar decomposition*, SIAM J. Sci. Statist. Comput. **11** (1990), no. 6, 1102–1115.
533. F. R. Gantmacher, *The Theory of Matrices*, vol. I and II, Chelsea, New York, 1959.
534. J. M. Garnett III, A. Ben-Israel, and S. S. Yau, *A hyperpower iterative method for computing matrix products involving the generalized inverse*, SIAM J. Numer. Anal. **8** (1971), 104–109.
535. M. K. Gavurin and Ju. B. Farforovskaja, *An iterative method for finding the minimum of sums of squares*, Ž. Vyčisl. Mat. i Mat. Fiz. **6** (1966), 1094–1097.
536. D. M. Gay, *Modifying singular values: existence of solutions to sytems of nonlinear equations having a possibly singular Jacobian matrix*, Math. Comp. **31** (1977), no. 140, 962–973.
537. ———, *Corrigenda: "Modifying singular values: existence of solutions to systems of nonlinear equations having a possibly singular Jacobian matrix" (Math. Comp.* **31** *(1977), no. 140, 962–973)*, Math. Comp. **33** (1979), no. 145, 432–433.
538. T. M. Gerig and A. R. Gallant, *Computing methods for linear models subject to linear parametric constraints*, J. Statist. Comput. Simulation **3** (1975), 283–296, (Errata: ibid 4 (1975), no. 1, 81–82).
539. B. Germain-Bonne, *Calcul de pseudo–inverses*, Rev. Francaise Informat. Recherce Opérationelle **3** (1969), 3–14.
540. A. J. Getson and F. C. Hsuan, *{2}-inverses and their statistical application*, Springer-Verlag, New York, 1988.
541. C. Giurescu and R. Gabriel, *Some properties of the generalized matrix inverse and semiinverse*, An. Univ. Timişoara Ser. Şti. Mat.-Fiz. No. **2** (1964), 103–111.
542. I. M. Glazman and Ju. I. Ljubich, *Finite Dimensional Linear Analysis*, Nauka, Moscow, 1969, (English translation published by MIT Press).
543. S. Goldberg, *Unbounded Linear Operators*, McGraw-Hill Book Co., New York, 1966.
544. A. J. Goldman and M. Zelen, *Weak generalized inverses and minimum variance linear unbiased estimation*, J. Res. Nat. Bur. Standards Sect. B **68B** (1964), 151–172.
545. A. A. Goldstein, *Constructive Real Analysis*, Harper and Row, New York, 1967.
546. G. R. Goldstein and J. A. Goldstein, *The best generalized inverse*, J. Math. Anal. Appl. **252** (2000), no. 1, 91–101.
547. H. Goller, *Shorted operators and rank decomposition matrices*, Linear Algebra and its Applications **81** (1986), 207–236.
548. G. H. Golub, *Numerical methods for solving linear least squares problems*, Numer. Math. **7** (1965), 206–216.
549. ———, *Least squares, singular values and matrix approximations*, Aplikace Mathematiky **13** (1968), 44–51.
550. ———, *Matrix decompositions and statistical calculations*, Tech. Report STAN-CS-124, Stanford University, Stanford, March 1969.
551. G. H. Golub, M. T. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), 215–223.
552. G. H. Golub, A. Hoffman, and G. W. Stewart, *A generalization of the Eckart–Young matrix approximation theorem*, Linear Algebra and Its Applications **88/89** (1987), 317–327.
553. G. H. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, J. Soc. Indust. Appl. Math. Ser. B Numer. Anal. **2** (1965), 205–224.
554. G. H. Golub and C. D. Meyer, Jr., *Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains*, SIAM J. Algebraic Discrete Methods **7** (1986), no. 2, 273–281.
555. G. H. Golub and V. Pereyra, *The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate*, SIAM J. Numer. Anal. **10** (1973), 413–432.
556. ———, *Differentiation of pseudoinverses, separable nonlinear least squares problems and other tales*, In Nashed [**1116**], pp. 303–324.
557. G. H. Golub and C. Reinsch, *Singular value decompositions and least squares solutions*, Numer.

Math. **14** (1970), 403–420, (republished, pp. 134–151 in [**1598**]).

558. G. H. Golub and G. P. H. Styan, *Numerical computations for univariate linear models*, Journal of Statistical Computations and Simulation **2** (1973), 253–274.
559. G. H. Golub and J. H. Wilkinson, *Note on the iterative refinement of least squares solutions*, Numer. Math. **9** (1966), 139–148.
560. Xiangyang Gong, Wanyi Chen, and Fengsheng Tu, *The stability and design of nonlinear neural networks*, Comput. Math. Appl. **35** (1998), no. 8, 1–7.
561. C. C. Gonzaga and H. J. Lara, *A note on properties of condition numbers*, Linear Algebra and its Applications **261** (1997), 269–273, (see [**767**]).
562. I. J. Good, *Some applications of the singular decomposition of a matrix*, Technometrics **11** (1969), 823–831.
563. ______, *Generalized determinants and generalized generalized variance*, J. Statist. Comput. Simulation **12** (1980/81), no. 3-4, 311–315, (see [**122**]).
564. M. C. Gouveia, *Generalized invertibility of Hankel and Toeplitz matrices*, Linear Algebra and its Applications **193** (1993), 95–106.
565. ______, *Group and Moore-Penrose invertibility of Bezoutians*, Linear Algebra and its Applications **197/198** (1994), 495–509.
566. M. C. Gouveia and R. Puystjens, *About the group inverse and Moore-Penrose inverse of a product*, Linear Algebra and its Applications **150** (1991), 361–369.
567. P. R. Graves-Morris, D. E. Roberts, and A. Salam, *The epsilon algorithm and related topics*, J. Comput. Appl. Math. **122** (2000), no. 1-2, 51–80.
568. F. A. Graybill, *An Introduction to Linear Statistical Models. Vol I*, McGraw-Hill Book Co., Inc., New York, 1961.
569. ______, *Theory and application of the linear model*, Duxbury Press, North Scituate, Mass., 1976.
570. ______, *Matrices with applications in statistics*, second ed., Wadsworth Advanced Books and Software, Belmont, Calif., 1983.
571. F. A. Graybill and G. Marsaglia, *Idempotent matrices and quadratic forms in the general linear hypothesis*, Ann. Math. Statist. **28** (1957), 678–686.
572. F. A. Graybill, C. D. Meyer, Jr., and R. J. Painter, *Note on the computation of the generalized inverse of a matrix*, SIAM Rev. **8** (1966), 522–524.
573. B. Green, *The orthogonal approximation of an oblique structure in factor analysis*, Psychometrika **17** (1952), 429–440.
574. W. L. Green and T. D. Morley, *Operator means and matrix functions*, Linear Algebra and its Applications **137/138** (1990), 453–465.
575. ______, *Operator means, norm convergence and matrix functions*, Signal processing, scattering and operator theory, and numerical methods (Amsterdam, 1989), Birkhäuser Boston, Boston, MA, 1990, pp. 551–556.
576. W. Greub and W. C. Rheinboldt, *On a generalization of an inequality of L. V. Kantorovich*, Proc. Amer. Math. Soc. **10** (1959), 407–415.
577. ______, *Non self–adjoint boundary value problems in ordinary differential equations*, J. Res. Nat. Bur. Standards Sect. B **1960** (64B), 83–90.
578. T. N. E. Greville, *On smoothing a finite table: A matrix approach*, J. Soc. Indust. Appl. Math. **5** (1957), 137–154.
579. ______, *The pseudoinverse of a rectangular matrix and its application to the solution of systems of linear equations*, SIAM Rev. **1** (1959), 38–43.
580. ______, *Some applications of the pseudoinverse of a matrix*, SIAM Rev. **2** (1960), 15–22.
581. ______, *Note on fitting functions of several independent variables*, J. Soc. Indust. Appl. Math. **9** (1961), 109–115, (Erratum, *ibid* **9**(1961), 317).
582. ______, *Note on the generalized inverse of a matrix product*, J. Soc. Indust. Appl. Math. **9** (1966), 109–115.
583. ______, *Spectral generalized inverses of square matrices*, Math. Research Center Technical Summary Report 823, University of Wisconsin, Madison, WI, October 1967.
584. ______, *Some new generalized inverses with spectral properties*, In Boullion and Odell [**207**], pp. 26–46.
585. ______, *The Souriau-Frame algorithm and the Drazin pseudoinverse*, Linear Algebra and Appl. **6** (1973), 205–208.
586. ______, *Solutions of the matrix equation $XAX = X$ and relations between oblique and orthogonal projectors*, SIAM J. Appl. Math. **26** (1974), 828–832.
587. T. N. E. Greville and N. Keyfitz, *Backward population projection by a generalized inverse*, Computational probability (Proc. Actuarial Res. Conf., Brown Univ., Providence, R.I., 1975), Academic Press, New York, 1980, pp. 173–183.
588. E. Griepentrog and R. März, *Basic properties of some differential-algebraic equations*, Z. Anal. Anwendungen **8** (1989), no. 1, 25–41.
589. C. W. Groetsch, *Steepest descent and least squares solvability*, Canad. Math. Bull. **17** (1974), 275–276.
590. ______, *A product integral representation of the generalized inverse*, Comment. Math. Univ. Carolinae **16** (1975), 13–20.
591. ______, *Generalized Inverses of Linear Operators: Representation and Approximation. monographs and textbooks in pure and applied mathematics, no. 37*, Marcel Dekker Inc., New York, 1977.
592. ______, *The Forsythe-Motzkin method for singular linear operator equations*, J. Optim. Theory Appl. **25** (1978), no. 2, 311–315.
593. ______, *On rates of convergence for approximations to the generalized inverse*, Numer. Funct. Anal. Optim. **1** (1979), no. 2, 195–201.

594. ———, *Generalized inverses and generalized splines*, Numer. Funct. Anal. Optim. **2** (1980), no. 1, 93–97.
595. ———, *Spectral methods for linear inverse problems with unbounded operators*, J. Approx. Theory **70** (1992), no. 1, 16–28.
596. ———, *Inclusions for the Moore-Penrose inverse with applications to computational methods*, Contributions in Numerical Mathematics, World Sci. Publishing, River Edge, NJ, 1993, pp. 203–211.
597. ———, *Inverse Problems in the Mathematical Sciences*, Friedr. Vieweg & Sohn, Braunschweig, 1993.
598. ———, *Inclusions and identities for the Moore-Penrose inverse of a closed linear operator*, Math. Nachr. **171** (1995), 157–164.
599. C. W. Groetsch and B. J. Jacobs, *Iterative methods for generalized inverses based on functional interpolation*, In Campbell [**267**], pp. 220–232.
600. C. W. Groetsch and J. T. King, *Extrapolation and the method of regularization for generalized inverses*, J. Approx. Theory **25** (1979), no. 3, 233–247.
601. R. Grone, *Certain isometries of rectangular complex matrices*, Linear Algebra and its Applications **29** (1980), 161–171.
602. R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, *Normal matrices*, Linear Algebra and its Applications **87** (1987), 213–225, (characterizations of normal matrices, continued in [**459**]).
603. J. Groß, *A note on a partial ordering in the set of Hermitian matrices*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 4, 887–892.
604. ———, *Some remarks concerning the reverse order law*, Discuss. Math. Algebra Stochastic Methods **17** (1997), no. 2, 135–141.
605. ———, *Special generalized inverse matrices connected with the theory of unified least squares*, Linear Algebra and its Applications **264** (1997), 325–327.
606. ———, *More on concavity of a matrix function*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 2, 365–368 (electronic).
607. J. Gross, *On contractions in linear regression*, J. Statist. Plann. Inference **74** (1998), no. 2, 343–351.
608. J. Groß, *Idempotency of the Hermitian part of a complex matrix*, Linear Algebra and its Applications **289** (1999), no. 1-3, 135–139.
609. ———, *On oblique projection, rank additivity and the Moore-Penrose inverse of the sum of two matrices*, Linear and Multilinear Algebra **46** (1999), no. 4, 265–275.
610. ———, *On the product of orthogonal projectors*, Linear Algebra and its Applications **289** (1999), no. 1-3, 141–150.
611. ———, *Solution to a rank equation*, Linear Algebra and its Applications **289** (1999), no. 1-3, 127–130.
612. ———, *The Moore-Penrose inverse of a partitioned nonnegative definite matrix*, Linear Algebra and its Applications **321** (2000), no. 1-3, 113–121.
613. ———, *Nonnegative-definite and positive-definite solutions to the matrix equation $AXA^* = B$—revisited*, Linear Algebra and its Applications **321** (2000), no. 1-3, 123–129.
614. ———, *Löwner partial ordering and space preordering of Hermitian non-negative definite matrices*, Linear Algebra and its Applications **326** (2001), no. 1-3, 215–223.
615. J. Groß, J. Hauke, and A. Markiewicz, *Partial orderings, preorderings, and the polar decomposition of matrices*, Linear Algebra and its Applications **289** (1999), no. 1-3, 161–168.
616. J. Groß and S. Puntanen, *Estimation under a general partitioned linear model*, Linear Algebra and its Applications **321** (2000), no. 1-3, 131–144.
617. J. Groß and G. Trenkler, *On the least squares distance between affine subspaces*, Linear Algebra and its Applications **237/238** (1996), 269–276.
618. ———, *Generalized and hypergeneralized projectors*, Linear Algebra and its Applications **264** (1997), 463–474.
619. ———, *On the equality of usual and Amemiya's partially generalized least squares estimator*, Comm. Statist. Theory Methods **26** (1997), no. 9, 2075–2086.
620. ———, *Restrictions and projections in linear regression*, Internat. J. Math. Ed. Sci. Tech. **28** (1997), no. 3, 465–468.
621. ———, *On the product of oblique projectors*, Linear and Multilinear Algebra **44** (1998), no. 3, 247–259.
622. ———, *Nonsingularity of the difference of two oblique projectors*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 2, 390–395 (electronic).
623. J. Groß, G. Trenkler, and S. O. Troschke, *On semi-orthogonality and a special class of matrices*, Linear Algebra and its Applications **289** (1999), no. 1-3, 169–182.
624. J. Groß and S. O. Troschke, *Some remarks on partial orderings of nonnegative definite matrices*, Linear Algebra and its Applications **264** (1997), 457–461.
625. Marek Gruszczyński, *The Moore-Penrose matrix inversion and estimation under additional constraints*, Przegląd Statyst. **28** (1981), no. 3-4, 311–319 (1982).
626. Chuanqing Gu, *Generalized inverse matrix Padé approximation on the basis of scalar products*, Linear Algebra and its Applications **322** (2001), no. 1-3, 141–167.
627. S. P. Gudder and M. Neumann, *Splittings and iterative methods for approximate solutions to singular operator equations in Hilbert spaces*, J. Math. Anal. Appl. **62** (1978), no. 2, 272–294.

628. P. G. Guest, *Orthogonal polynomials in the least squares fitting of observations*, Philos. Mag. (7) **41** (1950), 124–137.
629. E. A. Guillemin, *Theory of linear physical systems*, Wiley, New York, 1963.
630. M. E. Gulliksson, *On modified Gram–Schmidt for weighted and constrained linear least squares*, BIT **35** (1995), 458–473.
631. M. E. Gulliksson, P.-Å. Wedin, and Yimin Wei, *Perturbation identities for regularized Tikhonov inverses and weighted pseudoinverses*, BIT **40** (2000), no. 3, 513–523.
632. N. N. Gupta, *An iterative method for computation of generalized inverse and matrix rank*, IEEE Trans. Systems Man Cybernet. (1971), 89–90.
633. ———, *An optimum iterative method for the computation of matrix rank*, IEEE Trans. Systems Man Cybernet. (1972), 437–438.
634. S. J. Haberman, *How much do Gauss-Markov and least square estimates differ? A coordinate-free approach*, Ann. Statist. **3** (1975), no. 4, 982–990, (extension of [**885**].
635. F. J. Hall, *Generalized inverses of a bordered matrix of operators*, SIAM J. Appl. Math. **29** (1975), 152–163.
636. ———, *On the independence of blocks of generalized inverses of bordered matrices*, Linear Algebra and Appl. **14** (1976), no. 1, 53–61.
637. F. J. Hall and R. E. Hartwig, *Further results on generalized inverses of partitioned matrices*, SIAM J. Appl. Math. **30** (1976), no. 4, 617–624.
638. F. J. Hall, R. E. Hartwig, I. J. Katz, and M. Newman, *Pseudosimilarity and partial unit regularity*, Czechoslovak Math. J. **33(108)** (1983), no. 3, 361–372.
639. F. J. Hall and I. J. Katz, *On ranks of integral generalized inverses of integral matrices*, Linear and Multilinear Algebra **7** (1979), no. 1, 73–85.
640. ———, *More on integral generalized inverses of integral matrices*, Linear and Multilinear Algebra **9** (1980), no. 3, 201–209.
641. ———, *Nonnegative integral generalized inverses*, Linear Algebra and its Applications **39** (1981), 23–39.
642. F. J. Hall and C. D. Meyer, Jr., *Generalized inverses of the fundamental bordered matrix used in linear estimation*, Sankhyā Ser. A **37** (1975), no. 3, 428–438, (corrigendum in *Sankhyā Ser. A* **40**(1980), 399).
643. C. R. Hallum, T. L. Boullion, and P. L. Odell, *Best linear estimation in the restricted general linear model*, Indust. Math. **34** (1984), no. 1, 53–64.
644. C. R. Hallum, T. O. Lewis, and T. L. Boullion, *Estimation in the restricted general linear model with a positive semidefinite covariance matrix*, Comm. Statist. **1** (1973), 157–166.
645. P. R. Halmos, *Finite–Dimensional Vector Spaces*, 2nd ed., D. Van Nostrand, Co., Princeton, 1958.
646. ———, *A Hilbert Space Problem Book*, D. Van Nostrand, Co., Princeton, 1967.
647. P. R. Halmos and J. E. McLaughlin, *Partial isometries*, Pacific J. Math. **13** (1963), 585–596.
648. P. R. Halmos and L. J. Wallen, *Powers of partial isometries*, J. Math. Mech. **19** (1970), 657–663.
649. I. Halperin, *Closures and adjoints of linear differential operators*, Ann. of Math. (1937), 880–919.
650. H. Hamburger, *Non–symmetric operators in Hilbert space*, Proceedings Symposium on Spectral Theory and Differential Problems, Oklahoma A& M College, Stillwater, OK, 1951, pp. 67–112.
651. M. Hanke and M. Neumann, *Preconditionings and splittings for rectangular systems*, Numerische Mathematik **57** (1990), no. 1, 85–95.
652. ———, *The geometry of the set of scaled projections*, Linear Algebra and its Applications **190** (1993), 137–148.
653. G. W. Hansen and D. W. Robinson, *On the existence of generalized inverses*, Linear Algebra and its Applications **8** (1974), 95–104.
654. P. C. Hansen, *The truncated SVD as a method for regularization*, BIT **27** (1987), 534–553.
655. R. J. Hanson, *A numerical method for solving Fredholm integral equations of the first kind using singular values*, SIAM J. Numer. Anal. **8** (1971), 616–622.
656. R. J. Hanson and M. J. Norris, *Analysis of measurements based on the singular value decomposition*, SIAM J. Sci. Statist. Comput. **2** (1981), no. 3, 363–373.
657. B. Harris, *Theory of probability*, Addison-Wesley, Reading, Mass., 1966.
658. W. A. Harris, Jr. and T. N. Helvig, *Applications of the pseudoinverse to modeling*, Technometrics **8** (1966), 351–357.
659. R. Harte, *Polar decomposition and the Moore-Penrose inverse*, Panamer. Math. J. **2** (1992), no. 4, 71–76.
660. W. M. Hartmann and R. E. Hartwig, *Computing the Moore-Penrose inverse for the covariance matrix in constrained nonlinear estimation*, SIAM J. Optim. **6** (1996), no. 3, 727–747.
661. J. Hartung, *On a method for computing pseudoinverses*, Optimization and operations research (Proc. Conf., Oberwolfach, 1975). Lecture Notes in Econom. Math. Systems, Vol. 117, Springer, Berlin, 1976, pp. 115–125.
662. ———, *Zur Darstellung pseudoinverser Operatoren*, Arch. Math. (Basel) **28** (1977), no. 2, 200–208.
663. ———, *A note on restricted pseudoinverses*, SIAM J. Math. Anal. **10** (1979), no. 2, 266–273.
664. J. Hartung and H. -J. Werner, *Zur Verwendung der restringierten Moore-Penrose-Inversen beim Testen von linearen Hypothesen*, Z. Angew. Math. Mech. **60** (1980), no. 7, T344–T346.
665. ———, *Hypothesenprüfung im restringierten linearen Modell. theorie und anwendungen (with*

english and french summaries), Vandenhoeck & Ruprecht, Göttingen, 1984.

666. R. E. Hartwig, 1-2 *inverses and the invariance of* BA^+C, Linear Algebra and its Applications **11** (1975), no. 3, 271–275.
667. ______, $AX - XB = C$*, resultants and generalized inverses*, SIAM J. Appl. Math. **28** (1975), 154–183.
668. ______, *Block generalized inverses*, Arch. Rational Mech. Anal. **61** (1976), no. 3, 197–251.
669. ______, *More on the Souriau-Frame algorithm and the Drazin inverse*, SIAM J. Appl. Math. **31** (1976), no. 1, 42–46.
670. ______, *Rank factorization and Moore-Penrose inversion*, Indust. Math. **26** (1976), no. 1, 49–63.
671. ______, *Singular value decomposition and the Moore-Penrose inverse of bordered matrices*, SIAM J. Appl. Math. **31** (1976), no. 1, 31–41.
672. ______, *Generalized inverses, EP elements and associates*, Rev. Roumaine Math. Pures Appl. **23** (1978), no. 1, 57–60.
673. ______, *Schur's theorem and the Drazin inverse*, Pacific J. Math. **78** (1978), no. 1, 133–138.
674. ______, *Spectral inverses and the row-space equations*, Linear Algebra and its Applications **20** (1978), no. 1, 57–68.
675. ______, *An application of the Moore-Penrose inverse to antisymmetric relations*, Proc. Amer. Math. Soc. **78** (1980), no. 2, 181–186.
676. ______, *How to partially order regular elements*, Math. Japon. **25** (1980), no. 1, 1–13.
677. ______, *A method for calculating* A^d, Math. Japon. **26** (1981), no. 1, 37–43.
678. ______, *Applications of the Wrońskian and Gram matrices of* $\{t^i e^{\lambda_k t}\}$, Linear Algebra and its Applications **43** (1982), 229–241.
679. ______, *The reverse order law revisited*, Linear Algebra and its Applications **76** (1986), 241–246.
680. ______, *The group inverse of a block triangular matrix*, Current trends in matrix theory (Auburn, Ala., 1986), North-Holland, New York, 1987, pp. 137–146.
681. ______, *A remark on the characterization of the parallel sum of two matrices*, Linear and Multilinear Algebra **22** (1987), no. 2, 193–197.
682. ______, *The pyramid decomposition and rank minimization*, Linear Algebra and its Applications **191** (1993), 53–76.
683. ______, *EP perturbations*, Sankhyā Ser. A **56** (1994), no. 2, 347–357.
684. ______, *The weighted* $*$*-core-nilpotent decomposition*, Linear Algebra and its Applications **211** (1994), 101–111.
685. R. E. Hartwig and F. J. Hall, *Pseudo–similarity for matrices over a field*, Proc. Amer. Math. Soc. **71** (1978), no. 1, 6–10.
686. ______, *Applications of the Drazin inverse to Cesàro– Neumann iterations*, In Campbell [**267**], pp. 145–195.
687. R. E. Hartwig and I. J. Katz, *On products of EP matrices*, Linear Algebra and its Applications **252** (1997), 339–345.
688. R. E. Hartwig and J. Levine, *Applications of the Drazin inverse to the Hill cryptographic system. III*, Cryptologia **5** (1981), no. 2, 67–77.
689. ______, *Applications of the Drazin inverse to the Hill cryptographic system. IV*, Cryptologia **5** (1981), no. 4, 213–228.
690. R. E. Hartwig and Jiang Luh, *A note on the group structure of unit regular ring elements*, Pacific J. Math. **71** (1977), no. 2, 449–461.
691. R. E. Hartwig and J. M. Shoaf, *Group inverses and Drazin inverses of bidiagonal and triangular Toeplitz matrices*, J. Austral. Math. Soc. Ser. A **24** (1977), no. 1, 10–34.
692. ______, *On the derivative of the Drazin inverse of a complex matrix*, SIAM J. Math. Anal. **10** (1979), no. 1, 207–216.
693. ______, *Invariance, group inverses and parallel sums*, Rev. Roumaine Math. Pures Appl. **25** (1980), no. 1, 33–42.
694. R. E. Hartwig, Guorong Wang, and Yimin Wei, *Some additive results on Drazin inverse*, Linear Algebra and its Applications **322** (2001), no. 1-3, 207–217.
695. D. A. Harville, *Extension of the Gauss-Markov theorem to include the estimation of random effects*, Ann. Statist. **4** (1976), no. 2, 384–395.
696. ______, *Generalized inverses and ranks of modified matrices*, J. Indian Soc. Agricultural Statist. **49** (1996/97), 67–78, iv.
697. ______, *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York, 1997.
698. W. R. Harwood, V. Lovass-Nagy, and D. L. Powers, *A note on the generalized inverses of some partitioned matrices*, SIAM J. Appl. Math. **19** (1970), 555–559.
699. D. M. Hawkins and D. Bradu, *Application of the Moore-Penrose inverse of a data matrix in multiple regression*, Linear Algebra and its Applications **127** (1990), 403–425.
700. J. B. Hawkins and A. Ben-Israel, *On generalized matrix functions*, Linear and Multilinear Algebra **1** (1973), no. 2, 163–171.
701. E. V. Haynsworth and J. R. Wall, *Group inverses of certain nonnegative matrices*, Linear Algebra Appl. **25** (1979), 271–288.
702. ______, *Group inverses of certain positive operators*, Linear Algebra and its Applications **40** (1981), 143–159.
703. J. Z. Hearon, *Construction of* epr *generalized inverses by inversion of nonsingular matrices*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 57–60.
704. ______, *A generalized matrix version of Rennie's inequality*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 61–64.

