

Kohei Adachi

Matrix-Based Introduction to Multivariate Data Analysis

Second Edition

 Springer

Matrix-Based Introduction to Multivariate Data Analysis

Kohei Adachi

Matrix-Based Introduction to Multivariate Data Analysis

Second Edition

 Springer

Kohei Adachi
Graduate School of Human Sciences
Osaka University
Suita, Osaka, Japan

ISBN 978-981-15-4102-5 ISBN 978-981-15-4103-2 (eBook)
<https://doi.org/10.1007/978-981-15-4103-2>

1st edition: © Springer Nature Singapore Pte Ltd. 2016, corrected publication 2018

2nd edition: © Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface to the Second Edition

In this second edition, I have added new six chapters (Chaps. 17–22) and three Appendices (A.7–A.9) to the first edition (which spanned Chaps. 1–16 and Appendices A.1–A.6), together with correcting all known misprints and other errors in the first edition. Furthermore, I have made minor modifications to some parts of the first edition, in line with the additional chapters and appendices.

The chapters added in this second edition are as follows:

17. Advanced Matrix Operations
18. Exploratory Factor Analysis (Part 2)
19. Principal Component Analysis versus Factor Analysis
20. Three-way Principal Component Analysis
21. Sparse Regression Analysis
22. Sparse Factor Analysis

which form Part V (*Advance Procedures*) following Parts I–IV.

Chapter 17 serves as a mathematical preparation for the following chapters. In Chap. 17, the Moore–Penrose (MP) inverse in particular is covered in detail, emphasizing its definition through singular value decomposition (SVD). I believe that the MP inverse is of secondary importance among matrix operations, with SVD being of primary importance, as the SVD-based definition of the MP inverse allows us to easily derive its properties and various matrix operations. In this chapter, we also introduce an orthogonal complement matrix, as it is foreseeable that the need for this matrix will increase in multivariate analysis procedures.

Chapter 18 is titled “Exploratory Factor Analysis (Part 2)”, while “(Part 1)” was added to the title of Chap. 12 in the first edition. The contents of Chap. 12 remain unchanged in this second edition, but the exploratory factor analysis (EFA) in Chap. 18 is of a new type, i.e., the EFA procedure formulated as a matrix decomposition problem. This differs from EFA based on the latent variable model in Chap. 12. To emphasize the difference, the former (new) EFA is referred to as matrix decomposition FA (MDFA), while the latter is called latent variable FA

(LVFA) in Chap. 18. Its addition owes to recent developments after the publication of the first edition, as studies of MDFA advanced rapidly. I believe that MDFA is generally superior to LVFA in that the former makes the essence of FA more transparent.

In Chap. 19, answers are given to the question of how solutions from principal component analysis (PCA) and FA differ. No clear answer to this question is found in other books, to the best of my knowledge. The answers in Chap. 19 also are owing to advances in MDFA studies, with the MDFA formulation allowing for straightforward comparisons to be made between FA and PCA.

Three-way principal component analysis (3WPCA) is treated in Chap. 20. 3WPCA refers to a specially modified PCA designed for three-way data sets. The given example is a data array of inputs \times outputs \times boxes, whose elements are the magnitudes of output signals elicited by input signals for multiple black boxes. Three-way data are often encountered in various areas of sciences, and as such 3WPCA is a useful dimension reduction methodology. Its algorithms are very matrix-intensive and suitably treated in this book.

Sparse estimation procedures are introduced in Chaps. 21 and 22. Here, sparse estimation refers to estimating a number of parameters as zeros. Such procedures are popular topics in the field of machine learning. This field can be defined as learning attained by machines (in particular computers) as opposed to humans or living organisms. Statistical analysis procedures are useful methodologies for machine learning. Sparse estimation is also I believe a key property of human learning: our perception performs sparse estimation too in that usually we only cognize useful signals, neglecting useless ones as “zeros”. In this respect, it is very important to enable machines to perform sparse estimation, as a complement to humans. In Chap. 21, sparse regression analysis procedures are described, including Tibshirani’s (1996) procedure called *lasso* which spurred the developments in sparse estimation. Finally, sparse factor analysis (FA) procedures are introduced in Chap. 22.

The Appendices added in this second edition are as follows:

A.7. Scale Invariance of Covariance Structure Analysis

A.8. Probability Densities and Expected Values with EM Algorithm

A.9. EM Algorithm for Factor Analysis.

Though the scale invariance in A.7 had been described with short notes in the first edition, the notes were found too short and insufficient. Thus, the scale invariance is described in more detail in Appendix A.7: Notes 9.3 and 10.2 in the first edition have been expanded and moved to A.7 in this edition. The new Appendix A.9 is necessary for explaining one of the two sparse FA procedures in Chap. 22 and is also useful for deepening the understanding of the confirmatory and exploratory FA treated in Chaps. 10 and 12. The foundations of the algorithm in A.9 are introduced in the preceding new Appendix A.8. Further, this A.8 serves to deepen the understanding of the treatment in Chap. 8.

In the first edition, some parts of the bibliographical notes and exercises were provided to allow readers to extend their understanding beyond the scope covered in that edition. Such parts have become unnecessary in the second edition, as the advanced contents are now described in the additional chapters. Hence, sections of the bibliographical notes and exercises related to the new chapters (Chaps. [17–22](#)) have been deleted or moved to the relevant chapters in the second edition.

Yutaka Hirachi of Springer has encouraged me for publishing this revised version, as well as when I prepared the drafts for the first edition. I am most grateful to him. I am also thankful to the reviewers who read through drafts of this book.

Kyoto, Japan
February 2020

Kohei Adachi

Preface to the First Edition

A set of multivariate data can be expressed as a table, i.e., a matrix, of individuals (rows) by variables (columns), with the variables interrelated. Statistical procedures for analyzing such data sets are generally referred to as multivariate data analysis. The demand for this kind of analysis is increasing in a variety of fields. Each procedure in multivariate data analysis features a special purpose. For example, predicting future performance, classifying individuals, visualizing inter-individual relationships, finding a few factors underlying a number of variables, and examining causal relationships among variables are included in the purposes for the procedures.

