94-53

FA-9 8:50

Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

L_2-optimal linear system identification Structured Total Least Squares for SISO systems *

Bart De Moor
ESAT-SISTA/K.U.Leuven
Kard. Mercierlaan 94
B-3001 Leuven
Belgium
tel: 32/16/220931
fax: 32/16/221855
bart.demoor@esat.kuleuven.ac.be

Berend Roorda
Tinbergen Instituut
Oostmaaslaan 950-952 NL
3063 DM Rotterdam
The Netherlands
tel: 31/10/4088932
fax: 31/10/4527347
roorda@tir.few.eur.nl

Abstract

We treat the extension of Total Linear Least Squares to single-input single-output dynamic systems. It is shown that the L_2-optimal model for a given dataset has some remarkable properties:

- 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).
- It is shown how certain block Hankel matrices $\widetilde{W}_i$ constructed from the weighted residuals are always rank deficient and hence, the residuals themselves can be considered as being generated by a linear system.
- One can find completions $\widehat{W}_i^{\mathrm{com}}$ of the block Hankel matrices $\widehat{W}_i$ with the L_2-approximations, such that $\widehat{W}_i^{\mathrm{com}}\ \widetilde{W}_i^T = 0$ (which is a more general type of orthogonality than the first one mentioned).

The results of this paper are just a special application of a general theory to approximate in a least squares sense, a given structured matrix, by one which has the same structure and is rank deficient. The main result of this theory says that, when the matrix structure is affine, the solution is generated in terms of a Riemannian SVD, which is a 'nonlinear' generalized singular value decomposition (see De Moor B., Linear Algebra and its Applications, Vol.188-189, pp.163-207, 1993).

*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 NFWO under contract nr. G.0292.95 (Differential geometry and matrix algorithms), the European SCIENCE project ERNSI (European Research Network for System Identification) and the European Human Capital and Mobility network SIMONET (Systems and Modelling Network). The scientific responsibility is assumed by its authors, one of which (BDM) is a senior research associate with the NFWO.

0-7803-1968-0/94$4.00©1994 IEEE

1 Introduction and notation

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 $w_k = (u_k\ y_k)^T$ (a superscript capital T denotes 'transpose'). 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. Our task now is to find least squares approximations 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. Said in other words, there is no a priori distinction between inputs and outputs and causality is not imposed. The weights g_k and h_k are god-given or, for the areligious or agnostic reader, 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 the 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.

This paper is organized as follows: In Section 2, we state one of the main results, namely that the solution to the dynamic total least squares problem follows from a so-called Riemannian SVD, which is a nonlinear generalization of the Singular Value Decomposition (SVD). In the remaining sections, we treat several properties of the solution: Orthogonality of residuals and approximations (Sections 3 and 5) and the fact that the residuals themselves are generated by a *linear* system (which at least to one of the authors was quite surprising).

We will use the notations

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

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 $\hat{w}_k$:

$$\widehat{W}_i = \left(\begin{array}{cccccccccc} 0 & 0 & \dots & 0 & 0 & \hat{w}_0 & \hat{w}_1 & \dots \\ 0 & 0 & \dots & 0 & \hat{w}_0 & \hat{w}_1 & \dots & \hat{w}_{N-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \hat{w}_0 & \dots & \hat{w}_{N-1} & \hat{w}_N & 0 & 0 & \dots \\ \hat{w}_0 & \hat{w}_1 & \dots & \hat{w}_N & 0 & 0 & 0 & \dots \end{array} \right.$$

$$\left. \begin{array}{cc} \hat{w}_{N-1} & \hat{w}_N \\ \hat{w}_N & 0 \\ \dots & \dots \\ 0 & 0 \\ 0 & 0 \end{array} \right) . \tag{4}$$

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 $\mathbb{R}^{(N+1)\times 2}$:

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

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:

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

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 as 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

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

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 [5], [8]), i.e. the triplet (u, σ, v) corresponding to the smallest singular value σ, which satisfies

$$\begin{array}{ll} Av = u\sigma , & u^T u = 1 , \\ A^T u = v\sigma , & v^T v = 1 . \end{array} \tag{5}$$

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

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

It turns out (see [2] [3] [4]) 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

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

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*

$$u^T(D_{\overline{a}}+D_{\overline{b}})u=1 \text{ and } (\overline{a}^T\ \overline{b}^T)D_u\begin{pmatrix}\overline{a}\\ \overline{b}\end{pmatrix}=1\,. \tag{6}$$

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 [2] for detailed formulas).