705. ______, *On the singularity of a certain bordered matrix*, SIAM J. Appl. Math. **15** (1967), 1413–1421.
706. ______, *Partially isometric matrices*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 225–228.
707. ______, *Polar factorization of a matrix*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 65–67.
708. ______, *Symmetrizable generalized inverses of symmetrizable matrices*, J. Res. Nat. Bur. Standards Sect. B **71B** (1967), 229–231.
709. ______, *Generalized inverses and solutions of linear systems*, J. Res. Nat. Bur. Standards Sect. B **72B** (1968), 303–308.
710. J. Z. Hearon and J. W. Evans, *Differentiable generalized inverses*, J. Res. Nat. Bur. Standards Sect. B **72B** (1968), 109–113.
711. ______, *On spaces and maps of generalized inverses*, J. Res. Nat. Bur. Standards Sect. B **72B** (1968), 103–107.
712. C. Heil and D. Walnut, *Continuous and discrete wavelet transforms*, SIAM Rev. **31** (1989), 628–666.
713. G. Heinig, *The group inverse of the transformation* $S(X) = AX - XB$, Linear Algebra and its Applications **257** (1997), 321–342.
714. G. Heinig and F. Hellinger, *On the Bezoutian structure of the Moore–Penrose inverses of Hankel matrices*, SIAM Journal on Matrix Analysis and Applications **14** (1993), no. 3, 629–645.
715. ______, *Displacement structure of generalized inverse matrices*, Linear Algebra and its Applications **211** (1994), 67–83.
716. ______, *Displacement structure of generalized inverse matrices*, Linear Algebra and its Applications **211** (1994), 67–83.
717. ______, *Moore-Penrose inversion of square Toeplitz matrices*, SIAM Journal on Matrix Analysis and Applications **15** (1994), no. 2, 418–450.
718. B. W. Helton, *Logarithms of matrices*, Proceedings of the American Mathematical Society **19** (1968), 733–736.
719. H. V. Henderson and S. R. Searle, *The vec-permutation matrix, the vec operator and Kronecker products: a review*, Linear and Multilinear Algebra **9** (1980/81), no. 4, 271–288.
720. ______, *On deriving the inverse of a sum of matrices*, SIAM Rev. **23** (1981), no. 1, 53–60.
721. G. P. Herring, *A note on generalized interpolation and the pseudoinverse*, SIAM J. Numer. Anal. **4** (1967), 548–556.
722. M. R. Hestenes, *Applications of the theory of quadratic forms in Hilbert space to the calculus of variations*, Pacific J. Math. **1** (1951), 525–581.
723. ______, *Inversion of matrices by biorthogonalization and related results*, J. Soc. Indust. Appl. Math. **6** (1958), 51–90.
724. ______, *Relative Hermitian matrices*, Pacific J. Math. **11** (1961), 225–245.
725. ______, *Relative self-adjoint operators in Hilbert space*, Pacific J. Math. **11** (1961), 1315–1357.
726. ______, *A ternary algelbra with applications to matrices and linear transformations*, Arch. Rational Mech. Anal. **11** (1962), 138–194.
727. ______, *On a ternary algebra*, Scripta Math. **29** (1973), 253–272.
728. ______, *Pseudoinverses and conjugate gradients*, Comm. ACM **18** (1975), 40–43.
729. K. R. Hickey and G. R. Luecke, *Remarks on Marti's method for solving first kind equations*, SIAM J. Numer. Anal. **19** (1982), no. 3, 623–628.
730. N. J. Higham, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput. **7** (1986), no. 4, 1160–1174.
731. ______, *Computing real square roots of a real matrix*, Linear Algebra and its Applications **88/89** (1987), 405–430.
732. ______, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra and Its Applications **103** (1988), 103–118.
733. N. J. Higham and R. S. Schreiber, *Fast polar decomposition of an arbitrary matrix*, SIAM J. Sci. Comput. **11** (1990), no. 4, 648–655.
734. D. Hilbert, *Grundzüge einer algemeine Theorie der linearen Integralgleichungen*, B. G. Teubner, Leipzig and Berlin, 1912, (Reprint of six articles which appeared originally in the *Götingen Nachrichten* (1904), 49–51; (1904), 213–259; (1905), 307–338; (1906), 157–227; (1906), 439–480; (1910), 355–417).
735. B. L. Ho and R. E. Kalman, *Effective construction of linear state–variables models from input/output functions*, Regelungstechnik **14** (1966), 545–548.
736. Yu-chi Ho and R. L. Kashyap, *A class of iterative procedures for linear inequalities*, SIAM J. Control **4** (1966), 112–115.
737. H. W. Hodaway, *GINV, A subroutine in ANSI fortran for generalized matrix inversion*, Australian Computer J. **9** (1977), no. 4, 159–161.
738. A. E. Hoerl and R. W. Kennard, *Ridge regression—1980. Advances, algorithms, and applications*, Amer. J. Math. Management Sci. **1** (1981), no. 1, 5–83.
739. R. W. Hoerl, *Ridge analysis 25 years later*, Amer. Statist. **39** (1985), no. 3, 186–192.
740. A. Holder, *Desinging radiotherapy plans with elastic constraints and interior point methods*, Tech. Report No. 49, Trinity University, San Antonio, TX, 2000.
741. A. Holder, J. Sturm, and S. Zhang, *Marginal and parametric analysis of the central optimal solution*, Tech. Report No. 48, Trinity University Mathematics, 1999, (to appear in Information Systems and Operational Research).
742. A. G. Holder and R. J. Caron, *Uniform bounds on the limiting and marginal derivatives of the analytic center solution over a set of normalized weights*, Oper. Res. Lett. **26** (2000), no. 2, 49–54.

743. R. B. Holmes, *A Course on Optimization and Best Approximation*, Springer–Verlag, Berlin, 1972.

744. H. H. H. Homeier, *Extrapolationsverfahren für Zahlen-, Vektor- und Matrizenfolgen und ihre Anwendung in der Theoretischen und Physikalischen Chemie*, Habilitation thesis, Universität Regensburg, 1996.

745. H. H. H. Homeier, *The vector $\mathcal{J}$ extrapolation method*, Iterative Methods in Scientific Computation (Junping Wang, Myron B. Allen, III., Benito M. Chen, and Tarek Mathew, eds.), IMACS Series in Computational and Applied Mathematics, vol. 4, IMACS, Dept. of Computer Science, Rutgers University, New Brunswick, NJ 08903, USA, 1998, Proceedings of the *Third IMACS International Symposium on Iterative Methods in Scientific Computation Jackson Hole, Wyoming, USA - July 9-12, 1997*, pp. 375–380.

746. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985, (corrected reprint Cambridge University Press, Cambridge, 1990).

747. ———, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

748. A. Höskuldsson, *Data analysis, matrix decompositions, and generalized inverse*, SIAM J. Sci. Comput. **15** (1994), no. 2, 239–262.

749. H. Hotelling, *Relation between two sets of variates*, Biometrika **28** (1936), 322–377.

750. ———, *Some new methods in matrix calculations*, Annals of Mathematical Statistics **14** (1943), 1–34.

751. ———, *Relation of the newer multivariate statistical methods to factor analysis*, Br. J. Static. Psychol. **10** (1957), 69–79.

752. P. D. Hough and S. A. Vavasis, *Complete orthogonal decomposition for weighted least squares*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 2, 369–392.

753. A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

754. ———, *The Kantorovich and some related inequalities*, SIAM Rev. **7** (1965), 463–473.

755. A. S. Householder and G. Young, *Matrix approximation and latent roots*, Amer. Math. Monthly **45** (1938), 165–171.

756. F. C. Hsuan, P. Langenberg, and A. J. Getson, *The $\{2\}$-inverse with applications in statistics*, Linear Algebra and its Applications **70** (1985), 241–248.

757. K. Huang and J. Yu, *Remarks on a formula for the pseudoinverse by the modified Huang algorithm*, Computing **59** (1997), no. 2, 183–185.

758. Ching Hsiang Hung and T. L. Markham, *The Moore-Penrose inverse of a partitioned matrix $M = \left(\begin{smallmatrix} A & 0 \\ B & C \end{smallmatrix}\right)$*, Czechoslovak Math. J. **25(100)** (1975), no. 3, 354–361.

759. ———, *The Moore-Penrose inverse of a sum of matrices*, J. Austral. Math. Soc. Ser. A **24** (1977), no. 4, 385–392.

760. M. F. Hurt and C. Waid, *A generalized inverse which gives all the integral solutions to a system of linear equations*, SIAM J. Appl. Math. **19** (1970), 547–550.

761. W. A. Hurwitz, *On the pseudo–resolvent to the kernel of an integral equation*, Trans. Amer. Math. Soc. **13** (1912), 405–418.

762. D. Huylebrouck, *The generalized inverse of a sum with radical element: applications*, Linear Algebra and its Applications **246** (1996), 159–175.

763. D. Huylebrouck and R. Puystjens, *Generalized inverses of a sum with a radical element*, Linear Algebra and its Applications **84** (1986), 289–300.

764. D. Huylebrouck, R. Puystjens, and J. Van Geel, *The Moore-Penrose inverse of a matrix over a semisimple Artinian ring*, Linear and Multilinear Algebra **16** (1984), no. 1-4, 239–246.

765. ———, *The Moore-Penrose inverse of a matrix over a semi-simple Artinian ring with respect to an involution*, Linear and Multilinear Algebra **23** (1988), no. 3, 269–276.

766. Y. Ijiri, *On the generalized inverse of an incidence matrix*, J. Soc. Indust. Appl. Math. **13** (1965), 941–945.

767. K. D. Ikramov, *An algebraic proof of a result by Gonzaga and Lara*, Linear Algebra and its Applications **299** (1999), no. 1-3, 191–194.

768. I. C. F. Ipsen, *An overview of relative $\sin\theta$ theorems for invariant subspaces of complex matrices*, J. Comput. Appl. Math. **123** (2000), no. 1-2, 131–153.

769. I. C. F. Ipsen and C. D. Meyer, Jr., *The angle between complementary subspaces*, Amer. Math. Monthly **102** (1995), no. 10, 904–911.

770. ———, *The idea behind Krylov methods*, Amer. Math. Monthly **105** (1998), no. 10, 889–899.

771. C. Itiki, R. E. Kalaba, and F. E. Udwadia, *Appell's equations of motion and the generalized inverse form*, Recent trends in optimization theory and applications, World Sci. Publishing, River Edge, NJ, 1995, pp. 123–143.

772. Saichi Izumino, *Convergence of generalized inverses and spline projectors*, J. Approx. Theory **38** (1983), no. 3, 269–278.

773. N. Jacobson, *An application of E. H. Moore's determinant of a Hermitian matrix*, Bull. Amer. Math. Soc. **45** (1939), 745–748.

774. S. K. Jain, *Nonnegative rectangular matrices having certain nonnegative W-weighted group inverses*, Proc. Amer. Math. Soc. **85** (1982), no. 1, 1–9.

775. S. K. Jain, S. K. Mitra, and H. -J. Werner, *Extensions of $\mathcal{G}$-based matrix partial orders*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 4, 834–850.

776. D. James, *Implicit nullspace iterative methods for constrained least squares problems*, SIAM J. Matrix. Anal. Appl. **13** (1992), 962–978.

777. D. R. Jensen, *Minimal properties of Moore-Penrose inverses*, Linear Algebra and its Applications **196** (1994), 175–182.

778. J. W. Jerome and L. L. Schumaker, *A note on obtaining natural spline functions by the abstract*

approach of Atteia and Laurent, SIAM J. Numer. Anal. **5** (1968), 657–663.

779. Jun Ji, *The algebraic perturbation method for generalized inverses*, J. Comput. Math. **7** (1989), no. 4, 327–333.
780. ______, *An alternative limit expression of Drazin inverse and its application*, Appl. Math. Comput. **61** (1994), no. 2-3, 151–156.
781. Sheng Jiang, *Angles between Euclidean subspaces*, Geom. Dedicata **63** (1996), no. 2, 113–121.
782. L. Jódar, A. G. Law, A. Rezazadeh, J. H. Weston, and G. Wu, *Computations for the Moore Penrose and other generalized inverses*, Proceedings of the Twentieth Manitoba Conference on Numerical Mathematics and Computing (Winnipeg, MB, 1990), vol. 80, 1991, pp. 57–64.
783. J. A. John, *Use of generalized inverse matrices in MANOVA*, J. Roy. Statist. Soc. Ser. B **32** (1970), 137–143.
784. P. W. M. John, *Pseudo–inverses in the analysis of variance*, Ann. Math. Statist. **35** (1964), 895–896.
785. C. R. Johnson, M. K. Kerr, and D. P. Stanford, *Semipositivity of matrices*, Linear and Multilinear Algebra **37** (1994), no. 4, 265–271, (see [**1585**]).
786. C. R. Johnson, R. Loewy, D. D. Olesky, and P. van den Driessche, *Maximizing the spectral radius of fixed trace diagonal perturbations of nonnegative matrices*, Linear Algebra and its Applications **241/243** (1996), 635–654.
787. J. Jones, Jr., *On the Lyapunov stability criteria*, J. Soc. Indust. Appl. Math. **13** (1965), 941–945.
788. ______, *Solution of certain matrix equations*, Proc. Amer. Math. Soc. **31** (1972), 333–339.
789. Jon Jones, N. P. Karampetakis, and A. C. Pugh, *The computation and application of the generalized inverse via Maple*, J. Symbolic Comput. **25** (1998), no. 1, 99–124.
790. F. Jongmans, *Resolution numérique de l'équation matricielle* $AXB = C$, Simon Stevin **34** (1960/1961), 3–26.
791. ______, *Retour critique sur l'équation matricielle* $AXB = C$, Hommage au Professeur Lucien Godeaux, Librairie Universitaire, Louvain, 1968, pp. 127–134.
792. V. N. Joshi, *Remarks on iterative methods for computing the generalised inverse*, Studia Sci. Math. Hungar. **8** (1973), 457–461.
793. ______, *A determinant for rectangular matrices*, Bull. Austral. Math. Soc. **21** (1980), no. 1, 137–146.
794. V. N. Joshi and R. P. Tewarson, *On solving ill-conditioned systems of linear equations*, Trans. New York Acad. Sci. (2) **34** (1972), 565–571.
795. S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bulletin de l'Académie Polonaise des Sciences et Lettres **A35** (1937), 355–357.
796. D. G. Kaffes, *An inequality for matrices*, Bull. Soc. Math. Grèce (N.S.) **22** (1981), 143–159.
797. D. G. Kaffes, T. Mathew, M. Bhaskara Rao, and K. Subramanyam, *On the matrix convexity of the Moore-Penrose inverse and some applications*, Linear and Multilinear Algebra **24** (1989), no. 4, 265–271.
798. W. Kahan, *Huge generalized inverses of rank–defficient matrices*, University of Calfornia at Berkley, 2001, (online lecture notes).
799. S. Kakutani, *Some characterizations of Euclidean spaces*, Japan J. Math. **16** (1939), 93–97.
800. Radoslaw Kala, *Projectors and linear estimation in general linear models*, Comm. Statist. A—Theory Methods **10** (1981), no. 9, 849–873.
801. R. E. Kalaba and F. E. Udwadia, *On constrained motion*, Appl. Math. Comput. **51** (1992), no. 1, 85–86.
802. ______, *Lagrangian mechanics, Gauss' principle, quadratic programming, and generalized inverses: new equations for non-holonomically constrained discrete mechanical systems*, Quart. Appl. Math. **52** (1994), no. 2, 229–241.
803. R. E. Kalaba, F. E. Udwadia, and R. Xu, *Constrained motion revisited*, Appl. Math. Comput. **70** (1995), no. 1, 67–76.
804. R. E. Kalaba, F. E. Udwadia, R. Xu, and C. Itiki, *The equivalence of Lagrange's equations of motion of the first kind and the generalized inverse form*, Nonlinear World **2** (1995), no. 4, 519–526.
805. R. E. Kalaba and Rong Xu, *On the generalized inverse form of the equations of constrained motion*, Amer. Math. Monthly **102** (1995), no. 9, 821–825.
806. C. Kallina, *A Green's function approach to perturbations of periodic solutions*, Pacific J. Math. **29** (1969), 325–334.
807. R. E. Kalman, *Contributions to the theory of optimal control*, Bol. Soc. Mat. Mexicana **5** (1960), no. 2, 102–119.
808. ______, *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D. J. Basic Eng. **82** (1960), 35–45.
809. ______, *New results in linear filtering and prediction problems*, Trans. ASME Ser. D. J. Basic Eng. **83** (1961), 95–107.
810. ______, *Mathematical description of linear dynamical systems*, SIAM J. Control **1** (1963), 152–192.
811. ______, *Algebraic aspects of the generalized inverses of a rectangular matrix*, Generalized Inverses and Applications (Proc. Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1973) (M. Z. Nashed, ed.), Academic Press, New York, 1976, pp. 189–213.
812. R. E. Kalman, Y. C. Ho, and K. S. Narendra, *Controllability of linear dynamical systems*, Contributions to Differential Equations, Vol. I, Interscience, New York, 1963, pp. 189–213.
813. W. J. Kammerer and M. Z. Nashed, *A generalization of a matrix iterative method of g. cimmino to best approximate solution of linear integral equations of the first kind*, Atti Accad. Naz. Lincei

Rend. Cl. Sci. Fis. Mat. Natur. (8) **51** (1971), 20–25.

814. ______, *Steepest descent for singular linear operators with nonclosed range*, Applicable Anal. **1** (1971), no. 2, 143–159.
815. ______, *Iterative methods for best approximate solutions of linear integral equations of the first and second kinds*, J. Math. Anal. Appl. **40** (1972), 547–573.
816. ______, *On the convergence of the conjugate gradient method for singular linear operator equations*, SIAM J. Numer. Anal. **9** (1972), 165–181.
817. L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Pergamon Press, Oxford, 1964, (translated from Russian).
818. L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis*, Interscience, New York, 1958, (translated from Russian).
819. N. P. Karampetakis, *Computation of the generalized inverse of a polynomial matrix and applications*, Linear Algebra and its Applications **252** (1997), 35–60.
820. ______, *Generalized inverses of two-variable polynomial matrices and applications*, Circuits Systems Signal Process. **16** (1997), no. 4, 439–453.
821. I. Karasalo, *A criterion for truncation of the QR-decomposition algorithm for the singular linear least squares problem*, Nordisk Tidskr. Informationsbehandling (BIT) **14** (1974), 156–166.
822. S. Kass, *Spaces of closest fit*, Linear Algebra and its Applications **117** (1989), 93–97.
823. A. K. Katsaggelos and S. N. Efstratiadis, *A class of iterative signal restoration algorithms*, IEEE Trans. Acoust. Speech Signal Process. **38** (1990), no. 5, 778–786.
824. I. J. Katz, *Wiegmann type theorems for EP_r matrices*, Duke Math. J. **32** (1965), 423–427.
825. ______, *Remarks on a paper of Ben-Israel*, SIAM J. Appl. Math. **18** (1970), 511–513.
826. I. J. Katz and M. H. Pearl, *On EPr and normal EPr matrices*, J. Res. Nat. Bur. Standards Sect. B **70B** (1966), 47–77.
827. ______, *Solutions of the matrix equation $A^* = XA = AX$*, J. London Math. Soc. **41** (1966), 443–452.
828. L. Kaufman and V. Pereyra, *A method for separable nonlinear least squares problems with separable nonlinear equality constraints*, SIAM J. Numer. Anal. **15** (1978), no. 1, 12–20.
829. H. B. Keller, *On the solution of singular and semidefinite linear systems by iteration*, SIAM Journal on Numerical Analysis **2** (1965), 281–290.
830. S. Keller-McNulty and W. J. Kennedy, *Error-free computation of the Moore-Penrose inverse with application to linear least squares analysis*, J. Statist. Comput. Simulation **27** (1987), no. 1, 45–64.
831. J. D. Kelly, *A regularization approach to the reconciliation of constrained data sets*, Computers & Chemical Engineering **22** (1998), 1771–1788.
832. I. M. Khabaza, *An iterative least-square method suitable for solving large sparse matrices*, Comput. J. **6** (1963/1964), 202–206.
833. C. G. Khatri, *A note on a commutative g-inverse of a matrix*, Sankhyā Ser. A **32** (1970), 299–310.
834. ______, *A representation of a matrix and its use in the Gauss-Markoff model*, J. Indian Statist. Assoc. **20** (1982), 89–98.
835. ______, *A generalization of Lavoie's inequality concerning the sum of idempotent matrices*, Linear Algebra and its Applications **54** (1983), 97–108.
836. ______, *Commutative g-inverse of a matrix*, Math. Today **3** (1985), 37–40.
837. ______, *A note on idempotent matrices*, Linear Algebra and its Applications **70** (1985), 185–195.
838. ______, *Study of redundancy of vector variables in canonical correlations*, Comm. Statist. Theory Methods **18** (1989), no. 4, 1425–1440.
839. C. G. Khatri and S. K. Mitra, *Hermitian and nonnegative solutions of linear matrix equations*, SIAM J. Appl. Math. **31** (1976), 579–585.
840. C. G. Khatri and C. R. Rao, *Some extensions of the Kantorovich inequality and statistical applications*, J. Multivariate Anal. **11** (1981), no. 4, 498–505.
841. Byung Chun Kim and Jang Taek Lee, *The Moore-Penrose inverse for the classificatory models*, J. Korean Statist. Soc. **15** (1986), no. 1, 46–61.
842. Doh-Hyun Kim and Jun-Ho Oh, *The Moore-Penrose inverse for the classificatory models*, Control Engineering Practice **7** (1999), no. 3, 369–373.
843. S. Kim, *Generalized inverses in numerical solutions of Cauchy singular integral equations*, Commun. Korean Math. Soc. **13** (1998), no. 4, 875–888.
844. ______, *Numerical solutions of cauchy singular integral equations using generalized inverses*, Computers & Mathematics with Applications **38** (1999), no. 5–6, 183–195.
845. Chen F. King, *A note on Drazin inverses*, Pacific J. Math. **70** (1977), no. 2, 383–390.
846. M. J. L. Kirby, *Generalized Inverses and Chance Constrained Programming*, Applied math., Northwestern Univ., Evanston, IL, June 1965.
847. S. J. Kirkland, *The group inverse associated with an irreducible periodic nonnegative matrix*, SIAM J. Matrix Anal. Appl. **16** (1995), no. 4, 1127–1134.
848. S. J. Kirkland and M. Neumann, *Convexity and concavity of the Perron root and vector of Leslie matrices with applications to a population model*, SIAM J. Matrix Anal. Appl. **15** (1994), no. 4, 1092–1107.
849. ______, *Group inverses of M-matrices associated with nonnegative matrices having few eigenvalues*, Proceedings of the Workshop "Nonnegative Matrices, Applications and Generalizations" and the Eighth Haifa Matrix Theory Conference (Haifa, 1993), vol. 220, 1995, pp. 181–213.
850. ______, *The M-matrix group generalized inverse problem for weighted trees*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 1, 226–234 (electronic).

