ELSEVIER

Signal Processing 81 (2001) 1935–1945

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

IQML-like algorithms for solving structured total least squares problems: a unified view ☆

Philippe Lemmerling *,1, Leentje Vanhamme [2], Sabine Van Huffel, Bart De Moor

Department of Electrical Engineering, ESAT-SISTA/COSIC, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Heverlee, Leuven, Belgium

Received 15 May 2000; received in revised form 13 December 2000

Abstract

The structured total least squares (STLS) problem is an extension of the total least squares (TLS) problem for solving an overdetermined system of equations $Ax \approx b$. In many cases the extended data matrix $[A\ b]$ has a special structure (Hankel, Toeplitz, ...). In those cases the use of STLS is often required if a maximum likelihood (ML) estimate of x is desired. The main objective of this manuscript is to clarify the difference between several IQML-like algorithms—for solving STLS problems—that have been proposed by different authors and within different frameworks. Some of these algorithms yield suboptimal solutions while others yield optimal solutions. An important result is that the classicial IQML algorithm yields suboptimal solutions to the STLS problem. We illustrate this on a specific STLS problem, namely the estimation of the parameters of superimposed exponentially damped cosines in noise. We also indicate when this suboptimality starts playing an important role. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Structured total least squares; IQML; Parameter estimation

1. Introduction

The ordinary total least squares (TLS) method [11] is a frequently used method in parameter estimation problems. It can be formulated as follows:

$$\min_{x,\Delta A,\Delta b} ||[\Delta A\ \ \Delta b]||_F,$$

$$\text{such that } (A+\Delta A)x = b+\Delta b, \qquad (1)$$

where $||.||_F$ denotes the Frobenius norm, $A, \Delta A \in \mathbb{R}^{m\times n}$, $b, \Delta b \in \mathbb{R}^{m\times 1}$ and $x \in \mathbb{R}^{n\times 1}$. The matrix $S \equiv [A\ b]$ contains the measurements, whereas x is the parameter vector that characterizes the underlying linear(ized) system. In many signal processing applications the matrix S has a special structure (Hankel, Toeplitz, ...). A possible example might

☆ This work was supported by the FWO projects G.0256.97 and G.0240.99, the Research Communities ICCoS and ANMMM, the Brite/Euram Thematic Network NICONET, the IWT projects EUREKA 2063-IMPACT and STWW, the TMR Networks ALAPEDES and System Identification, the Belgian Programme on Interuniversity Poles of Attraction (IUAP-4/2 & 24), initiated by the Belgian State, Prime Minister's Office for Science, and by a Concerted Research Action (GOA) project of the Flemish Community, entitled "Mathematical Engineering for Information and Communication Systems Technology".

* Corresponding author. Tel.: +32-16-321796; fax: +32-16-321970.
E-mail address: philippe.lemmerling@esat.kuleuven.ac.be (P. Lemmerling).

[1] This author is supported by a post doctoral KUL scholarship.
[2] This author is supported by a post doctoral KUL scholarship.

PII: S0165-1684(01)00093-7

be that S is a Toeplitz matrix created by storing a signal vector $s \in \mathbb{R}^{(m+n)\times 1}$ in the first row and first column of S. Under the assumption that the error on s is independently identically distributed (i.i.d.) Gaussian noise, it is intuitively clear that $\Delta S \equiv [\Delta A\ \Delta b]$ should have the same structure as S if a maximum likelihood (ML) estimate of x is required (for a proof see [2]). This leads to a natural extension of the TLS problem called the structured total least squares (STLS) approach. The STLS problem can be formulated as follows:

$$\min_{x,\Delta s} \Delta s^{\mathrm{T}} W \Delta s, \quad \text{such that } (A + \Delta A)x = b + \Delta b$$
$$\text{and } [\Delta A\ \ \Delta b] \text{ has the same linear}$$
$$\text{structure as } [A\ \ b], \tag{2}$$

where $\Delta s \in \mathbb{R}^{q\times 1}$ contains the q *different* elements of $\Delta S \equiv [\Delta A\ \Delta b]$, and $W \in \mathbb{R}^{q\times q}$ is a weighting matrix, which for x to be a maximum likelihood estimate needs to have a special structure. We say that S is linearly structured if we can write it as follows: $S = \sum_{i=1}^{q} s(i)T_i$, where the $T_i \in \mathbb{R}^{m\times(n+1)}$, $i = 1,\dots,q$ are constant basis matrices and the elements of the vector $s \in \mathbb{R}^{q\times 1}$ represent the *different* elements of S (e.g. for a Toeplitz matrix, s would contain the $m+n$ different elements of S).

In the remainder of the paper we will adopt a Matlab like notation for vectors and matrices:

- $A(i,j)$: the entry in the jth column of the ith row of A.
- $A(i,:)$: the ith row of A.
- $A(:,j)$: the jth column of A.
- $A(p:q, r:s)$: the $(q-p+1)\times(s-r+1)$ submatrix of A containing the entries that belong to rows p till q *and* to columns r till s.
- $b(i)$: the entry on the ith row of column vector b.
- $b(p:q)$: the $(q-p+1)\times 1$ subvector of b containing the entries of row p till row q.
- $b(q:-1:p)$: this vector is equal to the previous one but with the elements in reversed order.

