Graduate Series in Mathematics **5**

Wang Guorong Wei Yimin Qiao Sanzheng

Generalized Inverses: Theory and Computations

(广义逆：理论与计算)

Science Press
Beijing/ New York

Wang Guorong   Wei Yimin   Qiao Sanzheng

Generalized Inverses: Theory and Computations

(广义逆:理论与计算)

Book published by the support of the Graduate Textbook publishing Foundation of Shanghai Education Committee

SCIENCE PRESS
Beijing/New York

Responsible Editors: Lin Peng Chen Yuzhuo

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, China

Printed in Beijing

ISBN 7-03-012437-5/O·1869(Beijing)

Preface

The concept of generalized inverses was first introduced by I. Fredholm[57] in 1903, where a generalized inverse of an integral operator was given and was called "pseudoinverse". Generalized inverses of differential operators were implied in D. Hilbert's[78] discussion of generalized Green's functions in 1904. See W. Reid's[131] paper in 1931 for a history of generalized inverses of differential operators.

The generalized inverse of matrices was first introduced by E. H. Moore[113] in 1920, who defined a unique generalized inverse by means of projectors of matrices. Little was done in the next 30 years until mid-1950s when discoveries of the least-squares properties of certain generalized inverses and the relationship of generalized inverses to solutions of linear systems brought new interests in the subject. In particular, R. Penrose[119] showed in 1955 that the Moore's inverse is the unique matrix satisfying four matrix equations. This important discovery revived the study of gene-ralized inverses. In honor of Moore and Penrose's contribution, this unique generalized inverse is called the Moore-Penrose inverse.

The theory, applications and computational methods of generalized inverses have been developing rapidly during the last 50 years. One milestone is the publication of several monographs ([7], [13], [65]and[129]) on the subject in 1970s, particularly, the excellent volume by A. Ben-Israel and T. N.E Greville[7] which has made a long lasting impact on the subject; the other milestone is the publications of two volumes of proceedings. The first is the volume of proceedings[114]of the Advanced Seminar on Generalized Inverses and Applications held at the University of Wisconsin-Madison in 1973 edited by M. Z. Nashed. It is an excellent and extensive survey book. It contains 14 survey papers on the theory, computations and applications of generalized inverses and an exhaustive bibliography that includes all related references up to 1975. The other is the volume of proceedings[11] of the AMS Regional Conference held in Columbia, South Carolina in 1976 edited by S. L. Campbell. It is a new survey book containing 12 papers on the latest applications of generalized inverses. The volume describes changes in research directions and types of generalized inverses since mid-1970s. Prior to this period, due to the applications in statistics, research often centered on generalized inverses for solving linear systems and generalized inverses with least-squares properties. Recent studies focus on such topics as: infinite dimensional theory, numerical computation, matrices of special types (Boolean, integral), matrices over algebraic structures other than real or complex fields, systems theory and non-equation solving generalized inverses.

I have been conducting teaching and research in generalized inverses of matrices since 1976. I gave a course "Generalized Inverses of Matrices" and held many seminars for graduate students majoring in Computational Mathematics in our department. Since 1979, my colleagues and I with graduate students have obtained a number of results on generalized inverses in the areas of perturbation theory, condition numbers, recursive algorithms, finite algorithms, imbedding algorithms, parallel algorithms, generalized inverses of rank-r modified matrices and Hessenberg matrices, extensions of the Cramer rules and the representation and approximation of generalized inverses of linear operators. Dozens of papers are published in refereed journals in China and other countries. They draw attentions from researchers around world. I have received letters from more than ten universities in eight countries, U.S.A., Germany, Sweden, etc. requesting papers or seeking academic contacts. Colleagues in China show strong interests and support in our work, and request systematic presentation of our work. With the support of the Academia Sinica Publishing Foundation and the National Natural Science Foundation of China, Science Press published my book "Generalized Inverses of Matrices and Operators"[169] in Chinese in 1994. That book is noticed and welcomed by researchers and colleagues in China. It has been adopted by several universities as textbook or reference book for graduate students. The book was reprinted in 1998.

