IMM

Algorithms and Applications of Huber Estimation

Hans Bruun Nielsen
Informatics and Mathematical Modelling
Technical University of Denmark
hbn@imm.dtu.dk
http://www.imm.dtu.dk/~hbn

Joint work with

- Kaj Madsen, IMM
- Håkan Ekblom, Luleå University, Sweden
- Mustafa Pinar, Bilkent University, Turkey
- Ove Edlund, Luleå University, Sweden

IMM

Outline

- Definitions
- Huber algorithm
- Example: one-sided Huber
- Linear ℓ_1 estimation
- LP algorithm
- QP algorithm
- Ongoing and future work

IMM

Huber's M-Estimator

Given $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ with $m > n$

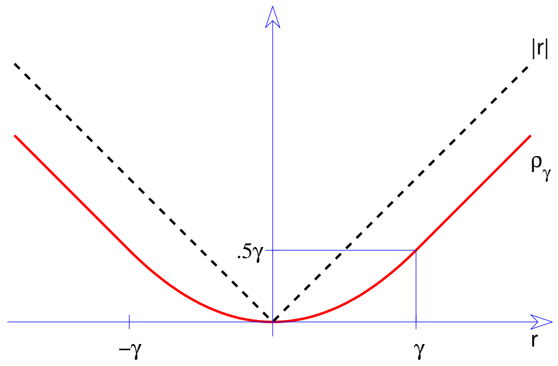
Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{r}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}$

Seek

$$\mathbf{x}_\gamma = \operatorname{argmin}\{F_\gamma(\mathbf{x}) \equiv \sum_{i=1}^{m} \varrho_\gamma(r_i(\mathbf{x}))\}$$

where

$$\varrho_\gamma(r) = \begin{cases} \frac{1}{2\gamma}r^2 & , \ |r| \le \gamma \\ |r| - \frac{1}{2}\gamma, & |r| > \gamma \end{cases}$$



IMM

ℓ_1 and ℓ_2 Estimators

$$\mathbf{x}_{(1)} = \operatorname{argmin}\left\{\|\mathbf{r}(\mathbf{x})\|_1 \equiv \sum_{i=1}^{m} |r_i(\mathbf{x})|\right\}$$

$$\mathbf{x}_{(2)} = \operatorname{argmin}\left\{\tfrac{1}{2}\|\mathbf{r}(\mathbf{x})\|_2^2 \equiv \sum_{i=1}^{m} (r_i(\mathbf{x}))^2\right\}$$

$$\mathbf{x}_\gamma = \operatorname{argmin}\{F_\gamma(\mathbf{x}) \equiv \sum_{i=1}^{m} \varrho_\gamma(r_i(\mathbf{x}))\}$$

$$\varrho_\gamma(r) = \begin{cases} \frac{1}{2\gamma}r^2 & , \ |r| \le \gamma \\ |r| - \frac{1}{2}\gamma, & |r| > \gamma \end{cases}$$

If $\|\mathbf{r}(\mathbf{x}_{(2)})\|_\infty \le \gamma$ then $\mathbf{x}_\gamma = \mathbf{x}_{(2)}$

Intuitively:

$$\mathbf{x}_{(1)} = \lim_{\gamma \to 0} \mathbf{x}_\gamma$$

IMM

Define

active set	$\mathcal{A}$
weights	$\mathbf{W} = \mathrm{diag}(w_i)$
generalized sign vector	$\mathbf{s}$

$$
\begin{array}{lll}
r_i < -\gamma & : & w_i = 0\ ,\quad s_i = -1 \\
-\gamma \le r_i \le \gamma & : & w_i = 1\ ,\quad s_i = 0 \qquad i \in \mathcal{A} \\
r_i > \gamma & : & w_i = 0\ ,\quad s_i = 1
\end{array}
$$

Matrix formulation $\quad (\ \mathbf{r}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}\ )$

$$F_\gamma(\mathbf{x}) = \tfrac{1}{2\gamma}\mathbf{r}^\top \mathbf{W}\mathbf{r} + \mathbf{s}^\top(\mathbf{r} - \tfrac{1}{2}\gamma\mathbf{s})$$

$$\mathbf{F}'_\gamma(\mathbf{x}) = -\mathbf{A}^\top(\tfrac{1}{\gamma}\mathbf{W}\mathbf{r} + \mathbf{s})$$

$$\mathbf{F}''_\gamma(\mathbf{x}) = \tfrac{1}{\gamma}\mathbf{A}^\top\mathbf{W}\,\mathbf{A} \qquad \text{(Pos. semidefinite)}$$

The hyperplanes $r_i(\mathbf{x}) = \pm\gamma$ split $\mathbb{R}^n$ into subregions, in each of which $F_\gamma(\mathbf{x})$ is a quadratic.

