SINBAD 2008

February 20–22, 2008

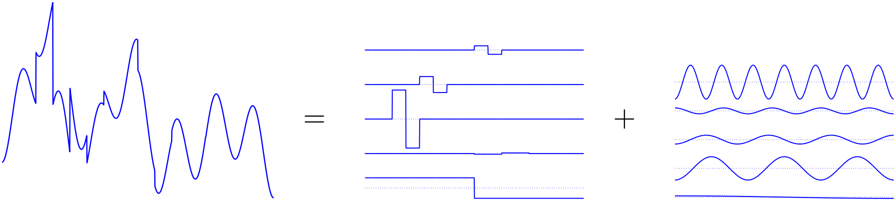
Algorithms for Large-Scale Sparse Reconstruction

Michael P. Friedlander
UBC Computer Science

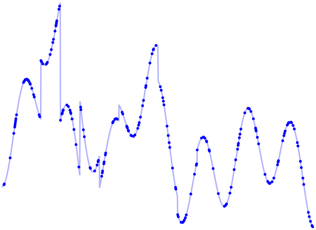
Collaborators:
Ewout van den Berg and Michael Saunders

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

Example 1: Missing Observations



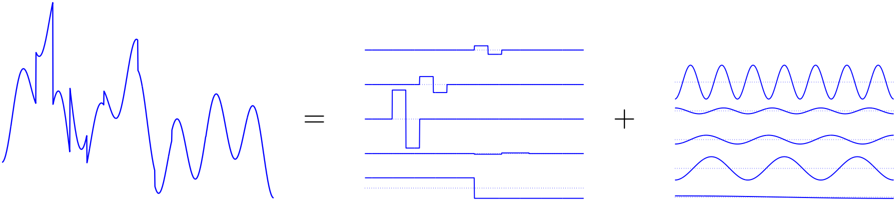
$\mathbf{R}b =$



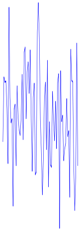
$\mathbf{R}b =$ [Restriction]

$\mathbf{R}b \approx \mathbf{R}$ [Haar | DCT] x

Example 2: Compressed Sensing



$\langle \phi_1, b \rangle, \ldots, \langle \phi_m, b \rangle$

$\boldsymbol{\Phi} b =$  $\boldsymbol{\Phi} =$ [Gaussian]

$$\boldsymbol{\Phi} b \approx \boldsymbol{\Phi} \begin{bmatrix} \text{Haar} & \text{DCT} \end{bmatrix} x$$

SPARSE SOLUTIONS

Sparse Solutions

$$\boxed{\quad A \quad} \quad x \approx b$$

Find the **sparsest** solution:

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

Drawbacks

- Combinatorial problem
- Tractable for only trivial problems

Alternative

Replace "nonzeros(x)" with $\|x\|_1$

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

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

$$A \quad x \approx b$$

Three regularization approaches, three parameters

$$\text{BP}_\sigma: \quad \text{minimize} \quad \|x\|_1 \quad \text{subj to} \quad \|Ax - b\|_2 \le \boldsymbol{\sigma}$$

$$\text{LS}_\tau: \quad \text{minimize} \quad \|Ax - b\|_2^2 \quad \text{subj to} \quad \|x\|_1 \le \boldsymbol{\tau}$$

$$\text{QP}_\lambda: \quad \text{minimize} \quad \|Ax - b\|_2^2 + \boldsymbol{\lambda}\|x\|_1$$

BP_σ := Basis pursuit denoise (BP_0 := basis pursuit) [Chen,...'98]

LS_τ := Lasso [Tibshirani '96]

QP_λ := Quadratic program [many!]