In order to improve graduate teaching and international academic exchange, I was encouraged to write this English version based on the Chinese version. This English version is not a direct translation of the Chinese version. In addition to the contents in the Chinese version, this book includes the contents from more than 100 papers since 1994. The final product is an entirely new book, while the spirit of the Chinese version still lives. For example, Sections 2, 3 and 5 of Chapter 3, Section 1 of Chapter 6, Sections 4 and 5 of Chapter 7, Sections 1, 4 and 5 of Chapter 8, Chapters 4, 10 and 11 are all new.

Dr. Wei Yimin of Fudan University in China and Dr. Qiao Sanzheng of McMaster University in Canada were two of my former excellent students. They have made many achievements in the area of generalized inverses and are recognized internationally. I would not possibly finish this book without their collaborations.

We would like to thank Professor A. Ben-Israel, Dr. Miao Jianming of Rutgers University, and Professors R.E. Hartwig, S.L. Campbell and C. D. Meyer, Jr. of North Carolina State University, and Professor C. W. Groetsch of University of Cincinnati. The texts[7], [13] and [65] undoubtedly have had an influence on this book. We also thank Professor Jiang Erxiong of Shanghai University, Professor Cao Zhihao of Fudan University, Professor Wei Musheng and Chen Guoliang of East-China Normal University and Professor Chen Yonglin of Nanjing Normal University for their help and advice in the subject for many years, and my doctoral student Yu Yaoming for typing this book.

I appreciate any comments and corrections from the readers.

Finally, I am indebted to the support by the Graduate Textbook Publishing Foundation of Shanghai Education Committee and Shanghai Normal University.

Wang Guorong
Shanghai Normal University
June 2003

Contents

Chapter 1

Equation Solving Generalized Inverses

1.1 The Moore-Penrose inverse

Let $\mathbb{C}$ ($\mathbf{R}$) be the field of complex (real) numbers, $\mathbb{C}^n(\mathbf{R}^n)$ the vector space of n-tuples of complex (real) numbers over $\mathbb{C}$ ($\mathbf{R}$), $\mathbb{C}^{m\times n}$ ($\mathbf{R}^{m\times n}$) the class of $m \times n$ complex (real) matrices, $\mathbb{C}_r^{m\times n}$ ($\mathbf{R}_r^{m\times n}$) the class of $m \times n$ complex (real) matrices of rank r, and $R(A) = \{\boldsymbol{y} \in \mathbb{C}^m : \boldsymbol{y} = A\boldsymbol{x},\ \boldsymbol{x} \in \mathbb{C}^n\}$ the range of $A \in \mathbb{C}^{m\times n}$. It is well known that every nonsingular matrix $A \in \mathbb{C}_n^{n\times n}$ has a unique matrix $X \in \mathbb{C}_n^{n\times n}$ satisfying

$$AX = I,\ XA = I, \tag{1.1.1}$$

where I is the identity matrix. This X is called the inverse of A, and is denoted by $X = A^{-1}$.

The nonsingular system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b} \quad (A \in \mathbb{C}_n^{n\times n},\ \boldsymbol{b} \in \mathbb{C}^n) \tag{1.1.2}$$

has a unique solution for $\boldsymbol{x}$, given by

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

In the general case, A may be singular or rectangular, there may exist no solution or multiple solutions. The consistent system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}\ (A \in \mathbb{C}^{m\times n},\ m < n,\ \boldsymbol{b} \in R(A)) \tag{1.1.3}$$

has many solutions. The inconsistent system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}\ (A \in \mathbb{C}^{m\times n},\ \boldsymbol{b} \notin R(A)) \tag{1.1.4}$$

has no solution. However, it has a least-squares solution.

