Statistics for Social and Behavioral Sciences

Haruo Yanai
Kei Takeuchi
Yoshio Takane

Springer

Statistics for Social and Behavioral Sciences

Advisors:

S.E. Fienberg
W.J. van der Linden

For other titles published in this series, go to
http://www.springer.com/series/3463

Haruo Yanai • Kei Takeuchi • Yoshio Takane

Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition

Springer

Haruo Yanai
Department of Statistics
St. Luke's College of Nursing
10-1 Akashi-cho Chuo-ku Tokyo
104-0044 Japan
hyanai@slcn.ac.jp

Kei Takeuchi
2-34-4 Terabun Kamakurashi
Kanagawa-ken
247-0064 Japan
kei.takeuchi@wind.ocn.ne.jp

Yoshio Takane
Department of Psychology
McGill University
1205 Dr. Penfield Avenue
Montreal Québec
H3A 1B1 Canada
takane@psych.mcgill.ca

ISBN 978-1-4419-9886-6 e-ISBN 978-1-4419-9887-3
DOI 10.1007/978-1-4419-9887-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011925655

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

All three authors of the present book have long-standing experience in teaching graduate courses in multivariate analysis (MVA). These experiences have taught us that aside from distribution theory, projections and the singular value decomposition (SVD) are the two most important concepts for understanding the basic mechanism of MVA. The former underlies the least squares (LS) estimation in regression analysis, which is essentially a projection of one subspace onto another, and the latter underlies principal component analysis (PCA), which seeks to find a subspace that captures the largest variability in the original space. Other techniques may be considered some combination of the two.

This book is about projections and SVD. A thorough discussion of generalized inverse (g-inverse) matrices is also given because it is closely related to the former. The book provides systematic and in-depth accounts of these concepts from a unified viewpoint of linear transformations in finite dimensional vector spaces. More specifically, it shows that projection matrices (projectors) and g-inverse matrices can be defined in various ways so that a vector space is decomposed into a direct-sum of (disjoint) subspaces. This book gives analogous decompositions of matrices and discusses their possible applications.

This book consists of six chapters. Chapter 1 overviews the basic linear algebra necessary to read this book. Chapter 2 introduces projection matrices. The projection matrices discussed in this book are general oblique projectors, whereas the more commonly used orthogonal projectors are special cases of these. However, many of the properties that hold for orthogonal projectors also hold for oblique projectors by imposing only modest additional conditions. This is shown in Chapter 3.

Chapter 3 first defines, for an n by m matrix $\boldsymbol{A}$, a linear transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ that maps an element $\boldsymbol{x}$ in the m-dimensional Euclidean space E^m onto an element $\boldsymbol{y}$ in the n-dimensional Euclidean space E^n. Let $\mathrm{Sp}(\boldsymbol{A}) = \{\boldsymbol{y}|\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}$ (the range or column space of $\boldsymbol{A}$) and $\mathrm{Ker}(\boldsymbol{A}) = \{\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$ (the null space of $\boldsymbol{A}$). Then, there exist an infinite number of the subspaces V and W that satisfy

$$E^n = \mathrm{Sp}(\boldsymbol{A}) \oplus W \ \text{ and } \ E^m = V \oplus \mathrm{Ker}(\boldsymbol{A}), \tag{1}$$

where $\oplus$ indicates a direct-sum of two subspaces. Here, the correspondence between V and $\mathrm{Sp}(\boldsymbol{A})$ is one-to-one (the dimensionalities of the two subspaces coincide), and an inverse linear transformation from $\mathrm{Sp}(\boldsymbol{A})$ to V can be uniquely defined. Generalized inverse matrices are simply matrix representations of the inverse transformation with the domain extended to E^n. However, there are infinitely many ways in which the generalization can be made, and thus there are infinitely many corresponding generalized inverses $\boldsymbol{A}^-$ of $\boldsymbol{A}$. Among them, an inverse transformation in which $W = \mathrm{Sp}(\boldsymbol{A})^\perp$ (the ortho-complement subspace of $\mathrm{Sp}(\boldsymbol{A})$) and $V = \mathrm{Ker}(\boldsymbol{A})^\perp = \mathrm{Sp}(\boldsymbol{A}')$ (the ortho-complement subspace of $\mathrm{Ker}(\boldsymbol{A})$), which transforms any vector in W to the zero vector in $\mathrm{Ker}(\boldsymbol{A})$, corresponds to the Moore-Penrose inverse. Chapter 3 also shows a variety of g-inverses that can be formed depending on the choice of V and W, and which portion of $\mathrm{Ker}(\boldsymbol{A})$ vectors in W are mapped into.

