CAIMS Annual Meeting

21 May 2007

In Pursuit of a Root

Ewout van den Berg **Michael P. Friedlander**
UBC Computer Science

[Felix Herrmann (UBC), Ozgur Yılmaz (UBC), Michael Saunders (Stanford)]

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2 \le \sigma \end{array}$$

Generic Regularization

$$A \quad x \approx b$$

Infinitely many solutions

- Encourage stability of solution
- Choose a solution with favorable properties — eg, **sparsity**

Regularization approaches

1. minimize $\|x\|$ subj to $\|Ax - b\| \leq \sigma$
2. minimize $\|Ax - b\|$ subj to $\|x\| \leq \lambda$
3. minimize $\|Ax - b\| + \tau\|x\|$

Sparse Solutions

$$\|x\|_0 := \textbf{sparsity} := \text{number of nonzeros in } x$$

The **sparsest** solution is given by

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_{\mathbf{0}} \\ \text{subject to} & \|Ax - b\|_2 \le \sigma \end{array}$$

Drawbacks

- Combinatorial problem: NP hard
- Tractable for only trivial problems

Alternative

Replace $\|x\|_{\mathbf{0}}$ with $\|x\|_{\mathbf{1}}$

[Chen et al'98], [Donoho, Elad '03], [Candés et al '04,'05], [Donoho et al, '05], [Donoho, Tanner, '05], [Fuchs '04,'05] [Tropp '04,'06], [many more!]

Basis Pursuit

Sparse Solutions via $\|\mathbf{x}\|_1$

$$A \quad x \approx b$$

Basis pursuit [Chen, Donoho, Saunders '98]

(BP) minimize $\|x\|_1$ subject to $Ax = b$

Basis pursuit denoise

(BPDN) minimize $\|x\|_1$ subject to $\|Ax - b\|_2 \leq \sigma$

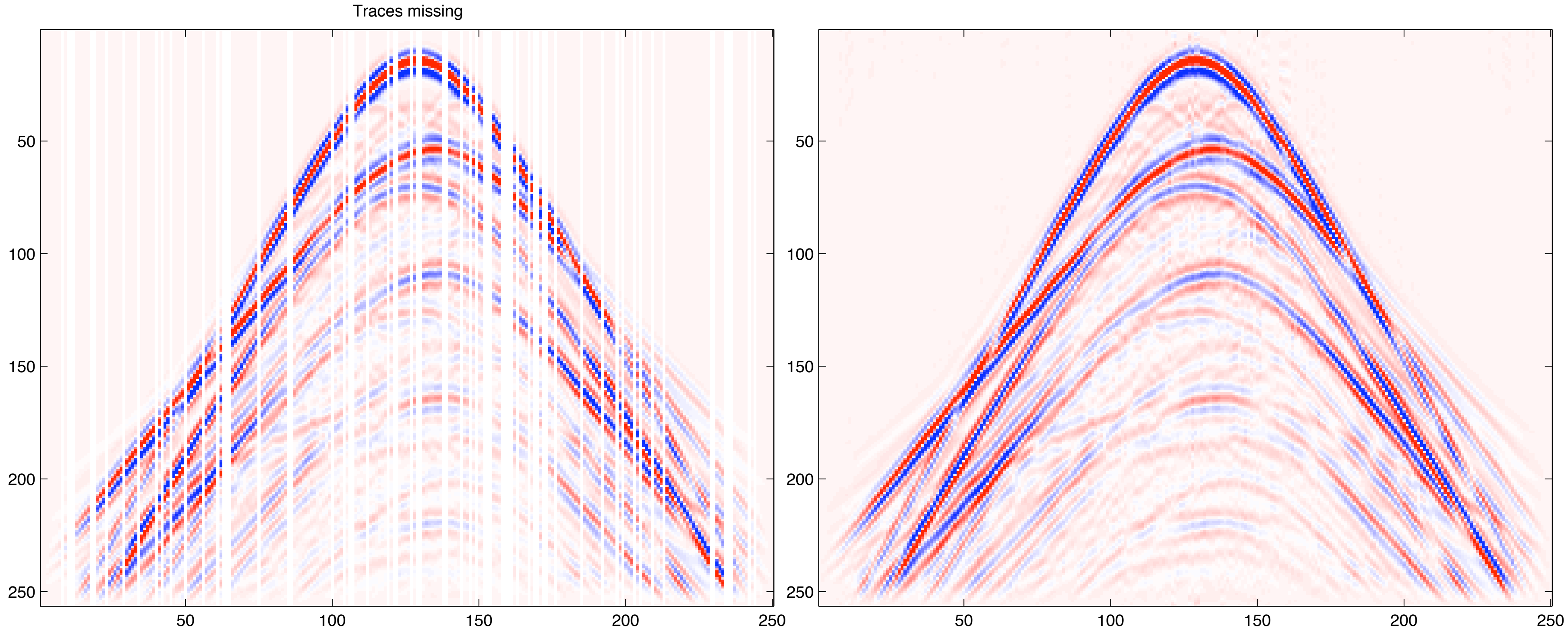
Lasso [Tibshirani '96]

