Proceedings 10th IFAC Symposium on System Identification,
Copenhagen, Denmark, Vol. 3,k July 1994, pp. 159-164

94-19C

DYNAMIC TOTAL LINEAR LEAST SQUARES

BART DE MOOR

ESAT, Electrical Engineering Department, Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium
Tel: 32/16/220931 ; Fax : 32/16/221855
email: bart.demoor@esat.kuleuven.ac.be

Abstract: The relationships between linear least squares (equation error), total linear least squares (errors-in-variables) and the Frisch scheme are well understood for the static case. For the dynamic case, much work has been done to explore similar relationships. What seemed to be lacking so far is the solution to the so-called dynamic total least squares problem, which itself is a special case of a structured total least squares problem. In this paper, we present a survey of some recent results for this dynamic total least squares problem (which in fact corresponds to the L_2-optimal modelling of linear SISO systems). We discuss the most relevant properties and make some suggestions on how these results might be useful in the context of the Frisch scheme.

Key Words. Estimation Theory, System Identification, Frisch Scheme, errors-in-variables, total linear least squares. Riemannian singular value decomposition

1. INTRODUCTION [2]

For the static case, the relationships between least squares (equation-error in the L_2 norm), total least squares (errors-in-variables in the L_2 norm) [14] and the Frisch scheme are well understood by now (see e.g. [5] [6] [7]). The least squares solutions play a fundamental role in the construction of the solution set of the Frisch scheme, at least in the case where the inverse of the data covariance matrix is sign-similar to an elementwise positive matrix [15] [16] [17]. In this case, the solution of the Frisch scheme in the so-called solution space (the null space of the data covariance matrix) can be represented as a polyhedral cone, the vertices of which are the least squares solutions. By appropriate sign changes, these can all be 'reflected' to lie in the positive orthant (we refer to [10] for details). In this case, the total least squares solution is a convex combination (a linear combination with positive weights) of the least squares solutions. It corresponds to the Perron-Frobenius eigenvector of the inverse data covariance matrix and is contained in the polyhedral cone of the least squares solutions [7]. Other relations between the least squares and total least squares solutions are explored in [7]. In the case where the inverse of the data covariance matrix is not sign-similar to an elementwise positive matrix, the geometry of the solution set is much more involved, as demonstrated in [2].

In the dynamic case, when the model is taken to be a SISO linear time-invariant system, the relationships between least squares methods (e.g. equation error in the L_2 norm) and linear dynamic errors-in-variables models (as e.g. explored in [1] [3]) are far less understood.

As a matter of fact, up to recently, there was no known solution to the so-called *dynamic total least squares problem*, which is the following (formulated and solved here for SISO systems):

Let $w_k \in \mathbb{R}^2, k = 0, \ldots, N$ be a given vector sequence of data where the scalar sequences u_k and y_k are defined as [3] $w_k = \begin{pmatrix} u_k \\ y_k \end{pmatrix}$. Find approxi-

[2]The following text presents research results obtained within the framework of the Belgian programme on interuniversity attraction poles (IUAP-17 and IUAP-50) initiated by the Belgian State - Prime Minister's Office - Science Policy Programming. The scientific responsibility is assumed by its author, who is a Senior Research Associate of the Belgian National Fund for Scientific Research. A preliminary version of this text was available as ESAT-SISTA TR-1994-19, Dept. of EE, K.U.Leuven, March 1994.

[3]One could consider the first component of w_k to be an input sequence u_k and the second component to be an output sequence y_k or the other way around.

mations v_k of the u_k, and z_k of the y_k, such that v_k and z_k are related by a linear model of given order n with real coefficients:

$$\sum_{i=0}^{n} \alpha_i z_{k-i} + \sum_{i=0}^{n} \beta_i v_{k-i} = 0 \ , \ k = n, \ldots, N \ , \quad (1)$$

and v_k and z_k minimize

$$J = \sum_{k=0}^{N} [(y_k - z_k)^2 h_k + (u_k - v_k)^2 g_k] \ , \quad (2)$$

where g_k and $h_k, k = 0, \ldots, N$ are given positive weights, subject to

$$\sum_{i=0}^{n} \alpha_i^2 + \sum_{i=0}^{n} \beta_i^2 = 1 \ . \quad (3)$$

It is assumed that $N \geq 3n + 1$ to make constraint (1) meaningful.

