
Terry Fest, May 2015

Level Set Methods in Convex Optimization

Objective-Constraint Interchange Methods
"Flippy"

James V. Burke
UW Mathematics
jvburke@uw.edu

Aleksandr Y. Aravkin
IBM, T.J.Watson Research
sasha.aravkin@gmail.com

Dmitriy Drusvyatskiy
UW Mathematics
ddrusv@uw.edu

Michael P. Friedlander
UC Davis Mathematics
mpf@math.ucdavis.edu

Scott Roy
UW Mathematics
scott.michael.roy@gmail.com

Thank you Samir, Vincent, and Henri!

Problem Framework

Assume ρ and ϕ are closed, proper, and convex

$$\mathcal{P}_1(\sigma): \min \phi(x) \quad \text{st} \quad \rho(b - Ax) \leq \sigma$$

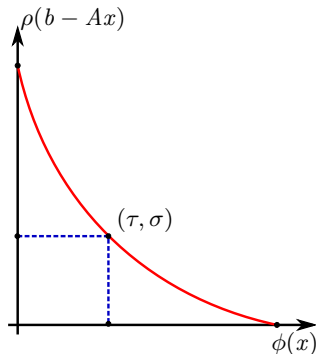
$$\mathcal{P}_2(\tau): \min \rho(b - Ax) \quad \text{st} \quad \phi(x) \leq \tau$$

$\mathcal{P}_1(\sigma)$ is the target problem

$\mathcal{P}_2(\tau)$ is the easier flipped problem.

Problems $\mathcal{P}_1(\sigma)$ and $\mathcal{P}_2(\tau)$ are linked by

$$v_2(\tau) := \min_x \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$



Problem Framework

Assume ρ and ϕ are closed, proper, and convex

$$\mathcal{P}_1(\sigma): \min \phi(x) \quad \text{st} \quad \rho(b - Ax) \leq \sigma$$

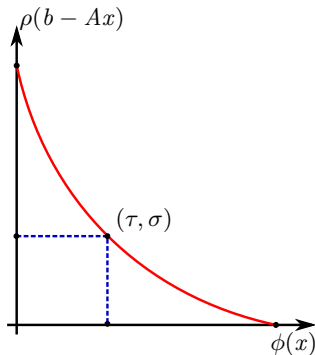
$$\mathcal{P}_2(\tau): \min \rho(b - Ax) \quad \text{st} \quad \phi(x) \leq \tau$$

$\mathcal{P}_1(\sigma)$ is the target problem

$\mathcal{P}_2(\tau)$ is the easier flipped problem.

Problems $\mathcal{P}_1(\sigma)$ and $\mathcal{P}_2(\tau)$ are linked by

$$v_2(\tau) := \min_x \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$



Broad summary of results:

- 1 $v_2(\tau)$ is always convex (inf-projection), but may not be differentiable.
- 2 Solve $v_2(\tau) = \sigma$ by an inexact secant (Newton) method.
- 3 We have precise knowledge of the variational properties of $v_2(\tau)$ for a large classes of problems $\mathcal{P}_2(\tau)$.

Basic Definitions

The **horizon function** of h

$$h^\infty(z) := \sup_{x \in \text{dom}(h)} [h(x+z) - h(x)].$$

The **perspective function** of h

$$\tilde{h}(z, \lambda) := \begin{cases} \lambda h(\lambda^{-1}z) & \text{if } \lambda > 0, \\ \delta(x \mid 0) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0. \end{cases}$$

The **closure of the perspective function** of h

$$h^\pi(z, \lambda) := \begin{cases} \lambda h(\lambda^{-1}z) & \text{if } \lambda > 0, \\ h^\infty(z) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0. \end{cases}$$

Support Functions for $\text{epi}(h)$ and $\text{lev}_h(\tau)$ (Rockafellar 1966)

$h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be closed proper and convex.