(LS_λ) minimize $\|Ax - b\|_2$ subject to $\|x\|_1 \leq \lambda$

GOAL

Seismic Imaging

Felix Herrmann (UBC Earth & Ocean Sciences)



Large-Scale BPDN Solver

$$\underset{x}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \sigma$$

Solver requirements

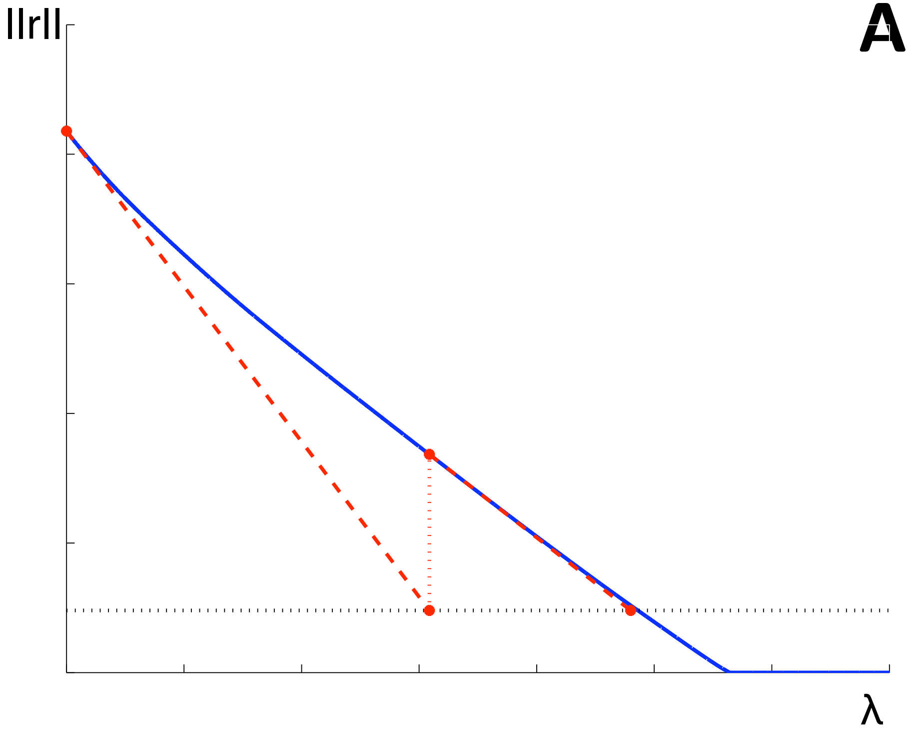
- Use A only as operator — Ax and $A^T y$
- Frugal with mat-vec projects — $\sim$ days CPU time
- "warm" starts — take advantage of $x_0 \approx x_\lambda$

Current approaches

- Log-barrier for SOCP — eg, MOSEK, SeDuMi
- QP formulation:

  minimize $\frac{1}{2}\|Ax - b\|_2^2 + \tau\|x\|_1$ — eg, PDCO, GPSR, Homotopy

Approach



$$\text{(LS}_\lambda\text{)} \quad \underset{x}{\text{minimize}} \quad \|Ax - b\|_2 \quad \text{subject to} \quad \|x\|_1 \le \lambda$$