$F_\gamma(\mathbf{x})$ is a convex, piecewise quadratic function

IMM

Huber algorithm

Seek $\mathbf{x}_\gamma$: $\mathbf{F}'_\gamma(\mathbf{x}_\gamma) = \mathbf{0}$. Newton iteration

Given initial guess $\mathbf{x}$ $\qquad$ (e.g. $\mathbf{x}_{(2)}$)

`repeat`

Solve $(\mathbf{A}^\top\mathbf{W}\,\mathbf{A})\mathbf{h} = \mathbf{A}^\top(\mathbf{W}\mathbf{r} + \gamma\mathbf{s}) \quad (*)$

`if` $\mathcal{A}(\mathbf{x}+\mathbf{h}) = \mathcal{A}(\mathbf{x})$

$\mathbf{x}_\gamma := \mathbf{x} + \mathbf{h}$; STOP

`else`

$\widetilde{\alpha} := \mathrm{argmin}\{F_\gamma(\mathbf{x}+\alpha\mathbf{h})\}$ (Line search)

$\mathbf{x} := \mathbf{x} + \widetilde{\alpha}\mathbf{h}$

Robust: Non-active residuals contribute just with γ times their sign.

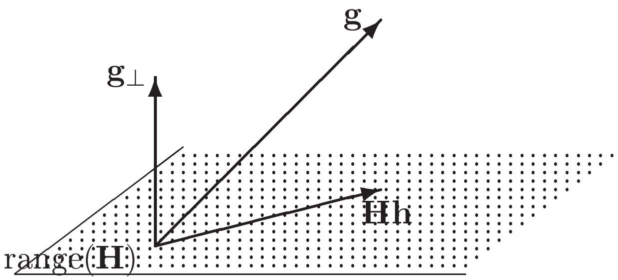
Newton step

$$(\mathbf{A}^\top \mathbf{W}\ \mathbf{A})\mathbf{h} \;=\; \mathbf{A}^\top(\mathbf{W}\mathbf{r} + \gamma \mathbf{s}) \qquad (*)$$

$$\mathbf{H}\mathbf{h} \;=\; \mathbf{g}$$

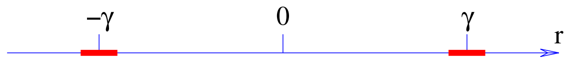
If $\mathbf{H}$ is singular:

$$\mathbf{h} = \begin{cases} \mathbf{h}_{\mathrm{min\_norm}} & \text{if } \ \mathbf{g} \in \mathrm{range}(\mathbf{H}) \\ \mathbf{g}_\perp & \text{otherwise} \end{cases}$$



Same active set

Fuzzy region



s =	−1	−.25	0	.25	1
w =	0	1	1	1	0

Width $2\tau = 20\varepsilon_{\mathrm{M}}\|\mathbf{b}\|_\infty$.

The check	$\mathcal{A}(\mathbf{x}) = \mathcal{A}(\mathbf{x}_{\mathrm{prev}})$
is implemented as	$\|\mathbf{s}(\mathbf{x}) - \mathbf{s}(\mathbf{x}_{\mathrm{prev}})\|_\infty < 1$

Line search

$$\widetilde{\alpha} := \mathrm{argmin}\{F_\gamma(\mathbf{x}{+}\alpha\mathbf{h})\}$$

$$\mathbf{F}'_\gamma(\mathbf{x} + \widetilde{\alpha}\mathbf{h}) = 0$$

Active set changes at *kink values*:

$$r_k - \alpha_k(\mathbf{A}\mathbf{h})_k = \pm\gamma$$

$\mathbf{F}'_\gamma(\mathbf{x} + \alpha\mathbf{h})$ is piecewise linear.
Simple updating of coefficients.