The transition from (1) to (2) can further be clarified by the following example. Suppose a noisy signal $s \in \mathbb{R}^{6\times 1}$ is measured. Knowing that the noiseless signal should satisfy a linear system of order 2 the following set of equations results:

$$Ax \approx b, \tag{3}$$

where $A \in \mathbb{R}^{4\times 2}$, $b \in \mathbb{R}^{4\times 1}$ and

$$[A\ \ b] \equiv \begin{bmatrix} s(3) & s(2) & s(1) \\ s(4) & s(3) & s(2) \\ s(5) & s(4) & s(3) \\ s(6) & s(5) & s(4) \end{bmatrix}.$$

Note that the "≈" symbol has to be used in (3), since s is the noisy and not the noiseless signal. Solving (3) in a TLS sense corresponds to solving (1) using the above defined matrix $[A\ b]$. Note that for this particular choice of $[A\ b]$, (1) can be rewritten as

$$\min_{x,\Delta s} \Delta s^{\mathrm{T}} W \Delta s, \quad \text{such that } (A + \Delta A)x = b + \Delta b,$$

$$[\Delta A\ \ \Delta b] = \begin{bmatrix} \Delta s(3) & \Delta s(2) & \Delta s(1) \\ \Delta s(6) & \Delta s(5) & \Delta s(4) \\ \Delta s(9) & \Delta s(8) & \Delta s(7) \\ \Delta s(12) & \Delta s(11) & \Delta s(10) \end{bmatrix}$$

with $W = I_{12\times 12}$ the 12×12 identity matrix. It is now very easy to understand that going from (1) to (2) consists of

(i) adding a constraint on the structure of $[\Delta A\ \Delta b]$ namely

$$[\Delta A\ \ \Delta b] \equiv \begin{bmatrix} \Delta s(3) & \Delta s(2) & \Delta s(1) \\ \Delta s(4) & \Delta s(3) & \Delta s(2) \\ \Delta s(5) & \Delta s(4) & \Delta s(3) \\ \Delta s(6) & \Delta s(5) & \Delta s(4) \end{bmatrix}.$$

Notice that in the STLS case $\Delta s \in \mathbb{R}^{6\times 1}$ instead of $\Delta s \in \mathbb{R}^{12\times 1}$ as in the TLS case.

(ii) introducing a general weighting matrix $W \in \mathbb{R}^{6\times 6}$.

The statistical meaning of W that appears in (2) is explained in Appendix A. However, intuitively it is also easy to understand that for (2) to be a ML estimator, W needs to be equal to the inverse of the noise covariance matrix [3]

$$R \equiv E(\Delta s \Delta s^{\mathrm{T}}),$$

[3] From formulation (2) it is clear that it is sufficient to know the inverse of the noise covariance matrix up to proportionality constant.

where $E(.)$ is the expected value operator. This is what is called a "prewhitening step" in many signal processing applications. In Section 3, we consider numerical examples where the data s consists of noiseless data perturbed by additive i.i.d. Gaussian noise. This explains the choice of $W = I$, since $R = I$. Summarizing we can say that the choice of W is determined by the statistical properties of the additive noise that perturbs the measurements.

In recent years many different formulations have been proposed for the STLS problem: the constrained total least squares (CTLS) approach [1,2], the structured total least norm (STLN) approach [8,10], the Riemannian singular value decomposition (RiSVD) approach [4,5] and the bootstrapped total least squares (BTLS) approach [9]. All these approaches start more or less from a formulation similar to (2), but the final formulation for which an algorithm is developed might be quite different. For example in the RiSVD approach problem (2) is reformulated into a nonlinear "Riemannian" SVD, which is then solved with an inverse iteration algorithm. In [6] it is proven—for the first three mentioned approaches—that, although the final formulations used in the different approaches are quite different, they are equivalent under mild conditions.

The main objective of this paper is to discuss different iterative quadratic maximum likelihood (IQML)-like algorithms for solving the STLS problem. Besides the classical IQML algorithm presented in [3], we will consider other IQML-like algorithms that originate from the CTLS, RiSVD and BTLS framework. All of these algorithms are very similar but nevertheless some of them yield suboptimal results for the STLS problem.

It is important to understand exactly what is meant here by "optimality of an algorithm". Therefore, we first look at the general form of the STLS problem (2). The STLS problem is clearly a constrained optimization problem with a quadratic cost function and nonlinear constraints. As will be indicated further on, the STLS problem (2) can be transformed into the equivalent formulation (4), which consists of a highly nonlinear cost function in the parameter vector y and a simple constraint on the same parameter vector. As pointed out in the next section, the cost function in (4) is scaling invariant, meaning that $y, 2y, \ldots$ $(y \neq 0)$ yield the same cost. The constraint on the parameter vector thus simply serves to select one of the solutions that lie in the direction where the cost function in (4) achieves its lowest value. An algorithm is now said to be optimal when the parameter vector y converges to the parameter vector y_{opt}, where y_{opt} is the parameter vector in which the *global minimum* of the cost function in (4) is reached, provided the initial estimate for y lies close enough to y_{opt}. It is important to make a distinction between the problem of determining good starting values and the question whether an algorithm yields optimal or suboptimal results. The starting value problem is discussed in [7] where it is shown that depending on the choice of the starting values even optimal STLS algorithms can get stuck in a local minimum. The latter solution could also be called "suboptimal", however the suboptimality we refer to, has nothing to do with the fact that an algorithm can end up in a local minimum. Finally, note that in Section 3 the Monte-Carlo simulations start from the optimal solution (which is known because we deal with simulation examples), precisely to avoid a mix up of the local minima problem and the suboptimality of the algorithms.

The paper is structured as follows. In Section 2 we present the different IQML-like algorithms for solving the STLS problem. It is explained why some of these algorithms are suboptimal. The last section illustrates how the choice of a suboptimal algorithm can affect the statistical accuracy of the obtained solution and provides some insight in the condition under which this suboptimality becomes more important.