The aim of this book is to enable readers who may not be familiar with matrix operations to understand major multivariate data analysis procedures in matrix forms. For that aim, this book begins with explaining fundamental matrix calculations and the matrix expressions of elementary statistics, followed by an introduction to popular multivariate procedures, with chapter-by-chapter advances in the levels of matrix algebra. The organization of this book allows readers without knowledge of matrices to deepen their understanding of multivariate data analysis.

Another feature of this book is its emphasis on the model that underlies each procedure and the objective function that is optimized for fitting the model to data. The author believes that the matrix-based learning of such models and objective functions is the shortest way to comprehend multivariate data analysis. This book is also arranged so that readers can intuitively capture for what purposes multivariate analysis procedures are utilized; plain explanations of the purposes with numerical examples precede mathematical descriptions in almost all chapters.

The preceding paragraph featured three key words: purpose, model, and objective function. The author considers that capturing those three points for each procedure suffices to understand it. This consideration implies that the mechanisms behind how objective functions are optimized must not necessarily be understood. Thus, the mechanisms are only described in appendices and some exercises.

This book is written with the following guidelines in mind:

- (1) Not using mathematics except matrix algebra
- (2) Emphasizing singular value decomposition (SVD)
- (3) Preferring a simultaneous solution to a successive one.

Although the exceptions to (1) are found in Appendix A.6, where differential calculus is used, and in some sections of Part III and Chap. 15, where probabilities are used, those exceptional parts only occupy a limited number of pages; the majority of the book is matrix-intensive. Matrix algebra is also exclusively used for formulating the optimization of objective functions in Appendix A.4. For matrix-intensive formulations, ten Berge's (1983, 1993) theorem is considered to be the best starting fact, as found in Appendix A.4.1.

Guideline (2) is due to the fact that SVD can be defined for any matrix, and a number of important properties of matrices are easily derived from SVD. In the former point, SVD is more general than eigenvalue decomposition (EVD), which is only defined for symmetric matrices. Thus, EVD is only mentioned in Sect. 6.2. Further, SVD takes on an important role in optimizing trace and least squares functions of matrices: The optimization problems are formulated with the combination of SVD and ten Berge's (1983, 1993) theorem, as found in Appendix A.4.2 and Appendix A.4.3.

Guideline (3) is particularly concerned with principal component analysis (PCA), which can be formulated as minimizing $\|\mathbf{X} - \mathbf{FA}'\|^2$ over PC score matrix \mathbf{F} and loading matrix \mathbf{A} for a data matrix \mathbf{X} . In some of the literature, PCA is described as obtaining the first component, the second, and the remaining components in turn (i.e., per column of \mathbf{F} and \mathbf{A}). This can be called a successive solution. On the other hand, PCA can be described as obtaining \mathbf{F} and \mathbf{A} matrix-wise, which can be called a simultaneous solution. This is preferred in this book, as the above formulation is actually made matrix-wise, and the simultaneous solution facilitates understanding PCA as a reduced rank approximation of \mathbf{X} .

This book is appropriate for undergraduate students who have already learned introductory statistics, as the author has used preliminary versions of the book in a course for such students. It is also useful for graduate students and researchers who are not familiar with the matrix-intensive formulations of multivariate data analysis.

I owe this book to the people who can be called the "matricians" in statistics, more exactly, the ones taking matrix-intensive approaches for formulating and developing data analysis procedures. Particularly, I have been influenced by the Dutch psychometricians, as found above, in that I emphasize the theorem by Jos M. F. ten Berge (Professor Emeritus, University of Groningen). Yutaka Hirachi of Springer has been encouraging me since I first considered writing this book. I am

most grateful to him. I am also thankful to the reviewers who read through drafts of this book. Finally, I must show my gratitude to Yoshitaka Shishikura of the publisher Nakanishiya Shuppan, as he readily agreed to the use of the numerical examples in this book, which I had originally used in that publisher's book.

Kyoto, Japan
May 2016

Kohei Adachi

Contents

Part I Elementary Statistics with Matrices

1	Elementary Matrix Operations	3
1.1	Matrices	3
1.2	Vectors	5
1.3	Sum of Matrices and Their Multiplication by Scalars	6
1.4	Inner Product and Norms of Vectors	7
1.5	Product of Matrices	8
1.6	Two Properties of Matrix Products	11
1.7	Trace Operator and Matrix Norm	11
1.8	Vectors and Matrices Filled with Ones or Zeros	13
1.9	Special Square Matrices	14
1.10	Bibliographical Notes	16
2	Intra-variable Statistics	17
2.1	Data Matrices	17
2.2	Distributions	19
2.3	Averages	20
2.4	Centered Scores	20
2.5	Variances and Standard Deviations	23
2.6	Standard Scores	25
2.7	What Centering and Standardization Do for Distributions	26
2.8	Matrix Representation	27
2.9	Bibliographical Notes	28
3	Inter-variable Statistics	31
3.1	Scatter Plots and Correlations	31
3.2	Covariances	32
3.3	Correlation Coefficients	34
3.4	Variable Vectors and Correlations	36
3.5	Covariances and Correlations for Standard Scores	37

3.6	Matrix Expressions of Covariances and Correlations	38
3.7	Unbiased Covariances	39
3.8	Centered Matrices	40
3.9	Ranks of Matrices: Intuitive Introduction	41
3.10	Ranks of Matrices: Mathematical Definition	42
3.11	Bibliographical Notes	44

Part II Least Squares Procedures

4	Regression Analysis	49
4.1	Prediction of a Dependent Variable by Explanatory Variables	49
4.2	Least Squares Method	52
4.3	Predicted and Error Values	54
4.4	Proportion of Explained Variance and Multiple Correlation	56
4.5	Interpretation of Regression Coefficients	58
4.6	Standardization	60
4.7	Geometric Derivation of Regression Coefficients	61
4.8	Bibliographical Notes	62
5	Principal Component Analysis (Part 1)	65
5.1	Reduction of Variables into Components	65
5.2	Singular Value Decomposition	67
5.3	Formulation with a Weight Matrix	70
5.4	Constraints for Components	71
5.5	Interpretation of Loadings	73
5.6	Interpretation of Weights	74
5.7	Percentage of Explained Variance	74
5.8	High-Dimensional Data Analysis	76
5.9	Bibliographical Notes	77
6	Principal Component Analysis (Part 2)	81
6.1	Reformulation with Different Constraints	81
6.2	Maximizing the Sum of Variances	82
6.3	Weighted Composite Scores with Maximum Variance	84
6.4	Projecting Three-Dimensional Vectors onto Two-Dimensional Ones	86
6.5	Visualization of Invisible Distributions	89
6.6	Goodness of Projection	92
6.7	Bibliographical Notes	93
7	Cluster Analysis	95
7.1	Membership Matrices	95
7.2	Example of Clustering Results	96

