Application of Structured Total Least Squares for System Identification and Model Reduction

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

***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 question leads to a problem that is known as the global total least squares problem 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. Special cases of the problem are autonomous system identification, approximate realization, and finite time optimal ℓ_2 model reduction. The identification problem is related to the structured total least squares problem. This paper presents an efficient software package that implements the theory. The proposed method and software are tested on data sets from the database for the identification of systems DAISY.**

***Index Terms*—DAISY, errors-in-variables, model reduction, MPUM, numerical software, system identification, structured total least squares.**

I. Introduction

A. The Structured Total Least Squares Problem

THE structured total least squares (STLS) problem originates [1], [2] 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 [3], [4] known in the early system identification literature as the Koopmans–Levin's method [5]. In this paper, we show the applicability of the STLS method for system identification. We extend previous results [6], [7] of the application of STLS for single-input–single-output (SISO) system identification to the multiple-input–multiple-output (MIMO) case and present numerical results on data sets from DAISY [8].

Manuscript received May 26, 2004; revised January 27, 2005 and May 31, 2005. Recommended by Guest Editor L. Ljung. The work of I. Markovsky was supported by a K.U. Leuven doctoral scholarship. Research was supported by: **Research Council KUL:** GOA-Mefisto 666, GOA-Ambiorics, IDO/99/003 and IDO/02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several Ph.D./postdoctoral and fellow grants; **Flemish Government: FWO:** Ph.D./postdoctoral grants, projects, G.0078.01 (structured matrices), 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 (QC), G.0499.04 (robust SVM), research communities (ICCoS, ANMMM, MLDM); AWI: Bil. Int. Collaboration Hungary/Poland; **IWT:** Ph.D. Grants; GBOU (McKnow) **Belgian Federal Government:** DWTC (IUAP IV-02 (1996–2001) and Belgian Federal Science Policy Office IUAP V-22 (2002–2006) (Dynamical Systems and Control: Computation, Identification and Modeling)); PODO-II (CP/01/40: TMS and Sustainibility); **EU:** PDT-COIL, BIOPATTERN, eTUMOUR, FP5-Quprodis; ERNSI; Eureka 2063-IMPACT; Eureka 2419-FliTE; Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, IPCOS, Mastercard.

I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor are with the Electrical Engineering Department, K. U. Leuven, B-3001 Leuven, Belgium (e-mail: Ivan.Markovsky@esat.kuleuven.be).

R. Pintelon is with Department ELEC, Vrije Universiteit Brussel, B-1050 Brussels, Belgium.

Digital Object Identifier 10.1109/TAC.2005.856643

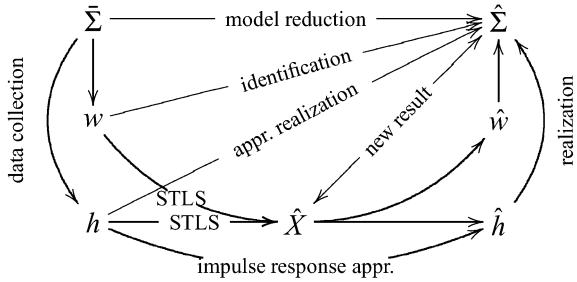

Fig. 1. Different problems aiming at a (low complexity) model $\hat{\Sigma}$ that approximates a given (high complexity) model $\bar{\Sigma}$. The time series w is an observed response and h is an observed impulse response.

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

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

The constraint of (1) enforces the structured matrix $\mathcal{S}(\hat{w})$ to be rank deficient, with rank at most $\operatorname{row\,dim}(X)$. 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 $\mathcal{S}(w)$ by $\mathcal{S}(\hat{w})$.

B. Approximate Modeling Problems

Fig. 1, shows several approximate modeling problems. On top is the model reduction problem: given a *linear time-invariant* (LTI) system $\bar{\Sigma}$, find an LTI approximation $\hat{\Sigma}$ of a desired lower complexity. A tractable solution that gives very good results in practice is balanced truncation [9]. We consider finite time ℓ_2 optimal model reduction: The sequence of the first T Markov parameters of $\bar{\Sigma}$ is approximated by the sequence of the corresponding Markov parameters of $\hat{\Sigma}$ in the ℓ_2 norm sense.

The identification problem is similar to the model reduction one, but starts instead from an observed response w. Various data collection models (the down arrows from $\bar{\Sigma}$ to w and h in Fig. 1) are possible. For example, the *errors-in-variables* model: $w = \bar{w} + \tilde{w}$, where $\bar{w}$ is a trajectory generated by $\bar{\Sigma}$ and $\tilde{w}$ is measurement noise.

Of independent interest are the identification problems from an observed impulse response h, called *approximate realization*

0018-9286/$20.00 © 2005 IEEE

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

problem, and the autonomous system identification problem, where $\bar{w}$ and $\hat{w}$ are free responses. A classical solution to these problems is Kung's algorithm [10].

The key observation that motivates the application of STLS for system identification and model reduction is that their kernel subproblem is to find a block-Hankel rank deficient matrix $\mathcal{S}(\hat{w})$ approximating a given full-rank matrix $\mathcal{S}(w)$ with the same structure. Methods like balanced model reduction, subspace identification, and Kung's algorithm solve the kernel problem via the *singular value decomposition* (SVD). For finite matrices, however, the SVD approximation of $\mathcal{S}(w)$ is unstructured. For this reason the algorithms based on the SVD are suboptimal with respect to an induced norm of the residual $\Delta w := w - \hat{w}$. The STLS method, on the other hand, preserves the structure and is optimal according to this criterion. Our purpose is to show how system theoretic problems with misfit optimality criterion are solved as equivalent STLS problems and subsequently make use of efficient numerical methods developed for the STLS problem [11]–[13].

C. The Global Total Least Squares Problem

Let $\mathcal{M}$ be a user specified model class and let w be an observed time series of length $T \in \mathbb{N}$. We view a model $\mathcal{B} \in \mathcal{M}$ 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 $\mathcal{M}$, we aim to find the model $\hat{\mathcal{B}}$ that best fits the data according to the misfit criterion

$$\hat{\mathcal{B}} := \arg\min_{\mathcal{B}\in\mathcal{M}} M(w,\mathcal{B}) \quad \text{with } M(w,\mathcal{B}) := \min_{\hat{w}\in\mathcal{B}} \|w-\hat{w}\|_{\ell_2}^2.$$

The resulting optimization problem is known as the *global total least squares* (GTLS) problem [14]. The described system identification framework is put forward in [15].

The approach of Roorda and Heij [14], [16] is based on solving the inner minimization problem, the misfit computation, by using isometric state representation and subsequently used alternating least squares or Gauss–Newton type algorithm for the outer minimization problem. They use a state space representation with driving input. Our approach of solving the GTLS problem, is different. We relate the identification problem to the STLS problem (1) and subsequently use solution methods developed for the STLS problem. Also we use a kernel representation of the system.

D. Link With the Most Powerful Unfalsified Model

In [17], the concept of the *most powerful unfalsified model* (MPUM) is introduced. A model $\mathcal{B}$ is unfalsified by the observation w if $w \in \mathcal{B}$. A model $\mathcal{B}_1$ is more powerful than $\mathcal{B}_2$ if $\mathcal{B}_1 \subset \mathcal{B}_2$. Thus the concept of the MPUM is to find the most powerful model consistent with the observations—a most reasonable and intuitive identification principle.

In practice, however, the MPUM could be unacceptably complex. For example, in the errors-in-variables setup the observation $w := (w(1),\ldots,w(T)), w(t) \in \mathbb{R}^{\mathrm{w}}$, is perturbed by noise, so that with probability one the MPUM is all of $(\mathbb{R}^{\mathrm{w}})^T$. Such a model is useless because it imposes no laws.