Before we present an outline of the proof of this result, 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 $(Y\ \ U)$ 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 [1]) 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* (Also because of the fact that it seems possible using ideas from differential geometry, to design optimization algorithms such as steepest descent or conjugate gradient, on the manifolds described by the constraints (6)).
In [2] 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 [2], 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 [2] [3] (which also contains much more details about the derivation and other additional properties).

Proof of Theorem 1 (outline):
The Lagrangean for the L_2-optimal modelling problem is

$$\begin{aligned}&\mathcal{L}(v_k,z_k,\alpha_i,\beta_j,l_k,\lambda)\\ =\ & J+\sum_{k=n}^{N} l_{k-n}\Big(\sum_{i=0}^{n}[\alpha_i z_{k-i}+\beta_i v_{k-i}]\Big)\\ &+\lambda\Big(1-\sum_{i=0}^{n}[\alpha_i^2+\beta_i^2]\Big)\,,\end{aligned}$$

where $l_k, k=0,\ldots,N-n$ and λ are scalar Lagrange multipliers and J is the object function (2).
Setting now to zero all the derivatives of the Lagrangean results in the following conditions (without loss of generality, we write out the equations for the case $n=2$):

Derivatives with respect to v_k:

$$\begin{aligned}(u_0-v_0)g_0 &= l_0\beta_2\\ (u_1-v_1)g_1 &= l_0\beta_1+l_1\beta_2\\ (u_2-v_2)g_2 &= l_0\beta_0+l_1\beta_1+l_2\beta_2\\ (u_3-v_3)g_3 &= l_1\beta_0+l_2\beta_1+l_3\beta_2\\ &\vdots\\ (u_{N-2}-v_{N-2})g_{N-2} &= l_{N-4}\beta_0+l_{N-3}\beta_1+l_{N-2}\beta_2\\ (u_{N-1}-v_{N-1})g_{N-1} &= l_{N-3}\beta_0+l_{N-2}\beta_1\\ (u_N-v_N)g_N &= l_{N-2}\beta_0\,.\end{aligned} \tag{7}$$

Derivatives with respect to z_k:

$$\begin{aligned}(y_0-z_0)h_0 &= l_0\alpha_2\\ (y_1-z_1)h_1 &= l_0\alpha_1+l_1\alpha_2\\ (y_2-z_2)h_2 &= l_0\alpha_0+l_1\alpha_1+l_2\alpha_2\\ &\vdots\\ (y_{N-2}-z_{N-2})h_{N-2} &= l_{N-4}\alpha_0+l_{N-3}\alpha_1+l_{N-2}\alpha_2\\ (y_{N-1}-z_{N-1})h_{N-1} &= l_{N-3}\alpha_0+l_{N-2}\alpha_1\\ (y_N-z_N)h_N &= l_{N-2}\alpha_0\,.\end{aligned} \tag{8}$$

Derivatives with respect to α_i:

$$\begin{aligned}l_0y_2+l_1y_3+\ldots+l_{N-2}y_N &= \alpha_0\lambda\\ l_0y_1+l_1y_2+\ldots+l_{N-2}y_{N-1} &= \alpha_1\lambda\\ l_0y_0+l_1y_1+\ldots+l_{N-2}y_{N-2} &= \alpha_2\lambda\end{aligned} \tag{9}$$

Derivatives with respect to β_i:

$$\begin{aligned}l_0u_2+l_1u_3+\ldots+l_{N-2}u_N &= \beta_0\lambda\\ l_0u_1+l_1u_2+\ldots+l_{N-2}u_{N-1} &= \beta_1\lambda\\ l_0u_0+l_1u_1+\ldots+l_{N-2}u_{N-2} &= \beta_2\lambda\end{aligned} \tag{10}$$

Derivatives with respect to l_k and λ: will just result in the constraints.

