1

STRUCTURED TOTAL LEAST SQUARES FOR HANKEL MATRICES

Bart L.R. De Moor

Department of Electrical Engineering (ESAT/SISTA)
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium
tel: 32/(0)16/321715, fax: 32/(0)16/321986
email: bart.demoor@esat.kuleuven.ac.be

To Thomas Kailath,
at the occasion of his sixtieth birthday.
With great admiration and gratitude.

ABSTRACT

In this paper[1], we present some recent results on structured total least squares problems, the solution of which can be derived from what we have called a Riemannian singular value decomposition. This is a nonlinear generalization of the singular value decomposition. There are many applications in statistics, signal processing and system identification, but we will concentrate on Hankel matrices, showing how to approximate, in a least squares sense, a given Hankel matrix by a rank deficient one. The importance of this problem originates in the well known equivalence between rank deficient Hankel matrices, realization theory and linear systems. Another application is the least squares fit of a given matrix by an observability matrix. We discuss several properties and characterizations of the optimal solution and illustrate our results on a certain time series related to Thomas Kailath's scientific activities.

[1]Research supported by the Flemish Government (Concerted Research Action MIPS: Model-based Information Processing Systems), the National Fund for Scientific Research (G.0292.95 Matrix algorithms and differential geometry for adaptive signal processing, system identification and control), the Belgian Government (Interuniversity Attraction Poles IUAP-17: Modelling and Control of Dynamical Systems, IUAP-50: Automation in Design and Production) and the European Commission (Human Capital and Mobility Network SIMONET: System Identification and Modelling Network). Bart De Moor is an associate professor at the Katholieke Universiteit Leuven, a Senior Research Associate with the National Fund for Scientific Research and the main advisor on Science and Technology of the Flemish minister-president.

1 INTRODUCTION

The *total linear least squares problem* for a given matrix $A \in R^{p \times q}$, which is of full column rank (without loss of generality, we assume throughout that $p \geq q$), is to find a rank deficient least squares approximation $B \in R^{p \times q}$:

$$\min_{B \in R^{p \times q}\ ,\ \mathrm{rank}(B)=q-1} \|A - B\|_F \ , \tag{1.1}$$

where $\|.\|_F$ is the matrix Frobenius norm. It is well known that the solution can be found from the singular value decomposition (SVD) of the matrix A, by removing from its dyadic decomposition, the term corresponding to the smallest singular value [12] [14] [23]. While the problem is certainly not new[2], the technique has quite recently been popularized in the mathematical engineering community by the books [13] [22].

Often, for applications in statistics, signal processing and system identification, the objective function $\|A - B\|_F^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} (a_{ij} - b_{ij})^2$ is extended with 'elementwise' nonnegative weights w_{ij} and, in addition, the approximating rank deficient matrix B is required to have some structure, such as for instance a Hankel or Toeplitz structure, in which case we have *the structured total least squares problem*:

$$\min_{B \in R^{p \times q}\ ,\ \mathrm{rank}(B)=q-1\ ,\ B \text{ structured}} \sum_{i=1}^{p} \sum_{j=1}^{q} (a_{ij} - b_{ij})^2 w_{ij} \ . \tag{1.2}$$

While in this paper, we will concentrate on Hankel matrices, most of our results carry through (mutatis mutandis) to so-called *affine matrix structures*, i.e. structured matrices B that can be written as an affine (linear) combination of a given set $\{B_k; k = 0, \ldots, N\}$ of basis matrices as $B = B_0 + B_1\beta_1 + \ldots + B_N\beta_N$. Here, the real coefficients $\beta_k, k = 1, \ldots, N$ 'parametrize' the structured matrix B. Examples of such structured matrices are (centro- and per-) symmetric matrices, (block-) Hankel, (block-) Toeplitz, (block-) circulant, Brownian, Hankel + Toeplitz matrices, matrices with a certain zero structure (sparsity pattern), etc ....

