93-215

Total least squares for affinely structured matrices and the noisy realization problem

Bart De Moor, Senior Member IEEE

Abstract— **Structured rank deficient matrices arise in many applications in signal processing, system identification and control theory. We discuss the Structured Total Least Squares (STLS) problem, which is the problem of approximating affinely structured matrices (i.e. matrices affine in the parameters) by similarly structured rank deficient ones, while minimizing an L_2-error criterion. It is shown that the optimality conditions lead to a nonlinear generalized singular value decomposition, which can be solved via an algorithm that is inspired by inverse iteration. Next we concentrate on the so-called L_2-optimal noisy realization problem, which is equivalent with approximating a given data sequence by the impulse response of a finite dimensional, time invariant linear system of a given order. This can be solved as a Structured Total Least Squares problem. It is shown with some simple counterexamples that 'classical' algorithms such as Steiglitz-McBride, Iterative Quadratic Maximum Likelihood and Cadzow's iteration do not converge to the optimal L_2 solution, despite misleading claims in the literature.**

Keywords— **Hankel matrix, (Restricted) singular value decomposition, realization theory, (structured) Total Least Squares, Iterative Quadratic Maximum Likelihood (IQML), Steiglitz-McBride iteration.**

I. Introduction

THE noisy realization problem is the problem of approximating a given data sequence by the impulse response of a linear time-invariant (LTI) system, such that the (weighted) sum of squares of residuals is minimized. Applications occur in system identification, modal analysis, biomedical signal processing (such as e.g.NMR), etc... In this paper, we treat the noisy realization problem in the context of the so-called *Structured Total Least Squares problem* (STLS): Let $B(b) = B_0 + b_1 B_1 + \ldots + b_n B_m \in \mathbb{R}^{p\times q}$ be an affine matrix function of the components b_i of the parameter vector $b \in \mathbb{R}^m$ where $B_i, i = 0, 1, \ldots, m$ are fixed given matrices. We assume throughout that $p \geq q$. Let $a \in \mathbb{R}^m$ be a data vector and w be a given vector of weights. The problem is then to find a rank deficient matrix in the affine set $B(b)$ such that a given quadratic function $[a, b, w]_2^2$ of the parameters b_i is minimized.

This problem often occurs in systems and control and we refer to [6] for a survey of applications, including total least squares with relative weightings and fixed elements, linear system identification, rank deficient Toeplitz matrix approximation and the calculation of the largest L_2 stability radius of a parametric uncertain system.

In this paper, we will first derive the optimality conditions which have to be satisfied by any L_2-optimal solution. This will be done in our main Theorem in Section II, which states that the solution can be constructed from a nonlinear generalized singular value decomposition. An algorithm to find a solution, inspired by the power method of numerical analysis, is proposed in Section III. Its asymptotic convergence rate seems to be linear as is suggested by the numerical experiments in Section VI. In Section IV, we concentrate on the L_2-optimal noisy realization problem as a special case of the STLS problem. In Section V we show that some well known 'classical' algorithms that have been proposed to solve the noisy realization problem, such as the Iterative Quadratic Maximum Likelihood method (IQML) [2], the Steiglitz-McBride iteration [11] (which in [10] was shown to be equivalent with IQML for the noisy realization problem) and Cadzow's method [3], do not converge to the L_2-optimal solution. This is shown on a simple numerical counterexample on which the global optimum can be calculated explicitly. For Cadzow's algorithm, formal convergence to a fixed point can be proven, but his method does not satisfy any optimality criterion. If they converge, IQML and Steiglitz-McBride do so to a suboptimal solution. This suboptimality is due to a misinterpretation of the freedom that one has to manipulate a certain constraint and is explained in detail in Section V.

We will confine ourselves to real data and real STLS problems, although all results can be generalized to the complex field. Our notation is fairly standard.

Structured Total Least Squares

Let us now consider in detail the STLS problem where we take the quadratic criterion

$$[\, a \,,\, b \,,\, w \,]_2^2 = \sum_{i=1}^{m} (a_i - b_i)^2 \,,$$

in which, for the moment, we do not consider weights w (An example with weights is treated in Section IV and is

ESAT - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium, tel 32/16/22 09 31, fax 32/16/22 18 55, email: *bart.demoor@esat.kuleuven.ac.be*.
This work was started when the author was a visitor at the Institute of Mathematics and Applications (IMA) of the University of Minnesota in Minneapolis in April 1992 (IMA Preprint Series nr 1036, September 1992). It was supported in part by the IMA with funds provided by the National Science Foundation. Additional support was provided by the Belgian Program on Interuniversity Attraction Poles, initiated by the Belgian Science Policy Programming (Prime Minister's Office) and by the European Community Research Program ESPRIT, Basic Research Action no.3280. The scientific responsibility is assumed by its author, who is a research associate of the Belgian National Fund for Scientific Research (NFWO).

a straightforward extension of the results obtained in this section).

Theorem 1 STLS as a non-linear generalized SVD

Consider the STLS problem

$$\min_{b \in \mathbb{R}^m,\, y \in \mathbb{R}^q} \sum_{i=1}^{m}(a_i - b_i)^2 \quad \text{subject to} \quad \begin{array}{c} B(b)y = 0, \\ y^t y = 1, \end{array} \tag{1}$$

where $a_i, i = 1, \ldots, m$ are the components of the data vector $a \in \mathbb{R}^m$ and $B(b) = B_0 + B_1 b_1 + \ldots + B_m b_m$, with $B_i, i = 0, 1, \ldots, m \in \mathbb{R}^{p \times q}$ fixed given matrices.
The solution is described as follows:
- Find the triplet (u, τ, v), $u \in \mathbb{R}^p$, $v \in \mathbb{R}^q, \tau \in \mathbb{R}$ corresponding to the smallest scalar τ that satisfies

$$\begin{array}{ll} Av = D_v u \tau, & u^t D_v u = 1, \\ A^t u = D_u v \tau, & v^t D_u v = 1, \end{array} \tag{2}$$

where $A = B_0 + \sum_{i=1}^{m} a_i B_i$. D_u is defined via

$$\sum_{i=1}^{m} B_i^t (u^t B_i v) u = D_u v,$$

and is a symmetric positive or nonnegative definite matrix, the elements of which are quadratic in the components of u. D_v is defined similarly via

$$\sum_{i=1}^{m} B_i (u^t B_i v) v = D_v u,$$

and is symmetric positive or nonnegative definite with elements that are quadratic in the components of v.
- The vector y is given as $y = v/||v||$.
- The components of b are obtained from

$$b_k = a_k - u^t B_k v \tau, \qquad k = 1, \ldots, m. \tag{3}$$

Before we present the proof of this Theorem, we will first spend a few words on its interpretation. First observe that, if D_u and D_v would be constant positive or nonnegative definite matrices (i.e. independent of u and v), the equations of Theorem 1 for a given matrix A would be satisfied by any triplet (u, τ, v) of the *Restricted Singular Value Decomposition* (RSVD) of the triplet $(A, D_v^{1/2}, D_u^{1/2})$ (where $D_v^{1/2}$ is a square root of D_v). The RSVD is just an SVD with different positive or nonnegative definite inner products in the column and row space and is extensively studied in [5]. Observe that, still under the assumption that D_u and D_v are constant (2) would be a generalized eigenvalue problem. The 'non-linear' aspect however is the explicit dependence of the weighting matrices D_u and D_v on the components of u and v. Still, as we will see, these matrices are always nonnegative definite (hence correspond to inner products, which are however 'position dependent') and the elements are quadratic functions of the components of u and v.