851. ______, *Cutpoint decoupling and first passage times for random walks on graphs*, SIAM J. Matrix Anal. Appl. **20** (1999), no. 4, 860–870 (electronic).
852. ______, *On group inverses of M-matrices with uniform diagonal entries*, Linear Algebra and its Applications **296** (1999), no. 1-3, 153–170.
853. ______, *Extremal first passage times for trees*, Linear and Multilinear Algebra **48** (2000), 21–33.
854. ______, *Regular Markov chains for which the transition matrix has large exponent*, Linear Algebra and its Applications **316** (2000), no. 1-3, 45–65.
855. S. J. Kirkland, M. Neumann, and B. L. Shader, *Distances in weighted trees and group inverse of Laplacian matrices*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 4, 827–841.
856. ______, *Applications of Paz's inequality to perturbation bounds for Markov chains*, Linear Algebra and its Applications **268** (1998), 183–196.
857. ______, *Bounds on the subdominant eigenvalue involving group inverses with applications to graphs*, Czechoslovak Math. J. **48(123)** (1998), no. 1, 1–20.
858. ______, *On a bound on algebraic connectivity: the case of equality*, Czechoslovak Math. J. **48(123)** (1998), no. 1, 65–76.
859. F. H. Kishi, *On line computer control techniques and their application to re–entry aerospace vehicle control*, Advances in Control Systems Theory and Applications (C. T. Leondes, Editor), Academic Press, New York, 1964, pp. 245–257.
860. V. Klee, *Review of 'linearity of best approximations: A characterization of ellipsoids' (Rudin and Smith)*, Math. Reviews **23** (1962), A2028.
861. A. Klinger, *Approximate pseudoinverse solutions to ill–conditioned linear systems*, J. Optimization Th. Appl. **2** (1968), 117–128.
862. A. V. Knyazev and M. E. Argentati, *An effective and robust algorithm for finding principal angles between subspaces using an A-based scalar product*, Tech. Report 163, Center for Computational Mathematics, University of Colorado at Denver, August 2000.
863. M. Koecher, *The generalized inverse of integral matrices*, Linear Algebra and its Applications **71** (1985), 187–198.
864. E. G. Kogbetliantz, *Solution of linear systems by diagonalization of coefficients matrix*, Quarterly of Applied Mathematics **13** (1955), 123–132.
865. J. J. Koliha, *Power convergence and pseudoinverses of operators in Banach spaces*, J. Math. Anal. Appl. **48** (1974), 446–469.
866. ______, *Pseudo-inverses of operators*, Bull. Amer. Math. Soc. **80** (1974), 325–328.
867. ______, *A generalized Drazin inverse*, Glasgow Math. J. **38** (1996), no. 3, 367–381.
868. ______, *A simple proof of the product theorem for EP matrices*, Linear Algebra and its Applications **294** (1999), no. 1-3, 213–215.
869. ______, *Elements of C^*-algebras commuting with their Moore-Penrose inverse*, Studia Math. **139** (2000), no. 1, 81–90.
870. J. J. Koliha and V. Rakočević, *Continuity of the Drazin inverse. II*, Studia Math. **131** (1998), no. 2, 167–177.
871. J. J. Koliha and Ivan Straškraba, *Power bounded and exponentially bounded matrices*, Appl. Math. **44** (1999), no. 4, 289–308.
872. A. Korányi, *Around the finite–dimensional spectral theorem*, Amer. Math. Monthly **108** (2001), 120–125.
873. André Korganoff and Monica Pavel-Parvu, *Méthodes de calcul numérique. Tome II: Éléments de théorie des matrices carrées et rectangles en analyse numérique*, Dunod, Paris, 1967.
874. V. M. Korsukov, *An application of iteration methods to the computation of semi-inverses of matrices*, Optimization methods and operations research, applied mathematics (Russian), Akad. Nauk SSSR Sibirsk. Otdel. Sibirsk. Ènerget. Inst., Irkutsk, 1976, pp. 171–173, 191.
875. ______, *Some properties of generalized inverse matrices*, Degenerate systems of ordinary differential equations, "Nauka" Sibirsk. Otdel., Novosibirsk, 1982, pp. 19–37.
876. Mathew Koshy and R. P. Tewarson, *On the use of restricted pseudo-inverses for the unified derivation of quasi-Newton updates*, IMA J. Numer. Anal. **5** (1985), no. 2, 141–151.
877. S. Kourouklis and C. C. Paige, *A constrained least squares approach to the general Gauss-Markov linear model*, J. Amer. Statist. Assoc. **76** (1981), no. 375, 620–625.
878. Olaf Krafft, *An arithmetic-harmonic-mean inequality for nonnegative definite matrices*, Linear Algebra and its Applications **268** (1998), 243–246.
879. R. G. Kreijger and H. Neudecker, *Exact linear restrictions on parameters in the general linear model with a singular covariance matrix*, J. Amer. Statist. Assoc. **72** (1977), no. 358, 430–432.
880. M. G. Kreĭn, *The theory of self-adjoint extensions of semi-bounded Hermitian transformations and its applications. I*, Rec. Math. [Mat. Sbornik] N.S. **20(62)** (1947), 431–495.
881. ______, *The theory of self-adjoint extensions of semi-bounded Hermitian transformations and its applications. II*, Mat. Sbornik N.S. **21(63)** (1947), 365–404.
882. R. Kress, *On the Fredholm alternative*, Integral Equations Operator Theory **6** (1983), no. 3, 453–457.
883. E. V. Krishnamurthy, *Fast parallel exact computation of the Moore-Penrose inverse and rank of a matrix*, Elektron. Informationsverarb. Kybernet. **19** (1983), no. 1-2, 95–98.
884. W. Kruskal, *The coordinate-free approach to Gauss-Markov estimation, and its application to*

missing and extra observations, Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, Univ. California Press, Berkeley, Calif., 1961, pp. 435–451.

885. ———, *When are Gauss-Markov and least squares estimators identical? A coordinate-free approach*, Ann. Math. Statist **39** (1968), 70–75, (see [**634**]).
886. ———, *The geometry of generalized inverses*, J. Roy. Statist. Soc. Ser. B **37** (1975), 272–283, (correction in *J. Roy. Statist. Soc. Ser. B* **48**(1986), 258).
887. Jiao Xun Kuang, *The representation and approximation of Drazin inverses of linear operators*, Numer. Math. J. Chinese Univ. **4** (1982), no. 2, 97–106.
888. ———, *Approximate methods for generalized inverses of operators in Banach spaces*, J. Comput. Math. **11** (1993), no. 4, 323–328.
889. V. N. Kublanovskaya, *On the calculation of generalized inverses and projections (Russian)*, Z. Vycisl. Mat. i Mat. Fiz. **6** (1966), 326–332.
890. H. W. Kuhn (ed.), *Proceedings Princeton Sympos. Math. Prog.*, Princeton, NJ, Princeton Univ. Press, 1970.
891. S. H. Kulkarni and K. C. Sivakumar, *Applications of generalized inverses to interval linear programs in Hilbert spaces*, Numer. Funct. Anal. Optim. **16** (1995), no. 7-8, 965–973.
892. ———, *Explicit solutions of a special class of linear programming problems in Banach spaces*, Acta Sci. Math. (Szeged) **62** (1996), no. 3-4, 457–465.
893. P. Kunkel and V. Mehrmann, *Generalized inverses of differential-algebraic operators*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 2, 426–442.
894. I Wen Kuo, *The Moore-Penrose inverses of singular M-matrices*, Linear Algebra and Appl. **17** (1977), no. 1, 1–14.
895. M. C. Y. Kuo and L. F. Mazda, *Mimimum energy problems in Hilbert function space*, J. Franklin Inst. **283** (1967), 38–54.
896. S. Kurepa, *Generalized inverse of an operator with a closed range*, Glasnik Mat. **3** (1968), no. 23, 207–214.
897. B. Kutzler and V. Kokol-Voljc, *Introduction to Derive 5*, Texas Instruments, Dallas, TX 75265, 2000.
898. C. D. LaBudde and G. R. Verma, *On the computation of a generalized inverse of a matrix*, Quart. Appl. Math. **27** (1969), 391–395.
899. B. F. Lamond, *An efficient factorization for the group inverse*, SIAM J. Algebraic Discrete Methods **8** (1987), no. 4, 797–808.
900. ———, *A generalized inverse method for asymptotic linear programming*, Math. Programming **43** (1989), no. 1 (Ser. A), 71–86.
901. B. F. Lamond and M. L. Puterman, *Generalized inverses in discrete time Markov decision processes*, SIAM J. Matrix Anal. Appl. **10** (1989), no. 1, 118–134.
902. P. Lancaster, *Explicit solutions of linear matrix equations*, SIAM Rev. **12** (1970), 544–566.
903. ———, *A fundemental theorem on lambda-matrices with applications—I. ordinary differential equations with constant coefficients*, Linear Algebra and Its Applications **18** (1977), 189–211.
904. ———, *A fundemental theorem on lambda-matrices with applications—II. difference equations with constant coefficients*, Linear Algebra and Its Applications **18** (1977), 213–222.
905. P. Lancaster and J. G. Rokne, *Solutions of nonlinear operator equations*, SIAM Journal on Mathematical Analysis **8** (1977), 448–457.
906. C. Lanczos, *Linear systems in self–adjoint form*, Amer. Math. Monthly **65** (1958), 665–679.
907. ———, *Linear Differential Operators*, D. Van Nostrand, Co., Princeton, 1961.
908. E. M. Landesman, *Hilbert–space methods in elliptic partial differential equations*, Pacific J. Math. **21** (1967), 113–131.
909. P. M. Lang, J. M. Brenchley, R. G. Nieves, and J. H. Kalivas, *Cyclic subspace regression*, J. Multivariate Anal. **65** (1998), no. 1, 58–70.
910. C. E. Langenhop, *On generalized inverses of matrices*, SIAM J. Appl. Math. **15** (1967), 1239–1246.
911. ———, *The Laurent expansion for a nearly singular matrix*, Linear Algebra and its Applications **4** (1971), 329–340.
912. ———, *On the invertibiliby of a nearly singular matrix*, Linear Algebra and its Applications **7** (1973), 361–365.
913. L. J. Lardy, *A series representation for the generalized inverse of a closed linear operator*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **58** (1975), no. 2, 152–157.
914. Ty A. Lasky and B. Ravani, *Sensor-based path planning and motion control for a robotic system for roadway crack sealing*, IEEE Transactions on Control Systems Technology **8** (2000), 609–622.
915. K. J. Latawiec, S. Bańka, and J. Tokarzewski, *Control zeros and nonminimum phase lti mimo systems*, Annual Reviews in Control **24** (2000), no. 1, 105–112.
916. P. -J. Laurent, *Approximation et Optimisation. collection enseignement des sciences, no. 13*, Hermann, Paris, 1972.
917. ———, *Quadratic convex analysis and splines*, Methods of Functional Analysis in Approximation Theory (Bombay, 1985), Birkhäuser, Basel, 1986, pp. 17–43.
918. J. -L. Lavoie, *A determinantal inequality involving the Moore-Penrose inverse*, Linear Algebra and its Applications **31** (1980), 77–80.
919. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1974, (reprinted, *Classics in Applied Mathematics*, 15. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995. xii+337 pp).

920. E. B. Leach, *A note on inverse function theorems*, Proc. Amer. Math. Soc. **12** (1961), 694–697.
921. ———, *On a related function theorem*, Proc. Amer. Math. Soc. **14** (1963), 687–689.
922. G. G. Lendaris, K. Mathia, and R. Saeks, *Linear Hopfield networks and constrained optimization*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **29** (1999), 114–118.
923. G. -S. Leng and Y. Zhang, *Vertex angles for simplices*, Applied Mathematics Letters **12** (1999), 1–5.
924. A. H. Lent, *Wiener–Hopf Operators and Factorizations*, Doctoral dissertation in applied mathematics, Northwestern Univertsity, Evanston, IL, June 1971.
925. A. S. Leonov, *Approximate calculation of a pseudoinverse matrix by means of the generalized residual principle*, Zh. Vychisl. Mat. i Mat. Fiz. **25** (1985), no. 6, 933–935, 959.
926. ———, *The method of a minimal pseudoinverse matrix for solving ill-posed problems of linear algebra*, Dokl. Akad. Nauk SSSR **285** (1985), no. 1, 36–40, (English translation: Soviet Math. Dokl. **32**(1985), no. 3, 628–632).
927. ———, *The method of a minimal pseudoinverse matrix*, Zh. Vychisl. Mat. i Mat. Fiz. **27** (1987), no. 8, 1123–1138, 1276.
928. ———, *The minimal pseudo-inverse matrix method: theory and numerical realization*, Zh. Vychisl. Mat. i Mat. Fiz. **31** (1991), no. 10, 1427–1443.
929. ———, *The method of minimal pseudoinversed matrix. Basic statements*, Ill-posed problems in natural sciences (Moscow, 1991), VSP, Utrecht, 1992, pp. 57–62.
930. Ö. Leringe and P.-Å. Wedin, *A comparison between different methods to compute a vector* $\mathbf{x}$ *which minimizes* $\|A\mathbf{x} - \mathbf{b}\|_2$ *when* $G\mathbf{x} = \mathbf{h}$, Department of computer science, Lund University, Lund, Sweden, March 1970.
931. G. Lešnjak, *Semigroups of EP linear transformations*, Linear Algebra and its Applications **304** (2000), no. 1-3, 109–118.
932. Y. Levin and A. Ben-Israel, *Inverse-free methods for systems of nonlinear equations*, RUTCOR–Rutgers Ctr. Oper. Res. 58-2000, Rutgers University, Piscataway, NJ, 2000.
933. ———, *Directional Halley and quasi-Halley methods in n variables*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (Amsterdam) (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier Science, 2001.
934. ———, *Directional Newton methods in n variables*, Mathematics of Computations (2001).
935. ———, *The Newton bracketing method for convex minimization*, Computational Optimization and Applications (2001).
936. ———, *A Newton method for systems of m equations in n variables*, Nonlinear Analysis (2001).
937. Y. Levin, M. Nediak, and A. Ben-Israel, *A direct approach to calculus of variations via Newton-Raphson method*, Comput. & Applied Mathematics (2001).
938. J. Levine and R. E. Hartwig, *Applications of the Drazin inverse to the Hill cryptographic system. I*, Cryptologia **4** (1980), no. 2, 71–85.
939. ———, *Applications of the Drazin inverse to the Hill cryptographic system. II*, Cryptologia **4** (1980), no. 3, 150–168.
940. B. C. Levy, *A note on the hyperbolic singular value decomposition*, Linear Algebra and its Applications **277** (1998), no. 1-3, 135–142.
941. A. S. Lewis, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal. **2** (1995), no. 1-2, 173–183.
942. T. O. Lewis, T. L. Boullion, and P. L. Odell, *A bibliography on generalized matrix inverses*, In Boullion and Odell [**207**], pp. 283–315.
943. T. O. Lewis and T. G. Newman, *Pseudoinverses of positive semidefinite matrices*, SIAM J. Appl. Math. **16** (1968), 701–703.
944. T. O. Lewis and P. L. Odell, *A generalization of the Gauss-Markov theorem*, J. Amer. Statist. Assoc. **61** (1966), 1063–1066.
945. Bin Li, Yinghui Li, and Xuegang Ying, *Dynamic modeling and simulation of flexible cable with large sag*, Applied Mathematics and Mechanics (English Edition) **21** (2000), 707–714.
946. Chi-Kwong Li and R. Mathias, *Extremal characterizations of the Schur complement and resulting inequalities*, SIAM Rev. **42** (2000), no. 2, 233–246 (electronic).
947. Chi-Kwong Li and Nam-Kiu Tsing, *Some isometries of rectangular complex matrices*, Linear and Multilinear Algebra **23** (1988), no. 1, 47–53.
948. Ren Cang Li, *Norms of certain matrices with applications to variations of the spectra of matrices and matrix pencils*, Linear Algebra and its Applications **182** (1993), 199–234.
949. Ren-Cang Li, *A perturbation bound for the generalized polar decomposition*, BIT **33** (1993), no. 2, 304–308.
950. ———, *Relative perturbation theory. III. More bounds on eigenvalue variation*, Linear Algebra and its Applications **266** (1997), 337–345.
951. ———, *Relative perturbation theory. I. Eigenvalue and singular value variations*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 4, 956–982 (electronic).
952. ———, *Relative perturbation theory. II. Eigenspace and singular subspace variations*, SIAM J. Matrix Anal. Appl. **20** (1999), no. 2, 471–492 (electronic).
953. ———, *Relative perturbation theory. IV.* $\sin 2\theta$ *theorems*, Linear Algebra and its Applications **311** (2000), no. 1-3, 45–60.
954. Ren-Cang Li and G. W. Stewart, *A new relative perturbation theorem for singular subspaces*, Linear Algebra and its Applications **313** (2000), no. 1-3, 41–51.

955. Ping Liang, Su Huan Chen, and Cheng Huang, *Moore–Penrose inverse method of topological variation of finite element systems*, Comput. & Structures **62** (1997), no. 2, 243–251.
956. Yi Liang and Xue Rong Yong, *Group inverses of block matrices*, J. Xinjiang Univ. Natur. Sci. **9** (1992), no. 4, 34–39.
957. A. E. Liber, *On the theory of generalized groups*, Doklady Akad. Nauk SSSR (N.S.) **97** (1954), 25–28.
958. D. K. Lika, *The application of generalized inverse operators in iteration processes*, Mat. Issled. **10** (1975), no. 2(36), 264–270, 289.
959. K. -W. Lin and A. R. Sanford, *Improving regional earthquake locations using a modified G matrix and fuzzy logic*, Bulletin of the Seismological Society of America **91** (2001), no. 1-2, 82–93.
960. Shwu-Yeng T. Lin and You Feng Lin, *The n-dimensional Pythagorean theorem*, Linear and Multilinear Algebra **26** (1990), no. 1-2, 9–13, (see [**164**]).
961. E. P. Liski, *On Löwner-ordering antitonicity of matrix inversion*, Acta Math. Appl. Sinica (English Ser.) **12** (1996), no. 4, 435–442.
962. E. P. Liski and S. Puntanen, *A further note on a theorem on the difference of the generalized inverses of two nonnegative definite matrices*, Comm. Statist. Theory Methods **18** (1989), no. 5, 1747–1751.
963. E. P. Liski and Song Gui Wang, *Another look at the naive estimator in a regression model*, Metrika **41** (1994), no. 1, 55–64.
964. ——, *On the $\{2\}$-inverse and some ordering properties of nonnegative definite matrices*, Acta Math. Appl. Sinica (English Ser.) **12** (1996), no. 1, 22–27.
965. W. G. Lister, *Ternary rings*, Trans. Amer. Math. Soc. **154** (1971), 37–55.
966. X. Liu and D. Liu, *Recursive computation of generalized inverses with application to optimal state estimation*, Control Theory Adv. Tech. **10** (1995), 1485–1497.
967. J. Locker, *An existence analysis for nonlinear equations in Hilbert space*, Trans. Amer. Math. Soc. **128** (1967), 403–413.
968. ——, *An existence analysis for nonlinear boundary value problems*, SIAM J. Appl. Math. **19** (1970), 199–207.
969. A. T. Lonseth, *Approximate solutions of Fredholm–type integral equations*, Bull. Amer. Math. Soc. **60** (1954), 415–430.
970. J. López and V. Guerra, *On the computing of the pseudoinverse by Leonov's minimal method*, Revista Investigación Operacional **17** (1996).
971. ——, *Maximum balance criterion for choosing the parameter λ in the minimal pseudoinverse method*, Revista Investigación Operacional **21** (2000).
972. W. S. Loud, *Some singular cases of the implicit function theorem*, Amer. Math. Monthly **68** (1961), 965–977.
973. ——, *Generalized inverses and generalized Green's functions*, SIAM J. Appl. Math. **14** (1966), 342–369.
974. ——, *Some examples of generalized Green's functions and generalized Green's matrices*, SIAM Rev. **12** (1970), 194–210.
975. ——, *A bifurcation application of the generalized inverse of a linear differential operator*, SIAM J. Math. Anal. **11** (1980), no. 3, 545–558.
976. V. Lovass-Nagy, R. J. Miller, and D. L. Powers, *On system realization by matrix generalized inverses*, Internat. J. Control **26** (1977), no. 5, 745–751, (see correction in [**454**]).
977. V. Lovass-Nagy and D. L. Powers, *On the commuting reciprocal inverse of some partitioned matrices*, Linear Algebra and its Applications **4** (1971), 183–190.
978. V. Lovass-Nagy and D. L. Powers, *A note on the "Y-inverse" of a matrix*, Internat. J. Control (1) **18** (1973), 1113–1115.
979. ——, *A relation between the Moore-Penrose and commuting reciprocal inverses*, SIAM J. Appl. Math. **24** (1973), 44–49.
980. ——, *On a relation among generalized inverses, with application to the Moore-Penrose inverse of certain Toeplitz matrices*, Indust. Math. **24** (1974), 67–76.
981. ——, *Matrix generalized inverses in the handling of control problems containing input derivatives*, Internat. J. Systems Sci. **6** (1975), 693–696.
982. ——, *On rectangular systems of differential equations and their application to circuit theory*, J. Franklin Inst. **299** (1975), no. 6, 399–407.
983. Per-Olov Löwdin, *Studies in perturbation theory. IV. Solution of eigenvalue problem by projection operator formalism*, J. Mathematical Phys. **3** (1962), 969–982.
984. Glenn R. Luecke, *A numerical procedure for computing the Moore-Penrose inverse*, Numer. Math. **32** (1979), no. 2, 129–137.
985. G. Lukács, *The generalized inverse matrix and the surface-surface intersection problem*, Theory and practice of geometric modeling (Blaubeuren, 1988), Springer, Berlin, 1989, pp. 167–185.
986. C. C. MacDuffee, *The theory of matrices*, Chelsea, New York, 1956.
987. G. Maeß, *A projection method solving general linear algebraic equations*, Rostock. Math. Kolloq. **12** (1979), 77–85, (extension of results of [**1420**]).
988. ——, *Iterative solution of rectangular systems of linear algebraic equations*, Computational mathematics (Warsaw, 1980) (Warsaw), PWN, 1984, pp. 527–533.
989. G. Maeß and W. Peters, *Lösung inkonsistenter linearer Gleichungssysteme und Bestimmung einer Pseudoinversen für rechteckige Matrizen durch*

Spaltenapproximation, Z. Angew. Math. Mech. **58** (1978), no. 4, 233–237.