This paper treats the optimal least squares approximation of a given Hankel matrix by a rank deficient one. We have omitted all proofs and have kept the discussion informal. For more details and proofs, we will refer to the literature. In Section 2, we elaborate on the relation between rank deficient

[2] In the statistics community, total least squares is known as 'orthogonal regression' and 'errors-in-variables', with first references going back to the previous century [2] [3] [19].

Hankel matrices, linear systems and realization theory. We also give a small motivating example on rank deficient approximation of Hankel matrices from given data, a problem which is one specific instance of a structured total least squares problem. In Section 3, we discuss the solution of structured total least squares problems in terms of a so-called Riemannian singular value decomposition, which is a nonlinear generalization of the SVD. We present its structure for the Hankel case and discuss several characterizations and properties of the optimal solution. In Section 4, a numerical example is given, where we model the cumulative number of publications of Thomas Kailath. In Section 5, we show how the least squares approximation of a given data matrix by an observability matrix, also leads to a structured total least squares problem, this time for a block Hankel matrix. Finally, some summarizing conclusions and perspectives for future research are presented in Section 6.

2 HANKEL MATRICES AND REALIZATION THEORY

Hankel matrices play an important role in linear system theory, modelling and identification, ever since the path breaking work by many authors in the seventies and eighties on realization theory for linear systems. One of the central problems is how to obtain a state space model of a linear system from observed impulse response data.

An intriguing result here states that the rank deficiency of a Hankel matrix with data is equivalent with the fact that these data can be modelled as the impulse response of a linear dynamic time-invariant system. This basic result links the properties of systems being linear and time-invariant to the field of (numerical) linear algebra, in which the rank of a matrix, and also its structure (such as Hankel structure) are central notions. Indeed, the minimal system order (i.e. the minimum number of states) for the 'realization' of the given data sequence as a linear system, is exactly equal to the rank of the Hankel matrix (see e.g. [15]) (It seems that this result was already known to Kronecker in the context of rational approximation).

In practical situations however, where measurements are obtained using finite precision sensors, data are always observed inaccurately, i.e. they are perturbed by measurement noise. Therefore, any $p \times q$ Hankel matrix with $p \geq q$, constructed from these data, either directly or after a deconvolution, will be of full column rank q. It is then meaningful to approximate the 'observed' Hankel matrix A, by a least squares rank deficient approximation B. However, just calculating the SVD of the given Hankel matrix A, and then putting to zero the

q-th singular value (the smallest one), will give the best least squares rank deficient approximation, but the reconstructed matrix will not be a Hankel matrix and hence, 'Kronecker's' result cited above does not apply. Let us illustrate this point with a simple numerical example. Consider the 3×2 Hankel matrix A, which obviously is of rank 2 (its 2 singular values being $\sigma_1 = 6.5468$ and $\sigma_2 = 0.37415$), and its best least squares rank deficient approximation C, which is obtained from the SVD of A by putting its smallest singular value to zero:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix} , \; C = \begin{pmatrix} 1.2608 & 1.8193 \\ 2.0534 & 2.9630 \\ 2.8460 & 4.1067 \end{pmatrix} .$$

Obviously, C, while being of rank 1, is **not** a Hankel matrix. Next consider the following two rank one Hankel matrices:

$$B_1 = \begin{pmatrix} 1.3299 & 1.9448 \\ 1.9448 & 2.8441 \\ 2.8441 & 4.1593 \end{pmatrix} , \; B_2 = \begin{pmatrix} 1.2734 & 1.8827 \\ 1.8827 & 2.7835 \\ 2.7835 & 4.1155 \end{pmatrix} , \qquad (1.3)$$

The first one is the solution to the optimization problem

$$\min_{B \text{ , rank}(B)=1 \text{ , } B \text{ Hankel}} \|A - B\|_F \ ,$$

and can be found using our power method for the Riemannian SVD, discussed below. Its largest singular value is equal to 6.5430 which is not too different from the largest singular value of the original A. Note that due to the Hankel structure, certain elements are counted twice in the objective function $\|A - B\|_F$. In the case that we only count them once, the least squares rank one approximation, which can also be obtained from the Riemannian SVD by introducing the appropriate weights, is given by B_2, the largest singular value of which is 6.4144.