Proof of Theorem 1: For the proof we will use Lagrange multipliers and proceed in several steps. The Lagrangean is given by

$$\begin{aligned} \mathcal{L}(b, y, l, \lambda) &= \sum_{i=1}^{m}(a_i - b_i)^2 + l^t(B_0 + b_1 B_1 + \ldots \\ &\quad + b_m B_m) y + \lambda(1 - y^t y), \end{aligned}$$

where $l \in \mathbb{R}^p$ is a vector of Lagrange multipliers and $\lambda \in \mathbb{R}$ is a scalar Lagrange multiplier.
Step 1: Derivatives: Setting all derivatives to 0 and absorbing irrelevant constants in the Lagrange multipliers, gives

$$\begin{aligned} &\forall k > 0 \;:\; a_k - b_k = l^t B_k y, \\ &(B_0 + b_1 B_1 + \ldots + b_m B_m) y = 0, \\ &(B_0^t + b_1 B_1^t + \ldots + b_m B_m^t) l = y \lambda, \\ &y^t y = 1. \end{aligned} \tag{4}$$

From $l^t B(b)\, y = 0$, it follows directly that $\lambda = 0$.
Step 2: Elimination of b: Next we eliminate the parameters b by using $b_k = a_k - l^t B_k y$ to find

$$\begin{aligned} (B_0 + a_1 B_1 + \ldots + a_m B_m) y &= ((l^t B_1 y) B_1 + \ldots \\ &\quad + (l^t B_m y) B_m) y \quad (5) \\ (B_0^t + a_1 B_1^t + \ldots + a_m B_m^t) l &= ((l^t B_1 y) B_1^t + \ldots \\ &\quad + (l^t B_m y) B_m^t) l \quad (6) \end{aligned}$$

Looking at the right hand sides, we observe that the first right hand side (5) is quadratic in the elements of y and linear in those of l, while the second right hand side (6) is quadratic in the elements of l and linear in those of y.
Let's now concentrate on one term in the right hand side of (5). Without loss of generality, we can take the first one. Define β_{ij} as element (i, j) of B_1 and let $\overline{b_i}^t$ be the i-th row of B_1. Then

$$\begin{aligned} \text{Element } k \text{ of } (B_1 y (l^t B_1 y)) &= \sum_{j=1}^{q} \beta_{kj} y_j \sum_{r=1}^{p} \sum_{s=1}^{q} \beta_{rs} l_r y_s \\ &= \sum_{r=1}^{p} (y^t \overline{b_k})\, (\overline{b_r}^t y) l_r . \end{aligned}$$

Hence,

$$B_1 y (l^t B_1 y) = \begin{pmatrix} y^t \overline{b_1} \\ y^t \overline{b_2} \\ \vdots \\ y^t \overline{b_p} \end{pmatrix} (\, \overline{b_1}^t y \;\; \ldots \overline{b_p}^t y \,)\, l .$$

Observe that in the right hand side, the matrix preceeding l is a rank one matrix and as it is the outer product of a vector with itself, it is nonnegative definite. Obviously, we can repeat this for each term of (5) to obtain the result that the right hand side of (5) can be written as

$$\sum_{i=1}^{m} (B_i (l^t B_i y)) y = D_y l . \tag{7}$$

Here, D_y is a symmetric matrix which is a sum of m rank one nonnegative definite matrices, hence D_y itself is nonnegative or positive definite. Its elements are quadratic functions of the components of the vector y. A similar derivation applies for the right hand side (6):

$$\sum_{i=1}^{m}(B_i^t(l^t B_i y))l = D_l y \ , \tag{8}$$

where D_l is symmetric nonnegative or positive definite, with elements that are quadratic functions of the components of l.

Step 3: Normalization: Next we define $x = l/||l||$ and call $\sigma = ||l||$. Let D_x be defined in the same way as D_l, by replacing every component of l appearing in D_l by its corresponding component in x. Since the elements of D_l are quadratic functions of the components of l, we find $D_l = D_x\sigma^2$. Next define $A \in \mathbb{R}^{p\times q}$ as $A = B_0 + \sum_{i=1}^m B_i a_i$. Then we find from (5), (6), (7) and (8):

$$\begin{aligned} A\,y &= D_y x\sigma, & x^t x &= 1, \\ A^t\,x &= D_x y\sigma, & y^t y &= 1. \end{aligned} \tag{9}$$

Step 4: Evaluation of the object function: From equations (9) it follows directly that

$$x^t A y = y^t A^t x = x^t D_y x\sigma = y^t D_x y\sigma. \tag{10}$$

Next observe that from (4)

$$\begin{aligned} \sum_{i=1}^{m}(a_i - b_i)^2 &= \sum_{i=1}^{m}(l^t B_i y)^2 \\ &= \sum_{i=1}^{m}(x^t B_i y)^2\sigma^2 = x^t \sum_{i=1}^{m}(B_i y(x^t B_i y))\ \sigma^2 \\ &= x^t D_y x\sigma^2 = x^t A y\sigma \ . \end{aligned} \tag{11}$$

The last equality follows from (7) and (10).

Step 5: Equations (2) imply (9) : Let the triplet (u,τ,v) solve (2). Then

$$u^t A v = (u^t D_v u)\tau = \tau \ . \tag{12}$$

Furthermore, by normalizing u and v, and recalling that D_u and D_v are quadratic in the components of u and v, we obtain:

$$\begin{aligned} A\frac{v}{||v||} &= (D_v/||v||^2)\frac{u}{||u||}(\tau||u||||v||) \ , \\ A^t\frac{u}{||u||} &= (D_u/||u||^2)\frac{v}{||v||}(\tau||u||||v||) \ . \end{aligned}$$

If we now put $x = u/||u||$, $y = v/||v||$ and $\sigma = \tau||u||||v||$, we have shown that equations (2) imply equations (9) .