990. J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons Ltd., Chichester, 1999, (revised reprint of the 1988 original).
991. P. J. Maher, *Some operator inequalities concerning generalized inverses*, Illinois J. Math. **34** (1990), no. 3, 503–514.
992. ______, *Some norm inequalities concerning generalized inverses*, Linear Algebra and its Applications **174** (1992), 99–110.
993. Timo Mäkeläinen, *Projections and generalized inverses in the general linear model*, Soc. Sci. Fenn. Comment. Phys.-Math. **38** (1970), 13–25.
994. R. K. Manherz, *New energy theorems in Fourier transform theory*, Proc. IEEE **57** (1969), 826–827.
995. R. K. Manherz and S. L. Hakimi, *The generalized inverse in network analysis and quadratic error-minimization problems*, IEEE Trans. Circuit Theory **CT-16** (1969), 559–562.
996. M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn & Bacon, Boston, Mass., 1964.
997. T. L. Markham, *An application of theorems of Schur and Albert*, Proc. Amer. Math. Soc. **59** (1976), no. 2, 205–210.
998. D. W. Marquardt, *Generalized inverses, ridge regression, biased linear estimation, and nonlinear regression*, Technometrics **12** (1970), 591–613.
999. G. Marsaglia, *Conditional means and variances of normal variables with singular convariance matrix*, J. Amer. Statist. Assoc. **58** (1964), 1203–1204.
1000. G. Marsaglia and G. P. H. Styan, *Equalities and inequalities for ranks of matrices*, Linear and Multilinear Algebra **2** (1974/75), 269–292.
1001. J. T. Marti, *An algorithm for computing minimum norm solutions of Fredholm integral equations of the first kind*, SIAM J. Numer. Anal. **15** (1978), no. 6, 1071–1076, (see [**729**]).
1002. R. S. Martin, G. Peters, and J. H. Wilkinson, *Iterative refinement of the solution of a positive definite system of equations*, Numer. Math. **8** (1966), 203–216, (republished, pp. 31–44 in [**1598**]).
1003. J. M. Martínez, *A bound for the Moore-Penrose pseudoinverse of a matrix*, Comment. Math. Univ. Carolin. **20** (1979), no. 2, 357–360.
1004. E. Martínez-Torres, *Formulation of the vibrational theory in terms of redundant internal coordinates*, Journal of Molecular Structure **520** (2000), 53–61.
1005. T. Mathew and S. K. Mitra, *Shorted operators and the identification problem—the real case*, IEEE Trans. Circuits and Systems **31** (1984), no. 3, 299–300.
1006. J. C. Maxwell, *Treatise of Electricity and Magnetism*, 3rd ed., vol. I, Oxford University Press, Oxford, 1892.
1007. D. Q. Mayne, *An algorithm for the calculation of the pseudo-inverse of a singular matrix*, Comput. J. **9** (1966), 312–317.
1008. ______, *On the calculation of pseudoinverses*, IEEE Trans. Automatic Control **AC-14** (1969), 204–205.
1009. N. H. McCoy, *Generalized regular rings*, Bull. Amer. Math. Soc. **45** (1939), 175–178.
1010. J. S. McMath and S. E. Sims, *Mean value theorems for a matrix valued derivative*, Linear and Multilinear Algebra **9** (1980/81), no. 4, 255–261.
1011. J. S. McMath, S. E. Sims, and C. R. Hallum, *On the properties of a matrix valued derivative utilizing the Moore-Penrose inverse*, Linear Algebra and its Applications **18** (1977), no. 3, 281–291, (see [**1010**]).
1012. ______, *A comparison of various derivatives for matrix-valued functions*, Comput. Math. Appl. **6** (1980), no. 2, 161–166.
1013. A. Meenakshi, *On integral EP_r matrices*, Period. Math. Hungar. **14** (1983), no. 3-4, 229–234.
1014. Ar. Meenakshi and N. Anandam, *Polynomial generalized inverses of a partitioned polynomial matrix*, J. Indian Math. Soc. (N.S.) **58** (1992), no. 1-4, 11–18.
1015. M. Meicler, *Chebyshev solution of an inconsistent system of $n+1$ linear equations in n unknowns in terms of its least squares solution*, SIAM Rev. **10** (1968), 373–375.
1016. ______, *A steepest ascent method for the Chebyshev problem*, Math. Comp. **23** (1969), 813–817.
1017. E. Meister and F. -O. Speck, *The Moore-Penrose inverse of Wiener-Hopf operators on the half axis and the quarter plane*, J. Integral Equations **9** (1985), no. 1, 45–61.
1018. G. Merz, *Über die Interpolationsaufgabe bei natürlichen Polynom-Splines mit äquidistanten Knoten*, J. Approximation Theory **10** (1974), 151–158.
1019. C. D. Meyer, Jr., *On ranks of pseudoinverses*, SIAM Rev. **11** (1969), 382–385.
1020. ______, *On the construction of solutions to the matrix equations $AX = A$ and $YA = A$*, SIAM Rev. **11** (1969), 612–615.
1021. ______, *Generalized inverses of block triangular matrices*, SIAM J. Appl. Math. **19** (1970), 741–750.
1022. ______, *Generalized inverses of triangular matrices*, SIAM J. Appl. Math. **18** (1970), 401–406.
1023. ______, *Some remarks on EP_r matrices, and generalized inverses*, Linear Algebra and its Applications **3** (1970), 275–278.
1024. ______, *Representations for* (1)- *and* (1, 2)-*inverses for partitioned matrices*, Linear Algebra and its Applications **4** (1971), 221–232.
1025. ______, *The Moore-Penrose inverse of a bordered matrix*, Linear Algebra and its Applications **5** (1972), 375–382.

1026. ______, *Generalized inverses and ranks of block matrices*, SIAM J. Appl. Math. **25** (1973), 597–602.
1027. ______, *Generalized inversion of modified matrices*, SIAM J. Appl. Math. **24** (1973), 315–323.
1028. ______, *Limits and the index of a square matrix*, SIAM J. Appl. Math. **26** (1974), 469–478, (see [**1309**]).
1029. ______, *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Rev. **17** (1975), 443–464.
1030. ______, *Analysis of finite Markov chains by group inversion techniques*, In Campbell [**267**], pp. 50–81.
1031. ______, *The character of a finite Markov chain*, Linear algebra, Markov chains, and queueing models (Minneapolis, MN, 1992), Springer, New York, 1993, pp. 47–58.
1032. C. D. Meyer, Jr. and R. J. Painter, *Note on a least squares inverse for a matrix*, J. Assoc. Comput. Mach. **17** (1970), 110–112.
1033. C. D. Meyer, Jr. and R. J. Plemmons, *Convergent powers of a matrix with applications to iterative methods for singular linear systems*, SIAM J. Numer. Anal. **14** (1977), no. 4, 699–705.
1034. C. D. Meyer, Jr. and N. J. Rose, *The index and the Drazin inverse of block triangular matrices*, SIAM J. Appl. Math. **33** (1977), no. 1, 1–7.
1035. C. D. Meyer, Jr. and J. M. Shoaf, *Updating finite Markov chains by using techniques of group matrix inversion*, J. Statist. Comput. Simulation **11** (1980), no. 3-4, 163–181.
1036. C. D. Meyer, Jr. and M. W. Stadelmaier, *Singular M-matrices and inverse positivity*, Linear Algebra and its Applications **22** (1978), 139–156.
1037. C. D. Meyer, Jr. and G. W. Stewart, *Derivatives and perturbations of eigenvectors*, SIAM J. Numer. Anal. **25** (1988), no. 3, 679–691.
1038. Jian Ming Miao, *The Moore-Penrose inverse of a rank-r modified matrix*, Numer. Math. J. Chinese Univ. **11** (1989), no. 4, 355–361.
1039. ______, *Representations for the weighted Moore-Penrose inverse of a partitioned matrix*, J. Comput. Math. **7** (1989), no. 4, 321–323.
1040. ______, *General expressions for the Moore-Penrose inverse of a 2×2 block matrix*, Linear Algebra and its Applications **151** (1991), 1–15.
1041. Jian Ming Miao and A. Ben-Israel, *On principal angles between subspaces in R^n*, Linear Algebra and its Applications **171** (1992), 81–98.
1042. ______, *Minors of the Moore-Penrose inverse*, Linear Algebra and its Applications **195** (1993), 191–207.
1043. ______, *On l_p-approximate solutions of linear equations*, Linear Algebra and its Applications **199** (1994), 305–327.
1044. ______, *The geometry of basic, approximate, and minimum-norm solutions of linear equations*, Linear Algebra and its Applications **216** (1995), 25–41.
1045. ______, *Product cosines of angles between subspaces*, Linear Algebra and its Applications **237/238** (1996), 71–81.
1046. Jian Ming Miao and D. W. Robinson, *Group and Moore-Penrose inverses of regular morphisms with kernel and cokernel*, Linear Algebra and its Applications **110** (1988), 263–270.
1047. C. Miehe and J. Schröder, *Comparative study of stress update algorithms for rate-independent and rate-dependent crystal plasticity*, Internat. J. Numer. Methods Engrg. **50** (2001), 273–298.
1048. L. Mihályffy, *A note on the matrix inversion by the partitioning technique*, Studia Sci. Math. Hungar. **5** (1970), 127–135.
1049. ______, *An alternative representation of the generalized inverse of partitioned matrices*, Linear Algebra and its Applications **4** (1971), 95–100.
1050. A. J. Miller and B. D. Cornuelle, *Forecasts from fits of frontal fluctuations*, Dynamics of Atmospheres and Oceans **29** (1999), no. 2–4, 305–333.
1051. G. A. Milliken and F. Akdeniz, *A theorem on the difference of the generalized inverses of two nonnegative matrices*, Comm. Statist.—Theory Methods **A6** (1977), no. 1, 73–79.
1052. R. D. Milne, *An oblique matrix pseudoinverse*, SIAM J. Appl. Math. **16** (1968), 931–944.
1053. H. W. Milnes, J. Amburgey, T. O. Lewis, and T. L. Boullion, *Spectral eigenvalue property of A^+ for rectangular matrices*, In Boullion and Odell [**207**], pp. 98–113.
1054. N. Minamide and K. Nakamura, *Minimum error control problem in banach space*, Research Report of Automatic Control Lab 16, Nagoya University, Nagoya, Japan, 1969.
1055. ______, *A restricted pseudoinverse and its applications to constrained minima*, SIAM J. Appl. Math. **19** (1970), 167–177.
1056. L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford **11** (1960), 50–59.
1057. S. K. Mitra, *A new class of g–inverse of square matrices*, Sankhyā Ser. A **30** (1968), 323–330.
1058. ______, *On a generalized inverse of a matrix and applications*, Sankhyā Ser. A **30** (1968), 107–114.
1059. ______, *Fixed rank solutions of linear matrix equations*, Sankhyā Ser. A **35** (1972), 387–392.
1060. ______, *Common solutions to a pair of linear matrix equations $A_1XB_1 = C_1$, $A_2XB_2 = C_2$*, Proceedings of the Cambridge Philosophical Society **74** (1973), 213–216.
1061. ______, *Shorted operators and the identification problem*, IEEE Trans. Circuits and Systems **29** (1982), no. 8, 581–583.
1062. ______, *Shorted matrices in star and related orderings*, Circuits Systems Signal Process. **9** (1990), no. 2, 197–212.
1063. ______, *Matrix partial order through generalized inverses: unified theory*, Linear Algebra and its Applications **148** (1991), 237–263.

1064. ______, *Separation theorems*, Linear Algebra and its Applications **208/209** (1994), 239–256.
1065. ______, *Erratum: "The nonunique shorted matrix" [Linear Algebra Appl.* **237/238** *(1996), 41–70; MR 97a:15012] by Mitra and K. M. Prasad*, Linear Algebra and its Applications **260** (1997), 323.
1066. ______, *Diagrammatic presentation of inner and outer inverses: S-diagrams*, Linear Algebra and its Applications **287** (1999), no. 1-3, 271–288.
1067. S. K. Mitra and P. Bhimasankaram, *Generalized inverses of partitioned matrices and recalculation of least squares estimates for data or model changes*, Sankhyā Ser. A **33** (1971), 395–410.
1068. S. K. Mitra and R. E. Hartwig, *Partial orders based on outer inverses*, Linear Algebra and its Applications **176** (1992), 3–20.
1069. S. K. Mitra and P. L. Odell, *On parallel summability of matrices*, Linear Algebra and its Applications **74** (1986), 239–255.
1070. S. K. Mitra and K. M. Prasad, *The nonunique shorted matrix*, Linear Algebra and its Applications **237/238** (1996), 41–70, (erratum in [**1065**]).
1071. ______, *The nonunique parallel sum*, Linear Algebra and its Applications **259** (1997), 77–99.
1072. S. K. Mitra and M. L. Puri, *Shorted operators and generalized inverses of matrices*, Linear Algebra and its Applications **25** (1979), 45–56.
1073. ______, *Shorted matrices—an extended concept and some applications*, Linear Algebra and its Applications **42** (1982), 57–79.
1074. ______, *The fundamental bordered matrix of linear estimation and the Duffin-Morley general linear electromechanical systems*, Applicable Anal. **14** (1983), no. 4, 241–258.
1075. S. K. Mitra and C. R. Rao, *Some results in estimation and tests of linear hypotheses under the Gauss–Markov model*, Sankhyā Ser. A **30** (1968), 281–290.
1076. ______, *Conditions for optimality and valididty of simple least squares theory*, Ann. Math. Statist. **40** (1969), 1617–1624.
1077. ______, *Projections under seminorms and generalized Moore Penrose inverses*, Linear Algebra and its Applications **9** (1974), 155–167.
1078. S. Mohideen and V. Cherkassky, *On recursive calculation of the generalized inverse of a matrix*, ACM Trans. Math. Software **17** (1991), no. 1, 130–147.
1079. C. B. Moler, *Iterative refinement in floating point*, J. Assoc. Comput. Mach. **14** (1967), 316–321.
1080. W. Mönch, *Monotone Einschliessung der Moore-Penrose Pseudoinversen einer Matrix*, Z. Angew. Math. Mech. **58** (1978), no. 2, 67–74.
1081. ______, *Iterative refinement of approximations to a generalized inverse of a matrix*, Computing **28** (1982), no. 1, 79–87.
1082. B. Mond, *Generalized inverse extensions of matrix inequalities*, Linear Algebra and its Applications **2** (1969), 393–399.
1083. B. Mond and J. E. Pečarić, *Inequalities with weights for powers of generalised inverses*, Bull. Austral. Math. Soc. **48** (1993), no. 1, 7–12.
1084. ______, *Inequalities with weights for powers of generalized inverses. II*, Linear Algebra and its Applications **210** (1994), 265–272.
1085. ______, *On matrix convexity of the Moore-Penrose inverse*, Internat. J. Math. Math. Sci. **19** (1996), no. 4, 707–710.
1086. ______, *Hadamard products and generalized inverses*, Austral. Math. Soc. Gaz. **25** (1998), no. 4, 194–197.
1087. E. H. Moore, *On the reciprocal of the general algebraic matrix*, Bull. Amer. Math. Soc. **26** (1920), 394–395, (Abstract).
1088. ______, *General Analysis*, Memoirs of the American Philosophical Society, I, American Philosophical Society, Philadelphia, Pennsylvania, 1935.
1089. R. H. Moore and M. Z. Nashed, *Approximations of generalized inverses of linear operators in Banach spaces*, Approximation theory (Proc. Internat. Sympos., Univ. Texas, Austin, Tex., 1973), Academic Press, New York, 1973, pp. 425–428.
1090. ______, *Approximations to generalized inverses of linear operators*, SIAM J. Appl. Math. **27** (1974), 1–16.
1091. C. F. Moppert, *On the Gram determinant*, Quart. J. Math. Oxford Ser. (2) **10** (1959), 161–164.
1092. T. D. Morley, *A Gauss-Markov theorem for infinite-dimensional regression models with possibly singular covariance*, SIAM J. Appl. Math. **37** (1979), no. 2, 257–260, (the uniqueness claim in Corollary 3 is incorrect).
1093. ______, *Parallel summation, Maxwell's principle and the infimum of projections*, J. Math. Anal. Appl. **70** (1979), no. 1, 33–41.
1094. ______, *An alternative approach to the parallel sum*, Adv. in Appl. Math. **10** (1989), no. 3, 358–369.
1095. G. L. Morris and P. L. Odell, *A characterization for generalized inverses of matrices*, SIAM Rev. **10** (1968), 208–211.
1096. ______, *Common solutions for n matrix equations with applications*, J. Assoc. Comput. Mach. **15** (1968), 272–274.
1097. L. Moura and R. Kitney, *A direct method for least-squares circle fitting*, Comput. Phys. Comm. **64** (1991), no. 1, 57–63, (see [**1140**]).
1098. W. D. Munn, *Pseudoinverses in semigroups*, Proc. Cambridge Phil. Soc. **57** (1961), 247–250.
1099. ______, *Moore–Penrose inversion in complex contracted inverse semigroup algebras*, J. Austral. Math. Soc. Ser. A **66** (1999), no. 3, 297–302.
1100. W. D. Munn and R. Penrose, *A note on inverse semigroups*, Proceedings of the Cambridge Philosophical Society **51** (1955), 396–399.
1101. K. Murakami and T. Aibara, *An improvement on the Moore-Penrose generalized inverse associative*

memory, IEEE Trans. Systems Man Cybernet. **17** (1987), no. 4, 699–707.

1102. F. D. Murnaghan and A. Wintner, *A canonical form for real matrices under orthogonal transformation*, Proc. Nat. Acad. Sci. U.S.A. **17** (1931), 417–420.
1103. F. J. Murray and J. von Neumann, *On rings of operators*, Ann. of Math. **37** (1936), 116–229.
1104. K. V. V. Murthy and V. Ramachandran, *On the ranks of certain matrices*, Matrix Tensor Quart. **31** (1980/81), no. 3, 67–69.
1105. K. N. Murty and P. V. S. Lakshmi, *On two-point boundary value problems*, J. Math. Anal. Appl. **153** (1990), no. 1, 217–225.
1106. M. S. N. Murty and B. V. Appa Rao, *Application of Moore-Penrose inverse to three point boundary value problems*, Ranchi Univ. Math. J. **29** (1998), 1–9 (1999).
1107. M. S. N. Murty and D. R. K. S. Rao, *Application of generalized inverses to two-point boundary value problems*, Bull. Inst. Math. Acad. Sinica **19** (1991), no. 3, 271–278.
1108. S. Nakagiri and K. Suzuki, *Finite element interval analysis of external loads identified by displacement input with uncertainty*, Computer Methods in Applied Mechanics and Engineering **168** (1999), no. 1–4, 63–72.
1109. K. S. Subramonian Nambooripad, *The natural partial order on a regular semigroup*, Proc. Edinburgh Math. Soc. (2) **23** (1980), no. 3, 249–260.
1110. V. C. Nanda, *A generalization of Cayley's theorem*, Math. Z. **101** (1967), 331–334.
1111. J. C. Nash, *Compact Numerical Methods for Computers*, second ed., Adam Hilger Ltd., Bristol, 1990.
1112. J. C. Nash and R. L. C. Wang, *Algorithm 645: subroutines for testing programs that compute the generalized inverse of a matrix*, ACM Trans. Math. Software **12** (1986), no. 3, 274–277.
1113. S. G. Nash, *Newton-type minimization via the Lánczos method*, SIAM J. Numer. Anal. **21** (1984), no. 4, 770–788.
1114. M. Z. Nashed, *Steepest descent for singular linear operator equations*, SIAM J. Numer. Anal. **7** (1970), 358–362.
1115. ———, *Generalized inverses, normal solvability, and iteration for singular operator equations*, Nonlinear Functional Anal. and Appl. (Proc. Advanced Sem., Math. Res. Center, Univ. of Wisconsin, Madison, Wis., 1970), Academic Press, New York, 1971, pp. 311–359.
1116. M. Z. Nashed (ed.), *Generalized Inverses and Applications (Proc. Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1973)*, New York, Academic Press, 1976.
1117. ———, *On moment-discretization and least-squares solutions of linear integral equations of the first kind*, J. Math. Anal. Appl. **53** (1976), no. 2, 359–366.
1118. ———, *On the perturbation theory for generalized inverse operators in Banach spaces*, Functional analysis methods in numerical analysis (Proc. Special Session, Annual Meeting, Amer. Math. Soc., St. Louis, Mo., 1977), Springer, Berlin, 1979, pp. 180–195.
1119. ———, *On generalized inverses and operator ranges*, Functional analysis and approximation (Oberwolfach, 1980), Birkhäuser, Basel, 1981, pp. 85–96.
1120. M. Z. Nashed and X. Chen, *Convergence of Newton-like methods for singular operator equations using outer inverses*, Numer. Math. **66** (1993), 235–257.
1121. M. Z. Nashed and L. B. Rall, *Annotated bibliography on generalized inverses and applications*, In Nashed [**1116**], pp. 771–1041.
1122. M. Z. Nashed and G. F. Votruba, *A unified approach to generalized inverses of linear operators. I. Algebraic, topological and projectional properties*, Bull. Amer. Math. Soc. **80** (1974), 825–830.
1123. ———, *A unified approach to generalized inverses of linear operators. II. Extremal and proximal properties*, Bull. Amer. Math. Soc. **80** (1974), 831–835.
1124. ———, *A unified operator theory of generalized inverses*, In Nashed [**1116**], pp. 1–109.
1125. M. Z. Nashed and G. Wahba, *Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind*, Math. Comp. **28** (1974), 69–80.
1126. ———, *Generalized inverses in reproducing kernel spaces: an approach to regularization of linear operator equations*, SIAM J. Math. Anal. **5** (1974), 974–987.
1127. M. Z. Nashed and Ya Gu Zhao, *The Drazin inverse for singular evolution equations and partial differential operators*, Recent trends in differential equations, World Sci. Publishing, River Edge, NJ, 1992, pp. 441–456.
1128. D. L. Nelson, T. O. Lewis, and T. L. Boullion, *A quadratic programming technique using matrix pseudoinverses*, Indust. Math. **21** (1971), 1–21.
1129. H. Neudecker, *A note on Kronecker matrix products and matrix equation systems*, SIAM J. Appl. Math. **17** (1969), 603–606.
1130. ———, *Mathematical properties of the variance of the multinomial distribution*, J. Math. Anal. Appl. **189** (1995), no. 3, 757–762, (see [**1426**]).
1131. H. Neudecker and Shuangzhe Liu, *The density of the Moore-Penrose inverse of a random matrix*, Linear Algebra and its Applications **237/238** (1996), 123–126.
1132. A. Neumaier, *Hybrid norms and bounds for overdetermined linear systems*, Linear Algebra and its Applications **216** (1995), 257–265.
1133. ———, *Solving ill-conditioned and singular linear systems: a tutorial on regularization*, SIAM Rev. **40** (1998), no. 3, 636–666 (electronic).

1134. M. Neumann, *On the Schur complement and the LU-factorization of a matrix*, Linear and Multilinear Algebra **9** (1980/81), no. 4, 241–254.
1135. M. Neumann and R. J. Plemmons, *Generalized inverse-positivity and splittings of M-matrices*, Linear Algebra and its Applications **23** (1979), 21–35.
1136. M. Neumann, G. D. Poole, and H. -J. Werner, *More on generalizations of matrix monotonicity*, Linear Algebra and its Applications **48** (1982), 413–435.
1137. M. Neumann and H. -J. Werner, *Nonnegative group inverses*, Linear Algebra and its Applications **151** (1991), 85–96.
1138. T. G. Newman, M. Meicler, and P. L. Odell, *On the concept of a $p-q$ generalized inverse of a matrix*, In Boullion and Odell [**207**], pp. 276–282.
1139. T. G. Newman and P. L. Odell, *On the concept of a $p-q$ generalized inverse of a matrix*, SIAM J. Appl. Math. **17** (1969), 520–525.
1140. Y. Nievergelt, *Computing circles and spheres of arithmetic least squares*, Comput. Phys. Comm. **81** (1994), no. 3, 343–350, (extension of [**1097**]).
1141. ———, *Schmidt-Mirsky matrix approximation with linearly constrained singular values*, Linear Algebra and its Applications **261** (1997), 207–219.
1142. ———, *A tutorial history of least squares with applications to astronomy and geodesy*, J. Comput. Appl. Math. **121** (2000), no. 1-2, 37–72.
1143. Yves Nievergelt, *Total least squares: state-of-the-art regression in numerical analysis*, SIAM Rev. **36** (1994), no. 2, 258–264.
1144. B. Noble, *A method for computing the generalized invrerse of a matrix*, SIAM J. Numer. Anal. **3** (1966), 582–584.
1145. ———, *Applied linear algebra*, Prentice–Hall, Inc., Englewood Cliffs, NJ, 1969.
1146. ———, *Methods for computing the Moore-Penrose generalized inverse, and related matters*, Generalized inverses and applications (Proc. Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1973), Academic Press, New York, 1976, pp. 245–301. Publ. Math. Res. Center Univ. Wisconsin, No. 32.
1147. K. Nordström, *Some further aspects of the Löwner-ordering antitonicity of the Moore-Penrose inverse*, Comm. Statist. Theory Methods **18** (1989), no. 12, 4471–4489 (1990).
1148. P. L. Odell and T. L. Boullion, *Simultaneous diagonalization of rectangular matrices*, Comput. Math. Appl. **33** (1997), no. 9, 93–96.
1149. P. L. Odell and H. P. Decell, Jr., *On computing the fixed-point probability vector of ergodic transition matrices*, J. Assoc. Comput. Mach. **14** (1967), 765–768.
1150. W. Oktaba, *Tests of hypotheses for the fixed model not of full rank*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **16** (1968), 409–413.
1151. D. P. O'Leary, *On bounds for scaled projections and pseudo-inverses*, Linear Algebra and its Applications **132** (1990), 115–117, (Answer to a question in [**1404**]).
1152. I. Olkin, *The density of the inverse and pseudo-inverse of a random matrix*, Statist. Probab. Lett. **38** (1998), no. 2, 131–135, (alternative proof to [**1631**]).
1153. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.
1154. E. E. Osborne, *On least squares solutions of linear equations*, J. Assoc. Comput. Mach. **8** (1961), 628–636.
1155. ———, *Smallest least squares solutions of linear equations*, SIAM J. Numer. Anal. **2** (1965), 300–307.
1156. M. R. Osborne, *A class of methods for minimising a sum of squares*, Austral. Comput. J. **4** (1972), 164–169.
1157. A. Ostrowski, *A new proof of Haynsworth's quotient formula for Schur complements.*, Linear Algebra and its Applications **4** (1971), 389–392.
1158. D. V. Ouellette, *Schur complements and statistics*, Linear Algebra and its Applications **36** (1981), 187–295.
1159. Ping Qi Pan and Zi Xiang Ouyang, *Moore–Penrose inverse simplex algorithms based on successive linear subprogramming approach*, Numer. Math. J. Chinese Univ. (English Ser.) **3** (1994), no. 2, 180–190.
1160. V. Pan and R. S. R. Schreiber, *An improved newton iteration for the generalized inverse of a matrix with applications*, SIAM J. Sci. Statist. Comput. **12** (1991), 1109–1130.
1161. B. N. Parlett, *The LU and QR algorithms*, In Ralston and Wilf [**1238**], pp. 116–130, Vol. II.
1162. K. H. Parshall and D. E. Rowe, *The emergence of the American mathematical research community, 1876–1900: J. J. Sylvester, Felix Klein, and E. H. Moore*, American Mathematical Society, Providence, RI, 1994.
1163. S. Pati, *Moore–Penrose inverse of matrices on idempotent semirings*, SIAM J. Matrix Anal. Appl. **22** (2000), no. 2, 617–626 (electronic).
1164. M. Pavel-Parvu and A. Korganoff, *Iteration functions for solving polynomial equations*, Constructive Aspects of the Fundamental Theorem of Algebra (New York) (B. Dejon and P. Henrici, eds.), John Wiley, 1969.
1165. G. Peano, *Intégration par séries des équations différentielles linéaires*, Mathematische Annallen **32** (1888), 450–456.
1166. M. H. Pearl, *On Cayley's parametrization*, Canad. J. Math. **9** (1957), 553–562.
1167. ———, *A further extension of Cayley's parametrization*, Canad. J. Math. **11** (1959), 48–50.
1168. ———, *On normal and EP_r matrices*, Michigan Math. J. **6** (1959), 1–5.