Then

$$\delta^*((y, \mu) \mid \text{epi}(h)) = (h^*)^\pi(y, -\mu)$$

and

$$\delta^*(y \mid \text{lev}_h(\tau)) = \text{cl} \left(\inf_{\mu \geq 0} [\tau\mu + (h^*)^\pi(y, \mu)] \right),$$

where

$$\text{epi}(h) := \{(x, \mu) \mid h(x) \leq \mu\}$$

$$\text{lev}_h(\tau) := \{x \mid h(x) \leq \tau\}$$

$$\delta^*(z \mid C) := \sup_{w \in C} \langle z, w \rangle$$

Perturbation Framework and Duality (Rockafellar (1970))

The perturbation function

$$f(x, b, \tau) := \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

Its conjugate

$$f^*(y, u, \mu) = (\phi^*)^\pi(y + A^T u, -\mu) + \rho^*(u) .$$

Perturbation Framework and Duality (Rockafellar (1970))

The perturbation function

$$f(x, b, \tau) := \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

Its conjugate

$$f^*(y, u, \mu) = (\phi^*)^\pi(y + A^T u, -\mu) + \rho^*(u).$$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min_x f(x, b, \tau).$$

The Dual Problem

$$\mathcal{D}(b, \tau) : \quad \hat{v}(b, \tau) := \sup_{u, \mu} \langle b, u \rangle + \tau \mu - f^*(0, u, \mu)$$

Perturbation Framework and Duality (Rockafellar (1970))

The perturbation function

$$f(x, b, \tau) := \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

Its conjugate

$$f^*(y, u, \mu) = (\phi^*)^\pi(y + A^T u, -\mu) + \rho^*(u).$$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min_x f(x, b, \tau).$$

The Dual Problem

$$\mathcal{D}(b, \tau) : \quad \hat{v}(b, \tau) := \sup_{u, \mu} \langle b, u \rangle + \tau \mu - f^*(0, u, \mu)$$

$$\text{(reduced dual)} \quad = \sup_u \langle b, u \rangle - \rho^*(u) - \delta^*(A^T u \mid \text{lev}_\phi(\tau)).$$

Perturbation Framework and Duality (Rockafellar (1970))

The perturbation function

$$f(x, b, \tau) := \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

Its conjugate

$$f^*(y, u, \mu) = (\phi^*)^\pi(y + A^T u, -\mu) + \rho^*(u).$$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min_x f(x, b, \tau).$$

The Dual Problem

$$\mathcal{D}(b, \tau) : \quad \hat{v}(b, \tau) := \sup_{u, \mu} \langle b, u \rangle + \tau \mu - f^*(0, u, \mu)$$

$$\text{(reduced dual)} \quad = \sup_u \langle b, u \rangle - \rho^*(u) - \delta^*(A^T u \mid \text{lev}_\phi(\tau)).$$

The Subdifferential: If $(b, \tau) \in \text{int}(\text{dom}(v))$, then $v(b, \tau) = \hat{v}(b, \tau)$ and

$$\emptyset \neq \partial v(b, \tau) = \arg \max_{u, \mu} \mathcal{D}(b, \tau)$$

The Constraint Qualification $(b, \tau) \in \text{int}(\text{dom}(v))$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min_x \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

The Constraint Qualification (Slater CQ)

$$0 \in \text{int}(\text{dom}(v)) \iff \exists \hat{x} \text{ st } \phi(\hat{x}) < \tau \text{ and } b - A\hat{x} \in \text{int}(\text{dom}(\rho))$$

The Constraint Qualification $(b, \tau) \in \text{int}(\text{dom}(v))$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min_x \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi))$$

The Constraint Qualification (Slater CQ)

$$0 \in \text{int}(\text{dom}(v)) \iff \exists \hat{x} \text{ st } \phi(\hat{x}) < \tau \text{ and } b - A\hat{x} \in \text{int}(\text{dom}(\rho))$$

Solution Existence: Coercivity Conditions

- The dual objective is coercive iff $b \in \text{int}(\text{dom}(\rho) + A(\text{lev}_\phi(\tau)))$.
- The primal objective is coercive iff

$$\text{hzn}(\phi) \cap [-A^{-1}\text{hzn}(\rho)] = \{0\},$$

where

$$\text{hzn}(p) := \{y \mid p^\infty(y) \leq 0\} = [\text{lev}_p(\tau)]^\infty \quad \forall \tau > \inf p$$

KKT Conditions:

$$\bar{u} \in \partial\rho(b - A\bar{x}) \quad \text{and} \quad A^T \bar{u} \in N(\bar{x} | \text{lev}_\phi(\tau))$$

Subdifferential Representation 1: If $\partial v(b, \tau) \neq \emptyset$, then

$$\partial v(b, \tau) = \left\{ \begin{pmatrix} \bar{u} \\ -\bar{\mu} \end{pmatrix} \mid \begin{array}{l} (\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ satisfy KKT cond. and} \\ \bar{\mu} \in \arg \min_{\mu \geq 0} [\tau\mu + (\phi^*)^\pi(A^T \bar{u}, \mu)] \end{array} \right\}.$$

