43rd IEEE Conference on Decision and Control
December 14-17, 2004
Atlantis, Paradise Island, Bahamas

Application of Structured Total Least Squares for System Identification

Ivan Markovsky, Jan C. Willems, Sabine Van Huffel, Bart De Moor, and Rik Pintelon

Abstract— **The following identification problem is considered: minimize the ℓ_2 norm of the difference between a given time series and an approximating one under the constraint that the approximating time series is a trajectory of a linear time invariant system of a fixed complexity. The complexity is measured by the input dimension and the maximum lag. The problem is known as the global total least squares and alternatively can be viewed as maximum likelihood identification in the errors-in-variables setup. Multiple time series and latent variables can be considered in the same setting.**

The identification problem is related to the structured total least squares problem. The paper presents an efficient software package that implements the theory in practice. The proposed method and software are tested on data sets from the database for the identification of systems DAISY.

Index Terms— **Errors-in-variables, system identification, structured total least squares, numerical software, DAISY.**

I. Introduction

A. The structured total least squares problem

The *structured total least squares (STLS)* problem originates [2], [3] from the signal processing and numerical linear algebra communities and is not widely known in the area of systems and control. It is a generalization to matrices with structure of the total least squares problem [4], [5], known in the system identification literature as the Koopmans–Levin's method [6]. In this paper, we show the applicability of the STLS method for system identification. We extend previous results [7], [8] of the application of STLS for SISO system identification to the MIMO case and present numerical results on data sets from DAISY [9].

The STLS problem is defined as follows: given a time series w and a structure specification $\mathscr{S}$, find the global minimum point of the optimization problem

$$\min_{X}\left(\min_{\hat{w}} \|w-\hat{w}\|_{\ell_2}^2 \quad \text{s.t.} \quad \mathscr{S}(\hat{w})\begin{bmatrix} X \\ -I \end{bmatrix} = 0\right). \quad (1)$$

The constraint of (1) enforces the structured matrix $\mathscr{S}(\hat{w})$ to be rank deficient, with rank at most $\mathrm{row\,dim}(X)$. The cost function measures the distance of the given data w to its approximation $\hat{w}$. Thus the STLS problem aims at optimal structured low rank approximation of $\mathscr{S}(w)$ by $\mathscr{S}(\hat{w})$.

An extended version of this paper is submitted for publication to *IEEE Trans. on Automatic Control*, see [1].

Ivan Markovsky, Jan C. Willems, Sabine Van Huffel, and Bart De Moor are with the Electrical Engineering Department, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
{Ivan.Markovsky,Jan.Willems}@esat.kuleuven.ac.be
{Sabine.VanHuffel,Bart.DeMoor}@esat.kuleuven.ac.be
Rik Pintelonis is with Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium, Rik.Pintelon@vub.ac.be

B. The global total least squares problem

Let $\mathscr{M}$ be a user specified model class and w a given time series. We view a model $\mathscr{B} \in \mathscr{M}$ as a collection of legitimate time series. The more the model forbids from the universe of possible time series, the less complex and therefore more powerful it is. The model class restricts the maximal allowed model complexity. Within $\mathscr{M}$, we aim to find the model $\hat{\mathscr{B}}$ that best fits the data according to the misfit criterion

$$M(w,\mathscr{B}) := \min_{\hat{w}\in\mathscr{B}} \|w-\hat{w}\|_{\ell_2}^2.$$

The resulting optimization problem $\min_{\mathscr{B}\in\mathscr{M}} M(w,\mathscr{B})$ is known as the *global total least squares* problem [10]. The described system identification framework is put forward by Willems in [11].

Our approach of solving the global total least squares problem, is different from the one of Roorda and Heij [10], [12]. We relate the identification problem to the STLS problem (1) and subsequently use solution methods developed for the STLS problem, while different algorithms are developed in the framework of the global total least squares problem. Another difference between the approach of this paper and the approach of Roorda and Heij is that we use a kernel representation of the system, while in [10], [12] a state space representation with driving input is used.

C. Outline of the paper

Section II gives background material on LTI systems described by kernel representation. Section III defines and solves the identification problems. Section IV describes some extensions of the identification problem. Section V shows results of the proposed method on data sets from DAISY, and Section VI gives conclusions.