The minimization problem (2) with constraints (1) and (3) will be called the *dynamic total linear least squares problem*.
Let us observe that all the observations are treated 'symmetrically' in the sense that not only the outputs are modified, as is the case in equation error methods, but that also the inputs are modified. The weights g_k and h_k are god-given or specified by the user. In the case that $g_k = h_k = 1, \forall k$, the formulation corresponds to the statistical errors-in-variables problem, which is maximum likelihood when the input-output data of an unknown linear system are corrupted by additive white Gaussian noise and one wants to find from the observations the difference equation that models the system.

It is the purpose of this paper to present the solution to the dynamic total linear least squares problem and some of its properties. We hope that this contribution will clarify some of the relations in the *terra incognita* between least squares, total least squares and the Frisch scheme (errors-in-variables) for dynamic systems. We will not give proofs in this paper. But details can be found in [10] [11] [12].
This paper is organized as follows: In Section 2, we briefly recall the static total linear least squares problem and its solution via the singular value decomposition (SVD). We then show how the dynamic total least squares problem is a special case of a structured total least squares problem, and hence its solution can be obtained via a so-called Riemannian singular value decomposition (which is one of the major results of [10]). In Section 3, it is shown how the given data sequence can be decomposed into two sequences, the L_2-approximations and the residuals, which are orthogonal to each other (in diagonal inner products derived from given weights). In Section 4, it is shown how certain block Hankel matrices $\widetilde{W}_i$ constructed from the weighted residuals are always non-singular and hence, the residuals themselves can be considered as being generated by a linear system. In Section 5, we discuss how one can find completions $\widehat{W}_i^{com}$ of the block Hankel matrices $\widehat{W}_i$ with the L_2-approximations, such that $\widehat{W}_i^{com} \widehat{W}_i^T = 0$.

We will use the notations

$$w_k = \begin{pmatrix} u_k \\ y_k \end{pmatrix} \ , \ \hat{w}_k = \begin{pmatrix} v_k \\ z_k \end{pmatrix} ,$$

and

$$\tilde{w}_k = \begin{pmatrix} (u_k - v_k)\, g_k \\ (y_k - z_k)\, h_k \end{pmatrix} ,$$

for the given data, the optimal approximations and the weighted residuals respectively. By $\widehat{W}_i$ we denote the $2i \times (N+i)$ block Hankel matrix constructed from the sequence $\hat{w}_k$:

$$\widehat{W}_i = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 & \hat{w}_0 & \hat{w}_1 & \ldots & \hat{w}_{N-1} & \hat{w}_N \\ 0 & 0 & \ldots & 0 & \hat{w}_0 & \hat{w}_1 & \ldots & \hat{w}_{N-1} & \hat{w}_N & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \hat{w}_0 & \ldots & \hat{w}_{N-1} & \hat{w}_N & 0 & 0 & \ldots & 0 & 0 \\ \hat{w}_0 & \hat{w}_1 & \ldots & \hat{w}_N & 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} . \quad (4)$$

Here, i is a user-defined integer which is assumed to be larger than n (the order in the difference equation (1)). The block Hankel matrix $\widetilde{W}_i$ is constructed similarly from the weighted residuals $\tilde{w}_k$. Note that the first block row has $i-1$ leading zero vectors. We'll also use the data sequences $w, \hat{w}$ and $\tilde{w}$, all of which are in $\mathbb{R}^{(N+1)\times 2}$, where

$$w^T = (\, w_0 \; w_1 \; \ldots \; w_{N-1} \; w_N \,) \ ,$$

and $\hat{w}$ and $\tilde{w}$ are constructed similarly. The sequences u, y, v and z, all in $\mathbb{R}^{(N+1)}$, are defined as:

$$\begin{aligned} u^T &= (\, u_0 \, u_1 \, \ldots u_{N-1} \, u_N) \ , \\ y^T &= (\, y_0 \, y_1 \, \ldots y_{N-1} \, y_N) \ , \\ v^T &= (\, v_0 \, v_1 \, \ldots v_{N-1} \, v_N) \ , \\ z^T &= (\, z_0 \, z_1 \, \ldots z_{N-1} \, z_N) \ . \end{aligned}$$

The vectors a and b contain the coefficients α_i and β_i of the linear model. The transfer function associated to the difference equation (1) will be denoted by $b(z)/a(z)$.