GOAL

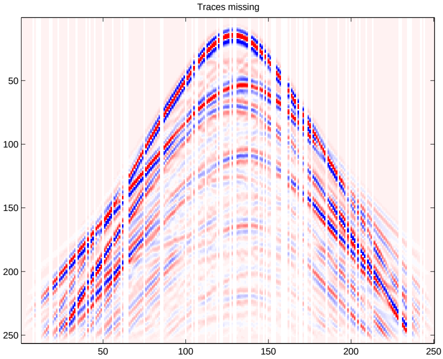
Seismic Imaging

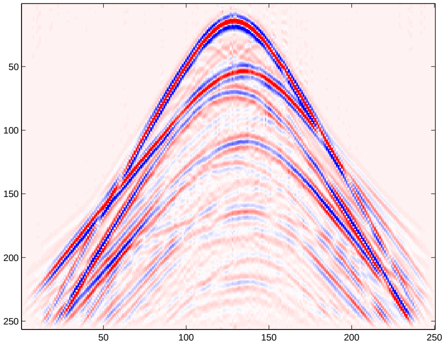
Recover missing traces via

$$\text{minimize} \ \|x\|_1 \quad \text{subj to} \quad \|RC^H x - b\|_2 \leq \sigma$$

Smallest 2D images: length(x) $\approx$ 1/2 million

Target 3D images: length(x) $\approx$ hundreds of millions





Large-Scale BPDN Solver

$$\text{BP}_\sigma : \quad \text{minimize} \quad \|x\|_1 \quad \text{subj to} \quad \|Ax - b\|_2 \le \sigma$$

Solver requirements

- Use A only as operator — Ax and $A^T y$
- Very frugal with mat-vec products — $\sim$ days CPU time
- "warm" starts — take advantage of $x_0 \approx x^*$

Approaches

BP_σ : minimize $\|x\|_1$ subj to $\|Ax - b\|_2 \le \boldsymbol{\sigma}$

- Log-barrier — ℓ_1-Magic [Candés & Romberg '05]
- Active-set — Homotopy [Osborne...00,Malioutov...05]
- Root-finding — SPGL1 [vandenBerg & F. '07]

QP_λ : minimize $\|Ax - b\|_2^2 + \boldsymbol{\lambda}\|x\|_1$

- Log-barrier — PDCO, ℓ_1-Magic, L1_LS
- Soft thresholding — FPC [Hale et al '07, Daubechies et al '04]
- Projected gradient — GPSR [Figuereido et al '07]
- Orthogonal blocks — BCR [Sarty et al '00]

LS_τ : minimize $\|Ax - b\|_2^2$ subj to $\|x\|_1 \le \boldsymbol{\tau}$

- Greedy — LARS [Efron...'04]
- Projected gradient — [Daubechies et al '07], SPGL1

Sparse $Ax \approx b$

- Greedy — OMP [Pati...'93,Davis...'97,Tropp...'07]
- Greedy++ — StOMP [Donoho...'06]

Outline

A dusty attic treasure

- active-set for quadratic programming
- improves on greedy approaches

Probing the Pareto frontier

- root finding
- large-scale basis pursuit

Orthogonal Matching Pursuit

Greedy approach to sparse $Ax = b$

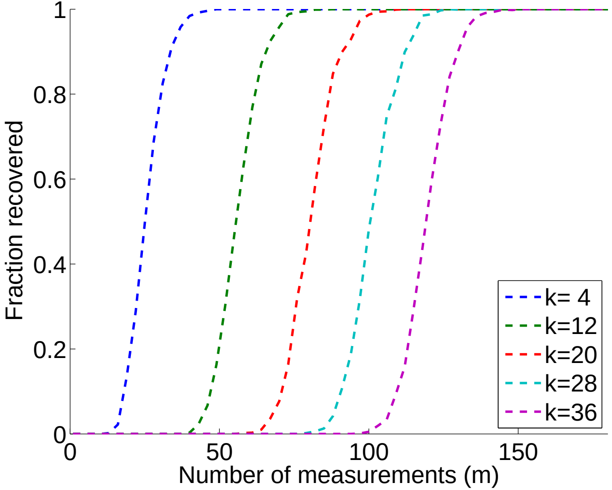
0.	Initialize	$r \leftarrow b, \quad B \leftarrow [\ ]$
1.	Largest correlation	find j st $\lvert a_j^T r \rvert = \lVert A^T r \rVert_\infty$
2.	Add column	$B \leftarrow [\ B \quad a_j\ ]$
3.	Least squares	min $\lVert b - Bx \rVert_2$
4.	Update residual	$r \leftarrow b - Bx$