II. Preliminaries

A. Kernel representation

Consider a time series $w := \big(w(1),\ldots,w(T)\big)$ with $\mathtt{w}$ variables. A block-Hankel matrix with l block rows, constructed from w, is denoted by

$$\mathscr{H}_l(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-l+1) \\ w(2) & w(3) & \cdots & w(T-l+2) \\ \vdots & \vdots & & \vdots \\ w(l) & w(l+1) & \cdots & w(T) \end{bmatrix}.$$

The time series w satisfies the set of difference equations

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_l w(t+l) = 0, \quad \text{for } t = 1,\ldots,T-l \quad (2)$$

0-7803-8682-5/04/$20.00 ©2004 IEEE

with maximum l delays or *lags*, if and only if

$$R\mathscr{H}_{l+1}(w) = 0, \quad \text{where } R := \begin{bmatrix} R_0 & R_1 & \cdots & R_l \end{bmatrix}.$$

The rank deficiency of $\mathscr{H}_{l+1}(w)$ is related to the existence of an LTI system Σ that "explains" the data and the Hankel structure is related to the dynamic nature of the model.

In the behavioral approach to system theory [13], (2) is called a *kernel representation* of the system Σ. A more compact notation is

$$R(\sigma)w = 0, \quad \text{where } R(\xi) := \sum_{i=0}^{l} R_i \xi^i, \tag{3}$$

and σ is the backwards shift operator, $(\sigma w)(t) = w(t+1)$. Let $\mathscr{B}$ be the set of all trajectories of a system Σ, described by (3), *i.e.*,

$$\mathscr{B} := \{ w : \mathbb{Z} \to \mathbb{R}^{\mathbb{w}} \mid R(\sigma)w = 0 \}.$$

We identify the *behavior* $\mathscr{B}$ of Σ with the system Σ itself. No a priori separation of the variables into inputs and outputs is imposed.

B. Shortest lag representation

A kernel representation (3) is not unique. There exists a minimal one, called *shortest lag representation* [14, Sec. 7], in which the number of equations $p = \operatorname{row\,dim}(R)$, the maximum lag l, and the total lag $n = \sum_{i=1}^{p} l_i$, where l_i is the lag of the ith equation, are all minimal. In a shortest lag representation the number of equations p is equal to the number of outputs in an input/output representation and the total lag n is equal to the state dimension in a minimal state space representation. The maximal lag l is called the lag of the system Σ.

C. Model class $\mathscr{L}_{m,l}$

The number of inputs in an input/output or input/state/output representation of a linear system $\mathscr{B}$ is invariant [14, Sec. 4]. Denote by $\mathscr{L}_{m,l}$ the set of all LTI systems with at most m inputs and lag at most l. The natural numbers m and l specify the complexity of the model class $\mathscr{L}_{m,l}$. For $\mathscr{B} \in \mathscr{L}_{m,l}$ and T sufficiently large, the restriction $\mathscr{B}|_{[1,T]}$ of the behavior $\mathscr{B}$ to the interval $[1,T]$ has dimension $mT + n \leq mT + lp$.

The specification of the complexity by the lag of the system does not fix the order. For a system $\mathscr{B} \in \mathscr{L}_{m,l}$, the order n is in the range $(l-1)p < n \leq lp$.

III. Identification in the misfit setting

A. Problem formulation

Considering the model class $\mathscr{M} = \mathscr{L}_{m,l}$, the identification problem introduced in Section I-B is the following double minimization problem:

$$\min_{\mathscr{B} \in \mathscr{L}_{m,l}} \left(\min_{\hat{w}} \|w - \hat{w}\|_{\ell_2}^2 \quad \text{s.t.} \quad \hat{w} \in \mathscr{B} \right). \tag{4}$$

The given data is the time series w and the complexity specification (m,l); m the number of inputs and l the lag.

The inner minimization problem, *i.e.*, the misfit $M(w,\mathscr{B})$ computation, has the system theoretic meaning of finding the best approximation $\hat{w}$ of the given time series w, that is a trajectory of the (fixed from the outer minimization problem) LTI system $\mathscr{B}$. This is a *smoothing problem*. The outer minimization problem, however, is a difficult non-convex optimization problem that requires iterative methods. Some details on the numerical method used can be found in Appendix A, where a software package for solving the STLS problem (1) is described.

B. Solution by structured total least squares

Our goal is to express (4) as an STLS problem (1). Therefore, we need to ensure that the constraint $\mathscr{S}(\hat{w}) \left[\begin{smallmatrix} X \\ -I \end{smallmatrix} \right] = 0$ is equivalent to $\hat{w} \in \mathscr{B}$. As a byproduct of doing this, we relate the parameter X, in the STLS problem formulation, to the system $\mathscr{B}$.