2. The IQML-like algorithms

In this section we describe several IQML-like algorithms for solving STLS problems. The different algorithms originate from different approaches to the STLS problem.

In order to clarify the objective of this paper, we give a more extensive explanation of the previous two sentences. First of all, we will study algorithms for solving the STLS problem (2). Often, the STLS problem as formulated in (2) is reformulated and only then an algorithm for solving the newly obtained formulation is proposed. These different reformulations correspond to what we indicate as the *different approaches* or *frameworks* for solving the STLS problem. In the following subsections the CTLS, RiSVD and BTLS approaches will be considered. In this paper we will compare *IQML-like* algorithms for solving the formulations proposed in the different approaches.

2.1. CTLS approach

The CTLS approach described in [1,2] transforms the original STLS problem (2) into the following optimization problem: [4]

$$\min_{y\in\mathbb{R}^{(n+1)\times 1}} y^{\mathrm{T}}S^{\mathrm{T}}D_y^{-1}Sy, \quad y(n+1)=-1, \tag{4}$$

with $D_y \equiv H_yW^{-1}H_y^{\mathrm{T}}$, with $W\in\mathbb{R}^{q\times q}$ a weighting matrix, $H_y \equiv [T_1y\ T_2y\ \ldots\ T_qy]\in\mathbb{R}^{m\times q}$ where q is again the number of different entries in S (and thus also in ΔS, since S and ΔS have the same structure) and T_i, $i=1,\ldots,q$ are the so-called "basis matrices" (as defined in Section 1), i.e. they are used to construct the linearly structured matrix S starting from the vector $s\in\mathbb{R}^{s\times 1}$ that contains the *different* elements of S: $S=\sum_{i=1}^{q} s(i)T_i$.

Note that the objective function in (4) looks like a Rayleigh quotient. The difference with a classical Rayleigh quotient is the introduction of the matrix D_y, which depends on y and on the specific structure that needs to be preserved in the STLS problem. If no structure were imposed on ΔS in (2), we would obtain (1) but above all, D_y would [5] become $||y||_2^2$ and (4) would yield the eigenvector corresponding to the smallest eigenvalue of $S^{\mathrm{T}}S$ (since then (4) would indeed correspond to the minimization of the well-known Rayleigh quotient). Normally, we would use the SVD of S in order to determine the solution of the TLS problem. More specifically we would determine the right singular vector corresponding to the smallest singular value of S, but in theory we could also look for the eigenvector corresponding to the smallest eigenvalue of $S^{\mathrm{T}}S$. This is exactly what is done through the Rayleigh quotient minimization. Intuitively we can thus summarize the previous as follows. The solution of the STLS problem is found by minimizing a Rayleigh quotient-like cost function, in which the specific structure to be preserved in the STLS problem is reflected by the specific form of D_y.

As proven in [6], (4) and (2) are equivalent formulations in the sense that they yield the same parameter vector x (or $y(1:n)$ in the CTLS notation). Note that problem (4) basically is an unconstrained optimization problem, since the constraint can easily be substituted in the objective function. We then can use standard unconstrained optimization techniques for solving problem (4). Note however that the objective function is highly nonlinear and can have many local minima. Due to the equivalence of (4) and (2) this is a common problem to all algorithms solving the STLS problem [7].

In [1,2] a Newton algorithm using analytically calculated gradients and Hessians is proposed for solving (4).

Looking at the formula for D_y it is clear that the objective function in (4) is scaling invariant in y. Therefore, under mild conditions, [6] the following optimization problem is equivalent to (4) and thus also to (2):

$$\min_{y\in\mathbb{R}^{(n+1)\times 1}} y^{\mathrm{T}}S^{\mathrm{T}}D_y^{-1}Sy, \quad y^{\mathrm{T}}y=1. \tag{5}$$

We add this extra formulation since we will introduce a heuristic algorithm for solving the latter formulation in the next paragraph.

Due to the above mentioned equivalences, it should be clear now that if the optimization algorithm chosen to solve (4) or (5) converges to the global minimum, the obtained solution is the optimal solution of the STLS problem (2). The iterative quadratic maximum likelihood (IQML) algorithm was initially [3] designed for estimating the parameters of superimposed complex damped exponentials in noise. For this purpose a cost function similar to the objective function in (4) was

[4] Note that in the remainder of the paper only structures having nonsingular D_y matrices are considered. This is not an overly stringent condition, since many popular structures such as Hankel and Toeplitz matrices belong to this class.

[5] Simply write out the formula for D_y in case S is unstructured.

[6] The equivalence between (4) and (5) is only true when at the solution of the STLS problem $y(n+1)\neq 0$. The latter problems are so-called generic STLS problems. Most STLS problems belong to this class.

derived.[7] The same algorithm as in [3] can thus be used for solving the STLS problem. We present two versions of the IQML algorithm, corresponding to the different non-triviality constraints on the parameter vector.

IQML1 algorithm

Input: data matrix S, user-defined precision ε, structure that has to be preserved in the STLS problem (i.e. the formula for calculating D_y)

Output: the parameter vector x

Step 1: Initialize $y^{[0]} = arg\min_{y,\, y(n+1)=-1} y^{\mathrm{T}} S^{\mathrm{T}} S y$, $k = 0$

Step 2: $y^{[k+1]} = arg\min_{y,\, y(n+1)=-1} y^{\mathrm{T}} S^{\mathrm{T}} D_{y^{[k]}}^{-1} S y$

Step 3: if $\|y^{[k+1]} - y^{[k]}\|_2 < \varepsilon$ then $x = y(1:n)$ else $k = k+1$ and goto Step 2

The latter is a heuristic algorithm for solving problem formulation (4), because in every iteration formulation (4) is solved while considering D_y to be constant and only updated at the end of each iteration. Since $y(n+1) = -1$ it is obvious that Step 2 of the algorithm IQML1 corresponds to the following least squares (LS) problem:

$$\min_{y(1:n)} \|L^{[k]\mathrm{T}} S(:, 1:n) y(1:n) - L^{[k]\mathrm{T}} S(:, n+1)\|_2,$$

where $L^{[k]} L^{[k]\mathrm{T}} = D_{y^{[k]}}^{-1}$, $L^{[k]}$ and $L^{[k]\mathrm{T}}$ being the Cholesky factors of $D_{y^{[k]}}^{-1}$.

