Application of structured total least squares for system identification

Ivan Markovsky
K.U.Leuven, SCD

Outline

- STLS and approximate modeling
- Kernel representation and lag structure
- The identification problem: global total least squares
- Solution by structured total least squares
- Extensions of the identification problem
- Software package for solving STLS problems
- Results on data sets from DAISY
- Conclusions

Structured matrices

given a injective mapping $\mathcal{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m\times(n+d)}$, we say that the matrix $C \in \mathbb{R}^{m\times n+d}$ is $\mathcal{S}$-structured if $C \in \text{image}(\mathcal{S})$

let C be $\mathcal{S}$-structured, then the vector $p \in \mathbb{R}^{n_p}$, such that $C = \mathcal{S}(p)$, is called the parameter vector of C

respectively, $\mathbb{R}^{n_p}$ is called the parameter space of the structure $\mathcal{S}$

in this talk, of interest is the block-Hankel structure, denoted by $\mathcal{H}$

Structured total least squares

the STLS problem: given a time series w and a structure specification $\mathcal{S}$, find the global minimum point of the optimization problem

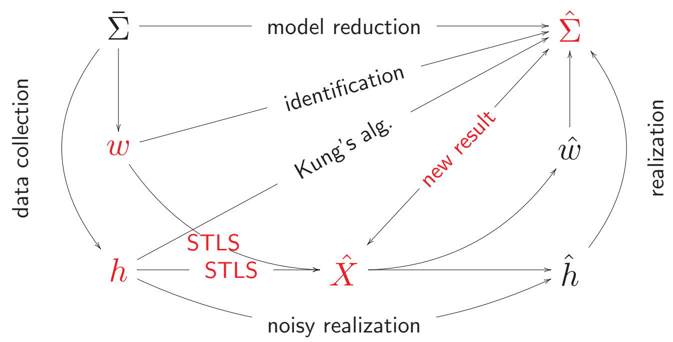
$$\min_X \left(\min_{\hat{w}} \ \|w - \hat{w}\|_{\ell_2}^2 \quad \text{s.t.} \quad \mathcal{S}(\hat{w}) \begin{bmatrix} X \\ -I \end{bmatrix} = 0 \right)$$

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

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

STLS finds optimal structured low rank approximation of $\mathcal{S}(w)$ by $\mathcal{S}(\hat{w})$

Approximate modeling problems



$\hat{\Sigma}$ — (low complexity) approximating model

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

w — observed general response

h — observed impulse response

SVD vs STLS

errors-in-variables model: $w = \bar{w} + \tilde{w}$ $\quad$ $\bar{w}$ trajectory of $\bar{\Sigma}$, $\tilde{w}$ measurement noise

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

non-iterative methods:
balanced model reduction, subspace identification, Kung's algorithm solve the kernel problem via SVD, suboptimal with respect to $\|w - \hat{w}\|_{\ell_2}^2$

our purpose: solve optimal approximate modeling problems by STLS

subsequently make use of efficient numerical methods for STLS

The global total least squares problem

$\mathcal{M}$ — user specified model class $\qquad$ 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: $M(w, \mathcal{B}) := \min_{\hat{w} \in \mathcal{B}} \|w - \hat{w}\|_{\ell_2}^2$ (smoothing problem)

the resulting optimization problem is the

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

Difference equation representation

let $\mathcal{H}_l(\bullet)$ be a Hankel matrix with l block-rows

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

$w = \big(w(1), \ldots, w(T)\big)$ satisfies the set of difference equations

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

with maximum l lags (lag=delay) if and only if

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

Kernel representation

more compact notation: $R(\sigma)w = 0$, where $R(\xi) := \sum_{i=0}^{l} R_i \xi^i$

where σ is the shift operator, $(\sigma w)(t) = w(t+1)$

$\mathcal{B}$ — the set of all trajectories of the system Σ described by $R(\sigma)w = 0$

$$\mathcal{B} := \{\, w : \mathbb{Z}_+ \to \mathbb{R}^{\mathtt{w}} \mid R(\sigma)w = 0 \,\}$$

(no a priori separation of the variables into inputs and outputs)