Subdifferential Representation 2: If $\partial v(b, \tau) \neq \emptyset$ and cone $(\partial\phi(\bar{x}))$ is closed for all $\bar{x} \in \arg \min_x f(x, b, \tau)$, then

$$\partial v(b, \tau) = \left\{ \begin{pmatrix} \bar{u} \\ -\bar{\mu} \end{pmatrix} \mid \begin{array}{l} \exists \bar{x} \text{ s.t. } 0 \in -A^T \partial\rho(b - A\bar{x}) + \bar{\mu}^+ \partial\phi(\bar{x}) \\ \text{with } \bar{\mu} \geq 0 \text{ and } \bar{\mu}(\phi(\bar{x}) - \tau) = 0 \end{array} \right\}.$$

KKT Conditions:

$$\bar{u} \in \partial\rho(b - A\bar{x}) \quad \text{and} \quad A^T \bar{u} \in N(\bar{x} | \text{lev}_\phi(\tau))$$

Subdifferential Representation 1: If $\partial v(b, \tau) \neq \emptyset$, then

$$\partial v(b, \tau) = \left\{ \begin{pmatrix} \bar{u} \\ -\bar{\mu} \end{pmatrix} \mid \begin{array}{l} (\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ satisfy KKT cond. and} \\ \bar{\mu} \in \arg \min_{\mu \geq 0} [\tau\mu + (\phi^*)^\pi(A^T \bar{u}, \mu)] \end{array} \right\}.$$

Subdifferential Representation 2: If $\partial v(b, \tau) \neq \emptyset$ and $\text{cone}(\partial\phi(\bar{x}))$ is closed for all $\bar{x} \in \arg \min_x f(x, b, \tau)$, then

$$\partial v(b, \tau) = \left\{ \begin{pmatrix} \bar{u} \\ -\bar{\mu} \end{pmatrix} \mid \begin{array}{l} \exists \bar{x} \text{ s.t. } 0 \in -A^T \partial\rho(b - A\bar{x}) + \bar{\mu}^+ \partial\phi(\bar{x}) \\ \text{with } \bar{\mu} \geq 0 \text{ and } \bar{\mu}(\phi(\bar{x}) - \tau) = 0 \end{array} \right\}.$$

$$\partial v_2(\tau) = \left\{ \omega \mid \exists u \text{ st } \begin{pmatrix} u \\ \omega \end{pmatrix} \in \partial v(b, \tau) \right\}$$

Piecewise Linear-Quadratic Penalties (Rockafellar 1988)

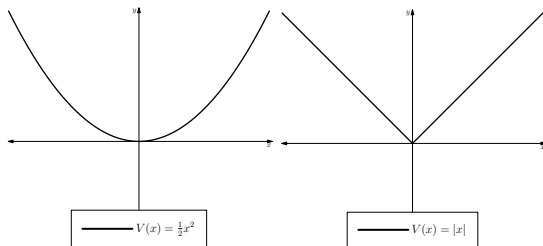
$$\phi(x) := \sup_{u \in U} [\langle x, u \rangle - \frac{1}{2} u^T B u]$$

$U \subset \mathbb{R}^n$ is nonempty, closed and convex with $0 \in U$ (not nec. poly.)
 $B \in \mathbb{R}^{n \times n}$ is symmetric positive semi-definite.

Examples:

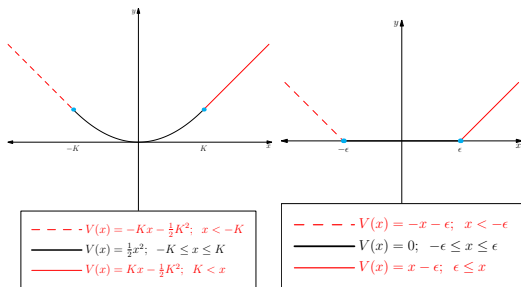
- 1 Support functionals: $B = 0$
- 2 Gauge functionals: $\gamma(\cdot \mid U^\circ) = \delta^*(\cdot \mid U)$
- 3 Norms: $\mathbb{B} =$ closed unit ball, $\|\cdot\| = \gamma(\cdot \mid \mathbb{B})$
- 4 Least-squares: $U = \mathbb{R}^n$, $B = I$
- 5 Huber: $U = [-\epsilon, \epsilon]^n$, $B = I$