Theorem 1: Assume that $\mathscr{B} \in \mathscr{L}_{m,l}$ is a system that admits a kernel representation $R(\sigma)w = 0$, $R(\xi) = \sum_{i=0}^{l} R_i \xi^i$ with $R_l =: \begin{bmatrix} Q_l & P_l \end{bmatrix}$, $P_l \in \mathbb{R}^{p \times p}$ of full row rank. Then the constraint $w \in \mathscr{B}|_{[1,T]}$ is equivalent to the constraint

$$\mathscr{H}_{l+1}^{\top}(w) \left[\begin{smallmatrix} X \\ -I \end{smallmatrix} \right] = 0,$$

where

$$X^{\top} = -P_l^{-1} \begin{bmatrix} R_0 & \cdots & R_{l-1} & Q_l \end{bmatrix}.$$

Proof: The proof is given in [1]. ∎

Theorem 1 states the desired equivalence of the identification problem and the STLS problem under the assumption that the optimal approximating system $\hat{\mathscr{B}}$ admits a kernel representation

$$\hat{\mathscr{B}} = \ker\Big(\sum_{i=0}^{l} \hat{R}_i \sigma^i \Big), \quad \hat{R}_l := \begin{bmatrix} \hat{Q}_l & \hat{P}_l \end{bmatrix}$$
$$\text{with } \hat{P}_l \in \mathbb{R}^{p \times p} \text{ full row rank.} \tag{5}$$

We conjecture that condition (5) holds true for almost all $w \in (\mathbb{R}^{\mathbb{w}})^T$. Define the subset of $(\mathbb{R}^{\mathbb{w}})^T$ consisting of all time series $w \in (\mathbb{R}^{\mathbb{w}})^T$ for which the identification problem is equivalent to the STLS problem, *i.e.*,

$$\Omega := \left\{ w \in (\mathbb{R}^{\mathbb{w}})^T \;\middle|\; \begin{array}{l} \text{problem (4) has a unique global} \\ \text{minimizer } \hat{\mathscr{B}} \text{ that satisfies (5)} \end{array} \right\}.$$

Conjecture 1: The set Ω is generic in $(\mathbb{R}^{\mathbb{w}})^T$, *i.e.*, it contains an open subset, whose complement has measure 0.

The existence and uniqueness part of the conjecture (see the definition of Ω) is motivated in [15, Sec. 5.1]. The motivation for (5) being generic is the following one. The highest possible order of a system in the model class $\mathscr{L}_{m,l}$ is pl. One can expect that generically in the data space $(\mathbb{R}^{\mathbb{w}})^T$, $n = pl$, where n is the order of $\hat{\mathscr{B}}$. By [1, Lemma 2], $n = pl$ implies that in a kernel representation $\hat{\mathscr{B}} = \ker\big(\sum_{i=0}^{l} \hat{R}_i \sigma^i\big)$, $\hat{R}_l$ is of full row rank. But generically in $\mathbb{R}^{p \times \mathbb{w}}$ the matrix $\hat{P}_l \in \mathbb{R}^{p \times p}$, defined by $\hat{R}_l =: \begin{bmatrix} \hat{Q}_l & \hat{P}_l \end{bmatrix}$, is of full row rank. Although the motivation for the conjecture is quite obvious, the proof seems to be rather involved.

C. Properties of the solution

The optimal value of the STLS problem is the minimal achievable misfit $M(w,\hat{\mathscr{B}})$ for a system in $\mathscr{L}_{m,l}$. With given $\hat{\mathscr{B}}$, the problem of finding $\hat{w}$ is a smoothing problem and can be solved independently of the identification problem.

Statistical properties of the identification problem (4), are studied in the literature. For the stochastic analysis, the errors-in-variables model is assumed and the basic results are consistency and asymptotic normality. Consistency in the SISO case is proven in [16]. Consistency in the MIMO case is proven in [15], in the framework of the global total least squares problem. Complete statistical theory with practical confidence bounds is presented in [17], in the setting of the Markov estimator for SISO semi-linear models. Consistency of the STLS estimator for the general structure specification described in Appendix A is proven in [18].

Deterministic properties of the optimal approximate trajectory $\hat{w}$ are also established. The following are properties of the smoothing problem: i) $\hat{w}$ is orthogonal to the correction $\Delta w := w - \hat{w}$ and ii) Δw is generated by an LTI system of order n with p inputs. Since the identification problem has as an inner minimization problem the smoothing problem, the same properties hold in particular for the optimal solution. These results are stated for the SISO case in [7] and then proven for the MIMO case in [10, Sec. VI].

IV. Extensions

A. Given input/output partitioning

A standard assumption in system identification is that an input/output partitioning of the variables is given. Let the first m variables of the given time series be inputs and the last p variables be outputs. Then with $\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] := w$ and

$$\begin{bmatrix} Q(\xi) & -P(\xi) \end{bmatrix} := R(\xi),$$

the kernel representation $R(\sigma)w = 0$ becomes a *left matrix fraction representation* $P(\sigma)y = Q(\sigma)u$. The transfer function of $\hat{\mathscr{B}}$ with the given input/output partitioning is $H(z) := P^{-1}(z)Q(z)$.

