

Simo Puntanen
George P. H. Styan
Jarkko Isotalo

Matrix Tricks for Linear Statistical Models

Our Personal Top Twenty

Simo Puntanen
School of Information Sciences
FI-33014 University of Tampere
Finland
simo.puntanen@uta.fi

George P. H. Styan
Department of Mathematics & Statistics
McGill University
805 ouest, Sherbrooke Street West
Montréal (Québec) H3A 2K6
Canada
styan@math.mcgill.ca

Jarkko Isotalo
School of Information Sciences
FI-33014 University of Tampere
Finland
jarkko.isotalo@uta.fi

ISBN 978-3-642-10472-5 e-ISBN 978-3-642-10473-2
DOI 10.1007/978-3-642-10473-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011935544

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Printed on acid-free paper

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

Preface

In teaching linear statistical models to first-year graduate students or to final-year undergraduate students there is no way to proceed smoothly without matrices and related concepts of linear algebra; their use is really essential. Our experience is that making some particular matrix tricks very familiar to students can substantially increase their insight into linear statistical models (and also multivariate statistical analysis). In matrix algebra, there are handy, sometimes even very simple “tricks” which simplify and clarify the treatment of a problem—both for the student and for the professor. Of course, the concept of a *trick* is not uniquely defined—by a trick we simply mean here a useful important handy result. Notice the three adjectives we use here, useful, important, and handy, to describe the nature of the tricks to be considered.

In this book we collect together our Top Twenty favourite matrix tricks for linear statistical models. Interestingly, nobody can complain that our title is wrong; someone else could certainly write a different book with the same title.

Structure of the Book

Before presenting our Top Twenty, we offer a quick tour of the notation, linear algebraic preliminaries, data matrices, random vectors, and linear models. Browsing through these pages will familiarize the reader with our style. There may not be much in our Introduction that is new, but we feel it is extremely important to become familiar with the notation to be used. To take a concrete example, we feel that it is absolutely necessary for the reader of this book to remember that throughout we use \mathbf{H} and \mathbf{M} , respectively, for the orthogonal projectors onto the column space of the model matrix \mathbf{X} and onto its orthocomplement. A comprehensive list of symbols with explanations is given on pages 427–434.

In our book we have inserted photographs of some matricians¹ and statisticians. We have also included images of some philatelic items—for more about mathematical philatelic items we recommend the book by Wilson (2001) and the website by Miller (2010); see also Serre (2007). For “stochastic stamps”—postage stamps that are related in some way to chance, i.e., that have a connection with probability and/or statistics, see Puntanen & Styan (2008b).

We do not claim that there are many new things in this book even after the Introduction—most of the results can be found in the literature. However, we feel that there are some results that might have appeared in the literature in a somewhat concealed form and here our aim is to upgrade their appreciation—to put them into “business class”.

Twenty chapters correspond to our Top Twenty tricks, one chapter for each trick. Almost every chapter includes sections which are really examples illustrating the use of the trick. Most but not all sections are statistical. Most chapters end with a set of exercises (without solutions). Each section itself, however, serves as an exercise—with solution! So if you want to develop your skills with matrices, please look at the main contents of a particular section that interests you and try to solve it on your own. This is highly recommended!

Our Tricks are not all of the same size² as is readily seen from the number of pages per chapter. There are some further matrix tools that might well have deserved a place in this book. For example, Lagrange multipliers and matrix derivatives, see, e.g., Ito & Kunisch (2008), Magnus & Neudecker (1999), as well as Hadamard products and Kronecker products, see, e.g., Graham (1981), Horn (1990), Styan (1973a), could have been added as four more Tricks.

Our twenty tricks are not necessarily presented in a strictly logical order in the sense that the reader must start from Chapter 1 and then proceed to Chapter 2, and so on. Indeed our book may not build an elegant mathematical apparatus by providing a construction piece by piece, with carefully presented definitions, lemmas, theorems, and corollaries. Our first three chapters have the word “easy” in their title: this is to encourage the reader to take a look at these chapters first. However, it may well be appropriate to jump directly into a particular chapter if the trick presented therein is of special interest to the reader.

¹ Farebrother (2000) observed that: “As the word matrician has not yet entered the English language, we may either use it to denote a member of the ruling class of a matriarchy or to denote a person who studies matrices in their medical, geological, or mathematical senses.”

² Unlike Greenacre (2007, p. xii), who in his Preface says he “wanted each chapter to represent a fixed amount to read or teach, and there was no better way to do that than to limit the length of each chapter—each chapter is exactly eight pages in length.”

Matrix Books for Statisticians

An interesting signal of the increasing importance of matrix methods for statistics is the recent publication of several books in this area—this is not to say that matrices were not used earlier—we merely wish to identify some of the many recently-published matrix-oriented books in statistics. The recent book by Seber (2008) should be mentioned in particular: it is a delightful handbook for a matrix-enthusiast-statistician. Other recent books include: Bapat (2000), Hadi (1996), Harville (1997, 2001), Healy (2000), Magnus & Neudecker (1999), Meyer (2000), A.R. Rao & Bhimasankaram (2000), C.R. Rao & M.B. Rao (1998), and Schott (2005) (first ed. 1996). Among somewhat older books we would like to mention Graybill (2002) (first ed. 1969), Seber (1980) (first ed. 1966), and Searle (1982).