$\mathcal{B} = \ker\big(R(\sigma)\big)$, so that $R(\sigma)w = 0$ is a kernel representation of $\mathcal{B}$

we will associate $\mathcal{B}$ with the system Σ itself

Shortest lag representation

facts about kernel representations:

- non-uniqueness

  $\ker\big(R(\sigma)\big) = \ker\big(U(\sigma)R(\sigma)\big)$, where $U \in \mathbb{R}^{p\times p}[\xi]$ is unimodular ($\det U$ is a nonzero constant)

- shortest lag representation

  let l_i be the lag of the ith equation of $R(\sigma)w = 0$

  $\exists$ a kernel repr., s.t. $\operatorname{row\,dim}(R)$, $\max l_i$, and $\sum l_i$ are minimal

  $R(\sigma)w = 0$ is shortest lag $\iff$ $R(\xi)$ is row proper
  (the leading row coefficient matrix is full row rank)

Invariants of $\mathcal{B}$

consider a minimal kernel representation $R(\sigma)w = 0$ of $\mathcal{B}$

the following are invariants of the behavior $\mathcal{B}$

$$\begin{array}{lllllll}
p & := & \mathbf{p}(\mathcal{B}) & := & \operatorname{row\,dim}(R) & — & \text{output cardinality} \\
m & := & \mathbf{m}(\mathcal{B}) & := & \mathtt{w} - p & — & \text{input cardinality} \\
n & := & \mathbf{n}(\mathcal{B}) & := & \sum_{i=1}^{p} l_i & — & \text{total lag} \quad (\text{or order of } \Sigma) \\
l & := & \mathbf{l}(\mathcal{B}) & := & \max_{i=1,\dots,p} l_i & — & \text{maximal lag (or lag of } \Sigma)
\end{array}$$

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, $\dim(\mathcal{B}|_{[1,T]}) = mT + n \le mT + lp$

complexity of $\mathcal{L}_{m,l} \simeq \dim(\mathcal{B}|_{[1,T]})$

the specification of the complexity by the lag does not fix the order

$\mathcal{B} \in \mathcal{L}_{m,l} \implies$ the order n of $\mathcal{B}$ is in the range: $(l-1)p < n \le lp$

for the identification problem we will use the model class $\mathcal{M} = \mathcal{L}_{m,l}$

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

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

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

our goal is to express the identification problem as an STLS problem

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

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

Lemma 1

given: $w = \big(w(1),\dots,w(T)\big)$, $w(t)\in\mathbb{R}^{\mathtt{w}}$, $m\le\mathtt{w}$, and $l\le T-1$

assume: $R\mathcal{H}_{l+1}(w)=0$, for certain $R =: \begin{bmatrix} R_0 & R_1 & \cdots & R_l \end{bmatrix}$, $R_i\in\mathbb{R}^{p\times\mathtt{w}}$

where $p := \mathtt{w}-m$, with R_l full row rank

then the system $\mathcal{B} := \ker\left(\sum_{i=0}^{l} R_i\sigma^i\right)$ satisfies:

1. $\mathcal{B}\in\mathcal{L}_{m,l}$
2. $\mathbf{n}(\mathcal{B}) = pl$
3. $w\in\mathcal{B}|_{[1,T]}$

Proof of Lemma 1

by definition $\mathcal{B}$ is linear with $\mathbf{l}(\mathcal{B})\le l$

$$\begin{aligned} R_l \text{ full row rank} \quad &\Longrightarrow \quad R(\xi) \text{ row proper} \\ &\Longrightarrow \quad \mathbf{p}(\mathcal{B}) = \text{row dim}(R) = p \\ &\Longrightarrow \quad \mathbf{m}(\mathcal{B}) = \mathtt{w} - \mathbf{p}(\mathcal{B}) = m \\ &\Longrightarrow \quad \mathcal{B}\in\mathcal{L}_{m,l} \end{aligned}$$

let l_i be the degree of the ith equation in $R(\sigma)w=0$

R_l full row rank $\implies$ $l_i = l$ for all i $\implies$ $\mathbf{n}(\mathcal{B}) = \sum_{i=1}^{p} l_i = pl$

$$\begin{aligned} R\mathcal{H}_{l+1}(w) = 0 \quad &\Longrightarrow \quad \textstyle\sum_{\tau=0}^{l} R_\tau w(t+\tau), \text{ for } t=1,\dots,T-l \\ &\Longrightarrow \quad w\in\mathcal{B}|_{[1,T]} \end{aligned}$$