IQML2 algorithm

Input: data matrix S, user-defined precision ε, structure that has to be preserved in the STLS problem (i.e. the formula for calculating D_y)

Output: the parameter vector x

Step 1: Initialize $y^{[0]} = arg\min_{y,\, \|y\|_2=1} y^{\mathrm{T}} S^{\mathrm{T}} S y$, $k = 0$

Step 2: $y^{[k+1]} = arg\min_{y,\, \|y\|_2=1} y^{\mathrm{T}} S^{\mathrm{T}} D_{y^{[k]}}^{-1} S y$

Step 3: if $\|y^{[k+1]} - y^{[k]}\|_2 < \varepsilon$ then $x = -y(1:n)/y(n+1)$ else $k = k+1$ and goto Step 2

The IQML2 algorithm is a heuristic algorithm for solving problem (5). Again it is straightforward to see that Step 2 of the IQML2 algorithm corresponds to finding the eigenvector corresponding to the smallest eigenvalue of $S^T D_{y^{[k]}}^{-1} S$. The following lemma shows that IQML2 yields suboptimal results. A similar result can be constructed to prove the suboptimality of IQML1.

Lemma 2.1. *The result y obtained with algorithm IQML2 yields a suboptimal solution* $(y(1:n))$ *to the STLS problem.*

Let y_{IQML} represent the vector y obtained at convergence of the IQML2 algorithm. By definition y_{IQML} thus solves the following problem: $y_{\mathrm{IQML}} = arg\min_{y,\|y\|_2=1} y^{\mathrm{T}} S^{\mathrm{T}} D_{y_{\mathrm{IQML}}}^{-1} S y$. As mentioned before this means that y_{IQML} is the eigenvector of $S^{\mathrm{T}} D_{y_{\mathrm{IQML}}}^{-1} S$ that corresponds to its smallest eigenvalue. We will now show that y_{IQML} is not the optimal solution to (2). Therefore we apply the method of Lagrange multipliers to the equivalent problem (5). The Lagrangian of the latter problem is $G = y^{\mathrm{T}} S^{\mathrm{T}} D_y^{-1} S y - \lambda(y^{\mathrm{T}} y - 1)$. Differentiating G with respect to y and λ gives the following necessary conditions for y to be an optimal solution of (5) and thus of the STLS problem:

$$2S^{\mathrm{T}} D_y^{-1} S y - z_y = 2\lambda y$$

$$\text{with } z_y = \begin{bmatrix} y^{\mathrm{T}} S^{\mathrm{T}} D_y^{-1} \frac{\delta D_y}{\delta y(1)} D_y^{-1} S y \\ \vdots \\ y^{\mathrm{T}} S^{\mathrm{T}} D_y^{-1} \frac{\delta D_y}{\delta y(n+1)} D_y^{-1} S y \end{bmatrix} \tag{6}$$

$$y^{\mathrm{T}} y = 1. \tag{7}$$

Let us represent the optimal solution of (5) by y_{CTLS}. From (6) we clearly see that y_{CTLS} is not an eigenvector of $S^{\mathrm{T}} D_{y_{\mathrm{CTLS}}}^{-1} S$. Since we know that y_{CTLS} is an optimal solution of the STLS problem (2) and we just proved that y_{IQML} differs from y_{CTLS}, we can conclude that algorithm IQML2 yields suboptimal solutions to the STLS problem.

Remember that the objective function of problem (5) is scaling invariant and thus the constraint $y^{\mathrm{T}} y = 1$ only serves to select one of the infinite number of solutions. The latter implies that $\lambda = 0$ in (6) and (7). Thus it is easy to derive a heuristic algorithm, similar to IQML2, by changing Step 2

[7] As a matter of fact, writing out the objective function in (4) for the STLS problem in case a Toeplitz structure has to be fixed and $W = I$, yields the cost function described in [3].

into

$$y^{[k+1]} = -(2S^{\mathrm{T}}D_{y^{[k]}}^{-1}S)^{-1}z_{y^{[k]}},$$

$$y^{[k+1]} = y^{[k+1]}/||y^{[k+1]}||_2.$$

Using this as Step 2, it is seen that upon convergence the stationary condition for optimality of the STLS problem is satisfied.

Summarizing we can say that the classical IQML algorithms IQML1 and IQML2 yield suboptimal solutions to the STLS problem (2), because a term (namely z_y) in the differentiation of the objective function w.r.t. y in (5) is not taken into account. Mostly the correct ML arguments are invoked, leading to the correct STLS formulation for the problem at hand (namely (4) or (5)). In many cases however (see e.g. [13]), the next step consists of applying a suboptimal IQML algorithm to the optimal STLS problem formulation. Evidently the statistical accuracy of the estimates can seriously be degraded by this wrong choice of algorithm. This will be illustrated in Section 3. It will also be shown that depending on the circumstances, this suboptimality will lead to a big loss of statistical accuracy.

