Structured Total Least Squares

Ivan Markovsky, Sabine Van Huffel, and Bart De Moor

Katholieke Universiteit Leuven, ESAT/SCD

Overview

Least Squares	Gauss	1820
Total Least Squares	Golub and Van Loan	1980
Structured Total Least Squares	Abatzoglou, De Moor	1990

Least Squares: $\min_{\Delta B, X} \|[\Delta B]\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad AX = B - \Delta B$

Structured Total Least Squares:

$$\min_{\Delta A, \Delta B, X} \|[\Delta A\ \Delta B]\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad (A - \Delta A)X = B - \Delta B \quad \text{and}$$

$[\Delta A\ \Delta B]$ has the same structure as $[A\ B]$

Approaches:

CTLS	constraint TLS		(Abatzoglou *et al.*, 1991)
RiSVD	Riemannian SVD		(De Moor, 1993)
STLN	structured total least norm		(Rosen *et al.*, 1996)

Types of structure:

nonlinear	(Rosen *et al.*, 1998; Lemmerling *et al.*, 2002)
affine	CTLS, RiSVD, STLN
Toeplitz/Hankel	all methods

Rank reduction $d = \operatorname{row}\dim(X)$:

$d \geq 1$	multivariate problems	(Van Huffel *et al.*, 1996)
$d = 1$	univariate problems	all methods

Efficiency of the algorithms:

affine structure	$O(m^3)$	CTLS, RiSVD, STLN
Toeplitz/Hankel	$O(m^2)$	STLN
Toeplitz/Hankel	$O(m)$	(Mastronardi *et al.*, 2000; Lemmerling *et al.*, 2000)

Applications:

- direction of arrival (DOA)
- nuclear magnetic resonance (NMR)
- image deblurring
- system identification (De Moor and Roorda, 1994)

No software available

New results

Main features:

approach	CTLS-like
structure	block-Toeplitz/Hankel, unstructured, exact
rank reduction	$d \geq 1$, multivariate
efficiency	$O(m)$
software	C-code with LAPACK/SLICOT calls (Markovsky *et al.*, 2003)

STLS problem: $\min_{\Delta p, X} \|\Delta p\|^2 \quad \text{s.t.} \quad \mathcal{S}(p - \Delta p) \begin{bmatrix} X \\ -I_d \end{bmatrix} = 0$

$\mathcal{S}$—structure specification, *e.g.*, affine $\mathcal{S}(p) = S_0 + \sum_{i=1}^{n_p} S_i p_i$

Define $X_{\mathrm{ext}} := \left[\begin{smallmatrix} X \\ -I \end{smallmatrix}\right]$, $r := \operatorname{vec}\big(([A\ B]\, X_{\mathrm{ext}})^\top\big)$, and $G := \big[\operatorname{vec}\big((S_1 X_{\mathrm{ext}})^\top\big) \cdots \operatorname{vec}\big((S_{n_p} X_{\mathrm{ext}})^\top\big)\big]$.

Equivalent problem:

$$\min_X r(X)\Gamma^{-1}(X) r(X), \qquad \text{where} \quad \Gamma(X) := G(X) G^\top(X) \qquad (*)$$

Theorem: Under the assumption that

$$C = \begin{bmatrix} C^{(1)} & \cdots & C^{(q)} \end{bmatrix}, \quad \text{where } C^{(i)} \text{ is } \begin{cases} \text{block-Hankel/Toeplitz,} \\ \text{unstructured, or} \\ \text{exact} \end{cases}$$

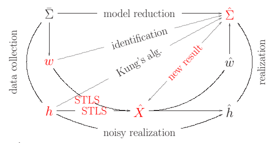
Γ is block-Toeplitz and block-banded with block size dK, where K is the row dimension of a block in block-Hankel/Toeplitz structured block $C^{(i)}$, and with half block bandwidth, the maximum number of block columns in a block-Hankel/Toeplitz structured block $C^{(i)}$.

the structure of Γ allows computation of $f_0(X)$ and $f_0'(X)$ in $O(m)$ flops $\implies$ fast algorithms for $(*)$

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

Application for MIMO system identification

Approximate modeling problems



$\hat{\Sigma}$ — (low complexity) approximating model
$\bar{\Sigma}$ — (high complexity) "true" model
w — observed general response
h — observed impulse response

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

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

$\mathcal{L}_{m,l}$ — the set of all LTI systems with m inputs and lag at most l (m and l specify the complexity of the model class $\mathcal{L}_{m,l}$)

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

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

Theorem: Assume that $\mathcal{B} \in \mathcal{L}_{m,l}$ admits a kernel representation

$$\mathcal{B} = \ker \left(\sum_{i=0}^{l} R_i \sigma^i \right), \qquad \text{with} \quad R_l =: [Q_l\ P_l], \qquad P_l \in \mathbb{R}^{p \times p} \text{ full rank}$$

and let

$$X^\top := -P_l^{-1} [R_0 \cdots R_{l-1}\ Q_l].$$

Then

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

The extensions of the identification problem:
- exact variables
- multiple time series
- latent inputs

and the special identification problems:
- model reduction
- noisy realization
- autonomous systems identification

are also solved as STLS problems.

Results on data sets from DAISY

DAISY (DAta base for Identification of SYstems), (De Moor, 1998)

Compared methods:

- `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 of the misfit $M(w, \hat{\mathcal{B}})$ scaled by $M(w, \hat{\mathcal{B}}_{\mathtt{stls}})$.

		parameters				scaled misfit		
#	Data set name	T	m	p	l	`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
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 of the execution time scaled by $M(w, \hat{\mathcal{B}}_{\mathtt{subid}})$.

		parameters				scaled exec. time		
#	Data set name	T	m	p	l	`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
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

References

Abatzoglou, T., J. Mendel and G. Harada (1991). The constrained total least squares technique and its application to harmonic superresolution. *IEEE Trans. on Signal Proc.* **39**, 1070–1087.

De Moor, B. (1993). Structured total least squares and L_2 approximation problems. *Lin. Alg. and Its Appl.* **188–189**, 163–207.

De Moor, B. (1998). Daisy: Database for the identification of systems. Department of Electrical Engineering, ESAT/SISTA, K.U.Leuven, Belgium, URL: `http://www.esat.kuleuven.ac.be/sista/daisy/`.

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

Lemmerling, P., N. Mastronardi and S. Van Huffel (2000). Fast algorithm for solving the Hankel/Toeplitz structured total least squares problem. *Numerical Algorithms* **23**, 371–392.

Lemmerling, P., S. Van Huffel and B. De Moor (2002). The structured total least squares approach for nonlinearly structured matrices. *Numerical Linear Algebra with Applications* **9**(1–4), 321–332.

Markovsky, I., S. Van Huffel and R. Pintelon (2003). Software for structured total least squares estimation: User's guide. Technical Report 03–136. Dept. EE, K.U. Leuven.

Mastronardi, N., P. Lemmerling and S. Van Huffel (2000). Fast structured total least squares algorithm for solving the basic deconvolution problem. *SIAM J. Matrix Anal.* **22**, 533–553.

Rosen, J. B., H. Park and J. Glick (1996). Total least norm formulation and solution of structured problems. *SIAM J. Matrix Anal.* **17**, 110–128.

Rosen, J. B., H. Park and J. Glick (1998). Structured total least norm for nonlinear problems. *SIAM J. Matrix Anal.* **20**(1), 14–30.

Van Huffel, S., H. Park and J. B. Rosen (1996). Formulation and solution of structured total least norm problems for parameter estimation. *IEEE Trans. on Signal Proc.* **44**(10), 2464–2474.