INTERNATIONAL IMS-ILAS WORKSHOP ON MATRIX METHODS FOR STATISTICS

4 & 5 December 1992
University of Auckland
New Zealand
International IMS-ILAS Workshop on Matrix Methods for Statistics

Auckland, New Zealand: 4-5 December 1992

Registration will take place from 9 a.m. onwards on Friday, 4 December 1992, in the
Department of Mathematics & Statistics Resource Centre, University of Auckland, 8th
Building, 36-38 Princess Street. All lectures will be given in Lecture Theatre B10 which is situated in the Basement of the General Library Building.

FRIDAY 4 DECEMBER

10:00 - 10:20 Opening Remarks by George P. H. Styan, Chair: Alastair J. Scott
10:20 - 10:30 Group Photo
10:30 - 11:00 Coffee break
11:00 - 11:30 SESSION 1, Chair: Shayle R. Searle
11:00 - 11:30 John S. Chipman
"Generalized matrix Schwarz inequality to evaluation of biased estimation in
linear regression"
11:30 - 12:00 George P. H. Styan
"On the efficiency of a linear unbiased estimator and on a matrix version of the
Cauchy-Schwarz inequality"
12:00 - 12:30 Simo Puntanen
"Matrix tricks related to deleting an observation in the general linear model"
12:30 - 14:00 Lunch
14:00 - 14:30 SESSION 2, Chair: George P. H. Styan
14:00 - 14:15 Garry J. Tee
"Alexander Craig Aitken: 1895-1967"
14:15 - 15:30 Richard William Farebrother
"Statistical contributions to matrix methods"
15:30 - 16:00 Coffee break
16:00 - 17:00 SESSION 3, Chair: Jeffrey J. Hunter
16:00 - 16:30 Peter Clifford
"On the distribution of Pearson’s correlation coefficient in the presence of
spatial autocorrelation"
16:30 - 17:00 David J. Vere-Jones
"Generalized permanents and their applications to multivariate negative binomial
distributions"

17:15 - 18:45 Reception
19:30 - late Workshop Dinner in Berlin

SATURDAY 5 DECEMBER

9:00 - 10:30 SESSION 4, Chair: Peter Clifford
9:00 - 9:30 David J. Saville and Graham R. Wood
"How not to use matrices when teaching statistics"
9:30 - 10:00 Thomas Mathew
"Combining independent tests for a common mean: an application of the parallel
sum of matrices"
10:00 - 10:30 Renae Meyer
"Invariant preorderings of matrices and approximation problems in multivariate
statistics and multidimensional scaling"
10:30 - 11:00 Coffee break
11:00 - 12:00 SESSION 5, Chair: Harold V. Henderson
11:00 - 11:30 Shayle R. Searle
"Further results and proofs for the singular linear model"
11:30 - 12:00 Michael G. Schimek
"Problems with direct solutions of the normal equations for nonparametric
models"
12:00 - 14:00 Lunch
14:00 - 15:15 SESSION 6, Chair: John H. Maindonald
14:00 - 15:15 Chris C. Paige
"The full CS-decomposition of a partitioned orthogonal matrix"
15:15 - 15:45 Coffee break
15:45 - 16:45 SESSION 7, Chair: Simo Puntanen
15:45 - 16:15 Jeffrey J. Hunter
"Stationary distributions and mean first passage times in Markov chains using
generalized inverses"
16:15 - 16:45 Alastair J. Scott
"Characterizing invariant convex functions of matrices"
INTERNATIONAL IMS-ILAS WORKSHOP
ON MATRIX METHODS
FOR STATISTICS

Auckland, New Zealand
4-5 December 1992

ABSTRACTS
Characterizing Invariant Convex Functions of Matrices

James V. Bondar and Alastair J. Scott
Carleton University and University of Auckland

Some years ago, C. Davis (1957) showed that the invariant convex functions on the class of symmetric matrices are precisely the functions that are convex increasing functions of the eigenvalues. (Here “invariant” means invariant under the similarity transformation $O \rightarrow OMO'$ where $O$ is orthogonal.) In this talk we show that the invariant convex functions on the class of all matrices are precisely those functions of the singular values that are both convex and weak Schur convex. (Here “invariant” means invariant under the transformation $M \rightarrow O_1MO_2$ where the $O_i$ are orthogonal.)