Can we find a suitable matrix X, such that $\boldsymbol{x} = X\boldsymbol{b}$ is some kind of solution of the equation $A\boldsymbol{x} = \boldsymbol{b}$? This X is called the equation solving generalized inverse. A generalized inverse reduces to the usual inverse when A is nonsingular. The Moore-Penrose inverse and $\{i,\ j,\ k\}$ inverses which will be discussed in this chapter are the classes of generalized inverses.

1.1.1 The definition and the basic properties of $A^\dagger$

Let $A \in \mathbb{C}_n^{m\times n}$ and A^* denote the conjugate transpose of A. Then A^*A is the nonsingular matrix of order n, and the least-squares solution $\boldsymbol{x}$ of the inconsistent system of linear equations (1.1.4) can be solved by the following normal equations

$$A^*A\boldsymbol{x} = A^*\boldsymbol{b}, \tag{1.1.5}$$

we have $\boldsymbol{x}=(A^*A)^{-1}A^*\boldsymbol{b}$. Let

$$X = (A^*A)^{-1}A^*. \tag{1.1.6}$$

It can be verified that X is the unique matrix satisfying the following four equations (usually called the Penrose conditions)

(1) $AXA = A$,
(2) $XAX = X$,
(3) $(AX)^* = AX$,
(4) $(XA)^* = XA$.

This X is called the Moore-Penrose generalized inverse of A, and is denoted by $X=A^\dagger$. Thus the least-squares solution of (1.1.5) is $\boldsymbol{x} = A^\dagger\boldsymbol{b}$.

Especially, if $m = n = \text{rank}(A)$, we have

$$A^\dagger = (A^*A)^{-1}A^* = A^{-1}(A^*)^{-1}A^* = A^{-1}.$$

This shows that Moore-Penrose inverse $A^\dagger$ reduces to usual inverse A^{-1} when A is nonsingular.

For general $m \times n$ matrices, we have

Definition 1.1.1 *Let $A \in \mathbb{C}^{m\times n}$. Then the matrix $X \in \mathbb{C}^{n\times m}$ satisfying the Penrose conditions (1) $\sim$ (4) is called the Moore-Penrose inverse of A (abbreviated as the M-P inverse), and is denoted by $X = A^\dagger$.*

In the following, we will show the existence and uniqueness of the matrix $X \in \mathbb{C}^{n\times m}$ in Definition 1.1.1.

Theorem 1.1.1 *The generalized inverse X satisfying the Penrose conditions (1) $\sim$ (4) is existent and unique.*

Proof Let $A \in \mathbb{C}_r^{m\times n}$. Then A can be decomposed as $A = Q^*RP$ (see for example [141]), where Q and P are the unitary matrices of orders m and n respectively and

$$R = \begin{pmatrix} R_{11} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{C}^{m\times n},$$

where R_{11} is the nonsingular upper triangular matrix of order r. Denote

$$R^\dagger = \begin{pmatrix} R_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{C}^{n\times m},$$

then $X = P^*R^\dagger Q$ satisfies the Penrose conditions (1) $\sim$ (4). In fact,

$$AXA = Q^*RPP^*R^\dagger QQ^*RP = Q^*RP = A,$$
$$XAX = P^*R^\dagger QQ^*RPP^*R^\dagger Q = P^*R^\dagger Q = X,$$
$$(AX)^* = (Q^*RPP^*R^\dagger Q)^* = (Q^*\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}Q)^* = AX,$$
$$(XA)^* = (P^*R^\dagger QQ^*RP)^* = (P^*\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}P)^* = XA.$$

Therefore, for any $A \in \mathbb{C}_r^{m\times n}$, $X = A^\dagger$ always exists. The uniqueness of X is proved as follows.

