Weighted structured total least squares

Ivan Markovsky and Sabine Van Huffel

K.U.Leuven, ESAT-SISTA

5th Int. Conf. on "Large-Scale Scientific Computations"
Sozopol, Bulgaria, June 6–10, 2005

Outline

Total least squares

Consider an over-determined linear system of equations

$$AX \approx B, \qquad \text{where} \quad A \in \mathbb{R}^{m\times n} \quad \text{and} \quad B \in \mathbb{R}^{m\times d}.$$

How to find an approximate solution? Two popular methods are:

Least squares method: correct B

$$\min_{\Delta B, X} \|\Delta B\|_{\mathrm{F}}^2 \quad \text{subject to} \quad AX = B - \Delta B \qquad \text{(LS)}$$

Total least squares method: correct both A and B

$$\min_{\Delta A, \Delta B, X} \|[\Delta A \quad \Delta B]\|_{\mathrm{F}}^2 \quad \text{subject to} \quad (A - \Delta A)X = B - \Delta B \qquad \text{(TLS)}$$

Weighted total least squares

- the TLS cost function $\|[\Delta A \quad \Delta B]\|_{\mathrm{F}}^2$ measures the correction size
- the Frobenius norm puts equal emphasis on all elements
- define $\|\Delta C\|_W^2 := \operatorname{vec}^\top(\Delta C^\top) W \operatorname{vec}(\Delta C^\top)$, where $W > 0$ is given

Weighted total least squares method

$$\min_{\Delta A, \Delta B, X} \|[\Delta A \quad \Delta B]\|_W^2 \quad \text{subject to} \quad (A - \Delta A)X = B - \Delta B \qquad \text{(WTLS)}$$

Structured total least squares

- the data matrix $C := \begin{bmatrix} A & B \end{bmatrix}$ was assumed unstructured
- in problems involving dynamic phenomena C is structured
- next we define a TLS problem with structure constraint

The structure is specified by an injective function $\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m\times(n+d)}$.

C is $\mathscr{S}$-structured if $C \in \operatorname{image}(\mathscr{S})$, *i.e.*, $C = \mathscr{S}(p)$, for some $p \in \mathbb{R}^{n_p}$.

Structured total least squares method

$$\min_{\Delta p, X} \|\Delta p\|^2 \quad \text{subject to} \quad \mathscr{S}(p - \Delta p) \begin{bmatrix} X \\ -I_d \end{bmatrix} = 0 \qquad \text{(STLS)}$$

The topic of this talk: structured weighted TLS

Structured weighted total least squares method

Given a data matrix $C \in \mathbb{R}^{m\times(n+d)}$, $m > n$, structure specification $\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m\times(n+d)}$, and weight matrix $W \in \mathbb{R}^{n_p \times n_p}$, $W > 0$, find

$$\hat{X} = \min_{\Delta p, X} \Delta p^\top W \Delta p \quad \text{subject to} \quad \mathscr{S}(p) \begin{bmatrix} X \\ -I \end{bmatrix} = 0 \qquad \text{(SWTLS}$$

- properties
- solution methods
- software

Affine structured TLS

Let $\mathscr{S}$ be affine, *i.e.*, $\mathscr{S}(p) = S_0 + \sum_{i=1}^{n_p} S_i p_i$, for some $S_i \in \mathbb{R}^{m\times(n+d)}$.

Then the constraint $\mathscr{S}(p - \Delta p) X_{\text{ext}} = 0$ is bilinear in X and Δp.

$$\mathscr{S}(p - \Delta p) X_{\text{ext}} = 0 \quad \Longleftrightarrow \quad G(X)\Delta p = r(X) \ ,$$

where

$$G(X) := \begin{bmatrix} \operatorname{vec}\left((S_1 X_{\text{ext}})^\top\right) & \cdots & \operatorname{vec}\left((S_{n_p} X_{\text{ext}})^\top\right) \end{bmatrix} \in \mathbb{R}^{md\times n_p} \ ,$$

and

$$r(X) := \operatorname{vec}\left(\left(\mathscr{S}(p) X_{\text{ext}}\right)^\top\right) \in \mathbb{R}^{md} \ .$$

The affine STLS problem is a double minimization problem

$$\min_X \left(\min_{\Delta p} \|\Delta p\|_2^2 \quad \text{subject to} \quad G(X)\Delta p = r(X) \right) \ .$$

Equivalent unconstrained problem

The inner minimization has an analytic solution.

Theorem (Equivalent optimization problem for affine STLS)

Assuming that $n_p \geq md$, an affine STLS problem is equivalent to

$$\min_X f(X) \quad \textit{where} \quad f(X) := r^\top(X) \Gamma^\dagger(X) r(X)$$

$$\textit{and} \quad \Gamma(X) := G(X) G^\top(X) \ .$$

- the constraint and the decision variable Δp are eliminated
- nonlinear least squares problem
- use classical optimization methods

Properties

Structure assumption

$\mathscr{S}(p) = \begin{bmatrix} C^1 & \cdots & C^q \end{bmatrix}$, where C^l, for $l = 1,\dots,q$, is block-Toeplitz, block-Hankel, unstructured, or exact.