The GTLS problem addresses this issue by restricting the model complexity by the constraint $\hat{\mathcal{B}} \in \mathcal{M}$, where $\mathcal{M}$ is an *a priori* specified model class. Whenever the MPUM does not belong to $\mathcal{M}$, an approximation is needed. The idea is to modify as little as possible the given time series, so that the MPUM of the modified time series belongs to $\mathcal{M}$—a most reasonable generalization of the MPUM to the bounded complexity case. The measure of closeness is chosen as the ℓ_2 norm, which weights equally all variables over all time instants. Weighted norms can be used in order to take into account prior knowledge about nonuniform variance among the variables and/or in time.

E. Outline of this 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 describes the special identification problems from impulse and free response. Section VI shows results of the proposed method on data sets from DAISY, and Section VII gives conclusions.

II. Preliminaries

A. Kernel Representation

Consider a time series $w := (w(1),\ldots,w(T))$. The block-Hankel matrix with l block rows, constructed from the time series w, is denoted by

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

The time series w satisfies the set of difference equations

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

with maximum l *lags* (i.e., unit delays in time), if and only if

$$R\mathcal{H}_{l+1}(w) = 0 \quad \text{where } R := [R_0 \ R_1 \ \ldots \ R_l].$$

The rank deficiency of $\mathcal{H}_{l+1}(w)$ 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 [18], (2) is called a *kernel representation* of the system Σ. A more compact notation is

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

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

$$\mathcal{B} := \{w : \mathbb{N} \to \mathbb{R}^{\mathrm{w}} \mid R(\sigma)w = 0\}.$$

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

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

B. Shortest Lag Representation

A kernel representation (3) for a given $\mathcal{B}$ is not unique. If the polynomial matrix $R(z)$ defines a kernel representation of $\mathcal{B}$, then for any unimodular matrix $U \in \mathbb{R}^{p\times p}[z], U(z)R(z)$ also defines a kernel representation of $\mathcal{B}$. The kernel representation is called minimal if R is full row rank. There exists a minimal one, called *shortest lag representation* [19, 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. A kernel representation $R(\sigma)w = 0$ is a shortest lag representation if and only if $R(z)$ is row proper. Let l_i be the degree of the ith row of R. The polynomial matrix R is called *row proper* if the leading row coefficient matrix, i.e., the matrix of which the (i, j)th entry is the coefficient of the term with power l_i of $R_{ij}(z)$, is full row rank.

In a shortest lag representation the number of equations is equal to the number of outputs in an input/output representation and the total lag is equal to the state dimension in a minimal state–space representation. These numbers are *invariants* of the system; see [19, Sec. 4]. The maximal lag of $\mathcal{B}$, denoted by $\mathbf{l}(\mathcal{B})$, is called the lag of the system, and its total lag, denoted by $\mathbf{n}(\mathcal{B})$, is called the order of the system.

C. Model Class $\mathcal{L}_{m,l}$

The number of inputs $\mathbf{m}(\mathcal{B})$ and the number of outputs $\mathbf{p}(\mathcal{B})$ in an input–output or input/state/output representation are also invariants of the system $\mathcal{B}$. We denote by $\mathcal{L}_{m,l}$ the set of LTI systems with m inputs and lag at most l. (Occasionally, we display the number of variables $\mathtt{w}$ as follows $\mathcal{L}^{\mathtt{w}}_{m,l}$.) The natural numbers m and l specify the maximum complexity of a model in the model class $\mathcal{L}_{m,l}$. For $\mathcal{B} \in \mathcal{L}_{m,l}$ and T sufficiently large, the restriction $\mathcal{B}|_{[1,T]}$ of the behavior $\mathcal{B}$ to the interval $[1, T]$ has dimension $mT + \mathbf{n}(\mathcal{B}) \leq mT + l\mathbf{p}(\mathcal{B})$.

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

In Section V, we use the notation $\mathcal{B}(A, B, C, D)$ for the input/state/output representation

$$\sigma x = Ax + Bu \quad y = Cx + Du \tag{4}$$

of the LTI system $\mathcal{B} := \{\operatorname{col}(u, y) \mid \exists x, \text{ such that (4) holds}\}$. The column vector with (block) entries $a_1, \ldots, a_n$ is denoted by $\operatorname{col}(a_1, \ldots, a_n)$.

III. Identification in the Misfit Setting by STLS

The considered identification problem is defined as follows.

Problem 1(GTLS): For given time series w and a complexity specification (m, l), where m is the number of inputs and l is the lag of the identified system, solve the optimization problem

$$\hat{\mathcal{B}} := \arg\min_{\mathcal{B}\in\mathcal{L}_{m,l}} \left(\min_{\hat{w}\in\mathcal{B}} \|w - \hat{w}\|^2_{\ell_2} \right). \tag{5}$$

The optimal approximating time series is $\hat{w}$, corresponding to a global minimum point of (5), and the optimal approximating system is $\hat{\mathcal{B}}$.

Problem 1 is the GTLS problem for the model class $\mathcal{M} = \mathcal{L}_{m,l}$. The inner minimization problem, i.e., the misfit $M(w, \mathcal{B})$ 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) system $\mathcal{B}$. This is a *smoothing problem*.

Our goal is to express (5) as an STLS problem (1). Therefore, we need to ensure that the constraint $\mathcal{S}(\hat{w}) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$ is equivalent to $\hat{w} \in \mathcal{B} \in \mathcal{L}_{m,l}$. As a byproduct of doing this, we relate the parameter X, in the STLS problem formulation, to the system $\mathcal{B}$. The equivalence is proven under an assumption that is conjectured to hold generically in the data space $(\mathbb{R}^{\mathtt{w}})^T$.

Lemma 1: Consider a time series $w := (w(1), \ldots, w(T))$, $w(t) \in \mathbb{R}^{\mathtt{w}}$, and natural numbers $m \leq \mathtt{w}$ and $l \leq T-1$. Assume that $R\mathcal{H}_{l+1}(w) = 0$ for certain matrix $R = [R_0 \ R_1 \ \cdots \ R_l]$, $R_i \in \mathbb{R}^{p\times \mathtt{w}}$, where $p := \mathtt{w} - m$, with R_l being full row rank. Then the system $\mathcal{B}$, defined by the kernel representation $R(\sigma)w = 0$ with $R(z) = \sum_{i=0}^{l} R_i z^i$, is such that $w \in \mathcal{B}|_{[1,T]}, \mathcal{B} \in \mathcal{L}_{m,l}$, and the order of $\mathcal{B}$ is $\mathbf{n}(\mathcal{B}) = pl$.

Proof: By definition $\mathcal{B}$ is a linear system with lag $\mathbf{l}(\mathcal{B}) \leq l$. The assumption that R_l is full-row rank implies that $R(z)$ is row proper. Then the number of outputs of $\mathcal{B}$ is $\mathbf{p}(\mathcal{B}) = p$ and, therefore, $\mathbf{m}(\mathcal{B}) = \mathtt{w} - \mathbf{p}(\mathcal{B}) = m$. This proves that $\mathcal{B} \in \mathcal{L}_{m,l}$.

Let l_i be the degree of the ith equation in $R(\sigma)w = 0$. The assumption that R_l is full row rank implies that $l_i = l$ for all i. Therefore, $\mathbf{n}(\mathcal{B}) = \sum_{i=1}^{p} l_i = pl$.

Finally, $R\mathcal{H}_{l+1}(w) = 0$ implies $\sum_{\tau=0}^{l} R_\tau w(t+\tau) = 0$, for $t = 1, \ldots, T-l$, so that $w \in \mathcal{B}|_{[1,T]}$. ∎

The next lemma states the reverse implication.

Lemma 2: Consider a time series

$$w := (w(1), \ldots, w(T)), w(t) \in \mathbb{R}^{\mathtt{w}}$$

and natural numbers $m \leq \mathtt{w}$ and $l \leq T-1$. Assume that there is a system $\mathcal{B} \in \mathcal{L}_{m,l}$ with order $\mathbf{n}(\mathcal{B}) = pl$, such that $w \in \mathcal{B}|_{[1,T]}$. Let $R(\sigma)w = 0$, where $R(z) = \sum_{i=0}^{l} R_i z^i$, be a shortest lag kernel representation of $\mathcal{B}$. Then, R_l is full-row rank and the matrix $R := [R_0 \ R_1 \ \cdots \ R_l]$ annihilates the Hankel matrix $\mathcal{H}_{l+1}(w)$, i.e., $R\mathcal{H}_{l+1}(w) = 0$.