It has been known for some time that increasing symmetric convex functions of the squares of the singular values are matrix convex. Bondar (1985) strengthens Schwartz’ results to increasing symmetric convex functions of the singular values. Here we show that to be a convex, weak Schur convex function of the singular values is necessary and sufficient for matrix convexity. We also derive sufficient conditions that are simpler to apply in practice.

References
C. Davis (1957). All convex invariant functions of Hermitian matrices. \textit{Arch. Math.}, 8: 276-278.
In the general linear regression model, three estimators are considered: (1) the Gauss-Markoff estimator; (2) the Gauss-Markoff estimator subject to a set of linear restrictions on the regression-coefficient vector; (3) the Theil-Goldberger estimator with uncertain linear restrictions corresponding to the certain ones of (2). This third estimator is styled the generalized ridge estimator. It is shown, generalizing a result of Toro-Vizcarrondo and Wallace, that a sufficient condition for both estimators (2) and (3) to have lower matrix-mean-square error than estimator (1) is that the noncentrality parameter arising in the $F$-test for the restrictions in (2) should be less than 1. A generalization is also obtained of the Hoerl-Kennard result that for sufficiently small but positive values of their $k$-parameter, estimator (3) has lower matrix-mean-square error than estimator (1).

References
Statistical Contributions to Matrix Methods

Richard William Farebrother

University of Manchester

The Formalisation of Matrix Algebra in the 1850s and 1860s was preceded by a number of significant developments, some of which (*) are to be found in statistics works:

1. Cauchy’s use of modern subscript notation for matrix elements.
2. Gauss’s use of matrix multiplication when transforming variables in a quadratic form.
3. Mayer’s prototype of the elimination procedure.
4. Gauss’s identification of a linearly dependent system of equations.
5. Gauss’s elimination procedure and the implicit LU decomposition of a square matrix.
7. Gauss’s derivation of a $LDL^T$ decomposition of a symmetric matrix.
8. Gauss’s least squares updating formulas.
9. Gauss’s proof that the inverse of a symmetric matrix is itself symmetric.
10. Cauchy’s latent value decomposition of a $3 \times 3$ matrix.
11. Laplace’s derivation of an orthogonalisation procedure subsequently rediscovered by Gram and Schmidt.
12. Donkin’s definition of an orthogonal basis for the orthogonal complement of a matrix.
13. Jacobi’s characterisation of the least squares solution as a weighted sum of subset estimates. This characterisation of the Least Squares solution as a weighted sum of subset estimates by Glaisher and Subrahmaniam.

Later Contributions Include:

14. Thiele’s discussion of the canonical form of the linear model.
15. The Singular value decomposition derived by Eckart and Young.
16. The so-called Gauss-Markov theorem is due to Gauss alone.
17. The generalisation of this theorem usually attributed to Aitken would seem to be due to Plackett.
Stationary Distributions and Mean First Passage Times in Markov Chains using Generalized Inverses

Jeffrey J. Hunter

Massey University

The determination of the mean first passage times in finite irreducible discrete time Markov chains requires the computation of a generalized inverse of $I - P$, where $P$ is the transition matrix of the Markov chain, and also knowledge of the stationary distribution of the Markov chain. Generalized inverses can be used to find stationary distributions. The linking of these two procedures and the computation of stationary distributions using various multi-condition generalized inverses is investigated.
Combining Independent Tests For A Common Mean  
– An Application of the Parallel Sum of Matrices

Thomas Mathew

Department of Mathematics and Statistics
University of Maryland
Baltimore County
Baltimore, Maryland 21228
U.S.A.