PLQ Densities: Gauss, Laplace, Huber, Vapnik



Gauss

Laplace



Huber

Vapnik

$$\phi(x) := \sup_{u \in U} [\langle x, u \rangle - \frac{1}{2} u^T B u]$$

$$\mathcal{P}(b, \tau) : \quad v(b, \tau) := \min \rho(b - Ax) \quad \text{st } \phi(x) \leq \tau$$

$$\partial v(b, \tau) = \left\{ \begin{array}{l} \left(\begin{array}{c} \bar{u} \\ -\bar{\mu} \end{array} \right) \left| \begin{array}{l} (\bar{x}, \bar{u}) \text{ satisfy the KKT cond. for } \mathcal{P}(b, \tau) \text{ and} \\ \bar{\mu} = \max \left\{ \gamma(A^T \bar{u} \mid U), \sqrt{\bar{u}^T A B A^T \bar{u}} / \sqrt{2\tau} \right\} \end{array} \right. \end{array} \right\}.$$

A Few Special Cases

$$v_2(\tau) := \min \frac{1}{2} \|b - Ax\|_2^2 \quad \text{st } \phi(x) \leq \tau$$

Optimal Solution: \bar{x}

Optimal Residual: $\bar{r} = A\bar{x} - b$

1 Support functionals: $\phi(x) = \delta^*(x | U)$, $0 \in U \implies$
 $v'_2(\tau) = -\delta^*(A^T \bar{r} | U^\circ) = -\gamma(A^T \bar{r} | U)$

2 Gauge functionals: $\phi(x) = \gamma(x | U)$, $0 \in U \implies$
 $v'_2(\tau) = -\gamma(A^T \bar{r} | U^\circ) = -\delta^*(A^T \bar{r} | U)$

3 Norms: $\phi(x) = \|x\| \implies v'_2(\tau) = -\|A^T \bar{r}\|_*$

4 Huber: $\phi(x) = \sup_{u \in [-\epsilon, \epsilon]^n} [\langle x, u \rangle - \frac{1}{2} u^T u] \implies$
 $v'_2(\tau) = -\max\{\epsilon \|A^T \bar{r}\|_\infty, \|A^T \bar{r}\|_2 / \sqrt{2\tau}\}$

5 Vapnik: $\phi(x) = \|(x - \epsilon)_+\|_1 + \|(-x - \epsilon)_+\|_1 \implies$
 $v'_2(\tau) = -(\|A^T \bar{r}\|_\infty + \epsilon \|A^T \bar{r}\|_1)$

Definition [Inexact function evaluation oracle]

Given $f : \mathbb{R} \rightarrow \mathbb{R}$, an inexact evaluation oracle is a mapping \mathcal{O}_v assigning to each pair $(x, \alpha) \in \mathbb{R} \times (1, \infty)$ real numbers $(l, u) \in \mathbb{R} \times \mathbb{R}$ satisfying

$$l \leq f(x) \leq u \quad \text{and} \quad 1 \leq \frac{u}{l} \leq \alpha.$$

Secant Input Data:

A non increasing convex function $v : \mathbb{R} \rightarrow \mathbb{R}$ and an associated inexact evaluation oracle \mathcal{O}_v , a target accuracy $\beta > 0$, initial points $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$ and $v(t_0) \geq v(t_1) > 0$, and a constant $\alpha \in (1, 2)$.

An Inexact Secant Method

Initialization: Set $(l_0, u_0) = \mathcal{O}_v(t_0, \alpha)$, $(l_1, u_1) = \mathcal{O}_v(t_1, \alpha)$ and $k = 1$.

Iteration:

While $u_k > \beta$

if $u_k > u_{k-1}$

$$u_k \leftarrow u_{k-1}$$

if $l_k > l_{k-1}$

$$l_k \leftarrow l_{k-1}$$

end-if:

$$\text{Set } \hat{m}_k := \frac{u_{k-1} - l_k}{t_{k-1} - t_k} \text{ and } t_{k+1} := t_k + \frac{l_k}{-\hat{m}_k}.$$

Compute lower and upper bounds $(l_{k+1}, u_{k+1}) := \mathcal{O}_v(t_{k+1}, \alpha)$.

Set $k \leftarrow k + 1$.

end-while.