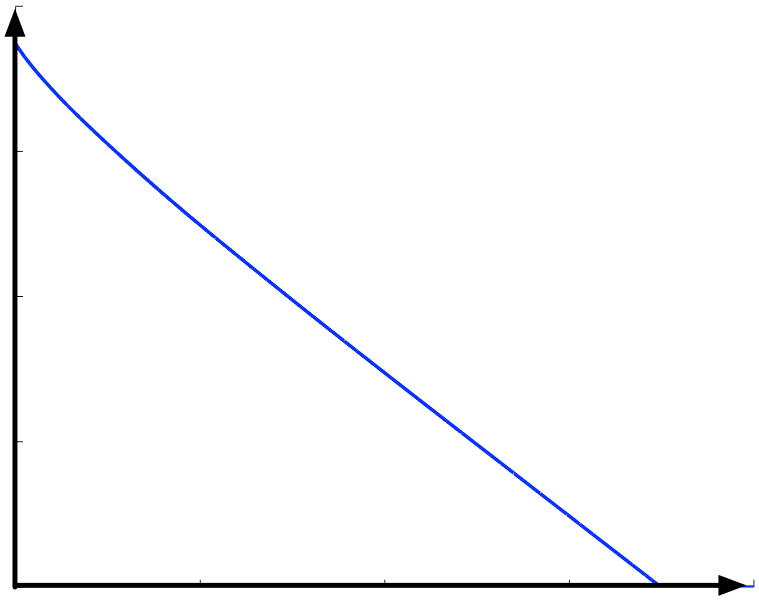
Value function

$$\psi(\lambda) = \|Ax_\lambda - b\|_2$$

Algorithm

1. Evaluate $\psi(\lambda)$ — gradient projection
2. Compute $\psi'(\lambda)$ — duality
3. Apply 1-dim Newton's method $$\lambda_{k+1} \leftarrow \lambda_k - \frac{\psi(\lambda_k) - \sigma}{\psi'(\lambda_k)}$$

Pareto Frontier



Lasso and its Dual

$$\psi(\lambda) := \|Ax_\lambda - b\|_2 = \|r_\lambda\|_2$$

Primal

$$(\mathrm{LS}_\lambda) \quad \underset{r,x}{\text{minimize}} \quad \|r\|_2 \quad \text{s.t.} \quad \mathbf{Ax} + \mathbf{r} = \mathbf{b}, \quad \|\mathbf{x}\|_1 \le \boldsymbol{\lambda}$$

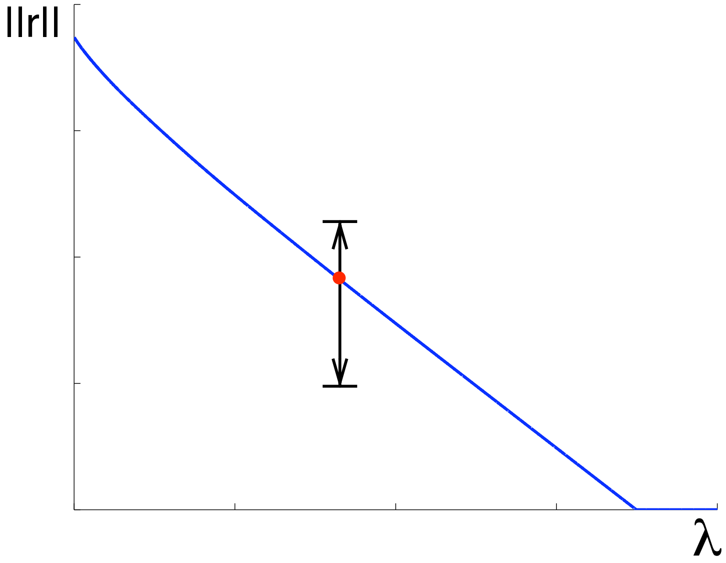
Dual

$$(\mathrm{LS}^D_\lambda) \quad \underset{y,\mu}{\text{maximize}} \quad b^T\mathbf{y} - \boldsymbol{\mu}\lambda \quad \text{s.t.} \quad \|\mathbf{y}\|_2 \le 1, \quad \|A^T\mathbf{y}\|_\infty \le \boldsymbol{\mu}$$

Solutions for $\lambda \in (0, \lambda_{\mathrm{BP}})$

$$\|x_\lambda\|_1 = \lambda \qquad \mathbf{y}_\lambda = \frac{r_\lambda}{\|r_\lambda\|_2}$$

$$\psi(\lambda) = \|r_\lambda\|_2 \qquad \boldsymbol{\mu}_\lambda = \|A^T y_\lambda\|_\infty$$