1169. ______, *On normal EP_r matrices*, Michigan Math. J. **8** (1961), 33–37.
1170. ______, *On generalized inverses of matrices*, Proc. Cambridge Philos. Soc. **62** (1966), 673–677.
1171. ______, *A decomposition theorem for matrices*, Canad. J. Math. **19** (1967), 344–349, (See [**1447**], [**1593**]).
1172. ______, *Generalized inverses of matrices with entries taken from an arbitrary field*, Linear Algebra and its Applications **1** (1968), 571–587, (see also [**1272**]).
1173. ______, *Automorphic transformations of an arbitrary matrix*, Linear and Multilinear Algebra **15** (1984), no. 3-4, 245–256.
1174. K. Pearson, *On lines and planes of closest fit to points in space*, Philosophical Magazine **2** (1901), 559–572.
1175. R. Peluso and G. Piazza, *Bounds for products of singular values of a matrix*, Rend. Mat. Appl. (7) **19** (1999), no. 4, 507–522 (2000).
1176. R. H. Pennington, *Introductory Computer Methods and Numerical Analysis*, MacMillan Co., New York, 1970.
1177. R. Penrose, *A generalized inverse for matrices*, Proceedings of the Cambridge Philosophical Society **51** (1955), 406–413.
1178. ______, *On best approximate solutions of linear matrix equations*, Proceedings of the Cambridge Philosophical Society **52** (1956), 17–19.
1179. V. Pereyra, *Iterative methods for solving nonlinear least squares problems*, SIAM J. Numer. Anal. **4** (1967), 27–36.
1180. ______, *Stability of general systems of linear equations*, Aequationes Mathematicae **2** (1969), 194–206.
1181. V. Pereyra and J. B. Rosen, *Computation of the pseudoinverse of a matrix of unknown rank*, Tech. Report CS 13, Dept. of Computer Science, Stanford University, Stanford, CA, Sept. 1964, (*Comp. Rev.***6**(1965), 259 #7948).
1182. G. Peters and J. H. Wilkinson, *The least squares problem and pseudo-inverses*, Comput. J. **13** (1970), 309–316.
1183. W. V. Petryshyn, *On generalized inverses and on the uniform convergence of $(I - \beta K)^n$ with application to iterative methods*, J. Math. Anal. Appl. **18** (1967), 417–439.
1184. R. Picón, *A systematic deduction of linear natural approach equations*, Computer Methods in Applied Mechanics and Engineering **177** (1999), 137–151.
1185. W. H. Pierce, *A self-correcting matrix iteration for the Moore-Penrose generalized inverse*, Linear Algebra and its Applications **244** (1996), 357–363.
1186. A. Pietsch, *Zur Theorie der σ–Transformationen in lokalkonvexen Vektorräumen*, Math. Nach. **21** (1960), 347–369.
1187. R. Piziak, P. L. Odell, and R. Hahn, *Constructing projections on sums and intersections*, Comput. Math. Appl. **37** (1999), no. 1, 67–74.
1188. R. L. Plackett, *Some theorems in least squares*, Biometrika **37** (1950), 149–157.
1189. R. J. Plemmons, *Generalized inverses of Boolean relation matrices*, SIAM J. Appl. Math. **20** (1971), 426–433.
1190. R. J. Plemmons and R. E. Cline, *The generalized inverse of a nonnegative matrix*, Proc. Amer. Math. Soc. **31** (1972), 46–50, (erratum, *ibid* **39**(1972), 651).
1191. Olga Pokorná, *On the pseudoinversion of matrix products*, Numerical methods (Third Colloq., Keszthely, 1977), North-Holland, Amsterdam, 1980, pp. 501–506.
1192. W. Pölzleitner and H. Wechsler, *Selective and focused invariant recognition using distributed associative memories (DAM)*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), no. 8, 809–814.
1193. G. D. Poole and G. P. Barker, *MP matrices*, Linear Algebra and its Applications **26** (1979), 165–174.
1194. G. D. Poole and T. L. Boullion, *The Drazin inverse for certain power matrices*, Indust. Math. **22** (1972), 35–37.
1195. ______, *Weak spectral inverses which are partial isometries*, SIAM J. Appl. Math. **23** (1972), 171–172.
1196. ______, *The Drazin inverse and a spectral inequality of Marcus, Minc, and Moyls*, J. Optimization Theory Appl. **15** (1975), 503–508.
1197. W. A. Porter, *Modern Foundations of System Engineering*, MacMillan, New York, 1966.
1198. ______, *A basic optimization problem in linear systems*, Math. Syst. Th. **5** (1971), 20–44.
1199. W. A. Porter and J. P. Williams, *Extension of the minimum effort control problem*, J. Math. Anal. Appl. **13** (1966), 536–549.
1200. ______, *A note on the minimum effort control problem*, J. Math. Anal. Appl. **13** (1966), 251–264.
1201. K. M. Prasad, *Generalized inverses of matrices over commutative rings*, In Bapat et al. [**80**], (special issue of *Linear Algebra and its Applications* **211** (1994)), pp. 35–52.
1202. K. M. Prasad and R. B. Bapat, *The generalized Moore-Penrose inverse*, Linear Algebra and its Applications **165** (1992), 59–69.
1203. ______, *A note on the Khatri–inverse*, Sankhyā Ser. A **54** (1992), no. 2, 291–295.
1204. K. M. Prasad, K. P. S. Bhaskara Rao, and R. B. Bapat, *Generalized inverses over integral domains. II. Group inverses and Drazin inverses*, Linear Algebra and its Applications **146** (1991), 31–47.
1205. G. B. Preston, *Inverse semi-groups*, J. London Math. Soc. **29** (1954), 396–403.
1206. C. M. Price, *The matrix pseudoinverse and minimal variance estimates*, SIAM Rev. **6** (1964), 115–120.

1207. R. M. Pringle and A. A. Rayner, *Expressions for generalized inverses of a bordered matrix with application to the theory of constrained linear models*, SIAM Rev. **12** (1970), 107–115.

1208. ______, *Generalized Inverse Matrices with Applications to Statistics. griffin's statistical monographs and courses, no. 28*, Hafner Publishing Co., New York, 1971.

1209. D. Przeworska-Rolewicz and S. Rolewicz, *Equations in Linear Spaces*, Polska Akad. Nauk Monog. Mat., vol. 47, PWN Polish Scientific Publishers, Warsaw, 1968.

1210. A. C. Pugh and Liansheng Tan, *A generalized chain-scattering representation and its algebraic system properties*, IEEE Trans. Automat. Control **45** (2000), no. 5, 1002–1007.

1211. P. Pulay and G. Fogarasi, *Geometry optimization in redundant internal coordinates*, The Journal of Chemical Physics **96** (1992), no. 4, 2856–2860, (see [**1004**]).

1212. S. Puntanen and G. P. H. Styan, *The equality of the ordinary least squares estimator and the best linear unbiased estimator (with comments by oscar kempthorne and shayle r. searle and a reply by the authors)*, Amer. Statist. **43** (1989), no. 3, 153–164.

1213. S. Puntanen, G. P. H. Styan, and H. -J. Werner, *Two matrix-based proofs that the linear estimator Gy is the best linear unbiased estimator*, J. Statist. Plann. Inference **88** (2000), no. 2, 173–179.

1214. M. L. Puri, C. T. Russell, and T. Mathew, *Convergence of generalized inverses with applications to asymptotic hypothesis testing*, Sankhyā Ser. A **46** (1984), no. 2, 277–286.

1215. R. Puystjens, *Moore-Penrose inverses for matrices over some Noetherian rings*, J. Pure Appl. Algebra **31** (1984), no. 1-3, 191–198.

1216. ______, *Some aspects of generalized invertibility*, Bull. Soc. Math. Belg. Sér. A **40** (1988), no. 1, 67–72.

1217. R. Puystjens and H. de Smet, *The Moore-Penrose inverse for matrices over skew polynomial rings*, Ring theory, Antwerp 1980 (Proc. Conf., Univ. Antwerp, Antwerp, 1980), Springer, Berlin, 1980, pp. 94–103.

1218. R. Puystjens and R. E. Hartwig, *The group inverse of a companion matrix*, Linear and Multilinear Algebra **43** (1997), no. 1-3, 137–150.

1219. R. Puystjens and D. W. Robinson, *The Moore-Penrose inverse of a morphism with factorization*, Linear Algebra and its Applications **40** (1981), 129–141.

1220. ______, *The Moore-Penrose inverse of a morphism in an additive category*, Comm. Algebra **12** (1984), no. 3-4, 287–299, (see [**640**]).

1221. ______, *Symmetric morphisms and the existence of Moore-Penrose inverses*, Linear Algebra and its Applications **131** (1990), 51–69.

1222. L. D. Pyle, *Generalized inverse computations using the gradient projection method*, J. Assoc. Comput. Mach. **11** (1964), 422–428.

1223. ______, *A generalized inverse ε-algorithm for constructing intersection projection matrices, with applications*, Numer. Math. **10** (1967), 86–102.

1224. ______, *The generalized inverse in linear programming. Basic structure*, SIAM J. Appl. Math. **22** (1972), 335–355.

1225. L. D. Pyle and R. E. Cline, *The generalized inverse in linear programming—interior gradient projection methods*, SIAM. J. Appl. Math. **24** (1973), 511–534.

1226. P. Rabinowitz (editor), *Numerical Methods for Nonlinear Algebraic Equations*, Gordon and Breach, London, 1970.

1227. G. Rabson, *The generalized inverse in set theory and matrix theory*, Tech. report, Dept. of Mathematics, Clarkson College of Technology, Potsdam, NY, 1969.

1228. M. Radić, *A definition of the determinant of a rectangular matrix*, Glasnik Mat. Ser. III **1 (21)** (1966), 17–22.

1229. ______, *On a generalization of the Arghiriade-Dragomir representation of the Moore-Penrose inverse*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **44** (1968), 333–336.

1230. R. Rado, *Note on generalized inverses of matrices*, Proceedings of the Cambridge Philosophical Society **52** (1956), 600–601.

1231. M. Abdur Rahman, *Representation of Moore-Penrose generalized inverse of the product of two matrices*, Ganit. Journal of Bangladesh Mathematical Society **13** (1993), no. 1-2, 83–85.

1232. V. Rakočević, *Moore–Penrose inverse in Banach algebras*, Proc. Roy. Irish Acad. Sect. A **88** (1988), no. 1, 57–60.

1233. ______, *On the continuity of the Moore-Penrose inverse in Banach algebras*, Facta Univ. Ser. Math. Inform. (1991), no. 6, 133–138.

1234. ______, *Continuity of the Drazin inverse*, J. Operator Theory **41** (1999), no. 1, 55–68.

1235. V. Rakočević and Yimin Wei, *The perturbation theory for the Drazin inverse and its applications II*, J. Austral. Math. Soc. **70** (2001), no. 2, 189–197.

1236. L. B. Rall, *Comutational Solution of Nonlinear Operator Equations*, Wiley, New York, 1969.

1237. A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis (2nd edition)*, McGraw-Hill, New York, 1978.

1238. A. Ralston and H. Wilf (eds.), *Mathematical Methods for Digital Computers*, New York, Wiley, 1967.

1239. J. H. Randall and A. A. Rayner, *The accuracy of least squares calculations with the Cholesky algorithm*, Linear Algebra and its Applications **127** (1990), 463–502.

1240. C. R. Rao, *A note on a generalized inverse of a matrix with applications to problems in mathematical statistics*, J. Roy. Statist. soc. Ser. B **24** (1962), 152–158.

1241. ———, *Linear Statistical Inference and its Applications*, J. Wiley & Sons, New York, 1965, (Second edition, 1973).

1242. ———, *Generalized inverse for matrices and its applications in mathematical statistics*, Research Papers in Statistics (Festschrift J. Neyman), Wiley, London, 1966, pp. 263–279.

1243. ———, *Calculus of generalized inverses of matrices, part I: General theory*, Sankhyā Ser. A **29** (1967), 317–342.

1244. ———, *Unified theory of linear estimation*, Sankhyā Ser. A **33** (1971), 371–394.

1245. ———, *Corrigenda: "Unified theory of linear estimation" (Sankhyā Ser. A* **33** *(1971), 371–394)*, Sankhyā Ser. A **34** (1972), 477.

1246. ———, *Corrigendum: "Unified theory of linear estimation" (Sankhyā Ser. A* **33** *(1971), 371–394)*, Sankhyā Ser. A **34** (1972), 194.

1247. ———, *Projectors, generalized inverses and the* blue*'s*, J. Roy. Statist. Soc. Ser. B **36** (1974), 442–448.

1248. ———, *A lemma on g-inverse of a matrix and computation of correlation coeffiecients in the singular case*, Communications in Statistics (A) **10** (1981), 1–10.

1249. ———, *Linear transformations, projection operators and generalized inverses: a geometric approach*, Contributions to stochastics, Wiley, New York, 1992, pp. 1–10.

1250. C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*, John Wiley, New York, 1971.

1251. ———, *Theory and application of constrained inverse of matrices*, SIAM J. Appl. Math. **24** (1973), 473–488.

1252. C. R. Rao, S. K. Mitra, and P. Bhimasankaram, *Determination of a matrix by its subclasses of generalized inverses*, Sankhyā Ser. A **34** (1972), 5–8.

1253. C. R. Rao and H. Yanai, *General definition and decomposition of projectors and some applications to statistical problems*, J. Statist. Plann. Inference **3** (1979), 1–17.

1254. ———, *Generalized inverse of linear transformations: a geometric approach*, Linear Algebra and its Applications **66** (1985), 87–98.

1255. ———, *Generalized inverses of partitioned matrices useful in statistical applications*, Linear Algebra and its Applications **70** (1985), 105–113.

1256. K. P. S. Bhaskara Rao, *On generalized inverses of matrices over principal ideal rings*, Linear and Multilinear Algebra **10** (1981), no. 2, 145–154.

1257. ———, *On generalized inverses of matrices over integral domains*, Linear Algebra and its Applications **49** (1983), 179–189.

1258. A. A. Rayner and R. M. Pringle, *A note on generalized inverses in the linear hypothesis not of full rank*, Ann. Math. Statist. **38** (1967), 271–273.

1259. ———, *Some aspects of the solution of singular normal equations with the use of linear restrictions*, SIAM J. Appl. Math. **31** (1976), no. 3, 449–460, (erratum, *ibid* **47**(1987), 1130).

1260. W. T. Reid, *Generalized Green's matrices for compatible systems of differential equations*, Amer. J. Math. **53** (1931), 443–459.

1261. ———, *Principal solutions of non–oscillatory linear differential systems*, J. Math. Anal. Appl. **9** (1964), 397–423.

1262. ———, *Generalized Green's matrices for two–point boundary problems*, SIAM J. Appl. Math. **15** (1967), 856–870.

1263. ———, *Generalized inverses of differential and integral operators*, In Boullion and Odell [**207**], pp. 1–25.

1264. ———, *Ordinary Differential Equations*, Wiley-Interscience, New York, 1970.

1265. ———, *Generalized polar coordinate transformations for differential systems.*, Rocky Mountain J. Math. **1** (1971), no. 2, 383–406.

1266. ———, *A result on the singularities of matrix functions*, Quart. Appl. Math. **35** (1977/78), no. 2, 293–296.

1267. B. C. Rennie, *Letter to the editor: "Rank factorization of a matrix and its applications" [Math. Sci.* **13** *(1988), no. 1, 4–14; MR 90a:15009a] by P. Bhimasankaram*, Math. Sci. **13** (1988), no. 2, 152, (see [**168**]).

1268. W. C. Rheinboldt, *A unified convergence theory for a class of iterative processes*, SIAM J. Numer. Anal. **5** (1968), 42–63.

1269. O. M. Ribits′ka, *A fractional-analytic method of finding Moore-Penrose and Drasin pseudo-inverse matrices*, Mat. Metodi Fiz.-Mekh. Polya **39** (1996), no. 2, 140–143.

1270. J. Rice, *Experiments on gram–schmidt orthogonalization*, Math. Comput. **20** (1966), 325–328.

1271. M. Q. Rieck, *Totally isotropic subspaces, complementary subspaces, and generalized inverses*, Linear Algebra and its Applications **251** (1997), 239–248, (extension of a result of [**1172**]).

1272. ———, *Maximal orthogonality and pseudo-orthogonality with applications to generalized inverses*, Linear Algebra and its Applications **315** (2000), no. 1-3, 155–173.

1273. K. S. Riedel, *A Sherman-Morrison-Woodbury identity for rank augmenting matrices with application to centering*, SIAM J. Matrix Anal. Appl. **13** (1992), no. 2, 659–662, (see [**492**]).

1274. R. F. Rinehart, *The equivalence of definitions of a matric function*, Amer. Math. Monthly **62** (1955), 395–414.

1275. W. Rising, *Applications of generalized inverses to Markov chains*, Adv. in Appl. Probab. **23** (1991), 293–302.

1276. P. D. Robers and A. Ben-Israel, *An interval programming algorithm for discrete linear L_1 approximation problems*, J. Approximation Theory **2** (1969), 323–336.
1277. ———, *A suboptimization method for interval linear programming: A new method for linear programming*, Linear Algebra and its Applications **3** (1970), 383–405.
1278. P. Robert, *On the group-inverse of a linear transformation*, J. Math. Anal. Appl. **22** (1968), 658–669.
1279. D. W. Robinson, *A proof of the composite function theorem for matric functions*, Amer. Math. Monthly **64** (1957), 34–35.
1280. ———, *On the genralized inverse of an arbitrary linear transformation*, Amer. Math. Monthly **69** (1962), 412–416.
1281. ———, *Gauss and generalized inverses*, Historia Mathematica **7** (1980), 118–125.
1282. ———, *On the covariance of the Moore-Penrose inverse*, Linear Algebra and its Applications **61** (1984), 91–99.
1283. ———, *Covariance of Moore-Penrose inverses with respect to an invertible matrix*, Linear Algebra and its Applications **71** (1985), 275–281.
1284. ———, *Nullities of submatrices of the Moore-Penrose inverse*, Linear Algebra and its Applications **94** (1987), 127–132.
1285. ———, *The determinantal rank idempotents of a matrix*, Linear Algebra and its Applications **237/238** (1996), 83–96.
1286. ———, *The image of the adjoint mapping*, Linear Algebra and its Applications **277** (1998), no. 1-3, 143–148.
1287. ———, *Separation of subspaces by volume*, Amer. Math. Monthly **105** (1998), no. 1, 22–27.
1288. D. W. Robinson and R. Puystjens, *EP morphisms*, Linear Algebra and its Applications **64** (1985), 157–174.
1289. ———, *Generalized inverses of morphisms with kernels*, Linear Algebra and its Applications **96** (1987), 65–85.
1290. D. W. Robinson, R. Puystjens, and J. Van Geel, *Categories of matrices with only obvious Moore-Penrose inverses*, Linear Algebra and its Applications **97** (1987), 93–102.
1291. S. M. Robinson, *A short proof of Cramer's rule*, Math. Mag. **43** (1977), 94–95, (Reprinted in *Selected Papers on Algebra* (S. Montgomery et al, editors), Math. Assoc. of Amer., 1977, pp. 313–314).
1292. S. Roch and B. Silbermann, *Asymptotic Moore-Penrose invertibility of singular integral operators*, Integral Equations Operator Theory **26** (1996), no. 1, 81–101.
1293. ———, *Continuity of generalized inverses in Banach algebras*, Studia Math. **136** (1999), no. 3, 197–227.
1294. Steffen Roch and Bernd Silbermann, *Index calculus for approximation methods and singular value decomposition*, J. Math. Anal. Appl. **225** (1998), no. 2, 401–426.
1295. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
1296. C. A. Rohde, *Contributions to the theory, computation and application of generalized inverses*, Ph.d., University of North Carolina, Raleigh, N.C., May 1964.
1297. ———, *Generalized inverses of partitioned matrices*, J. Soc. Indust. Appl. Math. **13** (1965), 1033–1035.
1298. ———, *Some results on generalized inverses*, SIAM Rev. **8** (1966), 201–205.
1299. ———, *Special applications of the theory of generalized matrix inversion to statistics*, In Boullion and Odell [**207**], pp. 239–266.
1300. C. A. Rohde and J. R. Harvey, *Unified least squares analysis*, J. Amer. Statist. Assoc. **60** (1965), 523–527.
1301. N. J. Rose, *A note on computing the Drazin inverse*, Linear Algebra and its Applications **15** (1976), no. 2, 95–98.
1302. ———, *The Laurent expansion of a generalized resolvent with some applications*, SIAM J. Math. Anal. **9** (1978), no. 4, 751–758.
1303. J. B. Rosen, *The gradient projection method for nonlinear programming. Part I: Linear Constraints*, J. Soc. Indust. Appl. Math. **8** (1960), 181–217.
1304. ———, *The gradient projection method for nonlinear programming. Part II: Nonlinear Constraints*, J. Soc. Indust. Appl. Math. **9** (1961), 514–532.
1305. ———, *Minimum and basic solutions to singular linear systems*, J. Soc. Indust. Appl. Math. **12** (1964), 156–162.
1306. ———, *Chebyshev solutions of large linear systems*, J. Comput. Syst. Sci. **1** (1967), 29–43.
1307. M. Rosenberg, *Range decomposition and generalized inverse of nonnegative Hermitian matrices*, SIAM Rev. **11** (1969), 568–571.
1308. P. C. Rosenbloom, *The method of steepest descent*, Numerical Analysis. Proceedings of the Sixth Symposium in Applied Mathematics, McGraw–Hill Book Co., New York, 1956, pp. 127–176.
1309. U. G. Rothblum, *A representation of the Drazin inverse and characterizations of the index*, SIAM J. Appl. Math. **31** (1976), no. 4, 646–648.
1310. ———, *Resolvent expansions of matrices and applications*, Linear Algebra and its Applications **38** (1981), 33–49.
1311. A. L. Rukhin, *Pattern correlation matrices and their properties*, Linear Algebra and its Applications **327** (2001), no. 1-3, 105–114.
1312. B. Rust, W. R. Burrus, and C. Schneeberger, *A simple algorithm for computing the generalized inverse of a matrix*, Comm. ACM **9** (1966), 381–385, 387.
1313. G. Salinetti, *The generalized inverse in parametric programming*, Calcolo **11** (1974), 351–363 (1975).

1314. W. Sautter, *A posteriori-Fehlerasbschätzungen für die Pseudoinverse und die Lösung minimaler Länge*, Computing **14** (1975), no. 1-2, 37–44.

1315. J. -P. Schellhorn, *Generalized inverses and generalized convexity*, Statistical data analysis and inference (Neuchâtel, 1989), North-Holland, Amsterdam, 1989, pp. 445–455.

1316. E. Schmidt, *Zur Theorie der linearen und nichlinearen Integralgleichungen, I. Entwicklung willküricher Funktionen nach Systemen vargeschriebener*, Math. Ann. **63** (1907), 433–476.

1317. ———, *Zur Theorie der linearen und nichlinearen Integralgleichungen, II. Auflösung der allgemeinen linearen Integralgleichung*, Math. Ann. **64** (1907), 161–174.

1318. P. H. Schönemann, *A generalized solution of the orthogonal Procrustes problem*, Psychmoetrika **31** (1966), 1–10.

1319. P. Schönfeld and H. -J. Werner, *A note on C. R. Rao's wider definition BLUE in the general Gauss-Markov model*, Sankhyā Ser. B **49** (1987), no. 1, 1–8.

1320. R. S. Schreiber, *Computing generalized inverses and eigenvalues of symmetric matrices using systolic arrays*, Computing methods in applied sciences and engineering, VI (Versailles, 1983), North-Holland, Amsterdam, 1984, pp. 285–295.

1321. R. S. Schreiber and B. N. Parlett, *Block reflectors: Theory and computation*, SIAM J. Numer. Anal. **25** (1988), no. 1, 189–205.