B. Exact variables

Another standard assumption is that the inputs are exact (in the errors-in-variables context noise free). Let $\hat{u}$ and $\hat{y}$ be the estimated input and output. The assumption that u is exact imposes the constraint $\hat{u} = u$.

More generally, if some variables of w are exact, then the corresponding elements in $\hat{w}$ are fixed. In the STLS problem formulation (1), the exact elements of w can be separated in a block of $\mathscr{S}(w)$ by permuting the columns of $\mathscr{H}_{l+1}^{\top}(w)$. The software package described in Appendix A allows specification of exact blocks in $\mathscr{S}(w)$ that are not modified in the solution $\mathscr{S}(\hat{w})$. After solving the modified problem, the solution $\hat{X}$ of the original problem, with exact variables, is obtained by applying the reverse permutation.

C. Latent inputs

The classical system identification framework [19] differs from the one in this paper in the choice of the optimization criterion and the model class. In [19], an unobserved input is assumed to act on the system that generates the observations and the optimization criterion is defined as the prediction error.

An unobserved input e, $\operatorname{row\,dim}\big(e(t)\big) = n_e$, called *latent input*, can be accommodated in the setting of Section III-A by augmenting the model class $\mathscr{M} = \mathscr{L}_{m,l}$ with n_e extra inputs and the cost function $\|w - \hat{w}\|_{\ell_2}^2$ with the term $\|e\|_{\ell_2}^2$. The resulting identification problem is

$$\min_{\mathscr{B} \in \mathscr{L}_{m+n_e,l}} \left(\min_{\hat{e},\hat{w}} \underbrace{\|w - \hat{w}\|_{\ell_2}^2}_{\text{misfit}} + \underbrace{\|\hat{e}\|_{\ell_2}^2}_{\text{latency}} \text{ s.t. } \begin{bmatrix} \hat{e} \\ \hat{w} \end{bmatrix} \in \mathscr{B} \right). \quad (6)$$

It unifies the misfit and latency description of the uncertainty and is put forward by Lemmerling and De Moor [8]. In [8], it is claimed that the pure latency identification problem

$$\min_{\mathscr{B} \in \mathscr{L}_{m+n_e,l}} \left(\min_{\hat{e}} \|\hat{e}\|_{\ell_2}^2 \text{ s.t. } \begin{bmatrix} \hat{e} \\ w \end{bmatrix} \in \mathscr{B} \right) \quad (7)$$

is equivalent to the prediction error approach.

The misfit–latency identification problem (6) can easily be reformulated as an equivalent pure misfit identification problem (4). Let $w_{\text{aug}} := \left[\begin{smallmatrix} e \\ w \end{smallmatrix}\right]$, where $e := 0$ is an n_e dimensional zero time series. Then the misfit minimization problem for the time series w_{aug} and the model class $\mathscr{L}_{m+n_e,l}$ is equivalent to (6).

The latent input amounts to increasing the complexity of the model class. Thus a better fit is achieved with a less powerful model. In our view, a latent input should be used in the identification problem only when there is a *prior knowledge* that the data w is generated by a system with unobserved input, *i.e.*, when e exists physically.

This viewpoint is in contrast to the classical by now system identification philosophy that considers the unobserved input as a tool to account for the cumulative effect of various sources—model class inadequacy, unobserved inputs, measurement errors, *etc.*

D. Multiple time series

In certain cases, *e.g.*, the noisy realization problem, not one but several observed time series $w^{(k)}$, $k = 1,\ldots,K$ are given. Assume that all time series are of the same length and define w to be the matrix valued time series $w = \begin{bmatrix} w^{(1)} & \cdots & w^{(K)} \end{bmatrix}$, so that $w(t) \in \mathbb{R}^{\mathtt{w} \times K}$. The only modification needed for this case is to consider block-Hankel matrix $\mathscr{H}_{l+1}(w)$ with size of the repeated blocks $K \times \mathtt{w}$ instead of $1 \times \mathtt{w}$, as for the case of a single observed time series. The software package described in Appendix A can treat such problems.

V. Performance on Data Sets from DAISY

Currently the data base for system identification DAISY [9] contains 28 real-life and simulated data sets, which are used for verification and comparison of identification algorithms. In this section, we apply the described identification method, implemented by the software tool of [20], on data sets from DAISY.

The first part of Table I gives information for the data sets (number of data points T, number of inputs m, the number of outputs p) and shows the selected lag l for the identified model. Since all data sets are with given input/output partitioning, the only user defined parameter selecting the complexity of the model class $\mathscr{L}_{m,l}$ is the lag l.