7.3	Formulation	98
7.4	Iterative Algorithm	100
7.5	Obtaining Cluster Features	102
7.6	Obtaining Memberships	103
7.7	Brief Description of Algorithm	104
7.8	Bibliographical Notes	105

Part III Maximum Likelihood Procedures

8	Maximum Likelihood and Multivariate Normal Distribution	111
8.1	Model, Parameter, Objective Function, and Optimization	111
8.2	Maximum Likelihood Method	112
8.3	Probability Density Function	115
8.4	Multivariate Normal Distribution	116
8.5	Maximum Likelihood Method for Normal Variables	119
8.6	Maximum Likelihood Estimates of Means and Covariances	121
8.7	Model Selection	123
8.8	Assessment of Between-Group Heterogeneity	124
8.9	Bibliographical Notes	127
9	Path Analysis	131
9.1	From Multiple Regression Analysis to Path Analysis	131
9.2	Matrix Expression	136
9.3	Distributional Assumptions	137
9.4	Likelihood for Covariance Structure Analysis	138
9.5	Maximum Likelihood Estimation	139
9.6	Estimated Covariance Structure	140
9.7	Unstandardized and Standardized Solutions	142
9.8	Other and Extreme Models	142
9.9	Model Selection	144
9.10	Bibliographical Notes	145
10	Confirmatory Factor Analysis	149
10.1	Example of Confirmatory Factor Analysis Model	150
10.2	Matrix Expression	150
10.3	Distributional Assumptions for Common Factors	156
10.4	Distributional Assumptions for Errors	157
10.5	Maximum Likelihood Method	158
10.6	Solutions	159
10.7	Other and Extreme Models	160
10.8	Model Selection	161
10.9	Bibliographical Notes	162

11	Structural Equation Modeling	165
11.1	Causality Among Factors	165
11.2	Observed Variables as Indicator of Factors	166
11.3	SEM Model	167
11.4	Matrix Expression	167
11.5	Distributional Assumptions	172
11.6	Maximum Likelihood Method	173
11.7	Solutions	174
11.8	Model Selection	175
11.9	Bibliographical Notes	176
12	Exploratory Factor Analysis (Part 1)	179
12.1	Example of Exploratory Factor Analysis Model	179
12.2	Matrix Expression	180
12.3	Distributional Assumptions	181
12.4	Maximum Likelihood Method	182
12.5	Indeterminacy of EFA Solutions	182
12.6	Two-Stage Procedure	184
12.7	Interpretation of Loadings	184
12.8	Interpretations of Unique Variances	186
12.9	Selecting the Number of Factors	186
12.10	Difference to Principal Component Analysis	188
12.11	Bibliographical Notes	192
 Part IV Miscellaneous Procedures		
13	Rotation Techniques	197
13.1	Geometric Illustration of Factor Rotation	197
13.2	Oblique and Orthogonal Rotation	200
13.3	Rotation to Simple Structure	200
13.4	Varimax Rotation	203
13.5	Geomin Rotation	204
13.6	Orthogonal Procrustes Rotation	206
13.7	Bibliographical Notes	207
14	Canonical Correlation and Multiple Correspondence Analyses	211
14.1	Block Matrices	211
14.2	Canonical Correlation Analysis	214
14.3	Generalized Canonical Correlation Analysis	217
14.4	Multivariate Categorical Data	221
14.5	Multiple Correspondence Analysis	222
14.6	Homogeneity Assumption	224
14.7	Bibliographical Notes	226

15	Discriminant Analysis	229
15.1	Modification of Multiple Correspondence Analysis	229
15.2	Canonical Discriminant Analysis	231
15.3	Minimum Distance Classification	233
15.4	Maximum Probability Classification	234
15.5	Normal Discrimination for Two Groups	236
15.6	Interpreting Solutions	238
15.7	Generalized Normal Discrimination	241
15.8	Bibliographical Notes	244
16	Multidimensional Scaling	247
16.1	Linking Coordinates to Quasi-distances	247
16.2	Illustration of an MDS Solution	249
16.3	Iterative Algorithm	250
16.4	Matrix Expression for Squared Distances	251
16.5	Inequality for Distances	253
16.6	Majorization Algorithm	255
16.7	Bibliographical Notes	256
Part V Advanced Procedures		
17	Advanced Matrix Operations	261
17.1	Introductory Systems of Linear Equations	261
17.2	Moore–Penrose Inverse and System of Linear Equations	262
17.3	Singular Value Decomposition and the Moore–Penrose Inverse	264
17.4	Least Squares Problem Solved with Moore–Penrose Inverse	266
17.5	Orthogonal Complement Matrix	268
17.6	Kronecker Product	271
17.7	Khatri–Rao Product	272
17.8	Vec Operator	274
17.9	Hadamard Product	275
17.10	Bibliographical Notes	276
18	Exploratory Factor Analysis (Part 2)	279
18.1	Matrix Decomposition Formulation	279
18.2	Comparisons to Latent Variable Formulation	282
18.3	Solution of Loadings and Unique Variances	283
18.4	Iterative Algorithm	284
18.5	Estimation of Covariances Between Variables and Factor Scores	285
18.6	Estimation of Loadings and Unique Variances	288
18.7	Identifiability of the Model Part and Residuals	289

