Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2007) 1076–1079
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Editorial

Total Least Squares and Errors-in-variables Modeling

The main purpose of this special issue is to present an overview of the progress of a modeling technique which is known as total least squares (TLS) in computational mathematics and engineering, and as errors-in-variables (EIV) modeling or orthogonal regression in the statistical community. The TLS method is one of several linear parameter estimation techniques that has been devised to compensate for data errors. The basic motivation is the following: let a set of multidimensional data points (vectors) be given. How can one obtain a linear model that explains these data? The idea is to modify all data points in such a way that some norm of the modification is minimized subject to the constraint that the modified vectors satisfy a linear relation. Although the name "TLS" appeared in the literature only 27 years (Golub and Van Loan, 1980) ago, this method of fitting is certainly not new and has a long history in the statistical literature, where the method is known as "orthogonal regression", "EIV regression" or "measurement error (ME) modeling". The univariate line fitting problem was already discussed since 1877 (Adcock, 1877). More recently, the TLS approach to fitting has also stimulated interests outside statistics. One of the main reasons for its popularity is the availability of efficient and numerically robust algorithms in which the singular value decomposition (SVD) plays a prominent role (Golub and Van Loan, 1980). Another reason is the fact that TLS is an application oriented procedure. It is suited for situations in which all data are corrupted by noise, which is almost always the case in engineering applications (Van Huffel et al., 2007). In this sense, TLS and EIV modeling are a powerful extension of classical least squares and ordinary regression, which corresponds only to a partial modification of the data.

The problem of *linear parameter estimation* arises in a broad class of scientific disciplines such as signal processing, automatic control, system theory and in general engineering, statistics, physics, economics, biology, medicine, etc. It starts from a model described by a linear equation:

$$\xi_1\beta_1 + \cdots + \xi_p\beta_p = \eta, \tag{1}$$

where $\xi_1, \ldots, \xi_p$ and η denote the variables and the p-dimensional vector $\beta=[\beta_1, \ldots, \beta_p]^{\mathrm{T}}$ plays the role of a parameter vector that characterizes the specific system. A basic problem of applied mathematics is to determine an estimate of the *true* but *unknown* parameters from certain measurements of the variables. This gives rise to an *overdetermined* set of n linear equations ($n > p$):

$$X\beta \approx y, \tag{2}$$

where the ith row of the $n \times p$ data matrix X and n-dimensional vector y contain, respectively, the measurements of the variables $\xi_1, \ldots, \xi_p$ and η.

In the classical least squares approach, as commonly used in ordinary regression, the measurements X of the variables ξ_i are assumed to be free of error and hence, all errors are confined to the observation vector y. However, this assumption is frequently unrealistic: sampling errors, human errors, modeling errors and instrument errors may imply inaccuracies of the data matrix X as well. One way to take errors in X into account is to introduce *perturbations also in X*. Therefore, the TLS method was introduced in the field of computational mathematics (Golub, 1973; Golub and Van Loan, 1980) as a numerical linear algebra tool for finding approximate solutions to overdetermined systems of equations $X\beta \approx y$, where both the vector y as well as the matrix X are assumed to be perturbed. Since its definition by Golub and Van Loan (1980), the classical TLS method has been extended to solve weighted, structured, and regularized TLS problems

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.07.001

and was applied in signal processing, system identification, computer vision, document retrieval, computer algebra, and other fields, see, e.g. Markovsky et al. (2006).

EIV models, also known as ME models, are an alternative to the classical regression model in statistics when both the dependent as well as the independent variables are subject to errors. EIV models are closely related to TLS methods but consider less restrictive assumptions on the error distributions, see, e.g. Kukush et al. (2002, 2004). The former provide statistical analysis, derive consistent estimators and provide possible justification for the deterministic approximation criteria used in the latter.

A comprehensive description of the state of the art on TLS from its conception up to the summer of 1990 and its use in parameter estimation has been presented in Van Huffel and Vandewalle (1991). While the latter book is entirely devoted to TLS, a second (Van Huffel, 1997) and third book (Van Huffel and Lemmerling, 2002) present the progress in TLS and in the broader field of EIV modeling, respectively, from 1990 to 1996 and from 1996 till 2001. A recent overview of weighted and structured TLS problems, solution methods, and applications is given in Markovsky and Van Huffel (2007). This paper presents the TLS modeling problem from a novel low-rank approximation point of view instead of the classical one, i.e., the approximate solution of an overdetermined system of equations $X\beta \approx y$.