Step 6: The object function equals τ^2: From (11) and (12) we find

$$\begin{aligned} x^t A y\sigma &= \frac{u^t}{||u||}A\frac{v}{||v||}(\tau||u||||v||) \\ &= \frac{u^t}{||u||}\frac{D_v}{||v^2||}\frac{u}{||u||}(\tau||u||||v||)^2 = \tau^2 \ . \end{aligned} \tag{13}$$

This shows that we need to find the minimal τ. It also delivers the result for $y = v/||v||$. For the components of b we find from (4)

$$b_k = a_k - \frac{u^t}{||u||}B_k\frac{v}{||v||}\sigma = a_k - u^t B_k v\tau \ .$$

Step 7: Equations (9) are equivalent with equations (2) : Because of Step 5, we only need to show that (9) implies (2). Let (x,σ,y) satisfy (9). Define u and v as $u = x/\alpha$ and $v = y/\beta$ for certain scalars α and β. Then from (9):

$$\begin{aligned} Av\beta = D_v\beta^2 u\alpha\sigma &\iff Av = D_v u(\sigma\alpha\beta) \ , \\ A^t u\alpha = D_u\alpha^2\beta\sigma &\iff A^t u = D_u v(\sigma\alpha\beta) \ . \end{aligned}$$

Call $\tau = \sigma\alpha\beta$. Recall that $x^t x = y^t y = 1$. Observe that $x^t D_y x = y^t D_x y$ (this holds for arbitary vectors x and y, see Property 4 below). Call $\gamma^2 = x^t D_y x = y^t D_x y$ (Recall that D_x and D_y are nonnegative or positive definite). Then we find that $u^t D_v u = v^t D_u v = \gamma^2/(\alpha^2\beta^2)$. Hence if we choose α and β such that $\gamma^2 = \alpha^2\beta^2$, we can go from (9) to (2).

This completes the proof. □

There are several properties of the solution that we would like to emphasize:

Property 1: Orthogonality

A useful characterization of the optimal solution can be obtained from (4) by observing that

$$l^t(\sum_{i=1}^{m} B_i b_i + B_0)y = 0 \implies \sum_{i=1}^{m}(a_i - b_i)b_i + l^t B_0 y = 0 \ .$$

If $B_0 = 0$ (as will be the case for the noisy realization problem, see below), this property says that the vector of residuals $a-b$ is orthogonal to b. This property will be used below to establish that certain methods do not deliver an L_2-optimal solution, since this orthogonality condition is not satisfied. Obviously, the orthogonality property is necessary, not sufficient for optimality.

Property 2: An equivalent minimization problem

An important observation is that, when D_y, or equivalently D_v is invertible, that

$$A^t D_v^{-1} A v = D_u v\tau^2 \ , \qquad v^t D_u v = 1 \ , \tag{14}$$

which implies that

$$\tau^2 = v^t A^t D_v^{-1} A v \tag{15}$$

is to be minimized, subject to

$$v^t D_u v = 1 \ . \tag{16}$$

Property 3: Scaling invariance

Observe that the object function (15) is scaling invariant, meaning that if we replace v by $w = v/\alpha$, that

$$v^t A^t D_v^{-1} A v = w^t A^t D_w^{-1} A w \,.$$

This is due to the fact that the elements of D_v are *quadratic* functions of the components of v.

Now compare the optimization problem (15)-(16) with expression (14) . It turns out that if we would *assume* that D_u and D_v were constant matrices (independent of u and v), the minimization of (15) subject to (16) would result in the generalized eigenvalue problem (14) . A nice feature is of course that now, even with the quadratic dependence of D_u and D_v on the elements of u and v, the same eigenvalue problem applies. It can however not be solved with 'classical linear' eigenvalue solvers, because the weighting matrices D_u and D_v depend on the components of u and v.

Equation (10) seems to suggest that $x^t D_y x = y^t D_x y$. As a matter of fact, we have:

Property 4: Normalization
With D_u and D_v as defined in Theorem 1, we have $u^t D_v u = v^t D_u v$ for arbitrary vectors u and v.

Property 5: Non-uniqueness of the vectors u and v
The vectors u and v that belong to the same value of τ in (2) are not unique. Indeed, replacing v by v/β and u by u/α with $\alpha\beta = 1$, results in

$$\begin{aligned} A(v/\beta) &= (D_v/\beta^2)(u/\alpha)\tau, & (v^t/\beta)(D_u/\alpha^2)(v/\beta) &= 1, \\ A^t(u/\alpha) &= (D_u/\alpha^2)(v/\beta)\tau, & (u^t/\alpha)(D_v/\alpha^2)(u/\alpha) &= 1, \end{aligned}$$

showing that the triplet $((u/\alpha), \tau, (v/\beta))$ also satisfies equations (2). The conclusion is that the 'direction' of u and v is uniquely determined but not their norm. Therefore, we have one more degree of freedom in imposing an additional constraint on the norm of u or v, which will be necessary in the algorithm to be presented now.

III. An inverse iteration algorithm

We now present an algorithm to solve the set of nonlinear equations (2). If D_u and D_v would be constant matrices independent of u and v, then the minimal eigenvalue could be computed with inverse iteration. This is exactly what we will do now: For given D_u and D_v, we perform one step of inverse iteration, using the QR-decomposition of the matrix A, in order to obtain new estimates of u and v. These are then used in updating D_u and D_v, etc.... In what follows, the iteration number will be indexed between square brackets. For the results upon convergence, we will use the index ∞. We will use the QR-decomposition of A:

$$A = \begin{pmatrix} \underbrace{Q_1}_{p\times q} & \underbrace{Q_2}_{p\times(p-q)} \end{pmatrix} \begin{pmatrix} \underbrace{R}_{q\times q} \\ 0 \end{pmatrix} \qquad (17)$$

A. Outline of the algorithm

We decompose u as $u = Q_1 z + Q_2 w$ for certain vectors $z \in \mathbb{R}^q$ and $w \in \mathbb{R}^{p-q}$. From (2) , we easily find that

$$\begin{pmatrix} R^t & 0 & 0 \\ Q_2^t D_v Q_1 & Q_2^t D_v Q_2 & 0 \\ Q_1^t D_v Q_1 \tau & Q_1^t D_v Q_2 \tau & -R \end{pmatrix} \begin{pmatrix} z \\ w \\ v \end{pmatrix} = \begin{pmatrix} D_u v \tau \\ 0 \\ 0 \end{pmatrix} \qquad (18)$$

This is a set of $(p+q)$ equations in $(p+q)$ unknowns. For the algorithm, we assume that D_u and D_v are constant, and solve the system of linear equations for z, w and v. The solution is easy because of the block triangular structure of the system matrix. Next u is computed from $u = Q_1 z + Q_2 w$, which is then used to update D_u and D_v. As a convergence test, we will reconstruct the affine matrix $B(b)$ from (3) and monitor the condition number of $B(b)$ in function of the iteration number. Let β_1 and β_q be the largest and smallest singular value of $B(b)$. Then numerical convergence occurs when $\frac{\beta_q}{\beta_1} \leq \epsilon_m$ where ϵ_m is the machine precision ($\epsilon_m = \sup \{ \epsilon \mid \text{fl}(1+\epsilon) = 1 \}$ where fl(.) denotes the floating point result of the expression between brackets).