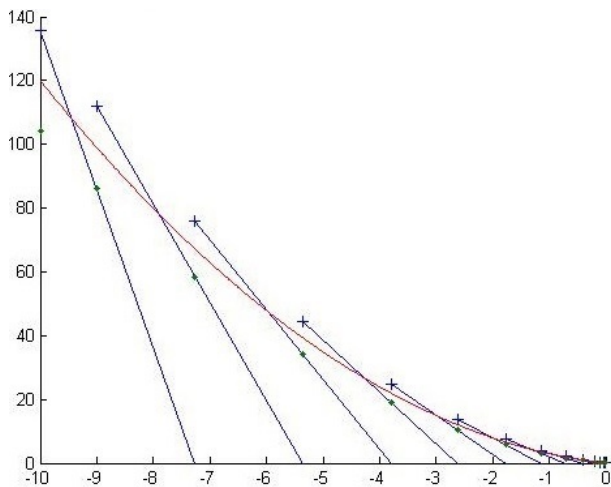
Suppose there exists $\bar{t} > t_1$ such that

$$v(\bar{t}) = 0 \quad \text{and} \quad \partial v(\bar{t}) \cap (-\infty, 0) \neq \emptyset.$$

If the inexact secant method is implemented with $\beta = 0$, then

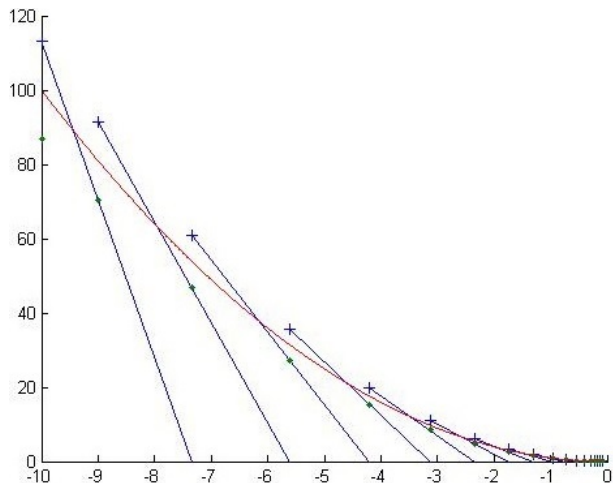
$$t_k \rightarrow \bar{t} \quad \text{at a superlinear rate.}$$

Illustration



$f(t) = (t - 1)^2 - 1$ with $\alpha = 1.3$, $\beta = 0.01$, #iterations = 16.

Illustration: Degenerate Case



$f(t) = t^2$ with $\alpha = 1.3$, $\beta = 0.01$, #iterations = 18.

Theorem: The inexact secant method terminates (i.e., $u_K \leq \beta$) after at most

$$K \leq \max \left\{ \log_{2/\alpha} \left(\frac{|\hat{m}_1|R}{\beta} \right) + \log_{2/\alpha}(2) \cdot \log_{2/\alpha} \left(\frac{2l_1}{\beta} \right) + 1, 2 \right\}$$

iterations, where we set $R = \bar{t} - t_1$ (even in the degenerate case).

In particular, in the case $\alpha = \sqrt{2} \approx 1.41$, the iteration bound simplifies:

$$K \leq \max \left\{ \log_{\sqrt{2}} \left(\frac{l_1^2 |\hat{m}_1|R}{\beta^3} \right) + 5, 2 \right\}.$$

The Primal Problem

$$\mathcal{P}(b, \tau) : \quad \min_x \rho(b - Ax) + \delta((x, \tau) \mid \text{epi}(\phi)) .$$

The Dual Problem

$$\mathcal{D}(b, \tau) : \quad \sup_u \langle b, u \rangle - \rho^*(u) - \delta^*(A^T u \mid \text{lev}_\phi(\tau)) =: h(u, \tau).$$

$h(u, \tau)$ is concave in u and convex in τ .

Let $(\bar{u}, \bar{\tau}) \in \text{dom}(h)$ and let $\bar{s} \in \partial_\tau h(\bar{u}, \bar{\tau})$, then, by weak duality, the line

$$\tau \mapsto h(\bar{u}, \bar{\tau}) + \bar{s}(\tau - \bar{\tau})$$

is a lower minorant of the value function $v(\tau)$.

Inexact Newton Method

Consider any two points $\hat{\tau} < \bar{\tau}$ along with lower and upper bounds,

$$\hat{l} \leq v(\hat{\tau}) \leq \hat{u} \quad \text{and} \quad \bar{l} \leq v(\bar{\tau}) \leq \bar{u}.$$

Suppose we are given $(\bar{u}, \bar{\tau}) \in \text{dom}(h)$ satisfying $\bar{l} = h(\bar{u}, \bar{\tau})$.