Proof: Let l_i be the degree of the ith equation in $R(\sigma)w = 0$. We have $l_i \leq l$ and $\mathbf{n}(\mathcal{B}) = \sum_{i=1}^{p} l_i$. The assumption $\mathbf{n}(\mathcal{B}) = pl$ is possible only if $l_i = l$, for all i. Because $R(z)$ is row proper (by the shortest lag assumption of the kernel representation), the leading row coefficient matrix L has full-row rank. But since $l_i = l$, for all i, $L = R_l$.

The fact that $R\mathcal{H}_{l+1}(w) = 0$ follows from $w \in \mathcal{B}|_{[1,T]}$. ∎

We have the following main result.

Theorem 1: Assume that $\mathcal{B} \in \mathcal{L}_{m,l}$ is a system that admits a kernel representation $R(\sigma)w = 0, R(z) = \sum_{i=0}^{l} R_i z^i$ with $R_l =: [Q_l \ -P_l], P_l \in \mathbb{R}^{p\times p}$ of full-row rank. Then, the constraint $w \in \mathcal{B}|_{[1,T]}$ is equivalent to the constraint $\mathcal{H}^{\top}_{l+1}(w) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$, where $X^{\top} = P_l^{-1}[R_0 \ \ldots \ R_{l-1}] \ Q_l$.

Proof: The assumption of the theorem is stronger than the assumptions of Lemmas 1 and 2 because not only R_l is required to be of full row rank but its submatrix P_l is required to have this

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

property. In the direction of assuming $w \in \mathcal{B}|_{[1,T]}$, by Lemma 2, it follows that $R\mathcal{H}_{l+1}(w) = 0$. Since P_l is of full-row rank, $R\mathcal{H}_{l+1}(w) = 0$ is equivalent to $\mathcal{H}_{l+1}^{\top}(w) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$, with $X^{\top} := P_l^{-1}[R_0 \ \cdots \ R_{l-1} \ Q_l]$. In the opposite direction, by Lemma 1, $\mathcal{B} = \ker(\sum_{i=0}^{l} R_i \sigma^i)$ with $[R_0 \ R_1 \ \ldots \ R_l] := [X^{\top} \ -I]$. Therefore, $P_l = I$ is of full-row rank. ∎

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

$$\hat{\mathcal{B}} = \ker\left(\sum_{i=0}^{l} \hat{R}_i \sigma^i\right) \quad \hat{R}_l := [\hat{Q}_l \ -\hat{P}_l] \quad \text{with } \hat{P}_l \in \mathbb{R}^{p\times p} \text{ full row rank.} \quad (6)$$

We conjecture that (6) holds true for almost all $w \in (\mathbb{R}^{\mathtt{w}})^T$. Define the subset of $(\mathbb{R}^{\mathtt{w}})^T$ consisting of all time series $w \in (\mathbb{R}^{\mathtt{w}})^T$ for which the GTLS problem is equivalent to the STLS problem, i.e.,

$$\Omega := \left\{ w \in (\mathbb{R}^{\mathtt{w}})^T \;\middle|\; \begin{array}{c} \text{Problem (5) has a unique global} \\ \text{minimizer } \hat{\mathcal{B}} \text{ that satisfies (6)} \end{array} \right\}.$$

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

The existence and uniqueness part of the conjecture (see the definition of Ω) is justified in [20, Sec. 5.1]. The justification for (6) being generic is the following one. The highest possible order of a system in the model class $\mathcal{L}_{m,l}$ is pl. One can expect that generically in the data space $(\mathbb{R}^{\mathtt{w}})^T, \mathbf{n}(\hat{\mathcal{B}}) = pl$. By Lemma 2, $\mathbf{n}(\hat{\mathcal{B}}) = pl$ implies that in a kernel representation $\hat{\mathcal{B}} = \ker(\sum_{i=0}^{l} \hat{R}_i \sigma^i)$, $\hat{R}_l$ is of full row rank. But generically in $\mathbb{R}^{p\times \mathtt{W}}$ the matrix $\hat{P}_l \in \mathbb{R}^{p\times p}$, defined by $\hat{R}_l =: [\hat{Q}_l \ -\hat{P}_l]$, is of full-row rank. Although the justification for the conjecture is quite obvious, the proof seems to be rather involved.

A. Properties of the Solution

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 $\mathcal{B}^{\perp} \in \mathcal{L}_{p,l}$. Since the GTLS problem has as an inner minimization problem, the smoothing problem, the same properties hold in particular for the optimal solution of (5). These results are stated for the SISO case in [6] and then proven for the MIMO case in [14, Sec. VI].

Statistical properties of the GTLS problem, 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 [21]. Consistency in the MIMO case is proven in [20], in the framework of the GTLS problem. Complete statistical theory with practical confidence bounds is presented in [22], in the setting of the Markov estimator for semilinear models. Consistency of the STLS estimator for the general structure specification described in Appendix A is proven in [23].

B. Numerical Implementation

A *recursive solution* of the smoothing problem $M(w, \mathcal{B})$ is obtained by dynamic programming in [24]. An alternative derivation by isometric state representation is given in [14]. Both solutions are derived from a system theoretic point of view. A related problem occurs in the STLS formulation, where it is viewed from a numerical linear algebra point of view and is solved in a different way.

Because of the flexible structure specification, the inner minimization problem in the STLS formulation (1) is more general than the smoothing problem, where the block-Hankel structure is fixed. A closed form expression is derived and a special structure of the involved matrices is recognized, see [12] for details. The structure is then used on the level of the computation by employing numerical linear algorithms for structured matrices [25]. The resulting computational complexity is linear in the length T of the given time series w.

The outer minimization problem $\min_{\mathcal{B}\in\mathcal{M}} M(w, \mathcal{B})$, however, is a difficult nonconvex optimization problem that requires iterative methods. Two methods are proposed in the framework of the GTLS problem. In [14], an alternating least squares method is used. Its convergence is linear and can be very slow in certain cases. In [16], a Gauss–Newton algorithm is proposed. For the solution of the STLS problem, a Levenberg-Marquardt algorithm is used. More details on the implementation of the latter algorithm can be found in Appendix A, where a software package for solving the STLS problem (1) is described. The convergence of all these algorithms to the desired global minimum is *not* guaranteed and depends on the provided initial approximation and the given data.

IV. Extensions

A. Input/Output Partitionings

A standard assumption in system identification is that an input/output partitioning of the variables is a priori given. Consider a $\mathtt{w} \times \mathtt{w}$ permutation matrix Π and redefine w as Πw. The first m variables of the redefined time series are assumed to be inputs and the remaining p variables outputs. With $\operatorname{col}(u, y) := w$ and $[Q(z) \ -P(z)] := R(z)$, the kernel representation $R(\sigma)w = 0$ becomes a *left matrix fraction representation* $Q(\sigma)u = P(\sigma)y$. The transfer function of $\hat{\mathcal{B}}$ for the fixed by Π input/output partitioning is $H(z) := P^{-1}(z)Q(z)$.

Let $R_i =: [Q_i \ P_i]$ with P_i squares. Under the assumption $\mathbf{n}(\hat{\mathcal{B}}) = \mathbf{p}(\hat{\mathcal{B}})\mathbf{l}(\hat{\mathcal{B}})$, the state–space representation

$$\hat{A} = \begin{bmatrix} 0 & \cdots & 0 & -\hat{P}_0 \\ I & & & -\hat{P}_1 \\ & \ddots & & \vdots \\ & & I & -\hat{P}_{l-1} \end{bmatrix} \quad \hat{B} = \begin{bmatrix} \hat{Q}_0 - \hat{P}_0\hat{Q}_l \\ \hat{Q}_1 - \hat{P}_1\hat{Q}_l \\ \vdots \\ \hat{Q}_{l-1} - \hat{P}_{l-1}\hat{Q}_l \end{bmatrix}$$

$$\hat{C} = [0 \ \ldots \ 0 \ I] \qquad \hat{D} = \hat{Q}_l$$

is minimal. Therefore, the transition from $\hat{P}$ and $\hat{Q}$ (which is the result obtained from the STLS optimization problem) to an input/state/output representation is trivial and requires extra computations only for the formation of the $\hat{B}$ matrix.