IMM

Factorization

$$\begin{aligned}\mathbf{H} &= \mathbf{A}^\top \mathbf{W}\,\mathbf{A}\\ &= \sum_{i\in\mathcal{A}} \mathbf{A}_{i,:}^\top \mathbf{A}_{i,:}\\ &= \mathbf{L}\mathbf{D}\mathbf{L}^\top\end{aligned}$$

Next iterate: normally only a few changes in active set.

$$\mathbf{H}_{\text{new}} = \mathbf{H} + \sum_{i\in\mathcal{E}} \mathbf{A}_{i,:}^\top \mathbf{A}_{i,:} - \sum_{i\in\mathcal{L}} \mathbf{A}_{i,:}^\top \mathbf{A}_{i,:}$$

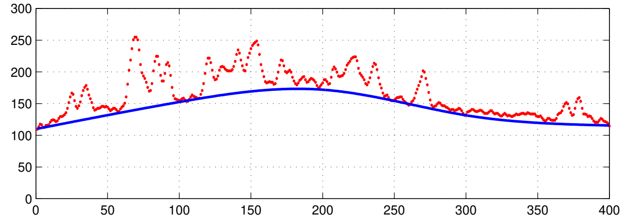
Cheap up– and downdates of factorization.

km & hbn: BIT (1990), 682–699.

IMM

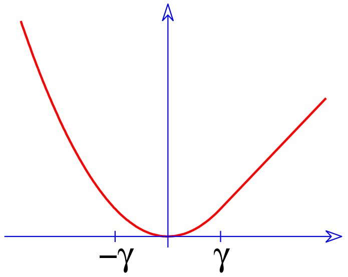
Example. One-sided Huber

Electrophoretic analysis of wheat



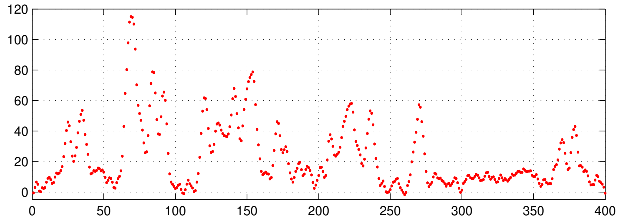
Smoothly varying background noise.
Model this by a cubic spline (with 5 equi-distant knots).
Coefficients determined by one-sided Huber:

$$\tilde{\varrho}_\gamma(r) = \begin{cases} \frac{1}{2\gamma} r^2 & , \quad r \le \gamma \\ |r| - \frac{1}{2}\gamma, & r > \gamma \end{cases}$$



$\gamma = .01.$
7 iterations to find the solution.

"Reduced data" were used to train a neural network.



hbn et *al*: Electrophoresis (1995), 921–926.

ℓ_1 Estimator

$$\mathbf{x}_{(1)} = \operatorname{argmin}\{\|\mathbf{r}(\mathbf{x})\|_1\} = \lim_{\gamma \to 0} \mathbf{x}_\gamma$$

Huber condition:

$$\mathbf{A}^\top \left(\frac{1}{\gamma}\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x}_\gamma) + \mathbf{s}\right) \;=\; \mathbf{0}$$

Reduce γ:

$$\mathbf{x}_{\gamma-\delta} = \mathbf{x}_\gamma + \delta\mathbf{u}; \quad (\mathbf{A}^\top\mathbf{W}\,\mathbf{A})\mathbf{u} = \mathbf{A}^\top\mathbf{s}$$

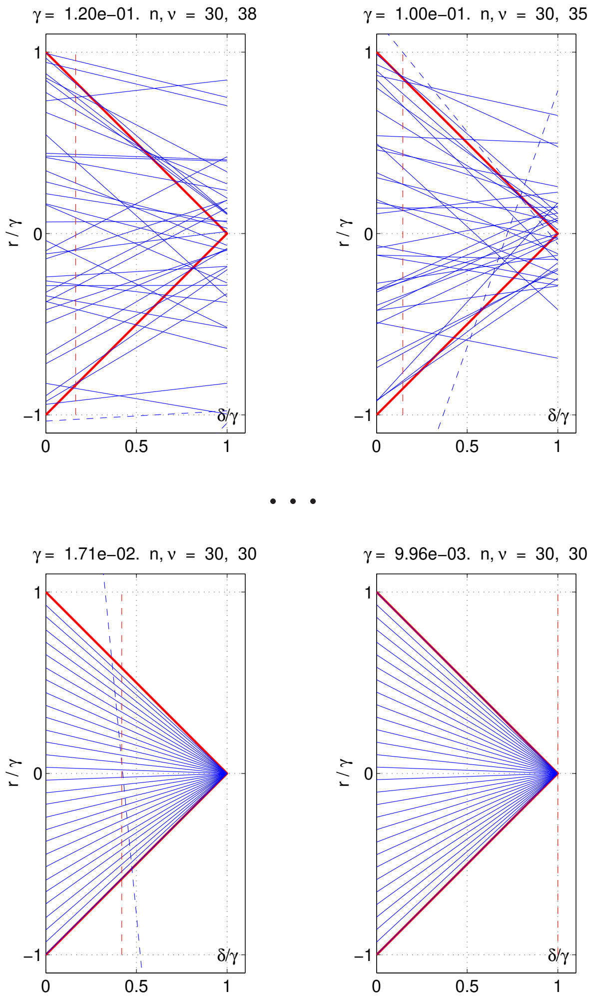
$$\mathbf{r}(\mathbf{x}_{\gamma-\delta}) = \mathbf{r}(\mathbf{x}_\gamma) - \delta\mathbf{A}\mathbf{u}$$

