Jin Zhang

Linear Statistical Models

（线性统计模型）

SCIENCE PRESS
Beijing

Linear Statistical Models

（线性统计模型）

Jin Zhang

SCIENCE PRESS
Beijing

Responsible Editor: Yulong Hao

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

Printed in Chengdu

ISBN 978-7-03-040266-0

Dedicated to my wife Jikun Yi
and my daughter Yili Zhang

In the memory of my father Junmo Zhang
and my mother Zhenying Zhu

Preface

This book, *Linear Statistical Models*, is designed as a textbook or for a one-semester course for graduate or senior undergraduate students. It is written primarily for students majoring in statistics or related fields, but can be served as a reference book for researchers to better understand linear models in statistics.

Although there are many excellent English textbooks on this subject, most of them contain lengthy explanations and examples, which are difficult and time-consuming for non-native English readers. My teaching experience in China and Canada has inspired me to write a concise textbook with simple language, reducing the language barrier for students and instructors from non-English-speaking countries. The main objective of this book is to introduce the theory of linear statistical models in a clear but rigorous format. Hopefully, students, instructors, researchers, and practitioners will find this text more comfortable than most other textbooks.

This book grew from my lecture notes for teaching linear statistical models at Yunnan University (China) and University of Manitoba (Canada). The contents and structure of the book are mainly taken from the textbook "*A First Course in Linear Model Theory*" (Chapman & Hall/CRC, 2002) by N. Ravishanker and D. K. Dey, with reference to other standard textbooks, such as "*A First Course in the Theory of Linear Statistical Models*" (2nd ed. McGraw-Hill, 1998) by R. H. Myers and J. S. Milton, "*Theory and Application of the Linear Statistical Inference and its Applications*" (2nd ed. John Wiley & Sons, 1973) by C. R. Rao, "*Theory and Application of the Linear Model*" (Duxbury Press, 1976) by F. A. Graybill, and "*Linear Models*" (John Wiley & Sons, 1972) by S. R. Searle.

The mathematical prerequisites for this book are multivariate calculus and matrix algebra, where the later plays a fundamental role in linear statistical models. Statistical prerequisites include statistical theory, multivariate regression, and analysis of variance.

This book uses seven chapters to introduce and develop the essential theories and methodologies of linear models in statistics, where important concepts and terminologies are italicized and indexed. At the end of each chapter, there are many exercises for readers, some of which are supplementary materials of the textbook.

Chapter 1 introduces generalized inverse matrices and solutions to linear system of equations, which are necessary for developing the general theory of linear models. Chapter 2 describes the general linear model and related topics, including the least squares method, estimable functions, and estimation subject to linear restrictions. Chapter 3 discusses multivariate normal and related distributions, especially the distributions of quadratic forms as theoretic foundations of statistical inference for the general linear model in Chapter 4, where linear hypothesis tests and confidence intervals are developed. Chapter 5 is devoted to linear regression models, presenting diagnostic tools for model assumptions, criteria for model selection, multicollinearity, and some related topics. Then Chapter 6 uses simple examples to briefly illustrate fixed-effects, random-effects and mixed-effects models. Finally, the generalized linear model is introduced in Chapter 7, with emphasis on discussing its components, link structures, parameter estimation, inference and diagnostics.

In addition, this book provides the readers with Table of Common Statistical Distributions in Appendix A, which includes the commonly used discrete distributions, continuous distributions and multivariate distributions. For each listed distribution, the table gives the detailed information about its pdf/pmf, moments, moment generating function, and important notes on associated distributions.

Like many statistical textbooks, the most commonly used statistical tables for the standard normal, t, χ^2 and F distributions are attached in Appendix B, where the table values are computed by using statistical software **R**, available on its official web site

http://www.r-project.org

In writing this book, I received great contribution from many of my students. I take this opportunity to thank those graduate students who took Liner Statistical Models from me and helped me in typewriting and proofreading the manuscript. Among them are Jie Li, Xiaojie Yang, Tianxia Ai, Hua Li, Yunqi Zhang, Xiaozhun Zhuang and Menglin Li.

I would like to sincerely thank Nalini Ravishanker and Dipak K. Dey, for writing an excellent textbook, from which I greatly benefited. I am grateful for help from my teachers, colleagues, friends and students, especially Xueren Wang and Niansheng Tang, Yunnan University; Yuehua Wu, York University; Jianxin Pan, University of Manchester; Xuming He and Peter Song, University of Michigan (Ann Arbor); Michael Stephens and Richard Lockhart, Simon Fraser University; James Fu, Liqun Wang and Xikui Wang, University of Manitoba; Gemai Chen, University of Calgary; Jiahua Chen, University of British Columbia; Keming Yu, Brunel University.