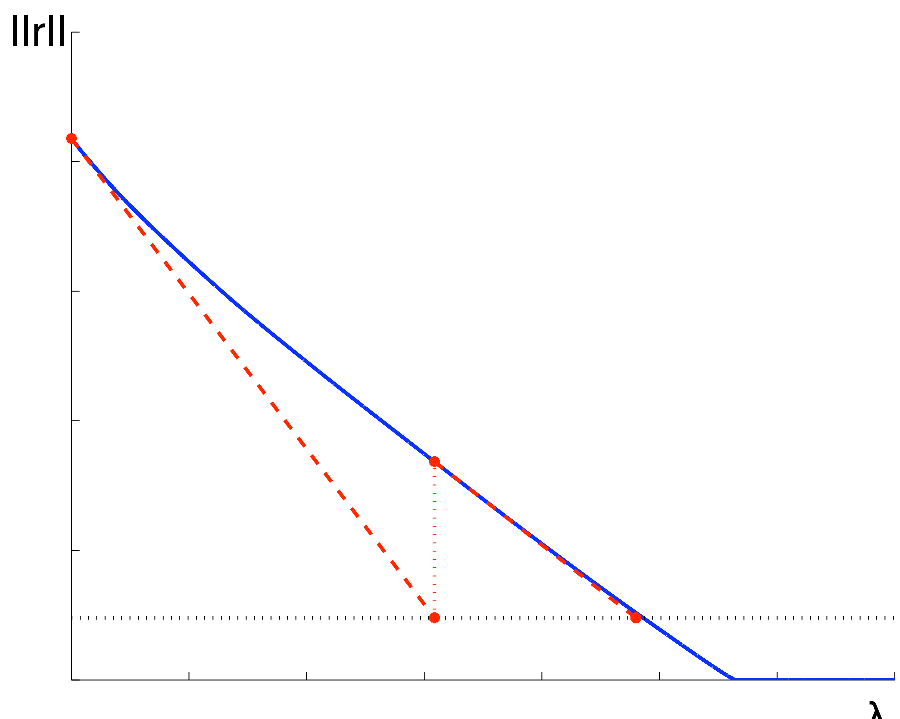


Useful Properties

$$\psi(\lambda) := \quad \text{optimal value of} \quad \underset{x}{\text{minimize}} \ \|Ax - b\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda$$

Theorem

1. ψ is convex and nonincreasing
2. For all $\lambda \in (0, \lambda_{\mathrm{BP}})$
   - ψ is **continuously differentiable**
   - $\psi'(\lambda) = -\mu_\lambda \quad$ where $\quad \mu_\lambda = \|A^T y_\lambda\|_\infty, \quad y_\lambda = r_\lambda / \|r_\lambda\|_2$



Spectral Projected Gradient

Projected Gradient

Evaluate $\psi(\lambda) \Longrightarrow$ $\boxed{\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \mathcal{C}}$

Projected gradient path

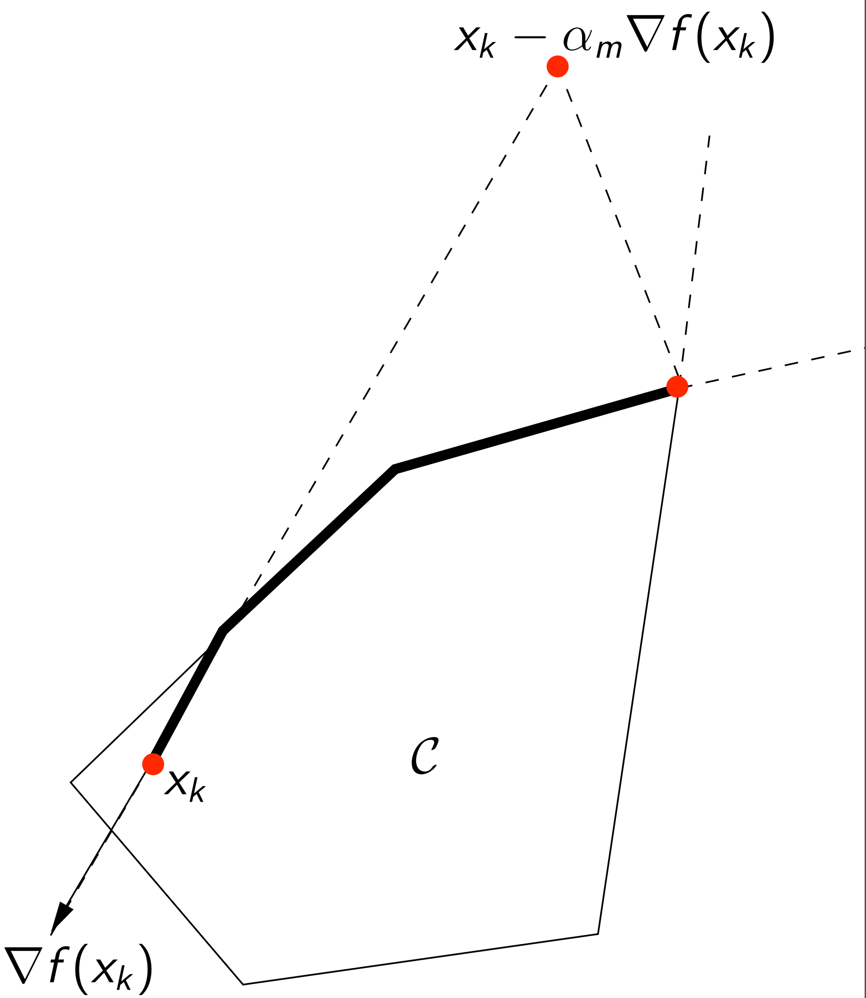
- $x_k(\alpha) = P[x_k + \alpha d], \qquad d = -g_k$