2. SOLUTION VIA A RIEMANNIAN SVD

The static total least squares problem for a given data matrix $A \in \mathbb{R}^{p\times q}$ can be formulated as

$$\min_{B\,,\,y} \|A - B\| \text{ subject to } \begin{matrix} By = 0 \ , \\ y^T y = 1 \ . \end{matrix}$$

The two constraints ensure the rank deficiency of the approximating matrix B. As is well known, the solution can be calculated via the 'smallest' singular triplet of A (see [14], [21]), i.e. the triplet (u, σ, v) corresponding to the smallest singular value σ, which satisfies

$$\begin{aligned} Av &= u\sigma , \qquad & u^T u &= 1 , \\ A^T u &= v\sigma , \qquad & v^T v &= 1 . \end{aligned} \qquad (5)$$

For σ the smallest singular value of A, the matrix B is given by a rank one update of A as

$$B = A - u\sigma v^T .$$

It turns out (see [10] [11] [13]) that the solution of the dynamic total least squares problem is given by the following Theorem:

Theorem 1
The vectors a and b that contain the coefficients of the difference equation (1) which is the solution of the dynamic total least squares problem, follow from the 'smallest' singular triplet of a generalized ('nonlinear') singular value decomposition of the form

$$\begin{aligned} \begin{pmatrix} Y & U \end{pmatrix} \begin{pmatrix} \overline{a} \\ \overline{b} \end{pmatrix} &= (D_{\overline{a}} + D_{\overline{b}})\, u\, \tau , \\ \begin{pmatrix} Y^T \\ U^T \end{pmatrix} u &= D_u \begin{pmatrix} \overline{a} \\ \overline{b} \end{pmatrix} \tau , \end{aligned}$$

in which Y and U are Hankel matrices of dimension $(N - n + 1) \times (n + 1)$ that are built up with the output and input data; $D_{\overline{a}}$, $D_{\overline{b}}$ and D_u are positive definite matrices, the elements of which are certain quadratic functions of the components of a, b resp.u. The vectors u and $(\overline{a}^T \ \overline{b}^T)^T$ are normalized such that

$$u^T(D_{\overline{a}} + D_{\overline{b}})u = 1 \ \text{ and } \ (\overline{a}^T \ \overline{b}^T) D_u \begin{pmatrix} \overline{a} \\ \overline{b} \end{pmatrix} = 1 .$$

The minimum value of the object function (2) is given by the smallest singular value τ. The vectors a and b from the difference equation (1) can be obtained from a simple scaling of $\overline{a}$ and $\overline{b}$ so that they satisfy (3). The data sequences z and v can be obtained from the smallest singular triplet and the original data (see [8] for detailed formulas).

Due to space limitations, we can not provide the full details here (for which we refer to [10] [11] [12] [13]). But let us make the following remarks:

