Maximum likelihood estimation in linear models with a Gaussian matrix

Ami Wiesel, Yonina C. Eldar and Amir Beck

ICASSP-2006 Special Session on convex optimization
Outline

- ML estimate in random model
  - Problem formulation
  - Efficient numerical solution

- Comparison to similar problems & models
  - Errors in Variables (EIV) model
  - Total Least Squares (TLS)
  - Regularized TLS

- Conclusions
Linear model

\[ y = Gx + w \]

- \( y \) received vector
- \( G \) model matrix
- \( x \) deterministic unknown vector
- \( w \) Gaussian noise vector \( \sim N(0, \sigma_w^2 I) \)

What is the ML estimate of \( x \)?
Background: \( G \) is known

- ML = LS (Least squares)

\[
\hat{x} = \arg \max_x \left\{ f(y; x) \right\} = \arg \min_x \|y - Gx\|^2
\]
Our model: $G$ is random

- What happens when $G$ is unknown?
  - In particular, when it is random.
  - $[G]_{ij}$ are independent and Gaussian

\[
E\{G\} = H \\
VAR\{[G]_{ij}\} = \sigma_h^2
\]
ML estimate in our model

\[ \hat{x} = \arg\max_x \{ f(y; x) \} \]

\[ = \arg\min_x \left\{ \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} + N \log\left(\sigma_h^2 \|x\|^2 + \sigma_w^2\right) \right\} \]

Highly non linear and non convex optimization problem!!!
Efficient numerical solution - I

\[ \hat{x} = \arg \min_x \left\{ \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} + N \log \left( \sigma_h^2 \|x\|^2 + \sigma_w^2 \right) \right\} \]

We add a slack variable:

\[ \hat{x} = \arg \min_t \left\{ \min_{x: \|x\|^2 = t} \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} + N \log \left( \sigma_h^2 t + \sigma_w^2 \right) \right\} \]
Efficient numerical solution - II

\[ \hat{x} = \arg \min_{t} \left\{ \min_{x:||x||^2=t} \| y - Gx \|^2 + N \log \left( \sigma_h^2 t + \sigma_w^2 \right) \right\} \]

Constrained LS

Solvable via hidden convexity!

Unimodal in \( t \)!

We can find the ML estimate by iteratively solving the constrained LS!
Relation to EIV model

The Errors in Variables (EIV) model

\[ y = Gx + w \]
\[ H = G + W \]

- Here, \( G \) is deterministic unknown.
- But we have a new observation - \( H \).
  - Similar to our model!!!
  - (in both \( y \) and \( H \) are available)
Relation to ML in EIV (TLS)

- Reminder: $\mathbf{G}$ is known ML = LS.
- $\mathbf{G}$ is unknown ML = TLS (Total Least Squares).

\[
\left\{ \hat{x}_{TLS}, \hat{\mathbf{G}} \right\} = \arg \max_x \left\{ f \left( \mathbf{y}, \mathbf{H}; x, \mathbf{G} \right) \right\} = \arg \min_{x, \mathbf{G}} \left\{ \frac{\| \mathbf{y} - \mathbf{G}x \|^2}{\sigma_w^2} + \frac{\| \mathbf{H} - \mathbf{G} \|^2_F}{\sigma_h^2} \right\}
\]
The interesting relation...

\[ \{ \hat{x}_{TLS}, \hat{G} \} = \arg \min_{x,G} \left\{ \frac{\|y - Gx\|^2}{\sigma_w^2} + \frac{\|H - G\|^2_F}{\sigma_h^2} \right\} \]

If we solve for \( G \):

\[ \hat{x}_{TLS} = \arg \min_x \left\{ \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} \right\} \]

The difference is just a regularization term!

Reminder: our new estimator (in random \( G \)) is

\[ \hat{x} = \arg \min_x \left\{ \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} + N \log \left( \frac{\sigma_h^2 \|x\|^2 + \sigma_w^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} \right) \right\} \]
It gets even more interesting..

- In many applications, TLS is not stable and must be regularized - RTLS!!
- The usual trick is

\[
\hat{x}_{\text{RTLS}} = \arg \min_x \left\{ \frac{\|y - Gx\|^2}{\sigma_h^2 \|x\|^2 + \sigma_w^2} + \frac{N \sigma_h^2}{\sigma_w^2} \|x\|^2 \right\}
\]

- Which is exactly our new ML estimator when

\[\sigma_h^2 \ll \sigma_w^2\]
To summarize

- **Known** $G$: $ML = LS$
- **Unknown** $G$: $ML = TLS$
- **Random** $G$: $ML = RTLS$

logarithmically
Generalizations

- Just like in TLS literature:
  - Independent rows with a known COV.
  - Some rows known.
  - Some columns known.
- Unfortunately, no structure 😞

- Asymptotic performance analysis via CRB.
Something to think about

- Two alternative models:
  - Random with known mean
  - EIV model

- When should we choose what model?
  - Which model describes reality better?
  - What problem is easier to solve?
For more details:


Generalization & performance analysis in ICASSP 2006

Thank you!!!
Numerical example
CRB in our model

\[
MSE \geq (\sigma_h^2 \|x\|^2 + \sigma_w^2)(H^T H)^{-1} - \Delta(\sigma_h^2, \sigma_w^2, x, H)
\]

\[
\Delta(\sigma_h^2, \sigma_w^2, x, H) \geq 0
\]

\[
MSE(\sigma_h^2 = 0) \geq \sigma_w^2 (H^T H)^{-1}
\]

- Randomness has negative and positive effect.
- CRB depends on the unknown parameter \(x\).
The CRB is a function of $G$ instead of $H$.

It is easier to estimate $x$ in our model than in the EIV model – (no $\Delta$ term)!