Kinks: $\quad r_k - d_k(\mathbf{A}\mathbf{u})_k = \gamma - d_k$

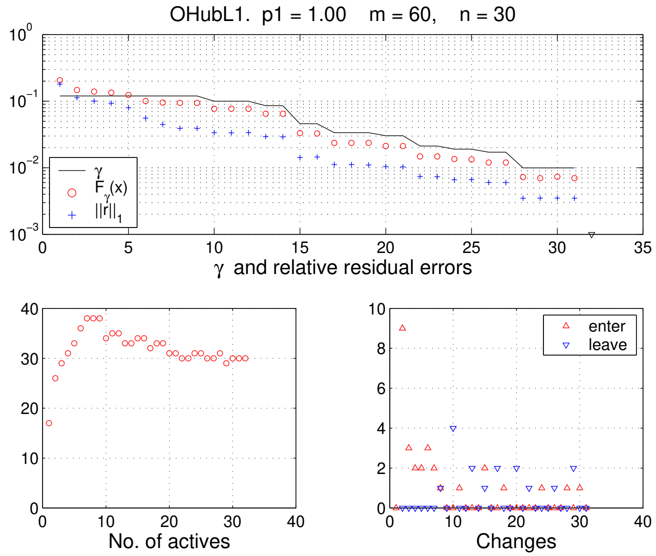
Reduced γ:

$$\delta \geq 0.1\gamma, \qquad {\scriptstyle \nu_{\rm red} < \frac{1}{2}(n+\nu)}$$

$\nu = \#\mathcal{A}$, number of active elements.



History



$$\mathbf{H}_{\rm new} = \mathbf{H} + \mathbf{A}_{\mathcal{E},:}^{\top}\mathbf{A}_{\mathcal{E},:} - \mathbf{A}_{\mathcal{L},:}^{\top}\mathbf{A}_{\mathcal{L},:}$$

The "final" active set is identified while γ is strictly positive.

km & hbn: SIAM J. Opt. (1993), 223–235.

IMM

LP Problems

Primal:

$$\mathbf{y}^* = \operatorname{argmax}\left\{\mathbf{b}^\top\mathbf{y} \;:\; \mathbf{A}^\top\mathbf{y} = \mathbf{c},\; -\mathbf{1} \le \mathbf{y} \le \mathbf{1}\right\}$$

Dual:

$$\mathbf{x}^* = \operatorname{argmin}\left\{\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_1 + \mathbf{c}^\top\mathbf{x}\right\}$$

"Huber formulation"

$$\mathbf{x}_\gamma = \operatorname{argmin}\{F_\gamma(\mathbf{x})\}$$

$$F_\gamma(\mathbf{x}) = \tfrac{1}{2\gamma}\mathbf{r}^\top\mathbf{W}\mathbf{r} + \mathbf{s}^\top(\mathbf{r} - \tfrac{1}{2}\gamma\mathbf{s}) + \mathbf{c}^\top\mathbf{x}$$

$$\mathbf{F}'_\gamma(\mathbf{x}) = -\mathbf{A}^\top(\tfrac{1}{\gamma}\mathbf{W}\mathbf{r} + \mathbf{s}) + \mathbf{c}$$

$$\mathbf{x}^* = \lim_{\gamma \to 0} \mathbf{x}_\gamma$$

$$\mathbf{y}^* = \lim_{\gamma \to 0} \frac{1}{\gamma}\mathbf{W}(\mathbf{x}_\gamma)\mathbf{r}(\mathbf{x}_\gamma) + \mathbf{s}(\mathbf{x}_\gamma)$$

km, hbn & mp: SIAM J. Opt. (1996), 600–616.
oe: Doctoral thesis (sparse), Luleå. 1999.