In addition, I would like to acknowledge the financial support of Yunnan University and the Natural Science Foundation of China (NSFC) for publishing this book.

Last, but not least, I would like to sincerely thank my wife Jikun Yi and my daughter Yili Zhang from the bottom of my heart for their patience, understanding, encouragement and steadfast support.

Jin Zhang
School of Mathematics and Statistics
Yunnan University
Kunming, China

Contents

Chapter 1

Generalized Inverse Matrices and Related Topics

As a powerful mathematical tool, linear and matrix algebra plays an important role in the foundation of linear statistical models. Thus, this book requires basic knowledge of linear and matrix algebra, which is essential for understanding and developing the theory of linear statistical models. The elements of matrix algebra useful for statistics can be found in many textbooks, such as Rao (1973), Searle (1982), Graybill (1976, 1983), Jørgensen (1993), Rao and Toutenberg (1995), Harville (1997), Myers and Milton (1998), Christensen (2002), and Ravishanker and Dey (2002).

In this chapter, we briefly introduce the essentials of generalized inverse matrices, solutions to systems of linear equations, and some related topics, which are necessary in the development of theory for the general linear model. Throughout this book, all numbers, vectors and matrices are assumed to be real, unless stated otherwise.

1.1 Generalized Inverse Matrices

The notion of generalized inverse matrices was originally established in the theory of linear equations. The generalized inverse for matrices has now become a very important mathematical tool in the theory of linear statistical models (e.g., Rao and Mitra, 1971; Pringle and Ragner, 1971; Searle, 1971, 1982; Rao, 1973; Rao and Toutenberg, 1995; Graybill, 1976; Jørgensen, 1993; Harville, 1997; Myers and mitton, 1998; Christensen, 2002; Ravishanker and Dey, 2002), making it easy to understand certain aspects of analysis procedures associated with linear models.

A *generalized inverse* or *g-inverse* of an $m \times n$ matrix $\boldsymbol{A}$ is defined as any $n \times m$

matrix $\boldsymbol{G}$, denoted by $\boldsymbol{A}^-$, satisfying the equation

$$\boldsymbol{AGA} = \boldsymbol{A} \quad \text{or} \quad \boldsymbol{AA}^-\boldsymbol{A} = \boldsymbol{A}.$$

The "generalized inverse" for matrices is sometimes called the "conditional inverse" or "pseudo inverse" in the literature. However, the term of "generalized inverse" or "g-inverse" will be used throughout this book, as did in many other related books.

Unlike the regular inverse, the generalized inverse of matrix $\boldsymbol{A}$ always exists, but it is not unique unless $\boldsymbol{A}$ is a square nonsingular matrix (see Example 1.1.1), according to the following theorem.

Theorem 1.1.1. *For any matrix $\boldsymbol{A}$, its g-inverse $\boldsymbol{A}^-$ always exists and is uniquely to be $\boldsymbol{A}^{-1}$ if and only if $\boldsymbol{A}$ is invertible.*

Proof. Let $\boldsymbol{A}$ be a matrix of rank r. Then there exist invertible (nonsingular) matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, such that

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \text{ or } \boldsymbol{A} = \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{Q}^{-1},$$

where $\boldsymbol{I}_r$ is an identity matrix of order r. It follows that

$$\begin{aligned} \boldsymbol{AGA} = \boldsymbol{A} \quad &\Longleftrightarrow \quad \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{Q}^{-1}\boldsymbol{G}\boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &\Longleftrightarrow \quad \boldsymbol{Q}^{-1}\boldsymbol{G}\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \\ &\Longleftrightarrow \quad \boldsymbol{G} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \boldsymbol{P}. \end{aligned}$$

Hence, a g-inverse $\boldsymbol{G}$ or $\boldsymbol{A}^-$ always exist, and $\boldsymbol{A}^- = \boldsymbol{A}^{-1}$ if and only if $\boldsymbol{A}$ is invertible.

□

Actually, the proof of Theorem 1.1.1 provides us an algorithm to compute any g-inverse $\boldsymbol{A}^-$. Below is a modified algorithm for computing a simple g-inverse of matrix $\boldsymbol{A}$, where $r(\boldsymbol{A})$ denotes the rank of $\boldsymbol{A}$.