If X_1 and X_2 both satisfy the Penrose conditions (1) $\sim$ (4), then

$$\begin{aligned} X_1 &= X_1AX_1 = X_1AX_2AX_1 \\ &= X_1(AX_2)^*(AX_1)^* = X_1(AX_1AX_2)^* \\ &= X_1(AX_2)^* = X_1AX_2 \\ &= X_1AX_2AX_2 = (X_1A)^*(X_2A)^*X_2 \\ &= (X_2AX_1A)^*X_2 = (X_2A)^*X_2 \\ &= X_2AX_2 = X_2. \end{aligned}$$

□

The next theorem lists some of properties of the M-P inverse.

Theorem 1.1.2 *Let $A \in \mathbb{C}^{m\times n}$. Then*

(1) $(A^\dagger)^\dagger = A$;

(2) $(\lambda A)^\dagger = \lambda^\dagger A^\dagger$, *where* $\lambda \in \mathbb{C}$, $\lambda^\dagger = \begin{cases} \dfrac{1}{\lambda}, & \lambda \neq 0, \\ 0, & \lambda = 0; \end{cases}$

(3) $(A^*)^\dagger = (A^\dagger)^*$;

(4) $(AA^*)^\dagger = (A^*)^\dagger A^\dagger$; $(A^*A)^\dagger = A^\dagger(A^*)^\dagger$;

(5) $A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger$;

(6) $A^* = A^*AA^\dagger = A^\dagger AA^*$;

(7) *If* $\text{rank}(A) = n$, *then* $A^\dagger A = I_n$;
if $\text{rank}(A) = m$, *then* $AA^\dagger = I_m$;

(8)$(UAV)^\dagger = V^*A^\dagger U^*$, *where U and V are unitary matrices.*

The above properties can be checked by using Definition 1.1.1. The proof is left as an exercise.

1.1.2 The range and null space of a matrix

Definition 1.1.2 *Let $A \in \mathbb{C}^{m\times n}$. We denote by*

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

We can prove

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

where $R(A)^{\perp}$ is the orthogonal complementary subspace of $R(A)$, i.e., the set of all vectors in $\mathbb{C}^n$ which are orthogonal to every vector in $R(A)$. Every $\boldsymbol{x} \in \mathbb{C}^n$ can be expressed uniquely as a sum

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}, \ \ \boldsymbol{y} \in R(A), \boldsymbol{z} \in R(A)^{\perp}.$$

Theorem 1.1.3 (*The basic properties of the range and null space*)
(1) $R(A) = R(AA^{\dagger}) = R(AA^*)$;
(2) $R(A^{\dagger}) = R(A^*) = R(A^{\dagger}A) = R(A^*A)$;
(3) $R(I - A^{\dagger}A) = N(A^{\dagger}A) = N(A) = R(A^*)^{\perp}$;
(4) $R(I - AA^{\dagger}) = N(AA^{\dagger}) = N(A^{\dagger}) = N(A^*) = R(A)^{\perp}$;
(5) $R(AB) = R(A) \ \Leftrightarrow \text{rank}(AB) = \text{rank}(A)$;
(6) $N(AB) = N(B) \Leftrightarrow \text{rank}(AB) = \text{rank}(B)$.

The proof is left as an exercise.
The following properties of rank are used in this book.

Lemma 1.1.1 *Let* $A \in \mathbb{C}^{m\times n}$, $E_A = I_m - AA^{\dagger}$, *and* $F_A = I_n - A^{\dagger}A$. *Then*
(1) $\text{rank}(A) = \text{rank}(A^{\dagger}) = \text{rank}(A^{\dagger}A) = \text{rank}(AA^{\dagger})$;
(2) $\text{rank}(A) = m - \text{rank}(E_A), \ \text{rank}(A) = n - \text{rank}(F_A)$;
(3) $\text{rank}(AA^*) = \text{rank}(A) = \text{rank}(A^*A)$.

The proof is left as an exercise.