Chapter 4 discusses generalized forms of oblique projectors and g-inverse matrices, and gives their explicit representations when V is expressed in terms of matrices.

Chapter 5 decomposes $\mathrm{Sp}(\boldsymbol{A})$ and $\mathrm{Sp}(\boldsymbol{A}') = \mathrm{Ker}(\boldsymbol{A})^\perp$ into sums of mutually orthogonal subspaces, namely

$$\mathrm{Sp}(\boldsymbol{A}) = E_1 \dot{\oplus} E_2 \dot{\oplus} \cdots \dot{\oplus} E_r$$

and

$$\mathrm{Sp}(\boldsymbol{A}') = F_1 \dot{\oplus} F_2 \dot{\oplus} \cdots \dot{\oplus} F_r,$$

where $\dot{\oplus}$ indicates an orthogonal direct-sum. It will be shown that E_j can be mapped into F_j by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and that F_j can be mapped into E_j by $\boldsymbol{x} = \boldsymbol{A}'\boldsymbol{y}$. The singular value decomposition (SVD) is simply the matrix representation of these transformations.

Chapter 6 demonstrates that the concepts given in the preceding chapters play important roles in applied fields such as numerical computation and multivariate analysis.

Some of the topics in this book may already have been treated by existing textbooks in linear algebra, but many others have been developed only recently, and we believe that the book will be useful for many researchers, practitioners, and students in applied mathematics, statistics, engineering, behaviormetrics, and other fields.

This book requires some basic knowledge of linear algebra, a summary of which is provided in Chapter 1. This, together with some determination on the part of the reader, should be sufficient to understand the rest of the book. The book should also serve as a useful reference on projectors, generalized inverses, and SVD.

In writing this book, we have been heavily influenced by Rao and Mitra's (1971) seminal book on generalized inverses. We owe very much to Professor

C. R. Rao for his many outstanding contributions to the theory of g-inverses and projectors. This book is based on the original Japanese version of the book by Yanai and Takeuchi published by Todai-Shuppankai (University of Tokyo Press) in 1983. This new English edition by the three of us expands the original version with new material.

January 2011

Haruo Yanai
Kei Takeuchi
Yoshio Takane

Contents

Chapter 1

Fundamentals of Linear Algebra

In this chapter, we give basic concepts and theorems of linear algebra that are necessary in subsequent chapters.

1.1 Vectors and Matrices

1.1.1 Vectors

Sets of n real numbers $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$, arranged in the following way, are called n-component column vectors:

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \tag{1.1}$$

The real numbers $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$ are called elements or components of $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. These elements arranged horizontally,

$$\boldsymbol{a}' = (a_1, a_2, \cdots, a_n), \quad \boldsymbol{b}' = (b_1, b_2, \cdots, b_n),$$

are called n-component row vectors.

We define the length of the n-component vector $\boldsymbol{a}$ to be

$$||\boldsymbol{a}|| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}. \tag{1.2}$$

H. Yanai et al., *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*, Statistics for Social and Behavioral Sciences, DOI 10.1007/978-1-4419-9887-3_1,

This is also called a norm of vector $\boldsymbol{a}$. We also define an inner product between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ to be

$$(\boldsymbol{a}, \boldsymbol{b}) = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n. \tag{1.3}$$

The inner product has the following properties:

(i) $||\boldsymbol{a}||^2 = (\boldsymbol{a}, \boldsymbol{a})$,

(ii) $||\boldsymbol{a} + \boldsymbol{b}||^2 = ||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2 + 2(\boldsymbol{a}, \boldsymbol{b})$,

(iii) $(a\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a}, a\boldsymbol{b}) = a(\boldsymbol{a}, \boldsymbol{b})$, where a is a scalar,

(iv) $||\boldsymbol{a}||^2 = 0 \iff \boldsymbol{a} = \boldsymbol{0}$, where $\iff$ indicates an equivalence (or "if and only if") relationship.

We define the distance between two vectors by

$$d(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||. \tag{1.4}$$

Clearly, $d(\boldsymbol{a}, \boldsymbol{b}) \geq 0$ and

(i) $d(\boldsymbol{a}, \boldsymbol{b}) = 0 \iff \boldsymbol{a} = \boldsymbol{b}$,

(ii) $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$,

(iii) $d(\boldsymbol{a}, \boldsymbol{b}) + d(\boldsymbol{b}, \boldsymbol{c}) \geq d(\boldsymbol{a}, \boldsymbol{c})$.

The three properties above are called the metric (or distance) axioms.