Algorithm to compute $\boldsymbol{A}^-$:

1. Find a nonsingular submatrix $\boldsymbol{M}$ of order $r(\boldsymbol{A})$.
2. Transpose $\boldsymbol{A}$ to get $\boldsymbol{A}'$.
3. Replace $\boldsymbol{M}'$ by $\boldsymbol{M}^{-1}$ and other elements of $\boldsymbol{A}'$ by 0 to obtain $\boldsymbol{A}^-$.

Proof. Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be elementary permutation matrices such that

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \boldsymbol{E}.$$

Since $\boldsymbol{P}$ and $\boldsymbol{Q}$ are orthogonal matrices ($\boldsymbol{P}^{-1} = \boldsymbol{P}'$ and $\boldsymbol{Q}^{-1} = \boldsymbol{Q}'$),

$$\boldsymbol{A}^- = (\boldsymbol{P}^{-1}\boldsymbol{E}\boldsymbol{Q}^{-1})^- = \boldsymbol{Q}\boldsymbol{E}^-\boldsymbol{P} = [\boldsymbol{P}^{-1}(\boldsymbol{E}^-)'\boldsymbol{Q}^{-1}]'.$$

It suffices to show that one choice of $\boldsymbol{E}^-$ is $\begin{pmatrix} \boldsymbol{M}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$. In fact,

$$\begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}\begin{pmatrix} \boldsymbol{M}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B} \end{pmatrix}$$

where $\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B} = \boldsymbol{D}$ can proved as follows. Note that

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C}\boldsymbol{M}^{-1} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{M}^{-1}\boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} - \boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B} \end{pmatrix}$$

and

$$r(\boldsymbol{A}) = r(\begin{pmatrix} \boldsymbol{M} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}) = r(\begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} - \boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B} \end{pmatrix})$$

so that

$$r(\boldsymbol{M}) = r(\boldsymbol{M}) + r(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B})$$

or

$$r(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{B}) = 0.$$

□

Example 1.1.1. To find a g-inverse of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 6 & 2 & 6 & 10 \end{pmatrix},$$

we can choose $\boldsymbol{M} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$ or $\begin{pmatrix} 2 & 0 \\ 5 & 15 \end{pmatrix}$ as a nonsingular submatrix of order $2 = r(\boldsymbol{A})$. It follows from the above algorithm that a g-inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^- = \begin{pmatrix} \boldsymbol{M}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{4}{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

or

$$\boldsymbol{A}^- = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{M}^{-1} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & \frac{1}{15} & 0 \end{pmatrix}.$$

□

Example 1.1.2. For a symmetric matrix

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

its g-inverse has the form (see the proof of Theorem 1.1.1)

$$\boldsymbol{A}^- = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix},$$

which can be symmetric or not.

□

Remark: The g-inverse of a symmetric matrix $\boldsymbol{A}$ may not necessarily be symmetric, but it has a symmetric g-inverse $\frac{1}{2}(\boldsymbol{A}^- + \boldsymbol{A}^{-\prime})$ no matter $\boldsymbol{A}^-$ is symmetric or not.

In matrix algebra, a square matrix $\boldsymbol{A}$ is said to be *idempotent* if $\boldsymbol{A}^2 = \boldsymbol{A}$, and $\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$ is called the *projection matrix*, which plays an important role in the linear model theory. The main properties of these matrices are given in the next four theorems, where $tr(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$.

Theorem 1.1.2. *If $\boldsymbol{A}$ is an $n \times n$ idempotent matrix, then*

1. *The eigenvalues of $\boldsymbol{A}$ are 1 or 0.*

2. *$\boldsymbol{I}_n - \boldsymbol{A}$ is idempotent.*

3. *$r(\boldsymbol{A}) = tr(\boldsymbol{A})$; $r(\boldsymbol{I}_n - \boldsymbol{A}) = n - r(\boldsymbol{A})$.*

Proof. Let λ be an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x}$ be a corresponding eigenvector ($\boldsymbol{x} \neq \boldsymbol{0}$).

1. $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \Longrightarrow \boldsymbol{A}^k\boldsymbol{x} = \lambda^k\boldsymbol{x}$ $(k = 1, 2, \cdots)$. For polynomial $f(t) = t^2 - t$, $f(\boldsymbol{A}) = \boldsymbol{0} \Longrightarrow f(\lambda)\boldsymbol{x} = f(\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \Longrightarrow f(\lambda) = 0$. That is, $\lambda = 1$ or 0.

2. $(\boldsymbol{I}_n - \boldsymbol{A})^2 = \boldsymbol{I}_n + \boldsymbol{A}^2 - 2\boldsymbol{A} = \boldsymbol{I}_n - \boldsymbol{A}$.

