DYSCO course on low-rank approximation and its applications

Generalized total least squares

Ivan Markovsky

Vrije Universiteit Brussel

Outline

Statistical properties of TLS

- errors-in-variables (EIV) model

$$A = \overline{A} + \widetilde{A} \quad \text{and} \quad y = \overline{y} + \widetilde{y}$$

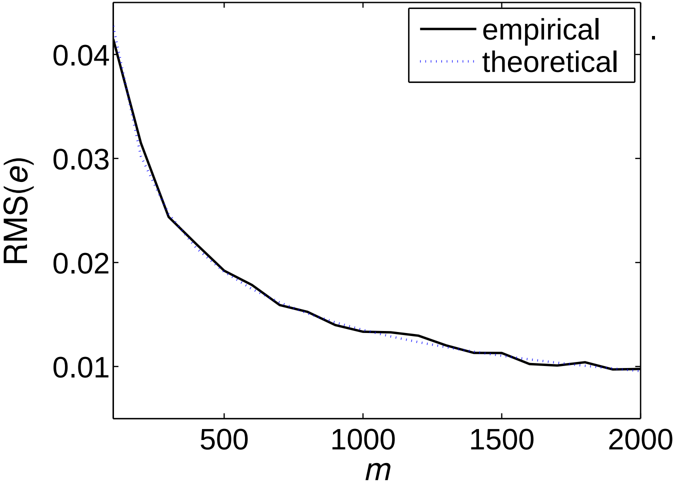
- true values $\overline{A}$, $\overline{y}$ satisfy $\overline{A}\overline{x} = \overline{y}$, for some $\overline{x} \in \mathbb{R}^n$

- perturbations $\widetilde{A}$, $\widetilde{y}$ are zero mean element-wise i.i.d.

- under additional mild assumptions the TLS approx. solution $\widehat{x}$ is a consistent estimator of the true value $\overline{x}$

- measurement errors model
  - A, y — measured data
  - $\overline{x}$ / $\widehat{x}$ — true/estimated model parameters

Estimation error $e = \overline{x} - \widehat{x}$



empirical — solid line, theoretical — dotted line

Notes

- TLS problem vs EIV model
  - TLS approx. can be used without EIV model
  - EIV model shows the correct testbed TLS approx.
- distinguish
  - corrections ΔA, Δy in the TLS problem, and
  - noise/perturbations $\widetilde{A}$, $\widetilde{y}$ in the EIV model

Confidence bounds

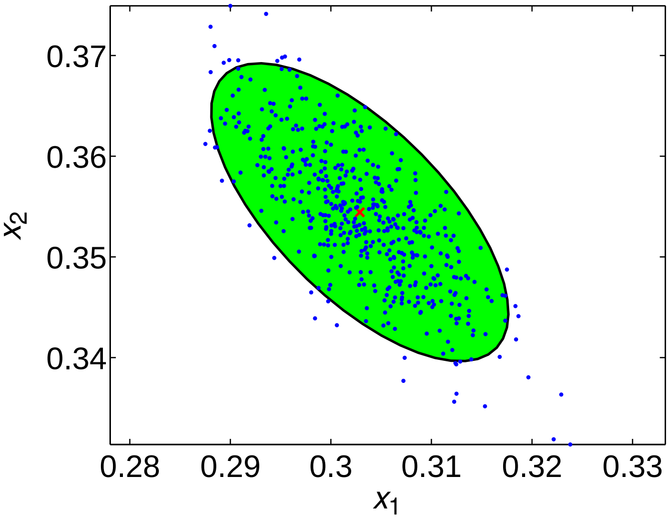
- assume that $\widetilde{A}$, $\widetilde{y}$ are i.i.d. normal with variance ξ^2
- the estimation error is e asymptorically normal
  $\rightsquigarrow$ confidence bounds for $\widehat{x}$
- the asymptotic error $e := \overline{x} - \widehat{x}$ covariance matrix is