Note that equations (7) and (8) imply that the weighted residuals are generated by linear FIR systems, with the Lagrange multipliers as inputs. In what follows, we will use the vectors

$$a = \begin{pmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{pmatrix} \quad , \quad b = \begin{pmatrix} \beta_2 \\ \beta_1 \\ \beta_0 \end{pmatrix} ,$$

the diagonal matrices $G, H \in \mathbb{R}^{(N+1)\times(N+1)}$:

$$G = \operatorname{diag}(g_k) \quad , \quad H = \operatorname{diag}(h_k) ,$$

and the banded Toeplitz matrices $T_a \in \mathbb{R}^{(N-n+1)\times(N+1)}$ and $T_b \in \mathbb{R}^{(N-n+1)\times(N+1)}$, where e.g. for $n = 2$:

$$T_a = \begin{pmatrix} \alpha_2 & \alpha_1 & \alpha_0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & \alpha_2 & \alpha_1 & \alpha_0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_2 & \alpha_1 & \alpha_0 & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \dots & \dots & \alpha_2 & \alpha_1 & \alpha_0 & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 & \alpha_2 & \alpha_1 & \alpha_0 \end{pmatrix}$$

and T_b is defined similarly. Define the vector of Lagrange multipliers $l \in \mathbb{R}^{N-n+1}$:

$$l^T = (\, l_0 \; l_1 \; \dots l_{N-n-1} \; l_{N-n}) ,$$

and the banded Toeplitz matrix $L_i \in \mathbb{R}^{i\times(N-n+i)}$ constructed from it, i.e.:

$$L_i = \left(\begin{matrix} l_0 & l_1 & l_2 & \dots & l_{i-1} & \dots & l_{N-n} \\ 0 & l_0 & l_1 & \dots & l_{i-2} & \dots & l_{N-n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & l_0 & \dots & l_{N-n-i+1} \end{matrix} \right.$$

$$\left. \begin{matrix} 0 & 0 & \dots & 0 \\ l_{N-n} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & l_{N-n} \end{matrix} \right) .$$

Observe that, if at least one of the elements of l is non-zero, $\operatorname{rank}(L_i) = i$.
Let U, Y, V and $Z \in \mathbb{R}^{(N+1-n)\times(n+1)}$ be the Hankel matrices constructed from the sequences u_k, y_k, v_k and z_k, for instance:

$$Z = \begin{pmatrix} z_0 & z_1 & \dots & z_{n-1} & z_n \\ z_1 & z_2 & \dots & z_n & z_{n+1} \\ z_2 & z_3 & \dots & z_{n+1} & z_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{N-n-1} & z_{N-n} & \dots & z_{N-2} & z_{N-1} \\ z_{N-n} & z_{N-n+1} & \dots & z_{N-1} & z_N \end{pmatrix} .$$

We will frequently use the following trivial

Lemma 1 The 'exchange' Lemma

$$\begin{aligned} Z^T l &= L_{n+1} z \quad , \\ L_{n+1}^T a &= T_a^T l \quad , \\ Za &= T_a z \; . \end{aligned}$$

In words, the transpose of the Hankel matrix Z times the vector l, equals the band Toeplitz matrix L_{n+1} times the vector z. The transpose of the band Toeplitz matrix L_{n+1} times the vector a equals the transpose of the band Toeplitz matrix T_a times the vector l and finally, the Hankel matrix Z times the vector a is the band Toeplitz matrix T_a times the vector z.

Equations (7)-(8) can be written as:

$$\begin{aligned} u - v &= G^{-1} L_{n+1}^T b , \\ y - z &= H^{-1} L_{n+1}^T a . \end{aligned} \tag{11}$$

Equations (9)-(10) together with the constraints (1) and (3) lead to the conclusion that $\lambda = 0$:

$$\begin{aligned} & V^T l = b\lambda , \\ & Z^T l = a\lambda , \\ & Za + Vb = 0 , \\ & a^T a + b^T b = 1 , \end{aligned} \quad \Longrightarrow \quad \lambda = 0 . \tag{12}$$