Conjecture 1 implies that generically the optimal approximation $\hat{\mathcal{B}}$ admits an input/output partitioning $\operatorname{col}(u, y) := w$, with

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

$\Pi = I$. Moreover, we conjecture that generically $\hat{\mathcal{B}}$ admits an arbitrary input–output partitioning (i.e., $\mathrm{col}(u, y) := \Pi w$, with any permutation matrix Π).

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 $\mathcal{S}(w)$ by permuting the columns of $\mathcal{H}_{l+1}^{\top}(w)$. The software package described in Appendix A allows specification of exact blocks in $\mathcal{S}(w)$ that are not modified in the solution $\mathcal{S}(\hat{w})$. After solving the modified problem, the solution $\hat{X}$ of the original problem, with exact variables, is obtained by applying the reverse permutation.

With a given input–output partition (defined by a permutation matrix Π) and exact inputs, the GTLS problem becomes the output error identification problem

$$\hat{\mathcal{B}}_{\mathrm{oe}} := \arg \min_{\mathcal{B} \in \mathcal{L}_{m,l}} \left(\min_{\hat{y}} \|y - \hat{y}\|_{\ell_2}^2 \text{ s.t. } \Pi^{\top} \mathrm{col}(u, \hat{y}) \in \mathcal{B} \right).$$

In Problem 5, the approximating trajectory $\hat{w}$ is *any* trajectory of $\hat{\mathcal{B}}$, while in the output error identification problem, the approximating trajectory is generated by the *given* input u and only the initial conditions are freely chosen. In the single output, output error identification problem, the misfit

$$M_{\mathrm{oe}}(\mathrm{col}(u, y), \mathcal{B}) := \min_{\hat{y}} \|y - \hat{y}\|_{\ell_2}^2 \text{ subject to } \mathrm{col}(u, \hat{y}) \in \mathcal{B}$$

is equivalent to the cost function minimized by the prediction error methods. Simulation results are shown in Section VI-A.

C. Multiple Time Series

In certain cases, e.g., the noisy realization problem, not one but several observed time series $w_k, k = 1, \ldots, N$ are given. Assume that all time series are of the same length and define w to be the matrix valued time series $w = [w_1 \ \cdots \ w_N]$, so that $w(t) \in \mathbb{R}^{\mathtt{w} \times N}$. The only modification needed for this case is to consider block-Hankel matrix $\mathcal{H}_{l+1}(w)$ with size of the blocks $\mathtt{w} \times N$ instead of $\mathtt{w} \times 1$, as for the case of a single observed time series. The software package described in Appendix A can deal with such problems.

D. Known Initial Conditions

In the GTLS problem, no prior knowledge about initial conditions is assumed. Thus, the best fitting trajectory $\hat{w}$ is searched in the whole behavior $\hat{\mathcal{B}}$ of the approximating system. If the initial conditions are a priori known, $\hat{w}$ should be searched only among the trajectories of $\hat{\mathcal{B}}$, generated with the specified initial conditions. A typical examples of identification problems with known initial conditions are approximate realization and identification from step response observations. In both cases, the initial conditions are *a priori* known to be zero.

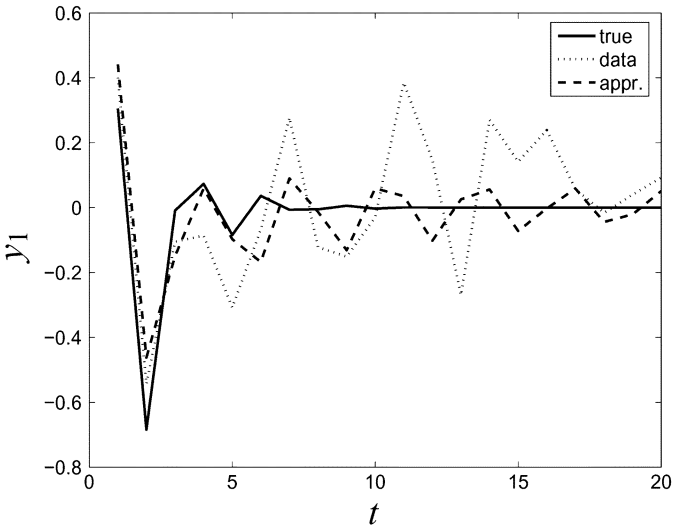

Fig. 2. Example of an autonomous identification problem in the errors-in-variables setting. Solid line—exact trajectory $\bar{y}$, dotted line—data y, dashed line—approximating trajectory $\hat{y}$.

Zero initial conditions can be taken into account in the identification problem by extending the given time series w with l zero samples. Let w_{ext} be the obtained in this way extended data sequences. In order to ensure that the approximation $\hat{w}_{\mathrm{ext}}$ is also obtained under zero initial conditions, the first l samples of w_{ext} should be preserved unmodified in $\hat{w}_{\mathrm{ext}}$.

Note 1: In the current software implementation of the GTLS method the specification that the l leading data samples are exact is *not* possible. This feature of the identification problem goes beyond the scope of the current STLS solution method and software.

E. Latent Inputs

The classical system identification framework [26] differs from the one in this paper in the choice of the optimization criterion and the model class. In [26], an unobserved input e, $\dim(e(t)) = n_e$ is assumed to act on the system that generates the observations and the optimization criterion is defined as the prediction error.

The unobserved input e, called *latent input*, plays the role of innovations. Written in a polynomial form, the model with latent inputs is the classical ARMAX model

$$P(\sigma)y = Q(\sigma)u + M(\sigma)e.$$

Latent input can be accommodated in the setting of Section III, by augmenting the model class $\mathcal{M} = \mathcal{L}_{m,l}$ 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

$$\min_{\mathcal{B} \in \mathcal{L}_{m+n_e,l}} \left(\min_{\hat{e}, \hat{w}} \underbrace{\|w - \hat{w}\|_{\ell_2}^2}_{\text{misfit}} + \underbrace{\|\hat{e}\|_{\ell_2}^2}_{\text{latency}} \text{ subject to } \begin{bmatrix} \hat{e} \\ \hat{w} \end{bmatrix} \in \mathcal{B} \right). \tag{7}$$

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

It unifies the misfit and latency description of the uncertainty and is put forward by Lemmerling and De Moor [7]. The pure latency identification problem

$$\min_{\mathcal{B}\in\mathcal{L}_{m+n_e,l}} \left(\min_{\hat{e}} \|\hat{e}\|_{\ell_2}^2 \text{ subject to } \begin{bmatrix} \hat{e} \\ w \end{bmatrix} \in \mathcal{B} \right) \tag{8}$$

corresponds to the prediction error approach.

The misfit-latency identification problem (7) can easily be reformulated as an equivalent pure misfit identification problem (5). Let $w_{\text{aug}} := \text{col}(e, w)$, 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 $\mathcal{L}_{m+n_e,l}$ is equivalent to (7). The pure latency identification problem (8) can also be treated in our framework by considering w exact (see Subsection IV-B) and modifying only e. Note that the latent input amounts to increasing the complexity of the model class, so that a better fit is achieved with a less powerful model.

V. Special Problems

In this section, we consider three special identification problems in an input/output setting. In the first one the data is an observed impulse response. In the second one the data is an observed free response. In the third one, the data is an exact impulse response of a high order system, i.e., a system that is not in the specified model class.