- Works **most of the time** [Tropp & Gilbert '07]
- Considered a **cheaper** alternative to $\underset{x}{\text{minimize}}\ \lVert Ax - b \rVert^2 + \lambda \lVert x \rVert_1$

Recovery Rates for OMP

$Ax = b$ where

- A is Gaussian $m \times 256$
- x has k nonzeros



Quadratic Programming

Basis pursuit approach to sparse $Ax = b$

$$\text{QPp}_\lambda: \quad \underset{r,x}{\text{minimize}} \quad \tfrac{1}{2}\|r\|_2^2 + \lambda\|x\|_1 \quad \text{st} \quad r = b - Ax$$

$$\text{QPd}_\lambda: \quad \underset{r}{\text{maximize}} \quad b^T r - \tfrac{1}{2}\|r\|_2^2 \quad \text{st} \quad \|A^T r\|_\infty \le \lambda$$

Dual is vanilla QP:

- interior-point — polynomial complexity
- active-set — exponential complexity, better in practice

Active-set for Dual

$$\underset{r}{\text{maximize}} \quad b^T r - \tfrac{1}{2}\|r\|_2^2 \quad \text{st} \quad -\lambda e \le A^T r \le \lambda e$$

Maintain partition of active & inactive constraints:

$$A = [\ B \quad N\ ], \qquad B^T r_k = \pm\lambda e, \qquad -\lambda e < N^T r_k < \lambda e$$

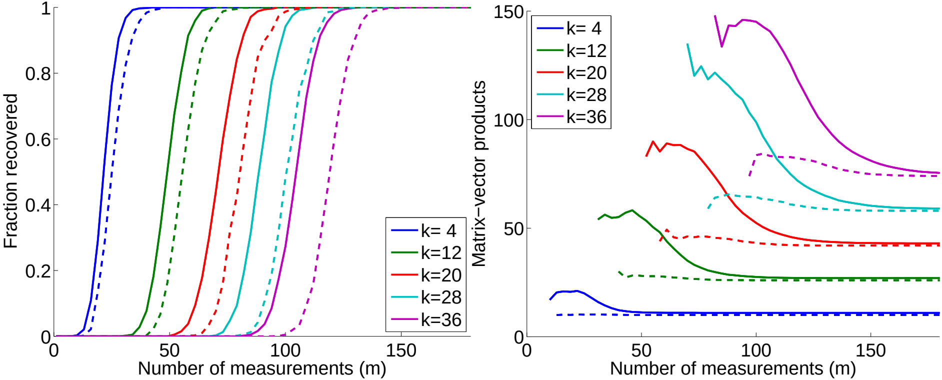
Main work per iteration:

1. Least squares $\quad \min\ \|g - Bx\|_2, \quad g = b - r$
2. Ascent direction $\quad \Delta r \leftarrow g - Bx$
3. Stay feasible $\quad -\lambda e \le A^T(r_k + \alpha\Delta r) \le \lambda e$
4. Add col to B $\quad B \leftarrow [\ B \quad a_j\ ]$

Main Work per Iteration

OMP		Active-set QP
1.	find j st $\lvert a_j^T r\rvert = \lVert \mathbf{A^T r} \rVert_\infty$	$\mathbf{\min \lVert g - Bx \rVert_2}$
2.	$\mathbf{B} \leftarrow [\ \mathbf{B} \quad \mathbf{a_j}\ ]$	$\Delta r \leftarrow g - Bx$
3.	$\mathbf{\min \lVert b - Bx \rVert_2}$	$-\lambda e \le \mathbf{A^T(r_k + \alpha \Delta r)} \le \lambda e$
4.	$r \leftarrow b - Bx$	$\mathbf{B} \leftarrow [\ \mathbf{B} \quad \mathbf{a_j}\ ]$