2.2. RiSVD and BTLS approach

As shown in [6] the RiSVD approach is equivalent to the STLS problem (2) under mild conditions. [8] The RiSVD approach is derived in [4,5] by using the technique of Lagrange multipliers. The result is the following equivalent problem formulation

Find the triplet (u, τ, v) corresponding the smallest

τ such that

$$Sv = D_v u\tau \quad u^{\mathrm{T}}D_v u = 1, \tag{8}$$

$$S^{\mathrm{T}}u = D_u v\tau \quad v^{\mathrm{T}}D_u v = 1, \quad v^{\mathrm{T}}v = 1 \tag{9}$$

with $D_v \equiv H_v W^{-1} H_v^{\mathrm{T}}, D_u \equiv H_u W^{-1} H_u^{\mathrm{T}}, W \in \mathbb{R}^{q\times q}$ a weighting matrix, $H_v \equiv [T_1 v\ T_2 v\ \dots\ T_q v] \in \mathbb{R}^{m\times q}$, $H_u \equiv [T_1^{\mathrm{T}}u\ T_2^{\mathrm{T}}u\ \dots\ T_q^{\mathrm{T}}u] \in \mathbb{R}^{n\times q}$, where q is again the number of different entries in S and T_i, $i = 1,\dots,q$ are the so-called "basis matrices" (as defined in Section 1), i.e. they are used to construct the linearly structured matrix S starting from the vector $s \in \mathbb{R}^{s\times 1}$ that contains the *different* elements of S: $S = \sum_{i=1}^{q} s(i)T_i$.

Note that the Eqs. (8) and (9) are very similar to the classical SVD equations. The difference with the classical SVD equations is the introduction of the matrices D_u and D_v, which depend on respectively u and v and on the specific structure that needs to be preserved in the STLS problem. If no structure were imposed on ΔS in (2), we would obtain (1) but above all, we would [9] obtain the classical SVD Eqs. (8) and (9). This should come as no surprise since the SVD is the standard way for solving the TLS problem. Intuitively we can summarize the previous as follows. The STLS problem is solved by determining the right singular vector corresponding to the smallest singular value of a nonlinear SVD (8) and (9), in which the specific structure to be preserved in the STLS problem is reflected by the specific form of D_u and D_v.

It is clear that the following two equations can be derived from and are equivalent to (8) and (9):

$$S^{\mathrm{T}}D_v^{-1}Sv = D_u v\tau^2, \quad v^{\mathrm{T}}v = 1, \tag{10}$$

$$u = D_v^{-1}Sv/\tau, \quad u^{\mathrm{T}}D_v u = 1. \tag{11}$$

As suggested in [5] (10) and (11) can be used to solve the STLS problem (2) in an iterative way. As a matter of fact we see that—at least for constant u—(10) is a "nonlinear" generalized eigenvalue problem of which we have to find the eigenvector v corresponding to the smallest eigenvalue τ^2.

The latter observation has lead to two different algorithms: one developed in the RiSVD [5] framework—further referred to as the RiSVD generalized eigenvalue (RiSVD-GE) algorithm—and another one developed in the BTLS [9] framework and further referred to as the BTLS generalized eigenvalue (BTLS-GE) algorithm. An outline of these two algorithms follows.

---

[8] The condition being again that the STLS problem has to be generic.

[9] Simply write out the formulas for D_u and D_v in case S is unstructured.

RiSVD-GE algorithm
Input: data matrix S, user-defined precision ε, structure that has to be preserved in the STLS problem (i.e. the formulas for calculating D_u and D_v)
Output: the parameter vector x

Step 1: Initialize $(u^{[0]}, \tau^{[0]}, v^{[0]})$ with the triplet corresponding to the smallest singular value of S; $k=0$

Step 2: $v^{[k+1]} = (S^{\mathrm{T}} D_{v^{[k]}}^{-1} S)^{-1} D_{u^{[k]}} v^{[k]} (\tau^{[k]})^2$

$v^{[k+1]} = v^{[k+1]}/||v^{[k+1]}||_2$
$u^{[k+1]} = D_{v^{[k+1]}}^{-1} S v^{[k+1]}/\tau^{[k]}$

$\gamma = (u^{[k+1]})^{\mathrm{T}} D_{v^{[k+1]}} u^{[k+1]}$
$u^{[k+1]} = u^{[k+1]}/\gamma^{1/2}$
$\tau^{[k+1]} = u^{[k+1]\mathrm{T}} S v^{[k+1]}$

Step 3: if $||v^{[k+1]} - v^{[k]}||_2 < \varepsilon$ then
$x = -v(1:n)/v(n+1)$
else $k = k+1$ and goto Step 2

BTLS-GE algorithm
Input: data matrix S, user-defined precision ε, structure that has to be preserved in the STLS problem (i.e. the formulas for calculating D_u and D_v)
Output: the parameter vector x

Step 1: Initialize $u^{[0]}$ and $v^{[0]}$ with the left respectively right singular vector corresponding to the smallest singular value of S; $k=0$

Step 2: Solve the following generalized eigenvalue problem
$S^{\mathrm{T}} D_{v^{[k]}}^{-1} S v = D_{u^{[k]}} v \tau^2$
for the eigenvector $v^{[k+1]}$ corresponding to the smallest eigenvalue τ
$v^{[k+1]} = v^{[k+1]}/||v^{[k+1]}||_2$
$u^{[k+1]} = D_{v^{[k+1]}}^{-1} S v^{[k+1]}/\tau$

Step 3: if $||v^{[k+1]} - v^{[k]}||_2 < \varepsilon$ then
$x = -v(1:n)/v(n+1)$
else $k = k+1$ and goto Step 2

If we compare the RiSVD-GE and BTLS-GE algorithm to the IQML2 algorithm, we see that the former two solve a *generalized* eigenvalue problem whereas the IQML2 algorithm solves an ordinary eigenvalue problem. The other major difference between the RiSVD-GE and the BTLS-GE algorithm on the one hand and the IQML2 algorithm on the other hand is the introduction of the vector u in the former two. As shown in [5] this vector u is in fact a vector of Lagrange multipliers. From the previous subsection we know that algorithm IQML2 yields suboptimal results. Upon convergence of algorithm RiSVD-GE (and also algorithm BTLS-GE), the first order optimality conditions—namely (8) and (9)—are satisfied, and thus both algorithms yield optimal results.