In this special issue, we are aiming at the synergy of statistics and computations that provide better computational methods for statistically meaningful estimators.

The following four papers present new ideas on how to robustify linear ME modeling. Watson (2007) investigates robust counterparts of linear data fitting problems with uncertain data which lie in a given uncertainty set and shows how the original problem can be replaced by a convex optimization problem in fewer variables for which standard software exists. Lampe and Voss (2007) considerably improved a computational approach for solving regularized TLS problems, based on a sequence of quadratic eigenvalue problems and presented recently in Sima et al. (2004), by reusing information from the previous quadratic problems and early updates in a nonlinear Arnoldi method. Truncation methods are another class of methods for regularizing ill-posed linear problems in the presence of MEs. In essence, they aim to limit the contribution of noise on rounding errors by cutting off a certain number of terms in an expansion such as the SVD. Sima and Van Huffel (2007) show that the filter factors associated with the truncated TLS solution provide more information for choosing the truncation level compared to truncated SVD and illustrate the merits of a modified generalized cross validation method. Mastronardi and O'Leary (2007) present new ideas on fast robust regression for Toeplitz-like structured problems. In particular, they show how to include regularization and efficiently compute the regularized solution if the data matrix has low displacement rank.

New extensions in ME modeling, including consistent estimators, are described in the following papers. De Castro et al. (2007) investigate functional heteroscedastic ME models. Methods of local influence are used to assess the effects of perturbation of data on some inferential procedure. Schneeweiss and Shalabh (2007) derive three consistent estimators in a linear ultrastructural model with not-necessarily normally distributed measurement errors and analyze their efficiency properties. Shalabh et al. (2007) further extend these models to include available prior knowledge about the regression coefficients in the form of exact linear restrictions. Some consistent estimators are presented and their asymptotic properties are analyzed. Kukush et al. (2007) derive a consistent estimate for an extended EIV model $X\beta = Y$ with finite dependent rows clustered into two groups with essentially different second order empirical moments, in which the total error covariance structure of the data matrix $D = [X\ Y]$ is known up to two scalar factors. Since Toeplitz/Hankel structure is allowed, the results are applicable to system identification with a change point in the input data.

In addition, Vehkalahti et al. (2007) deal with the problem of selection of predictors in a multivariate model in the presence of ME in the data. These errors affect the predictor selection and can be measured using various measurement scales. Factor scores are shown to be the best choice for predictor selection. Barros et al. (2007) deal with restricted and unrestricted testing a structural measurement error model. In particular, they aim to consider hypothesis testing for testing the equality of slopes in ME models, including the additional assumption that the slopes lie in a closed interval.

Furthermore, attention is focused to geometric fitting. Kanatani and Sugaya (2007) compare the convergence performance of four typical numerical algorithms for fitting algebraic–geometric objects to data, called here geometric fitting and yet another application area for EIV modeling, as arising in computer vision applications. Petras and Podlubny (2007) provide a brief comparison of two methods for fitting linear manifolds (lines, planes, etc.) to data, in essence ordinary least squares versus orthogonal (geometric) TLS, followed by an interesting example for econometrics.

Finally, Ramos (2007) presents a nice overview of the applications of TLS and related techniques in environmental sciences. Several important applications like rainfall-runoff modeling, specific time series in the mentioned domain, chirp signal modeling, etc. are addressed.

The guest editors would like to acknowledge the help of Ivan Markovsky (Dept. of EE, K.U. Leuven, Belgium till January 2007 and from then on School of Electronics and Computer Science, Univ. of Southampton, UK) and Ida Tassens (Department of EE, K.U. Leuven, Belgium) in organizing this special issue. Special thanks also go to the referees for their continuous efforts in the reviewing process. Finally, the guest editors would like to express their sincere thanks to the Co-Editor Erricos John Kontoghiorghes.

References

Adcock, R.J., 1877. A problem in least squares. Analyst 4, 183–184.

Barros, M., Giampaoli, V., Lima, C.R.O.P., 2007. Hypothesis testing in the unrestricted and restricted parametric spaces of structural models. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.022.

