Total Least Squares and Chebyshev Norm

Milan Hladík[1] & Michal Černý[2]

[1] Department of Applied Mathematics
Faculty of Mathematics & Physics
Charles University in Prague, Czech Republic

[2]Department of Econometrics & DYME Research Center
Faculty of Computes Science & Statistics
University of Economics in Prague, Czech Republic

ICCS 2015, Reykjavik

Classical model:

$$b = Ax - \Delta b$$

Classical model:

$$b = \underbrace{A}_{\substack{\text{known}\\\text{matrix}\\\text{of regressors}}}x - \Delta b$$

To recall: linear regression, OLS and TLS

Classical model:

$$b = Ax - \Delta b$$

known observations of the dependent variable

known matrix of regressors

Classical model:

$$b = Ax - \Delta b$$

known observations of the dependent variable

known matrix of regressors

unknown regression parameters
→ **to estimate!**

To recall: linear regression, OLS and TLS

Classical model:

unknown

$$b = Ax - \Delta b$$

known observations of the dependent variable

known matrix of regressors

unknown regression parameters $\rightarrow$ **to estimate!**

Classical model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known observations of the dependent variable

known matrix of regressors

unknown regression parameters $\rightarrow$ **to estimate!**

Classical model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known observations of the dependent variable

known matrix of regressors

unknown regression parameters $\rightarrow$ **to estimate!**

"Good" estimator: Ordinary Least Squares (OLS)

Find $\Delta b, x^*$ s.t.:
- $Ax^* = b + \Delta b$ is solvable
- and $\|\Delta b\|_2$ is minimal

$$x^* = (A^T A)^{-1} A^T b$$

To recall: linear regression, OLS and TLS

Classical model:

unknown

$$b \;=\; Ax \;-\; \Delta b$$

unknown errors

known known unknown

Errors-in-variables (EIV) model:

$$b \;=\; Ax \;-\; \Delta b$$

To recall: linear regression, OLS and TLS

Classical model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known known unknown

Errors-in-variables (EIV) model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known unknown

To recall: linear regression, OLS and TLS

Classical model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known

known

unknown

Errors-in-variables (EIV) model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known

unknown: known is

$$A' = A - \Delta A$$

unknown errors

unknown

Classical model:

unknown

$$b = Ax - \Delta b$$

unknown errors

known known unknown

Errors-in-variables (EIV) model:

$$b = Ax - \Delta b \qquad A' = A - \Delta A$$

"Good" estimator: Total Least Squares (TLS)

Find $\Delta A, \Delta b, x^*$ s.t.:
- $(A' + \Delta A)x^* = b + \Delta b$ is solvable
- and $\|(\Delta A, \Delta b)\|_F$ is minimal

A reformulation

TLS finds: $\Delta A, \Delta b$ s.t.

- $(A' + \Delta A)x = b + \Delta b$ is solvable and
- $\|(\Delta A, \Delta b)\|_F$ is minimal, where

$$\|Q\|_F = \sqrt{\sum_{i,j} Q_{ij}^2} = \sqrt{\operatorname{trace}(Q^{\mathrm{T}} Q)}$$ is the Frobenius norm.

Our problem (Chebyshev Norm Problem, CNP): find $\Delta A, \Delta b$ s.t.

- $(A' + \Delta A)x = b + \Delta b$ is solvable and
- $\|(\Delta A, \Delta b)\|_{\max}$ is minimal, where

$$\|Q\|_{\max} = \max_{i,j} |Q_{ij}|$$ is the Chebyshev norm.

Motivation

Why to replace $\|\cdot\|_F$ by another norm?

- Robustness arguments — a usage of different norms is a usual method in robust statistics ($\|\cdot\|_F$ is sensitive to outliers and often ill-conditioned);
- Estimation theory arguments — under certain probabilistic assumptions on the errors $\Delta A, \Delta b$, the solution obtained from the Chebyshev Norm Problem gives a consistent estimator for the EIV model.

Definition. Interval $(m \times n)$-matrix is a system of matrices

$$\mathbf{A} = [\underline{A}, \overline{A}] = \{A \in \mathbb{R}^{m \times n} : \ \underline{A} \leqslant A \leqslant \overline{A}\},$$

where $\underline{A} \leqslant \overline{A} \in \mathbb{R}^{m \times n}$ are given and "$\leqslant$" is understood componentwise.