Inverse iteration algorithm

Initialization: Choose $u^{[0]}, v^{[0]}, \tau^{[0]}$. Construct $D_{u^{[0]}}, D_{v^{[0]}}$ and normalize such that

$$(v^{[0]})^t D_{u^{[0]}} v^{[0]} = (u^{[0]})^t D_{v^{[0]}} u^{[0]} = 1 \,.$$

For $k = 1$ till convergence:

1. $z^{[k]} = R^{-t} D_{u^{[k-1]}} v^{[k-1]} \tau^{[k-1]}$
2. $w^{[k]} = -(Q_2^t D_{v^{[k-1]}} Q_2)^{-1} (Q_2^t D_{v^{[k-1]}} Q_1) z^{[k]}$
3. $u^{[k]} = Q_1 z^{[k]} + Q_2 w^{[k]}$
4. $v^{[k]} = R^{-1} Q_1^t D_{v^{[k-1]}} u^{[k]}$
5. $v^{[k]} = v^{[k]} / \|v^{[k]}\|$
6. $\gamma^{[k]} = ((u^{[k]})^t D_{v^{[k]}} u^{[k]})^{1/4}$
7. $u^{[k]} = u^{[k]}/\gamma^{[k]}$; $v^{[k]} = v^{[k]}/\gamma^{[k]}$
8. $D_{u^{[k]}} = D_{u^{[k]}}/(\gamma^{[k]})^2$; $D_{v^{[k]}} = D_{v^{[k]}}/(\gamma^{[k]})^2$
9. $\tau^{[k]} = (u^{[k]})^t A v^{[k]}$
10. Convergence test: Calculate $B^{[k]}$ using (3) and its largest and smallest singular value $\beta_1^{[k]}$ and $\beta_q^{[k]}$. If $\beta_q^{[k]}/\beta_1^{[k]} \geq \epsilon_m$, go to 1, else stop.

Remarks:

- One iteration corresponds precisely to solving the set of linear equations (18) with constant D_u and D_v, using the QR-decomposition of A.
- The normalization of v in Step 5 is necessary. Without it, the algorithm grows unstable in the sense that $\|v^{[k]}\| \to 0$ and $\|u^{[k]}\| \to \infty$ as $k \to \infty$ while $\|u^{[k]}\|\|v^{[k]}\|$ is constant. We can however 'regularize' the algorithm by normalizing v in an intermediate step. This is perfectly possible because of Property 5 in the previous section.

- A natural choice for the initial guess is the singular triplet of A, corresponding to the smallest singular value. This triplet would provide the solution if A were unstructured. This choice however does not guarantee convergence to a *global* minimum as will be shown with a numerical example below. The initial normalization of the vectors $u^{[0]}$ and $v^{[0]}$ is straightforward because of Property 4.
- A variation of this algorithm could be to perform $t > 1$ steps of inverse iteration with fixed D_u and D_v. However here we will only use $t = 1$.
- The algorithm as presented above is far from efficient. For instance, if p is much larger that q, the inversion of the $(p-q) \times (p-q)$ matrix $(Q_2^t D_v Q_2)$ requires a lot of work. If A and $B(b)$ are Hankel matrices (as is the case in this paper), we could apply inverse iteration to the generalized eigenvalue problem (14) (be careful with the normalization constraint on v, see below) and fully exploit the complexity reducing ideas described in [4], in which DFTs and FFTs are used to obtain fast inversion of D_v. Also, the convergence test could be replaced by a much more efficient criterion than the one we propose here. However, we need a precise convergence criterion as we want to show by a numerical example that 'classical' algorithms for the same problem do not converge to the optimal solution.

B. Discussion

We have no formal proof of the convergence of this algorithm. However, as the numerical experiments will demonstrate, if it convergences, its convergence rate seems to be linear. Intuitively, this can be understood when we assume that D_u and D_v are invertible as follows: Assume that D_u and D_v are factorized as $D_u = C_u^t C_u$ and $D_v = C_v^t C_v$ (for instance via Cholesky factorization). Then, since D_u and D_v are invertible, we can replace (2) by

$$\begin{aligned} (C_v^{-t} A C_u^{-1})\,(C_u v) &= (C_v u)\tau, & (u^t C_v^t)(C_v u) &= 1, \\ (C_u^{-t} A^t C_v^{-1})\,(C_v u) &= (C_u v)\tau, & (v^t C_u^t)(C_u v) &= 1\,. \end{aligned}$$

Obviously, we now obtain a (nonlinear) (ordinary) singular value decomposition. We could also eliminate the vector $u = D_v^{-1} A v/\tau$ in (2) and obtain the generalized eigenvalue problem (14). Using the factorization of D_u and D_v, we find

$$\begin{aligned} T_{uv}\,[C_u v] &= [C_u^{-t} A^t D_v^{-1} A C_u^{-1}][C_u v] = [C_u v]\tau^2\,, \\ [v^t C_u^t][C_u v] &= 1\,, \end{aligned}$$

with an obvious definition of T_{uv}. Note that T_{uv} is a symmetric positive definite matrix, which implies that all its eigenvalues are real positive. As we will see in the numerical examples below, $D_{u^{[k]}}$ and $D_{v^{[k]}}$ converge to constant matrices, which implies that also T_{uv} is approximately constant so that we are basically iterating with T_{uv}^{-1} in the inverse iteration algorithm. This observation implies that, asymptotically, the convergence rate is linear and will be governed by the two leading eigenvalues λ_1 and λ_2 of T_{uv}^{-1} (This follows from the well-known convergence properties of the classical power method). It directly follows that also $\log_{10} \|v^{[k]} - v^{[k-1]}\|$ decreases linearly as a function of the iteration number. A third implication is that $\log_{10}[\sigma_{min}(B^{[k]})]$ will decrease linearly as a function of the iteration number k, where $B^{[k]} = B_0 + \sum_{i=1}^{m} B_i b_i^{[k]}$ with $b_i^{[k]} = a_i - (u^{[k]})^t B_i v^{[k]} \sigma^{[k]}$.

IV. Rank deficient Hankel matrices

While the results of the previous section apply to general affinely structured matrices, from now on, we will concentrate on Hankel matrices. In Section IV.A, we describe the so-called noisy realization problem in terms of rank deficient Hankel matrices. We re-do the general proof with Lagrange multipliers for this special case to establish the structure of the weighting matrices D_u and D_v. When additional weights are introduced, one has to modify the structure of D_u and D_v as explained in Section IV.B. In Section IV.C, we show how to solve the approximation problem when the approximating model is of first order.