Theorem (Structure of the weight matrix Γ)

If the structure assumption holds, the weight matrix $\Gamma(X)$ is

$$\Gamma(X) = \begin{bmatrix} \Gamma_0 & \Gamma_1^\top & \cdots & \Gamma_s^\top & & \mathbf{0} \\ \Gamma_1 & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \Gamma_s^\top \\ \Gamma_s & \ddots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \Gamma_1^\top \\ \mathbf{0} & & \Gamma_s & \cdots & \Gamma_1 & \Gamma_0 \end{bmatrix} \in \mathbb{R}^{md\times md} .$$

Efficient cost function and first derivative evaluation

From X to $f(X)$ and $f'(X)$

- cost function evaluation:

$$f(X) = r^\top(X)y(X), \qquad \text{where} \quad \Gamma(X)y(X) = r(X)$$

  the structure of Γ is exploited in solving for $y \rightsquigarrow O(m)$ flops
- $f'(X)$ can also be evaluated from $y(X)$ in $O(m)$ flops

From $\mathscr{S}$ to Γ

$$\Gamma_k(X) = (I_K \otimes X_{\text{ext}}^\top) S_k (I_K \otimes X_{\text{ext}}^\top)^\top \ , \quad k = 0,1,\dots,s \ ,$$

where the matrices $S_k \in \mathbb{R}^{K(n+d)\times K(n+d)}$ depend on the structure $\mathscr{S}$

Equivalent unconstrained problem for (SWTLS)

The affine structured weighted TLS problem

$$\min_{\Delta p, X} \Delta p^\top W \Delta p \quad \text{subject to} \quad \mathscr{S}(p) \begin{bmatrix} X \\ -I \end{bmatrix} = 0 \qquad \text{(SWTLS)}$$

again can be solved partially by eliminating Δp.

Theorem (Equivalent optimization problem for weighted STLS)

The affine weighted STLS problem (SWTLS) is equivalent to

$$\min_X f_{\mathrm{w}}(X) \quad \textit{where} \quad f_{\mathrm{w}}(X) := r^\top(X)\Gamma_{\mathrm{w}}^\dagger(X) r(X)$$

$$\textit{and} \quad \Gamma_{\mathrm{w}}(X) := G(X)W^{-1}G^\top(X) \ .$$

Properties of the weight matrix Γ_{w}

In general neither the block-Toeplitz nor the block-banded properties o $\Gamma = GG^\top$ are present in $\Gamma_{\mathrm{w}} = GW^{-1}G^\top$.

Assumption: block-diagonal weight matrix W

Let the blocks C^l be parameterized by $p_l \in \mathbb{R}^{n_{p,l}}$, and let $p = \operatorname{col}(p_1,\dots,p_q)$. In addition, W is assumed to be block-diagonal

$$W = \operatorname{diag}(W^1,\dots,W^q), \qquad \text{where} \quad W^l \in \mathbb{R}^{n_{p,l}\times n_{p,l}} \ .$$

The assumption forbids cross-weighting among the parameters of the blocks $C^1,\dots,C^q$.

Main result

Define $V := W^{-1}$ and $V^l := (W^l)^{-1}$.

Lemma

Under the structure and weight matrix assumptions, if all blocks V^l are Toeplitz, then $\Gamma_{\mathrm{w}} = GVG^\top$ is Toeplitz.

Lemma

Under the structure and weight matrix assumptions, if W is banded with bandwidth p, then $\Gamma_{\mathrm{w}} = GVG^\top$ is banded with bandwidth $s+p$, where s is the bandwidth in the unstructured case.

Implication of the main result

Under the structure and weight matrix assumptions, if W is banded the
cost function and first derivative can still be evaluated in $O(m)$ flops.

Therefore this type of problems can still be solved efficiently.

The only difference with the unweighted case is that now

$$\Gamma_{ij}(X) = (I_K \otimes X_{\text{ext}}^\top) S_{ij} (I_K \otimes X_{\text{ext}}^\top)^\top \ ,$$

where

$$S_{ij} := \begin{cases} \operatorname{diag}(V_i^1, \dots, V_i^q) S_{i-j} & \text{if } 0 \le i-j \le s \\ S_{ji}^\top & \text{if } -s \le i-j < 0 \\ 0 & \text{otherwise} \end{cases}$$

Conclusions

- main question: properties of Γ in the equivalent problem
- unstructured problems: Γ block-Toeplitz and banded
- structured problems with W block-diagonal: Γ banded
- this implies $O(m)$ computational complexity of the algorithm
- software is available at
  `http://www.esat.kuleuven.be/~imarkovs`