1.1.3 Full-rank factorization

A non-null matrix that is of neither full column rank nor full row rank can be expressed as the product of a matrix of full column rank and a matrix of full row rank. We call a factorization with the above property a full-rank factorization of a non-null matrix. This factorization turns out to be a powerful tool in the study of the generalized inverses.

Theorem 1.1.4 *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*

$$A = FG. \tag{1.1.7}$$

Proof Let F be any matrix whose columns are a basis for $R(A)$. Then $F = (\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_r) \in \mathbb{C}_r^{m\times r}$. Let $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$. Then every column $\boldsymbol{a}_i$ of A is uniquely representable as a linear combination of the columns of F,

$$\boldsymbol{a}_i = g_{1i}\boldsymbol{f}_1 + g_{2i}\boldsymbol{f}_2 + \cdots + g_{ri}\boldsymbol{f}_r \quad (i = 1, 2, \cdots, n). \tag{1.1.8}$$

Hence

$$\begin{aligned} A &= (\ \boldsymbol{a}_1, \ \ \boldsymbol{a}_2, \ \ \cdots, \ \ \boldsymbol{a}_n\ ) \\ &= (\ \boldsymbol{f}_1, \ \ \boldsymbol{f}_2, \ \ \cdots, \ \ \boldsymbol{f}_r\ ) \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{r1} & g_{r2} & \cdots & g_{rn} \end{pmatrix} \\ &\equiv FG. \end{aligned}$$

The matrix $G \in \mathbb{C}^{r\times n}$ is uniquely determined by the above equality. It is obvious, $\text{rank}(G) \le r$. Since

$$\text{rank}(G) \ge \text{rank}(FG) = r,$$

thus $\text{rank}(G) = r$. □

Let $A = FG$ be a full-rank factorization of A and $C \in \mathbb{C}_r^{r\times r}$. Then

$$A = (FC)(C^{-1}G) \equiv F_1G_1$$

is also a full-rank factorization of A. Thus the full-rank factorization of A is not unique. A practical algorithm of a full-rank factorization is given in Chapter 4. MacDuffe[97]pointed out that a full-rank factorization of A leads to an explicit formula for its M-P inverse $A^\dagger$.

Theorem 1.1.5 *Let $A \in \mathbb{C}_r^{m\times n}$, $r > 0$, and its full-rank factorization $A = FG$. Then*

$$A^\dagger = G^*(F^*AG^*)^{-1}F^* = G^*(GG^*)^{-1}(F^*F)^{-1}F^*. \tag{1.1.9}$$

Proof First we show that F^*AG^* is nonsingular. By $A = FG$,

$$F^*AG^* = (F^*F)(GG^*),$$

and F^*F and GG^* are $r \times r$ matrices. Also by Lemma 1.1.1, both are of rank r. Thus F^*AG^* is the product of two nonsingular matrices, and therefore F^*AG^* is nonsingular and

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

Denoting by X the right member of (1.1.9), we have

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

and it is easy to verify that this expression for X satisfies the Penrose conditions (1) $\sim$ (4). By the uniqueness of M-P inverse $A^\dagger$, (1.1.9) is therefore established. □

1.1.4 Moore-Penrose inverse and the minimum-norm least-squares solution of an inconsistent system of linear equations

Let $\boldsymbol{x} = (x_1,\ x_2,\ \cdots,\ x_p)^* \in \mathbb{C}^p$. Then

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{p} |x_i|^2\right)^{\frac{1}{2}} = (\boldsymbol{x}^*\boldsymbol{x})^{\frac{1}{2}} \tag{1.1.10}$$

is the 2-norm of $\boldsymbol{x}$, for convenience, set $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|$.

Let $\boldsymbol{u},\ \boldsymbol{v} \in \mathbb{C}^p$, and $(\boldsymbol{u},\ \boldsymbol{v})= 0$, i. e., $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal. Then

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = (\boldsymbol{u}+\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v}) = (\boldsymbol{u},\boldsymbol{u}) + (\boldsymbol{v},\boldsymbol{u}) + (\boldsymbol{u},\boldsymbol{v}) + (\boldsymbol{v},\boldsymbol{v}) = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2. \tag{1.1.11}$$

This is the Pythagorean theorem.