A. The noisy realization problem

Consider the problem of approximating a given data sequence $a \in \mathbb{R}^{p+q-1}$ by $b \in \mathbb{R}^{p+q-1}$ so as to minimize

$$\sum_{i=1}^{p+q-1} (a_i - b_i)^2 \quad \text{subject to} \quad \begin{array}{l} By = 0\,, \\ y^t y = 1\,, \end{array} \tag{19}$$

where B is a $p \times q$ Hankel matrix constructed from the elements of b. The rank deficiency of the Hankel matrix B ensures that b is the impulse response of a finite dimensional linear system of order $q-1$ at most. Hence, the number q of columns of the Hankel matrix B, which can be chosen by the user, will define the order of the approximating system, which is $q-1$ at most. The characteristic polynomial of the system that models the approximating sequence b is given by $y(z) = y_q z^{q-1} + \ldots + y_2 z + y_1$. Its roots are the poles of the approximating system. This implies that the z-transform of the sequence b will be of the form

$$b(z) = t(z)/y(z)\,, \tag{20}$$

for a certain polynomial $t(z)$ of degree smaller than q. If the sequence a is itself the impulse response of a higher dimensional system, our problem corresponds to *model reduction*. For $p \to \infty$ we get model reduction in the H_2-norm, which is treated in [7], using the z-transform and based on the insights obtained in [6] and this paper. If the sequence a is a given data sequence (which is not an impulse response, but for instance a noise corrupted one), one might consider this problem as a *noisy realization problem*.

The Lagrangean function for this problem is $\mathcal{L}(b,y,l) = \sum_{i=1}^{p+q-1}(a_i - b_i)^2 + l^t B y + \lambda(y^t y - 1)$ with B Hankel. Setting all derivatives to zero results in the set of equations (a convolution is denoted by a '$\star$'): $a - b = l \star y$, which means that

$$a_1 - b_1 \;=\; l_1 y_1$$

$$
\begin{aligned}
a_2 - b_2 &= l_1 y_2 + l_2 y_1 \\
a_3 - b_3 &= l_1 y_3 + l_2 y_2 + l_3 y_1 \\
\cdots \quad &\quad \cdots \\
a_{p+q-1} - b_{p+q-1} &= l_p y_q
\end{aligned}
$$

and in addition

$$B^t l = y\lambda \quad , \quad y^t y = 1 \quad , \quad By = 0 \, .$$

Note that we have $2p + 2q$ unknowns (the elements of b, l, y and λ) and exactly $2p + 2q$ equations. The first equation is a convolution which represents $p + q - 1$ equations. It is straightforward to find that $\lambda = 0$ because $l^t By = \lambda = 0$. Let B be the $p \times q$ Hankel matrix formed with the elements of b. Then

$$A - B = \left(\begin{matrix} l_1 & l_2 & \cdots & \cdots & \cdots & l_p & 0 & \cdots \\ l_2 & l_3 & \cdots & \cdots & l_p & 0 & 0 & \cdots \\ l_3 & l_4 & \cdots & l_p & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ l_p & 0 & \cdots & \cdots & 0 & l_1 & \cdots & \cdots \end{matrix} \right.$$

$$\left. \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & l_1 \\ 0 & l_1 & l_2 \\ \cdots & \cdots & \cdots \\ \cdots & l_{p-2} & l_{p-1} \end{matrix} \right) \times \begin{pmatrix} y_1 & y_2 & \cdots & y_q \\ 0 & y_1 & \cdots & y_{q-1} \\ 0 & 0 & \cdots & y_{q-2} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & y_1 \\ \cdots & \cdots & \cdots & \cdots \\ y_q & 0 & \cdots & 0 \\ y_{q-1} & y_q & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ y_2 & y_3 & \cdots & 0 \end{pmatrix} \qquad (21)$$

which means that the difference $A - B$ is the product of a Hankel and a Toeplitz matrix (note the 'circulant' structure in both matrices). A useful property of this factorization, when it is postmultiplied with a vector z is illustrated here for the case $p = 4, q = 3$:

$$\begin{pmatrix} l_1 & l_2 & l_3 & l_4 & 0 & 0 \\ l_2 & l_3 & l_4 & 0 & 0 & l_1 \\ l_3 & l_4 & 0 & 0 & l_1 & l_2 \\ l_4 & 0 & 0 & l_1 & l_2 & l_3 \end{pmatrix} \begin{pmatrix} y_1 & y_2 & y_3 \\ 0 & y_1 & y_2 \\ 0 & 0 & y_1 \\ 0 & 0 & 0 \\ y_3 & 0 & 0 \\ y_2 & y_3 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

$$\begin{pmatrix} z_1 & z_2 & z_3 & 0 & 0 & 0 \\ 0 & z_1 & z_2 & z_3 & 0 & 0 \\ 0 & 0 & z_1 & z_2 & z_3 & 0 \\ 0 & 0 & 0 & z_1 & z_2 & z_3 \end{pmatrix} \begin{pmatrix} y_1 & 0 & 0 & 0 \\ y_2 & y_1 & 0 & 0 \\ y_3 & y_2 & y_1 & 0 \\ 0 & y_3 & y_2 & y_1 \\ 0 & 0 & y_3 & y_2 \\ 0 & 0 & 0 & y_3 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{pmatrix}$$

$$= T_z \, T_y^t \, l \qquad (22)$$

where T_z and T_y are banded Toeplitz matrices with the elements of z and y. It shows that the Hankel-Toeplitz-vector product is converted into a Toeplitz-Toeplitz-vector product. We now use this property to elimate the matrix B. Postmultiplying $A - B$ with y results in $By = D_y l$ where D_y is $p \times p$ banded symmetric positive definite Toeplitz of the form $D_y = T_y T_y^t$. Hence, its elements are quadratic functions of the components of y. Similarly, we find by postmultiplying $A^t - B^t$ with l that $A^t l = D_l y$ where D_l is a $q \times q$ symmetric positive definite Toeplitz matrix of a form generated in the same way from the elements of l as D_y from the elements of y: $D_l = T_l T_l^t$. If we normalize l so that $l/\|l\| = x$ and $\|l\| = \sigma$, we have exactly the equations as in (9). Next we can renormalize x and y to finally obtain the equations as in (2).