The estimates obtained by the following methods are compared:

- `subid`, a MATLAB implementation of the robust combined subspace algorithm of [21, Fig. 4.8];
- `detss`, a MATLAB implementation of the deterministic balanced subspace algorithm of [22];
- `pem`, the prediction error method of the Identification Toolbox of MATLAB;
- `stls`, the proposed method based on STLS.

Note that $l+1$ is the user supplied parameter i in the combined subspace algorithm `subid`. The order specified for the methods `subid`, `detss`, and `pem` is pl (the maximum possible in the model class $\mathscr{L}_{m,l}$).

The comparison is in terms of the relative percentage misfit

$$M_{\text{rel}}(w,\hat{\mathscr{B}}) := 100\, M(w,\hat{\mathscr{B}})/\|w(t)\|_{\ell_2}.$$

M_{rel} is computed by solving the smoothing problem $M(w,\hat{\mathscr{B}})$ for the estimated models $\hat{\mathscr{B}}$. For `detss` and `pem`, $\hat{\mathscr{B}}$ is the deterministic part of the identified stochastic system.

The second part of Table I shows the relative misfits M_{rel} and execution time for the compared methods. The STLS solver is initialized with the approximation of the non-iterative method `subid` or `detss` that achieves smaller misfit on the particular data set. The time needed for the computation of the initial approximation is *not* added in the timing of `stls`. The prediction error method is called with the data w and the order $n = pl$ specification only, so that the appropriate model structure and computational method are selected automatically by the function.

Since M_{rel} is up to a scaling factor equal to the cost function of `stls`, it is not surprising that the proposed method outperforms with respect to this criterion the alternative methods. The purpose of doing the comparison is to verify that the numerical tool needed for the solution of the optimization problem (1) is robust and efficient.

Indeed, identification problems with a few thousands data points can be solved with the STLS software package. Such problems are infeasible for direct application of optimization methods without exploiting the special structure. Also the computation time of `stls` is similar to that of `pem`, which is also an optimization based method. On all examples, initialization of the STLS solver with the estimate obtained by a subspace identification method, leads to improved solution, in terms of the misfit criterion. Thus for the expense of some extra computation time, the subspace approximation is improved by `stls`.

In a stochastic setting, under suitable conditions, the STLS estimate is consistent and asymptotically normal. Therefore, an error estimate in the form of confidence bound is available for the parameters of the identified system [17, Ch. 17.4.7]. As a by product of the optimization algorithm, the STLS solution package returns the asymptotic covariance matrix of the error of estimation. In the SISO case, this information gives confidence bounds for the coefficients of the obtained transfer function.

VI. Conclusions

We generalized previous results on the application of STLS for system identification to multivariable systems. The STLS method allows to treat identification problems, without input/output partitioning of the variables and errors-in-variables identification problems. Multiple time series, latent variables, and prior knowledge about noise free variables can be taken into account.

The classical identification problem, where the uncertainty is attributed solely to unobserved inputs and the observed inputs are assumed exact is a special case of the proposed method. The relation and comparison with classical identification methods, however, is not yet investigated.

A robust and efficient software tool for solving STLS problems is presented. It makes the proposed identification method practically applicable. The performance of the software package was tested on data sets from DAISY. The results show that examples with a few thousands data points can be solved routinely and the optimization method is robust with respect to initial approximation obtained from a non-optimization based method.

Appendix

A. Software package for solving STLS problems [20]

Efficient numerical methods for STLS problems were developed in the past, see [23]. Applied to the system identification problem, however, they cover only SISO problems and the input/output identification problem can not be treated. On the other hand, the more general STLS methods of [2], [3] have cubic computational complexity and are restricted to rank reduction with one. Thus they are applicable only for small size SISO problems.

The described package solves the STLS problem (1) with $\mathscr{S}(\cdot) = \begin{bmatrix} C^{(1)} & \cdots & C^{(q)} \end{bmatrix}$, where $C^{(i)}$ is block-Toeplitz (T), block-Hankel (H), unstructured (U), or exact (E). All block-Toeplitz/Hankel structured blocks $C^{(i)}$ have blocks of the same row size K and possibly different column dimensions t_i. Such a structure specification is more general than necessary for the considered basic identification problem. For the extensions of Section IV, however, as well as for other approximation problems, the additional flexibility in

TABLE I

RELATIVE MISFITS M_{REL} AND EXECUTION TIMES t IN SECONDS FOR THE EXAMPLES AND THE METHODS.