So far for the small example, which illustrates the fact that the singular value decomposition does not preserve the Hankel structure (nor does it with any other structure, except for symmetry) when used as a rank deficient approximation tool. Existing realization methods for 'inaccurate' data, such as the SVD based methods of Zeiger-McEwen [24] and Kung [16] are therefore heuristic, although they often seem to work surprisingly well in practice. Other authors, such as Cadzow [4], have tried to retain the idea of a rank deficient approximation via the SVD while preserving the Hankel structure. The resulting iteration converges, but not to a fixed point which is optimal in any sense (as shown in [8] where we also show that 'classical' iterative methods, such as Steiglitz-McBride and Iterative Quadratic Maximum Likelihood, do not converge to an optimal solution).

In this paper, we formulate and solve the following least squares realization problem:

> Given a sequence of scalar data points $a_i \in R, i = 1, \ldots, N$ (collected in a vector $a \in R^N$) and a user-defined integer $q \leq (N+1)/2$. Find the impulse response $b_i, i = 1, \ldots, N$ (components of the vector $b \in R^N$) of a linear time-invariant system of order at most $q-1$, which is nearest to the given data in a least squares sense.

Mathematically, we have the following problem:

Structured Total Least Squares problem:

$$\min_{b_i, i=1,\ldots,N, y \in R^q} \sum_{i=1}^{N} (a_i - b_i)^2 \text{ subject to } \begin{cases} B\, y = 0\ , \\ y^T y = 1\ , \\ B \text{ is a } p \times q \\ \text{Hankel matrix} \\ \text{with } N = p + q - 1 \end{cases} \tag{1.4}$$

Observe that the rank deficiency of B is guaranteed by the existence of a non-trivial vector y in its null space, and from the Hankel structure of B it is guaranteed that the data b_i, which are the elements of B, can be realized exactly by the impulse response of a linear time-invariant system of order at most $q-1$. As a matter of fact, the $q-1$ poles of the approximating linear system could be found from the components of the vector y if it were known, because they are the coefficients of the characteristic polynomial of the underlying linear system. Observe also that the number q of columns of B, is a *user-defined integer* that determines the order of the 'approximating' linear system, whose minimal order will be at most $q-1$. This implies that the user is free to specify q (as long as $p \geq (N+1)/2$ or equivalently $(N+1)/2 \geq q$, a condition that guarantees that $p \geq q$).

3 THE RIEMANNIAN SVD

Problem (1.4) is a constrained optimization problem and necessary conditions for an optimal solution in the unknowns $b_i, i = 1, \ldots, N$, $y \in R^q$ can be obtained by introducing a vector $l \in R^p$ of Lagrange multipliers for the p constraints, $By = 0$, and a scalar Lagrange multiplier λ for the constraint $y^T y = 1$. From the Lagrangian, by setting all the derivatives to zero, one obtains a set of nonlinear equations in the unknowns $b \in R^N$, $y \in R^q$, $l \in R^p$ and $\lambda \in R$ (see [7] [8] for details). It turns out that $\lambda = 0$ and that the vector b can be eliminated. After some rescalings and renormalizations, one obtains a set of nonlinear equations, which was called the Riemannian SVD in [11]:

Riemannian SVD for Hankel matrices

$$\begin{aligned} A\,v &= D_v\,u\,\tau\,, & u^T D_v u &= 1\,,\\ A^T\,u &= D_u\,v\,\tau\,, & v^T D_u v &= 1\,,\\ & & v^T v &= 1\,. \end{aligned} \tag{1.5}$$

Here $u \in R^p$ and $v \in R^q$ are a left, resp. right singular vector and $\tau \in R$ is the singular value. $D_u \in R^{q\times q}$ and $D_v \in R^{p\times p}$ are banded symmetric Toeplitz positive definite matrix functions, the elements of which are quadratic in the components of u, resp. v as

$$D_u = T_u T_u^T\ ,\ D_v = T_v T_v^T\ ,$$

where $T_u \in R^{p\times(p+q-1)}$ is a banded Toeplitz matrix with the components of u as

$$T_u = \begin{pmatrix} u_1 & u_2 & \ldots & \ldots & u_{p-1} & u_p & 0 & \ldots & 0 & 0 \\ 0 & u_1 & u_2 & \ldots & u_{p-2} & u_{p-1} & u_p & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 0 & u_1 & \ldots & \ldots & u_p \end{pmatrix} \tag{1.6}$$

and $T_v \in R^{p\times(p+q-1)}$ is constructed similarly from the components of v.