We can eliminate the matrices V and Z by using the exchange Lemma:

$$\begin{aligned} Z^T l &= L_{n+1} z = L_{n+1}(y - H^{-1} L_{n+1}^T a) = 0 \\ \Longrightarrow \quad & Y^T l = (L_{n+1} H^{-1} L_{n+1}^T) a = D_l^H a, \quad (13) \\ V^T l &= L_{n+1} v = L_{n+1}(u - G^{-1} L_{n+1}^T b) = 0 \\ \Longrightarrow \quad & U^T l = (L_{n+1} G^{-1} L_{n+1}^T) b = D_l^G b \; , \quad (14) \end{aligned}$$

with an obvious definition for D_l^H and D_l^G, which are in $\mathbb{R}^{(n+1)\times(n+1)}$ and are both symmetric positive definite matrices. Also

$$\begin{aligned} Za + Vb &= T_a z + T_b v \\ &= T_a(y - H^{-1} L_{n+1}^T a) \\ &\quad + T_b(u - G^{-1} L_{n+1}^T b) \\ &= 0 \\ \Longrightarrow (\, Y \;\; U \,) \begin{pmatrix} a \\ b \end{pmatrix} &= (T_a H^{-1} T_a^T + T_b G^{-1} T_b^T) l \\ &= (D_a + D_b) l , \end{aligned}$$

with obvious definitions for the matrices D_a and D_b (which are in $\mathbb{R}^{(N-n+1)\times(N-n+1)}$ and which are both symmetric positive definite). We can now write

$$\begin{aligned} & (Y \; U) \begin{pmatrix} a \\ b \end{pmatrix} = (D_a + D_b) l , \\ & \begin{pmatrix} Y^T \\ U^T \end{pmatrix} l = \begin{pmatrix} D_l^H & 0 \\ 0 & D_l^G \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} , \\ & a^T a + b^T b = 1 . \end{aligned} \tag{15}$$

The object function (2) can be written as

$$\begin{aligned} J &= (y - z)^T H (y - z) + (u - v)^T G (u - v) \\ &= a^T L_{n+1} H^{-1} L_{n+1}^T a + b^T L_{n+1} G^{-1} L_{n+1}^T b \\ &= a^T D_l^H a + b^T D_l^G b . \end{aligned}$$

By using another normalization (see [2] for details), the set of equations (15) can be converted to the Riemannian SVD of Theorem 1, for which the minimum singular value the object function (2). This normalization turns the vectors a, b and l into the vectors $\overline{a}$, $\overline{b}$ and u of Theorem 1.

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. We will not present the proofs here in full detail, as they will be written out elsewhere and because they are more or less a straightforward application of the expressions obtained in the proof of Theorem 1.

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:

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

This is equivalent with

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

(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

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

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

4 Rank deficiency of the residuals

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. 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

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

Here, $a^{\mathrm{rev}}(z)$ and $b^{\mathrm{rev}}(z)$ are the polynomials obtained from $a(z)$ and $b(z)$ by reversing the order 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 be surprised 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 residuals and 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

$$\widehat{W}_i^{\mathrm{com}} \widetilde{W}_i^T = 0 \ , \ \forall i > n. \tag{16}$$

This characterization of orthogonality of the optimal solution is quite remarkable. As a special case it implies the orthogonality stated in 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

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

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 been describing the extension of the well-understood Total Linear Least Squares problem for matrices, to the identification of linear dynamic SISO systems. The main idea is to rephrase this problem as one of approximating an affinely structured matrix (in this case the 'double' Hankel matrix $(Y \ U)$ with the data as in Theorem 1), by a rank deficient one with the same structure. Algorithms to solve the Riemannian SVD of Theorem 1 are described in [2] and [3], but there are also very interesting connections with some recent work on continuous time algorithms and optimization on Riemannian manifolds by Brockett and co-authors.
When one tries to apply the ideas of this paper to systems with several outputs (instead of just one), there is a problem of *parametrizations*. Therefore, the algorithms in [6] [7], where full state space models are identified (and not just transfer functions or matrices) seem to provide good alternatives. The relations with the results presented here are still not completely worked out yet.

References

[1] 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.

[2] 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.

[3] 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.*

[4] De Moor B. *Dynamic Total Linear Least Squares.* 10th IFAC Symposium on System Identification, SYSID94, Copenhague, Volume 3, pp. 159-164.

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

[6] Roorda B. (1993). *Global total least squares modelling of multivariable time series.* Report 9343 Tinbergen Institute, Rotterdam, The Netherlands, 1993.

[7] Roorda B. *Algorithms for global total least squares modelling of finite multivariable time series.* Preprint Tinbergen Institute, Rotterdam, The Netherlands, 1994.

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