There are also some other recent books using or dealing with matrix algebra which is helpful for statistics: for example, Abadir & Magnus (2005), Bernstein (2009) (first ed. 2005), Christensen (2001, 2002), Fujikoshi, Ulyanov & Shimizu (2010), Gentle (2007), Groß (2003), Härdle & Hlávka (2007), Kollo & von Rosen (2005), S.K. Mitra, Bhimasankaram & Malik (2010), Pan & Fang (2002), C.R. Rao, Toutenburg, Shalabh et al. (2008) (first ed. 1995), Seber & A. J. Lee (2003) (first ed. 1977), Sengupta & Jammalamadaka (2003), Srivastava (2002), S.-G. Wang & Chow (1994), F. Zhang (1999; 2009), and the *Handbook of Linear Algebra* by Hogben (2007).

New arrivals of books on linear statistical models are appearing at a regular pace: Khuri (2009), Monahan (2008), Rencher & Schaalje (2008) (first ed. 1999), Ryan (2009) (first ed. 1997), Stapleton (2009) (first ed. 1995), Casella (2008), and Toutenburg & Shalabh (2009), to mention a few; these last two books deal extensively with experimental design which we consider only minimally in this book. As Draper & Pukelsheim (1996, p. 1) point out, “the topic of statistical design of experiments could well have an entire encyclopedia devoted to it.” In this connection we recommend the recent *Handbook of Combinatorial Designs* by Colbourn & Dinitz (2007).

There are also some books in statistics whose usefulness regarding matrix algebra has long been recognized: for example, the two classics, *An Introduction to Multivariate Statistical Analysis* by T.W. Anderson (2003) (first ed. 1958) and *Linear Statistical Inference and its Applications* by C.R. Rao (1973a) (first ed. 1965), should definitely be mentioned in this context. Both books also include excellent examples and exercises related to matrices in statistics. For generalized inverses, we would like to mention the books by C.R. Rao & S.K. Mitra (1971b) and by Ben-Israel & Greville (2003) (first ed. 1974), and the recent book by Piziak & Odell (2007). For Schur complements, see the recent book by Zhang (2005b) and the articles in this book by Puntanen & Styan (2005a,b).

The Role of Matrices in Statistics

For interesting remarks related to matrices in statistics, see the papers by Farebrother (1996, 1997), Olkin (1990, 1998), Puntanen, Seber & Styan (2007), Puntanen & Styan (2007), Searle (1999, 2000). Of special interest also are those papers in which statistical ideas are used to prove some matrix theorems, especially matrix inequalities; see, e.g., S.K. Mitra (1973a), S.K. Mitra & Puntanen (1991), Dey, Hande & Tiku (1994), and C.R. Rao (2000; 2006). As for the development of the use of matrices in statistics, we would like to refer to Searle (2005) and to the conversation by Wells (2009, p. 251), with Shayle R. Searle:

WELLS: “You were an early advocate of using matrices in statistics, looking back this prospective seems obvious. Do you have a conjecture why early progress on the application of matrices was so slow?”

SEARLE: “The first of my *Annals* papers of (1956), 1958 and 1961 was ‘Matrix methods in variance and covariance components analysis’. Its title begs the question: Why has it taken so long for matrices to get widely adopted where they are so extremely useful? After all, matrices are two hundred and some years old and their use in statistics is only slowly becoming commonplace. But this was not so, even as recently as the 1950s. Even at Cambridge, in lectures on regression in 1952 there was no use of matrices.”

Many thanks to Kimmo Vehkalahti for alerting us to the following interesting remarks by Bock (2007, p. 41):

“The year 1934 and part of 1935 was a period of intense work for Thurstone. In 1935, the University of Chicago Press published *The Vectors of Mind*, his extended treatise on multiple factor analysis. [...] It also includes, for the first time in the psychological or statistical literature, an introductory section containing definitions and results of matrix algebra and their geometrical interpretations. [...] As a matter of interest, I reviewed both the *Journal of the American Statistical Association* and *The Annals of Mathematical Statistics* looking for applications of matrix algebra before 1935. Even as late as 1940, *JASA* contained not only no matrix algebra, but hardly any algebra at all; it was still largely a journal of statistics in the old sense—the presentation and analysis of tables of economic and social indicators. The earliest instance of matrix algebra I could find was in the *AMS*, Volume 6, 1935, in an article by Y.K. Wong entitled, ‘Application of orthogonalization processes to the theory of least-squares’.”

Our Aim

In summary: our aim is not to go through all the steps needed to develop the use of matrices in statistics. There are already several books that do that in a nice way. Our main aim is to present *our personal favourite tools* for the interested student or professor who wishes to develop matrix skills for linear models. We assume that the reader is somewhat familiar with linear algebra, matrix calculus, linear statistical models, and multivariate statistical analysis,

although a thorough knowledge is not needed, one year of undergraduate study of linear algebra and statistics is expected. A short course in regression would also be welcome before going deeply into our book. Here are some examples of smooth introductions to regression: Chatterjee & Hadi (2006) (first ed. 1977), Draper & Smith (1998) (first ed. 1966), and Weisberg (2005) (first ed. 1980).