Definition. Solution set of a system of interval-valued linear equations $\mathbf{A}x = \mathbf{b}$ is defined as

$$\mathfrak{S}(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^n : (\exists A \in \mathbf{A})(\exists b \in \mathbf{b}) \ Ax = b\}.$$

Interval-theoretic reformulation of the Chebyshev Norm Problem (CNP): Find the minimum δ such that

$$\mathfrak{S}([A' - \delta E, A' + \delta E], [b - \delta e, b + \delta e]) \neq \emptyset,$$

where E is the all-one matrix and e is the all-one vector.

Oettli-Prager Theorem

Lemma (Oettli-Prager).

$$\mathfrak{S}(\mathbf{A},\mathbf{b}) = \{x \in \mathbb{R}^n : |A^C x - b^C| \leqslant A^\Delta |x| + b^\Delta\},$$

where $A^C = \frac{1}{2}(\overline{A} + \underline{A})$ is the center matrix and $A^\Delta = \frac{1}{2}(\overline{A} - \underline{A})$ is the radius matrix.

Corollary — characterization of the CNP system:

$$\begin{aligned}
&\mathfrak{S}([A' - \delta E, A' + \delta E], [b - \delta e, b + \delta e]) \\
&\qquad = \{x \in \mathbb{R}^n : |A'x - b| \leqslant \delta E|x| + \delta e\} \\
&\qquad = \bigcup_{s \in \{\pm 1\}^n} \left\{ x \in \mathbb{R} : \begin{array}{rcl} (A' - \delta E D_s)x & \leqslant & b + \delta e, \\ (-A' - \delta E D_s)x & \leqslant & -b + \delta e, \\ D_s x & \geqslant & 0 \end{array} \right\},
\end{aligned}$$

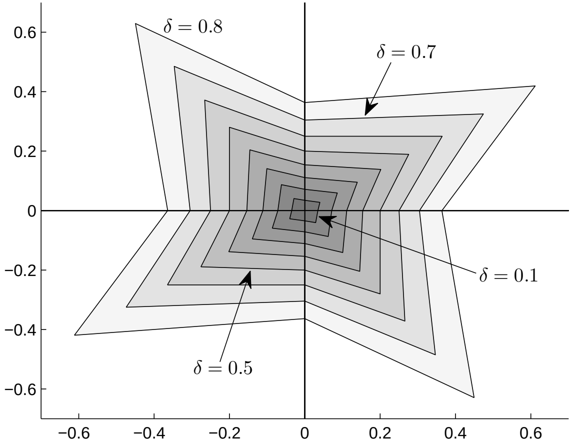
where $D_s = \mathrm{diag}(s)$.

Example of the CNP system

$$\bigcup_{s\in\{\pm1\}^n} \{(A' - \delta ED_s)x \leqslant b + \delta e,\ (-A' - \delta ED_s)x \leqslant -b + \delta e,\ D_s x \geqslant 0\}$$

$$A' = \begin{pmatrix} 3 & -0.5 \\ 0.5 & 3 \\ 0.6 & 3 \end{pmatrix}$$

$$b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

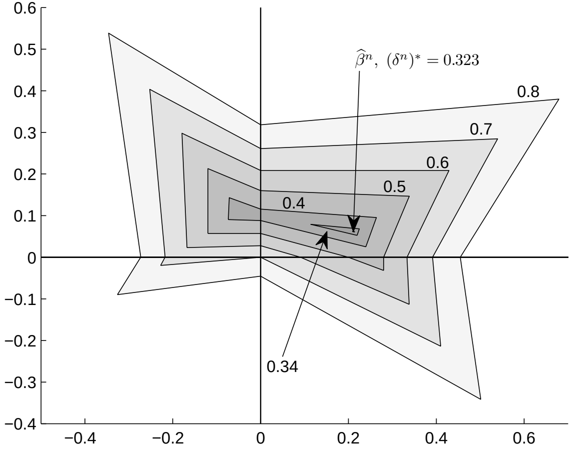


Another example of the CNP system

$$\bigcup_{s\in\{\pm1\}^n} \{(A' - \delta E D_s)x \leqslant b + \delta e,\ (-A' - \delta E D_s)x \leqslant -b + \delta e,\ D_s x \geqslant 0\}$$