Notice that the difference between the RiSVD-GE and the BTLS-GE algorithm mainly consists in the frequency in which D_u is updated. In Step 2 of the BTLS-GE algorithm the generalized eigenvalue problem is solved *completely* before D_u is updated, whereas in the RiSVD-GE algorithm D_u is updated after each step of the inverse iteration algorithm, used for solving the generalized eigenvalue problem. This is also an intuitive explanation for the convergence problems observed in the case of the BTLS-GE algorithm: both algorithms are alternating coordinates optimization algorithms, but in the case of RiSVD-GE, the alternation between u and v is more frequent.

Table 1 summarizes this section. It shows the framework from which the different IQML-like algorithms originate, the kernel problem (EVD = eigen value decomposition and GEVD = generalizedEVD) that has to be solved in each iteration and whether the obtained results are optimal or not.

Table 1
This table summarizes the different IQML-like algorithms for solving the STLS problem. It shows the framework from which the algorithms originate, the kernel problem (EVD = eigen value decomposition and GEVD = generalized EVD) that has to be solved in each iteration and whether the obtained results are optimal or not.

Approach	IQML-like algorithm	Kernel problem	Optimal?
CTLS	IQML1	LS	No
	IQML2	EVD	No
RiSVD	RiSVD-GE	GEVD	Yes
BTLS	BTLS-GE	GEVD	Yes

3. Numerical experiments

In this section we illustrate the suboptimality of the IQML2 algorithm on a small example of a special STLS problem: the estimation of the parameters of superimposed exponentially damped cosines in i.i.d. Gaussian noise. We first briefly explain why the latter is an STLS problem.

Let $u\in\mathbb{R}^{(m+n)\times 1}$ be a vector that is a sum of exponentially damped cosines: $u(i)=\sum_{k=1}^{K/2} a_k e^{d_k(i-1)\Delta t}\cos(2\pi f_k(i-1)\Delta t + p_k)$, $i=1,\dots,m+n$, where Δt is the sampling interval and chosen equal to 1 in this example. When u is placed in an $(m+n-K)\times(K+1)$ Toeplitz matrix (with first row $u(K+1:-1:1)$ and first column $u(K+1:m+n)$), it is obvious that this Toeplitz matrix will be rank deficient since all $K+1$ consecutive samples of u satisfy a linear prediction equation (this is due to the fact that u is a sum of exponentially damped cosines). Furthermore, the parameters of the exponentially damped cosines can be derived from the prediction error coefficients in a similar way as described in [3] (the prediction error coefficients are the elements of the null vector of the rank deficient Toeplitz matrix).

We now consider the noisy signal case. Let $e\in\mathbb{R}^{(m+n)\times 1}$ be a noise vector containing i.i.d. Gaussian noise entries of standard deviation σ_e. The goal is now to determine the signal parameters (i.e. $a_k,\ d_k,\ f_k,\ p_k,\ k=1,\dots,K/2$) starting from the noisy measurement vector $s\ \equiv\ u+e$. From the previous it should be clear that under the given noise circumstances, maximum likelihood (ML) estimates of the parameters can be obtained in the following way: store the noisy signal s in a Toeplitz matrix S (with first row $s(K+1:-1:1)$ and first column $s(K+1:m+n)$) and solve the STLS problem (2) in which the Toeplitz structure of S is preserved and W is the identity matrix. The resulting vector $[x^{\mathrm{T}}\ -1]^{\mathrm{T}}$ allows us to determine the signal parameters. Note that this specific STLS problem is thus completely equivalent to methods as described e.g. in [12], where through the use of a different parametrization this STLS problem is reformulated into a nonlinear least squares minimization. For finding the true solution of the STLS problem, we will use algorithm RiSVD-GE. We first consider a small example.

Example 3.1. $m=4$, $n=2$, $K=2$, $f_1=0.3$, $d_1=-0.08$, $a_1=2$, $p_1=0.2$, $\sigma_e=0.6$.

To illustrate the suboptimality of the IQML2 algorithm we compute $2S^{\mathrm{T}}D_y^{-1}Sy - z_y$ for both y_{IQML} (the solution vector obtained with algorithm IQML2) and $y_{\mathrm{RiSVD\text{-}GE}}$ (the solution vector obtained with algorithm RiSVD-GE) for one realization of the noisy signal s from Example 3.1. When filling in y_{IQML} this yields $[0.0756\ -0.1193\ -0.0015]^{\mathrm{T}}$ and when filling in $y_{\mathrm{RiSVD\text{-}GE}}$ we get $10^{-13}[-0.2626\ 0.0366\ 0.1998]^{\mathrm{T}}$, which clearly illustrates the suboptimality of the IQML2 result because the resulting vector should be $[0\ 0\ 0]^{\mathrm{T}}$ (see (6)). The result of algorithm RiSVD-GE, shows that this algorithm solves the STLS problem in an optimal way.