18.8	Factor Scores as Higher Rank Approximations	291
18.9	Bibliographical Notes	293
19	Principal Component Analysis Versus Factor Analysis	297
19.1	Motivational Examples	297
19.2	Comparisons of Models	298
19.3	Solutions and Decomposition of the Sum of Squares	300
19.4	Larger Common Part of Principal Component Analysis	304
19.5	Better Fit of Factor Analysis	305
19.6	Largeness of Unique Variances in Factor Analysis	306
19.7	Inequalities for Latent Variable Factor Analysis	307
19.8	Inequalities After Nonsingular Transformation	308
19.9	Proofs for Inequalities	309
19.10	Bibliographical Notes	310
20	Three-Way Principal Component Analysis	311
20.1	Tucker3 and Parafac Models	311
20.2	Hierarchical Relationships Among PCA and 3WPCA	313
20.3	Parafac Solution	319
20.4	Tucker3 Solution	322
20.5	Unconstrained Parafac Algorithm	324
20.6	Constrained Parafac Algorithm	327
20.7	Tucker3 Algorithm: The Optimal Core Array	329
20.8	Tucker3 Algorithm: Iterative Solution	331
20.9	Three-Way Rotation in Tucker3	333
20.10	Bibliographical Notes	336
21	Sparse Regression Analysis	341
21.1	Illustration of Sparse Solution	341
21.2	Penalized Least Squares Method and Lasso	343
21.3	Coordinate Descent Algorithm for Lasso	344
21.4	Selection of Penalty Weight	348
21.5	L_0 Sparse Regression	351
21.6	Standard Regression in Ordinary and High-Dimensional Cases	353
21.7	High-Dimensional Variable Selection by Sparse Regression	356
21.8	Bibliographical Notes	358
22	Sparse Factor Analysis	361
22.1	From Confirmatory FA to Sparse FA	361
22.2	Formulation of Penalized Sparse LVFA	363
22.3	Algorithm for Penalized Sparse LVFA	364
22.4	M-Step for Penalized Sparse LVFA	365
22.5	Using Penalized Sparse LVFA	369

22.6	Formulation of Cardinality Constrained MDFA	372
22.7	Algorithm for Cardinality Constrained MDFA	374
22.8	Using Cardinality Constrained MDFA	375
22.9	Sparse FA Versus Factor Rotation in Exploratory FA	377
22.10	Bibliographical Notes	379
Appendices.		383
A.1	Geometric Understanding of Matrices and Vectors	383
A.2	Decomposition of Sums of Squares	389
A.3	Singular Value Decomposition	392
A.4	Matrix Computations Using SVD	398
A.5	Normal Maximum Likelihood Estimates	411
A.6	Iterative Algorithms	415
A.7	Scale Invariance of Covariance Structure Analysis	420
A.8	Probability Densities and Expected Values with EM Algorithm . . .	425
A.9	EM Algorithm for Factor Analysis	431
References		443
Index		451

Part I

Elementary Statistics with Matrices

This part begins with introducing elementary matrix operations, followed by explanations of fundamental statistics with their matrix expressions. These initial chapters serve as preparation for learning the multivariate data analysis procedures that are described in Part II and thereafter.

Chapter 1

Elementary Matrix Operations



The mathematics for studying the properties of matrices is called *matrix algebra* or *linear algebra*. This first chapter treats the introductory part of matrix algebra required for learning multivariate data analysis. We begin by explaining what a matrix is, in order to describe elementary matrix operations.

In later chapters, more advanced properties of matrices are described, where necessary, with references to Appendices for more detailed explanations.

1.1 Matrices

Let us note that Table 1.1 is a 6 teams \times 4 items table. When such a table (i.e., a two-way array) is treated as a unit entity and expressed as

$$\mathbf{X} = \begin{bmatrix} 0.617 & 731 & 140 & 3.24 \\ 0.545 & 680 & 139 & 4.13 \\ 0.496 & 621 & 143 & 3.68 \\ 0.493 & 591 & 128 & 4.00 \\ 0.437 & 617 & 186 & 4.80 \\ 0.408 & 615 & 184 & 4.80 \end{bmatrix},$$

this is called a 6 (rows) \times 4 (columns) *matrix*, or a matrix of 6 rows by 4 columns. “*Matrices*” is the plural of “matrix”. Here, a horizontal array and a vertical one are called a *row* and a *column*, respectively. For example, the fifth row of \mathbf{A} is “0.437, 617, 0.260, 4.80”, while the third column is “140, 139, 143, 128, 186, 184”. Further, the cell at which the fifth row and third column intersect is occupied by 186, which is called “the (5,3) *element*”. Rewriting the rows of a matrix as columns (or its columns as rows) is referred to as a *transpose*. The transpose of \mathbf{X} is denoted as \mathbf{X}' :

Table 1.1 Averages of the six-teams in Japanese Central Baseball League 2005

Team	Item			
	Win %	Runs	HR	ERA
Tigers	0.617	731	140	3.24
Dragons	0.545	680	139	4.13
BayStars	0.496	621	143	3.68
Swallows	0.493	591	128	4.00
Giants	0.437	617	186	4.80
Carp	0.408	615	184	4.80

$$\mathbf{X}' = \begin{bmatrix} 0.617 & 0.545 & 0.496 & 0.493 & 0.437 & 0.408 \\ 731 & 680 & 621 & 591 & 617 & 615 \\ 140 & 139 & 143 & 128 & 186 & 184 \\ 3.24 & 4.13 & 3.68 & 4.00 & 4.80 & 4.80 \end{bmatrix}$$

Let us describe a matrix in a generalized setting. The array of a_{ij} ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$) arranged in n rows and m columns, i.e.,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad (1.1)$$

is called an $n \times m$ matrix with a_{ij} its (i, j) element. The transpose of \mathbf{A} is an $m \times n$ matrix

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}. \quad (1.2)$$

The transpose of a transposed matrix is obviously the original matrix, with $(\mathbf{A}')' = \mathbf{A}$.

The expression of matrix \mathbf{A} as the right-hand side in (1.1) takes a large amount of space. For economy of space, the matrix \mathbf{A} in (1.1) is also expressed as

$$\mathbf{A} = (a_{ij}), \quad (1.3)$$

using the general expression a_{ij} for the elements of \mathbf{A} . The statement “We define an $n \times m$ matrix as $\mathbf{A} = (a_{ij})$ ” stands for the matrix \mathbf{A} being expressed as (1.1).