In the context of the recovery of inter-block information in a BIBD, the problem of combining two independent tests is addressed in Cohen and Sackrowitz (1989, Journal of the American Statistical Association). Apart from the intra- and inter-block $F$-tests, their combined test uses a correlation type statistic, and this statistic has a positive expected value when the null hypothesis of equality of the treatment effects is not true. We show that a similar statistic can be defined in the context of hypothesis testing for a common mean in several independent linear models. Using the properties of the parallel sum of matrices, we also show that this statistic has a positive expected value when the null hypothesis is not true.
Invariant Preorderings of Matrices and Approximation Problems in Multivariate Statistics and MDS

Renate Meyer

Institute of Statistics and Documentation
Aachen University of Technology, Germany

The singular value decomposition of a complex $n \times k$-matrix $A = \sum_{i=1}^{k} \sigma_i(A) u_i v_i^*$, with $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_k(A) > 0$ the nonnegative square roots of the eigenvalues of the Hermitian matrix $A^* A$, plays an important role in various multivariate descriptive statistical methods, as for example in principal components and analysis, canonical correlation analysis, discriminant analysis, and correspondence analysis. It is applied to solve the key problem of approximating a given $n \times k$-matrix $(1 < k \leq n)$ by a matrix of lower rank $r < k$. Define $A_{(r)} = \sum_{i=1}^{r} \sigma_i(A) u_i v_i^*$. The minimum norm rank $r$ approximation result

$$\psi(A - A_{(r)}) \leq \psi(A - G)$$

for all $G$ with rank $(G) \leq r$, (0.1)

was obtained by Eckart and Young (1936) for the Euclidean norm $\psi(A = \{\text{tr}AA^*\}^{\frac{1}{2}}$. The solution $A_{(r)}$ turned out to be "robust" with respect to the choice of the approximation criterion, as it was extended by Mirsky (1960) to the class of unitarily invariant norms.

The objective in the first part of this paper is to generalize this result to an even larger class of real-valued loss functions, encompassing the unitarily invariant norms. Certain preorderings $\preceq$ of complex matrices, that occur naturally in the context of multivariate statistics, will be considering in proving universal optimality of $A_{(r)}$, i.e.

$$A - A_{(r)} \preceq A - G$$

for all $G$ with rank $(G) \leq r$, (0.2)

Of course, this immediately entails (1) for all real-valued $\preceq$-monotone functions $\psi$.

The second part treats the problem of approximating a Hermitian $n \times n$ matrix by a positive-semidefinite matrix of given rank, which is of major relevance in the context of multidimensional scaling (MDS). Thereby, the hitherto most general result of Mathar (1985) is extended and further universally optimal properties of the classical MDS solution are provided.

Key words: singular value decomposition, group induced preorderings, invariant orderings, weak majorization, principal components analysis, multidimensional scaling.

References


A Compact Expression For Variance of Sample Second-order Moments in Multivariate Linear Relations

Heinz Neudecker

Department of Actuarial Science and Econometrics
University of Amsterdam
The Netherlands

Satorra (1992) considered the random \((p \times l)\) vector

\[ z = \sum_{i=1}^{m} B_i \delta_i + \mu, \]

where the random \((n_i \times l)\) vectors \(\delta_i\) are independent with mean \(E(\delta_i) = 0\) and variance \(D(\delta_i)\), and \(B_i\) are \(\mu\) are constants

\[(i = 1 \ldots m).\]

He derived the variance of \(v(zz')\), where \(v(.)\) is the short version of vec(.)

It is the aim of this note to give a compact expression for the variance and subsequently derive Satorra's result.
Matrix Tricks Related to Deleting an Observation in the General Linear Model

Markku Nurhonen and Simo Puntanen

Department of Mathematical Sciences
University of Tampere

Consider the linear model $y, X\beta, V$, where $X$ has full column rank and $V$ is positive definite. When estimating the parameters of the model, it is natural to consider the consequences of some changes or perturbations in the data on the estimates: regression diagnostics measure these consequences. One fundamental perturbation is omitting one or several of the observations from the model. In this paper our interest focuses on some helpful matrix formulas while studying regression diagnostics.
The full CS-decomposition of a partitioned orthogonal matrix

Chris C. Paige

School of Computer Science, McGill University
Montréal, Québec, Canada H3A 2A7

The CS-decomposition (CSD) of a 2-block by 2-block partitioned unitary matrix $Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$ reveals the relationships between the singular value decompositions (SVDs) of each of the 4 subblocks of $Q$. The CSD shows each SVD has the form $Q_{ij} = U_i D_{ij} V_j^H$, for $i, j = 1, 2$, where each $U_i$ and $V_j$ is unitary, and each $D_{ij}$ is essentially diagonal. Here we give a simple proof of this which has no restrictions on the dimensions of $Q_{11}$.