#	Data set name	parameters				subid		detss		pem		stls	
		T	m	p	l	t	M_{rel}	t	M_{rel}	t	M_{rel}	t	M_{rel}
1	Destillation column	90	5	3	1	0.11	0.0089	0.57	0.0306	1.66	0.0505	0.45	0.0029
2	Destillation column n10	90	5	3	1	0.10	0.0089	0.57	0.0306	1.58	0.0505	0.45	0.0029
3	Destillation column n20	90	5	3	1	0.10	0.4309	0.57	0.1187	0.54	1.8574	1.17	0.0448
4	Destillation column n30	90	5	3	1	0.10	0.4357	0.61	0.1848	0.57	7.3600	1.18	0.0522
5	Glass furnace (Philips)	1247	3	6	1	0.15	33.6782	1.72	29.3373	43	31.5416	84	11.4120
6	120 MW power plant	200	5	3	2	0.14	8.9628	0.63	4.2906	2.68	35.4524	0.77	1.2427
7	pH process	2001	2	1	6	0.23	4.2564	0.64	4.4113	5.66	9.8727	1.93	3.2203
8	Hair dryer	1000	1	1	5	0.12	1.0437	0.18	1.0359	3.33	0.8311	1.19	0.8208
9	Winding process	2500	5	2	2	0.29	11.4838	1.15	10.1473	17	20.2908	62	7.1731
10	Ball-and-beam setup	1000	1	1	2	0.10	2.8962	0.17	28.3637	0.55	2.7708	0.11	2.6718
11	Industrial dryer	867	3	3	1	0.11	0.5586	0.59	0.5519	2.35	0.5553	5.11	0.4447
12	CD-player arm	2048	2	2	1	0.13	9.4629	0.59	8.7653	4.79	11.3623	9.37	7.7980
13	Wing flutter	1024	1	1	5	0.12	20.2766	0.19	21.0214	3.14	35.2727	0.91	11.6501
14	Robot arm	1024	1	1	4	0.12	3.8855	0.21	26.0082	2.66	36.1531	0.08	1.3905
15	Lake Erie	57	5	2	1	0.11	0.1423	0.14	0.2205	0.52	2.1548	0.43	0.0908
16	Lake Erie n10	57	5	2	1	0.10	0.0505	0.14	0.0538	0.85	0.1992	0.41	0.0221
17	Lake Erie n20	57	5	2	1	0.10	0.0607	0.16	0.0671	0.67	0.2677	0.45	0.0268
18	Lake Erie n30	57	5	2	1	0.10	0.0798	0.17	0.0564	0.52	0.1862	0.41	0.0329
19	Heat flow density	1680	2	1	2	0.13	0.7779	0.30	0.5651	3.73	4.1805	0.49	0.4219
20	Heating system	801	1	1	2	0.10	0.4913	0.17	0.4441	0.94	0.4973	0.09	0.3658
21	Steam heat exchanger	4000	1	1	2	0.14	0.1521	0.54	0.1499	3.37	0.6723	0.40	0.0822
22	Industrial evaporator	6305	3	3	1	0.32	37.7809	3.27	27.6341	40	40.6798	15	24.0065
23	Tank reactor	7500	1	2	1	0.19	0.1768	1.89	0.1621	24	3.9620	2.18	0.0749
24	Steam generator	9600	4	4	1	0.66	0.4175	8.45	0.5341	132	0.5751	118	0.1704

the problem formulation is useful. Also, as shown below, the extra flexibility of the problem formulation does not lead to significant complication in the solution method.

The structure of $\mathscr{S}(\cdot)$ is specified by K and the array $\mathscr{D} \in (\{\mathtt{T},\mathtt{H},\mathtt{U},\mathtt{E}\} \times \mathbb{N} \times \mathbb{N})^q$ that describes the blocks $\{C^{(i)}\}_{i=1}^q$; $\mathscr{D}_i$ specifies the block $C^{(i)}$ by giving its type $\mathscr{D}_i(1)$, the number of columns $n_i = \mathscr{D}_i(2)$, and (if $C^{(i)}$ is block-Hankel or block-Toeplitz) the column dimension $\mathscr{D}_i(3)$ of the repeated block. The input data for the problem is $\mathscr{S}(w)$ and the structure specification K and $\mathscr{D}$.

We solve the optimization problem (1) in its equivalent formulation, see [24],

$$\min_X \; r^\top(X)\Gamma^{-1}(X)r(X), \tag{8}$$

where $r(X) := \operatorname{vec}\left(\left(\mathscr{S}(w)\begin{bmatrix} X \\ -I \end{bmatrix}\right)^\top\right)$. The weight matrix Γ is block-Toeplitz and block-banded structured, see [25],

$$\Gamma(X) = \begin{bmatrix} \Gamma_0 & \Gamma_{-1} & \cdots & \Gamma_{-s} & & 0 \\ \Gamma_1 & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \Gamma_{-s} \\ \Gamma_s & \ddots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \Gamma_{-1} \\ 0 & & \Gamma_s & \cdots & \Gamma_1 & \Gamma_0 \end{bmatrix}.$$

The block size is dK and the upper/lower block bandwidth s equals the maximum number of block columns in a block-Hankel/Toeplitz structured block $C^{(i)}$.