- First note the ressemblance between the SVD for the static case in (5) and the 'generalized' SVD in Theorem 1. Both are in terms of the given data (A in the static case and the matrix $(Y \quad U)$ in the dynamic case). As a matter of fact, the SVD of Theorem 1 would be a well-known generalized SVD (the restricted SVD, see [9]) in case that D_u, $D_{\overline{a}}$ and $D_{\overline{b}}$ would be constant matrices. Because of the fact that, on the one hand these matrices are not constant as they are a function of the singular vectors to be found, and because they are always positive definite, we propose to call the generalized SVD of Theorem 1 a *Riemannian singular value decomposition* (see [13]).
- In [10] we demonstrate that the dynamic total least squares problem is a special case of the more general *structured* total least squares problem (STLS), which can be solved via a Riemannian SVD. The main conclusion of Theorem 1 is that the transfer function $b(z)/a(z)$, associated with the optimal model (1) that solves the dynamic total least squares problem, can be obtained from a Riemannian SVD.
- The intermediate steps in the proof of this result [10], which is obtained via the technique of Lagrange multipliers, are instrumental in the derivation of the properties of the dynamic total least squares solution to be stated below. To mention just one property: The difference of the Hankel matrices Z and V, which contain the 'modified' output and input data of equation (1), and the Hankel matrices Y and Y is a multilinear function of the singular triplet $(u, \tau, (\overline{a}^T \ \overline{b}^T)^T)$ Recall that in the static case this is similar as the difference $A - B$ is a rank one matrix. In the dynamic case, the rank is however larger than one.
- A heuristic algorithm which is inspired by the method of inverse iteration is described in [10] [11] (which also contains much more details about the derivation and other additional properties). In [13] we will be describing a continuous time method (a gradient flow) which employs some ideas from differential geometry.

We will now turn to the enumeration of some of the properties that are satisfied by a solution to the dynamic total least squares problem.

3. ORTHOGONALITY OF $\hat{w}$ AND $\tilde{w}$

When in optimization problems, a criterion to be optimized is a sum of squares, *orthogonality* is never far away. For instance, for the static total least squares problem, there is a property of orthogonality of the residuals and the data in the approximation matrix B, as the column and row spaces of both matrices are perpendicular:

$$(A - B)B^T = 0 \ \text{ and } \ (A - B)^T B = 0 .$$

$$\mathrm{vec}(A-B)^T\ \mathrm{vec}(B) = 0\ .$$

(The operator vec(.) stores the columns of the matrix between brackets in a long column vector).
A similar orthogonality property holds true for the dynamic total least squares problem, for which one can prove:

Theorem 2

$$\hat{w}^T\tilde{w} = 0 \text{ or } \begin{pmatrix} v^T \\ z^T \end{pmatrix} (\, G(u-v) \;\; H(y-z)\,) = 0\,,$$

where G and H are diagonal matrices with the weights g_k and h_k.

This property can be exploited in the derivation of algorithms for a closely related problem (within Willem's 'behavior' context [18] [19]).

4. RANK DEFICIENCY OF THE RESIDUAL MATRIX

Let i be a given integer for which it is assumed that $i > n$ where n is the order in the difference equation (1). Consider the *weighted* residual matrix $\widetilde{W}_i \in \mathbb{R}^{(2i\times(N+i))}$ constructed from the sequence $\tilde{w}_k$ as in (4). Then one can show:

Theorem 3
If $\mathrm{rank}[a\ b] = 2$, then:

$$\mathrm{rank}(\widetilde{W}_i) = n + i\ ,\ \forall i > n\ .$$

The fact that $\mathrm{rank}[a\ b] = 2$ is quite natural since it implies that the transfer function of the system is not just a constant.
This Theorem has some farreaching consequences. It is well known that, when the rank of a block Hankel matrix behaves as in the Theorem for increasing i, the data in the matrix (which in this case are the column vectors of the sequence $\tilde{w}$ padded with the appropriate number of zeros) are generated by a linear system of order n [8]. Hence, Theorem 3 really means that the vector sequence $\tilde{w}$ with the weighted residuals (and padded with zeros) is generated by a linear time invariant system. As a matter of fact, we can prove that this system is described as follows:

Corollary 1
The linear system that generates the weighted residuals $\tilde{w}$ is given by

$$-\frac{a^{\mathrm{rev}}(z)}{b^{\mathrm{rev}}(z)}\ .$$

of the coefficients.
From Theorem 2 and 3 we can now give the following interesting interpretation to the solution of the dynamic total linear least squares problem (where for simplicity we assume that the weights g_k and h_k are all 1): The given data w_k (which can be 'arbitrary') are split into two sequence $\hat{w}_k$ and $\tilde{w}_k$, which are not only orthogonal to one another (Theorem 2) but which are also generated by two linear systems: The transfer function $b(z)/a(z)$ for $\hat{w}_k$ can be obtained from a Riemannian SVD as in Theorem 1, while that for $\tilde{w}_k$ is then given by $-a^{\mathrm{rev}}(z)/b^{\mathrm{rev}}(z)$.