Now consider the problem of finding a solution $\boldsymbol{x}$ to an inconsistent system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}\ (A \in \mathbb{C}^{m\times n},\ \boldsymbol{b} \notin R(A)). \tag{1.1.4}$$

We look for $\boldsymbol{x}$ that makes $\|A\boldsymbol{x} - \boldsymbol{b}\|$ as small as possible.

Definition 1.1.3 *Let $A \in \mathbb{C}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{C}^m$. Then a vector $\boldsymbol{u} \in \mathbb{C}^n$ is called a least-squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ if $\|A\boldsymbol{u} - \boldsymbol{b}\| \le \|A\boldsymbol{v} - \boldsymbol{b}\|$ for all $\boldsymbol{v} \in \mathbb{C}^n$.*

Definition 1.1.4 *Let $A \in \mathbb{C}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{C}^m$. Then a vector $\boldsymbol{u} \in \mathbb{C}^n$ is called a minimum-norm least-squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ if $\boldsymbol{u}$ is a least-squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ and $\| \boldsymbol{u} \| < \| \boldsymbol{w} \|$ for any other least-squares solution $\boldsymbol{w}$.*

If $\boldsymbol{b} \in R(A)$, the system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then notations of solution and the least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ obviously coincide.

The next theorem shows the relation between the minimum-norm least-squares solution (1.1.4) and M-P inverse $A^\dagger$.

Theorem 1.1.6 *Let $A \in \mathbb{C}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{C}^m$. Then $A^\dagger \boldsymbol{b}$ is the minimum-norm least-squares solution of* (1.1.4).

Proof Let $\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2$, where

$$\boldsymbol{b}_1 = AA^\dagger \boldsymbol{b} \in R(A),\ \boldsymbol{b}_2 = (I - AA^\dagger)\boldsymbol{b} \in R(A)^\perp.$$

Then $A\boldsymbol{x} - \boldsymbol{b}_1 \in R(A)$ and

$$\|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \|A\boldsymbol{x} - \boldsymbol{b}_1 + (-\boldsymbol{b}_2)\|^2 = \|A\boldsymbol{x} - \boldsymbol{b}_1\|^2 + \|\boldsymbol{b}_2\|^2.$$

Thus $\boldsymbol{x}$ will be a least-squares solution if and only if $\boldsymbol{x}$ is a solution of the consistent system $A\boldsymbol{x} = AA^\dagger \boldsymbol{b}$. It is obvious that $A^\dagger \boldsymbol{b}$ is a particular solution. From Theorem 1.1.3,

$$N(A) = \{(I - A^\dagger A)\boldsymbol{h}\ :\ \boldsymbol{h} \in \mathbb{C}^n\},$$

thus the general solution of the consistent system $A\boldsymbol{x} = AA^\dagger \boldsymbol{b}$ is

$$\boldsymbol{x} = A^\dagger \boldsymbol{b} + (I - A^\dagger A)\boldsymbol{h},\ \boldsymbol{h} \in \mathbb{C}^n.$$

Because

$$\begin{aligned} \|A^\dagger \boldsymbol{b}\|^2 &< \|A^\dagger \boldsymbol{b}\|^2 + \|(I - A^\dagger A)\boldsymbol{h}\|^2 \\ &= \|A^\dagger \boldsymbol{b} + (I - A^\dagger A)\boldsymbol{h}\|^2,\ (I - A^\dagger A)\boldsymbol{h} \neq \boldsymbol{0}, \end{aligned}$$

$\boldsymbol{x} = A^\dagger \boldsymbol{b}$ is the minimum-norm least-squares solution of (1.1.4). □

In some applications, the minimality of a least-squares solution is important, in others it is not important. If the minimality is not important, then the next theorem can be very useful.