1322. O. Schreier and E. Sperner, *Introduction to Modern Algebra and Matrix Theory*, Chelsea Publishing Company, New York, N. Y., 1951, Translated by Martin Davis and Melvin Hausner.

1323. G. Schulz, *Iterative Berechnung der Reziproken Matrix*, Z. Angew. Math. Mech. **13** (1933), 57–59.

1324. I. Schur, *Potenzreihen im innern des einheitskreises*, J. Reine Angew. Math. **147** (1917), 205–232.

1325. J. Schwartz, *Perturbations of spectral operators, and applications*, Pacific J. Math. **4** (1954), 415–458.

1326. H. Schwerdtfeger, *Introduction to Linear Algebra and the Theory of Matrices*, P. Noordhoff, Groningen, 1950.

1327. ———, *Remarks on the generalized inverse of a matrix*, Linear Algebra and its Applications **1** (1968), 325–328.

1328. ———, *On the covariance of the Moore-Penrose inverse*, Linear Algebra and its Applications **52/53** (1983), 629–643.

1329. ———, *On the covariance of the Moore-Penrose inverse of a matrix*, C. R. Math. Rep. Acad. Sci. Canada **5** (1983), no. 2, 75–77.

1330. R. K. Scott, *Existence of strong solutions to the generalized inverse of the quasi-geostrophic equations*, Inverse Problems **16** (2000), no. 4, 891–907.

1331. J. E. Scroggs and P. L. Odell, *An alternate definition of a pseudoinverse of a matrix*, SIAM J. Appl. Math. **14** (1966), 796–810, (see [**208**]).

1332. H. L. Seal, *Studies in the history of probability and statistics. XV. The historical development of the Gauss linear model*, Biometrika **54** (1967), 1–24.

1333. S. R. Searle, *Additional results concerning estimable functions and generalized inverse matrices*, J. Roy. Statist. Soc. Ser. B **27** (1965), 486–490.

1334. ———, *Linear Models*, John Wiley & Sons Inc., New York, 1971.

1335. ———, *Extending some results and proofs for the singular linear model*, Linear Algebra and its Applications **210** (1994), 139–151.

1336. G. A. F. Seber, *The linear hypothesis: A general theory*, Charles Griffin & Co. Ltd., London, 1966.

1337. J. Seidel, *Angles and distances in n-dimensional euclidean and noneuclidean geometry. I, II, III*, Nederl. Akad. Wetensch. Proc. Ser. A. **58** = Indag. Math. **17** (1955), 329–335, 336–340, 535–541.

1338. S. K. Sen and S. S. Prabhu, *Optimal iterative schemes for computing the Moore-Penrose matrix inverse*, Internat. J. Systems Sci. **7** (1976), no. 8, 847–852.

1339. E. Seneta, *Perturbation of the stationary distribution measured by ergodicity coefficients*, Adv. in Appl. Probab. **20** (1988), no. 1, 228–230.

1340. ———, *Sensitivity of finite Markov chains under perturbation*, Statist. Probab. Lett. **17** (1993), no. 2, 163–168.

1341. A. Sengupta, *Multifunction and generalized inverse*, J. Inverse Ill-Posed Probl. **5** (1997), 265–285.

1342. Jia-Yu Shao and Hai-Ying Shan, *Matrices with signed generalized inverses*, Linear Algebra and its Applications **322** (2001), 105–127.

1343. G. E. Sharpe and G. P. H. Styan, *Circuit duality and the general network inverse*, IEEE Trans. Circuit Th. **12** (1965), 22–27.

1344. ———, *A note on the general network inverse*, IEEE Trans. Circuit Th. **12** (1965), 632–633.

1345. ———, *Circuit duality and the general network inverse*, Proc. IEEE **55** (1967), 1226–1227.

1346. R. D. Sheffield, *On pseudo–inverses of linear transformations in Banach space*, Tech. Report 2133, Oak Ridge National Laboratory, 1956.

1347. ———, *A general theory for linear systems*, Amer. Math. Monthly **65** (1958), 109–111.

1348. Ji Lin Shi and Xiao Fa Shi, *Two iterative methods for computing generalized inverses of matrices*, J. Dalian Inst. Tech. **25** (1986), no. Special Issue on Mathematics, suppl., 37–45.

1349. N. Shinozaki and M. Sibuya, *The reverse order law* $(AB)^- = B^-A^-$, Linear Algebra and its Applications **9** (1974), 29–40.

1350. ———, *Further results on the reverse-order law*, Linear Algebra and its Applications **27** (1979), 9–16.

1351. N. Shinozaki, M. Sibuya, and K. Tanabe, *Numerical algorithms for the Moore-Penrose inverse of a* *matrix: direct methods*, Ann. Inst. Statist. Math. **24** (1972), 193–203.

1352. ———, *Numerical algorithms for the Moore-Penrose inverse of a matrix: iterative methods*, Ann. Inst. Statist. Math. **24** (1972), 621–629.
1353. D. W. Showalter, *Representation and computation of the pseudoinverse*, Proc. Amer. Math. Soc. **18** (1967), 584–586.
1354. D. W. Showalter and A. Ben-Israel, *Representation and computation of the generalized inverse of a bounded linear operator between Hilbert spaces*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **48** (1970), 184–194.
1355. M. Sibuya, *Subclasses of generalized inverses of matrices*, Ann. Inst. Statist. Math. **22** (1970), 543–556.
1356. ———, *Generalized inverses of mappings*, Sankhyā Ser. A **33** (1971), 301–310, (corrigendum: *Sankhyā Ser. A***37**(1975), 309).
1357. ———, *The Azumaya-Drazin pseudoinverse and the spectral inverses of a matrix*, Sankhyā Ser. A **35** (1973), 95–102.
1358. A. Sidi, *Development of iterative techniques and extrapolation methods for Drazin inverse solution of consistent or inconsistent singular linear systems*, Linear Algebra and its Applications **167** (1992), 171–203.
1359. ———, *A unified approach to Krylov subspace methods for the Drazin-inverse solution of singular nonsymmetric linear systems*, Linear Algebra and its Applications **298** (1999), no. 1-3, 99–113.
1360. A. Sidi and V. Kluzner, *A Bi-CG type iterative method for Drazin-inverse solution of singular inconsistent nonsymmetric linear systems of arbitrary index*, Proceedings of the Eleventh Haifa Matrix Theory Conference (1999), vol. 6, 1999/00, pp. 72–94 (electronic).
1361. C. L. Siegel, *Über die analytische theorie der quadratischen formen III*, Ann. Math. **38** (1937), 212–291, (See in particular, pp. 217–229).
1362. ———, *Equivalence of quadratic forms*, Amer. J. Math. **63** (1941), 658–680.
1363. Reinhard Siegmund-Schultze, *Eliakim Hastings Moore's "general analysis"*, Arch. Hist. Exact Sci. **52** (1998), no. 1, 51–89.
1364. B. Silbermann, *Asymptotic Moore–Penrose inversion of Toeplitz operators*, Linear Algebra and its Applications **256** (1997), 219–234.
1365. I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Springer–Verlag, Berlin, 1970.
1366. Inderjit Singh, G. Poole, and T. L. Boullion, *A class of Hessenberg matrices with known pseudoinverse and Drazin inverse*, Math. Comp. **29** (1975), 615–619.
1367. K. C. Sivakumar, *Nonnegative generalized inverses*, Indian J. Pure Appl. Math. **28** (1997), no. 7, 939–942.
1368. R. E. Skelton, T. Iwasaki, and K. M. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*, Taylor & Francis Ltd., London, 1998, (see Chapter 2).
1369. L. Skula, *Involutions for matrices and generalized inverses*, Linear Algebra and its Applications **271** (1998), 283–308.
1370. P. Slavík, *The least extremal solution of the operator equation $AXB = C$*, J. Math. Anal. Appl. **148** (1990), no. 1, 251–262.
1371. Per A. Smeds, *Line digraphs and the Moore-Penrose inverse*, Linear Algebra and its Applications **36** (1981), 165–172.
1372. J. H. Smith, *A geometric treatment of nonnegative generalized inverses*, Linear Algebra and its Applications **2** (1974), 179–184.
1373. R. L. Smith, *The Moore-Penrose inverse of a retrocirculant*, Linear Algebra and its Applications **22** (1978), 1–8.
1374. F. Smithies, *The eigen–values and singular values of integral equations*, Proc. London. Math. Soc. **43** (1937), 255–279.
1375. ———, *Integral Equations*, Cambridge University Press, Cambridge, England, 1958.
1376. T. Söderstörm and G. W. Stewart, *On the numerical proerties of an iterative method for computing the Moore–Penrose generalized inverse*, SIAM J. Numer. Anal. **11** (1974), 61–74.
1377. E. D. Sontag, *On generalized inverses of polynomial and other matrices*, IEEE Trans. Automat. Control **25** (1980), no. 3, 514–517.
1378. A. G. Spera, *Radical of a Hestenes ring*, Atti Accad. Sci. Lett. Arti Palermo Parte I (4) **35** (1975/76), 283–296 (1978).
1379. J. Springer, *Die exakte Berechnung der Moore Penrose-Inversen einer Matrix durch Residuenarithmetik*, Z. Angew. Math. Mech. **63** (1983), no. 3, 203–210.
1380. ———, *Exact solution of general integer systems of linear equations*, ACM Trans. Math. Software **12** (1986), no. 1, 51–61.
1381. ———, *Verallgemeinerte Inversen ganzzahliger Matrizen*, Z. Angew. Math. Mech. **67** (1987), no. 10, 503–506.
1382. M. D. Springer, *The algebra of random variables*, John Wiley & Sons, New York-Chichester-Brisbane, 1979, Wiley Series in Probability and Mathematical Statistics.
1383. V. P. Sreedharan, *Least squares algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm*, Numer. Math. **17** (1971), 387–401.
1384. R. P. Srivastav, *An L_2-theory of dual integral equations*, J. of M. A. C. T. **9** (1976), 1–21.
1385. M. W. Stadelmaier, N. J. Rose, G. D. Poole, and C. D. Meyer, Jr., *Nonnegative matrices with power invariant zero patterns*, Linear Algebra and its Applications **42** (1982), 23–29.

1386. P. Stahlecker and G. Trenkler, *Linear and ellipsoidal restrictions in linear regression*, Statistics **22** (1991), no. 2, 163–176.

1387. I. Stakgold, *Branching of solutions of nonlinear equations*, SIAM Rev. **13** (1971), 289–332, (errata: SIAM Rev.**14**(1972), 492).

1388. W. T. Stallings and T. L. Boullion, *Computation of pseudoinverse matrices using residue arithmetic*, SIAM Rev. **14** (1972), 152–163.

1389. P. S. Stanimirović, *General determinantal representation of pseudoinverses and its computation*, Rev. Acad. Cienc. Zaragoza (2) **50** (1995), 41–49.

1390. ———, *Determinantal representation of $\{I, J, K\}$ inverses and solution of linear systems*, Math. Slovaca **49** (1999), no. 3, 273–286.

1391. ———, *Limit representations of generalized inverses and related methods*, Appl. Math. Comput. **103** (1999), no. 1, 51–68.

1392. P. S. Stanimirović and D. S. Djordjević, *Full-rank and determinantal representation of the Drazin inverse*, Linear Algebra and its Applications **311** (2000), no. 1-3, 131–151.

1393. P. S. Stanimirović and M. Stanković, *Determinantal representation of weighted Moore-Penrose inverse*, Mat. Vesnik **46** (1994), no. 1-2, 41–50.

1394. ———, *Generalized algebraic complement and Moore-Penrose inverse*, Filomat (1994), no. 8, 57–64.

1395. ———, *Determinants of rectangular matrices and Moore-Penrose inverse*, Novi Sad J. Math. **27** (1997), no. 1, 53–69.

1396. P. Stein, *Some general theorems on iterants*, J. Res. Nat. Bur. Standards **48** (1952), 82–82.

1397. T. E. Stern, *Extremum relations in nonlinear networks and their applications to mathematical programming*, Journées d'Études sur le Contrôle Optimum el les Systèmes Nonlinéaires, Institut National des Sciences et Techniques Nucleaires, Saclay, France, pp. 135–156.

1398. ———, *Theory of Nonlinear Networks and Systems*, Addison–Wesley, Reading, Mass., 1965.

1399. G. W. Stewart, *On the continuity of the generalized inverse*, SIAM J. Appl. Math. **17** (1969), 33–45, (see [**1214**]).

1400. ———, *Projectors and generalized inverses*, Tech Report TNN-97, University of Texas at Austin Computation Center, October 1969.

1401. ———, *Introduction to Matrix Computations*, Academic Press, New York-London, 1973.

1402. ———, *On the perturbation of pseudo-inverses, projections, and linear least squares problems*, SIAM Rev. **19** (1977), 634–662.

1403. ———, *An iterative method for solving linear inequalities*, Technical Report TR-1833, Computer Science Department, University of Maryland, 1987.

1404. ———, *On scaled projections and pseudo-inverses*, Linear Algebra and its Applications **112** (1989), 189–194, (see [**1151**], [**1455**], [**561**], [**188**]).

1405. ———, *On the early history of the singular value decomposition*, SIAM Rev. **35** (1993), 551–566.

1406. J. Stoer, *On the characterization of least upper bound norms in matrix space*, Numer. Math. **6** (1964), 302–314.

1407. M. Stojaković, *Determinanten rechteckiger Matrizen*, Bull. Soc. Mat. Phys. Serbie **4** (1952), no. nos. 1-2, 9–23.

1408. ———, *Sur les matrices quasi-inverses et les matrices quasi-unités*, C. R. Acad. Sci. Paris **236** (1953), 877–879.

1409. G. Strang, *A framework for equilibrium equations*, SIAM Rev. **30** (1988), no. 2, 283–297.

1410. ———, *Patterns in linear algebra*, Amer. Math. Monthly **96** (1989), no. 2, 105–117.

1411. ———, *Inverse problems and derivatives of determinants*, Arch. Rational Mech. Anal. **114** (1991), no. 3, 255–265.

1412. ———, *The fundamental theorem of linear algebra*, Amer. Math. Monthly **100** (1993), no. 9, 848–855.

1413. Ji Guang Sun, *Perturbation of angles between linear subspaces*, J. Comput. Math. **5** (1987), no. 1, 58–61.

1414. Wenyu Sun and Yimin Wei, *Inverse order rule for weighted generalized inverse*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 3, 772–775 (electronic).

1415. ———, *Researches on inverse order rule for weighted generalized inverse*, Numer. Math. J. Chinese Univ. (English Ser.) **9** (2000), no. 2, 234–240.

1416. ———, *Researches on inverse order rule for weighted generalized inverse*, Numer. Math. J. Chinese Univ. (English Ser.) **9** (2000), no. 2, 234–240.

1417. R. Sušanj and M. Radić, *Geometrical meaning of one generalization of the determinant of a square matrix*, Glas. Mat. Ser. III **29(49)** (1994), no. 2, 217–233.

1418. Y. Takane and H. Yanai, *On oblique projectors*, Linear Algebra and its Applications **289** (1999), no. 1-3, 297–310.

1419. W. Y. Tan, *Note on an extension of the Gauss–Markov theorems to multivariate linear regression models*, SIAM J. Appl. Math. **20** (1971), 24–29.

1420. K. Tanabe, *Projection method for solving a singular system of linear equations and its applications*, Numer. Math. **17** (1971), 203–214, (see also [**987**]).

1421. ———, *Characterization of linear stationary iterative process for solving a singular system of linear equations*, Numer. Math. **22** (1974), 349–359.

1422. ———, *Neumann-type expansions of reflexive generalized inverses of a matrix and the hyperpower iterative method*, Linear Algebra and its Applications **10** (1975), 163–175.

1423. ———, *Conjugate-gradient method for computing the Moore-Penrose inverse and rank of a matrix*, J. Optimization Theory Appl. **22** (1977), no. 1, 1–23.

1424. ———, *Continuous Newton-Raphson method for solving an underdetermined system of nonlinear* *equations*, Nonlinear Anal. **3** (1979), no. 4, 495–503.

1425. ———, *Differential geometric approach to extended GRG methods with enforced feasibility in nonlinear programming: global analysis*, In Campbell [**267**], pp. 100–137.

1426. K. Tanabe and M. Sagae, *An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications*, J. Roy. Statist. Soc. Ser. B **54** (1992), no. 1, 211–219, (see [**1130**]).

1427. X. Tang and A. Ben-Israel, *Two consequences of Minkowski's 2^n theorem*, Discrete Math. **169** (1997), 279–282.

1428. R. A. Tapia, *An application of a Newton-like method to the Euler-Lagrange equation.*, Pacific J. Math. **29** (1969), 235–246.

1429. ———, *The weak Newton method and boundary value problems*, SIAM J. Numer. Anal. **6** (1969), 539–550.

1430. A. Tarantola, *Inverse Problem Theory. methods for data fitting and model parameter estimation*, Elsevier Science Publishers B.V., Amsterdam, 1987.

1431. ———, *Inversion of elastic seismic reflection data*, Theory and Applications of Inverse Problems (Helsinki, 1985), Longman Sci. Tech., Harlow, 1988, pp. 97–142.

1432. A. Tarantola and B. Valette, *Generalized nonlinear inverse problems solved using the least squares criterion*, Rev. Geophys. Space Phys. **20** (1982), no. 2, 219–232.

1433. O. E. Taurian and Per-Olov Löwdin, *Some remarks on the projector associated with the intersection of two linear manifolds*, Acta Phys. Acad. Sci. Hungar. **51** (1981), no. 1-2, 5–12 (1982).

1434. O. Taussky, *Note on the condition of matrices*, Math, Tables Aids Comput. **4** (1950), 111–112.

1435. ———, *Matrices C with $C^n \to O$*, J. Algebra **1** (1964), 5–10.

1436. A. E. Taylor, *Introduction to Functional Analysis*, J. Wiley & Sons, New York, 1958.

1437. A. E. Taylor and C. J. A. Halberg, Jr., *General theorems about a bounded linear operator and its conjugate*, J. Reine Angew. Math. **198** (1957), 93–111.

1438. R. P. Tewarson, *On the product form of inverses of sparse matrices*, SIAM Rev. **8** (1966), 336–342.

1439. ———, *A direct method for generalized matrix inversion*, SIAM J. Numer. Anal. **4** (1967), 499–507.

1440. ———, *The product form of inverses of sparse matrices and graph theory*, SIAM Rev. **9** (1967), 91–99.

1441. ———, *A computational method for evaluating generalized inverses*, Comput. J. **10** (1967/1968), 411–413.

1442. ———, *A least squares iterative method for singular equations*, Comput. J. **12** (1969), 388–392.

1443. ———, *On computing generalized inverses*, Computing (Arch. Elektron. Rechnen) **4** (1969), 139–152.

1444. ———, *On some representations of generalized inverses*, SIAM Rev. **11** (1969), 272–276.

1445. ———, *On two direct methods for computing generalized inverses*, Computing **7** (1971), 236–239.

1446. ———, *On minimax solutions of linear equations*, Comput. J. **15** (1972), 277–279.

1447. R. C. Thompson, *On Pearl's paper "A decomposition theorem for matrices"*, Canad. Math. Bull. **12** (1969), 805–808, (See [**1171**]).

1448. ———, *Principal submatrices IX: Interlacing inequalities for singular values*, Linear Algebra and its Applications **5** (1972), 1–12.

1449. W. Thomson (Lord Kelvin), Cambridge and Dublin Math. J. (1848), 84–87.

1450. Yong Ge Tian, *Reverse order laws for the generalized inverses of multiple matrix products*, Linear Algebra and its Applications **211** (1994), 85–100.

1451. ———, *The Moore-Penrose inverses of $m \times n$ block matrices and their applications*, Linear Algebra and its Applications **283** (1998), no. 1-3, 35–60.

1452. A. N. Tihonov, *On the stability of algorithms for the solution of degenerate systems of linear algebraic equations*, Ž. Vyčisl. Mat. i Mat. Fiz. **5** (1965), 718–722.

1453. A. R. Tipton and H. W. Milnes, *Least squares solution of linear equations*, Indust. Math. **22** (1972), 11–16.

1454. W. M. To and D. J. Ewins, *The role of the generalized inverse in structural dynamics*, J. Sound Vibration **186** (1995), 185–195.

1455. M. J. Todd, *A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm*, Oper. Res. **38** (1990), no. 6, 1006–1018, (See [**1404**]).

1456. M. J. Todd, L. Tunçel, and Yinyu Ye, *Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems*, Math. Program. **90** (2001), no. 1, Ser. A, 59–69.

1457. J. Tokarzewski, *Geometric characterization of system zeros and zero directions by the Moore-Penrose inverse of the first non-zero Markov parameter*, Arch. Control Sci. **5** (1996), no. 3-4(41), 245–264.

1458. G. Trenkler, *Biased Estimators in the Linear Regression Model*, Verlagsgruppe Athenäum/Hain/Hanstein, Königstein/Ts., 1981.

1459. ———, *Generalizing Mallows' C_L and optimal ridge constants*, VII. symposium on operations research, Sektionen 4–9 (St. Gallen, 1982), Athenäum/Hain/Hanstein, Königstein, 1983, pp. 157–166.

1460. ———, *Characterizations of oblique and orthogonal projectors*, Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93 (Poznań, 1993) (Dordrecht), Kluwer Acad. Publ., 1994, pp. 255–270.

1461. Yuan-Yung Tseng, *The Characteristic Value Problem of Hermitian Functional Operators in a Non–Hilbert Space*, Ph.d. in mathematics, University of Chicago, Chicago, 1933, (Published by the University of Chicago Libraries, 1936).
1462. ______, *On generalized biorthogonal expansions in metric and unitary spaces*, Proc. Nat. Acad. Sci. U. S. A. **28** (1942), 170–175.
1463. ______, *Expansions according to an arbitrary system of functions in hyper-Hilbert space*, Sci. Rep. Nat. Tsing Hua Univ. **4** (1947), 286–312.
1464. ______, *Generalized inverses of unbounded operators between two unitary spaces*, Doklady Akad. Nauk SSSR (N.S.) **67** (1949), 431–434, (Reviewed in *Math. Rev.* **11**(1950), P. 115).
1465. ______, *Properties and classification of generalized inverses of closed operators*, Doklady Akad. Nauk SSSR (N.S.) **67** (1949), 607–610, (Reviewed in *Math. Rev.* **11**(1950), P. 115).
1466. ______, *Sur les solutions des équations opératrices fonctionnelles entre les espaces unitaires. Solutions extrémales. Solutions virtuelles*, C. R. Acad. Sci. Paris **228** (1949), 640–641, (Reviewed in *Math. Rev.* **11**(1950), P. 115).
1467. ______, *Virtual solutions and general inversions*, Uspehi Mat. Nauk (N.S.) **11** (1956), no. 6(72), 213–215, (Reviewed in *Math. Rev.* **18**(1957), P. 749).
1468. D. H. Tucker, *Boundary value problems for linear differential systems*, SIAM J. Appl. Math. **17** (1969), 769–783.
1469. F. E. Udwadia and R. E. Kalaba, *A new perspective on constrained motion*, Proc. Roy. Soc. London Ser. A **439** (1992), no. 1906, 407–410.
1470. ______, *On motion*, J. Franklin Inst. **330** (1993), no. 3, 571–577.
1471. ______, *An alternate proof for the equation of motion for constrained mechanical systems*, Appl. Math. Comput. **70** (1995), no. 2-3, 339–342.
1472. ______, *The geometry of constrained motion*, Z. Angew. Math. Mech. **75** (1995), no. 8, 637–640.
1473. ______, *Analytical Dynamics*, Cambridge University Press, Cambridge, 1996.
1474. ______, *Response to: R. S. Bucy's "Comments on a paper by F. E. Udwadia and R. E. Kalaba: 'A new perspective on constrained motion'" [Proc. Roy. Soc. London Ser. A **444** (1994), no. 1920, 253–255; MR 95c:70024]*, Proc. Roy. Soc. London Ser. A **452** (1996), no. 1948, 1055–1056, (see [**242**]).
1475. ______, *An alternative proof of the Greville formula*, J. Optim. Theory Appl. **94** (1997), no. 1, 23–28.
1476. ______, *The explicit Gibbs-Appell equation and generalized inverse forms*, Quart. Appl. Math. **56** (1998), no. 2, 277–288.
1477. ______, *General forms for the recursive determination of generalized inverses: unified approach*, J. Optim. Theory Appl. **101** (1999), no. 3, 509–521.
1478. ______, *A unified approach for the recursive determination of generalized inverses*, Comput. Math. Appl. **37** (1999), no. 1, 125–130.
1479. S. Ul′m, *Iteration methods with successive approximation of the inverse operator*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat. **16** (1967), 403–411.
1480. N. S. Urquhart, *Computation of generalized inverse matrices which satisfy specified conditions*, SIAM Rev. **10** (1968), 216–218.
1481. ______, *The nature of the lack of uniqueness of generalized inverse matrices*, SIAM Rev. **11** (1969), 268–271.
1482. O. Vaarmann, *The application of generalized inverse operators and their approximations to the solution of nonlinear equations*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat. **19** (1970), 265–274.
1483. ______, *Approximations of pseudo-inverse operators as applied to the solution of nonlinear equations*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat. **20** (1971), 386–394.
1484. ______, *On methods with successive approximation of the pseudoinverse operator*, Progress in operations research, Vols. I, II (Proc. Sixth Hungarian Conf., Eger, 1974), North-Holland, Amsterdam, 1976, pp. 941–946. Colloq. Math. Soc. János Bolyai, Vol. 12.
1485. ______, *Obobshchennye Obratnye Otobrazheniya (Generalized Inverse Mappings)*, "Valgus", Tallinn, 1988, With English and Estonian summaries.
1486. O. Vaarmann and Marika Lomp, *Extension of the domain of convergence of methods with approximation of a pseudoinverse operator*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat. **31** (1982), no. 4, 410–417, 474.
1487. V. V. Vagner, *The theory of generalized heaps and generalized groups*, Mat. Sbornik N.S. **32(74)** (1953), 545–632.
1488. K. Vala, *Sur les projections des bases orthogonales*, Ann. Acad. Sci. Fenn. Ser. A I No. **285** (1960), 6.
1489. ______, *On the Gram determinant and linear transformations of Hilbert space.*, Ann. Acad. Sci. Fenn. Ser. A I No. **306** (1961), 8.
1490. ______, *On a decomposition for operators with closed range*, Ann. Acad. Sci. Fenn. Ser. A I No. **364** (1965), 8.
1491. J. Van hamme, *Generalized inverses of linear operators in Hilbert spaces*, Bull. Soc. Math. Belg. Sér. B **41** (1989), no. 1, 83–93.
1492. ______, *On the generalized inverse of a matrix partial differential operator*, Simon Stevin **66** (1992), no. 1-2, 185–194.
1493. C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal. **13** (1976), no. 1, 76–83.
1494. ______, *Computing the CS and the generalized singular value decompositions*, Numer. Math. **479–491** (1985), 479–491.
1495. ______, *The ubiquitous Kronecker product*, J. Comput. Appl. Math. **123** (2000), no. 1-2, 85–100.