Each iteration

- Project steepest descent onto $\mathcal{C}$
- Minimize along piecewise linear path

Properties

- Large changes to active set possible
- $x_k \to x^*$, but slow (steepest descent)



Spectral Step Length

$$\text{Reduce} \quad f(x) \quad \text{along} \quad x_k(\alpha) = P[x_k + \alpha d]$$

Classical gradient projection

- $d = -g_k$
- Find $\alpha > 0$ to **minimize** along $x_k(\alpha)$ (ie, the Cauchy point)

Spectral step length

- $H_k d = -g_k \quad \text{with} \quad H_k := \gamma_k I \approx \nabla^2 f(x)$ (**!!**)
- Find $\alpha \in (0,1]$ to sufficiently reduce f along $x_k(\alpha)$
- Secant equation:

$$H_{k+1}s_k = y_k \qquad s_k := x_{k+1} - x_k$$

$$\gamma_{k+1}s_k = y_k \qquad y_k := g_{k+1} - g_k$$

$$\gamma_{k+1} := (y_k^T s_k)/(s_k^T s_k)$$

[Barzilai/Borwein '88], [Birgin,...'00], [Dai/Fletcher '05], [Figueiredo,...'07]

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \iff \boxed{\underset{x}{\text{minimize}} \ \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \iff \boxed{\underset{x}{\text{minimize}} \;\; \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0
28
-4
29
-38
21
120

- $n = 5,\ \Delta x_i = 20$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \iff \boxed{\underset{x}{\text{minimize}} \;\; \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$
28	8
-4	16
29	9
-38	-18
21	1
120	52

- $n = 5, \; \Delta x_i = 20$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \quad \Longleftrightarrow \quad \boxed{\underset{x}{\text{minimize}} \ \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$		x_1
28	8		28
-4	16		
29	9	$\Rightarrow$	29
-38	-18		-38
21	1		21
120	52		116

- $n = 5,\ \Delta x_i = 20$
- $n = 4,\ \Delta x_i = 24$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \iff \boxed{\underset{x}{\text{minimize}} \;\; \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$		x_1	$\bar{x}_1$
28	8		28	4
-4	16			
29	9	$\Rightarrow$	29	5
-38	-18		-38	-14
21	1		21	-3
120	52		116	26

- $n = 5,\ \Delta x_i = 20$
- $n = 4,\ \Delta x_i = 24$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \quad \Longleftrightarrow \quad \boxed{\underset{x}{\text{minimize}} \ \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$		x_1	$\bar{x}_1$		x_2
28	8		28	4		28
-4	16					
29	9	$\Rightarrow$	29	5	$\Rightarrow$	29
-38	-18		-38	-14		-38
21	1		21	-3		
120	52		116	26		95

- $n = 5,\ \Delta x_i = 20$
- $n = 4,\ \Delta x_i = 24$
- $n = 3,\ \Delta x_i = 25$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \iff \boxed{\underset{x}{\text{minimize}} \ \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$		x_1	$\bar{x}_1$		x_2	x^*
28	8		28	4		28	3
-4	16						
29	9	$\Rightarrow$	29	5	$\Rightarrow$	29	4
-38	-18		-38	-14		-38	-13
21	1		21	-3			
120	52		116	26		95	20

- $n = 5,\ \Delta x_i = 20$
- $n = 4,\ \Delta x_i = 24$
- $n = 3,\ \Delta x_i = 25$

Orthogonal Projection onto One-Norm Ball

$$P[x_0] \quad \Longleftrightarrow \quad \boxed{\underset{x}{\text{minimize}} \ \|x_0 - x\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \lambda}$$