A. Approximate Realization

Identification from exact impulse response is the topic of (partial) realization theory. When the given data $h \in (\mathbb{R}^{p\times m})^{T+1}$ (impulse response observations) is not exact, an approximation is needed. Kung's algorithm is a well known solution for this problem. However, Kung's algorithm is suboptimal in terms of the misfit criterion

$$M_{\text{imp}}(h, \mathcal{B}) := \sqrt{\sum_{t=0}^{T} \|h(t) - \hat{h}(t)\|_{\Gamma}^2} \quad \text{where}$$

$$\hat{h} \text{ is an impulse response of } \mathcal{B}.$$

The GTLS problem can be used to find optimal in terms of the misfit $M_{\text{imp}}(h, \cdot)$ approximate model.

Problem 2 (Approximate realization [27]): Given a matrix valued time series $h \in (\mathbb{R}^{p\times m})^{T+1}$ and a natural number l, solve the optimization problem $\min_{\mathcal{B}\in\mathcal{L}_{m,l}} M_{\text{imp}}(h, \mathcal{B})$.

The approximate realization problem is a special GTLS problem and can be treated as such. Now, however, the given matrix valued trajectory is an observed impulse response, so that the input is a pulse and the initial conditions are zeros. For this reason the direct approach is inefficient. In the rest of this section, we describe an indirect solution that exploits the special features of the data.

The following statement is a corollary of Theorem 1.

Corollary 1: Consider a shortest lag kernel representation

$$\mathcal{B} := \ker(R(\sigma)) \in \mathcal{L}_{m,l} \quad R(z) = \sum_{i=0}^{l} R_i z^i$$

(i.e., R row proper) and define $\begin{bmatrix} Q_l & -P_l \end{bmatrix} := R_l$, where P_l is square. If P_l is nonsingular, then $h \in (\mathbb{R}^{p\times m})^{T+1}$ is an impulse response of $\mathcal{B}$ if and only if $\mathcal{H}_{l+1}^{\top}(\sigma h) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$, where

$$X^{\top} = -P_l^{-1}[P_0 \ P_1 \ \cdots \ P_l].$$

Corollary 1 shows that under assumption (6), the approximate realization problem can be solved as an STLS problem with structured data matrix $\mathcal{H}_{l+1}^{\top}(\sigma h)$. Next we specify how one can obtain an input/state/output representation $\mathcal{B}(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ of the optimal approximating system $\hat{\mathcal{B}}$ from $\hat{X}$ and the l approximated Markov parameters $\hat{h}(1), \ldots, \hat{h}(l)$.

By Corollary 1, $\text{rank}(\mathcal{H}_{l+1}^{\top}(\sigma\hat{h})) =: n = lp$. Let

$$\mathcal{H}_{l+1}(\sigma\hat{h}) = \Gamma\Delta$$

be a rank revealing factorization. Since $\hat{h}$ is an impulse response of $\hat{\mathcal{B}}, \Gamma$ and Δ must be of the form

$$\Gamma = \text{col}(\hat{C}, \hat{C}\hat{A}, \ldots, \hat{C}\hat{A}^l), \quad \Delta = [\hat{B} \ \hat{B}\hat{A} \ \cdots \ \hat{B}\hat{A}^l].$$

(The basis of the representation is fixed by the rank revealing factorization.) We have

$$\mathcal{H}_{l+1}^{\top}(\sigma\hat{h}) \begin{bmatrix} \hat{X} \\ -I \end{bmatrix} = 0 \Longrightarrow (\Gamma\Delta)^{\top} \begin{bmatrix} \hat{X} \\ -I \end{bmatrix} = 0$$
$$\Longrightarrow [\hat{X}^{\top} \ -I]\Gamma = 0$$

so that

$$\text{col span}(\Gamma) \subset \ker([\hat{X}^{\top} \ -I]).$$

However, $\dim(\text{col span}(\Gamma)) = n$. On the other hand

$$\dim(\ker([\hat{X}^{\top} \ -I])) = (l+1)p - p = n$$

so that $\text{col span}(\Gamma) = \ker([\hat{X}^{\top} \ -I])$. Therefore a basis for the null space of $[\hat{X}^{\top} \ -I]$ defines an observability matrix of $\hat{\mathcal{B}}$, from which $\hat{C}$ and $\hat{A}$ can be obtained up to a similarity transformation. $\hat{B}$ is the unique solution of the system

$$\Gamma\hat{B} = \text{col}(\hat{h}(1), \ldots, \hat{h}(l))$$

and $\hat{D} = h(0)$.

Example 1 (Approximate realization): Consider a simulation example in the errors-in-variables setup, i.e., the data $h = \bar{h} + \tilde{h}$ is obtained as a noise corrupted impulse response $\bar{h}$ of an LTI system $\bar{\mathcal{B}}$. The time horizon is $T = 50$ and the the additive noise standard deviation is 0.25. The true system $\bar{\mathcal{B}}$ is random stable (obtained via MATLAB's `drss` function) with $m = 2$ inputs, $p = 2$ outputs, and lag $l = 2$. The approximate model $\hat{\mathcal{B}}$ is searched in the model class $\mathcal{L}_{2,2}^4$.

We apply both a direct identification from input/output data (the impulse response data is extended with l zeros) and the indirect procedure described above. In the first case the optimization algorithm converges in 1.13 s and in the second case in 0.63 s, which shows the better efficiency of the indirect algorithm. The relative estimation error $\|\bar{h} - \hat{h}\|/\|\bar{h}\|$ in the first case is