3. Let $\boldsymbol{A} = \boldsymbol{P}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{Q} = (\boldsymbol{P}_1, \boldsymbol{P}_2)\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \end{pmatrix} = \boldsymbol{P}_1\boldsymbol{Q}_1$. Then,

$$\begin{aligned} \boldsymbol{A} = \boldsymbol{A}^2 \quad &\Longrightarrow \quad \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{Q}\boldsymbol{P}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_1\boldsymbol{P}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &\Longrightarrow \quad \boldsymbol{Q}_1\boldsymbol{P}_1 = \boldsymbol{I}_r. \end{aligned}$$

It follows that

$$tr(\boldsymbol{A}) = tr(\boldsymbol{P}_1\boldsymbol{Q}_1) = tr(\boldsymbol{Q}_1\boldsymbol{P}_1) = tr(\boldsymbol{I}_r) = r = r(\boldsymbol{A})$$

and

$$r(\boldsymbol{I}_n - \boldsymbol{A}) = tr(\boldsymbol{I}_n - \boldsymbol{A}) = n - tr(\boldsymbol{A}) = n - r(\boldsymbol{A}).$$

□

Theorem 1.1.3. *Let $\boldsymbol{A}$ be an $m \times n$ matrix of rank r. Then*

1. *$\boldsymbol{A}^-\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^-$ are idempotent of rank r.*

2. *$\boldsymbol{I}_n - \boldsymbol{A}^-\boldsymbol{A}$ and $\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^-$ are idempotent of ranks $n - r$ and $m - r$.*

Proof.

It is straightforward to show that A^-A and AA^- are idempotent. The rest results follow from Theorem 1.1.2 and

$$r(A) = r(AA^-A) \leqslant r(A^-A) \leqslant r(A).$$

□

Theorem 1.1.4. *Let A be an $m \times n$ matrix. Then*

1. $A(A'A)^-A'A = A$ *and* $A'A(A'A)^-A' = A'$.
2. $A(A'A)^-A'$ *is unique, symmetric and idempotent of rank* $r(A)$.
3. $I_m - A(A'A)^-A'$ *is unique, symmetric and idempotent of rank* $m - r(A)$.

Proof.

1. Let $M = (A'A)^-A'A - I$. Then

$$A'AM = 0 \Longrightarrow (AM)'(AM) = 0 \Longrightarrow AM = 0.$$

That is,

$$AM = A(A'A)^-A'A - A = 0 \quad \text{or} \quad A(A'A)^-A'A = A.$$

2. Let $M = (A'A)_1^-A' - (A'A)_2^-A'$ where $(A'A)_1^-$ and $(A'A)_2^-$ denote any two g-inverses of $A'A$. Then

$$A'AM = 0 \Longrightarrow (AM)'(AM) = 0 \Longrightarrow AM = 0.$$

That is,

$$A(A'A)_1^-A' = A(A'A)_2^-A'.$$

Hence, $A(A'A)^-A'$ is unique, and it is symmetric since we can choose a symmetric g-inverse $(A'A)^-$ (refer to the Remark behind Example 1.1.2).

Obviously, $A(A'A)^-A'$ is idempotent, and its rank is $r(A)$ due to

$$r(A) = r(A(A'A)^-A'A) \leqslant r(A(A'A)^-A') \leqslant r(A).$$

The rest results follow immediately.

□

Theorem 1.1.5. *Let A be an $m \times n$ matrix. Then*

1. *The projection matrix $P = A(A'A)^-A'$ represents the orthogonal projection from R^m onto the column space*

$$\mathcal{C}(A) = \{Ay \mid y \in R^n\}.$$

2. *The matrix $I - P$ represents the orthogonal projection from R^m onto the null space*

$$\mathcal{N}(A') = \{z \mid A'z = 0,\ z \in R^m\},$$

where $\mathcal{N}(\boldsymbol{A}') = \mathcal{C}(\boldsymbol{A})^{\perp}$*, the orthogonal compliment of* $\mathcal{C}(\boldsymbol{A})$.

Proof. For any $\boldsymbol{x} \in R^m$, we have

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2,$$

where $\boldsymbol{x}_1 \in \mathcal{C}(\boldsymbol{A}),\ \boldsymbol{x}_2 \in \mathcal{N}(\boldsymbol{A}'),\ \boldsymbol{x}_1 \perp \boldsymbol{x}_2$ and $\mathcal{C}(\boldsymbol{A}) \perp \mathcal{N}(\boldsymbol{A}')$.

□

1.2 Solutions to Linear Equations

In this section, we briefly discuss the system of linear equations and its solutions. A convenient way for discussing the solutions of linear equations is to employ the generalized inverse for matrices, which plays a fundamental role in the development of linear model theory.