1.2 Vectors

A vertical array,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad (1.4)$$

is called a *column vector* or simply a *vector*. In exactness, (1.4) is said to be an $n \times 1$ *vector*, since it contains n elements. Vectors can be viewed as a special case of matrices; (1.4) can also be called an $n \times 1$ matrix. Further, a *scalar* is a 1×1 matrix. The right side of (1.4) is vertically long, and for the sake of the economy of space, (1.4) is often expressed as

$$\mathbf{a} = [a_1, a_2, \dots, a_n]' \text{ or } \mathbf{a}' = [a_1, a_2, \dots, a_n], \quad (1.5)$$

using a transpose. A horizontal array as \mathbf{a}' is called a *row vector*.

We can use vectors to express a matrix: by using $n \times 1$ vectors $\mathbf{a}_j = [a_{1j}, a_{2j}, \dots, a_{nj}]'$, $j = 1, 2, \dots, m$, and $m \times 1$ vectors $\tilde{\mathbf{a}}_i = [a_{i1}, a_{i2}, \dots, a_{im}]'$, $i = 1, 2, \dots, n$, and the matrix (1.1) or (1.3) is expressed as

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] = \begin{bmatrix} \tilde{\mathbf{a}}_1' \\ \tilde{\mathbf{a}}_2' \\ \vdots \\ \tilde{\mathbf{a}}_n' \end{bmatrix} = [\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{a}}_n]' = (a_{ij}). \quad (1.6)$$

In this book, a **bold uppercase** letter such as \mathbf{X} is used for denoting a *matrix*, a **bold lowercase** letter such as \mathbf{x} is used for a *vector*, and an *italic* letter (not bold) such as x is used for a *scalar*. Though a *series of integers* has so far been expressed as $i = 1, 2, \dots, n$, this may be rewritten as $i = 1, \dots, n$, omitting 2 when it obviously follows 1. With this notation, (1.1) or (1.6) is rewritten as

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = [\mathbf{a}_1, \dots, \mathbf{a}_m] = [\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_n]'$$

1.3 Sum of Matrices and Their Multiplication by Scalars

The sum of matrices can be defined when they are of the *same size*. Let matrices **A** and **B** be equivalently $n \times m$. Their *sum* $\mathbf{A} + \mathbf{B}$ yields the $n \times m$ matrix, each of whose *elements is the sum of the corresponding ones* of $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$: The sum is defined as

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij}), \quad (1.7)$$

using the notation in (1.3). For example, when $\mathbf{X} = \begin{bmatrix} 3 & -2 & 6 \\ 8 & 0 & -2 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 2 & 1 & -9 \\ -7 & 2 & -3 \end{bmatrix}$,

$$\mathbf{X} + \mathbf{Y} = \begin{bmatrix} 3+2 & -2+1 & 6-9 \\ 8-7 & 0+2 & -2-3 \end{bmatrix} = \begin{bmatrix} 5 & -1 & -3 \\ 1 & 2 & -5 \end{bmatrix}.$$

The multiplication of matrix $\mathbf{A} = (a_{ij})$ by scalar s is defined as *all elements* of **A** being *multiplied* by s :

$$s\mathbf{A} = (s \times a_{ij}), \quad (1.8)$$

using the notation in (1.3). For example, when $\mathbf{Z} = \begin{bmatrix} 8 & -2 & 6 \\ -5 & 0 & -3 \end{bmatrix}$

$$\begin{aligned} -0.1\mathbf{Z} &= \begin{bmatrix} -0.1 \times 8 & -0.1 \times (-2) & -0.1 \times 6 \\ -0.1 \times (-5) & -0.1 \times 0 & -0.1 \times (-3) \end{bmatrix} \\ &= \begin{bmatrix} -0.8 & 0.2 & -0.6 \\ 0.5 & 0 & 0.3 \end{bmatrix}. \end{aligned}$$

The *sum of the matrices multiplied by scalars* is defined simply as the combination of (1.7) and (1.8):

$$v\mathbf{A} + w\mathbf{B} = (va_{ij} + wb_{ij}). \quad (1.9)$$

For example, when $\mathbf{X} = \begin{bmatrix} 4 & -2 & 6 \\ 8 & 0 & -2 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 2 & 1 & -9 \\ -7 & 2 & -3 \end{bmatrix}$,

$$0.5\mathbf{X} + (-2)\mathbf{Y} = \begin{bmatrix} 2-4 & -1-2 & 3+18 \\ 4+14 & 0-4 & -1+6 \end{bmatrix} = \begin{bmatrix} -2 & -3 & 21 \\ 18 & -4 & 5 \end{bmatrix}.$$

Obviously, setting $v = 1$ and $w = -1$ in (1.9) leads to the definition of the matrix difference $\mathbf{A} - \mathbf{B}$.

The above definition is generalized as

$$\sum_{k=1}^K v_k \mathbf{A}_k = v_1 \mathbf{A}_1 + \cdots + v_K \mathbf{A}_K = \left(\sum_{k=1}^K v_k a_{ijk} \right), \quad (1.10)$$

where $\mathbf{A}_1, \dots, \mathbf{A}_K$ are of the same size and a_{ijk} is the (i, j) element of \mathbf{A}_k ($k = 1, \dots, K$).

1.4 Inner Product and Norms of Vectors

The *inner product* of the vectors $\mathbf{a} = [a_1, \dots, a_m]'$ and $\mathbf{b} = [b_1, \dots, b_m]'$ is defined as

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = [a_1, \dots, a_m] \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = a_1 b_1 + \cdots + a_m b_m = \sum_{k=1}^m a_k b_k. \quad (1.11)$$

Obviously, this can be defined only for the vectors of the same size. The inner product is expressed as $\mathbf{a}'\mathbf{b}$ or $\mathbf{b}'\mathbf{a}$, i.e., the form of a *transposed column vector* (i.e., *row vector*) followed by a *column vector*, so as to be congruous to the matrix product introduced in the next section.