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

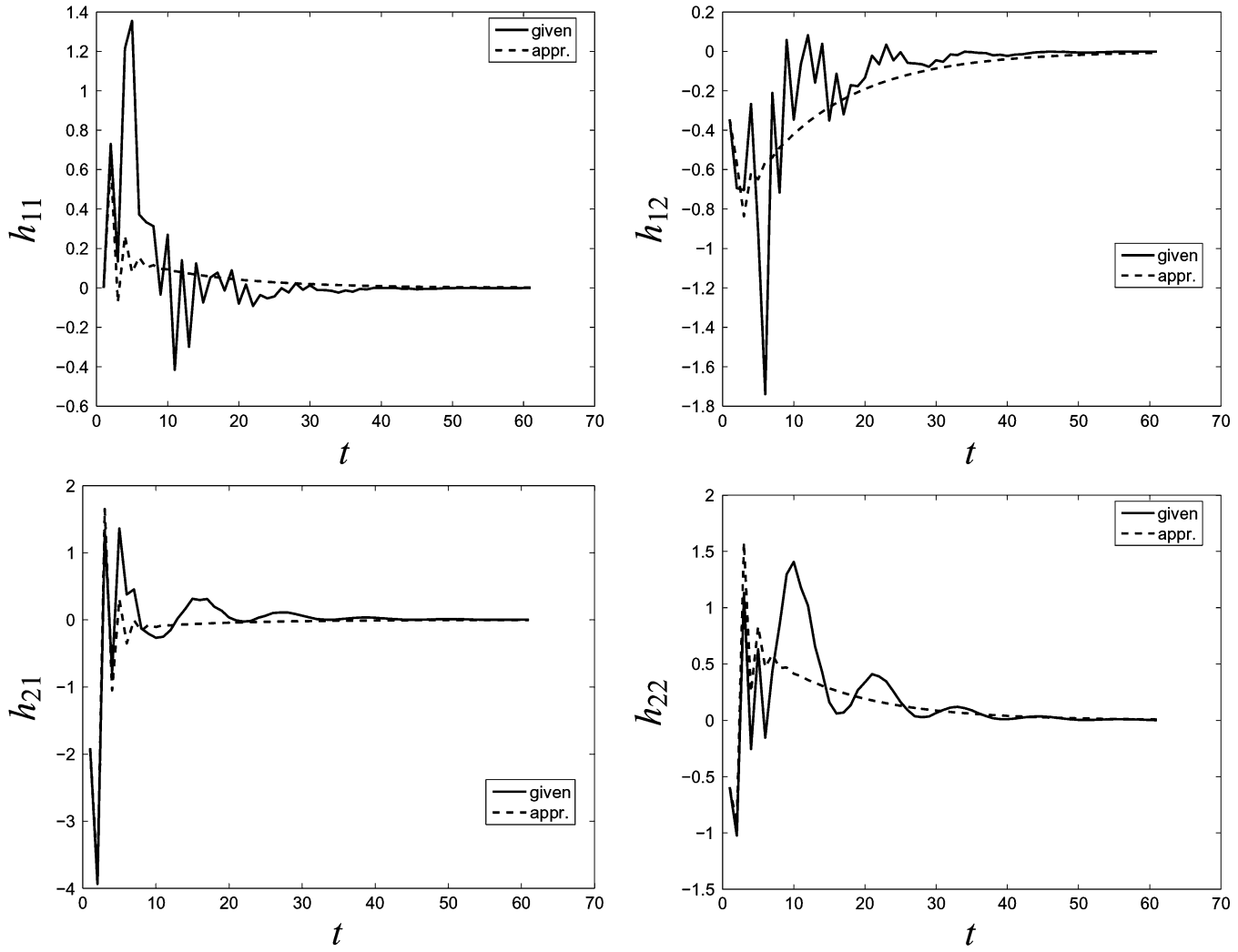

Fig. 3. Example of a finite time ℓ_2 model reduction problem. Solid line—impulse response of the given (high-order) system, dashed line—impulse response of the reduced order system.

0.2716 and in the second case 0.2608. (The difference is due to the wrong treatment, see Note 1, of the initial conditions in the direct method.) For comparison, the relative error with respect to the data h is 0.9219.

B. Identification of an Autonomous System

The autonomous system identification problem is a special case of the GTLS problem when the model class is $\mathcal{L}^p_{0,l}$—the set of autonomous systems with lag at most l.

Problem 3 (Autonomous System Identification): Given a time series $y \in (\mathbb{R}^p)^T$ and a natural number l, find a system $\hat{\mathcal{B}} \in \mathcal{L}_{0,l}$ and a free response $\hat{y}^*$ of $\hat{\mathcal{B}}$, such that $\hat{y}^*$ minimizes the approximation error $\|y - \hat{y}\|$ over all free responses of systems in the model class $\mathcal{L}^p_{0,l}$.

It is easy to reformulate Problem 3 as a special case of the approximate realization problem. The shifted impulse response σh of the system $\mathcal{B}(A, x_{\rm ini}, C, \cdot)$ is equal to the free response of the system $\mathcal{B}(A, \cdot, C, \cdot)$, obtained under initial condition $x_{\rm ini}$. Therefore, Problem 3 can be solved as an approximate realization problem with the obvious substitution. One can consider in the same way autonomous system identification for multiple time series $y_1, \ldots, y_N$.

Example 2 (Identification of an autonomous system): Consider the same simulation setup as in Example 1 with the only difference that the true data $\bar{y}$ is a single free response of length $T = 20$, obtained under random initial condition. The relative error of approximation $\|\bar{y} - \hat{y}\|/\|\bar{y}\|$ is 0.4184 versus 0.7269 for the given data y. Fig. 2, shows the fitting of $\bar{y}_1$ by $\hat{y}_1$.

C. Finite Time ℓ_2 Model Reduction

The finite time T, ℓ_2 norm of a system $\mathcal{B} \in \mathcal{L}_{m,l}$ with an impulse response h is defined as

$$\|\mathcal{B}\|_{\ell_2,T} := \|h|_{[0,T]}\| = \sqrt{\sum_{t=0}^{T} \|h(t)\|_{\rm F}^2}.$$

For a strictly stable system $\mathcal{B}$, $\|\mathcal{B}\|_{\ell_2,\infty}$ is well defined and is equal to its $\mathcal{H}_2$ norm.

Without loss of generality assume that the given time series h in the approximate realization problem is the *exact* impulse response of a (high order) system $\bar{\mathcal{B}}$. (Any *finite* time series $h \in (\mathbb{R}^{m\times p})^{T+1}$ can be considered as an impulse response of

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

TABLE I
EXAMPLES FROM DAISY. T—TIME HORIZON, m—NUMBER OF INPUTS, p—NUMBER OF OUTPUTS, l—LAG

#	Data set name	T	m	p	l
1	Heating system	801	1	1	2
2	Hair dryer	1000	1	1	5
3	Flexible robot arm	1024	1	1	4
4	Heat flow density	1680	2	1	2
5	Steam heat exchanger	4000	1	1	2

a system in the model class $\mathcal{L}_{m,pT}$.) Then, the approximate realization problem can be interpreted as the following finite time ℓ_2 model reduction problem.

Problem 4 (Finite Time ℓ_2 Model Reduction): Given a system $\bar{\mathcal{B}} \in \mathcal{L}_{m,l}$, a natural number $l_{\rm red} < l$, and a time horizon T, find a system $\hat{\mathcal{B}} \in \mathcal{L}_{m,l_{\rm red}}$, that minimizes the finite time T, ℓ_2 norm $||\bar{\mathcal{B}} - \hat{\mathcal{B}}||_{\ell_2,T}$ of the error system.

In the model reduction problem, the misfit is due to the low order approximation. In the approximate realization problem, assuming that the data is generated by an errors-in-variables model, the misfit is due to the measurement error $\tilde{h}$. The solution methods, however, are equivalent, so in this section we actually give an alternative interpretation of the approximate realization problem.

Example 3 (Finite Time ℓ_2 Model Reduction): The high-order system $\bar{\mathcal{B}}$ is a random stable system (obtained via MATLAB's function `drss`) with $m = 2$ inputs, $p = 2$ outputs, and lag $l = 10$. A reduced-order model $\hat{\mathcal{B}}$ with lag $l_{\rm red} = 1$ is searched. The time horizon T is chosen large enough for a sufficient decay of the impulse response of $\bar{\mathcal{B}}$. (It is selected automatically by MATLAB's function `impulse`.)

Fig. 3 shows the fitting of the impulse response of the high-order system $\bar{\mathcal{B}}$ by the impulse response of the reduced-order system $\hat{\mathcal{B}}$.

VI. PERFORMANCE ON DATA SETS FROM DAISY

The data base for system identification DAISY [8] is used for verification and comparison of identification algorithms. In this section, we apply the GTLS method, described in the paper and implemented by the software, presented in Appendix A, on data sets from DAISY. In Section VI-A, we solve output error identification problems and in Section VI.B, we consider the data set "Step response of a fractional distillation column," which consists of multiple vector time series.

A. Single Time Series Data Sets

The considered data sets are listed in Table I. Since all data sets are with a given input/output partitioning, the only user defined parameter that selects the complexity of the model class $\mathcal{M} = \mathcal{L}^{m+p}_{m,l}$ is the lag l.

The data is detrended and split into identification and validation data sets. The first 70% of the data, denoted by $w_{\rm id}$, is used for identification, and the remaining 30%, denoted by $w_{\rm val}$, is used for validation.

Approximate models are computed via the following methods:

- `n4sid` the N4SID method implemented in the System Identification Toolbox of MATLAB;
- `gtls` the GTLS method implemented by the STLS solver;
- `pem` the prediction error method of the System Identification Toolbox of MATLAB.

TABLE II
COMPARISON OF THE MODELS OBTAINED BY `n4sid`, `gtls`, AND `pem`

#	Data set name	Function	Fit %	Misfit
1	Heating system	`n4sid`	51.9971	140.8018
		`gtls`	76.0491	70.2527
		`pem`	76.0491	70.2527
2	Hair dryer	`n4sid`	88.3265	1.5219
		`gtls`	90.8722	1.1900
		`pem`	90.8772	1.1893
3	Flexible robot arm	`n4sid`	29.5496	3.2480
		`gtls`	96.5454	0.1593
		`pem`	96.5454	0.1593
4	Heat flow density	`n4sid`	40.7249	11.2233
		`gtls`	83.8574	3.0565
		`pem`	83.8574	3.0565
5	Steam heat exchanger	`n4sid`	29.6890	25.5047
		`gtls`	60.4452	14.3481
		`pem`	60.1575	14.4525

The inputs are assumed exact, so that identification in the output error setting is considered. The validation is performed in terms of the misfit $M_{\rm oe}(w_{\rm val}, \hat{\mathcal{B}})$ obtained on the validation data set and the simulation fit computed by the function `compare` from the System Identification Toolbox.

$$\texttt{compare}'\text{s simulation fit} = 100(1 - M_{\rm oe}/||y||)$$

so that these two criteria are equivalent.