Theorem 1.1.7 *Let $A \in \mathbb{C}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{C}^m$. Then the following statements are equivalent:*

(1) *$\boldsymbol{u}$ is a least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$;*

(2) *$\boldsymbol{u}$ is a solution of $A\boldsymbol{x} = AA^\dagger \boldsymbol{b}$;*

(3) *$\boldsymbol{u}$ is a solution of $A^* A\boldsymbol{x} = A^* \boldsymbol{b}$;*

(4) *$\boldsymbol{u}$ is of the form $A^\dagger \boldsymbol{b} + \boldsymbol{h}$, where $\boldsymbol{h} \in N(A)$.*

Proof We know from the proof of Theorem 1.1.6 that (1), (2) and (4) are equivalent. If (2) holds, then multiplying $A\boldsymbol{u} = AA^+\boldsymbol{b}$ on the left by A^* gives (3). On the other hand, multiplying $A^*A\boldsymbol{u} = A^*\boldsymbol{b}$ on the left by $A^{*\dagger}$ gives

$$A\boldsymbol{u} = AA^\dagger\boldsymbol{b},$$

$$\boldsymbol{u} = A^\dagger(AA^\dagger\boldsymbol{b}) + \boldsymbol{h} = A^\dagger\boldsymbol{b} + \boldsymbol{h}, \boldsymbol{h} \in N(A),$$

thus (4) holds. □

Notice that the equations in statement (3) of Theorem 1.1.7 do not involve $A^\dagger$ and are consistent. They are called the normal equations and play an important role in certain areas of statistics.

Exercises 1.1

1. Prove Theorem 1.1.2.
2. Prove that $R(A) = N(A^*)^\perp$.
3. Prove that $\mathrm{rank}(AA^*) = \mathrm{rank}(A) = \mathrm{rank}(A^*A)$.
4. Prove that $R(AA^*) = R(A),\ N(A^*A) = N(A)$.
5. Prove that $R(AB) = R(A) \Leftrightarrow \mathrm{rank}(AB) = \mathrm{rank}(A)$;
$N(AB) = N(B) \Leftrightarrow \mathrm{rank}(AB) = \mathrm{rank}(B)$.
6. Prove Theorem 1.1.3.
7. Show that if $A = FG$ is a full-rank factorization, then
$$A^\dagger = G^\dagger F^\dagger.$$
8. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are column vectors, then
(1) $\boldsymbol{a}^\dagger = (\boldsymbol{a}^*\boldsymbol{a})^\dagger\boldsymbol{a}^*$;
(2) $(\boldsymbol{a}\boldsymbol{b}^*)^\dagger = (\boldsymbol{a}^*\boldsymbol{a})^\dagger(\boldsymbol{b}^*\boldsymbol{b})^\dagger\boldsymbol{b}\boldsymbol{a}^*$.
9. Show that $H^\dagger = H$ if and only if $H^* = H$ and $H^2 = H$.
10. 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.
11. Show that if $A \in \mathbb{C}^{m\times n}$ and $\mathrm{rank}(A) = 1$, then $A^\dagger = \dfrac{1}{\alpha}A^*$, where $\alpha = \mathrm{tr}(A^*A) = \sum\limits_{i,j}|a_{ij}|^2$.
12. Show that if $X \in \mathbb{C}^{m\times n}$,
$$\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{C}^m,\ X_1 = (\boldsymbol{x}_0 \vdots X) \in \mathbb{C}^{m\times(n+1)},$$
$\boldsymbol{b} \in \mathbb{C}^n, \beta_0 \in \mathbb{C}, \boldsymbol{b}_1 = \begin{pmatrix} \beta_0 \\ \boldsymbol{b} \end{pmatrix} \in \mathbb{C}^{n+1}$, then $\boldsymbol{b}_1$ is a least-squares solution of $X_1\boldsymbol{b}_1 = \boldsymbol{y}$ if and only if $\beta_0 = \dfrac{1}{m}\boldsymbol{x}_0^*(\boldsymbol{y} - X\boldsymbol{b})$ and $\boldsymbol{b}$ is a least-squares

solution of

$$\left(I - \frac{1}{m}\boldsymbol{x}_0\boldsymbol{x}_0^*\right) X\boldsymbol{b} = \left(I - \frac{1}{m}\boldsymbol{x}_0\boldsymbol{x}_0^*\right)\boldsymbol{y}.$$