The relation between this Riemannian SVD and the optimal least squares solution to the noisy realization problem, follows from the following observations.

Fact 1: The singular value squared is exactly equal to the value of the objective function in (1.4):

$$\tau^2 = \sum_{i=1}^{N} (a_i - b_i)^2 \ ,$$

which implies that we need to find the minimal τ satisfying (1.5).

Fact 2: The structured rank deficient least squares approximation B can be reconstructed as

$$B = A - \begin{pmatrix} u_1 & u_2 & \dots & \dots & \dots & u_p & 0 & \dots & 0 & 0 & 0 \\ u_2 & u_3 & \dots & \dots & u_p & 0 & 0 & \dots & 0 & 0 & u_1 \\ u_3 & u_4 & \dots & u_p & 0 & 0 & 0 & \dots & 0 & u_1 & u_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ u_p & 0 & \dots & \dots & 0 & u_1 & \dots & \dots & \dots & u_{p-2} & u_{p-1} \end{pmatrix}$$

$$\times \tau \times \begin{pmatrix} v_1 & v_2 & \dots & v_q \\ 0 & v_1 & \dots & v_{q-1} \\ 0 & 0 & \dots & v_{q-2} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & v_1 \\ \dots & \dots & \dots & \dots \\ v_q & 0 & \dots & 0 \\ v_{q-1} & v_q & \dots & 0 \\ \dots & \dots & \dots & \dots \\ v_2 & v_3 & \dots & 0 \end{pmatrix} \tag{1.7}$$

which means that the difference $A - B$ is the product of a Hankel matrix with the elements of u, a Toeplitz matrix with the elements of v (note the 'circulant' structure in both matrices) and the scalar τ [3].

Fact 3: There is an interpretation to the singular vectors u and v. The right singular vector v satisfies

$$Bv = 0 \ , \ v^T v = 1 \ . \tag{1.8}$$

and therefore, its components are the coefficients of the characteristic polynomial of the approximating linear system (Actually, v can be taken equal to y

[3] It is interesting to observe that every $p \times q$ Hankel matrix A of full column rank q can be factorized in such a Hankel $\times$ Toeplitz product, i.e. when one puts $B = 0$ in (1.7), there exist $u \in R^p$, $v \in R^q$ and $\tau \in R$ such that A is factorized as prescribed. Such a triplet (u, τ, v) will satisfy the Riemannian SVD equations (1.5), with $\tau = \sqrt{\sum_{i=1}^{N} a_i^2}$ but is not optimal in any sense. For instance, for the 3×2 Hankel matrix A in Section 2, we have $u^T = (0.35235, 0.12310, 0.85386)$,$v^T = (0.51816, 0.85529)$, $\tau = 5.4772 = \sqrt{(}1 + 2^2 + 3^2 + 4^2)$, which satisfies the equations (1.5). This solution however does not correspond to the minimizing one.

in (1.4)). The left singular vector u is the vector of Lagrange multipliers l, normalized so as to satisfy the norm constraint in (1.5). This implies that u can be used in analysing the first order sensitivity properties of the structured total least squares optimization problem.

Fact 4: It can be shown that the optimal approximating impulse response, which is contained in the vector b, and the original data vector a, satisfies the following orthogonality property: $(a-b)^T b = 0$. As this is a (structured) least squares approximation problem, it is not too surprising to see that the vector of residuals $a - b$ is orthogonal to the vector of approximants b.