B. Weighted rank deficient Hankel approximation

One might also consider to minimize a weighted error criterion as

$$\sum_{i=1}^{p+q-1} (a_i - b_i)^2 w_i \quad \text{subject to} \quad \begin{matrix} By = 0 \, , \\ y^t y = 1 \, , \end{matrix} \qquad (23)$$

with B Hankel and $w_i \in \mathbb{R}_0^+$ positive weights. Let $W = \text{diag}(w_i)$. Then, the solution is generated as in Theorem 1 by the generalized SVD problem (2) with

$$D_u = T_u W^{-1} T_u^t \, , \qquad D_v = T_v W^{-1} T_v^t \, .$$

The orthogonality property now becomes

$$(a - b)^t W b = 0 \, .$$

When the time horizon goes to infinity ($p \to \infty$) and when $w_i = i, i = 1, 2, \ldots$, we obtain the so-called Hilbert-Schmidt-Hankel norm (HSH-norm, which has certain interesting features, see [8]).

C. A special case: Approximation by first order systems

There is one special case for which the global optimum of (23) can be found explicitly: This is when one wants to approximate a given data sequence a by a first order LTI impulse response, which can be parametrized as $b_k = \alpha\beta^{k-1}$. We then obtain as a minimization problem (setting $q = 2$):

$$\min_{\alpha \in \mathbb{R}\, , \, \beta \in \mathbb{R}} \sum_{k=1}^{p+1} (a_k - \alpha\beta^{k-1})^2 w_k \, . \qquad (24)$$

Setting the derivatives with respect to α and β to zero, we obtain:

$$\frac{\partial}{\partial \alpha} = 0 \implies \sum_{k=1}^{p+1} (a_k - \alpha\beta^{k-1}) w_k (-\beta^{k-1}) = 0 \, ,$$

$$\frac{\partial}{\partial \beta} = 0 \implies \sum_{k=2}^{p+1} (a_k - \alpha\beta^{k-1}) w_k (-(k-1)\beta^{k-2}) = 0 \, .$$

Note that the first of these two equations is nothing else than the orthogonality property (with additional weightings w_k). The second equation states that also the sequence $\{(k-1)\beta^{k-2}\}$ is orthogonal to the weighted residual vector. Eliminating α from the first equation and substituting it in the second, we get the polynomial in β:

$$\begin{aligned} & [\sum_{k=2}^{p+1} (k-1) a_k w_k \beta^{k-2}][\sum_{k=1}^{p+1} \beta^{2k-2} w_k] \\ - & [\sum_{k=1}^{p+1} a_k \beta^{k-1} w_k][\sum_{k=2}^{p+1} (k-1) w_k \beta^{2k-3}] = 0 \end{aligned} \qquad (25)$$

This is a polynomial of degree $3(p+1)-5$ in β. One has to select the root which gives the minimum in (24).

V. Three other algorithms: If they converge, they do so to a suboptimal solution !

In this section, we discuss three well known algorithms that have been proposed in the literature to solve the noisy realization problem. The main conclusion will be that they converge to a suboptimal solution.

A. Cadzow's algorithm

Cadzow's algorithm [3] obtains a rank deficient Hankel matrix approximation to a given Hankel matrix. It is well known that the best rank deficient least squares approximation to a given matrix A can be obtained from the SVD of A [1]. If A is structured, for instance Hankel, the rank deficient approximation, obtained by truncating the dyadic decomposition of A to $q-1$ terms, will typically not have the required structure, or better, no structure at all. One could then try to find the best least squares fit to the obtained rank deficient matrix, that has the required structure. For Hankel matrices, it turns out that this fit is found by replacing the elements on the anti-diagonals by their average. This new matrix will however not be rank deficient anymore. Therefore one could again obtain the SVD, truncate it to find a rank deficient approximation etc.... This procedure can be summarized as follows:

Cadzow's algorithm

Initialization: Call $B^{[0]}=A$.

For $k=1$ till convergence:

1. Truncate the SVD of $B^{[k-1]}$ by omitting the dyad corresponding to the smallest singular value in the dyadic decomposition. Call the resulting rank deficient matrix $C^{[k]}$ (It will not have the Hankel structure).
2. Obtain $B^{[k]}$, the best least squares fit to $C^{[k]}$, which has Hankel structure (It will not be rank deficient).
3. Convergence test: Let $\beta_1^{[k]}$ and $\beta_q^{[k]}$ be the largest and smallest singular value of $B^{[k]}$. Verify whether $\beta_q^{[k]}/\beta_1^{[k]} \le \epsilon_m$. If so, stop, else, go to 1.

Cadzow in [3] formally proves that this process converges to a rank deficient Hankel matrix. Our numerical simulations show that it does so linearly. But nothing is known about its optimality. One could expect that Cadzow's algorithm minimizes $||A-B||_F^2$ over all rank deficient Hankel matrices B, but below we will show with a small numerical example that Cadzow's method does **not** converge to the optimal solution.

[1] It is interesting to see what happens to our equations (2) when A has no special structure. Then D_u and D_v reduce to identity matrices and the equations become $Av=u\tau, A^tu=v\tau, v^tv=1, u^tu=1$ which can only be satisfied for a singular triplet (u,τ,v) of A. The minimum corresponds to the singular triplet associated with the smallest singular value of A.

B. Iterative Quadratic Maximum Likelihood (IQML)

The fact that expression (14) is to be minimized with respect to v, has been observed before in e.g. [2] [4] [9]. In these papers, a linear constraint on v is used, such as the requirement that one of its components be 1, e.g.:

$$v_q = 1\,. \tag{26}$$

The Iterative Quadratic Maximum Likelihood (IQML) method then proceeds by minimizing in each step the quadratic form (15), where D_v is assumed to be constant, subject to the linear constraint (26). We now describe the basic algorithm of [2]. Define $C^{[k-1]}\in\mathbb{R}^{(q-1)\times(q-1)}, c^{[k-1]}\in\mathbb{R}^q$ and $\gamma^{[k-1]}\in\mathbb{R}$ as

$$\begin{pmatrix} C^{[k-1]} & c^{[k-1]} \\ (c^{[k-1]})^t & \gamma^{[k-1]} \end{pmatrix} = A^tD^{-1}_{v^{[k-1]}}A\,,$$

and partition $v^{[k]}$ accordingly as

$$v^{[k]}=\begin{pmatrix} w^{[k]} \\ 1 \end{pmatrix}.$$

Then the minimization of (15) subject to (26) results in

$$w^{[k]}=-(C^{[k-1]})^{-1}c^{[k-1]}\,,$$

and

$$\begin{aligned}(A^tD^{-1}_{v^{[k-1]}}A)v^{[k]} &= \begin{pmatrix} C^{[k-1]} & c^{[k-1]} \\ (c^{[k-1]})^t & \gamma^{[k-1]} \end{pmatrix}\begin{pmatrix} w^{[k]} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \gamma^{[k-1]}-(c^{[k-1]})^tC^{[k-1]}c^{[k-1]} \end{pmatrix}\end{aligned}$$

If the algorithm converges, then this equation is satisfied for $v^{[\infty]}, C^{[\infty]}, c^{[\infty]}, \gamma^{[\infty]}$ such that

$$\begin{aligned}(A^tD^{-1}_{v^{[\infty]}}A)v^{[\infty]} &= \begin{pmatrix} C^{[\infty]} & c^{[\infty]} \\ (c^{[\infty]})^t & \gamma^{[\infty]} \end{pmatrix}\begin{pmatrix} w^{[\infty]} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \gamma^{[\infty]}-(c^{[\infty]})^tC^{[\infty]}c^{[\infty]} \end{pmatrix}\end{aligned} \tag{27}$$

This is not at all equivalent with the generalized eigenvalue problem (14).