1496. R. J. Vanderbei and J. C. Lagarias, *I. I. Dikin's convergence result for the affine-scaling algorithm*, Mathematical developments arising from linear programming (Brunswick, ME, 1988), Amer. Math. Soc., Providence, RI, 1990, pp. 109–119.

1497. J. M. Varah, *On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems*, SIAM J. Numer. Anal. **10** (1973), 257–267.

1498. S. A. Vavasis, *Stable numerical algorithms for equilibrium systems*, SIAM J. Matrix Anal. Appl. **15** (1994), 1108–1131.

1499. S. A. Vavasis and Yinyu Ye, *Condition numbers for polyhedra with real number data*, Oper. Res. Lett. **17** (1995), no. 5, 209–214.

1500. ______, *A primal-dual interior point method whose running time depends only on the constraint matrix*, Math. Programming **74** (1996), no. 1, Ser. A, 79–120.

1501. D. Veljan, *The sine theorem and inequalities for volumes of simplices and determinants*, Linear Algebra and its Applications **219** (1995), 79–91.

1502. G. C. Verghese, *A "Cramer rule" for the least-norm, least-squared-error solution of inconsistent linear equations*, Linear Algebra and its Applications **48** (1982), 315–316, (extension of [**120**]).

1503. A. Vogt, *On the linearity of form isometries*, SIAM J. Appl. Math. **22** (1972), 553–560.

1504. J. von Neumann, *Über adjungierte Funktionaloperatoren*, Ann. of Math. **33** (1932), 294–310.

1505. ______, *On regular rings*, Proc. Nat. Acad. Sci. U.S.A. **33** (1936), 294–310.

1506. ______, *Some matrix inequalities and metrization of matric–space*, Tomsk Univ. Rev. **1** (1937), 286–300, (Republished in *John von Neumann Collected Works*, MacMillan, New York, Vol IV, pp. 205–219).

1507. ______, *Functional Operators. Vol II: The Geometry of Orthogonal Spaces*, Annals of Math. Studies, vol. 29, Princeton University Press, Princeton, 1950.

1508. ______, *Continuous Geometry*, Princeton University Press, Princeton, 1960.

1509. D. von Rosen, *A matrix formula for testing linear hypotheses in linear models*, Linear Algebra and its Applications **127** (1990), 457–461.

1510. ______, *Some results on homogeneous matrix equations*, SIAM J. Matrix Anal. Appl. **14** (1993), no. 1, 137–145.

1511. G. Votruba, *Generalized Inverses and Singular Equations in Functional Analysis*, doctoral dissertation in mathematics, The University of Michigan, Ann Arbor, MI, 1963.

1512. G. Wahba and M. Z. Nashed, *The approximate solution of a class of constrained control problems*, Proceedings of the Sixth Hawaii International Conference on Systems Sciences (Hawaii), 1973.

1513. D. Walker and C. R. Hallum, *Pseudoinverses in generalizing Newton's method for obtaining the square root of a symmetric positive semidefinite matrix*, Indust. Math. **34** (1984), no. 2, 137–146.

1514. H. F. Walker, *Newton-like methods for underdetermined systems*, Computational Solution of Nonlinear Systems of Equations (Fort Collins, CO, 1988, Lectures in Applied Math. 26) (E. L. Allgower and K. Georg, eds.), American Mathematical Society, Providence, RI, 1990, pp. 679–699.

1515. H. F. Walker and L. T. Watson, *Least-change secant update methods for underdetermined systems*, SIAM J. Numer. Anal. **27** (1990), 1227–1262.

1516. J. R. Wall and R. J. Plemmons, *Spectral inverses of stochastic matrices*, SIAM J. Appl. Math. **22** (1972), 22–26.

1517. Bo-Ying Wang and Bo-Yan Xi, *Some inequalities for singular values of matrix products*, Linear Algebra and its Applications **264** (1997), 109–115.

1518. Bo-Ying Wang, Xiuping Zhang, and Fuzhen Zhang, *Some inequalities on generalized Schur complements*, Linear Algebra and its Applications **302/303** (1999), 163–172.

1519. Guorong Wang, *A Cramer rule for minimum-norm* (T) *least-squares* (S) *solution of inconsistent linear equations*, Linear Algebra and its Applications **74** (1986), 213–218, (see [**120**], [**1502**]).

1520. ______, *A finite algorithm for computing the weighted Moore-Penrose inverse* A^+_{MN}, Appl. Math. Comput. **23** (1987), no. 4, 277–289.

1521. ______, *Approximation methods for the* W*-weighted Drazin inverses of linear operators in Banach spaces*, Numer. Math. J. Chinese Univ. **10** (1988), no. 1, 74–81.

1522. ______, *A Cramer rule for finding the solution of a class of singular equations*, Linear Algebra and its Applications **116** (1989), 27–34.

1523. ______, *Iterative methods for computing the Drazin inverse and the* W*-weighted Drazin inverse of linear operators based on functional interpolation*, Numer. Math. J. Chinese Univ. **11** (1989), no. 3, 269–280.

1524. ______, *An imbedding method for computing the generalized inverses*, J. Comput. Math. **8** (1990), no. 4, 353–362.

1525. ______, *Weighted Moore-Penrose, Drazin, and group inverses of the Kronecker product* $A \otimes B$*, and some applications*, Linear Algebra and its Applications **250** (1997), 39–50.

1526. Guorong Wang and Yong Lin Chen, *A recursive algorithm for computing the weighted Moore-Penrose inverse* A^+_{MN}, J. Comput. Math. **4** (1986), no. 1, 74–85.

1527. Guorong Wang and Jiao Xun Kuang, *A new measure of the degree of ill-conditioning of a matrix*, Numer. Math. J. Chinese Univ. **1** (1979), no. 1, 20–30.

1528. Guorong Wang and Hongjiong Tian, *Projection methods for computing generalized inverses* $A^{(1,2)}_{T,S}$

and $A^{(2)}_{T,S}$ and their applications, Math. Appl. **9** (1996), no. 4, 464–469, Projection methods.

1529. Guorong Wang and Yimin Wei, *Limiting expression for generalized inverse $A^{(2)}_{T,S}$ and its corresponding projectors*, Numer. Math. J. Chinese Univ. (English Ser.) **4** (1995), no. 1, 25–30.
1530. Jia Song Wang and Xiao Wen Chang, *Some notes on the generalized inverse under the L_1 norm*, Nanjing Daxue Xuebao Shuxue Bannian Kan **7** (1990), no. 2, 228–237.
1531. Song-Gui Wang and Wai-Cheung Ip, *A matrix version of the Wielandt inequality and its applications to statistics*, Linear Algebra and its Applications **296** (1999), no. 1-3, 171–181.
1532. Yu Wen Wang, *Geometry of Banach space and generalized inverses of operators*, Natur. Sci. J. Harbin Normal Univ. **14** (1998), no. 5, 1–7.
1533. Yu Wen Wang and Da Qin Ji, *The Tseng–metric generalized inverse for linear operators in Banach spaces*, J. Systems Sci. Math. Sci. **20** (2000), no. 2, 203–209.
1534. Yu Wen Wang and Zhi Wei Li, *Moore–Penrose generalized inverses in Banach spaces and ill-posed boundary value problems*, J. Systems Sci. Math. Sci. **15** (1995), no. 2, 175–185.
1535. Yu Wen Wang and Run Jie Wang, *Pseudoinverse and two-objective optimal control in Banach spaces*, Funct. Approx. Comment. Math. **21** (1992), 149–160.
1536. J. F. Ward, T. L. Boullion, and T. O. Lewis, *A note on the oblique matrix pseudoinverse*, SIAM J. Appl. Math. **20** (1971), 173–175.
1537. ______, *Weak spectral inverses*, SIAM J. Appl. Math. **22** (1972), 514–518.
1538. J. H. M. Wedderburn, *Lectures on Matrices*, Colloq. Publ., vol. XVII, Amer. Math. Soc., Providence, RI, 1934.
1539. P.-Å. Wedin, *The non–linear least squares problem from a numerical point of view. I. geometrical properties*, Dept. of computer sciences, Lund University, Lund, Sweden, August 1972.
1540. ______, *Perturbation bounds in connection with singular value decomposition*, BIT **12** (1972), 99–111.
1541. ______, *Pertubation theory for pseudo-inverses*, BIT **13** (1973), 217–232.
1542. W. G. Wee, *Generalized inverse approach to adaptive multiclass pattern classification*, IEEE Trans. Electron. Comput. **C17** (1968), 1157–1164.
1543. ______, *A generalized inverse approach to clustering pattern selection and classification*, IEEE Trans. Info. Th. **17** (1971).
1544. Musheng Wei, *Upper bounds and stability of scaled pseudoinverses*, Numer. Math. **72** (1995), 285–293.
1545. ______, *Equivalent conditions for generalized inverses of products*, Linear Algebra and its Applications **266** (1997), 347–363.
1546. ______, *Reverse order laws for generalized inverses of multiple matrix products*, Linear Algebra and its Applications **293** (1999), 273–288.
1547. Yimin Wei, *A characterization and representation of the Drazin inverse*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 4, 744–747.
1548. ______, *A characterization and representation of the generalized inverse $A^{(2)}_{T,S}$ and its applications*, Linear Algebra and its Applications **280** (1998), 87–96.
1549. ______, *Expressions for the Drazin inverse of a* 2×2 *block matrix*, Linear and Multilinear Algebra **45** (1998), no. 2-3, 131–146.
1550. ______, *Index splitting for the Drazin inverse and the singular linear system*, Appl. Math. Comput. **95** (1998), no. 2-3, 115–124.
1551. ______, *On the perturbation of the group inverse and oblique projection*, Appl. Math. Comput. **98** (1999), no. 1, 29–42.
1552. ______, *Perturbation bound of singular linear systems*, Appl. Math. Comput. **105** (1999), no. 2-3, 211–220.
1553. ______, *The Drazin inverse of updating of a square matrix with application to perturbation formula*, Appl. Math. Comput. **108** (2000), no. 2-3, ’77–83.
1554. ______, *Perturbation analysis of singular linear systems with index one*, Int. J. Comput. Math. **74** (2000), no. 4, 483–491.
1555. ______, *Recurrent neural networks for computing weighted Moore–Penrose inverse*, Appl. Math. Comput. **116** (2000), no. 3, 279–287.
1556. ______, *Representations and perturbations of Drazin inverses in Banach spaces*, Chinese Ann. Math. Ser. A **21** (2000), no. 1, 33–38.
1557. ______, *Successive matrix squaring algorithm for computing the Drazin inverse*, Appl. Math. Comput. **108** (2000), no. 2-3, 67–75.
1558. Yimin Wei and Guoliang Chen, *Perturbation of least squares problem in Hilbert space*, Appl. Math. Comput. **121** (2001), 171–177.
1559. Yimin Wei and Jiu Ding, *Representations for Moore–Penrose inverses in Hilbert spaces*, Appl. Math. Letters **14** (2001), 599–604.
1560. Yimin Wei and Jiao Xun Kuang, *Iterative methods for computing the Drazin inverses of linear operators in Banach spaces*, J. Fudan Univ. Nat. Sci. **35** (1996), no. 4, 407–413.
1561. Yimin Wei and Guorong Wang, *The perturbation theory for the Drazin inverse and its applications*, Linear Algebra and its Applications **258** (1997), 179–186.
1562. ______, *Approximate methods for the generalized inverse $A^{(2)}_{T,S}$*, J. Fudan Univ. Nat. Sci. **38** (1999), no. 2, 234–239.
1563. ______, *Perturbation theory for the generalized inverse $A^{(2)}_{T,S}$*, J. Fudan Univ. Nat. Sci. **39** (2000), no. 5, 482–488.

1564. Yimin Wei and Hebing Wu, *Expression for the perturbation of the weighted Moore-Penrose inverse*, Comput. Math. Appl. **39** (2000), no. 5-6, 13–18.

1565. ______, *The perturbation of the Drazin inverse and oblique projection*, Appl. Math. Lett. **13** (2000), no. 3, 77–83.

1566. ______, *The representation and approximation for Drazin inverse*, J. Comput. Appl. Math. **126** (2000), no. 1-2, 417–432.

1567. ______, *Additional results on index splittings for drazin inverse solutions of singular linear systems*, The Electronic Journal of Linear Algebra **8** (2001), no. 1, 83–93.

1568. ______, *The representation and approximation for the weighted Moore–Penrose inverse*, Appl. Math. Comput. **121** (2001), no. 1, 17–28.

1569. ______, *$(T-S)$ splitting methods for computing the generalized inverse $A^{(2)}_{T,S}$ and rectangular systems*, International Journal of Computer Mathematics **77** (2001), no. 3, 401–424.

1570. H. -J. Werner, *On the matrix monotonicity of generalized inversion*, Linear Algebra and its Applications **27** (1979), 141–145.

1571. ______, *Die Bott-Duffin Inversen*, Z. Angew. Math. Mech. **60** (1980), no. 7, T275–T277.

1572. ______, *Zur Darstellung traditioneller BLU-Schätzer im singulären linearen Modell*, Proceedings of the Annual Meeting of the Gesellschaft für Angewandte Mathematik und Mechanik, Würzburg 1981, Part II (Würzburg, 1981), vol. 62, 1982, pp. T389–T390.

1573. ______, *Charakterisierungen von monotonen Matrizen*, Linear Algebra and its Applications **60** (1984), 79–90.

1574. ______, *More on BLIMB-estimation*, Contributions to operations research and mathematical economics, Vol. II, Athenäum/Hain/Hanstein, Königstein, 1984, pp. 629–638.

1575. ______, *On extensions of Cramer's rule for solutions of restricted linear systems*, Linear and Multilinear Algebra **15** (1984), no. 3-4, 319–330.

1576. ______, *Drazin-monotonicity characterizations for property-n matrices*, Linear Algebra and its Applications **71** (1985), 327–337.

1577. ______, *More on BLU estimation in regression models with possibly singular covariances*, Linear Algebra and its Applications **67** (1985), 207–214.

1578. ______, *Generalized inversion and weak bi-complementarity*, Linear and Multilinear Algebra **19** (1986), no. 4, 357–372.

1579. ______, *C. R. Rao's IPM method: a geometric approach*, New perspectives in theoretical and applied statistics (Bilbao, 1986), Wiley, New York, 1987, pp. 367–382.

1580. ______, *On weak r-monotonicity*, Linear Algebra and its Applications **86** (1987), 199–209.

1581. ______, *Some recent results on Drazin-monotonicity of property–n matrices*, Linear and Multilinear Algebra **21** (1987), no. 3, 243–251.

1582. ______, *A closed form formula for the intersection of two complex matrices under the star order*, Linear Algebra and its Applications **140** (1990), 13–30.

1583. ______, *On inequality constrained generalized least-squares estimation*, Linear Algebra and its Applications **127** (1990), 379–392.

1584. ______, *Some further results on matrix monotonicity*, Linear Algebra and its Applications **150** (1991), 371–392.

1585. ______, *Characterizations of minimal semipositivity*, Linear and Multilinear Algebra **37** (1994), no. 4, 273–278.

1586. ______, *When is B^-A^- a generalized inverse of AB?*, Linear Algebra and its Applications **210** (1994), 255–263.

1587. H. -J. Werner and C. Yapar, *A BLUE decomposition in the general linear regression model*, Linear Algebra and its Applications **237/238** (1996), 395–404.

1588. ______, *On inequality constrained generalized least squares selections in the general possibly singular Gauss-Markov model: a projector theoretical approach*, Linear Algebra and its Applications **237/238** (1996), 359–393.

1589. H. Weyl, *Repartición de corriente en una red conductora*, Revista Matemática Hispano–Americana **5** (1923), 153–164.

1590. ______, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 408–411.

1591. T. M. Whitney and R. K. Meany, *Two algorithms related to the method of steepest descent*, SIAM J. Numer. Anal. **4** (1967), 109–118.

1592. E. A. Wibker, R. B. Howe, and J. D. Gilbert, *Explicit solutions to the reverse order law $(AB)^+ = B^-_{mr}A^-_{lr}$*, Linear Algebra and its Applications **25** (1979), 107–114.

1593. N. A. Wiegmann, *Some analogs of the generalized principal axis transformation*, Bull. Amer. Math. Soc. **54** (1948), 905–908.

1594. H. S. Wilf and F. Harary (eds.), *Mathematical Aspects of Electrical Network Analysis*, SIAM–AMS Proc., vol. III, Providence, RI, Amer. Math. Soc., 1971.

1595. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford Univertsity Press, London, 1965.

1596. ______, *The solution of ill–conditioned linear equations*, In Ralston and Wilf [**1238**], pp. 65–93, Vol. II.

1597. ______, *Note on the practical significance of the Drazin inverse*, In Campbell [**267**], pp. 82–99.

1598. J. H. Wilkinson and C. Reinsch (eds.), *Handbook for Automatic Computation, Vol. II: Linear Algebra*, Springer-Verlag, Berlin, 1971.

1599. J. Williamson, *A polar representation of singular matrices*, Bull. Amer. Math. Soc. **41** (1935), 118–123.

1600. ______, *Note on a principal axis transformation for non–Hermitian matrices*, Bull. Amer. Math. Soc. **45** (1939), 920–922.

1601. L. B. Willner, *An elimination method for computing the generalized inverse*, Math. Comput. **21** (1967), 227–229.

1602. H. K. Wimmer, *Linear matrix equations, controllability and observability, and the rank of solutions*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 570–578.

1603. ______, *Bezoutians of polynomial matrices and their generalized inverses*, Linear Algebra and its Applications **122/123/124** (1989), 475–487.

1604. ______, *Generalized singular values and interlacing inequalities*, J. Math. Anal. Appl. **137** (1989), no. 1, 181–184.

1605. ______, *Explicit solutions of the matrix equation* $\sum A^i X D_i = C$, SIAM J. Matrix Anal. Appl. **13** (1992), no. 4, 1123–1130.

1606. ______, *Canonical angles of unitary spaces and perturbations of direct complements*, Linear Algebra and its Applications **287** (1999), no. 1-3, 373–379.

1607. ______, *Lipschitz continuity of oblique projections*, Proc. Amer. Math. Soc. **128** (2000), no. 3, 873–876.

1608. H. K. Wimmer and A. D. Ziebur, *Solving the matrix equation* $\sum_{\rho=1}^{r} f_\rho(A) X g_\rho(B) = C$, SIAM Rev. **14** (1972), 318–323.

1609. ______, *Blockmatrizen und lineare Matrizengleichungen*, Math. Nachr. **59** (1974), 213–219.

1610. H. Wolkowicz, *Explicit solutions for interval semidefinite linear programs*, Linear Algebra and its Applications **236** (1996), 95–104.

1611. H. Wolkowicz and S. Zlobec, *Calculating the best approximate solution of an operator equation*, Math. Comp. **32** (1978), no. 144, 1183–1213.

1612. Chi Song Wong, Hua Cheng, and J. Masaro, *Multivariate versions of Cochran theorems*, Linear Algebra and its Applications **291** (1999), no. 1-3, 227–234.

1613. Edward T. Wong, *Polygons, circulant matrices, and Moore-Penrose inverses*, Amer. Math. Monthly **88** (1981), no. 7, 509–515.

1614. ______, *Involutory functions and Moore–Penrose inverses of matrices in an arbitrary field*, Linear Algebra and its Applications **48** (1982), 283–291.

1615. James S. W. Wong, *Remarks on a result of Gram determinants and generalized Schwartz inequality*, Matrix Tensor Quart. **14** (1963/1964), 77–80.

1616. Chuan-Kun Wu and E. Dawson, *Existence of generalized inverse of linear transformations over finite fields*, Finite Fields Appl. **4** (1998), no. 4, 307–315.

1617. ______, *Generalised inverses in public key cryptosystem design*, IEEE Proceedings: Computers and Digital Techniques **145** (1998), no. 5, 321–326.

1618. O. Wyler, *Green's operators*, Ann. Mat. Pura Appl. (4) **66** (1964), 252–263, (see [**1662**]).

1619. ______, *On two-point boundary-value problems*, Ann. Mat. Pura Appl. (4) **67** (1965), 127–142.

1620. P. Wynn, *Upon the generalised inverse of a formal power series with vector valued coefficients*, Compositio Math. **23** (1971), 453–460.

1621. H. Yanai and S. Puntanen, *Partial canonical correlations associated with the inverse and some generalized inverses of a partitioned dispersion matrix*, Statistical sciences and data analysis (Tokyo, 1991), VSP, Utrecht, 1993, pp. 253–264.

1622. H. Ying, *A canonical form for pencils of matrices with applications to asymptotic linear programs*, Linear Algebra and its Applications **234** (1996), 97–123.

1623. K. Yosida, *Functional Analysis*, 2nd ed., Springer–Verlag, Berlin–New York, 1958.

1624. S. Zacks, *Generalized least squares estimators for randomized replication designs*, Ann. Math. Statist. **35** (1964), 696–704.

1625. L. A. Zadeh and C. A. Desoer, *Linear Syetem Theory*, McGraw–Hill Book Co., New York, 1963.

1626. E. H. Zarantonello, *Differentoids*, Advances in Math. **2** (1968), 187–306.

1627. H. Zassenhaus, *'Angles of inclination' in correllation theory*, Amer. Math. Monthly **71** (1964), 218–219.

1628. A. Zellner, *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*, J. Amer. Statist. Assoc. **57** (1962), 348–368.

1629. ______, *Estimators for seemingly unrelated equations: Some exact finite sample results*, J. Amer. Statist. Assoc. **58** (1963), 977–992.

1630. Bao-Xue Zhang and Xian-Hai Zhu, *Gauss-markov and weighted least-squares estimation under a general growth curve model*, Linear Algebra and its Applications **321** (2000), 387–398.

1631. Yao Ting Zhang, *The exact distribution of the Moore-Penrose inverse of X with a density*, Multivariate analysis VI (Pittsburgh, Pa., 1983), North-Holland, Amsterdam, 1985, (see [**1152**]), pp. 633–635.

1632. Dao-Sheng Zheng, *Further study and generalization of Kahan's matrix extension theorem*, SIAM J. Matrix Anal. Appl. **17** (1996), no. 3, 621–631, (See [**1633**]).

1633. ______, *Note on: "Further study and generalization of Kahan's matrix extension theorem" [SIAM J. Matrix Anal. Appl.* **17** *(1996), no. 3, 621–631; MR 97f:15008]*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 1, 277 (electronic).

1634. P. Zieliński and K. Ziętak, *The polar decomposition—properties, applications and algorithms*, Mat. Stos. **38** (1995), 23–49.