$$V_e = \xi^2 (1 + \widehat{x}\widehat{x}^\top)(A^\top A - m\sigma^2 I)^{-1}$$

- the noise variance ξ^2 can be estimated from the data

$$\widehat{\xi} = \frac{1}{m}\sigma_n^2$$

95% confidence ellipsoid



Outline

Weighted total least squares problem

- replace the Frobenius norm by the weighted 2-norm

$$\|D\|_W := \sqrt{\mathrm{vec}^\top(D) W \mathrm{vec}(D)}$$

- $W =$ inverse noise ($\mathrm{vec}([\widetilde{A} \;\; \widetilde{y}])$) covariance matrix
- WTLS doesn't have analytic solution
- special cases $\rightsquigarrow$ structure of W
  - column/row-wise weighting
  - element-wise weighting
  - generalized TLS
  - restricted TLS

Hierarchy of WTLS problems

1. fully weighted $W \geq 0$
2. column-wise weighted

$$W = \text{diag}(W_1, \ldots, W_m), \quad W_i \in \mathbb{R}_+^{(n+1)\times(n+1)}$$

3. element-wise weighted

$$W = \text{diag}(w), \quad w \in \mathbb{R}_+^{m(n+1)}$$

4. column-wise GTLS: case 2, with W_i's equal
5. column-wise scaled: case 3, with W_i — diagonal

Relative error TLS

- consider the element-wise weighted case

$$\|D\|_w = \|D\|_\Sigma := \|\Sigma \odot D\|_{\mathrm{F}}$$

($\odot$ — element-wise product)

- $\Sigma_{ij} = 1/d_{ij} \rightsquigarrow$ approximation in relative error sense

$$e_{ij} = \frac{d_{ij} - \widehat{d}_{ij}}{d_{ij}}$$

GTLS problem

- TLS approximation with criterion

$$\|D\|_{\Sigma_l,\Sigma_r} := \|\Sigma_l D \Sigma_r\|_F$$

- link to WTLS

$$\begin{aligned}
\|\Sigma_l(D-\widehat{D})\Sigma_r\|_F^2 &= \left\|\operatorname{vec}(\Sigma_l(D-\widehat{D})\Sigma_r)\right\|^2 \\
&= \left\|(\Sigma_r \otimes \Sigma_l)\operatorname{vec}(D-\widehat{D})\right\|^2 \\
&= \operatorname{vec}^\top(D-\widehat{D})\big(W_r \otimes W_l\big)\operatorname{vec}(D-\widehat{D})
\end{aligned}$$

where $\sqrt{W_r} = \Sigma_r$ and $\sqrt{W_l} = \Sigma_l$

- WTLS problem with weight matrix $W = W_r \otimes W_l$

Element-wise GTLS

- element-wise weighted total least squares

$$\|D\|_w = \|D\|_\Sigma := \|\Sigma \odot D\|_\mathsf{F}$$

- element-wise generalized total least squares

$$W_\mathrm{r} = \mathrm{diag}(w_\mathrm{r}) \quad \text{and} \quad W_\mathrm{l} = \mathrm{diag}(w_\mathrm{l})$$

- $\rightsquigarrow$ rank-1 matrix $\Sigma = w_\mathrm{l} w_\mathrm{r}^\top$

GTLS solution

- $\sqrt{W_\mathrm{r}} = \Sigma_\mathrm{r}$, w.l.o.g. we can choose Σ_r upper triangular, *e.g.*, the Cholesky factor of W_r

- modified data matrix: $D_\mathrm{m} := \Sigma_\mathrm{l} D \Sigma_\mathrm{r}$

- TLS approximation of D_m: $\widehat{D}_{\mathrm{m,tls}}$ and $\widehat{x}_{\mathrm{m,tls}}$

- partition Σ_r as $\begin{bmatrix} \Sigma_{\mathrm{r},11} & \Sigma_{\mathrm{r},12} \\ 0 & \Sigma_{\mathrm{r},22} \end{bmatrix}$, with $\Sigma_{\mathrm{r},11} \in \mathbb{R}^{n \times n}$