Theorem 1.1 *The following properties hold:*

$$(\boldsymbol{a}, \boldsymbol{b})^2 \leq ||\boldsymbol{a}||^2 ||\boldsymbol{b}||^2, \tag{1.5}$$

$$||\boldsymbol{a} + \boldsymbol{b}|| \leq ||\boldsymbol{a}|| + ||\boldsymbol{b}||. \tag{1.6}$$

Proof. (1.5): The following inequality holds for any real number t:

$$||\boldsymbol{a} - t\boldsymbol{b}||^2 = ||\boldsymbol{a}||^2 - 2t(\boldsymbol{a}, \boldsymbol{b}) + t^2 ||\boldsymbol{b}||^2 \geq 0.$$

This implies

$$\text{Discriminant} = (\boldsymbol{a}, \boldsymbol{b})^2 - ||\boldsymbol{a}||^2 ||\boldsymbol{b}||^2 \leq 0,$$

which establishes (1.5).

(1.6): $(||\boldsymbol{a}|| + ||\boldsymbol{b}||)^2 - ||\boldsymbol{a} + \boldsymbol{b}||^2 = 2\{||\boldsymbol{a}|| \cdot ||\boldsymbol{b}|| - (\boldsymbol{a}, \boldsymbol{b})\} \geq 0$, which implies (1.6). Q.E.D.

Inequality (1.5) is called the Cauchy-Schwarz inequality, and (1.6) is called the triangular inequality.

For two n-component vectors $\boldsymbol{a}$ ($\neq \mathbf{0}$) and $\boldsymbol{b}$ ($\neq \mathbf{0}$), the angle between them can be defined by the following definition.

Definition 1.1 *For two vectors* $\boldsymbol{a}$ *and* $\boldsymbol{b}$*,* θ *defined by*

$$\cos\theta = \frac{(\boldsymbol{a}, \boldsymbol{b})}{||\boldsymbol{a}|| \cdot ||\boldsymbol{b}||} \tag{1.7}$$

is called the angle between $\boldsymbol{a}$ *and* $\boldsymbol{b}$*.*

1.1.2 Matrices

We call nm real numbers arranged in the following form a matrix:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}. \tag{1.8}$$

Numbers arranged horizontally are called rows of numbers, while those arranged vertically are called columns of numbers. The matrix $\boldsymbol{A}$ may be regarded as consisting of n row vectors or m column vectors and is generally referred to as an n by m matrix (an $n \times m$ matrix). When $n = m$, the matrix $\boldsymbol{A}$ is called a square matrix. A square matrix of order n with unit diagonal elements and zero off-diagonal elements, namely

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

is called an identity matrix.

Define m n-component vectors as

$$\boldsymbol{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, \boldsymbol{a}_m = \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix}.$$

We may represent the m vectors collectively by

$$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m]. \tag{1.9}$$

The element of $\boldsymbol{A}$ in the ith row and jth column, denoted as a_{ij}, is often referred to as the (i, j)th element of $\boldsymbol{A}$. The matrix $\boldsymbol{A}$ is sometimes written as $\boldsymbol{A} = [a_{ij}]$. The matrix obtained by interchanging rows and columns of $\boldsymbol{A}$ is called the transposed matrix of $\boldsymbol{A}$ and denoted as $\boldsymbol{A}'$.

Let $\boldsymbol{A} = [a_{ik}]$ and $\boldsymbol{B} = [b_{kj}]$ be n by m and m by p matrices, respectively. Their product, $\boldsymbol{C} = [c_{ij}]$, denoted as

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}, \tag{1.10}$$

is defined by $c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj}$. The matrix $\boldsymbol{C}$ is of order n by p. Note that

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{O} \Longleftrightarrow \boldsymbol{A} = \boldsymbol{O}, \tag{1.11}$$

where $\boldsymbol{O}$ is a zero matrix consisting of all zero elements.

Note An n-component column vector $\boldsymbol{a}$ is an n by 1 matrix. Its transpose $\boldsymbol{a}'$ is a 1 by n matrix. The inner product between $\boldsymbol{a}$ and $\boldsymbol{b}$ and their norms can be expressed as

$$(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}'\boldsymbol{b}, \;\; ||\boldsymbol{a}||^2 = (\boldsymbol{a}, \boldsymbol{a}) = \boldsymbol{a}'\boldsymbol{a}, \text{ and } ||\boldsymbol{b}||^2 = (\boldsymbol{b}, \boldsymbol{b}) = \boldsymbol{b}'\boldsymbol{b}.$$