IMM

QP Problems

Primal:

$$\min\left\{q(\mathbf{y}) \equiv \tfrac{1}{2}\mathbf{y}^\top\mathbf{H}\,\mathbf{y} - \mathbf{c}^\top\mathbf{y} \;:\; -\mathbf{1} \le \mathbf{y} \le \mathbf{1}\right\}$$

$\mathbf{H} \in \mathbb{R}^{n\times n}$ symmetric, positive definite (SPD).
Eigenvalues $0 < \lambda_1(\mathbf{H}) \le \ldots \le \lambda_n(\mathbf{H})$
Find γ and $\mathbf{A}$:

$$0 < \gamma < \lambda_1(\mathbf{H}), \qquad \mathbf{A}^\top\mathbf{A} = \mathbf{H} - \gamma\mathbf{I} \quad \text{(SPD)}$$

Dual:

$$-\min\left\{F(\mathbf{x}) \equiv \tfrac{1}{2}\mathbf{r}^\top\mathbf{W}\,\mathbf{r} + \mathbf{s}^\top(\mathbf{r} - \tfrac{1}{2}\gamma\mathbf{s}) + \tfrac{1}{2}\mathbf{x}^\top\mathbf{x}\right\}$$

with $\mathbf{r} = \mathbf{r}(\mathbf{x}) = \mathbf{c} - \mathbf{A}^\top\mathbf{x}$.

Dual solution $\mathbf{x}^*$: $\quad \mathbf{F}'(\mathbf{x}^*) = \mathbf{0}$,

$$\mathbf{F}'(\mathbf{x}) = -\mathbf{A}^\top(\tfrac{1}{\gamma}\mathbf{W}\mathbf{r} + \mathbf{s}) + \mathbf{x}$$

$$\mathbf{F}''(\mathbf{x}) = \tfrac{1}{\gamma}\mathbf{A}^\top\mathbf{W}\,\mathbf{A} + \mathbf{I} \qquad \text{(SPD)}$$

Primal solution: $\mathbf{y}^* = \tfrac{1}{\gamma}\mathbf{W}(\mathbf{x}^*)\mathbf{r}(\mathbf{x}^*) + \mathbf{s}(\mathbf{x}^*)$

IMM

Algorithmic details

$\gamma = \frac{1}{2}\overline{\sigma}_n^2$. σ_n is the smallest singular value of $\mathbf{R}$, $\mathbf{R}^\top\mathbf{R} = \mathbf{H}$, and $\overline{\sigma}_n$ is an estimate, cf. condition estimators.

$\mathbf{A}^\top\mathbf{A} = \mathbf{H} - \gamma\mathbf{I} = \mathbf{R}^\top\mathbf{R} - \gamma\mathbf{I}$. Via orthogonal transformation of

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \mathbf{A} \\ \sqrt{\gamma}\mathbf{I} \end{bmatrix}$$

$\mathbf{AWA}^\top + \gamma\mathbf{I} = \mathbf{L}^\top\mathbf{L}$. Untraditional factorization; cheaper to up- and downdate.

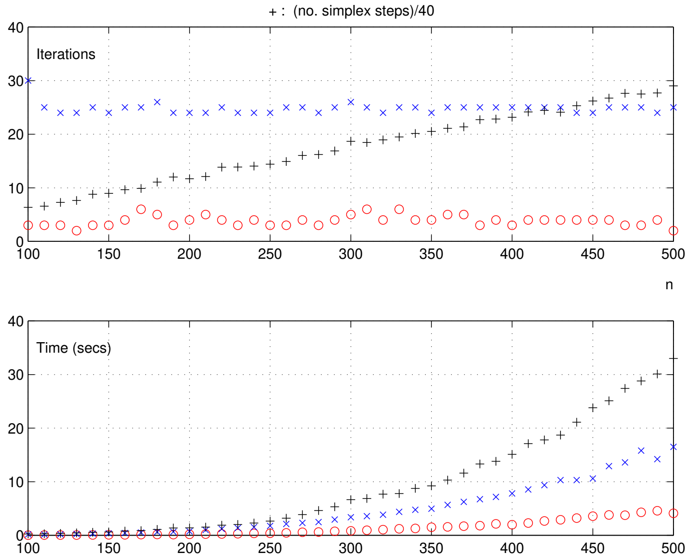
$\mathbf{x}_0$ Careful choice, taking into account both primal and dual problem.

hbn: IMM-REP-1996-21

IMM

Competing methods

- Interior point: Han, Pardalos & Ye (1990). (Marked by $\times$)
- Simplex type: Fletcher `bqpd` (1993). (Marked by $+$)



km, hbn & mp: Math. Prog. (1999), 135–156.

IMM

Ongoing and future projects

- Combine with regularization tools (with P.C. Hansen and M. Jacobsen, IMM).
- Extend QP algorithm to general constraints.
- Library in C/C++ for minimizing a piecewise quadratic (MATLAB version: IMM-REP-1998-14).
- Sparse library.
- Preprocessing and starting point for LP algorithm.