The CSD was originally proposed by C. Davis and W. Kahan, and is important in finding the principal angles between subspaces (Davis and Kahan, Björck and Golub), such as in computing canonical correlations between two sets of variates. It also arises in, for example, the Total Least Squares (TLS) problem. The relationships between the 4 subblock SVDs (in particular the way that each unitary matrix $U_i$ or $V_j$ appears in 2 different SVDs) has made the CSD a powerful tool for providing simple and elegant proofs of many useful results involving partitioned unitary matrices or orthogonal projectors. Here we also show the CSD makes several nice rank relations obvious, and can be used to prove some interesting results involving general nonsingular matrices.

Key words: CS decomposition, unitary matrices, rank relations.

---

1 Joint Research with Musheng Wei, Department of Mathematics, East China Normal University, Shanghai 200062, China.
How Not to Use Matrices When Teaching Statistics

D.J. Saville
NZ Pastoral Agriculture Research Institute Ltd

and

G.R. Wood
Department of Mathematics, University of Canterbury

Hidden in the appendices of some statistical texts is a geometric approach to analysis of variance and regression. This approach has been largely un-used in teaching for two reasons. First, the approach has been considered too hard, and second, this century has been a period in which algebraic methods have been dominant. It is the aim of this paper to show that the first reason offered is nonsense: the geometric approach can make things easy.

How does the geometric method make things easy? The mathematical framework necessary for the geometric development of analysis of variance and regression can be reduced to a handful of straightforward vector ideas. Once these are learnt a concrete method is followed in any situation. Three objects are isolated: the observation vector, the model space and hypothesis directions. All lie in a finite-dimensional Euclidean space. Two routine processes then serve to fit the model and test hypotheses: to fit the model we project the observation vector onto the model space, while to test hypotheses we compare averages of squared lengths of projections.

Why bother to do things differently? Conventional methods fail to convey a satisfactory understanding of the principles which unify these basic statistical methods. The cookbook approach remains mysterious, while the matrix approach is accessible only to those with mathematical maturity. The geometric approach provides an elementary but rigorous path through the material. It has been used successfully by the authors to teach second-year students in statistics as well as postgraduate students in the applied sciences. These ideas will be discussed, and an example of the method in action presented.
Problems with direct solutions of the normal equations for non-parametric additive models

M. G. Schimek

Medical Biometrics Group, University of Graz Medical Schools, A-8036 Graz, Austria

To study the functional dependence between variables of a multivariate regression problem in a data-driven fashion, non-parametric techniques like Generalized Additive Models (Buja, A., Hastie, T. and Tibshirani, R.: Linear smoothers and additive models. Ann. Statist. 1989, 17, 453-555) are very useful. Instead of solving the associated normal equations a Jacobi-type iterative procedure called backfitting is usually applied. This is primarily done for the purpose of computational efficiency and to avoid singularity problems. But little is known from a theoretical and even less from a practical point of view about the quality of the obtained results.

