Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 50 (2006) 1–4
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Editorial

2nd Special issue on matrix computations and statistics

A number of interesting and important ideas have resulted from the relationship between matrix computations and statistics. Well-known examples include the solution of least-squares problems, computation of the singular value decomposition and its generalizations, estimation of principal components, computation of canonical correlations, several cluster analysis algorithms, and the solution of total least-squares problems.

Computational statistics and data analysis has published papers in matrix computations and statistics (for instance, Doornik and Ooms, 2003; Eldén, 2004; Foschi et al., 2003; Kawai, 2004; Molinari et al., 2004; Pollock, 2003; Wilkinson and Yeung, 2004) and encourages submissions in this area. A previous special issue on this area featured papers on multidimensional scaling, an application to web search engines, an algorithm for seemingly unrelated regression models, an error measurement model for motion analysis, and a survey on alternating least-squares problems (Chang and Paige, 2002; Dhillon et al., 2002; Doornik and O'Brien, 2002; Erbay et al., 2002; Fodor and Kamath, 2002; Foschi and Kontoghiorghes, 2002; Genton and Gorsich, 2002; Georgiou et al., 2002; He et al., 2002; Kiers, 2002; Kukush et al., 2002; Malone et al., 2002; Vigneau and Qannari, 2002).

This second special issue on matrix computations and statistics continues in this direction. It contains topics such as new data analysis models, fast computational methods, and theoretical developments using matrix computations in statistics: a new model for principal components of binary data, high-speed smoothing for large grids, a theoretical and practical comparison of canonical correlation analysis and Procrustes analysis, a generalization of constrained correspondence analysis, a new generalization of total least squares, a study of the conditions for uniqueness of estimates in three-way models, a study of the use of the modified Leverrier–Faddeev algorithm to the spectral decomposition of symmetric block-circulant matrices, and a sensitivity analysis of the Strain criterion for multidimensional scaling.

Several papers developed matrix computational algorithms that were inspired by statistical applications. Li (2004) discusses the definition and application of sign eigenvectors. Sign eigenanalysis can be applied to the development of statistical inference procedures in the ℓ_1 norm. Dax (2004) presents a method for solving a system of linear inequalities $A\mathbf{x} \geqslant \mathbf{b}$. This problem is a ℓ_1 minimization problem or ℓ_1 regression problem. Gower (2004)

0167-9473/$ - see front matter
doi:10.1016/j.csda.2004.08.003

applies the Leverrier–Faddeev algorithm for deriving the algebraic structure of matrices to find the singular value decomposition of block-circulant matrices. Chang (2004) proposes a method for the efficient computation of Huber's M-estimator. Updating/downdating methods for matrix factorizations and special structure of the problem are exploited to make the algorithm efficient.

Aspects of dimension reduction are considered in four other papers in this volume. Trendafilov and Joliffe (2004) consider a projected gradient approach to a principal component problem where the principal components satisfy a one-norm constraint. Li and Zha (2004) use a mixture of Poisson distributions to model document classification and word clustering in term-document models. The parameters are regularized by imposing a clustering strategy on the set of words. Yamada et al. (2004) discuss mixed-level supersaturated design. This is a first step in reducing the number of factors in a large experiment. De Leeuw (2004) combines ideas from latent structure analysis with principal component analysis and multiple component analysis for binary data. The algorithms they propose can be fitted efficiently to large matrices.

Canonical correlations were among the themes of three papers. Gardner et al. (2004) synthesize canonical variate analysis, Procrustes analysis, and generalized canonical correlation. Using a data mining example consisting of 1425 samples, they isolate a four class structure using generalized canonical correlation, and use its synthesis with canonical variate analysis to discern the structure of the data in more detail. Takhane et al. (2004) develop a method for generalized constrained canonical correlation analysis based upon two orthogonal decomposition of projectors. This work improves upon previous work by two of the authors on a related technique.

Two papers concern scaling in a model and lead to solving an optimization problem. Lewis and Trosset (2004) offer a perturbation analysis of the strain criterion of multidimensional scaling. Their analysis builds gradients and Hessians of the appropriate objective functions from eigenvalue perturbation theory. Markovsky et al. (2004) develop of version of the total least squares (or errors-in-variables) model. This model allows for more general covariance structures for incorporating certain kinds of scaling in the total least-squares problem.

A final set of papers concerns the multidimensional representation of data. Eilers et al. (2004) present what they call a P-spline approach to multidimensional grids. A smooth multidimensional surface is constructed out of tensor products of B-splines. They approach this problem in a manner that reduces the memory required and works for any number of dimensions. Stegeman and Ten Berge (2004) analyze the Condecomp/Parafac method for component analysis of three-way arrays. They disprove a conjecture about this approach. Evangelaras and Koukouvinos (2004) show how the theory of Gröbner bases allow one to identify estimable effects of the factors in an experimental design.

As can be seen from this special issue and the papers that CSDA has already published in this area, the interaction between matrix computations and statistics will continue to be a rich and intellectually challenging area of research.

References

Chang, X.-W., 2004. Computation of Huber's m-estimates for a block-angular regression problem. Comput. Statist. Data Anal. (this issue).