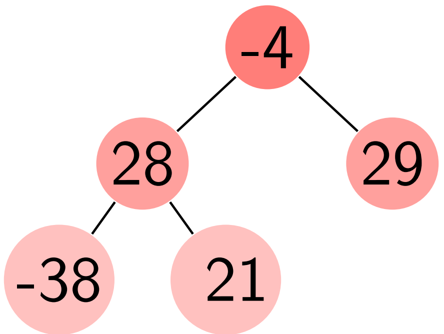
Stages

- Reduce all components x_i equally by $\Delta x := \|x\|_\lambda - \lambda$
- Do not let components x_i change sign

Example: $\lambda = 20$

x_0	$\bar{x}_0$		x_1	$\bar{x}_1$		x_2	x^*
28	8		28	4		28	3
-4	16						
29	9	$\Rightarrow$	29	5	$\Rightarrow$	29	4
-38	-18		-38	-14		-38	-13
21	1		21	-3			
120	52		116	26		95	20

- $n = 5, \ \Delta x_i = 20$
- $n = 4, \ \Delta x_i = 24$
- $n = 3, \ \Delta x_i = 25$



Algorithm cost

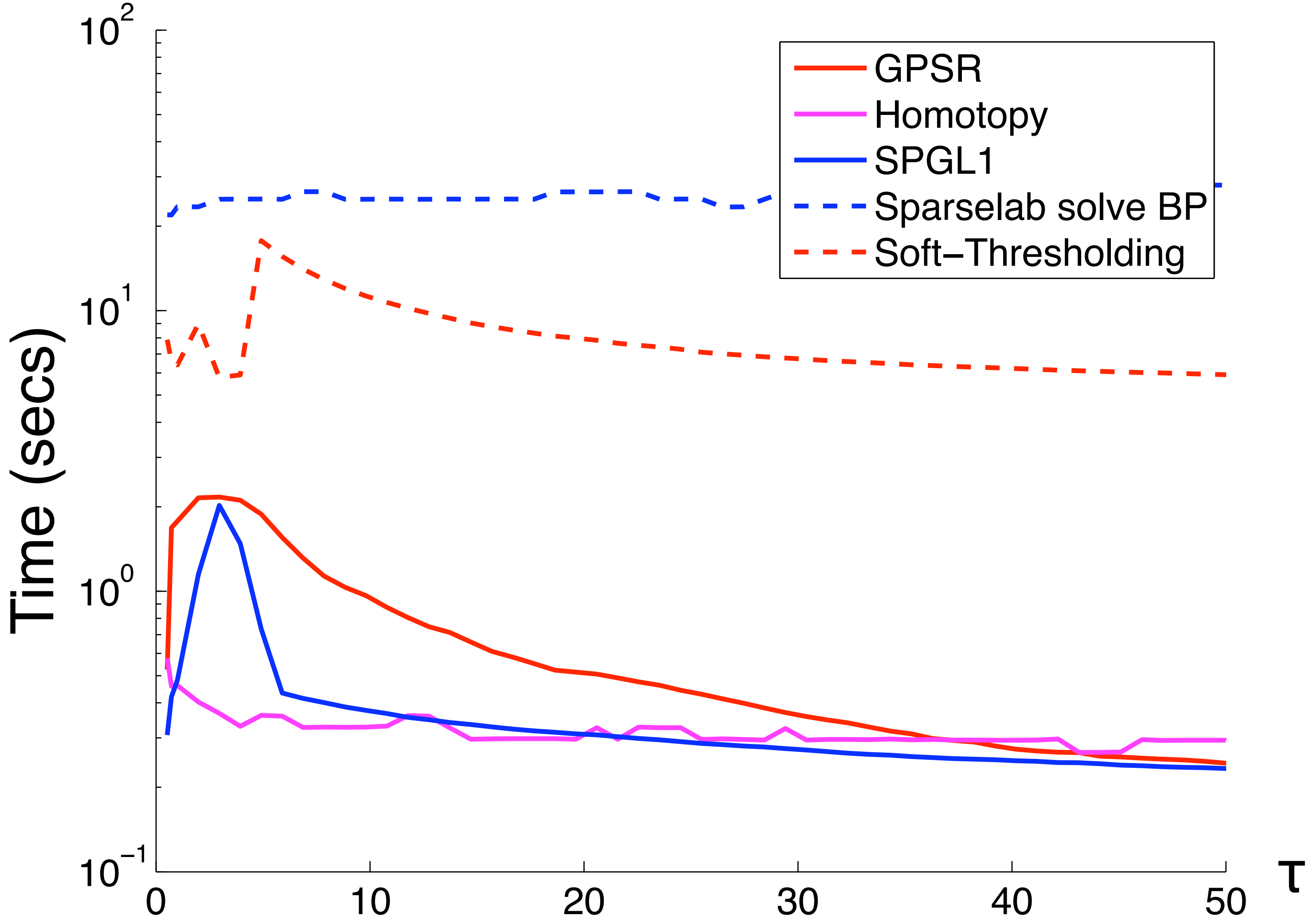
- Maintain elements in min-abs-val heap
- $\mathcal{O}(n \log n)$ iterations