Fact 5: The minimal singular value $\tau_{\min}$ and the corresponding left and right singular vectors u and v, and hence the optimal Hankel matrix B in (1.7), can be calculated by an algorithm, described in [7] [8]. The source of inspiration for this algorithm is the good old power method [25] for calculating the dominant eigenvalue and corresponding eigenvector of a symmetric matrix and seems to work well in practice [4]. Its asymptotic convergence rate is linear, as asymptotically it behaves exactly like the power method. A numerical example is given in Section 4.

Fact 6: When D_v is invertible, it is straightforward to show that

$$\tau^2 = v^T A^T D_v^{-1} A v \ , \tag{1.9}$$

which would be a quadratic form in v, if D_v were a matrix independent of v, which is not the case though. Yet, this observation is the basis of the *constrained total least squares* method, described in [1], which is basically applying Gauss-Newton to the nonlinear optimization problem. The equivalence with the problem (1.4) is discussed in [10].
Another possible algorithm is suggested in [21]. It starts from the observation that the Riemannian SVD can be rewritten as a generalized eigenvalue problem of the form

$$A^T D_v^{-1} A v = D_u v \tau^2 \ , \ v^T D_u v = 1 \ .$$

When fixing D_u and D_v, one can solve this problem for the minimal eigenvalue τ^2 and corresponding eigenvector v, determine a corresponding u, update D_u and D_v and then restart the whole process until convergence [5]. Yet, our experience shows that convergence here occurs less often than with our power method. An intuitive explanation for this is that the power method only does

[4] Several Matlab .m files are available from the author.

[5] Remark that, when in (1.5) D_u and D_v would be 'constant' matrices (i.e. independent of u and v), then, the Riemannian SVD would coincide with the Restricted SVD discussed in [5].

one little step towards the minimal τ and corresponding eigenvector v, for fixed D_u and D_v instead of going 'all the way'. In a certain sense, the power method updates the matrices D_u and D_v more frequently than in the eigenvalue method outlined above. Of course, this is only an intuitive explanation which deserves further investigation.
In [11] we have discussed continuous time methods for structured total least squares problems.
Finally, in [8] we have investigated other methods, such as Cadzow's method and Iterative Quadratic Maximum Likelihood, with a demonstration that they deliver suboptimal results. In [17] we have shown the equivalence between Abatzoglou's [1] constrained total least squares solution and the structured total least squares approach presented here.

Fact 7: Let $\hat{a} \in R^N$ be the vector of 'exact' data (i.e. containing the impulse response of a linear system of order $q-1$), $\tilde{a} \in R^N$ be a vector of i.i.d. normally distributed zero mean white noise with variance σ^2 and let the observed data sequence be $a = \hat{a} + \tilde{a}$. The structured total least squares problem (1.4) is then the maximum likelihood formulation of the realization problem.
It's interesting to observe that one can also motivate the objective function (1.9) in a heuristic way as follows. Let $A = \hat{A} + \tilde{A}$ be the $p \times q$ Hankel matrix with observed data (with obvious definitions for $\hat{A}$ and $\tilde{A}$ as Hankel matrices). Let's assume that $\text{rank}(\hat{A}) = q-1$, $\text{rank}(A) = \text{rank}(\tilde{A}) = q$. Let $v \in R^q$ be a vector in the null space of $\hat{A}$. Then

$$Av = \hat{A}v + \tilde{A}v = \tilde{A}v = e \ ,$$

where the vector $e \in R^p$ can be considered as a vector with the equation errors. It seems natural to minimize the so-called equation error, which is the norm of e. However, we have to take into account the correlation structure of the elements of e, which is induced by the Hankel structure in the matrix $\tilde{A}$. The covariance matrix is obtained as

$$\mathbf{E}(e.e^T) = \mathbf{E}(\tilde{A}.v.v^T.\tilde{A}^T) = D_v\sigma^2 \ ,$$

which is precisely the banded symmetric positive definite Toeplitz matrix of the Riemannian SVD (1.5), up to within the scalar σ^2, which is the white noise variance. Since we now know the structure of the noise covariance matrix, we can minimize the weighted equation error as

$$e^T D_v^{-1} e = v^T A^T D_v^{-1} A v \ ,$$

which is exactly the expression (1.9). Actually, this heuristic insight leads to a 'quick and dirty' way of calculating the structure of D_u and D_v for any structured total least squares problem of the form (1.2) as long as the matrix structure involved is affine.

