Application of structured total least squares for system identification

Ivan Markovsky

K.U.Leuven, ESAT-SISTA

Joint work with

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

Approximate modeling problems

$\bar{\Sigma}$ — "true" (high order) model

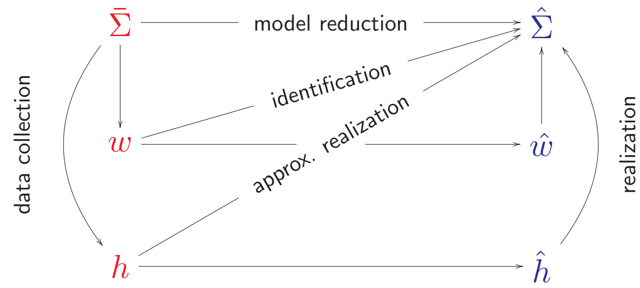
$\hat{\Sigma}$ — approximate (low order) model

w — observed response

h — observed impulse resp.

$\hat{w}$ — response of $\hat{\Sigma}$

$\hat{h}$ — impulse resp. of $\hat{\Sigma}$



Solution methods

kernel subproblem:
find a block-Hankel rank deficient matrix $\mathcal{H}(\hat{w})$ approximating a given full rank matrix $\mathcal{H}(w)$

SVD-based methods:
balanced model reduction, subspace identification, and Kung's algorithm use the singular value decomposition in order to solve the kernel problem

our purpose:
solve optimal according to the misfit criterion $\|w-\hat{w}\|_{\ell_2}^2$ approximate modeling problems

note that SVD is suboptimal in terms of the misfit criterion

Structured total least squares

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

STLS problem: given a time series w, find

$$\hat{X} := \arg\min_X \left(\min_{\hat{w}} \; \|w-\hat{w}\|_{\ell_2}^2 \quad \text{s.t.} \quad \mathcal{H}(\hat{w}) \begin{bmatrix} X \\ -I \end{bmatrix} = 0 \right)$$

note: $\mathcal{H}(\hat{w}) \left[\begin{smallmatrix} X \\ -I \end{smallmatrix}\right] = 0 \iff \text{rank}\big(\mathcal{H}(\hat{w})\big) \leq \text{row}\dim(X)$

$$\|w-\hat{w}\|_{\ell_2}^2 = \text{dist}(w, \hat{w})$$

Algorithm and software

- nonconvex optimization problem, we use local optimization methods
- the inner minimization is solved analytically $\rightsquigarrow$ nonlinear LS
- the cost function and Jacobian are evaluated in $O(T)$ flops by exploiting the structure of the involved matrices
- C software is available at:

  `http://www.esat.kuleuven.ac.be/~imarkovs/stls.html`

Approximate system identification

$\mathcal{M}$ — user specified model class $\quad w$ — given time series

the model $\mathcal{B} \in \mathcal{M}$ is a collection of legitimate time series

the more the model forbids, the less complex and more powerful it is

problem: find a $\hat{\mathcal{B}} \in \mathcal{M}$ that best fits the data according to the

misfit criterion: $\quad M(w, \mathcal{B}) := \min_{\hat{w} \in \mathcal{B}} \|w - \hat{w}\|_{\ell_2}^2 \quad$ (smoothing problem)

the resulting identification problem is:

global total least squares (GTLS) problem: $\quad \min_{\mathcal{B} \in \mathcal{M}} M(w, \mathcal{B})$

Model class $\mathcal{L}_{m,l}$

$\mathcal{L}_{m,l}$ — the set of all LTI systems with m inputs and lag at most l

m and l specify the complexity of the model class $\mathcal{L}_{m,l}$

$\mathcal{B}|_{[1,T]}$ — the restriction of $\mathcal{B}$ to the interval $[1, T]$

for $\mathcal{B} \in \mathcal{L}_{m,l}$ and T sufficiently large, $\dim(\mathcal{B}|_{[1,T]}) = mT + n \leq mT + lp$