To show the effect that this suboptimality can have on a larger example, we consider a new example, containing the component of the previous example and an additional one.

Example 3.2. f_1, a_1, d_1, and p_1 as in Example 3.1.
$d_2=0, a_2=1,\ p_2=0.1$. Furthermore $m=54$, $n=4$, $K=4$ and again $\sigma_e=0.6$.

Using this example, we perform 500 Monte-Carlo simulations, with both the IQML2 and RiSVD-GE algorithm. Since we know the exact parameters, we can calculate the "true" y vector, represented by y_{ex}. As mentioned in Section 2.1 the STLS problem is highly nonlinear and as a result it can have many local minima. Since the goal of this paragraph is to estimate the statistical accuracy of the algorithms IQML2 and RiSVD-GE and not their sensitivity w.r.t. the choice of initial values, the initial values are set equal to the exact values (obtained from y_{ex}). In this way, we avoid that suboptimality resulting from different local minima interferes with our experiment.

To get an idea of the statistical performance of both methods, we average the following relative errors over the 500 runs: $||y_{\mathrm{IQML}}-y_{\mathrm{ex}}||_2/||y_{\mathrm{ex}}||_2$ and $||y_{\mathrm{RiSVD\text{-}GE}}-y_{\mathrm{ex}}||_2/||y_{\mathrm{ex}}||_2$. To illustrate when the statistical suboptimality starts playing an important role, f_2 from Example 3.2 is varied from 0.301 to 0.329 in steps of 0.001. The resulting relative errors are shown in Fig. 1 (the dashed line represents the relative errors of the solutions obtained with algorithm IQML2, whereas the full line represents the

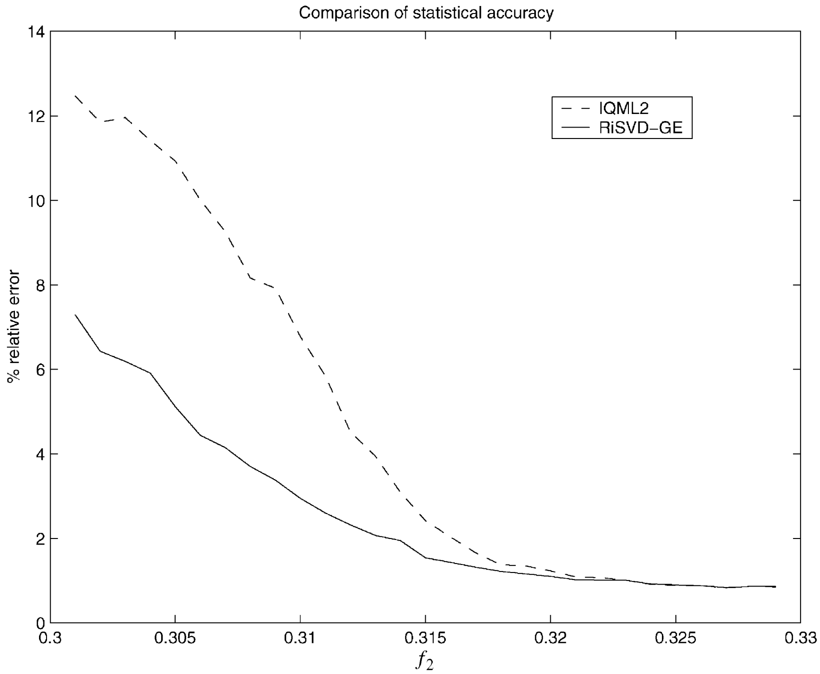

Fig. 1. This figure illustrates the difference in statistical accuracy obtained with algorithm IQML2 (dashed line) and algorithm RiSVD-GE (full line). It shows the relative errors (in %) of the solutions obtained by the algorithms as a function of the parameter f_2. Note that the first component lies at $f_1 = 0.3$ and thus the closer f_2 gets to f_1, the bigger the difference in statistical accuracy.

relative errors of the solutions obtained with algorithm RiSVD-GE).

Looking at Fig. 1, we see that when f_2 gets close to f_1 (i.e. when f_2 gets close to 0.3; note that the latter means that the difficulty of the parameter estimation problem increases), the suboptimal algorithm IQML2 performs a lot worse than the optimal RiSVD-GE algorithm. Therefore, the figure clearly shows that for high-frequency resolution applications, the suboptimality of IQML2 starts playing an important role.

4. Conclusion

We have given an overview of different existing IQML-like algorithms for solving the STLS problem. We have proven the suboptimality of the classical IQML algorithms (IQML1 and IQML2) for solving STLS problems. This suboptimality is shown to affect the statistical performance of the classical IQML algorithms (i.e. IQML1 and IQML2) when compared to the optimal STLS algorithm RiSVD-GE. Furthermore it is shown that in very demanding applications—e.g. when high-frequency resolution is needed—the suboptimality of suboptimal IQML algorithms starts playing an important role.

Appendix A.

In order to discuss the statistical properties of the STLS estimator (2), we first have to define the statistical "measurement" model that we consider. For the ordinary TLS case the corresponding model was shown to be the classical errors-in-variables (EIV) model (see [11]). The classical EIV model is described by

$$b_0 = A_0 x_0$$

$$\text{with } b = b_0 + \Delta b_0 \text{ and } A = A_0 + \Delta A_0, \qquad (A.1)$$

where $[A_0\ b_0]$ contains the true unobservable quantities, $[A\ b]$ contains the measured quantities and $[\Delta A_0\ \Delta b_0]$ contains the random variables corresponding to the noise on the measurements. Furthermore, the rows of the matrix $[\Delta A_0\ \Delta b_0]$ are assumed to be i.i.d. with common zero mean vector and common covariance matrix $C = \sigma_v^2 I_{n+1}$, with σ_v^2 unknown.