We have not included any real data or any discussion of computer software—these are beyond the scope of this book. In real life, when facing a class of statistics students, there is no way to ignore the need of modern high-calibre statistical software, allowing, in particular, access to flexible matrix manipulation. Nowadays, pen and paper are simply not enough—clever software can uncover or increase our understanding of lengthy formulas.

As regards the references, our attempt is to be rather generous—but not necessarily thorough. Speed (2009, p. 13) asks “Do you ever wonder how we found references before the www?” and follows with interesting remarks on today’s methods of finding references. We recommend *Current Index to Statistics*, *MathSciNet* and *ZMATH*, as well as *Google*!

The material of this book has been used in teaching statistics students at the University of Tampere for more than 10 years. Warm thanks to all those students for inspiring cooperation! The idea has been to encourage students to develop their skills by emphasizing the tricks that we have learnt to be used again and again. Our belief is that it is like practicing a sport: practice makes perfect.

One possible way to use this book as a textbook for a course might be just to go through the exercises in this book. Students should be encouraged to solve and discuss them on the blackboard without any notes. The idea would be to push the student into the waves of individual thinking.

Kiitos!

We are most grateful to Theodore W. Anderson for introducing matrix methods for statistics to the second author in the early 1960s, and for supervising his Ph.D. thesis (Styan, 1969) at Columbia University; this then led to the Ph.D. theses at the University of Tampere by the first author (Puntanen, 1987) and by the third author (Isotalo, 2007).

Much of the work reported in this book follows the extensive collaboration by the first authors with Jerzy K. Baksalary (1944–2005) during the past 30 years or so; we feel he had an enormous effect on us, and if he were still alive, he would certainly be very interested (and critical as always) in this book. Some of our joint adventures in the column and row spaces are described in Puntanen & Styan (2008a) and in Isotalo, Puntanen & Styan (2008b). For an appreciation of Jerzy by many scholars and for a complete bibliography of his publications see Oskar Maria Baksalary & Styan (2007).

Sincere thanks go also to Oskar Maria Baksalary, Ka Lok Chu, Maria Ruíz Delmàs, S.W. Drury, Katarzyna Filipiak, Stephen Haslett, Augustyn Markiewicz, Ville Puntanen, Daniel J.-H. Rosenthal, Evelyn Matheson Styan, Götz Trenkler, and Kimmo Vehkalahti, and to the anonymous reviewers for their help. Needless to say that ...³

We give special thanks to Jarmo Niemelä for his outstanding help in setting up the many versions of this book in L^AT_EX. The figures for scatter plots were prepared using Survo software, online at <http://www.survo.fi>, (again thanks go to Kimmo Vehkalahti) and the other figures using PSTricks (again thanks go to Jarmo Niemelä).

We are most grateful to John Kimmel for suggesting that we write this monograph, and to Niels Peter Thomas and Lilith Braun of Springer for advice and encouragement.

Almost all photographs were taken by Simo Puntanen and are based on his collection (Puntanen, 2010a); exceptions are the photographs of Gene H. Golub and Shayle R. Searle, which were taken by George P.H. Styan and Harold V. Henderson, respectively. The photograph of the three authors with their mentor (p. xi) was taken by Soile Puntanen. The photographs of the two Markovs are taken from Wikipedia and the photograph of a Tappara defenceman is based on a hockey card. The images of philatelic items are based on items in the collection of George P.H. Styan (2010); we owe our gratitude also to Somesh Das Gupta (1935–2006) for providing us with the first-day cover for Mahalanobis (p. 90). *Scott* catalogue numbers are as given in the *Scott Standard Postage Stamp Catalogue* (Kloetzel, 2010).

The website <http://mtl.uta.fi/matrixtricks> supports the book by additional material.

The research for this monograph was supported in part by the Natural Sciences and Engineering Research Council of Canada.

Dedication

To Soile and Evelyn

and

to Theodore W. Anderson—the academic grandfather, father and great grandfather, respectively, of the three authors.

SP, GPHS & JI
11 April 2011

MSC 2000: 15-01, 15-02, 15A09, 15A42, 15A99, 62H12, 62J05.

³ cf. Flury (1997, p. ix).

Key words and phrases: Best linear unbiased estimation, Cauchy–Schwarz inequality, column space, eigenvalue decomposition, estimability, Gauss–Markov model, generalized inverse, idempotent matrix, linear model, linear regression, Löwner ordering, matrix inequalities, oblique projector, ordinary least squares, orthogonal projector, partitioned linear model, partitioned matrix, rank cancellation rule, reduced linear model, Schur complement, singular value decomposition.



Photograph 0.1 (From left to right) Jarkko Isotalo, Simo Puntanen, George P.H. Styan and Theodore W. Anderson, during the 15th International Workshop on Matrices and Statistics in Uppsala, Sweden: 15 June 2006.