The structure in the weight matrix $\Gamma(X)$ is exploited for efficient cost function and first derivative evaluation. The cost function evaluation requires solving the system of equations $\Gamma(X)y(X) = r(X)$ and computing the inner product $r^\top(X)y(X)$. Straightforward implementation of these steps results in $O(T^3)$ floating point operations. By solving the system of equations, taking into account the structure of $\Gamma(X)$, we reduce the computational complexity to $O(T)$. In addition, the first derivative of the cost function can be computed with the same computational complexity. There is no closed form expression for the Jacobian matrix $J = \left[\partial r_i/\partial x_j\right]$, where $x = \operatorname{vec}(X)$, so that the pseudo-Jacobian J_+ proposed in [26] is used instead of J. The cost function and pseudo-Jacobian evaluation is performed with computational complexity $O(T)$.

The input data for the STLS algorithm is the structure specification $K, \mathscr{D}$ and the data matrix C. The flexible structure specification $\mathscr{D}$ is utilized in the computations for the construction of the weight matrix $\Gamma(X)$. Because of its structure, $\Gamma(X)$ is specified by the nonzero part of its first block row, *i.e.*, by the $s+1$ matrices $\{\Gamma_k(X)\}_{k=0}^s$. In [25, Sec. 4], we prove that the Γ_k's are quadratic in X. Define the shift matrix

$$J_{n_i} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n_i \times n_i}$$

and let δ be the Kronecker delta function: $\delta(0) = 1$ and $\delta(k) = 0$ for $k \neq 0$. It can be shown that

$$\Gamma_k(X) = \mathbf{X}_{\text{ext}} W_{\mathtt{t},k} \mathbf{X}_{\text{ext}}^\top, \quad \text{for } k = 0,\ldots,s, \tag{9}$$

where

$$\mathbf{X}_{\text{ext}} := \left(I_K \otimes \begin{bmatrix} X^\top & -I \end{bmatrix}\right),$$

$$W_{\bar{\epsilon},k} := \text{blk diag}(W_k^{(1)}, \ldots, W_k^{(q)}), \tag{10}$$

and

$$W_k^{(i)} = \begin{cases} (J_{n_i}^\top)^{t_i k} & \text{,if } C^{(i)} \text{ is Toeplitz} \\ (J_{n_i})^{t_i k} & \text{,if } C^{(i)} \text{ is Hankel} \\ \delta(k) I_{n_i} & \text{,if } C^{(i)} \text{ is unstructured} \\ 0_{n_i} & \text{,if } C^{(i)} \text{ is exact.} \end{cases}$$

By computing the $W_{\bar{\epsilon},k}$'s defined in (10), $\Gamma(X)$ can be constructed for any X via (9). Expressions (9) and (10) completely "decode" the structure specification $\mathscr{D}$ and utilize it in the subsequent computations of the cost function and the pseudo Jacobian.

The software is written in ANSI C language. For the vector–matrix manipulations and for a C version of MINPACK's Levenberg–Marquardt algorithm, the *GNU Scientific library (GSL)* is used. The computationally most intensive step of the algorithm—the Cholesky decomposition of the block-Toeplitz, block-banded weight matrix $\Gamma(X)$—is performed via the subroutine `MB02GD` from the SLICOT library [27]. MATLAB interface via C-mex file is available.

ACKNOWLEDGMENTS

We would like to thank Diana Sima for careful reading of the drafts and valuable suggestions.

I. Markovsky is supported by a K.U. Leuven doctoral scholarship. Research supported by: **Research Council KUL:** GOA-Mefisto 666, IDO/99/003 and IDO/02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants, GOA AMBioRICS; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), G.0452.04 (new quantum algorithms), G.0499.04 (Robust SVM), G.0499.04 (Statistics) research communities (ICCoS, ANMMM, MLDM); **IWT:** PhD Grants, GBOU (McKnow); **Belgian Federal Government:** DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling); PODO-II (CP/40: TMS and Sustainability); **EU:** PDT-COIL, BIOPATTERN, eTUMOUR, FP5-Quprodis; ERNSI; Eureka 2063-IMPACT; Eureka 2419-FliTE; **AWI:** Bil. Int. Collaboration Hungary/Poland; Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, Mastercard.