goal: solve the GTLS problem as an STLS problem

GTLS $\overset{?}{\equiv}$ STLS

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

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

need to ensure that $\quad \mathcal{H}(\hat{w}) \left[\begin{smallmatrix} X \\ -I \end{smallmatrix} \right] = 0 \iff \hat{w} \in \mathcal{B} \in \mathcal{L}_{m,l}$

as a byproduct of doing this, we relate the parameter X to $\mathcal{B}$

Main result

assume that $\mathcal{B} \in \mathcal{L}_{m,l}$ admits a representation

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

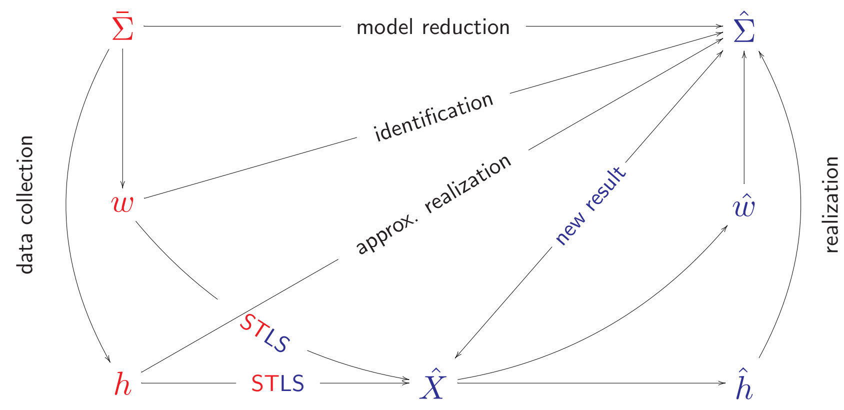
with $R_l =: \begin{bmatrix} Q_l & P_l \end{bmatrix}$, $P_l \in \mathbb{R}^{p \times p}$ full rank, and let

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

then $\quad w \in \mathcal{B}|_{[1,T]} \quad \iff \quad \mathcal{H}_{l+1}^\top(w) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$

conjecture: the assumption of the theorem holds generically in the data space $(\mathbb{R}^{\mathtt{w}})^T$ for the optimal approximate model $\hat{\mathcal{B}}$

Main result



Extensions of the identification problem

given input/output partitioning: $\quad w = \begin{bmatrix} u \\ y \end{bmatrix} \qquad$ (u input, y output)

with $R(\xi) =: \begin{bmatrix} Q(\xi) & P(\xi) \end{bmatrix}$, $\quad R(\sigma)w = 0 \implies P(\sigma)y = -Q(\sigma)u$

the transfer function of $\hat{\mathcal{B}}$ is $H(z) := -P^{-1}(z)Q(z)$

exact variables: $\quad w = \begin{bmatrix} u \\ y \end{bmatrix}$, $\hat{w} = \begin{bmatrix} \hat{u} \\ \hat{y} \end{bmatrix} \qquad$ u exact $\implies \hat{u} = u$

such a constraint can be specified in the STLS software package

Extensions of the identification problem

multiple time series:

given N time series $w_1, \ldots, w_N$ of the same length T

define a matrix valued time series $\quad w(t) = \begin{bmatrix} w_1(t) & \cdots & w_N(t) \end{bmatrix}$

$$w \in \mathcal{B}|_{[1,T]} \quad :\iff \quad w_1 \in \mathcal{B}|_{[1,T]}, \ldots, w_N \in \mathcal{B}|_{[1,T]}$$

find $\mathcal{B} \in \mathcal{L}_{m,l}$ that approximates simultaneously $w_1, \ldots, w_N$

misfit for matrix valued w: $M(w, \mathcal{B}) = \min_{\hat{w} \in \mathcal{B}} \sum_{i=1}^{N} \|w^{(i)} - \hat{w}^{(i)}\|_{\ell_2}^2$

$\rightsquigarrow$ STLS problem with $N \times \mathtt{w}$ size block of the block-Hankel matrix

Results on data set "power plant" from DAISY

DAISY — data base for system identification, available from

`http://www.esat.kuleuven.ac.be/~tokka/daisydata.html`

real-life and simulated data for verification and comparison of ident. alg.