Fact 8: Another interesting property is the following

$$\sum_{i=1}^{N}(a_i - b_i) = (\sum_{i=1}^{p} u_i)\ \tau\ (\sum_{i=1}^{q} v_i)\ .$$

4 A NUMERICAL EXAMPLE: PREDICTING THE CUMULATIVE NUMBER OF TK'S PUBLICATIONS

Consider the data sequence of Figure 1 which is the cumulative number of journal publications of Thomas Kailath, starting from 1960 (actually the year when I was born) up to 1995. There are 36 data points. We decide to fit these data by impulse responses of linear systems of increasing order, starting with order 1 (in which case the data Hankel matrix A of the Riemannian SVD is $(p = 35) \times (q = 2)$ [6], up to order 17 (where $p = 19$ and $q = 18$). Therefore, we have calculated the minimal singular value of the Riemannian SVD (1.5), for $p \times q$ data Hankel matrices A, where q varies from 2 up to 17. As can be expected, we get a better and better approximation as the order of the 'approximating' system increases. As is obvious from Figure 1, all of these realizations are unstable. For instance, the poles for the fifth-order approximation, which corresponds to the case $p = 31, q = 6$ are $\{1.0972, 1.0062 \pm 0.20505j, 0.91204 \pm 0.63262j\ \}$. They can easily be obtained by rooting the characteristic polynomial, the coefficients of which are given by the components of the right singular vector v of the Riemannian SVD (1.5).

[6] Actually, an impulse response of a first order linear system can be parametrized by two real numbers β and λ so that the optimization problem becomes an unconstrained one: $\min_{\beta,\lambda} \sum_{i=1}^{N}(a_i - \beta\lambda^{i-1})^2$. Setting to zero the derivative with respect to β, results in $\sum_{i=1}^{N}(a_i-\beta\lambda^{i-1})\lambda^{i-1} = 0$, and the derivative w.r.t. λ gives $\sum_{i=2}^{N}(a_i-\beta\lambda^{i-1})(i-1)\lambda^{i-2} = 0$. Observe that β appears linearly in both equations. When eliminated from one and substituted in the other, one obtains a polynomial in λ, which can be rooted. One of its real roots will generate the minimizing first order impulse response (see [8] for details). For instance, for the example (1.3), one then obtains the polynomial of degree 7 in λ: $3\lambda^7 + 0\lambda^6 + 3\lambda^5 - 6\lambda^4 - \lambda^3 - 12\lambda^2 - 5\lambda - 2 = 0$, which has 3 pairs of complex conjugated roots and only 1 real one, which is 1.4785, corresponding to the 3×2 optimal Hankel matrix B in (1.3)

5 LEAST SQUARES APPROXIMATION BY OBSERVABILITY MATRICES

A structured total least squares problem that often occurs in signal processing (e.g. in Direction of Arrival problems such as ESPRIT) or in system identification (subspace methods such as N4SID), is the following:
Given a data matrix $A \in R^{m \times n}$, $m > n$. Find the least squares approximation $G \in R^{m \times n}$ that is an observability matrix, i.e. a matrix for which the i-th row $g_i^T \in R^{1 \times n}$ is given by $g_i^T = h^T.H^{i-1}$ where $h \in R^n$ and $H \in R^{n \times n}$.
Stated as an optimization problem, this becomes:

$$\min_{h \in R^n \ , \ H \in R^{n \times n}} \|A - \begin{pmatrix} h^T \\ h^T H \\ h^T H^2 \\ \vdots \\ h^T H^{m-1} \end{pmatrix} \|_F^2 \ . \tag{1.10}$$