$$A' = \begin{pmatrix} 3 & -0.5 \\ 0.5 & 3 \\ 0.6 & 3 \end{pmatrix}$$

$$b = \begin{pmatrix} 0.2 \\ 0.7 \\ -0.1 \end{pmatrix}$$



Reduction to 2^n GLFPs

To recall: To solve CNP, we are to find the minimum δ such that the CNP system

$$\bigcup_{s\in\{\pm 1\}^n} \left\{(A' - \delta E D_s)x \leqslant b + \delta e,\ (-A' - \delta E D_s)x \leqslant -b + \delta e,\ D_s x \geqslant 0\right\}$$

is nonempty.

The main observation: In a given orthant $s \in \{\pm 1\}^n$, it suffices to solve the following generalized linear-fractional programming (GLFP) problem:

$$\min_{x\in\mathbb{R}^n} \max_{\substack{i\in\{1,\dots,m\}\\ j\in\{0,1\}}} \frac{(-1)^{1-j} A'_i x + (-1)^j b_i}{e^T D_s x + 1} \quad \text{s.t.} \quad D_s x \geqslant 0,$$

where A'_i is the i-th row of A'.

An important (well-known) fact. GLFP can be solved in polynomial time via interior point methods.

Remarks

We have shown:

- Bad news. The algorithm is exponential in n, the number of regression parameters.
- Good news. The algorithm is not exponential in m, the number of observations.
- Since usually $n \ll m$, we can say:

Corollary. As long as $n = O(1)$ (i.e., n is a constant independent of m), the method runs in polynomial time.

Comment. In practive we work with regression models with up to $n = 20$ (say) regression parameters. And 2^{20} is large, but still tractable.

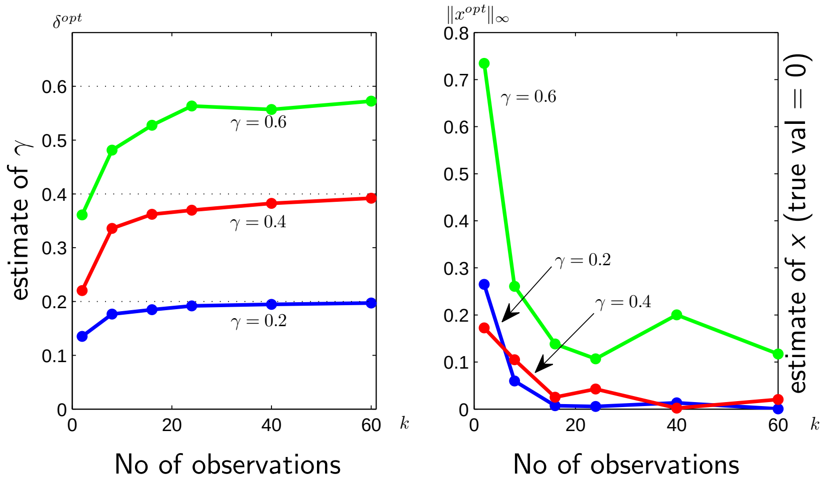
The main question. Can we achieve a better algorithm?

Theorem. The answer is NO. (CNP is NP-hard.)

A probabilistic setup:

- two regression parameters, their true values are zero
- the observations of the regressors are contaminated by independent errors sampled from Unif$(-\gamma, \gamma)$, where $\gamma > 0$ is a parameter
- the observations of the dependent variable are contaminated by independent errors sampled from Unif$(-\gamma, \gamma)$

Example: a simulation study



Remarks and conclusions

- **Further results.** Since the CNP problem is NP-hard, we are interested in designing heuristics. We have also designed some methods for
  - poly-time computable lower bounds,
  - poly-time computable upper bounds.
- **Current work.** Now we are investigating under which probabilistic assumptions on the errors $\Delta A, \Delta b$ the CNP problem gives a consistent estimator of the regression parameters and what is the speed of convergence.
- **Other norms.** The TLS problem is interesting not only with the Chebyshev norm. Other matrix norms are of interest as well.

Thank you for your attention.