To demonstrate the statistical properties of the STLS estimator, we consider a similar model, based on the same equations as in (A.1) but with different statistical assumptions on the elements of $\Delta S_0 = [\Delta A_0\ \Delta b_0]$. As an example we consider the statistical measurement model for the Hankel STLS problem (i.e. an STLS problem (2) in which ΔS needs to preserve the Hankel structure of S). In the latter case, a vector of samples $s \in \mathbb{R}^{q\times 1}$ is measured and afterwards a Hankel matrix S is constructed using this measured vector, simply by storing it in the first column and last row of the Hankel matrix. It is obvious that in the statistical measurement model for this STLS problem, ΔS_0 should have a Hankel structure too. Furthermore ΔS_0 can be represented by a vector $\Delta s_0 \in \mathbb{R}^{q\times 1}$ in a similar way as S can be represented by s. Thus, along the antidiagonals of ΔS_0 the same random variables occur, thereby violating the independency assumption between the rows of the EIV model. Furthermore the random variables in Δs_0 are assumed to be i.i.d. with covariance matrix R.

Having proposed the model and the assumptions on the "measurement" errors, we can now derive a ML estimator for the model parameters in x. For ease of notation we let $y = [x^{\mathrm{T}}\ -1]^{\mathrm{T}}$. If the matrix $S = [A\ b]$ contains the measured values, we know that the unobservable true values should obey a linear relation or

$$(S - \Delta S_0)y = Sy - H_y \Delta s_0 = 0$$

or

$$Sy = H_y \Delta s_0 \equiv e, \qquad (A.2)$$

where H_y is defined by $H_y \Delta s_0 \equiv \Delta S_0 y$.[10] The e defined in the last equation can be seen as an "equation error". Since the elements of e are linear combinations of Gaussian random variables, they obey Gaussian distributions themselves. It is then well known that a ML estimate for y can be obtained by minimizing

$$e^{\mathrm{T}} U^{-1} e = y^{\mathrm{T}} S^{\mathrm{T}} U^{-1} S y,$$

where U is the covariance matrix of e and thus

$$U = E(ee^{\mathrm{T}}) = H_y E(\Delta s_0 \Delta s_0^{\mathrm{T}}) H_y^{\mathrm{T}} = H_y R H_y^{\mathrm{T}},$$

where E is the expected value operator and the last equation followed from the noise assumptions we made in our model about Δs_0. Summarizing we see that for the proposed model, a ML estimate for y can be obtained from the observations in S by solving

$$\min_y y^{\mathrm{T}} S^{\mathrm{T}} (H_y R H_y^{\mathrm{T}})^{-1} S y, \qquad (A.3)$$

where $y = [x^{\mathrm{T}}\ -1]^{\mathrm{T}}$. If we compare (4) and (A.3) we clearly see that the ML estimate is thus obtained by solving the STLS problem involving S, where W is set to R^{-1}.

[10] Note that such a matrix H_y always exists due to the fact that ΔS_0 is a linearly structured matrix.

References

[1] T.J. Abatzoglou, J.M. Mendel, Constrained total least squares, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, 1987, pp. 1485–1488.

[2] T.J. Abatzoglou, J.M. Mendel, G.A. Harada, The constrained total least squares technique and its applications to harmonic superresolution, IEEE Trans. Signal Process. 39 (1991) 1070–1086.

[3] Y. Bresler, A. Macovski, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, IEEE Trans. Acoust. Speech Signal Process. 34 (5) (October 1986) 1081–1089.

[4] B. De Moor, Structured total least squares and l_2 approximation problems, in: Van Dooren, Ammar, Nichols, Mehrmann (Eds.), Linear Algebra and its Applications, Special Issue on Numerical Linear Algebra Methods in Control, Signals and Systems, vols. 188–189, July 1993, pp. 163–207.

[5] B. De Moor, Total least squares for affinely structured matrices and the noisy realization problem, IEEE Trans. Signal Process. 42 (November 1994) 3004–3113.

[6] P. Lemmerling, B. De Moor, S. Van Huffel, On the equivalence of constrained total least squares and structured total least squares, IEEE Trans. Signal Process. 44 (11) (November 1996) 2908–2910.
[7] P. Lemmerling, S. Van Huffel. Analysis of the structured total least squares problem for Hankel/Toeplitz matrices, Technical Report Internal Report 98-100, ESAT-SISTA/COSIC, K.U. Leuven, 1998.
[8] J.B. Rosen, H. Park, J. Glick, Total least norm formulation and solution for structured problems, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 110–128.
[9] H. Van Hamme, Identification of linear systems from time-or frequency-domain measurements, Ph.D. Thesis, Vrije Universiteit Brussel, Department of Electrical Engineering, ELEC, 1992.
[10] S. Van Huffel, H. Park, J.B. Rosen, Formulation and solution of structured total least norm problems for parameter estimation, IEEE Trans. Signal Process. 44 (10) (1996) 2464–2474.
[11] S. Van Huffel, J. Vandewalle. The Total Least Squares Problem: Computational Aspects and Analysis, Vol. 9, SIAM, Philadelphia, 1991.
[12] L. Vanhamme, A. van den Boogaart, S. Van Huffel, Improved method for accurate and efficient quantification of mrs data with use of prior knowledge, J. Magn. Reson. 129 (1997) 35–43.
[13] G. Zhu, W.Y. Choy, B.C. Sanctuary, Spectral parameter estimation by an iterative quadratic maximum likelihood method, J. Magn. Reson. 135 (1998) 37–43.