Note 2 (About the Usage of the Methods): The `pem` function is called with the option

```
'DisturbanceModel', 'None'
```

which specifies output error model structure. In addition, the options

```
'nk', 0, 'LimitError', 0
```

and `'alg'` are used to disable the default for `pem` feedthrough term set to zero, robustification of the cost function, and stability constraint. (The GTLS method does not constrain the model class by enforcing stability.) With these options (for the single output case) `pem` minimizes the output error misfit $M_{\rm oe}$. The `gtls` function is called with the specification that the inputs are exact, so that the GTLS and PEM methods solve equivalent identification problems. For both functions, we set the same convergence tolerance (`'Tolerance',1e-10`), maximum number of iterations (`'Maxiter',100`), and initial approximation (the model obtained by `n4sid`).

The identified systems by `n4sid`, `gtls`, and `pem` are compared in Table II. In all examples there is a good match between the models obtained with the `gtls` and `pem` functions. In addition, the output error optimal model outperforms the model computed by the N4SID method. Since the criterion is checked on a part of the data that is not used for identification, there is no *a priori* guarantee that this will be the case.

B. Identification From Step Response Measurements

Next, we consider the data set "Step response of a fractional distillation column" from DAISY. It consists of three indepen-

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

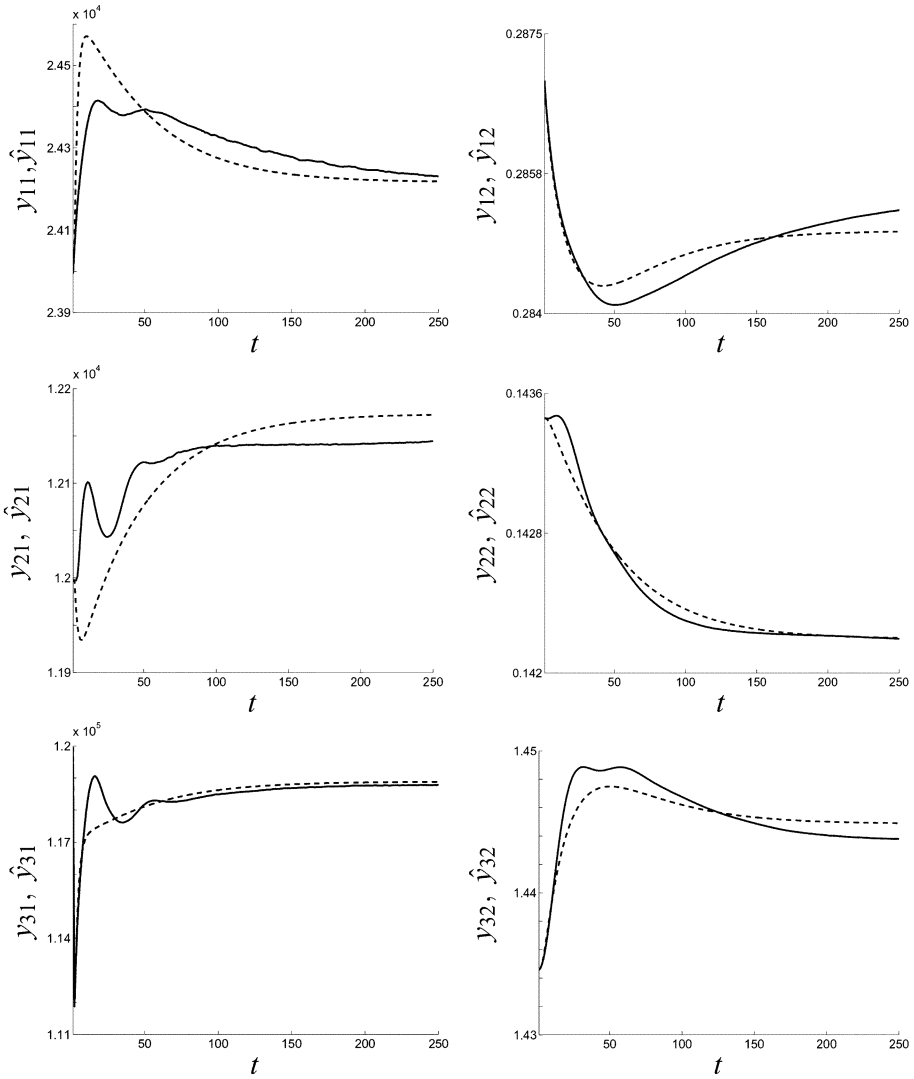

Fig. 4. Identification from step response measurements. Solid line—given data y, dashed line—GTLS approximation $\hat{y}$. (y_{ij} is the step response from input i to output j.)

dent time series, each one with $T = 250$ data points. The given data has a fixed input/output partitioning with $m = 3$ inputs and $p = 2$ outputs. We further bound the complexity of the model class by choosing the lag $l = 2$, so that an approximate model is searched in the model class $\mathcal{L}^5_{2,2}$.

The step response data is special because, it consists of multiple time series, the input is exactly known, and the initial conditions are also exactly known. In order to take into account the known zero initial conditions, we precede the given time series with l zero samples. In order to take into account the exactly known inputs, we use the modification of the GTLS method for time series with exact variables. Multiple time series are processed as explained in Section IV.

Fig. 4 shows the given data y (the measured step responses) with superimposed on them the step responses of the optimal approximating system, computed by the GTLS method.

VII. Conclusion

We generalized previous results on the application of STLS for system identification, approximate realization, and model reduction 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 exact variables can be taken into account.

A robust and efficient software tool for solving STLS problems is presented. It makes the proposed identification method and its extensions practically applicable. The performance of the software package is tested on data sets from DAISY. The results show that examples with a few thousands data points and low order model can be solved routinely.

Appendix

A. Software Package for Solving STLS Problems [13]

Efficient numerical methods for STLS problems were developed in the past, see [28]. 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 [1], [2] have

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.

cubic computational complexity and are restricted to rank reduction of order one. Thus, they are applicable only for small size SISO problems.

The described package solves the STLS problem (1) with $\mathcal{S}(\cdot) = [C^{(1)} \ \cdots \ C^{(q)}]$, 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 elements of the same row dimension N.

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

The package uses MINPACK's Levenberg–Marquardt algorithm to solve the optimization problem (1) in its equivalent formulation, see [11]

$$\min_X \ r^\top(X)\Gamma^{-1}(X)r(X)$$
$$\text{where } r(X) := \operatorname{vec}\left(\left(\mathcal{S}(w)\begin{bmatrix} X \\ -I \end{bmatrix}\right)^\top\right).$$

The weight matrix Γ is block-Toeplitz and block-banded structured, see [12], and this structure is exploited in the cost function and first derivative evaluation. There is no closed form expression for the Jacobian matrix $J = [\partial r_i/\partial x_j]$, where $x = \operatorname{vec}(X)$, so that the pseudo-Jacobian J_+ proposed in [29] is used instead of J. The cost function and pseudo-Jacobian evaluation is performed with computational complexity $O(T)$.

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 [30]. MATLAB interface via C-mex file is available.

An interface to the STLS solver for the purpose of approximate system identification is developed in [31]. The MATLAB `gtls` implements the mapping $(w,m,l) \mapsto (\hat{A},\hat{B},\hat{C},\hat{D})$, where $\mathcal{B}(\hat{A},\hat{B},\hat{C},\hat{D})$ is a GTLS optimal model. The function works with data consisting of multiple time series of equal length and allows for specification of exact variables. The software is available from http://www.esat.kuleuven.be/~imarkovs

ACKNOWLEDGMENT

The authors would like to thank D. Sima, the Guest Editors L. Ljung and A. Vicino, and the referees for carefully reading the paper and for giving them valuable suggestions.