Chang, X.-W., Paige, C.C., 2002. Numerical linear algebra in the integrity theory of the global positioning system. Comput. Statist. Data Anal. 41 (1), 123–142.
Dax, A., 2004. The ℓ_1 solution of linear inequalities. Comput. Statist. Data Anal. (this issue).
De Leeuw, J., 2004. Principal component analysis of binary data by iterated singular value decomposition. Comput. Statist. Data Anal. (this issue).
Dhillon, I.S., Modha, D.S., Spangler, W.S., 2002. Class visualization of high-dimensional data with applications. Comput. Statist. Data Anal. 41 (1), 59–90.
Doornik, J.A., O'Brien, R.J., 2002. Numerically stable cointegration analysis. Computat. Statist. Data Anal. 41 (1), 185–193.
Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. Comput. Statist. Data Anal. 42 (3), 333–348.
Eilers, P.H.C., Currie, I.D., Durbán, M., 2004. Fast and compact smoothing on large multidimensional grids. Comput. Statist. Data Anal. (this issue).
Eldén, L., 2004. Partial least-squares vs. Lanczos bidiagonalization I: analysis of a projection method for multiple regression. Comput. Statist. Data Anal. 46 (1), 11–31.
Erbay, H., Barlow, J.L., Zhang, Z., 2002. A modified Gram–Schmidt-based downdating technique for ulv decompositions with applications to recursive tls problems. Comput. Statist. Data Anal. 41 (1), 195–209.
Evangelaras, H., Koukouvinos, C., 2004. A comparison between the Gröbner bases approach and hidden projection properties in factoral designs. Comput. Statist. Data Anal. (this issue).
Fodor, I.K., Kamath, C., 2002. Dimension reduction techniques and the classification of bent double galaxies. Comput. Statist. Data Anal. 41 (1), 91–122.
Foschi, P., Belsley, D.A., Kontoghiorghes, E.J., 2003. A comparative study of algorithms for solving seemingly unrelated regressions models. Comput. Statist. Data Anal. 44 (1–2), 3–35.
Foschi, P., Kontoghiorghes, E.J., 2002. Seemingly unrelated regression model with unequal size observations: computational aspects. Comput. Statist. Data Anal. 41 (1), 211–229.
Gardner, S., Gower, J.C., le Roux, N.J., 2004. A synthesis of canonical variate analysis, generalized canonical correlation, and procrustes analysis. Comput. Statist. Data Anal. (this issue).
Genton, M.G., Gorsich, D.J., 2002. Nonparametric variogram and covariogram estimation with Fourier–Bessel matrices. Comput. Statist. Data Anal. 41 (1), 47–57.
Georgiou, S., Koukouvinos, C., Stylianou, S., 2002. On good matrices, skew Hadamard matrices and optimal designs. Comput. Statist. Data Anal. 41 (1), 171–184.
Gower, J.C., 2004. An application of the modified Leverrier–Faddeev algorithm to the spectral decomposition of symmetric block-circulant matrices. Comput. Statist. Data Anal. (this issue).
He, X., Zha, H., Ding, C.H.Q., Simon, H.D., 2002. Web document clustering using hyperlink structures. Comput. Statist. Data Anal. 41 (1), 19–45.
Kawai, S., 2004. Approximation of some moments in eigenvalue problem including some random variables and its application. Comput. Statist. Data Anal. 46 (2), 201–212.
Kiers, H.A.L., 2002. Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. Comput. Statist. Data Anal. 41 (1), 157–170.
Kukush, A., Markovsky, I., Van Huffel, S., 2002. Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. Comput. Statist. Data Anal. 41 (1), 3–18.
Lewis, R.M., Trosset, M.W., 2004. Sensitivity analysis of the strain criterion for multidimensional scaling. Comput. Statist. Data Anal. (this issue).
Li, B., 2004. Sign eigenanalysis and its applications to optimization and robust statistics. Comput. Statist. Data Anal. (this issue).
Li, J., Zha, H., 2004. Two-way Poisson mixture models for simultaneous document classifications and word clustering. Comput. Statist. Data Anal. (this issue).
Malone, S.W., Tarazaga, P., Trosset, M.W., 2002. Better initial configurations for metric multidimensional scaling. Comput. Statist. Data Anal. 41 (1), 143–156.
Markovsky, I., Rastello, M.L., Premoli, A., Kukush, A., Van Huffel, S., 2004. The element-wise weighted total least squares problem. Comput. Statist. Data Anal. (this issue).
Molinari, N., Durand, J.-F., Sabatier, R., 2004. Bounded optimal knots for regression splines. Comput. Statist. Data Anal. 45 (2), 159–178.
Pollock, D.S.G., 2003. Recursive estimation in econometrics. Comput. Statist. Data Anal. 44 (1–2), 37–75.

Stegeman, A., Ten Berge, J.M.F., 2004. Kruskal's condition for uniqueness in Candecomp/Parafac when ranks and *k*-ranks coincide. Comput. Statist. Data Anal. (this issue).

Takhane, Y., Yanai, H., Hwang, H., 2004. An improved method for generalized constrained canonical correlation analysis. Comput. Statist. Data Anal. (this issue).

Trendafilov, N., Joliffe, I., 2004. Projected gradient approach to the numerical solution of the SCoTLASS. Comput. Statist. Data Anal. (this issue).

Vigneau, E., Qannari, E., 2002. A new algorithm for latent root regression analysis. Comput. Statist. Data Anal. 41 (1), 231–242.

Wilkinson, D.J., Yeung, S.K.-H., 2004. A sparse matrix approach to bayesian computation in large linear models. Comput. Statist. Data Anal. 44 (3), 493–516.

Yamada, S., Matsui, M., Lin, D.K.J., Takahashi, T., 2004. A general construction method for mixed-level supersaturated design. Comput. Statist. Data Anal. (this issue).

Jesse L. Barlow[a],
Haesun Park[b],
Patrick J.F. Groenen[c],
Hongyuan Zha[a]

[a] *Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802-6822, USA*

[b] *4-192 EE/CS, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA*

[c] *Econometric Institute, Erasmus University Rotterdam, Room H11.23, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

E-mail addresses: barlow@cse.psu.edu (J.L. Barlow), hpark@cs.umn.edu (H. Park), groenen@few.eur.nl (P.J.F. Groenen), zha@cse.psu.edu (H. Zha).