Lemma 2

(reverse implication to the one of Lemma 1)

given: $w = \big(w(1),\dots,w(T)\big)$, $w(t)\in\mathbb{R}^{\mathtt{w}}$, $m\le\mathtt{w}$, and $l\le T-1$

assume: there is $\mathcal{B}\in\mathcal{L}_{m,l}$, $\mathbf{n}(\mathcal{B}) = pl$, s.t. $w\in\mathcal{B}|_{[1,T]}$

let $R(\sigma)w=0$, $R(\xi)=\sum_{i=0}^{l} R_i\xi^i$, be a shortest lag repr. of $\mathcal{B}$

then

1. R_l is full row rank
2. $R\mathcal{H}_{l+1}(w)=0$, where $R := \begin{bmatrix} R_0 & R_1 & \cdots & R_l \end{bmatrix}$

Proof of Lemma 2

let l_i be the degree of the ith equation in $R(\sigma)w = 0$

$l_i \leq l$ and $\mathbf{n}(\mathcal{B}) = \sum_{i=1}^{p} l_i = pl \implies l_i = l$, for all i

$R(\sigma)w = 0$ shortest lag $\implies R(\xi)$ row proper
$\implies$ the leading coef. matrix L of $R(\xi)$ is full rank

but $l_i = l$, for all $i \implies L = R_l \implies R_l$ is full rank

$w \in \mathcal{B}|_{[1,T]} \implies R\mathcal{H}_{l+1}(w) = 0$

Theorem 1

assume: $\mathcal{B} \in \mathcal{L}_{m,l}$ admits a kernel representation $R(\sigma)w = 0$ with

$$R(\xi) = \sum_{i=0}^{l} R_i \xi^i, \qquad R_l =: \begin{bmatrix} Q_l & P_l \end{bmatrix}, \quad \text{and} \quad P_l \in \mathbb{R}^{p \times p} \text{ full rank}$$

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

then

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

Proof of Theorem 1

assume $w \in \mathcal{B}|_{[1,T]}$

by Lemma 2, $R\mathcal{H}_{l+1}(w) = 0$

since P_l is full rank, $R\mathcal{H}_{l+1}(w) = 0 \implies \mathcal{H}_{l+1}^\top(w) \begin{bmatrix} X \\ -I \end{bmatrix} = 0$

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

by Lemma 1, $\mathcal{B} = \ker\left(\sum_{i=0}^{l} R_i \sigma^i\right)$ with

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_l \end{bmatrix} := \begin{bmatrix} X^\top & -I \end{bmatrix}$$

$\implies P_l = -I$ is full rank

Solution by structured total least squares

theorem 1 states the desired equivalence under the assumption that the optimal solution $\hat{\mathcal{B}}$ admits a kernel representation

$$\hat{\mathcal{B}} = \ker\left(\sum_{i=0}^{l} \hat{R}_i \sigma^i\right), \quad \hat{R}_l := \begin{bmatrix} \hat{Q}_l & \hat{P}_l \end{bmatrix}, \quad \hat{P}_l \in \mathbb{R}^{p \times p} \text{ full rank} \qquad (*)$$

conjecture:

$$\Omega := \left\{ w \in (\mathbb{R}^{\mathtt{w}})^T \;\middle|\; \begin{array}{l} \text{there is a unique global minimizer } \hat{\mathcal{B}} \\ \text{that satisfies } (*), \textit{ i.e.}, \text{ GTLS} \equiv \text{STLS} \end{array} \right\}$$

is generic in $(\mathbb{R}^{\mathtt{w}})^T$, *i.e.*, it contains an open and dense subset, which complement has measure zero