One might wonder about the fact that the sequence of weighted residuals is highly structured, in the sense that it can be modelled by a linear system. However, compared to the static case this is not really surprising: There, the matrix of residuals $A - B \in \mathbb{R}^{p\times q}$ is a rank one matrix, hence can be 'modelled' with $q-1$ linear relations. So also in the static case the matrix with residuals is highly structured. In the dynamic case however it is surprising that the residuals themselves also behave as a linear system.

5. ORTHOGONALITY OF THE WEIGHTED RESIDUAL MATRIX AND THE COMPLETED MATRIX WITH APPROXIMATIONS

Let i be a given integer satisfying $i > n$. Let $\widehat{W}_i^{\mathrm{com}} \in \mathbb{R}^{2i\times(N+i)}$ be a block Hankel matrix constructed from an extended sequence

$$(\hat{w}^{\mathrm{com}})^T = (\ \hat{w}_{-i+1}\ \hat{w}_{-i+2}\ \ldots\ \hat{w}_{-1}\ \hat{w}_0\ \ldots\hat{w}_N \\ \hat{w}_{N+1}\ \ldots\hat{w}_{N+i-1}\ )\,,$$

in which we call $(\hat{w}_{-i+1},\ldots,\hat{w}_{-1})$ a *past* extension and $(\hat{w}_{N+1},\ldots,\hat{w}_{N+i-1})$ a *future* extension.

Theorem 4
There exists a unique extension $(\hat{w}_{-n},\ldots,\hat{w}_{-1})$ in the past and an extension $(\hat{w}_{N+1},\ldots,\hat{w}_{N+n})$ in the future such that

$$\widehat{W}_i^{\mathrm{com}}\widetilde{W}_i^T = 0\ ,\ \forall i > n. \qquad (6)$$

This characterization of orthogonality of the optimal solution is quite remarkable. As a special case it implies the orthogonality of Theorem 2. But Theorem 4 says that in addition certain sums of products of vectors of $\hat{w}^{\mathrm{com}}$ and $\tilde{w}$ (which, because of the block Hankel structure of $\widehat{W}_i^{\mathrm{com}}$ and $\widetilde{W}_i$, are finite discrete convolutions) are zero. Hence this corresponds to a certain 'dynamic' orthogonality.

The required extensions of the past and the future can be obtained from a recursion, which is completely characterized by the following

Corollary 2

$$\text{rank } \widehat{W}_i^{\text{com}} = n + i \ , \ i > n \ .$$

Hence, the extensions can be calculated from the difference equation (1) once the solution to the dynamic total least squares problem has been obtained from the Riemannian SVD in Theorem 1.

6. CONCLUSIONS

We have presented the L_2-optimal solution to the so-called dynamic total least squares problem for SISO systems. There are nice properties that characterize the optimal solution such as a 'dynamic' orthogonality of the approximating data and the residuals and the fact that also the residuals are highly structured as they are generated by a linear system.

Much work remains to be done to find the relations between the existing identification methods for linear dynamic systems in which all observations are assumed to have been corrupted by additive noise. It remains to be investigated how this particular solution to the dynamic errors-in-variables problem fits into 'Frisch-like' descriptions of the solution sets.

Acknowledgment

I would like to thank Berend Roorda from the Tinbergen Institute (Rotterdam) for some stimulating discussions which lead to the results in this paper.

7. REFERENCES

[1] Anderson B.D.O., Deistler M.(1984) Identifiability in dynamic errors-in-variables models. *J. Time Series Analysis*, **5**, 1-13.

[2] Anderson B.D.O., Deistler M. (1994). Solution set properties for errors-in-variables problems. *Proc. of the 10th IFAC Symposium on System Identification*, July 4-6 1994, Copenhagen, Denmark, Session on Errors-in-variables models, to appear.