Let $\boldsymbol{A} = [a_{ij}]$ be a square matrix of order n. The trace of $\boldsymbol{A}$ is defined as the sum of its diagonal elements. That is,

$$\text{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn}. \tag{1.12}$$

Let c and d be any real numbers, and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same order. Then the following properties hold:

$$\text{tr}(c\boldsymbol{A} + d\boldsymbol{B}) = c\text{tr}(\boldsymbol{A}) + d\text{tr}(\boldsymbol{B}) \tag{1.13}$$

and

$$\text{tr}(\boldsymbol{A}\boldsymbol{B}) = \text{tr}(\boldsymbol{B}\boldsymbol{A}). \tag{1.14}$$

Furthermore, for $\boldsymbol{A}$ $(n \times m)$ defined in (1.9),

$$||\boldsymbol{a}_1||^2 + ||\boldsymbol{a}_2||^2 + \cdots + ||\boldsymbol{a}_n||^2 = \text{tr}(\boldsymbol{A}'\boldsymbol{A}). \tag{1.15}$$

Clearly,

$$\text{tr}(\boldsymbol{A}'\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2. \tag{1.16}$$

Thus,

$$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A}) = 0 \Longleftrightarrow \boldsymbol{A} = \boldsymbol{O}. \tag{1.17}$$

Also, when $\boldsymbol{A}_1'\boldsymbol{A}_1, \boldsymbol{A}_2'\boldsymbol{A}_2, \cdots, \boldsymbol{A}_m'\boldsymbol{A}_m$ are matrices of the same order, we have

$$\mathrm{tr}(\boldsymbol{A}_1'\boldsymbol{A}_1 + \boldsymbol{A}_2'\boldsymbol{A}_2 + \cdots + \boldsymbol{A}_m'\boldsymbol{A}_m) = 0 \Longleftrightarrow \boldsymbol{A}_j = \boldsymbol{O}\ (j = 1, \cdots, m). \tag{1.18}$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be n by m matrices. Then,

$$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2,$$

$$\mathrm{tr}(\boldsymbol{B}'\boldsymbol{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij}^2,$$

and

$$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij},$$

and Theorem 1.1 can be extended as follows.

Corollary 1

$$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{B}) \le \sqrt{\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}'\boldsymbol{B})} \tag{1.19}$$

and

$$\sqrt{\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})'(\boldsymbol{A}+\boldsymbol{B})} \le \sqrt{\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})} + \sqrt{\mathrm{tr}(\boldsymbol{B}'\boldsymbol{B})}. \tag{1.20}$$

Inequality (1.19) is a generalized form of the Cauchy-Schwarz inequality.

The definition of a norm in (1.2) can be generalized as follows. Let $\boldsymbol{M}$ be a nonnegative-definite matrix (refer to the definition of a nonnegative-definite matrix immediately before Theorem 1.12 in Section 1.4) of order n. Then,

$$||\boldsymbol{a}||_M^2 = \boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}. \tag{1.21}$$

Furthermore, if the inner product between $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined by

$$(\boldsymbol{a}, \boldsymbol{b})_M = \boldsymbol{a}'\boldsymbol{M}\boldsymbol{b}, \tag{1.22}$$

the following two corollaries hold.

Corollary 2

$$(\boldsymbol{a}, \boldsymbol{b})_M \leq ||\boldsymbol{a}||_M ||\boldsymbol{b}||_M. \tag{1.23}$$

Corollary 1 can further be generalized as follows.

Corollary 3

$$\text{tr}(\boldsymbol{A}'\boldsymbol{M}\boldsymbol{B}) \leq \sqrt{\text{tr}(\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A})\text{tr}(\boldsymbol{B}'\boldsymbol{M}\boldsymbol{B})} \tag{1.24}$$

and

$$\sqrt{\text{tr}\{(\boldsymbol{A}+\boldsymbol{B})'\boldsymbol{M}(\boldsymbol{A}+\boldsymbol{B})\}} \leq \sqrt{\text{tr}(\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A})} + \sqrt{\text{tr}(\boldsymbol{B}'\boldsymbol{M}\boldsymbol{B})}. \tag{1.25}$$

In addition, (1.15) can be generalized as

$$||\boldsymbol{a}_1||_M^2 + ||\boldsymbol{a}_2||_M^2 + \cdots + ||\boldsymbol{a}_m||_M^2 = \text{tr}(\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A}). \tag{1.26}$$