The confusion arises from a 'misinterpretation' of the scaling property. Indeed, the scaling or normalization of v has no effect on $v^tA^tD_v^{-1}Av$ (Property 3), but, one may not replace the constraint $v^tD_uv=1$ by another one (in this case the linear one (26)). The appearance of D_u is a crucial requirement for the optimum solution.

A possible variation is to replace the linear constraint (26) with a quadratic one as

$$v^tv=1\,. \tag{28}$$

If in each step, it is again assumed that D_v is constant, the minimization of (15) subject to (28) requires the calculation of the minimal eigenvalue $\lambda^{[k]}$ and corresponding eigenvector $v^{[k]}$, which will satisfy

$$(A^tD_{v^{[k-1]}}{}^{-1}A)\,v^{[k]}=v^{[k]}\lambda^{[k]}\,,\qquad (v^{[k]})^tv^{[k]}=1\,.$$

Upon convergence, we get the equation

$$(A^t D^{-1}_{v^{[\infty]}} A)\, v^{[\infty]} = v^{[\infty]} \lambda^{[\infty]} , \qquad (v^{[\infty]})^t v^{[\infty]} = 1 . \quad (29)$$

Again, this equation differs from the generalized eigenvalue problem (14). Comparing (29) to (14) shows that the weighting matrix D_u, which occurs in (14) is absent in (29). Said in other words, the two eigenvalue problems are fundamentally different [2] !

The conclusion is that, since our generalized eigenvalue (14) was obtained via a formal proof, IQML with a linear or quadratic constraint does **not** deliver the right solution! This will be confirmed by a numerical simulation below.

C. Steiglitz-McBride iteration

The Steiglitz-McBride algorithm is an iterative method for computing a rational approximation of a given data sequence. Originally, this method was formulated as a system identification technique based on input-output measurements. However, it also applies to the noisy impulse response realization problem. Let the model be as in (20). On the k-th iteration, an estimate $t^{[k]}(z)$ and $y^{[k]}(z)$ is obtained by minimizing the error criterion

$$e^{[k]} = \sum_{i=1}^{p+q-1} [y^{[k]}(z)[\frac{1}{y^{[k-1]}(z)}[a(i)]] - t^{[k]}(z)[\frac{1}{y^{[k-1]}(z)}[\delta(i)]]]^2$$

Typically, a constraint such as $y_q^{[k]} = 1$ is imposed to avoid the trivial zero solution. The notation $y^{[k]}[.]$ and $t^{[k]}[.]$ denotes filtering by finite impulse response (FIR) filters. Likewise, the operator $\frac{1}{y^{[k]}(z)}[a(k)]$ represents an all-pole inverse filter operation applied to the signal $a(k)$. $\delta(i)$ is an impulse and hence $\frac{1}{y^{[k-1]}(z)}[\delta(i)]$ is the impulse response of the all-pole filter. The error equation can be converted to matrix-vector form as

$$e^{[k]} = \|C^{[k-1]} y^{[k]} - H^{[k-1]} t^{[k]}\|_2^2 , \quad (30)$$

where $C^{[k-1]}$ and $H^{[k-1]}$ are lower triangular Toeplitz matrices made from the data sequence $a(k)$ and the impulse response of the all-pole filter $1/(y^{[k-1]}(z))$ (see [10] for full details).

Obviously, the minimization of (30) subject to a linear constraint on $y^{[k]}$ is a least squares problem. In [10] it is shown that at each iteration step, the solution is exactly equal to the solution given by IQML with a linear constraint. Hence, when applied to the noisy realization problem, the Steiglitz-McBride algorithm coincides with linearly constrained IQML. Since one of our conclusions so far was that IQML converges to a sub-optimal solution, the same observation applies for Steiglitz-McBride.

VI. Numerical examples

We discuss two numerical examples. The first one shows by a first order approximation that the methods of Cadzow, IQML and Steiglitz-McBride deliver suboptimal solutions. The second example illustrates the convergence behavior of the inverse iteration algorithm of this paper.

All simulations are done in MATLAB.

A. A simple counterexample

We take as a data sequence $a \in \mathbb{R}^6$ where $a^t = [\,6\;5\;4\;3\;2\;1\,]$, which we want to approximate with a first order linear system. (Any decreasing sequence of natural numbers can be modelled as the impulse response of an LTI system as $x_{k+1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} x_k$, $y_k = \begin{pmatrix} 1 & 0 \end{pmatrix} x_k$. For this data sequence, $x_0 = \begin{pmatrix} 6 & 1 \end{pmatrix}^t$.). The impulse response of the first order system (over this finite time horizon) can be parametrized as $b^t = [\,\alpha\;\alpha\beta\;\alpha\beta^2\;\alpha\beta^3\;\alpha\beta^4\;\alpha\beta^5\,]$. We will use the 5×2 Hankel matrix

$$A = \begin{pmatrix} 6 & 5 \\ 5 & 4 \\ 4 & 3 \\ 3 & 2 \\ 2 & 1 \end{pmatrix} .$$

In the first table, we take the unweighted case (i.e.all weights in (23) are equal to 1).

Some remarks that follow from Table I:

- It is clear that IQML and Steiglitz-McBride (equivalent with linearly constrained IQML) do not give the optimal order 1 approximation.
- When we apply the method of subsection IV.C to this example, we obtain a polynomial of degree 13, which has only one real root.
- The inverse iteration method of Section III converges to the exact solution in 9 iterations. Initial vectors are $(\,1\;1\;1\;1\;1\,)^t$ and $(\,1\;1\,)$ scaled appropriately, using property 4. At convergence, the orthogonality condition is satisfied (at least numerically): $(a-b)^t b \approx 10^{-15} \approx \epsilon_m$.
- IQML with a linear constraint requires 9 iterations for convergence while IQML with a quadratic constraint needs 19 iterations (both with the same starting vector $v^{[0]}$ as for inverse iteration).