A linear system of m equations in n unknown variables $x_1,\ x_2,\ \ldots,\ x_n$ can be written as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

or

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where $\boldsymbol{A}$ is an $m \times n$ coefficient matrix, $\boldsymbol{x}$ is the vector of unknown variables, and $\boldsymbol{b}$ is the right side of the system.

A linear system (of equations) $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is said to be *consistent* if it has one or more solutions. Otherwise, if no solution exists, the system is *inconsistent.* Let $(\boldsymbol{A}, \boldsymbol{b})$ denote the *augmented matrix* of the system. Then the following is a well-known theorem in linear algebra, in regard to the system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Theorem 1.2.1. *A linear system of equations* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *is consistent if and only if* $r(\boldsymbol{A}, \boldsymbol{b}) = r(\boldsymbol{A})$.

The solution of a consistent linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be expressed in terms of the generalized inverse of coefficient matrix $\boldsymbol{A}$, according to the next two theorems.

Theorem 1.2.2. *Let* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *be a consistent linear system. Then any solution to the system is given by*

$$\boldsymbol{x} = \boldsymbol{A}_0^- \boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{z},$$

where $\boldsymbol{A}_0^-$ *is a specific g-inverse of* $\boldsymbol{A}$*, and* $\boldsymbol{z}$ *is an arbitrary vector.*

Proof. Let $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$. If $\boldsymbol{x} = \boldsymbol{A}_0^- \boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{z}$, then

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{A}_0^- \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{A}\boldsymbol{A}_0^- \boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}.$$

Conversely, if $\boldsymbol{Ax} = \boldsymbol{b}$, then

$$\boldsymbol{x} = \boldsymbol{A}_0^- \boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{x}.$$

□

Note that in Theorem 1.2.2, $\boldsymbol{A}_0^- \boldsymbol{b}$ is a specific solution to the *nonhomogeneous system* $\boldsymbol{Ax} = \boldsymbol{b}$, while $(\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{z}$ is an arbitrary solution to the *homogeneous system* $\boldsymbol{Ax} = \boldsymbol{0}$.

Theorem 1.2.3. *Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a consistent linear system with $\boldsymbol{b} \neq \boldsymbol{0}$. Then $\boldsymbol{x}$ is a solution to the system if and only if $\boldsymbol{x} = \boldsymbol{A}^- \boldsymbol{b}$.*

Proof. By Theorem 1.2.2, $\boldsymbol{x}$ is a solution to the system if and only if

$$\boldsymbol{x} = \boldsymbol{A}_0^- \boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{z} = \boldsymbol{Gb},$$

where $\boldsymbol{G} = \boldsymbol{A}_0^- + (\boldsymbol{I} - \boldsymbol{A}_0^- \boldsymbol{A})\boldsymbol{z}\boldsymbol{b}'/\|\boldsymbol{b}\|^2$ is a g-inverse of $\boldsymbol{A}$.

□

1.3 Exercises

1.1. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ be respectively $n \times n$, $n \times m$, $m \times m$ and $m \times n$ matrices. Verify that

(a)

$$(\boldsymbol{A} - \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} - \boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{D}\boldsymbol{A}^{-1},$$

provided all involved inverses exit.

(b)

$$(\boldsymbol{A} - \boldsymbol{a}\boldsymbol{b}')^{-1} = \boldsymbol{A}^{-1} + \frac{\boldsymbol{A}^{-1}\boldsymbol{a}\boldsymbol{b}'\boldsymbol{A}^{-1}}{1 - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{a}},$$

where $\boldsymbol{A}$ is invertible, $\boldsymbol{a}$ and $\boldsymbol{b}$ are n-dimensional vectors, and $1 - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{a} \neq 0$.

1.2. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ and $n \times m$ matrices. Show that

(a)

$$|\lambda \boldsymbol{I}_m - \boldsymbol{AB}| = \lambda^{m-n} |\lambda \boldsymbol{I}_n - \boldsymbol{BA}|.$$

(b) The nonzero eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$ are the same.

Hint:

Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be nonsingular matrices such that

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \text{ and } \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{B}_3 & \boldsymbol{B}_4 \end{pmatrix}.$$

The result in part (a) follows from

$$|\lambda \boldsymbol{I}_m - \boldsymbol{AB}| = |\lambda \boldsymbol{I}_m - (\boldsymbol{PAQ})(\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})| = |\lambda \boldsymbol{I}_m - \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}| = \lambda^{m-r} |\lambda \boldsymbol{I}_r - \boldsymbol{B}_1|$$