1635. G. Zielke, *Numerische Berechnung von benachbarten inversen Matrizen und linearen Gleichungssystemen*, Friedr. Vieweg & Sohn, Braunschweig, 1970.

1636. ______, *Zur historischen Entwicklung von verallgemeinerten inversen Matrizen*, Wiss. Z.

Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **27** (1978), no. 4, 109–118.

1637. ______, *Motivation und Darstellung von verallgemeinerten Matrixinversen*, Beiträge Numer. Math. **7** (1979), 177–218.

1638. ______, *Die Auflösung beliebiger linearer algebraischer Gleichungssysteme durch Blockzerlegung*, Beiträge Numer. Math. **8** (1980), 181–199.

1639. ______, *Die Auflösung linearer Gleichungssysteme mittels verallgemeinerter inverser Matrizen*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **29** (1980), no. 1, 5–15.

1640. ______, *Verallgemeinerungen einer Testmatrix von Rutishauser mit exakten Moore Penrose-Inversen*, Z. Angew. Math. Mech. **61** (1981), no. 12, 662–663.

1641. ______, *Verallgemeinerte inverse Matrizen*, Yearbook: Surveys of mathematics 1983, Bibliographisches Inst., Mannheim, 1983, pp. 95–116.

1642. ______, *Report on test matrices for generalized inverses*, Computing **36** (1986), no. 1-2, 105–162.

1643. ______, *Some remarks on matrix norms, condition numbers, and error estimates for linear equations*, Linear Algebra and its Applications **110** (1988), 29–41.

1644. ______, *Lineare Gleichungssysteme und verallgemeinerte Inversen: Grundlagen und numerische Verfahren*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **40** (1991), no. 1, 45–59.

1645. K. Ziętak, *On a particular case of the inconsistent linear matrix equation $AX + YB = C$*, Linear Algebra and its Applications **66** (1985), 249–258.

1646. ______, *Properties of the approximate generalized inverses of a class of matrices*, Linear Algebra and its Applications **92** (1987), 259–272.

1647. ______, *Properties of the approximations of a matrix which lower its rank*, IMA J. Numer. Anal. **9** (1989), no. 4, 545–554.

1648. ______, *Subdifferentials, faces, and dual matrices*, Linear Algebra and its Applications **185** (1993), 125–141.

1649. ______, *Orthant-monotonic norms and overdetermined linear systems*, J. Approx. Theory **88** (1997), no. 2, 209–227.

1650. S. Zlobec, *On computing the generalized inverse of a linear operator*, Glasnik Mat. Ser. III **2 (22)** (1967), 265–271.

1651. ______, *Contributions to mathematical programming and generalized inversion*, Applied math., Northwestern Univ., Evanston, IL, 1970.

1652. ______, *An explicit form of the Moore-Penrose inverse of an arbitrary complex matrix*, SIAM Rev. **12** (1970), 132–134.

1653. ______, *On computing the best least squares solutions in Hilbert space*, Rend. Circ. Mat. Palermo (2) **25** (1976), no. 3, 256–270 (1977).

1654. S. Zlobec and A. Ben-Israel, *On explicit solutions of interval linear programs*, Israel J. Math. **8** (1970), 265–271.

1655. ______, *Explicit solutions of interval linear programs*, Operations Res. **21** (1973), 390–393.

1656. V. M. Zubov, *Certain properties of the generalized Green's matrices of linear differential operators*, Differencialnye Uravnenija **10** (1974), 996–1002, 1147.

1657. ______, *On the question of generalized Green's matrices*, Mat. Zametki **15** (1974), 113–120.

1658. ______, *The closedness of the generalized Green matrices of linear differential operators*, Mat. Zametki **24** (1978), no. 2, 217–229, 302, (English translation: Math. Notes 24 (1978), no. 1–2, 617–624 (1979)).

1659. ______, *On the theory of a nonclosed generalized Green matrix*, Izv. Vyssh. Uchebn. Zaved. Mat. (1979), no. 1, 21–26, (English translation: Soviet Math. (Iz. VUZ) 23 (1979), no. 1, 16–21).

1660. ______, *On the theory of the generalized Green function*, Izv. Vyssh. Uchebn. Zaved. Mat. (1981), no. 9, 69–73.

1661. ______, *Generalized inverse for a Fredholm operator*, Izv. Vyssh. Uchebn. Zaved. Mat. (1982), no. 5, 76–78, (English translation: Soviet Math. (Iz. VUZ) 26 (1982), no. 5, 100–103).

1662. ______, *Some properties of a generalized inverse operator in a vector space*, Izv. Vyssh. Uchebn. Zaved. Mat. (1983), no. 12, 67–69.

1663. ______, *A generalized inverse operator and a class of ill-posed problems*, Izv. Vyssh. Uchebn. Zaved. Mat. (1987), no. 6, 77–79, 91, (English translation: Soviet Math. (Iz. VUZ) 31 (1987), no. 6, 100–103).

1664. ______, *A method for the approximate reconstruction of an operator*, Vestsī Akad. Navuk BSSR Ser. Fīz.-Mat. Navuk (1989), no. 3, 39–43, 124.

1665. E. L. Žukovskiĭ, *The generalized solution of systems of linear algebraic equations*, Dokl. Akad. Nauk SSSR **232** (1977), no. 2, 269–272, (English translation: *Soviet Math. Dokl.* **18** (1977), no. 1, 49–53).

1666. ______, *The method of least squares for degenerate and ill-conditioned systems of linear algebraic equations*, Ž. Vyčisl. Mat. i Mat. Fiz. **17** (1977), no. 4, 814–827, 1091.

1667. E. L. Žukovskiĭ and R. Š. Lipcer, *A recurrence method for computing the normal solutions of linear algebraic equations*, Ž. Vyčisl. Mat. i Mat. Fiz. **12** (1972), 843–857, 1084.

1668. ______, *The computation of pseudoinverse matrices*, Ž. Vyčisl. Mat. i Mat. Fiz. **15** (1975), no. 2, 489–492, 542.

1669. G. Zyskind, *On canonical forms, nonnegative covariance matrices and best and simple least squares linear estimators in linear models*, Ann. Math. Statist. **38** (1967), 1092–1109.

1670. G. Zyskind and F. B. Martin, *On best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure*, SIAM J. Appl. Math. **17** (1969), 1190–1202.

Subject Index

$N(S)$–restricted pseudoinverse of T, 255
S–inverse, 131
S–restricted
  $\{1,3\}$–inverse, 98
  $\{1,4\}$–inverse, 98
  $\{i,j,\dots,k\}$–inverse, 73
S'–inverse, 136
UDV^*–decomposition, 176
α–β generalized inverse, 107
α–approximate solution, 108
ℓ_p norm, 10, 113
λ–vector, 125
ϕ–metric projector, 104
$\{1,2,3\}$–inverse, 35
$\{1,2,4\}$–inverse, 35
$\{1,2\}$–inverse, 34
$\{1,3\}$–inverse, 87, 97
$\{1,4\}$–inverse, 97
$\{1\}$–inverse, 28, 29
$\{2\}$–inverse, 40, 42
$\{i,j,\dots,k\}$–inverse, 27

admittance matrix, 85
affine set, 151
angle, 10

best rank-k approximation, 182
Bott–Duffin inverse, 76, 82, 122
branches
  currents, 83

carrier, 226
Cauchy–Schwartz inequality, 9, 113
  generalized, 113
Cholesky factorization, 100
closure, 225
complementary subspaces, 7
condition number, 178
consistent norms, 17
constrained
  inverse, 76
  least–square solution, 92
contraction, 191
convergent matrix, 20
convex
  body, 112
convex function, 104
convex set
  rotund, 115
  smooth, 115

Cramer's rule, 25, 172, 173
current, 83

decomposable, 225
derivative, 41
determinant, 23
differentiable, 41
direct sum, 225
discriminant, 78
domain, 226
Drazin inverse, 127, 132, 134
dual
  function, 111
  norms, 113
  set, 113
  vectors, 113

eigenfunction, 238
eigenvalue, 238
electrical network, 82, 123
  currents, 82
  dual transfer matrix, 84
  transfer matrix, 84
  voltages, 82
elementary
  matrices, 21
  row operations, 21
elementary operations, 81
EP matrix, 128
EP_r matrix, 128
equivalent matrices, 15
equivalent norms, 11
Erdelyi g.i., 236
essentially strictly convex norm, 104
Euclidean norm, 9
extension, 73
extremal g.i., 252
extremal solution, 249

factorization, 89
  Cholesky, 100
  full–rank, 37, 128
  QR, 89
Fredholm integral operators, 230
full–rank factorization, 37, 128
function
  convex, 104
  strictly convex, 104

gauge function, 111, 197

symmetric, 111
Gaussian elimination, 23
generalized
power, 206
generalized Green's function, 242
generalized inverse, 1
S–inverse, 131
S–restricted, 73, 98
S'–inverse, 136
α–β, 107, 120
$\{1,2,3\}$–inverse, 35, 149
$\{1,2,4\}$–inverse, 35, 149
$\{1,2,5\}$–inverse, 127
$\{1,2\}$–inverse, 34, 149
$\{1,3\}$–inverse, 87, 97
$\{1,4\}$–inverse, 97
$\{1\}$–inverse, 29
$\{1^k,2,5\}$–inverse, 125
$\{2\}$–inverse, 40, 42
$\{i,j,\dots,k\}$–inverse, 27
associated with α,β, 107, 120
constrained, 76
Drazin inverse, 127, 132, 134
maximal g.i., 233
Moore–Penrose inverse, 149
quasi–commuting inverse, 139
reverse order property, 141
strong spectral inverse, 139
Tseng, 231
weighted, 98
generalized resolvent, 203
grade, 125
Gram–Schmidt orthogonalization process, 211, 213
graph, 82, 226
branches, 82
connected, 83
incidence matrix, 82
nodes, 82
connected, 83
directly connected, 83
Green's function, 242
group inverse, 127
GSO, 211, 213

Hermite normal form, 22

idempotent, 29, 51
ill–conditioned, 88
incidence matrix, 82, 85
index, 127
of nilpotency, 140
inequality
Cauchy–Schwartz, 9, 113
generalized Cauchy–Schwartz, 113
Minkowski, 10
triangle, 8
inner product, 8, 225
standard, 8
integral matrix, 80
integral vector, 80
interval linear program, 78
bounded, 79
consistent, 79
invariant factors, 81
inverse
Bott–Duffin, 76, 82, 122
Drazin, 134
generalized, 1
Moore–Penrose, 27, 97, 103, 248
weighted, 101
inverse graph, 226
isometry, 186
linearity of, 191
partial, 186, 191
iterative method, 213
$p\,\underline{\text{th}}$–order, 214

Kirchhoff, 84
current law, 84, 123
voltage law, 84, 123
Kronecker product, 46

least extremal solution, 249
least upper bound, 116
least–squares solution
constrained, 92
minimum–norm, 94
length, 8
linear equations
approximate solution, 87
ill–conditioned, 88
least–squares solution, 87
linear manifold, 151
orthogonal representation, 151
linear operator
adjoint, 227
bounded, 226
carrier, 226
closable, 228
closed, 226
closure, 228
dense, 227
domain, 226
extension, 227
graph, 226
inverse graph, 226
non–negative, 228
normally solvable, 228
null space, 226
orthogonal projector, 228
range, 226
restriction, 227
self–adjoint, 228
symmetric, 228
linear transformation, 11
extension, 73
intrinsic property, 16
inverse, 12

inverse image, 12
invertible, 12
matrix representation, 12
nonsingular, 12
null space, 12
range, 11
restriction, 73

matrices
convergent, 20
diagonable, 53
EP, 128
EP_r, 128
equivalent, 15
equivalent over $\mathbb{Z}$, 81
idempotent, 29, 51
ill–conditioned, 88
normal, 61
orthogonally similar, 14
range–Hermitian, 128
similar, 14
unitarily equivalent, 15
unitarily similar, 14
matrix
UDV^*–decomposition, 176
admittance, 85
condition number, 178
incidence, 82, 85
index, 132
integral, 80
invariant factors, 81
nilpotent part, 138
normal form
Hermite, 22
Smith, 81
permutation, 21
polar decomposition, 188
reduced row–echelon form, 22
singular value decomposition, 176
singular values, 176
square root, 100, 190
transfer, 84
unit, 80
matrix norm, 13
multiplicative, 13
spectral, 18, 178
unitarily invariant, 19
matrix norms
unitarily invariant, 197
maximal g.i., 233
minimum–norm least–squares solution, 94
minimum–norm solution, 93
Minkowski functional, 111
Minkowski's inequality, 10
Moore–Penrose inverse, 27, 97, 103, 149, 248
iterative methods, 213
multiplicative norm, 13

Newton's method, 40, 42
nilpotent part, 138
Noble's method, 90, 212
norm, 8, 225
ℓ_p, 10, 113
consistent, 17
e.s.c., 104, 119
ellipsoidal, 9, 117
equivalent, 11
essentially strictly convex, 104
Euclidean, 9
matrix, 13
of homogeneous transformation, 116
projective, 117
unitarily invariant, 113
weighted Euclidean, 9
normal matrix, 61
norms
dual, 120
null space, 12, 31, 96, 226

Ohm's law, 84, 123
orthogonal, 10
projection, 60
projector, 60
orthogonal complement, 225
orthogonal direct sum, 225
orthogonally similar matrices, 14

partial isometry, 186, 191
Penrose equations, 27, 125, 235, 248
permutation
even, 22
inverse, 21
matrix, 21
odd, 22
sign, 22
permutation matrix, 21
polar decomposition, 188
potential, 78, 83
principal idempotents, 54
principal vector, 125
projection, 7
orthogonal, 60
projective
bound, 116
norm, 117
projector
ϕ–metric, 104
orthogonal, 60
pseudo inverse, 239
pseudo resolvent, 239
Hurwitz construction, 239
pseudoinverse, 1

QR factorization, 89
quasi–commuting inverse, 139

range, 11, 31, 96, 226
range–Hermitian matrix, 128
reduced row–echelon form, 22

regular value, 238
residual, 87, 213
resolvent, 203
  first equation, 203
  generalized, 203
restriction, 73
reverse order property, 141
rotund convex set, 115

similar matrices, 14
singular value decomposition, 176
singular values, 176
Smith normal form, 81
smooth convex set, 115
solution
  α–approximate, 108
  approximate, 87
  least–squares, 87
  minimum–norm, 93
  Tchebycheff, 113
spectral
  condition number, 179
  decomposition, 54, 100
  norm, 18, 178
  radius, 19
square root of a matrix, 190
standard
  basis, 7
  inner product, 8
standard basis, 12
stationary
  point, 122
  value, 122
strictly convex function, 104
strong spectral inverse, 139
SVD, 176

Tchebycheff
  approximate solution, 113
  norm, 10
tree, 86
triangle inequality, 8
Tseng generalized inverse, 231

unit matrix, 80
unitarily equivalent matrices, 15
unitarily invariant matrix norms, 197
unitarily invariant norm, 19
unitarily similar matrices, 14

vector
  integral, 80
  length, 8
  norm, 8, 113
  principal, 125

weighted generalized inverse, 98
weighted inverse, 101

Author Index

Verghese, G. C., 172

Afriat, S. N., 86, 173
Albert, A., 151, 169, 206, 207, 213, 224
Altman, M., 40
Anderson, Jr., W. N., 86, 152
Arghiriade, E., 141–143, 230–232
Atkinson, F. V., 260
Atkinson, K. E., 260
Autonne, L., 179, 180, 188

Balakrishnan, A. V., 96, 124
Barnett, S., 124
Bauer, F. L., 115
Beckenbach, E. F., 10
Bellman, R., 10, 15
Beltrami, E. J., 86
Ben–Israel, A., 40, 86, 97, 124, 135, 143, 154, 159, 172, 173, 176, 195, 207, 220, 224, 229, 249, 250, 252, 256, 258, 259
Berman, A., 260
Beutler, F. J., 260
Björck, Å., 88, 89, 93
Bjerhammar, A., 2, 34, 45, 86
Blattner, J. W., 171, 173
Bohnenblust, F., 117
Bonnesen, T., 111
Bott, R., 76, 82, 122
Boullion, T. L., 86, 102, 107, 143, 208, 224
Bounitzky, E., 242
Bowman, V. J., 86
Bradley, J. S., 260
Brand, L., 149, 209
Burdet, C.- A., 86
Burns, F., 173
Burrus, W. R., 213
Businger, P. A., 88, 208, 213, 224

Carlson, D., 173
Charnes, A., 78, 86, 95, 136, 207
Cheney, E. W., 114
Chipman, J. S., 86, 101
Cimmino, G., 173
Clarkson, J. A., 104
Cline, R. E., 60, 78, 128, 132, 139, 142, 162, 212, 213
Coddington, E. A., 245
Cohen, D., 220
Cooper, W. W., 86
Courant, R., 260
Cudia, D. F., 115
Cullen, C. G., 23

Davis, D. L., 260
Decell, Jr., H. P., 60, 207, 224
den Broeder Jr., C. G., 95, 136
Dennis, J. B., 123
Desoer, C. A., 124, 237
Drazin, M. P., 127, 132, 134
Duffin, R. J., 76, 82, 86, 122, 152
Dunford, N., 191, 200, 203, 205

Eckart, C., 179, 180, 183, 185
Elliott, W. W., 242
Englefield, M. J., 127, 134
Erdelsky, P. J., 103, 108, 116–121
Erdélyi, I., 124, 127, 129, 139, 186, 192, 193, 208, 229, 231, 235, 236, 249, 250, 252

Fan, Ky, 184, 194
Fenchel, W., 111
Fisher, A. G., 29
Fletcher, R., 40, 169, 170
Forsythe, G. E., 93
Frame, J. S., 205
Franck, P., 183
Fredholm, I., 4, 239
Funderlic, R. E., 49

Gaches, J., 183
Gale, K. J., 23
Gantmacher, F. R., 195, 205
Garnett III, J. M., 224
Germain–Bonne, B., 173, 224
Glazman, I. M., 86
Goldberg, S., 228
Goldstein, A. A., 110
Golub, G. H., 88, 89, 93, 143, 183, 184, 208, 213, 224
Good, I. J., 208
Granot, F., 86
Graybill, F. A., 86, 224
Green, B., 185
Greub, W., 260
Greville, T. N. E., 43, 47, 64, 65, 86, 132, 135, 139, 141, 165, 167, 211, 213
Guillemin, E. A., 123

Halmos, P. R., 51, 186, 192, 193, 208, 229, 241
Halperin, I., 244
Hamburger, H., 260
Hansen, G. W., 260
Hartwig, R. E., 43, 173, 208

Harwood, W. R., 173
Hawkins, J. B., 176
Haynsworth, E., 173
Hearon, J. Z., 86, 173, 193, 208
Hestenes, M. R., 176, 186, 187, 194, 206–208, 228, 234, 235, 243, 260
Hilbert, D., 4, 242, 260
Ho, B. L., 124
Ho, Y. C., 124
Hoffman, A. J., 184, 194
Holmes, R. B., 103, 238, 255, 260
Householder, A. S., 13, 20, 113, 115, 183, 185, 214, 220
Hurt, M. F., 80
Hurwitz, W. A., 4, 239

Ijiri, Y., 224

Jones, Jr., J., 86

Kahan, W., 183, 208, 213
Kakutani, S., 117
Kallina, C., 260
Kalman, R. E., 124
Kammerer, W. J., 224, 230, 259, 260
Kantorovich, L. V., 241
Katz, I. J., 130, 143, 159
Kirby, M. J. L., 86
Kishi, F. H., 124, 169
Korganoff, A., 260
Krylov, V. I., 241
Kublanovskaya, V. N., 224
Kuo, M. C. Y., 124
Kurepa, S., 237

Lancaster, P., 203, 205
Lanczos, C., 176, 208
Landesman, E. M., 234, 237, 244
Langenhop, C. E., 51, 57
Leach, E. B., 40, 260
Lent, A. H., 153, 154, 260
Leringe, Ö., 124
Levinson, N., 245
Lewis, T. O., 86, 102, 143
Ljubich, Ju. I., 86
Locker, J., 244, 246, 260
Lonseth, A. T., 221, 260
Loud, W. S., 248, 249
Lovass–Nagy, V., 173

MacDuffee, C. C., 37, 127, 179, 213
Marcus, M., 14, 22, 23, 46, 81, 177, 179, 209
Markham, T., 173
Marsaglia, G., 86
Mazda, L. F., 124
McLaughlin, J. E., 186, 192
Meany, R. K., 224
Meicler, M., 114
Meyer, Jr., C. D., 149, 173, 224
Mihályffy, L., 159
Miller, F. R., 208
Milne, R. D., 57
Minamide, N., 75, 98, 124, 255
Minc, H., 14, 22, 23, 46, 81, 177, 179, 209
Mirsky, L., 184, 197
Mitra, S. K., 59, 86
Moler, C. B., 91
Moore, E. H., 4, 27
Moore, R. H., 143
Morris, G. L., 86, 165
Murray, F. J., 188

Nakamura, K., 75, 98, 124, 255
Nanda, V. C., 195
Narendra, K. S., 124
Nashed, M. Z., 124, 143, 224, 230, 259, 260
Nelson, D. L., 86
Neumann, M., 260
Newman, T. G., 103, 108, 119
Noble, B., 19, 88, 90, 149, 212, 213, 224

Odell, P. L., 86, 103, 107, 108, 119, 143, 165
Osborne, E. E., 124

Painter, R. J., 149, 224
Parlett, B. N., 89
Pavel-Parvu, M., 260
Pearl, M. H., 128, 130, 134, 143, 195
Pennington, R. H., 210
Penrose, R., 2, 27, 45, 47, 60, 94, 97, 176, 181, 188, 207
Pereyra, V., 143, 224
Peters, G., 124, 224
Petryshyn, W. V., 59, 86, 220, 235, 238, 259
Pietsch, A., 260
Plemmons, R. J., 260
Poole, G. D., 143, 208
Porter, W. A., 92, 124, 256, 260
Powers, D. L., 173
Przeworska–Rolewicz, D., 260
Pyle, L. D., 78, 157, 158, 224

Rall, L. B., 260
Rao, C. R., 43, 86, 101, 149, 209
Rao, M. M., 86
Reid, W. T., 171–173, 225, 242, 246–248
Reinsch, C., 208, 213, 224
Rheinboldt, W. C., 40, 260
Rice, J., 89
Rigal, J.-L., 183
Rinehart, R. F., 205
Robers, P. D., 78
Robert, P., 143, 150
Robinson, D. W., 57, 260
Robinson, S. M., 25
Rockafellar, R. T., 103, 111, 123
Rohde, C. A., 46, 151
Rolewicz, S., 260
Rosen, J. B., 86, 124, 224
Rosenbloom, P. C., 257
Rousset de Pina, X., 183

Rust, B., 213

Schneeberger, C., 213
Schulz, G., 220
Schwartz, J., 244
Schwartz, J. T., 191, 200, 203, 205
Schwerdtfeger, H., 128
Schönemann, P. H., 185
Scroggs, J. E., 143
Sheffield, R. D., 86, 260
Shinozaki, N., 224
Showalter, D. W., 224, 256, 258, 259
Sibuya, M., 224
Singer, I., 104, 106
Sittler, R. W., 169
Smithies, F., 230, 238, 241
Stakgold, I., 260
Stallings, W. T., 224
Stein, P., 21
Stern, T. E., 123
Stewart, G. W., 49, 59, 86, 143, 220
Söderstörm, T., 220

Tanabe, K., 224
Taussky, O., 21, 88, 179
Taylor, A. E., 225, 228, 259
Tewarson, R. P., 224
Thompson, G. L., 86
Tseng, Y. Y., 4, 231, 232, 249, 255
Tucker, D. H., 260

Urquhart, N. S., 35, 58, 224

Vogt, A., 191
von Neumann, J., 111, 157, 186, 188, 196, 197, 199, 226, 228
Votruba, G. F., 260

Wahba, G., 124
Waid, C., 80
Wallen, L. J., 208
Wang, Guorong, 173
Ward, J. F., 86, 102, 143
Wedderburn, J. H. M., 86, 137
Wedin, P.-Å., 124, 143, 208
Werner, H. -J., 173
Weyl, H., 179
Whalen, B. H., 124, 237
Whitney, T. M., 224
Wilkinson, J. H., 88, 89, 100, 124, 125, 224
Williams, J. P., 124, 256, 260
Williamson, J., 185, 188
Willner, L. B., 31, 224
Wimmer, H. K., 205
Wyler, O., 260

Yau, S. S., 224
Yosida, K., 225
Young, G., 179, 180, 183, 185

Zadeh, L. A., 124
Zarantonello, E. H., 260
Ziebur, A. D., 205
Zlobec, S., 59, 78, 86, 150, 211, 213, 220, 224, 259, 260