The inner product of the identical vectors \mathbf{a} and \mathbf{a} is in particular called the *squared norm* of \mathbf{a} and denoted as $\|\mathbf{a}\|^2$:

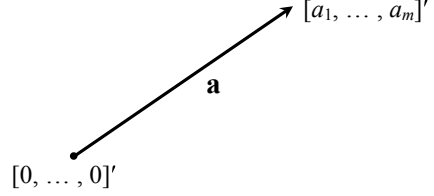
$$\|\mathbf{a}\|^2 = \mathbf{a}'\mathbf{a} = [a_1, \dots, a_m] \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = a_1^2 + \cdots + a_m^2 = \sum_{k=1}^m a_k^2. \quad (1.12)$$

The *square root* of $\|\mathbf{a}\|^2$, that is, $\|\mathbf{a}\|$ is simply called the *norm* of the vector $\mathbf{a} = [a_1, \dots, a_m]'$ with

$$\|\mathbf{a}\| = \sqrt{a_1^2 + \cdots + a_m^2}. \quad (1.13)$$

It is also called the *length* of \mathbf{a} , for the following reason. If $m = 3$ with $\mathbf{a} = [a_1, a_2, a_3]'$ and \mathbf{a} is viewed as the line extending from the origin to the point whose coordinate is $[a_1, a_2, a_3]'$, as illustrated in Fig. 1.1: (1.13) expresses the length of the line. It also holds for $m = 1, 2$. If $m > 3$, the line cannot be depicted or seen by those of us (i.e., the human beings living in three-dimensional world), but the length of \mathbf{a} is also defined as (1.13) for $m > 3$ in mathematics (in which the entities that do not exist in the real world are also considered if they are treated logically).

Fig. 1.1 Graphical representation of a vector



1.5 Product of Matrices

Let $n \times m$ and $m \times p$ matrices be defined as

$$\mathbf{A} \begin{bmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \text{ and } \mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_p] = \begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \vdots & \vdots \\ b_{m1} & \cdots & b_{mp} \end{bmatrix},$$

respectively, with $\mathbf{a}'_i = [a_{i1}, \dots, a_{im}] (i = 1, \dots, n)$ and $\mathbf{b}_j = \begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix} (j = 1, \dots, p)$.

Then, the *post-multiplication* of \mathbf{A} by \mathbf{B} is defined as

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}'_1 \mathbf{b}_1 & \cdots & \mathbf{a}'_1 \mathbf{b}_p \\ \vdots & \cdots & \vdots \\ \mathbf{a}'_n \mathbf{b}_1 & \cdots & \mathbf{a}'_n \mathbf{b}_p \end{bmatrix} = (\mathbf{a}'_i \mathbf{b}_j), \quad (1.14)$$

using the *inner products* of the row vectors of the preceding matrix \mathbf{A} and the column vectors of the following matrix \mathbf{B} . The resulting matrix \mathbf{AB} is the $n \times p$ matrix whose (i, j) element is the inner product of the i th row of \mathbf{A} and the j th column of \mathbf{B} :

$$\mathbf{a}'_i \mathbf{b}_j = [a_{i1}, \dots, a_{im}] \begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix} = a_{i1}b_{1j} + \cdots + a_{im}b_{mj} = \sum_{k=1}^m a_{ik}b_{kj}. \quad (1.15)$$

For example, if $\mathbf{A} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ 1 & 7 \end{bmatrix}$, $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2] = \begin{bmatrix} -3 & 1 \\ 2 & -5 \end{bmatrix}$, then

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \mathbf{a}'_1 \mathbf{b}_1 & \mathbf{a}'_1 \mathbf{b}_2 \\ \mathbf{a}'_2 \mathbf{b}_1 & \mathbf{a}'_2 \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 2 \times (-3) + (-4) \times 2 & 2 \times 1 + (-4) \times (-5) \\ 1 \times (-3) + 7 \times 2 & 1 \times 1 + 7 \times (-5) \end{bmatrix} \\ &= \begin{bmatrix} -14 & 22 \\ 11 & -34 \end{bmatrix}. \end{aligned}$$

As found above, the matrix product \mathbf{AB} is defined only when the following holds:

$$\text{the number of columns in } \mathbf{A} = \text{the number of rows in } \mathbf{B}. \quad (1.16)$$

The resulting matrix \mathbf{AB} is

$$(\text{the number of rows in } \mathbf{A}) \times (\text{the number of columns in } \mathbf{B}). \quad (1.17)$$

Thus, the product is sometimes expressed as

$$\mathbf{A}_{n \times m} \mathbf{B}_{m \times p} = \mathbf{C}_{n \times p}, \text{ or, more simply, } {}_n\mathbf{A}_m \mathbf{B}_p = {}_n\mathbf{C}_p, \quad (1.18)$$

with which we can easily verify (1.16) and (1.17). If $n = p$, we can define products \mathbf{AB} and \mathbf{BA} . Here, we should note

$$\mathbf{AB} \neq \mathbf{BA}, \quad (1.19)$$

except for special \mathbf{A} and \mathbf{B} , which is different from the product of scalars with $st = ts$, the inner product (1.11), and that of scalar s and matrix \mathbf{A} with

$$s\mathbf{A} = \mathbf{A} \times s. \quad (1.20)$$

For this reason, we call \mathbf{AB} “the *post-multiplication* of \mathbf{A} by \mathbf{B} ” or “the *pre-multiplication* of \mathbf{B} by \mathbf{A} ”, so as to clarify the order of the matrices.

Here, four examples of matrix products are presented as follows:

Ex. 1. For $\mathbf{X} = \begin{bmatrix} 2 & 3 & -1 \\ -2 & 0 & 4 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 3 & 5 & 4 \\ -1 & 0 & -2 \\ 0 & 6 & 0 \end{bmatrix}$,

$$\begin{aligned} \mathbf{XY} &= \begin{bmatrix} 2 \times 3 + 3 \times (-1) + (-1) \times 0 & 2 \times 5 + 3 \times 0 + (-1) \times 6 & 2 \times 4 + 3 \times (-2) + (-1) \times 0 \\ -2 \times 3 + 0 \times (-1) + 4 \times 0 & -2 \times 5 + 0 \times 0 + 4 \times 6 & -2 \times 4 + 0 \times (-2) + 4 \times 0 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 4 & 2 \\ -6 & 14 & -8 \end{bmatrix}. \end{aligned}$$