One could also introduce elementwise weights into the objective function but we will consider the unweighted case only. We now show that this problem is equivalent with a structured total least squares problem, this time for a block Hankel matrix, as follows. Let the components of the vector $y \in R^{n+1}$ be the coefficients of the characteristic polynomial of the matrix H. Then, it is easy to show that

$$T_y \begin{pmatrix} h^T \\ h^T H \\ \vdots \\ h^T H^{m-1} \end{pmatrix} = 0 \ ,$$

where $T_y \in R^{(m-n) \times m}$ is a banded Toeplitz matrix as in (1.6). If we put $g_i^T = h^T H^{i-1}$, then this relation is equivalent with

$$T_y G = 0 \Longleftrightarrow \begin{pmatrix} g_1 & g_2 & \cdots & g_n & g_{n+1} \\ g_2 & g_3 & \cdots & g_{n+1} & g_{n+2} \\ g_3 & g_4 & \cdots & g_{n+2} & g_{n+3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{m-n} & g_{m-n+1} & \cdots & g_{m-1} & g_m \end{pmatrix} . \ y = 0 \ , \tag{1.11}$$

where G is a matrix for which the i-th row is g_i^T. The block matrix to the right is a block Hankel matrix with $m-n$ block rows, each of dimension n, and $n+1$ columns, which will be denoted by $B \in R^{(m-n)n \times (n+1)}$. Such a matrix is an affinely structured matrix, which is required to be rank deficient. Hence, we have shown that the least squares approximation of a given matrix A by an observability matrix as in (1.10) is a structured total least squares problem

$$\min_{\substack{g_i \in R^n, i=1,\dots,m\ , \\ y \in R^{n+1}}} \|A - G\|_F^2 \text{ subject to } \begin{cases} B\,y = 0\ , \\ y^T y = 1\ , \\ B \text{ block Hankel as in (1.11).} \end{cases} .$$

The solution can be obtained from a Riemannian SVD as in (1.5), but with metric matrices D_u and D_v, whose structure is different from the ones in (1.5). It can easily be obtained, for instance with the heuristic 'noise' trick, outlined under 'Fact 7' above.

6 CONCLUSIONS

In this short paper, we have discussed the total least squares approximation of Hankel matrices, with as an additional constraint that the approximant be Hankel too. The insights here have been elaborated upon in detail in [7] [8] [9] to which we refer the reader for more details. Actually, the Riemannian SVD is quite a general result for rank deficient matrix approximation for weighted and structured matrix problems, as long as the matrix structure is affine. For instance, in [7], we also show how instead of minimizing the sum of absolute errors squared in the objective function, one can also minimize (a first order approximation to) the sum of *relative* errors squared. If the data a_i are already samples from an impulse response of a (high-order, say r) linear system and if the number of rows $p \to \infty$ while $q < r$ is fixed, the Hankel approximation problem treated here is equivalent to the H_2 model reduction problem. Z-domain iterations have been described in [11] [20]. In [9] [10] we have discussed the application of this framework for the identification of linear dynamic systems from input-output data, when both the inputs and the outputs are corrupted by noise (errors-in-variables).
The structured total least squares problem could be treated as a nonlinear constrained optimization problem and 'classical' minimization algorithms could be applied to come up with numerical solutions. Yet we feel that our interpretation in terms of a Riemannian SVD reveals that there is quite some structure in the solution, which is reminiscent of the solution of the unstructured case. We hope to reveal in the future more useful properties which maybe ultimately lead to a globally convergent algorithm.

Acknowledgements

From our first meeting in the Roman terms of Bath, England, at an evening reception of the Conference on Mathematics and Signal Processing in September 1985, over our discussion in Valencia, Spain, at the International Conference on Linear Algebra in September 1987, 15 minutes before TK was going to address the plenary audience about displacement ranks, over my post-doc period in Stanford from August 1988 till September 1989, over several visits of me and my family to Stanford and of TK to Leuven, to the fabulous 5-day birthday party in Stanford in June 1995, is a period of only 10 years in which I have learned to appreciate Tom as a brilliant teacher, a devoted researcher and a friendly and caring person.
My students (Johan David, Jeroen Dehaene, Philippe Lemmerling, Christiaan Moons, Geert Schelfhout, Peter Van Overschee), who in one way or another contributed to the work presented here, and I, we all sincerely hope that the explicit instability of Figure 1 may be an appropriate model for TK's scientific production over a considerable time horizon in the future !