Numerical Experiments

Lasso vs. Weighted Least-Squares

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|Ax - b\|_2 \\ \text{subject to} & \|x\|_1 \le \lambda \end{array} \quad \text{vs.} \quad \underset{x}{\text{minimize}} \quad \|Ax - b\|_2 + \tau\|x\|_1$$

A is 512-by-4069



Basis Pursuit De-noise

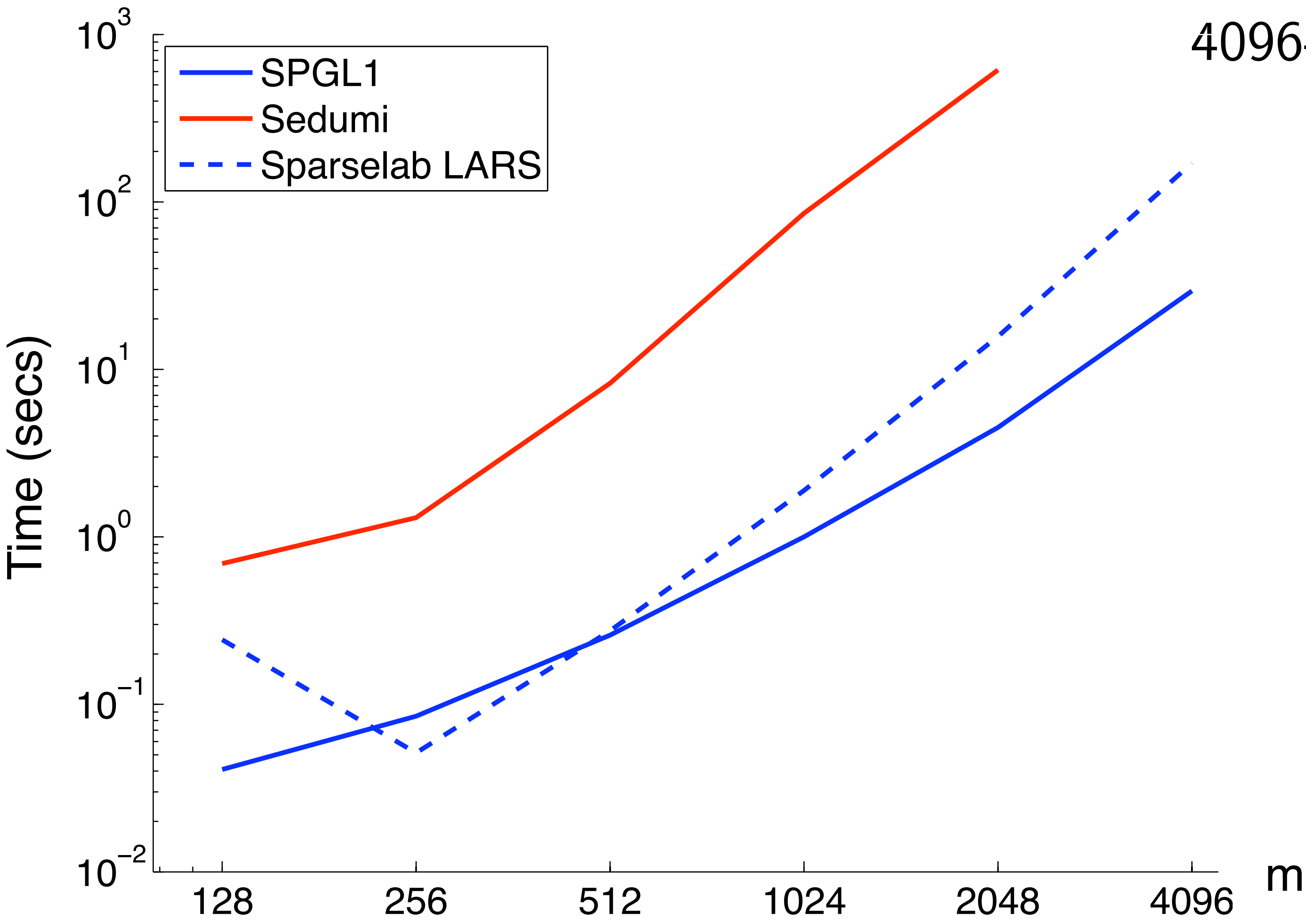
$$\text{minimize} \quad \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_2 \le \sigma$$