Ex. 2. For $\mathbf{F} = \begin{bmatrix} 2 & -1 \\ -3 & 0 \\ 1 & 3 \\ -2 & -3 \end{bmatrix}$ and $\mathbf{A} = \begin{bmatrix} -4 & 1 \\ 6 & -3 \\ 2 & 5 \end{bmatrix}$,

$$\begin{aligned}
\mathbf{FA}' &= \begin{bmatrix} 2 & -1 \\ -3 & 0 \\ 1 & 3 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} -4 & 6 & 2 \\ 1 & -3 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 2 \times (-4) + (-1) \times 1 & 2 \times 6 + (-1) \times (-3) & 2 \times 2 + (-1) \times 5 \\ -3 \times (-4) + 0 \times 1 & -3 \times 6 + 0 \times (-3) & -3 \times 2 + 0 \times 5 \\ 1 \times (-4) + 3 \times 1 & 1 \times 6 + 3 \times (-3) & 1 \times 2 + 3 \times 5 \\ -2 \times (-4) + (-3) \times 1 & -2 \times 6 + (-3) \times (-3) & -2 \times 2 + (-3) \times 5 \end{bmatrix} \\
&= \begin{bmatrix} -9 & 15 & -1 \\ 12 & -18 & -6 \\ -1 & -3 & 17 \\ 5 & -3 & -19 \end{bmatrix},
\end{aligned}$$

where it should be noted that \mathbf{A} has been transposed in the product.

Ex. 3. In statistics, the product of a matrix and its transpose is often used.

For $\mathbf{A} = \begin{bmatrix} -4 & 1 \\ 6 & -3 \\ 2 & 5 \end{bmatrix}$, the post-multiplication of \mathbf{A} by \mathbf{A}' , which we denote by \mathbf{S} , is

$$\begin{aligned}
\mathbf{S} = \mathbf{AA}' &= \begin{bmatrix} (-4)^2 + 1^2 & -4 \times 6 + 1 \times (-3) & -4 \times 2 + 1 \times 5 \\ 6 \times (-4) + (-3) \times 1 & 6^2 + (-3)^2 & 6 \times 2 + (-3) \times 5 \\ 2 \times (-4) + 5 \times 1 & 2 \times 6 + 5 \times (-3) & 2^2 + 5^2 \end{bmatrix} \\
&= \begin{bmatrix} 17 & -27 & -3 \\ -27 & 45 & -3 \\ -3 & -3 & 29 \end{bmatrix}.
\end{aligned}$$

The pre-multiplication of \mathbf{A} by \mathbf{A}' , which we denote by \mathbf{T} , is

$$\begin{aligned}
\mathbf{T} = \mathbf{A}'\mathbf{A} &= \begin{bmatrix} (-4)^2 + 6^2 + 2^2 & (-4) \times 1 + 6 \times (-3) + 2 \times 5 \\ 1 \times (-4) + (-3) \times 6 + 5 \times 2 & 1^2 + (-3)^2 + 5^2 \end{bmatrix} \\
&= \begin{bmatrix} 56 & -12 \\ -12 & 35 \end{bmatrix}.
\end{aligned}$$

Ex. 4. The product of vectors is a special case of that of matrices:

$$\text{For } \mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix},$$

the inner product yields a scalar as

$$\mathbf{u}'\mathbf{v} = 2 \times (-2) + (-1) \times 3 + 3 \times (-4) = -19,$$

but the post-multiplication of 3×1 vector \mathbf{u} by 1×3 \mathbf{v}' gives a 3×3 matrix with

$$\begin{aligned} \mathbf{uv}' &= \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \begin{bmatrix} -2 & 3 & -4 \end{bmatrix} = \begin{bmatrix} 2 \times (-2) & 2 \times 3 & 2 \times (-4) \\ (-1) \times (-2) & (-1) \times 3 & (-1) \times (-4) \\ 3 \times (-2) & 3 \times 3 & 3 \times (-4) \end{bmatrix} \\ &= \begin{bmatrix} -4 & 6 & -8 \\ 2 & -3 & 4 \\ -6 & 9 & -12 \end{bmatrix}. \end{aligned}$$

1.6 Two Properties of Matrix Products

The *transposed product* of matrices satisfies

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'; (\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}' \quad (1.21)$$

Let \mathbf{A} and \mathbf{B} be matrices of size $n \times m$; let \mathbf{C} and \mathbf{D} be those of $m \times l$. Then, the *product of their sums multiplied by scalars* s, t, u , and v satisfies

$$(s\mathbf{A} + t\mathbf{B})(u\mathbf{C} + v\mathbf{D}) = su\mathbf{AC} + sv\mathbf{AD} + tu\mathbf{BC} + tv\mathbf{BD}. \quad (1.22)$$

1.7 Trace Operator and Matrix Norm

A matrix with the number of rows equivalent to that of columns is said to be

square. For a square matrix $\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}$, the elements on the

diagonal, i.e., s_{11}, \dots, s_{nn} , are called the *diagonal elements* of \mathbf{S} . Their sum is called a *trace* and is denoted as

$$\text{tr}\mathbf{S} = s_{11} + s_{22} + \cdots + s_{nn}. \quad (1.23)$$

Obviously,

$$\text{tr} \mathbf{S}' = \text{tr} \mathbf{S} \quad (1.24)$$

The trace fulfills important roles when it is defined for the product of matrices.

Let us consider $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_m] = \begin{bmatrix} \tilde{\mathbf{a}}'_1 \\ \vdots \\ \tilde{\mathbf{a}}'_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$ and $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_n] = \begin{bmatrix} \tilde{\mathbf{b}}'_1 \\ \vdots \\ \tilde{\mathbf{b}}'_m \end{bmatrix} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$. Then, \mathbf{AB} and \mathbf{BA} are $n \times n$ and $m \times m$ square matrices, respectively, for which traces can be defined, with

$$\mathbf{AB} = \begin{bmatrix} \tilde{\mathbf{a}}'_1 \mathbf{b}_1 & & \# \\ & \ddots & \\ \# & & \tilde{\mathbf{a}}'_n \mathbf{b}_n \end{bmatrix} \text{ and } \mathbf{BA} = \begin{bmatrix} \tilde{\mathbf{b}}'_1 \mathbf{a}_1 & & \# \\ & \ddots & \\ \# & & \tilde{\mathbf{b}}'_m \mathbf{a}_m \end{bmatrix}.$$