1.2 $\{i, j, k\}$inverses

We discussed the relations between the minimum-norm least-squares solution of an inconsistent system of linear equations (1.1.4) and the M-P inverse in Section 1.1. The relations between the solutions of other linear equations and the matrix equation, and the $\{i, j, k\}$ inverses are given in this section.

1.2.1 {1}inverse and the solution of a consistent system of linear equations

If $A \in \mathbb{C}_n^{n\times n}$, then one of the characteristics of A^{-1} is that for every $\boldsymbol{b}$, $A^{-1}\boldsymbol{b}$ is the solution of $A\boldsymbol{x} = \boldsymbol{b}$. One might ask for $A \in \mathbb{C}^{m\times n}$, what are the characteristics of a matrix $X \in \mathbb{C}^{n\times m}$ such that $X\boldsymbol{b}$ is a solution of the consistent system of linear equations (1.1.3)?

If $AX\boldsymbol{b} = \boldsymbol{b}$ is true for every $\boldsymbol{b} \in R(A)$, it is clear that

$$AXA = A,$$

i.e., the Penrose condition (1) holds. Conversely, suppose X satisfies $AXA = A$. For every $\boldsymbol{b} \in R(A)$ there exists an $\boldsymbol{x}_{\mathbf{b}} \in \mathbb{C}^n$ such that $A\boldsymbol{x}_{\mathbf{b}} = \boldsymbol{b}$. Therefore $AX\boldsymbol{b} = AXA\boldsymbol{x}_{\mathbf{b}} = A\boldsymbol{x}_{\mathbf{b}} = \boldsymbol{b}$ for every $\boldsymbol{b} \in R(A)$. The next theorem is a formal statement of the above observations.

Theorem 1.2.1 *For $A \in \mathbb{C}^{m\times n}$, $X \in \mathbb{C}^{n\times m}$ has the property that $X\boldsymbol{b}$ is a solution of $A\boldsymbol{x} = \boldsymbol{b}$ for every $\boldsymbol{b} \in \mathbb{C}^m$ for which $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if*

$$AXA = A. \tag{1.2.1}$$

Definition 1.2.1 *A matrix X satisfying the Penrose condition (1) $AXA = A$ is called the equation solving generalized inverse for $AXA = A$ or $\{1\}$ inverse of A, and is denoted by $X = A^{(1)}$ or $X \in A\{1\}$, where $A\{1\}$ denotes the set of all $\{1\}$ inverses of A.*

1.2.2 {1, 4} inverse and the minimum-norm solution of a consistent system of linear equations

Suppose we seek $X \in \mathbb{C}^{n\times m}$ such that, in addition to being an equation solving inverse for consistent linear equations (1.1.3), we also require that for each $\boldsymbol{b} \in R(A)$, $\|X\boldsymbol{b}\| < \|\boldsymbol{z}\|$ for all $\boldsymbol{z} \neq X\boldsymbol{b}$ and $\boldsymbol{z} \in \{\boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{b}\}$. That is, for each $\boldsymbol{b} \in R(A)$ we want $X\boldsymbol{b}$ to be the solution of minimal-norm.

If $\boldsymbol{b} \in R(A)$, $AA^{(1,4)}\boldsymbol{b} = \boldsymbol{b}$ and the solutions and the least-squares solutions of the consistent system of linear equations (1.1.3) coincide. Therefore