[3] Beghelli S., Guidorzi R., Soverini U.(1990). The Frisch scheme in dynamic system identification. *Automatica*, **26**, 171-176.

[4] Beghelli S., Castaldi P., Guidorzi R., Soverini U.(1994) The Frisch identification scheme: Properties of the solution in the dynamic case. *Proc. of the 10th IFAC Symposium on System Identification*, July 4-6 1994, Copenhagen, Denmark, Session on Errors-in-variables models, to appear.

[5] De Moor B., Vandewalle J. (1986). A geometrical approach to the maximal corank problem in the analysis of linear relations. *Proc. of the 25th IEEE Conference on Decision and Control*, Athens, Greece, **3**, 1990-1996.

[6] De Moor B., Vandewalle J.(1986) The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data. *Proc. of the 25th IEEE conference on Decision and Control*, Athens, Greece, **3**, 1663-1666.

[7] De Moor B., Vandewalle J.(1990). A unifying theorem for linear and total linear least squares identification. *IEEE Transactions on Automatic Control*, Vol.35, **5**, 563-566.

[8] Moonen M., De Moor B., Vandenberghe L., Vandewalle J.(1990). On- and off-line identification of linear state space models. *International Journal of Control*, Vol.49, **1**, 219-239 [4].

[9] De Moor B., Golub G.H.(1991). The restricted singular value decomposition: properties and applications. *Siam Journal on Matrix Analysis and Applications*, Vol.12, **3**, 401-425.

[10] De Moor B. (1993). Structured total least squares and L_2 approximation problems. *Special issue of Linear Algebra and its Applications, on Numerical Linear Algebra Methods in Control, Signals and Systems (eds: Van Dooren, Ammar, Nichols, Mehrmann)*, Volume 188-189, 163-207.

[11] De Moor B. (1994). Total least squares for affinely structured matrices and the noisy realization problem. ESAT-SISTA Report 1993-21, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, 30 pp., May 1993, Accepted for publication in the *IEEE Transactions on Signal Processing*.

[12] De Moor B. (1994). L_2-optimal linear system identification: Dynamic Total Linear Least Squares for SISO systems. Internal Report ESAT-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven.

[13] De Moor B. (1994). The Riemannian SVD: Continuous-time algorithms that solve structured total least squares problems. Department of Electrical Engineering, Internal Report ESAT-SISTA, Katholieke Universiteit Leuven.

[14] Golub G., Van Loan C. (1983). *Matrix Computations.* Johns Hopkins University Press, Baltimore.

[15] Kalman R.E. (1982). Identification from real data. In: *Current developments in the Interface : Economics, Econometrics, Mathematics '(edited by M.Hazewinkel and A.H.G. Rinrooy*

[4]Re-printed in *Numerical Linear Algebra Techniques for Systems and Control*, R.V.Patel, A.J.Laub, P.Van Dooren (Eds.), Reprints Book, IEEE Press, New York, 1993.

Kan), D.Reidel Publishing Co., Dordrecht, 161-196.

[16] Kalman R.E. (1982). System Identification from noisy data. *Proc. Int. Symp.on Dynamical Systems*, (Gainesville, Florida 1981), edited by A.Bednarek, Academic Press.

[17] Kalman R.E. (1983). Identifiability and modeling in econometrics. In: *P.R. Krishnaiah (Ed.). Developments in Statistics 4*, New York, Academic Press.

[18] Roorda B. (1993). Global total least squares modelling of multivariable time series. Preprint Tinbergen Institute, Erasmus University Rotterdam.

[19] Roorda B.(1994). Algorithms for global total least squares modelling of finite multivariable time series. Preprint Tinbergen Institute, Erasmus University Rotterdam.

[20] Scherrer W., Deistler M. (1994). A structure theory for linear dynamic errors-in-variables models. *Proc. of the 10th IFAC Symposium on System Identification*, July 4-6 1994, Copenhagen, Denmark, Session on Errors-in-variables models, to appear.

[21] Van Huffel S., Vandewalle J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis.* Frontiers in Applied Mathematics 9, SIAM, Philadelphia, 300 pp.