data set "Data of a power plant (Pont-sur-Sambre, France) of 120MW"

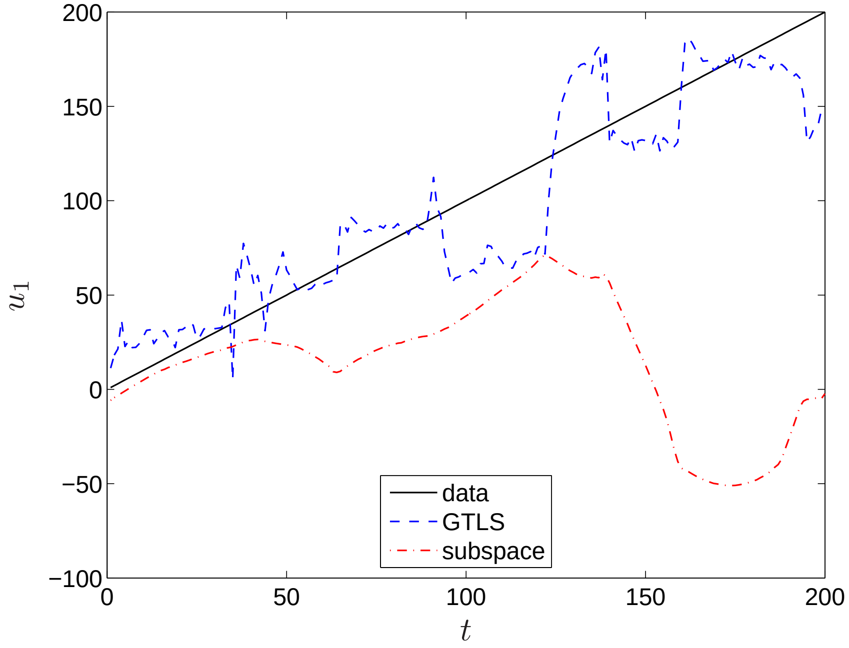
$m = 5$ inputs, $p = 3$ outputs, $T = 200$ data points

find a GTLS optimal model $\hat{\mathcal{B}}$ with lag $l = 2$

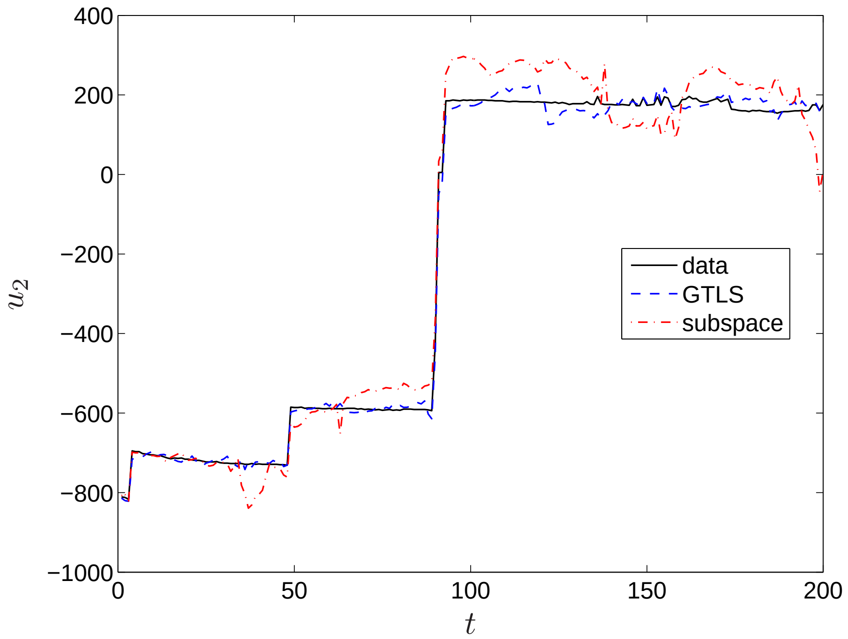
result: $M(w, \mathcal{B}_{\text{ini}}) = 8973$, $M(w, \hat{\mathcal{B}}) = 607$