REFERENCES

[1] I. Markovsky, J. C. Willems, S. Van Huffel, B. D. Moor, and R. Pintelon, "Application of structured total least squares for system identification and model reduction," Dept. EE, K.U. Leuven, Tech. Rep. 04–51, 2004.

[2] T. Abatzoglou, J. Mendel, and G. Harada, "The constrained total least squares technique and its application to harmonic superresolution," *IEEE Trans. on Signal Proc.*, vol. 39, pp. 1070–1087, 1991.

[3] B. De Moor, "Structured total least squares and L_2 approximation problems," *Lin. Alg. and Its Appl.*, vol. 188–189, pp. 163–207, 1993.

[4] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, pp. 883–893, 1980.

[5] S. Van Huffel and J. Vandewalle, *The total least squares problem: Computational aspects and analysis.* SIAM, Philadelphia, 1991.

[6] M. J. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. on Aut. Control*, vol. 9, pp. 229–235, 1964.

[7] B. De Moor and B. Roorda, "L_2-optimal linear system identification structured total least squares for SISO systems," in *In the proceedings of the CDC*, 1994, pp. 2874–2879.

[8] P. Lemmerling and B. De Moor, "Misfit versus latency," *Automatica*, vol. 37, pp. 2057–2067, 2001.

[9] B. De Moor, "Daisy: Database for the identification of systems," Department of Electrical Engineering, ESAT/SISTA, K.U.Leuven, Belgium, `http://www.esat.kuleuven.ac.be/sista/daisy/`

[10] B. Roorda and C. Heij, "Global total least squares modeling of multivariate time series," *IEEE Trans. on Aut. Control*, vol. 40, no. 1, pp. 50–63, 1995.

[11] J. C. Willems, "From time series to linear system—Part III. Approximate modelling," *Automatica*, vol. 23, no. 1, pp. 87–115, 1987.

[12] B. Roorda, "Algorithms for global total least squares modelling of finite multivariable time series," *Automatica*, vol. 31, no. 3, pp. 391–404, 1995.

[13] J. W. Polderman and J. C. Willems, *Introduction to mathematical systems theory.* Springer–Verlag, New York, 1998.

[14] J. C. Willems, "From time series to linear system—Part I. Finite dimensional linear time invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

[15] C. Heij and W. Scherrer, "Consistency of system identification by global total least squares," *Automatica*, vol. 35, pp. 993–1008, 1999.

[16] M. Aoki and P. C. Yue, "On certain convergence questions in system identification," *SIAM J. Control*, vol. 8, no. 2, pp. 239–256, 1970.

[17] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach.* IEEE Press, Piscataway, 2001.

[18] A. Kukush, I. Markovsky, and S. Van Huffel, "Consistency of the structured total least squares estimator in a multivariate errors-in-variables model," Dept. EE, K.U. Leuven, Tech. Rep. 02–192, 2002, accepted for publication in *J. of Stat. Planning and Inference.*

[19] L. Ljung, *System Identification: Theory for the user.* Prentice Hall, Upper Saddle River, 1999.

[20] I. Markovsky and S. Van Huffel, "Software for structured total least squares estimation: User's guide," Dept. EE, K.U. Leuven, Tech. Rep. 03–136, 2003.

[21] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications.* Kluwer Academic Publishers, 1996.

[22] I. Markovsky, J. C. Willems, P. Rapisarda, and B. D. Moor, "Algorithms for deterministic balanced subspace identification," Dept. EE, K.U. Leuven, Tech. Rep. 04–13, 2004.

[23] N. Mastronardi, "Fast and reliable algorithms for structured total least squares and related matrix problems," Ph.D. dissertation, ESAT/SISTA, K.U. Leuven, 2001.

[24] I. Markovsky, S. Van Huffel, and A. Kukush, "On the computation of the structured total least squares estimator," *Numerical Linear Algebra with Applications*, vol. 11, pp. 591–608, 2004.

[25] I. Markovsky, S. Van Huffel, and R. Pintelon, "Block-Toeplitz/Hankel structured total least squares," Dept. EE, K.U. Leuven, Tech. Rep. 03–135, 2003, accepted for publication in *SIAM J. On Matrix Analysis and Appl.*

[26] P. Guillaume and R. Pintelon, "A Gauss–Newton-like optimization algorithm for "weighted" nonlinear least-squares problems," *IEEE Trans. on Signal Proc.*, vol. 44, no. 9, pp. 2222–2228, 1996.

[27] S. Van Huffel, V. Sima, A. Varga, S. Hammarling, and F. Delebecque, "High-performance numerical software for control," *IEEE Control Systems Magazine*, vol. 24, pp. 60–76, 2004.