De Castro, M., Galea-Rojas, M., Bolfarine, H., 2007. Local influence assessment in heteroscedastic measurement error models. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.012.

Golub, G.H., 1973. Some modified matrix eigenvalue problems. SIAM Rev. 15, 318–344.

Golub, G.H., Van Loan, C.F., 1980. An analysis of the total least squares problem. SIAM J. Numer. Anal. 17, 883–893.

Kanatani, K., Sugaya, Y., 2007. Performance evaluation of iterative geometric fitting algorithms. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.013.

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.

Kukush, A., Markovsky, I., Van Huffel, S., 2004. Consistent estimation in an implicit quadratic measurement error model. Comput. Statist. Data Anal. 47 (1), 123–147.

Kukush, A., Markovsky, I., Van Huffel, S., 2007. Estimation in a linear multivariate measurement error model with a change point in the data. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.06.010.

Lampe, J., Voss, H., 2007. On a quadratic eigenproblem occurring in regularized total least squares. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.020.

Markovsky, I., Van Huffel, S., 2007. Overview of total least squares methods. Signal Process. 87, 2283–2302.

Markovsky, I., Rastello, M.-L., Premoli, A., Kukush, A., Van Huffel S., 2006. The element-wise weighted total least squares problem. Comput. Statist. Data Anal. 50 (1), 181–209.

Mastronardi, N., O'Leary, D.P., 2007. Robust regression and l_1 approximations for Toeplitz problems. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.008.

Petras, I., Podlubny, I., 2007. State space description of national economies: the V4 countries. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.014.

Ramos, J., 2007. Applications of TLS and related methods in the environmental sciences. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.06.009.

Sima, D.M., Van Huffel, S., 2007. Level choice in truncated total least squares. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.015.

Schneeweiss, H., Shalabh, 2007. On the estimation of the linear relation when the error variances are known. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.06.018.

Shalabh, Garg, G., Misra, N., 2007. Restricted regression estimation in measurement error models. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.011.

Sima, D.M., Van Huffel, S., Golub, G.H., 2004. Regularized total least Squares based on quadratic eigenvalue problem solvers. BIT Numer. Math. 44 (4), 793–812.

Van Huffel, S. (Ed.), 1997. Recent advances in total least squares techniques and errors-in-variables modeling, SIAM Proceedings Series, SIAM, Philadelphia.

Van Huffel, S., Lemmerling P. (Eds.), 2002. Total Least Squares and Errors-in-variables Modeling: Analysis, Algorithms and Applications, Kluwer Academic Publishers, Dordrecht.

Van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problem: Computational Aspects and Analysis. SIAM, Philadelphia.

Van Huffel S., Markovsky I., Vaccaro R. J., Söderström, T. (Guest Eds.), 2007. Total Least Squares and Errors-in-Variables Modeling. Signal Process. 87(10), 2281–2490 (Special Issue).

Vehkalahti, K., Puntanen, S., Tarkkonen, L., 2007. Effects of measurement errors in predictor selection of linear regression model. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.005.

Watson, G.A., 2007. Robust counterparts of errors-in-variables problems. Comput. Statist. Data Anal. doi:10.1016/j.csda.2007.05.006.

Sabine Van Huffel
Department of Electrical Engineering,
Katholieke Universiteit Leuven,
Kasteelpark Arenberg 10, B–3001 Leuven, Belgium
E-mail address: Sabine.Vanhuffel@esat.kuleuven.be

Chi-Lun Cheng
Institute of Statistical Science,
Academica Sinica,
Taipei, Taiwan, ROC
*E-mail address:*clcheng@stat.sinica.edu.tw

Nicola Mastronardi
Istituto per le Applicazioni del Calcolo "M. Picone" sez. Bari,
National Research Council of Italy,
via G. Amendola 122/D, I-70126 Bari, Italy
*E-mail address:*n.mastronardi@ba.iac.cnr.it

Chris Paige
McGill University,
School of Computer Science,
3480 University Street, Montreal, PQ, Canada H3A 2A7
*E-mail address:*chris@cs.mcgill.ca

Alexander Kukush
Kyiv National Taras Shevchenko University,
Volodymyrska st. 60, 01033 Kyiv, Ukraine
*E-mail address:*alexander_kukush@mail.univ.kiev.ua