Motivation for the conjecture

the existence and uniqueness part of the conjecture is motivated in the Ph.D. thesis of B. Roorda

motivation for $(*)$ being generic:

the highest possible order of a system in the model class $\mathcal{L}_{m,l}$ is pl

one can expect that $\mathbf{n}(\hat{\mathcal{B}}) = pl$, generically in the data space $(\mathbb{R}^{\mathtt{w}})^T$

by Lemma 2, $\mathbf{n}(\hat{\mathcal{B}}) = pl \implies \hat{R}_l$ is full rank

generically in $\mathbb{R}^{p\times \mathtt{w}}$, $\hat{P}_l \in \mathbb{R}^{p\times p}$, defined by $\hat{R}_l =: \begin{bmatrix} \hat{Q}_l & \hat{P}_l \end{bmatrix}$, is full rank

Extensions of the identification problem

given input/output partitioning: $\quad w = \begin{bmatrix} u \\ y \end{bmatrix} \quad$ (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 K time series $w^{(1)}, \dots, w^{(K)}$ of the same length T

define a matrix valued time series $\quad w(t) = \begin{bmatrix} w^{(1)}(t) & \cdots & w^{(K)}(t) \end{bmatrix}$

$$w \in \mathcal{B}|_{[1,T]} \quad :\iff \quad w^{(1)} \in \mathcal{B}|_{[1,T]}, \dots, w^{(K)} \in \mathcal{B}|_{[1,T]}$$

find $\mathcal{B} \in \mathcal{L}_{m,l}$ that approximates simultaneously $w^{(1)}, \dots, w^{(K)}$

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

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

Extensions of the identification problem

latent inputs: assume that there is an unobserved input e, $e(t) \in \mathbb{R}^{n_e}$

it is natural to modify the misfit identification problem as follows:

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

the misfit problem with $w_{\text{aug}} := \begin{bmatrix} 0 \\ w \end{bmatrix}$ and $\mathcal{M} = \mathcal{L}_{m+n_e,l}$

is equivalent to the misfit–latency problem $(*)$

pure latency identification problems can be solved by considering w exact

Software package for solving STLS problems

the inner minimization in the STLS problem is solved analytically

this gives a nonlinear least squares prob. $\min_X r^\top(X)\Gamma^{-1}(X)r(X)$

we use MINPACK's Levenberg–Marquardt algorithm

the cost function and Jacobian are evaluated efficiently (in $O(T)$ flops) by exploiting structure of the weight matrix Γ (block-Toeplitz banded)

we use numerical algorithms for structured matrices (from SLICOT)

the package is written in ANSI C and has MATLAB interface

available from:

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

Results on data sets 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.

the estimates obtained by the following methods are compared:

- `subid` — robust combined subspace algorithm
- `detss` — deterministic balanced subspace algorithm
- `pem` — the prediction error method of the Identification Toolbox
- `stls` — the proposed method based on STLS

the order specified for the methods `subid`, `detss`, and `pem` is $n = pl$

$\hat{\mathcal{B}}$ for `detss` and `pem` is the deterministic part of the identified system

comparison is in the misfit $M(w, \hat{\mathcal{B}})$ scaled by $M(w, \hat{\mathcal{B}}_{\texttt{stls}})$

#	Data set name	parameters T	m	p	l	scaled misfit `subid`	`detss`	`pem`
1	Destillation column	90	5	3	1	2.8	9.6	15.9
2	Destillation column n10	90	5	3	1	2.8	9.6	15.9
3	Destillation column n20	90	5	3	1	8.3	2.3	36.1
4	Destillation column n30	90	5	3	1	7.8	3.3	132.2
5	Glass furnace (Philips)	1247	3	6	1	2.9	2.5	2.7
6	120 MW power plant	200	5	3	2	7.2	3.4	28.5
7	pH process	2001	2	1	6	1.3	1.3	3.0
8	Hair dryer	1000	1	1	5	1.2	1.2	1.0
9	Winding process	2500	5	2	2	1.5	1.4	2.8
10	Ball-and-beam setup	1000	1	1	2	1.0	10.6	1.0
11	Industrial dryer	867	3	3	1	1.2	1.1	1.1