Then the root of the secant line

$$l_s(\tau) := \bar{l} + \frac{\hat{u} - \bar{l}}{\hat{\tau} - \bar{\tau}}(\tau - \bar{\tau})$$

is no greater than the root of the Newton line

$$l_N(\tau) = \bar{l} + \bar{s}(\tau - \bar{\tau})$$

for any $\bar{s} \in \partial_{\tau} h(\bar{u}, \bar{\tau})$ (use subdifferential formulas for PLQ functions).

Sparse and Robust Formulation

$$\text{HBP}_{\sigma}: \min \|x\|_1 \text{ st } \rho(b - Ax) \leq \sigma$$

Problem Specification

x 20-sparse spike train in \mathbb{R}^{512}

b measurements in \mathbb{R}^{120}

A Measurement matrix satisfying RIP

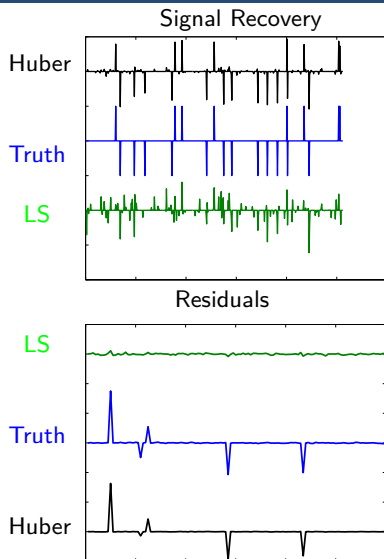
ρ Huber function

σ error level set at .01

5 outliers

Results

In the presence of outliers, the robust formulation recovers the spike train, while the standard formulation does not.



Sparse and Robust Formulation

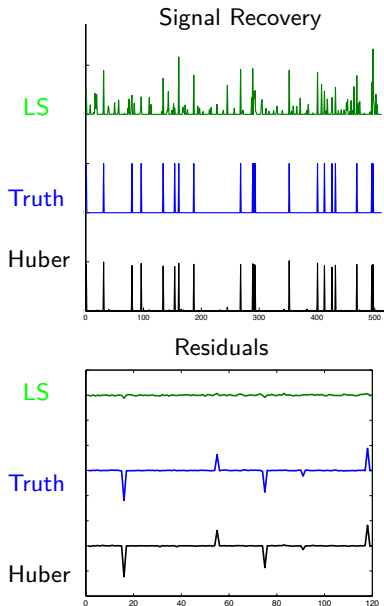
$$\text{HBP}_{\sigma}: \min_{0 \leq x} \|x\|_1 \quad \text{st} \quad \rho(b - Ax) \leq \sigma$$

Problem Specification

- x 20-sparse spike train in \mathbb{R}_+^{512}
- b measurements in \mathbb{R}^{120}
- A Measurement matrix satisfying RIP
- ρ Huber function
- σ error level set at .01
- 5 outliers

Results

In the presence of outliers, the robust formulation recovers the spike train, while the Huber standard formulation does not.



Least-Squares, Huber, and Student's t

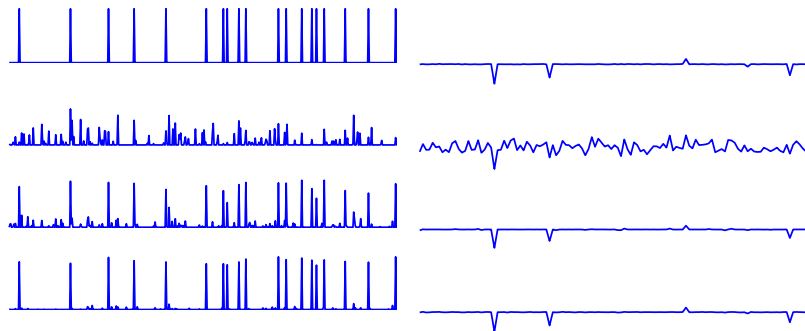


Figure : Left, top to bottom: True signal, and reconstructions via least-squares, Huber, and Student's t. Right, top to bottom: true errors, and least-squares, Huber, and Student's t residuals.

Go to the course homework webpage for the Huber-Laplace-Student t comparison for non RIP data.

Go to the course homework webpage for the Huber-Laplace-Student t comparison for non RIP data.

Thank you Terry for being my
adventure guide both in mathematics
and the wilderness!