1.2 Vector Spaces and Subspaces

For m n-component vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$, the sum of these vectors multiplied respectively by constants $\alpha_1, \alpha_2, \cdots, \alpha_m$,

$$\boldsymbol{f} = \alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_m\boldsymbol{a}_m,$$

is called a linear combination of these vectors. The equation above can be expressed as $\boldsymbol{f} = \boldsymbol{A}\boldsymbol{a}$, where $\boldsymbol{A}$ is as defined in (1.9), and $\boldsymbol{a}' = (\alpha_1, \alpha_2, \cdots, \alpha_m)$. Hence, the norm of the linear combination $\boldsymbol{f}$ is expressed as

$$||\boldsymbol{f}||^2 = (\boldsymbol{f}, \boldsymbol{f}) = \boldsymbol{f}'\boldsymbol{f} = (\boldsymbol{A}\boldsymbol{a})'(\boldsymbol{A}\boldsymbol{a}) = \boldsymbol{a}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{a}.$$

The m n-component vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ are said to be linearly dependent if

$$\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_m\boldsymbol{a}_m = \boldsymbol{0} \tag{1.27}$$

holds for some $\alpha_1, \alpha_2, \cdots, \alpha_m$ not all of which are equal to zero. A set of vectors are said to be linearly independent when they are not linearly dependent; that is, when (1.27) holds, it must also hold that $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$.

When $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ are linearly dependent, $\alpha_j \neq 0$ for some j. Let $\alpha_i \neq 0$. From (1.27),

$$\boldsymbol{a}_i = \beta_1\boldsymbol{a}_1 + \cdots + \beta_{i-1}\boldsymbol{a}_{i-1} + \beta_{i+1}\boldsymbol{a}_{i+1} + \beta_m\boldsymbol{a}_m,$$

where $\beta_k = -\alpha_k/\alpha_i$ $(k = 1, \cdots, m; k \neq i)$. Conversely, if the equation above holds, clearly $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ are linearly dependent. That is, a set of vectors are linearly dependent if any one of them can be expressed as a linear combination of the other vectors.

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ be linearly independent, and let

$$W = \left\{ \boldsymbol{d} | \boldsymbol{d} = \sum_{i=1}^{m} \alpha_i \boldsymbol{a}_i \right\},$$

where the α_i's are scalars, denote the set of linear combinations of these vectors. Then W is called a linear subspace of dimensionality m.

Definition 1.2 *Let E^n denote the set of all n-component vectors. Suppose that $W \subset E^n$ (W is a subset of E^n) satisfies the following two conditions:*

(1) *If $\boldsymbol{a} \in W$ and $\boldsymbol{b} \in W$, then $\boldsymbol{a} + \boldsymbol{b} \in W$.*

(2) *If $\boldsymbol{a} \in W$, then $\alpha \boldsymbol{a} \in W$, where α is a scalar.*

Then W is called a linear subspace or simply a subspace of E^n.

When there are r linearly independent vectors in W, while any set of $r + 1$ vectors is linearly dependent, the dimensionality of W is said to be r and is denoted as $\dim(W) = r$.

Let $\dim(W) = r$, and let $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r$ denote a set of r linearly independent vectors in W. These vectors are called basis vectors spanning (generating) the (sub)space W. This is written as

$$W = \text{Sp}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r) = \text{Sp}(\boldsymbol{A}), \tag{1.28}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r]$. The maximum number of linearly independent vectors is called the rank of the matrix $\boldsymbol{A}$ and is denoted as $\text{rank}(\boldsymbol{A})$. The following property holds:

$$\dim(\text{Sp}(\boldsymbol{A})) = \text{rank}(\boldsymbol{A}). \tag{1.29}$$

The following theorem holds.

Theorem 1.2 *Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r$ denote a set of linearly independent vectors in the r-dimensional subspace W. Then any vector in W can be expressed uniquely as a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r$.*

(Proof omitted.)

The theorem above indicates that arbitrary vectors in a linear subspace can be uniquely represented by linear combinations of its basis vectors. In general, a set of basis vectors spanning a subspace are not uniquely determined.

If $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r$ are basis vectors and are mutually orthogonal, they constitute an orthogonal basis. Let $\boldsymbol{b}_j = \boldsymbol{a}_j/||\boldsymbol{a}_j||$. Then, $||\boldsymbol{b}_j|| = 1$ $(j = 1, \cdots, r)$. The normalized orthogonal basis vectors $\boldsymbol{b}_j$ are called an orthonormal basis. The orthonormality of $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r$ can be expressed as

$$(\boldsymbol{b}_i, \boldsymbol{b}_j) = \delta_{ij},$$

where δ_{ij} is called Kronecker's δ, defined by

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

Let $\boldsymbol{x}$ be an arbitrary vector in the subspace V spanned by $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r$, namely

$$\boldsymbol{x} \in V = \text{Sp}(\boldsymbol{B}) = \text{Sp}(\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r) \subset E^n.$$

Then $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{b}_1)\boldsymbol{b}_1 + (\boldsymbol{x}, \boldsymbol{b}_2)\boldsymbol{b}_2 + \cdots + (\boldsymbol{x}, \boldsymbol{b}_r)\boldsymbol{b}_r. \tag{1.30}$$

Since $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r$ are orthonormal, the squared norm of $\boldsymbol{x}$ can be expressed as

$$||\boldsymbol{x}||^2 = (\boldsymbol{x}, \boldsymbol{b}_1)^2 + (\boldsymbol{x}, \boldsymbol{b}_2)^2 + \cdots + (\boldsymbol{x}, \boldsymbol{b}_r)^2. \tag{1.31}$$

The formula above is called Parseval's equality.