REFERENCES

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

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

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

[4] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

[5] M. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. Autom. Control*, vol. AC-9, pp. 229–235, Jul. 1964.

[6] B. De Moor and B. Roorda, "L_2-optimal linear system identification structured total least squares for SISO systems," in *Proc. Conf. Decision and Control*, 1994, pp. 2874–2879.

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

[8] B. De Moor, "DaISy: Database for the identification of systems,", Dept. Elect. Eng, K. U. Leuven, Leuven, Belgium, 1998, www.esat.kuleuven.be/sista/daisy/.

[9] B. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–31, Jan. 1981.

[10] S. Kung, "A new identification method and model reduction algorithm via singular value decomposition," in *Proc. 12th Asilomar Conf. on Circuits, Systems, and Computing*, 1978, pp. 705–714.

[11] I. Markovsky, S. Van Huffel, and A. Kukush, "On the computation of the structured total least squares estimator," *Num. Lin. Alg. Appl.*, vol. 11, pp. 591–608, 2004.

[12] I. Markovsky, S. Van Huffel, and R. Pintelon, "Block-Toeplitz/Hankel structured total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 4, pp. 1083–1099, 2005.

[13] I. Markovsky and S. Van Huffel, "High-performance numerical algorithms and software for structured total least squares," *J. Comput. Appl. Math.*, vol. 180, no. 2, pp. 311–331, 2005.

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

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

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

[17] J. C. Willems, "From time series to linear system—Part II. Exact modeling," *Automatica*, vol. 22, no. 6, pp. 675–694, 1986.

[18] J. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory*. New York: Springer-Verlag, 1998.

[19] 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.

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

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

[22] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. Piscataway, NJ: IEEE Press, 2001.

[23] A. Kukush, I. Markovsky, and S. Van Huffel, "Consistency of the structured total least squares estimator in a multivariate errors-in-variables model," *J. Statist. Planning Inference*, vol. 133, no. 2, pp. 315–358, 2005.

[24] I. Markovsky and B. De Moor, "Linear dynamic filtering with noisy input and output," *Automatica*, vol. 41, no. 1, pp. 167–171, 2005.

[25] T. Kailath and A. Sayed, *Fast Reliable Algorithms for Matrices with Structure*. Philadelphia, PA: SIAM, 1999.

[26] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[27] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," in *IEEE Trans. Signal Process.*, vol. 42, Nov. 1994, pp. 3104–3113.

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

[29] P. Guillaume and R. Pintelon, "A Gauss–Newton-like optimization algorithm for "weighted" nonlinear least-squares problems," in *IEEE Trans. Signal Process.*, vol. 44, Sep. 1996, pp. 2222–2228.

[30] S. Van Huffel, V. Sima, A. Varga, S. Hammarling, and F. Delebecque, "High-performance numerical software for control," *IEEE Control Syst. Mag.*, vol. 24, no. 1, pp. 60–76, Feb. 2004.

[31] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, "Software for approximate linear system identification," in *Proc. Conf. Decision Control*, Seville, Spain, 2005.

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.



Ivan Markovsky was born in Sofia, Bulgaria, in 1974. He received the M.S. degree in control and systems engineering from the Technical University of Sofia, in 1998, and the Ph.D. degree in electrical engineering from K. U. Leuven, Leuven, Belgium, in 2005.

Since February 2005, he has been a Postdoctoral Researcher in the Electrical Engineering Department of K.U. Leuven. His current research work is focused on identification methods in the behavioral setting and errors-in-variables estimation problems.



Jan C. Willems was born in Bruges, Flanders, Belgium. He studied engineering at the University of Ghent. After his graduation in 1963, he received the M.Sc. degree from the University of Rhode Island, Providence, in 1965, and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1968, both in electrical engineering.

He was an Assistant Professor in the Department of Electrical Engineering, MIT from 1968 to 1973, with a one-year leave of absence with the Department of Applied Mathematics and Theoretical Physics of Cambridge University, Cambridge, U.K. In 1973, he was appointed Professor of Systems and Control in the Mathematics Department of the University of Groningen, Groningen, The Netherlands. In 2003, he became Professor Emeritus at the University of Groningen. Currently, he is a full-time Visiting Professor at the Department of Electrical Engineering, with the research group on Signals, Identification, System Theory and Automation (SISTA), K.U. Leuven, Leuven, Belgium. During the academic year 2003–2004, he held the Francqui Chair at the Faculty of Applied Sciences of the Université Catholique de Louvain. His research interests involve various aspects of systems theory and control, especially the development of the behavioral approach.

Dr. Willems has served terms as Chairperson of the European Union Control Association and of the Dutch Mathematical Society. He has been on the Editorial Board of a number of journals, in particular, as Managing Editor of the *SIAM Journal of Control and Optimization* as and Founding and Managing Editor of *Systems and Control Letters*. In 1998, he received the IEEE Control Systems award.



Sabine Van Huffel received the M.D. degree in computer science engineering in June 1981, the M.D. degree in biomedical engineering in July 1985, and the Ph.D. degree in electrical engineering in June 1987, all from K. U. Leuven, Leuven, Belgium.

She is a Full Professor at the Department of Electrical Engineering from K. U. Leuven. Her research interests are in signal processing, numerical linear algebra, errors-in-variables regression, system identification, pattern recognition, (non)linear modeling, software, statistics, applied to biomedicine. In these areas, she has authored one book, entitled *The Total Least Squares Problem: Computational Aspects and Analysis* (Philadelphia, PA: SIAM, 1991), and more than 130 papers in international journals and 140 conference contributions.



Bart De Moor received the M.S. degree and a Ph.D. in electrical engineering at the K. U. Leuven, Leuven, Belgium, in 1983 and 1988, respectively.

He was a Visiting Research Associate at Stanford University, Stanford, CA (1988–1990). Currently, he is a Full Professor at the Department of Electrical Engineering of the K. U. Leuven. His research interests are in numerical linear algebra and optimization, system theory, control and identification, quantum information theory, data-mining, information retrieval, and bio-informatics, in which he (co-)authored more than 400 papers and three books.

Dr. De Moor's work has won him several scientific awards [Leybold-Heraeus Prize (1986), Leslie Fox Prize (1989), Guillemin-Cauer best paper Award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1990), Laureate of the Belgian Royal Academy of Sciences (1992), bi-annual Siemens Award (1994), best paper award of *Automatica* (IFAC, 1996), IEEE Signal Processing Society Best Paper Award (1999)]. From 1991 to 1999, he was the Chief Advisor on Science and Technology of several ministers of the Belgian Federal and the Flanders Regional Governments. He is on the board of three spin-off companies, of the Flemish Interuniversity Institute for Biotechnology, the Study Center for Nuclear Energy, and several other scientific and cultural organizations. Since 2002, he also makes regular television appearances in the Science Show "Hoe?Zo!" on national television in Belgium. Full biographical details can be found at www.esat.kuleuven.be/~demoor.



Rik Pintelon (M'90–SM'96–F'98) was born in Gent, Belgium, on December 4, 1959. He received the degree of electrical engineer (burgerlijk ingenieur) in July 1982, the degree of Doctor in applied sciences in January 1988, and the qualification to teach at university level (geaggregeerde voor het hoger onderwijs) in April 1994, all from the Vrije Universiteit Brussel (VUB), Brussels, Belgium.

From October 1982 to September 2000, he was a Researcher of the Fund for Scientific Research—Flanders at the VUB. Since October 2000, he has been a Professor at the VUB in the Electrical Measurement Department (ELEC). His main research interests are in the field of parameter estimation/system identification, and signal processing.

Authorized licensed use limited to: Katholieke Universiteit Leuven. Downloaded on October 14, 2008 at 07:26 from IEEE Xplore. Restrictions apply.