#	Data set name	parameters T	m	p	l	scaled misfit `subid`	`detss`	`pem`
12	CD-player arm	2048	2	2	1	1.2	1.1	1.4
13	Wing flutter	1024	1	1	5	1.6	1.7	2.8
14	Robot arm	1024	1	1	4	2.7	18.7	26.0
15	Lake Erie	57	5	2	1	1.5	2.3	23.1
16	Lake Erie n10	57	5	2	1	2.1	2.2	8.4
17	Lake Erie n20	57	5	2	1	2.2	2.4	9.8
18	Lake Erie n30	57	5	2	1	2.4	1.6	5.6
19	Heat flow density	1680	2	1	2	1.8	1.3	9.8
20	Heating system	801	1	1	2	1.3	1.2	1.3
21	Steam heat exchanger	4000	1	1	2	1.8	1.8	8.1
22	Industrial evaporator	6305	3	3	1	1.5	1.1	1.6
23	Tank reactor	7500	1	2	1	2.3	2.1	52.9
24	Steam generator	9600	4	4	1	2.4	3.1	3.3

comparison in the execution time scaled by $M(w, \hat{\mathcal{B}}_{\texttt{subid}})$

#	Data set name	parameters T	m	p	l	scaled exec. time detss	stls	pem
1	Destillation column	90	5	3	1	3.3	6.4	11.1
2	Destillation column n10	90	5	3	1	7.3	12.5	23.1
3	Destillation column n20	90	5	3	1	7.2	12.8	7.2
4	Destillation column n30	90	5	3	1	7.0	12.1	7.2
5	Glass furnace (Philips)	1247	3	6	1	13.5	361.2	373.3
6	120 MW power plant	200	5	3	2	6.3	15.5	27.3
7	pH process	2001	2	1	6	2.9	7.4	32.3
8	Hair dryer	1000	1	1	5	1.5	5.8	36.4
9	Winding process	2500	5	2	2	4.4	37.1	74.8
10	Ball-and-beam setup	1000	1	1	2	1.9	4.1	7.2
11	Industrial dryer	867	3	3	1	6.6	25.5	27.3

#	Data set name	parameters T	m	p	l	scaled exec. time subid	stls	pem
12	CD-player arm	2048	2	2	1	6.4	19.5	49.4
13	Wing flutter	1024	1	1	5	1.7	4.7	33.5
14	Robot arm	1024	1	1	4	1.8	3.8	30.7
15	Lake Erie	57	5	2	1	1.4	4.6	7.0
16	Lake Erie n10	57	5	2	1	1.4	4.6	11.4
17	Lake Erie n20	57	5	2	1	1.6	4.8	9.1
18	Lake Erie n30	57	5	2	1	1.7	4.8	7.0
19	Heat flow density	1680	2	1	2	2.6	6.3	39.7
20	Heating system	801	1	1	2	1.7	3.7	12.4
21	Steam heat exchanger	4000	1	1	2	4.3	8.4	31.1
22	Industrial evaporator	6305	3	3	1	10.5	59.9	134.4
23	Tank reactor	7500	1	2	1	11.0	25.2	146.0
24	Steam generator	9600	4	4	1	13.6	192.0	220.1

Conclusions

- we generalized previous results on the application of STLS for system identification to multivariable systems
- the STLS method allows to treat behavioral and EIV ident. problems
- multiple time series
latent variables
exact variables
can be taken into account
- also model reduction, noisy realization, and autonomous systems identification problems can be solved by STLS
- there is an efficient software tool that covers all these problems

Acknowledgments

Jan C. Willems for very helpful discussions on the identification problem and the parameterization issue for multivariable systems

Diana Sima for carefully reading all drafts of the manuscript and for many valuable suggestions

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. De Moor and B. Roorda. L_2-optimal linear system identification structured total least squares for SISO systems. In the proc. of *CDC 1994*, pages 2874–2879.

- B. Roorda and C. Heij. Global total least squares modeling of multivariate time series. *IEEE Trans. on Aut. Control*, 40(1):50–63, 1995.

References

- B. Roorda. Algorithms for global total least squares modelling of finite multivariable time series. *Automatica*, 31(3):391–404, 1995.

- P. Lemmerling and B. De Moor. Misfit versus latency. *Automatica*, 37:2057–2067, 2001.