Next, we consider relationships between two subspaces. Let $V_A = \text{Sp}(\boldsymbol{A})$ and $V_B = \text{Sp}(\boldsymbol{B})$ denote the subspaces spanned by two sets of vectors collected in the form of matrices, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_p]$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_q]$. The subspace spanned by the set of vectors defined by the sum of vectors in these subspaces is given by

$$V_A + V_B = \{\boldsymbol{a} + \boldsymbol{b} | \boldsymbol{a} \in V_A, \boldsymbol{b} \in V_B\}. \tag{1.32}$$

The resultant subspace is denoted by

$$V_{A+B} = V_A + V_B = \text{Sp}(\boldsymbol{A}, \boldsymbol{B}) \tag{1.33}$$

and is called the sum space of V_A and V_B. The set of vectors common to both V_A and V_B, namely

$$V_{A\cap B} = \{\boldsymbol{x} | \boldsymbol{x} = \boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\beta} \text{ for some } \boldsymbol{\alpha} \text{ and } \boldsymbol{\beta}\}, \tag{1.34}$$

also constitutes a linear subspace. Clearly,

$$V_{A+B} \supset V_A \text{ (or } V_B) \supset V_{A\cap B}. \tag{1.35}$$

The subspace given in (1.34) is called the product space between V_A and V_B and is written as

$$V_{A\cap B} = V_A \cap V_B. \tag{1.36}$$

When $V_A \cap V_B = \{\mathbf{0}\}$ (that is, the product space between V_A and V_B has only a zero vector), V_A and V_B are said to be disjoint. When this is the case, V_{A+B} is written as

$$V_{A+B} = V_A \oplus V_B \tag{1.37}$$

and the sum space V_{A+B} is said to be decomposable into the direct-sum of V_A and V_B.

When the n-dimensional Euclidean space E^n is expressed by the direct-sum of V and W, namely

$$E^n = V \oplus W, \tag{1.38}$$

W is said to be a complementary subspace of V (or V is a complementary subspace of W) and is written as $W = V^c$ (respectively, $V = W^c$). The complementary subspace of $\mathrm{Sp}(A)$ is written as $\mathrm{Sp}(A)^c$. For a given $V = \mathrm{Sp}(A)$, there are infinitely many possible complementary subspaces, $W = \mathrm{Sp}(A)^c$.

Furthermore, when all vectors in V and all vectors in W are orthogonal, $W = V^\perp$ (or $V = W^\perp$) is called the ortho-complement subspace, which is defined by

$$V^\perp = \{\boldsymbol{a}|(\boldsymbol{a}, \boldsymbol{b}) = 0, \forall \boldsymbol{b} \in V\}. \tag{1.39}$$

The n-dimensional Euclidean space E^n expressed as the direct sum of r disjoint subspaces W_j $(j = 1, \cdots, r)$ is written as

$$E^n = W_1 \oplus W_2 \oplus \cdots \oplus W_r. \tag{1.40}$$

In particular, when W_i and W_j $(i \neq j)$ are orthogonal, this is especially written as

$$E^n = W_1 \dot{\oplus} W_2 \dot{\oplus} \cdots \dot{\oplus} W_r, \tag{1.41}$$

where $\dot{\oplus}$ indicates an orthogonal direct-sum.

The following properties hold regarding the dimensionality of subspaces.

Theorem 1.3

$$\dim(V_{A+B}) = \dim(V_A) + \dim(V_B) - \dim(V_{A\cap B}), \tag{1.42}$$

$$\dim(V_A \oplus V_B) = \dim(V_A) + \dim(V_B), \tag{1.43}$$

$$\dim(V^c) = n - \dim(V). \tag{1.44}$$

(Proof omitted.)

Suppose that the n-dimensional Euclidean space E^n can be expressed as the direct-sum of $V = \mathrm{Sp}(\boldsymbol{A})$ and $W = \mathrm{Sp}(\boldsymbol{B})$, and let $\boldsymbol{Ax} + \boldsymbol{By} = \boldsymbol{0}$. Then, $\boldsymbol{Ax} = -\boldsymbol{By} \in \mathrm{Sp}(\boldsymbol{A}) \cap \mathrm{Sp}(\boldsymbol{B}) = \{\boldsymbol{0}\}$, so that $\boldsymbol{Ax} = \boldsymbol{By} = \boldsymbol{0}$. This can be extended as follows.