Here, $\#$ is used for all elements other than the diagonal ones. In this book, the matrix product precedes the trace operation:

$$\text{tr} \mathbf{AB} = \text{tr}(\mathbf{AB}). \quad (1.25)$$

Thus,

$$\text{tr} \mathbf{AB} = \sum_{i=1}^n \tilde{\mathbf{a}}'_i \mathbf{b}_i = \sum_{i=1}^n (a_{i1} b_{1i} + \cdots + a_{im} b_{mi}) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} b_{ji}, \quad (1.26)$$

$$\text{tr} \mathbf{BA} = \sum_{j=1}^m \tilde{\mathbf{b}}'_j \mathbf{a}_j = \sum_{j=1}^m (b_{j1} a_{1j} + \cdots + b_{jn} a_{nj}) = \sum_{j=1}^m \sum_{i=1}^n b_{ji} a_{ij} = \sum_{i=1}^n \sum_{j=1}^m a_{ij} b_{ji}. \quad (1.27)$$

Both are found to be equivalent, i.e.,

$$\text{tr} \mathbf{AB} = \text{tr} \mathbf{BA}, \quad (1.28)$$

and express the sum of $a_{ij} b_{ji}$ over all pairs of i and j .

It is an important property of the trace that (1.28) implies

$$\text{tr} \mathbf{ABC} = \text{tr} \mathbf{CAB} = \text{tr} \mathbf{BCA}; \text{tr} \mathbf{ABCD} = \text{tr} \mathbf{BCDA} = \text{tr} \mathbf{CDAB} = \text{tr} \mathbf{DABC}. \quad (1.29)$$

Using (1.21), (1.28), and (1.29), we also have

$$\text{tr}(\mathbf{AB})' = \text{tr}\mathbf{B}'\mathbf{A}' = \text{tr}\mathbf{A}'\mathbf{B}'; \text{tr}(\mathbf{ABC})' = \text{tr}\mathbf{C}'\mathbf{B}'\mathbf{A}' = \text{tr}\mathbf{A}'\mathbf{C}'\mathbf{B}' = \text{tr}\mathbf{B}'\mathbf{A}'\mathbf{C}'. \quad (1.30)$$

Substituting \mathbf{A}' into \mathbf{B} in (1.25), we have $\text{tr}\mathbf{AA}' = \text{tr}\mathbf{A}'\mathbf{A} = \sum_{i=1}^n \sum_{j=1}^m a_{ij}^2$ which is the sum of the squared elements of \mathbf{A} . This is called the *squared norm* of \mathbf{A} , i.e., the matrix version of (1.12), and is denoted as $\|\mathbf{A}\|^2$:

$$\|\mathbf{A}\|^2 = \text{tr}\mathbf{AA}' = \text{tr}\mathbf{A}'\mathbf{A} = \sum_{i=1}^n \sum_{j=1}^m a_{ij}^2. \quad (1.31)$$

This is also referred to as the *squared Frobenius norm* of \mathbf{A} , with Frobenius (1849–1917) a German mathematician. The squared norm of the sum of matrices weighted by scalars is expanded as

$$\begin{aligned} \|s\mathbf{X} + t\mathbf{Y}\|^2 &= \text{tr}(s\mathbf{X} + t\mathbf{Y})'(s\mathbf{X} + t\mathbf{Y}) \\ &= \text{tr}(s^2\mathbf{X}'\mathbf{X} + st\mathbf{X}'\mathbf{Y} + ts\mathbf{Y}'\mathbf{X} + t^2\mathbf{Y}'\mathbf{Y}) \\ &= s^2\text{tr}\mathbf{X}'\mathbf{X} + st\text{tr}\mathbf{X}'\mathbf{Y} + st\text{tr}\mathbf{X}'\mathbf{Y} + t^2\text{tr}\mathbf{Y}'\mathbf{Y} \\ &= s^2\text{tr}\mathbf{X}'\mathbf{X} + 2st\text{tr}\mathbf{X}'\mathbf{Y} + t^2\text{tr}\mathbf{Y}'\mathbf{Y} \\ &= s^2\|\mathbf{X}\|^2 + 2st\text{tr}\mathbf{X}'\mathbf{Y} + t^2\|\mathbf{Y}\|^2. \end{aligned} \quad (1.32)$$

1.8 Vectors and Matrices Filled with Ones or Zeros

A *zero vector* refers to a vector filled with zeros. The $p \times 1$ zero vector is denoted as $\mathbf{0}_p$, using the boldfaced zero:

$$\mathbf{0}_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (1.33)$$

A zero matrix refers to a matrix whose elements are all zeros. In this book, the $n \times p$ zero matrix is denoted as ${}_n\mathbf{O}_p$, using the boldfaced “O”:

$${}_n\mathbf{O}_p = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}. \quad (1.34)$$

A *vector of ones* refers to a vector filled with ones. The $n \times 1$ vector of ones is denoted as $\mathbf{1}_n$, with the boldfaced one:

$$\mathbf{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (1.35)$$

The $n \times p$ matrix filled with ones is given by

$$\mathbf{1}_n \mathbf{1}_p' = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \cdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}. \quad (1.36)$$

1.9 Special Square Matrices

A square matrix $\mathbf{S} = (s_{ij})$ satisfying

$$\mathbf{S} = \mathbf{S}', \text{ i.e., } s_{ij} = s_{ji} \quad (1.37)$$

is said to be *symmetric*. An example of a 3×3 symmetric matrix is

$$\mathbf{S} = \begin{bmatrix} 2 & -4 & 9 \\ -4 & 6 & -7 \\ 9 & -7 & 3 \end{bmatrix}. \text{ The products of a matrix } \mathbf{A} \text{ and its transpose, } \mathbf{A}\mathbf{A}' \text{ and } \mathbf{A}'\mathbf{A}, \text{ are symmetric; using (1.21), we have}$$

$$(\mathbf{A}\mathbf{A}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{A}\mathbf{A}' \text{ and } (\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}. \quad (1.38)$$

This has already been exemplified in Ex. 3 (Sect. 1.5).

The elements of $\mathbf{A} = (a_{ij})$ with $i \neq j$ are called the off-diagonal elements of \mathbf{A} . A square matrix \mathbf{D} whose off-diagonal elements are all zeros,

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & d_p \end{bmatrix}, \quad (1.39)$$

is called a *diagonal matrix*. The products of two diagonal matrices are easily obtained as