Let us observe \((d + 1)\)-dimensional data \((x_i, Y_i)\) with \(x_i = (x_{i1}, \ldots, x_{id})\). The \(x_{ij}, \ldots, x_{nj}\) represent independent observations drawn from a random vector \(X = (X_1, \ldots, X_d)\) and the \(x_i\) and \(Y_i\) fulfil \(Y_i = g(x_i) + \varepsilon_i\) for \(1 \leq i \leq n\), where \(g\) is an unknown smooth function from \(R^d\) to \(R\), and \(\varepsilon_1, \ldots, \varepsilon_n\) are independent errors. An additive approximation to \(g\)

\[
g(X) \approx g_0 + \sum_{j=1}^{d} g_j(X_j)
\]

is aimed at, where \(g_0\) is a constant and the \(g_j\)s are any smooth functions. To make these functions \(g_j\) identifiable it is required that \(E(g_j(X_j)) = 0\) for \(1 \leq j \leq d\). For instance we can apply the popular cubic polynomial smoothening splines. Let \(S_k = (I + \lambda_k K_k)^{-1}\) be the smoother matrix (operator) and \(K_k\) a penalty matrix of such a spline. For \(k = 1, 2, \ldots, d\) the normal equations form a \((nd) \times (nd)\) system

\[
\begin{pmatrix}
I & S_1 & S_1 & \ldots & S_1 \\
S_2 & I & S_2 & \ldots & S_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_d & S_d & S_d & \ldots & I
\end{pmatrix}
\begin{pmatrix}
g_1 \\
g_2 \\
\vdots \\
g_d
\end{pmatrix}
= 
\begin{pmatrix}
S_{1y} \\
S_{2y} \\
\vdots \\
S_{dy}
\end{pmatrix}
\]

denoted by \(Pg = Qy\), where \(P\) and \(Q\) are block matrices of smoothing operators \(S_k\).

One way to solve the system would be to apply the method of successive overrelaxation, but for \(d > 2\) we do not know how to choose the relaxation parameter. Instead we propose taking advantage of the specific block structure (especially the position of the \(Is\)) of \(P\). It allows us to derive a recursive scheme for the calculation of \(P^{-1}\). Under non-singularity a direct solution can be obtained by \(\hat{g} = P^{-1}Qy\). This approach is cheap but depends on the chosen scatterplot smoother.
In case of singularity we propose another approach based on a specific type of Tichonow regularization assuming measurement errors of the dependent variable. Let us have the singular system $Px = Qy$. Let $\alpha$ be a regularization parameter, then the disturbed system takes the form $(P^*P + \alpha I)\tilde{x} = P^*Qy$. The deviation $\|\tilde{x} - x\|$ can be estimated in terms of $\sqrt{\alpha}$ and the measurement error of the right-hand side of the equation. The system can be solved by standard techniques but is computationally expensive.

Research supported by Austrian Science Research Fund grant P8153-FKY.
On the Efficiency of a Linear Unbiased Estimator and on a Matrix Version of the Cauchy-Schwarz Inequality

George P.H. Styan

Department of Mathematics and Statistics,
McGill University
Montréal, Québec, Canada H3A 2A7

We consider the efficiency of a linear unbiased estimator in the general linear model and its connection to a determinant version of the Cauchy-Schwarz inequality; we illustrate our results with an example from simple linear regression.
Alexander Craig Aitken: 1895-1967

Garry J. Tee

Department of Mathematics and Statistics
University of Auckland

Alexander Craig Aitken was born at Dunedin on 1895 April 1, and he attended Otago Boys High School. On holiday at his grandparents dairy farm on Otago Peninsula in 1904, he discovered the now-famous breeding colony of the Royal Albatross at Taiaoa Head. His Calvinist grandparents punished him for telling such an unlikely tale but later one of his uncles was appointed as a Ranger to protect the colony. After 2 years at the University of Otago he joined the army and was severely wounded at the Battle of the Somme.

He completed his studies at the University of Otago, and graduated M.A. in 1919. In 1920 he married Mary Betts, who lectured in Botany at Otago University. He taught at Otago Boys High School until 1923, when Professor R. J. T. Bell persuaded him to study at the University of Edinburgh, where he spent the rest of his life. Professor E. T. Whittaker assigned him the problem of smoothing of data, which had practical importance in actuarial work. His thesis was of such merit that he was awarded the degree of Doctor of Science, rather than Ph.D.