Theorem 1.4 *The necessary and sufficient condition for the subspaces* $W_1 = \mathrm{Sp}(\boldsymbol{A}_1), W_2 = \mathrm{Sp}(\boldsymbol{A}_2), \cdots, W_r = \mathrm{Sp}(\boldsymbol{A}_r)$ *to be mutually disjoint is*

$$\boldsymbol{A}_1\boldsymbol{a}_1 + \boldsymbol{A}_2\boldsymbol{a}_2 + \cdots + \boldsymbol{A}_r\boldsymbol{a}_r = \boldsymbol{0} \Longrightarrow \boldsymbol{A}_j\boldsymbol{a}_j = \boldsymbol{0} \text{ for all } j = 1, \cdots, r.$$

(Proof omitted.)

Corollary *An arbitrary vector* $\boldsymbol{x} \in W = W_1 \oplus \cdots \oplus W_r$ *can uniquely be expressed as*

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \cdots + \boldsymbol{x}_r,$$

where $\boldsymbol{x}_j \in W_j$ $(j = 1, \cdots, r)$.

Note Theorem 1.4 and its corollary indicate that the decomposition of a particular subspace into the direct-sum of disjoint subspaces is a natural extension of the notion of linear independence among vectors.

The following theorem holds regarding implication relations between subspaces.

Theorem 1.5 *Let* V_1 *and* V_2 *be subspaces such that* $V_1 \subset V_2$, *and let* W *be any subspace in* E^n. *Then,*

$$V_1 + (V_2 \cap W) = (V_1 + W) \cap V_2. \tag{1.45}$$

Proof. Let $\boldsymbol{y} \in V_1 + (V_2 \cap W)$. Then $\boldsymbol{y}$ can be decomposed into $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$, where $\boldsymbol{y}_1 \in V_1$ and $\boldsymbol{y}_2 \in V_2 \cap W$. Since $V_1 \subset V_2$, $\boldsymbol{y}_1 \in V_2$, and since $\boldsymbol{y}_2 \subset V_2$, $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 \in V_2$. Also, $\boldsymbol{y}_1 \in V_1 \subset V_1 + W$, and $\boldsymbol{y}_2 \in W \subset V_1 + W$, which together imply $\boldsymbol{y} \in V_1 + W$. Hence, $\boldsymbol{y} \in (V_1 + W) \cap V_2$. Thus, $V_1 + (V_2 \cap W) \subset$

$(V_1 + W) \cap V_2$. If $\boldsymbol{x} \in (V_1 + W) \cap V_2$, then $\boldsymbol{x} \in V_1 + W$ and $\boldsymbol{x} \in V_2$. Thus, $\boldsymbol{x}$ can be decomposed as $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{y}$, where $\boldsymbol{x}_1 \in V_1$ and $\boldsymbol{y} \in W$. Then $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}_1 \in V_2 \cap W \Longrightarrow \boldsymbol{x} \in V_1 + (V_2 \cap W) \Longrightarrow (V_1 + W) \cap V_2 \subset V_1 + (V_2 \cap W)$, establishing (1.45). Q.E.D.

Corollary (a) *For* $V_1 \subset V_2$*, there exists a subspace* $\tilde{W} \subset V_2$ *such that* $V_2 = V_1 \oplus \tilde{W}$.
(b) *For* $V_1 \subset V_2$,

$$V_2 = V_1 \dot{\oplus} (V_2 \cap V_1^{\perp}). \tag{1.46}$$

Proof. (a): Let W be such that $V_1 \oplus W \supset V_2$, and set $\tilde{W} = V_2 \cap W$ in (1.45).
(b): Set $W = V_1^{\perp}$. Q.E.D.

Note Let $V_1 \subset V_2$, where $V_1 = \mathrm{Sp}(\boldsymbol{A})$. Part (a) in the corollary above indicates that we can choose $\boldsymbol{B}$ such that $W = \mathrm{Sp}(\boldsymbol{B})$ and $V_2 = \mathrm{Sp}(\boldsymbol{A}) \oplus \mathrm{Sp}(\boldsymbol{B})$. Part (b) indicates that we can choose $\mathrm{Sp}(\boldsymbol{A})$ and $\mathrm{Sp}(\boldsymbol{B})$ to be orthogonal.