the initial approx. is obtained from a subspace ident. method

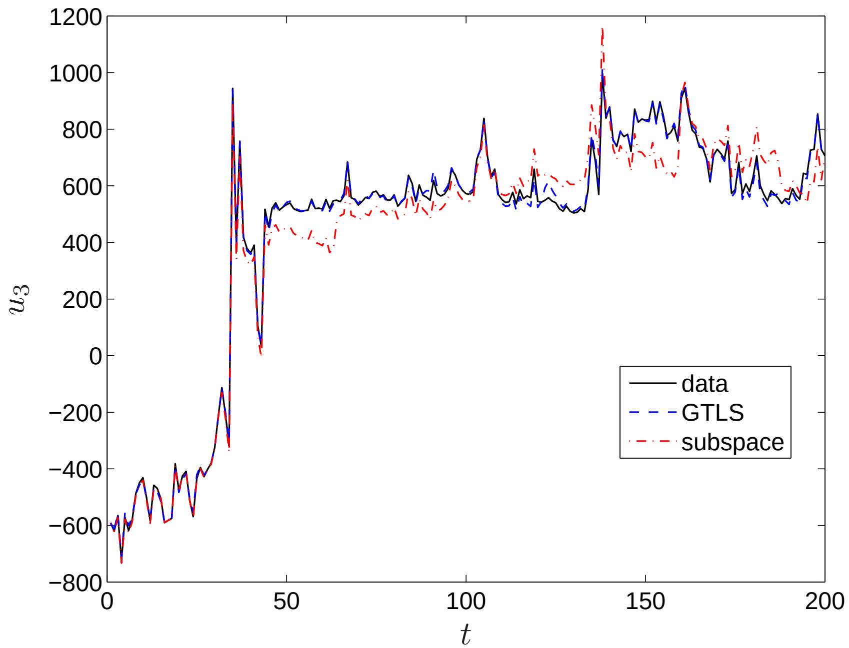
Input 1



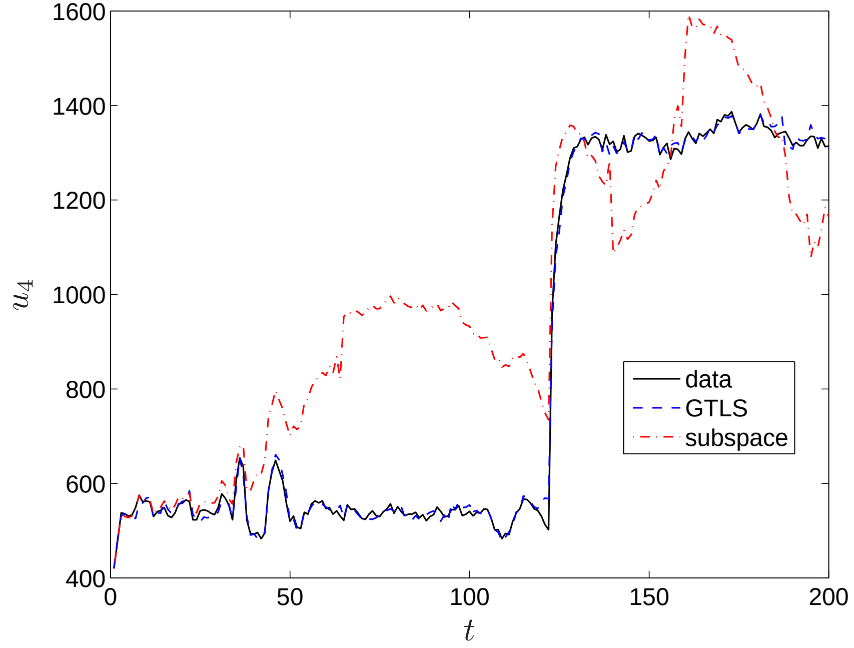
Input 2



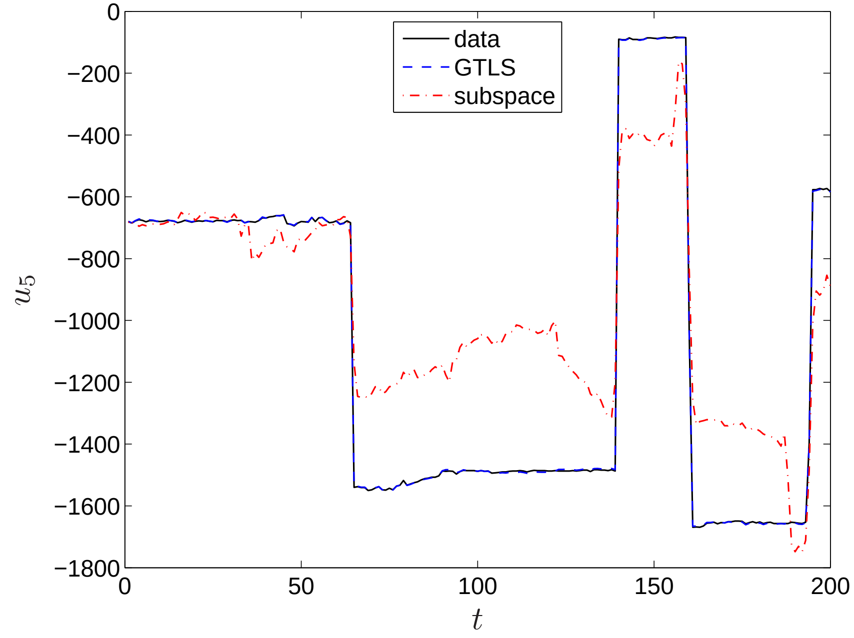
Input 3



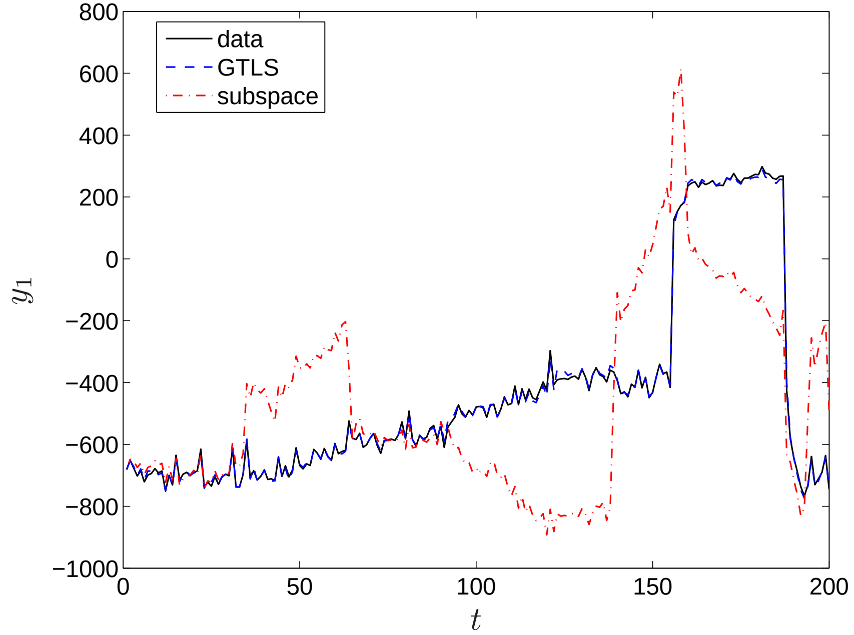
Input 4



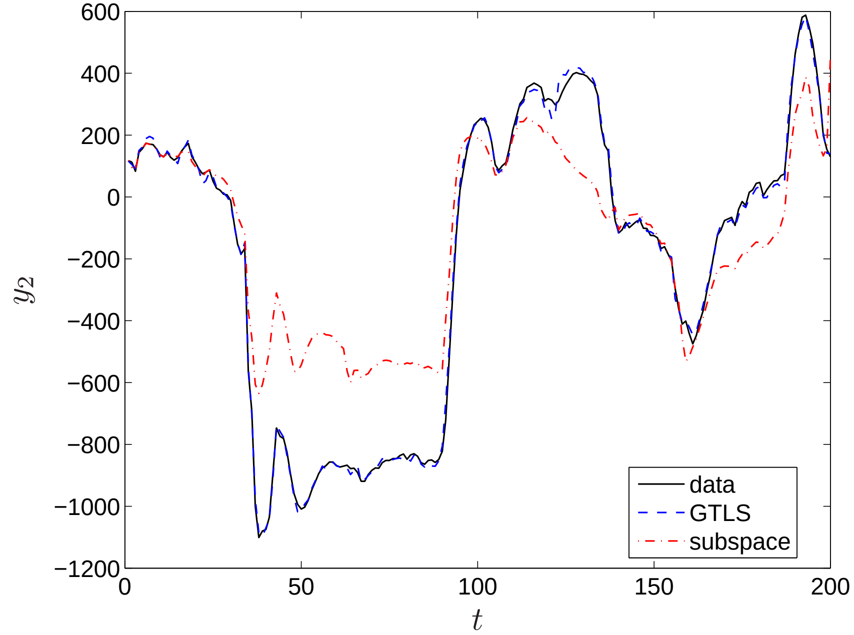
Input 5



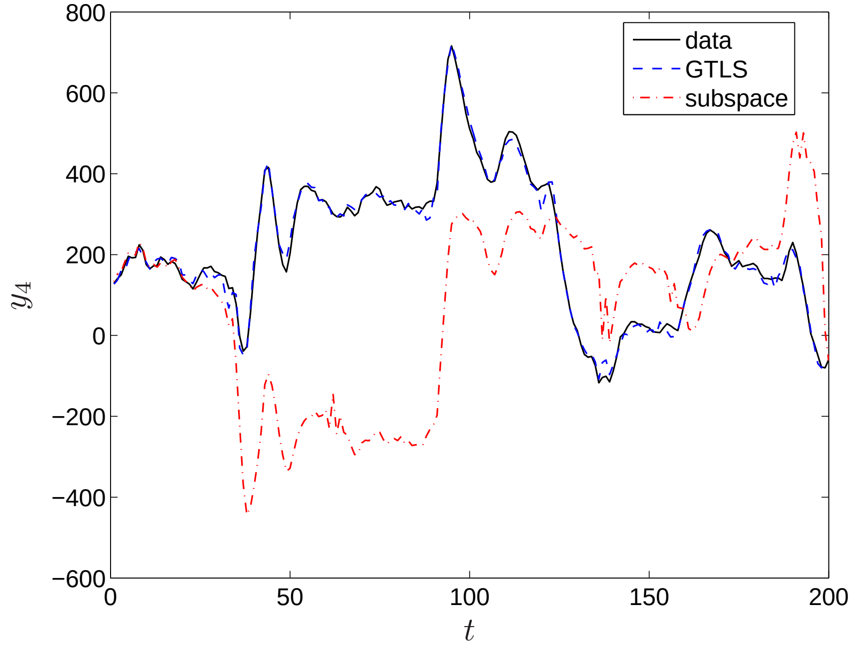
Output 1



Output 2



Output 3



Conclusions

- Hankel low rank approximation ($\approx$ STLS) is a kernel problem for approximate LTI modeling (model reduction, system ident., *etc.*)
- there is an efficient software tool that solves the kernel problem
- it allows to solve non-toy identification problems

future work:

- extend the GTLS framework with unobserved (latent) variables
- find link with the prediction error methods

References

- J. C. Willems. From time series to linear system—Part I. Finite dimensional linear time invariant systems. Part II. Exact modelling. Part III. Approximate modelling. *Automatica*, 22(5):561–580, 1986, 22(6):675–694, 1986, 23(1):87–115, 1987.
- B. De Moor. Structured total least squares and L_2 approximation problems. *Lin. Alg. and Its Appl.*, 188–189:163–207, 1993.
- B. De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Trans. on Signal Proc.*, 42(11):3104–3113, 1994.
- B. Roorda and C. Heij. Global total least squares modeling of multivariate time series. *IEEE Trans. on Aut. Control*, 40(1):50–63, 1995.
- B. Roorda. Algorithms for global total least squares modelling of finite multivariable time series. *Automatica*, 31(3):391–404, 1995.