Increasing sizes, $n = 4m$:

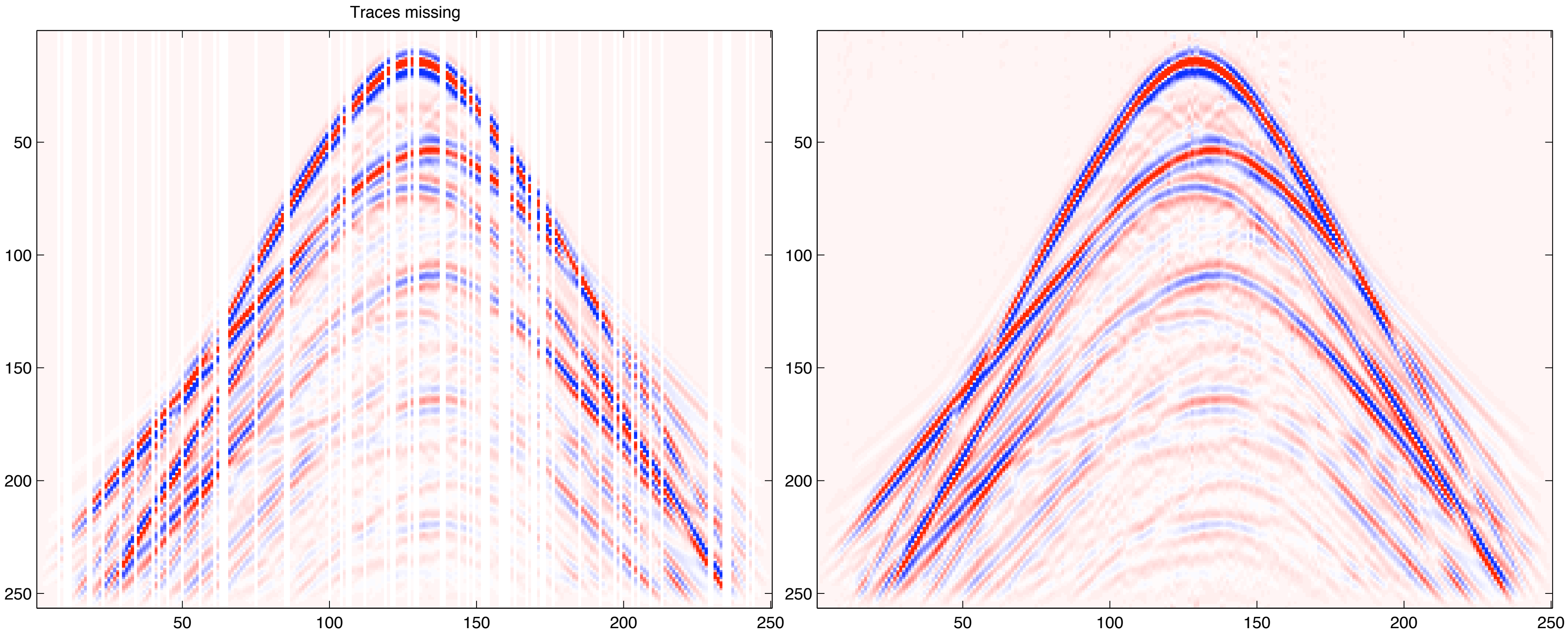
128-by-512

⋮

4096-by-16384



Seismic Imaging
Felix Herrmann (UBC Earth & Ocean Sciences)



- 30% missing traces
- $n = 481K$
- $\mathrm{nnz}(x^*) \approx 70K$ (15%)

- # mat-vec prod's Ax — 913
- # mat-vec prod's $A^T y$ — 585
- Total time for Ax or $A^T y$ — 622 secs
- Total time for $P[\cdot]$ — 60 secs

Thanks!

Looking Ahead

Accelerate Lasso subproblem

- Newton-type method for subproblem

$$(\text{LS}_\lambda) \quad \underset{x}{\text{minimize}} \quad \|Ax - b\|_2 \quad \text{subject to} \quad \|x\|_1 \le \lambda$$

Better root-finding

- Reformulation of Pareto curve is piecewise quadratic
- Can we use higher-order Newton approximations?