In addition, the following relationships hold among the subspaces V, W, and K in E^n:

$$V \supset W \Longrightarrow W = V \cap W, \tag{1.47}$$

$$V \supset W \Longrightarrow V + K \supset W + K, \text{ (where } K \in E^n\text{)}, \tag{1.48}$$

$$(V \cap W)^{\perp} = V^{\perp} + W^{\perp}, \ \ V^{\perp} \cap W^{\perp} = (V + W)^{\perp}, \tag{1.49}$$

$$(V + W) \cap K \supseteq (V \cap K) + (W \cap K), \tag{1.50}$$

$$K + (V \cap W) \subseteq (K + V) \cap (K + W). \tag{1.51}$$

Note In (1.50) and (1.51), the distributive law in set theory does not hold. For the conditions for equalities to hold in (1.50) and (1.51), refer to Theorem 2.19.

1.3 Linear Transformations

A function ϕ that relates an m-component vector $\boldsymbol{x}$ to an n-component vector $\boldsymbol{y}$ (that is, $\boldsymbol{y} = \phi(\boldsymbol{x})$) is often called a mapping or transformation. In this book, we mainly use the latter terminology. When ϕ satisfies the following properties for any two n-component vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and for any constant a, it is called a linear transformation:

$$\text{(i) } \phi(a\boldsymbol{x}) = a\phi(\boldsymbol{x}), \quad \text{(ii) } \phi(\boldsymbol{x} + \boldsymbol{y}) = \phi(\boldsymbol{x}) + \phi(\boldsymbol{y}). \tag{1.52}$$

If we combine the two properties above, we obtain

$$\phi(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \cdots + \alpha_m \boldsymbol{x}_m) = \alpha_1 \phi(\boldsymbol{x}_1) + \alpha_2 \phi(\boldsymbol{x}_2) + \cdots + \alpha_m \phi(\boldsymbol{x}_m)$$

for m n-component vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$, and m scalars, $\alpha_1, \alpha_2, \cdots, \alpha_m$.

Theorem 1.6 *A linear transformation ϕ that transforms an m-component vector $\boldsymbol{x}$ into an n-component vector $\boldsymbol{y}$ can be represented by an n by m matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m]$ that consists of m n-component vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$.* (Proof omitted.)

We now consider the dimensionality of the subspace generated by a linear transformation of another subspace. Let $W = \mathrm{Sp}(\boldsymbol{A})$ denote the range of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ when $\boldsymbol{x}$ varies over the entire range of the m-dimensional space E^m. Then, if $\boldsymbol{y} \in W$, $\alpha \boldsymbol{y} = \boldsymbol{A}(\alpha \boldsymbol{x}) \in W$, and if $\boldsymbol{y}_1, \boldsymbol{y}_2 \in W$, $\boldsymbol{y}_1 + \boldsymbol{y}_2 \in W$. Thus, W constitutes a linear subspace of dimensionality $\dim(W) = \mathrm{rank}(\boldsymbol{A})$ spanned by m vectors, $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$.

When the domain of $\boldsymbol{x}$ is V, where $V \subset E^m$ and $V \neq E^m$ (that is, $\boldsymbol{x}$ does not vary over the entire range of E^m), the range of $\boldsymbol{y}$ is a subspace of W defined above. Let

$$W_V = \{\boldsymbol{y} | \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \in V\}. \tag{1.53}$$

Then,

$$\dim(W_V) \leq \min\{\mathrm{rank}(\boldsymbol{A}), \dim(W)\} \leq \dim(\mathrm{Sp}(\boldsymbol{A})). \tag{1.54}$$

Note The W_V above is sometimes written as $W_V = \mathrm{Sp}_V(\boldsymbol{A})$. Let $\boldsymbol{B}$ represent the matrix of basis vectors. Then W_V can also be written as $W_V = \mathrm{Sp}(\boldsymbol{A}\boldsymbol{B})$.

We next consider the set of vectors $\boldsymbol{x}$ that satisfies $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for a given linear transformation $\boldsymbol{A}$. We write this subspace as

$$\mathrm{Ker}(\boldsymbol{A}) = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}. \tag{1.55}$$

Since $\boldsymbol{A}(\alpha \boldsymbol{x}) = \boldsymbol{0}$, we have $\alpha \boldsymbol{x} \in \mathrm{Ker}(\boldsymbol{A})$. Also, if $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{Ker}(\boldsymbol{A})$, we have $\boldsymbol{x} + \boldsymbol{y} \in \mathrm{Ker}(\boldsymbol{A})$ since $\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{0}$. This implies $\mathrm{Ker}(\boldsymbol{A})$ constitutes a subspace of E^m, which represents a set of m-dimensional vectors that are