REFERENCES

[1] Abatzoglou T.J., Mendel J.M., Harada G.A., *The constrained total least squares technique and its application to harmonic superresolution*, IEEE Transactions on Signal Processing, SP-39, no.5, pp. 1070-1087, 1991.

[2] Adcock R.J. *Note on the method of least squares.* The Analyst, Vol IV, no.6, Nov. 1877, pp.183-184.

[3] Adcock R.J. *A problem in least squares.* The Analyst, March 1878, Vol V, no.2, pp.53-54.

[4] Cadzow J., *Signal enhancement: a composite property mapping algorithm*, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-36, pp. 49-62, 1988.

[5] De Moor B., Golub G.H. *The restricted singular value decomposition: properties and applications.* Siam Journal on Matrix Analysis and Applications, Vol.12, no.3, July 1991, pp.401-425.

[6] De Moor B., Van Overschee P., Schelfhout G. *H_2-model reduction for SISO systems.* Proc. of the 12th World Congress International Federation of

Automatic Control, Sydney, Australia, July 18-23 1993, Vol. II pp.227-230. [7]

[7] De Moor B. *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, July 1993, pp.163-207.

[8] De Moor B. *Total least squares for affinely structured matrices and the noisy realization problem* . IEEE Transactions on Signal Processing, Vol.42, no.11, November 1994.

[9] De Moor B. *Dynamic Total Linear Least Squares.* SYSID '94, Proc. of the 10th IFAC Symposium of System Identification, 4-6 July 1994, Copenhagen, Denmark. Vol.3., pp.159-164.

[10] De Moor B., Roorda B. *L_2-optimal linear system identification: Dynamic total least squares for SISO systems.* Proc. of 33rd IEEE CDC, Florida, December 1994.

[11] De Moor B. *The Riemannian Singular Value Decomposition.* Proc. 3rd Int. Workshop on SVD and Signal Processing, 1995, pp. 61-78, Elsevier Science Publisher.

[12] Eckart C., Young G. *The approximation of one matrix by another of lower rank.* Psychometrika, 1, pp.211-218, 1936.

[13] Golub G.H., Van Loan C. *Matrix Computations.* Johns Hopkins University Press, Baltimore 1989 (2nd edition).

[14] Householder A.S., Young G. *Matrix approximation and latent roots.* Americ. Math. Monthly, 45, pp.165-171, 1938.

[15] Kailath T. *Linear Systems.* Prentice Hall, Englewood Cliffs, New Jersey, 1980.

[16] Kung S.Y. *A new identification method and model reduction algorithm via the singular value decomposition.* 12th Asilomar Conference on Circuits, Systems and Computers, pp.705-714, Asilomar, California, 1978.

[17] Lemmerling P., De Moor B., Van Huffel S. *On the equivalence of constrained and structured total least squares.* ESAT-SISTA Report 1995-55, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, October 1995.

[7] Reprinted in the Automatic Control World Congress 1993 5-Volume Set, Volume 1: Theory.

[18] Nagelkerke N.J.D., *Maximum Likelihood Estimation of Functional Relationships.* Lecture Notes in Statistics, Ed. J. Berger et al., Springer-Verlag, 1992.

[19] Pearson K. *On lines and planes of closest fit to systems of points in space.* Phil. Mag., 2, 6-th series, pp.559-572.

[20] Schelfhout G. *Model reduction for control design.* PhD Thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, November 1995.

[21] Van hamme H. *Identification of linear systems from time- or frequency domain measurements.* PhD Thesis, April 1992, Department of Electrical Engineering, Vrije Universiteit Brussel, Belgium, pp. 294.

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

[23] Young G. *Matrix approximation and subspace fitting.* Psychometrika, vol.2, no.1, March 1937, pp.21-25.

[24] Zeiger H., McEwen A. *Approximate linear realizations of given dimension via Ho's algorithm.* IEEE Transactions on Automatic Control, 19, pp.153, 1974.

[25] Wilkinson J. *The algebraic eigenvalue problem.* New York, Oxford University Press, 1965.