Aitken's mathematical work was devoted mainly to numerical analysis, statistics and linear algebra. He founded the renowned Oliver & Boyd series of textbooks and wrote the first two himself: both Determinants and Matrices and Statistical Mathematics are recognized as classic textbooks. In numerical analysis he devised many methods which exploited the capabilities of the calculating machines which were then available, and which have proved to be fundamental to much later work in scientific computing. He gained wide fame as the greatest mental calculator for whom detailed and reliable records exist.

Whittaker retired in 1946, and Aitken, without any move on his part, was elected to the Chair of Mathematics at Edinburgh. He was elected Fellow of the Royal Societies of London and of Edinburgh, and Honorary Fellow of the Royal Society of New Zealand, of the Society of Engineers, and of the Faculty of Actuaries of Edinburgh. He was awarded Honorary Degrees by the Universities of New Zealand and of Glasgow, and he was awarded several major prizes in mathematics. He was an inspiring lecturer and an extremely gentle person, intensely devoted to music, and he wrote some poetry of distinction.

In 1963 he published his memoir Gallipoli to the Somme: Recollections of a New Zealand Infantryman, which was acclaimed as a classic account of death and life in the trenches. The Royal Society of Literature elected him as a Fellow, in recognition of his achievement in writing that memoir.

Aitken retired in 1965 (in poor health), and he died at Edinburgh on 1967 November 3, aged 72.
Use of permanents and their analogues in the representation of multivariate gamma, binomial and negative binomial distributions

D. Vere-Jones

Institute of Statistics and Operations Research
Victoria University of Wellington
Wellington, New Zealand

Powers of determinantal expressions of the form \( \det[I - TA] \), \( T = \text{diag}(t_1, t_2, ..., t_n) \) occur in expressions for the moment or probability generating functions for several types of multivariate distribution - gamma, negative binomial, binomial at least. The fundamental identity relating the logarithm of the determinant to the trace of the logarithm of the matrix allows the coefficients in the determinantal expressions to be characterised as multilinear forms in the elements of \( A \), reducing to the determinant of \( A \) when the power is +1, to its permanent when the power is -1, and to an extension of both concepts ("alpha-permanents" in the terminology of Vere-Jones (1988)), for general powers. This paper will review some properties of these representations and the associated multivariate distributions, including extensions to mixtures of binomial or negative binomial distributions, and to stochastic processes. Some open questions will be mentioned, including the problem of developing effective numerical techniques.

Reference
Andrew Balemi  
Mathematics and Statistics  
University of Auckland  
38 Princes St  
Private Bag 92019  
Auckland  
New Zealand  
balemi@mat.auckuni.ac.nz

Richard William Farebrother  
Department of Econometrics and Social Statistics  
University of Manchester  
Manchester  
M13 9FL  
United Kingdom  
mrbsrf@cms.mcc.ac.uk

Rod Ball  
Hort Research  
Hort Research Institute NZ  
P.O. Box 92169  
Auckland  
New Zealand  
rod@marc.am.nz

Arthur R. Gilmour  
NSW Agriculture Centre  
Forest Road  
Orange  
NSW 2800  
Australia

John S. Chipman  
Department of Economics  
University of Minnesota  
1035 Management and Economics  
271 19th Avenue South  
Minneapolis, MN 55455  
U.S.A.  
jchipman@uminn.bitnet

Harold V. Henderson  
Statistics and Computing Centre  
Ruakura Agriculture Centre  
Hamilton  
New Zealand  
henderson@ruakura.cri.nz

Ronald Christensen  
Department of Mathematics and Statistics  
University of New Mexico  
Albuquerque  
NM 87131  
U.S.A.  
fletcher@gopher.unm.edu OR

Jeffrey J. Hunter  
Department of Mathematics & Statistics  
Massey University  
Palmerston North  
New Zealand  
jhunter@massey.ac.nz

Peter Clifford  
Oxford University  
Jesus College  
Oxford  
OX1 3DW  
United Kingdom  
clifford@vax.oxford.ac.uk

Snehalata Huzurbazar  
Department of Statistics  
University of Georgia  
Athens  
GA 30602-1952  
U.S.A  
lata@rolf.stat.uga.edu