- GTLS solution

$$\widehat{x}_{\mathrm{gtls}} = \frac{\Sigma_{\mathrm{r},11} \widehat{x}_{\mathrm{tls}} - \Sigma_{\mathrm{r},11}}{\Sigma_{\mathrm{r},22}}, \quad \widehat{D}_{\mathrm{gtls}} = \left(\Sigma_\mathrm{l}\right)^{-1} \widehat{D}_{\mathrm{m,tls}} \left(\Sigma_\mathrm{r}\right)^{-1}$$

Singular weight matrix

- consider the element-wise weighted case

$$\|D\|_w = \|D\|_\Sigma := \|\Sigma \odot D\|_\mathrm{F}$$

- Σ is a matrix of element-wise nonnegative weights
- $\sigma_{ij} = 0 \implies$ the solution doesn't depend on d_{ij}
- zero weights allow us to consider missing data

Restricted total least squares problem

- impose structured correction ΔD

$$\begin{array}{rl} \text{minimize} & \|E\|_\mathsf{F} \\ \text{subject to} & (A+\Delta y)x = y + \Delta y \\ \text{and} & \begin{bmatrix}\Delta A & \Delta y\end{bmatrix} = LER \end{array}$$

- link to WTLS: RTLS is a GTLS problem with

$$W_\mathsf{l} = (LL^\top)^+ \quad \text{and} \quad W_\mathsf{r} = (RR^\top)^+$$

(A^+ is the pseudo-inverse of A)

Outline

Structured total least squares

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

$$\begin{array}{ll} \text{minimize} & \text{over } x, \Delta A, \Delta y \quad \left\| \begin{bmatrix} \Delta A & \Delta y \end{bmatrix} \right\|_{\mathrm{F}} \\ \text{subject to} & (A + \Delta A)x = y + \Delta y \text{ and} \\ \multicolumn{2}{l}{\begin{bmatrix} \Delta A & \Delta y \end{bmatrix} \text{ has the same structure as } \begin{bmatrix} A & y \end{bmatrix}} \end{array}$$

- types of structure
  - linear: Hankel/Toeplitz, Sylvester
  - nonlinear: Vandermonde
- link to structured low-rank approximation

Link to structured low-rank approximation

- STLS is equivalent to structured low-rank approx.

$$\begin{array}{ll} \text{minimize} \quad \text{over } \Delta D & \|\Delta D\|_{\mathrm{F}} \\ \text{subject to} & \operatorname{rank}(D + \Delta D) \leq r \text{ and} \\ & \color{red}{\Delta D \text{ has the same structure as } D} \end{array}$$

with $D := \begin{bmatrix} A & y \end{bmatrix}$, $r = n$, and

$$e_{n+1} \notin \operatorname{kernel}(\widehat{D}) \qquad (*)$$

- generically, the condition (**??**) is satisfied
- in nongeneric cases, the STLS solution does not exist

History of the problem

- Errors-in-variables system identification

  M. Aoki and P. Yue. On a priori error estimates of some identification methods. *IEEE Trans. Automat. Control*, 15(5):541–548, 1970

- Sum-of-exponetials estimation

  Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acust., Speech, Signal Proc.*, 34:1081–1089, 1986

  J. Cadzow. Signal enhancement—A composite property mapping algorithm. *IEEE Trans. Signal Proc.*, 36:49–62, 1988

- Rimmanian SVD algorithm

  B. De Moor. Structured total least squares and L_2 approximation problems. *Linear Algebra Appl.*, 188–189:163–207, 1993

- Structured total least norm algorithm

  J. Rosen, H. Park, and J. Glick. Total least norm formulation and solution of structured problems. *SIAM J. Matrix Anal. Appl.*, 17:110–126, 1996

- Variable projection algorithm

  I. Markovsky, S. Van Huffel, and R. Pintelon. Block-Toeplitz/Hankel structured total least squares. *SIAM J. Matrix Anal. Appl.*, 26(4):1083–1099, 2005