[2] The fact that the normalization of an eigenvector and the presence of another matrix are important in eigenvalue-based optimization problems, can already be seen from a constant matrix case. First, consider the eigenvalue problem $Ax = Dx\lambda$. It's easy to see that the value of the quadratic form x^tAx where x is an eigenvector, depends on the normalization of x. Consider the example $A = \begin{pmatrix} 2.75 & 1.50 \\ 1.50 & 2.75 \end{pmatrix}$ and let D be $D = \begin{pmatrix} 0.5088 & 0.4715 \\ 0.4715 & 0.4637 \end{pmatrix}$. The eigenvalue decomposition $AX = DX\Lambda$ gives $X = \begin{pmatrix} 0.7741 & -0.6892 \\ 0.6331 & 0.7246 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 4.1669 & 0 \\ 0 & 70.3639 \end{pmatrix}$. Both eigenvectors are normalized to have norm 1. We now rescale the vectors of X to Y so that $Y^tDY = I_2$. Let's now check the quadratic forms: $X^tAX = \begin{pmatrix} 3.9702 & 0 \\ 0.0000 & 1.0019 \end{pmatrix}$. The minimum is obtained for the second column of X, the maximum for the first column. If we verify $Y^tAY = \begin{pmatrix} 4.1669 & 0 \\ 0 & 70.3639 \end{pmatrix}$, we see that the minimum is obtained for the first column of Y, the maximum for the second column!! Next, consider the optimization problems: $\min_{x \in \mathbb{R}^p} x^tAx$ subject to $x^tDx = 1$. This is equivalent with $\min_{x \in \mathbb{R}^p} \frac{x^tAx}{x^tDx}$ subject to $x^tx = 1$, for which the solution is given by the minimal eigenvalue satisfying $Ax = Dx\lambda$. This is **not** equivalent with the solution to $\min_{x \in \mathbb{R}^p} x^tAx$ subject to $x^tx = 1$, for which the solution is given by the minimal eigenvalue of A.

Next we consider the quadratic object function (23) with weights $w^t = [\ 1\ 2\ 2\ 2\ 2\ 1]$, i.e. we minimize $||A - B||_F^2$ subject to B Hankel and rank deficient. The results are summarized in Table II.

Remarks that follow from Table II:

- The inverse iteration method was started with initial vectors $u^{[0]^t} = [\ 1\ 1\ 1\ 1\ 1\ 1\ ]$ and $v^{[0]^t} = [\ 1\ 1\ ]$ (scaled appropriately) and required 10 iterations for convergence. One can verify that the orthogonality property $(a - b)^t Wb = 0$ is satisfied (numerically), where $W = \text{diag}(w)$.
- Cadzow's method converges in 89 steps. The corresponding 5×2 Hankel matrix is rank deficient. However, the orthogonality condition (which is necessary for L_2 optimality) is NOT satisfied for the Cadzow approximating sequence: $(a - b)^t Wb \approx 10^{-3} \not\approx \epsilon_m$.

B. Numerical example for the inverse iteration algorithm

Consider the impulse response

$$d(k) = 0.91^k \sin(0.9k) + 0.97^k \sin(0.7k+1)\ ,\ k = 1, 2, \ldots, 74, \quad (31)$$

and let it be perturbed by a white noise Gaussian random sequence $n(k)$ with zero mean and variance 0.2. The data that will be approximated consist of the sequence $a(k) = d(k) + n(k), k = 1, \ldots, 74$. We take $p = 70, q = 5$. In this example, we also demonstrate that the inverse iteration algorithm does not necessarily converge to the global minimum. There may be other stationary points as well. '*Estimate 1*' is obtained by taking as starting vectors, vectors with all ones, appropriately scaled. '*Estimate 2*' is obtained by taking as starting vectors the singular vectors of A, corresponding to the smallest singular value. The objective function for *Estimate 1* is smaller than that for *Estimate 2*. Some results and comments are shown in Figures 1, 2 and 3. The roots of the characteristic equation of the system (31) are $7.4190e-01 \pm 6.2489e-01i$ and $5.6567e-01 \pm 7.1283e-01i$. The results obtained from the estimates are summarized in Table III.

VI. Conclusions

In this paper, we have shown that some classical algorithms that were designed to solve the optimal L_2 noisy realization problem, do not generate L_2-optimal solutions.

Using Lagrange multipliers, we have derived necessary conditions for optimality and shown that these can be interpreted as a nonlinear generalized SVD problem. An inverse iteration method to solve it was proposed.

When the data come from the impulse response of a linear time invariant system, which is corrupted by additive Gaussian white noise, then the L_2 approximant as derived here, is the maximum likelihood estimate. Obviously, adding other linear constraints (such as e.g. the requirement that the polynomial $y(z)$ should be symmetric in its coefficients, a necessary condition for stability) present no additional conceptual nor algorithmic problems.

Many other applications for affinely structured matrices are analysed in [6].

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] Bresler Y., Macovski A, 'Exact maximum likelihood parameter estimation of superimposed exponential signals in noise', IEEE Trans. Acoust., Speech and Signal Processing, Vol.ASSP-34, no.5, pp.1081-1089, Oct. 1986.

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

[4] Clark M.P., Scharf L.L., ' On the complexity of IQML algorithms', IEEE Trans. SP., Vol.40, no.7, pp.1811-1813, 1992.

[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, pp.401-425, 1991.

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

[7] 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.

[8] Hanzon B., 'The area enclosed by the oriented Nyquist diagram and the Hilbert-Schmidt-Hankel norm of a linear system', IEEE Transactions on Automatic Control, Vol.AC-37, no.6, June 1992.

[9] Kumaresan R., Scharf L., Shaw A., 'An algorithm for pole-zero modeling and spectral estimation', IEEE Trans. ASSP, Vol.ASSP-34, pp.637-640, June 1986.

[10] McClellan J.H., Lee D., 'Exact equivalence of the Steiglitz-McBride iteration and IQML',IEEE Trans. SP, Vol.39, pp.509-512, Feb 1991.

[11] Steiglitz K., McBride L.E., ' A technique for identification of linear systems', IEEE Trans. Aut. Control., Vol.AC-10, pp.461-464, 1965.

Bart L.R. De Moor was born in Halle, Brabant, Belgium on July 12, 1960, and received his doctoral degree in Applied Sciences in 1988 at the Katholieke Universiteit Leuven, Belgium. He was a visiting research associate (1988-1989) at the Departments of Computer Science and Electrical Engineering of Stanford University, California, USA. Since October 1989 he is a senior research associate of the National Fund for Scientific Research of Belgium (NFWO) and associate professor at the Department of Electrical Engineering of the Katholieke Universiteit Leuven. In 1991-1992 he was the Chief of Staff of the Belgian National Minister of Science Ms.Wivina Demeester-DeMeyer. He received the Leybold-Heraeus Prize (1986), the Leslie Fox Prize (1989), the Guillemin-Cauer Best Paper Award (1990) of the IEEE Transactions on Circuits and Systems and became a Laureate of the Belgian Royal Academy of Sciences (1992).