Active-set vs OMP



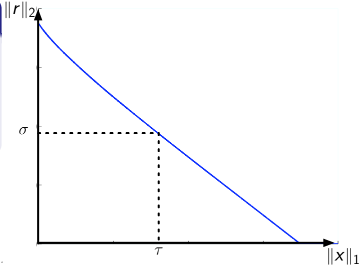
PROBING THE PARETO FRONTIER

Pareto Curve

Pareto curve

BP$_\sigma$: min $\|x\|_1$ st $\|Ax - b\|_2 \le \boldsymbol{\sigma}$

LS$_\tau$: min $\|Ax - b\|_2$ st $\|x\|_1 \le \boldsymbol{\tau}$



Value function

$$\phi(\tau) := \|Ax_\tau - b\|_2 = \|r_\tau\|_2$$

Algorithm

1. Evaluate $\phi(\tau)$ — projected gradient
2. Compute $\phi'(\tau)$ — duality
3. Root-finding on $\phi(\tau) = \sigma$ — Newton's method / Interpolation

Dual Problem

Lasso primal $\phi(\tau)$

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2}\|Ax-b\|_2^2 \\ \text{subject to} \quad & \|x\|_1 \le \boldsymbol{\tau} \end{aligned}$$

QP primal

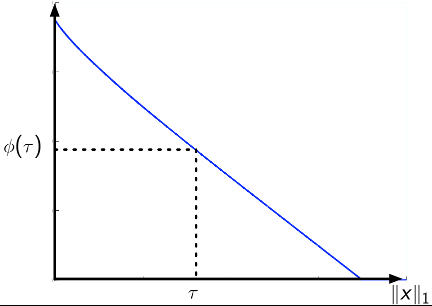
$$\underset{x}{\text{minimize}} \quad \tfrac{1}{2}\|Ax-b\|_2^2 + \boldsymbol{\lambda}\|x\|_1$$

Lasso dual

$$\begin{aligned} \underset{r,\boldsymbol{\lambda}}{\text{maximize}} \quad & b^T r - \tfrac{1}{2} r^T r - \boldsymbol{\tau\lambda} \\ \text{subject to} \quad & -\boldsymbol{\lambda} e \le A^T r \le \boldsymbol{\lambda} e \end{aligned}$$

QP dual

$$\begin{aligned} \underset{r}{\text{maximize}} \quad & b^T r - \tfrac{1}{2} r^T r \\ \text{subject to} \quad & -\boldsymbol{\lambda} e \le A^T r \le \boldsymbol{\lambda} e \end{aligned}$$



Pareto Curve: Useful Properties

$$\phi(\tau) \;:=\; \text{optimal value of} \quad \text{minimize} \;\; \tfrac{1}{2}\|Ax-b\|_2^2 \quad \text{subj to} \quad \|x\|_1 \le \tau$$

Theorem

1. ϕ is convex
2. For all $\tau \in (0, \tau_{\mathrm{BP}})$
   - ϕ is **continuously differentiable**
   - $\phi'(\tau) = -\lambda_\tau$ with $\lambda_\tau = \|A^T r_\tau\|_\infty$

Generic regularization

- Analogous result holds for the generic problem

$$\text{minimize} \quad \|Ax-b\|_{\mathbf{p}} \quad \text{subj to} \quad \|x\|_{\mathbf{q}} \le \tau \quad (1 \le p, q \le \infty)$$

- $\phi'(\tau) = -\lambda_\tau$ with $\lambda_\tau = \|A^T y_\tau\|_{\bar{\mathbf{p}}}, \quad y_\tau = r_\tau / \|r_\tau\|_{\bar{\mathbf{q}}}$

Root Finding: $\phi(\tau) = \sigma$

Approximately solve

minimize $\frac{1}{2}\|Ax - b\|_2^2$

subj to $\|x\|_1 \leq \tau_k$

Newton update

$$\tau_{k+1} \leftarrow \tau_k - (\phi_k - \sigma)/\phi'_k$$

Early termination

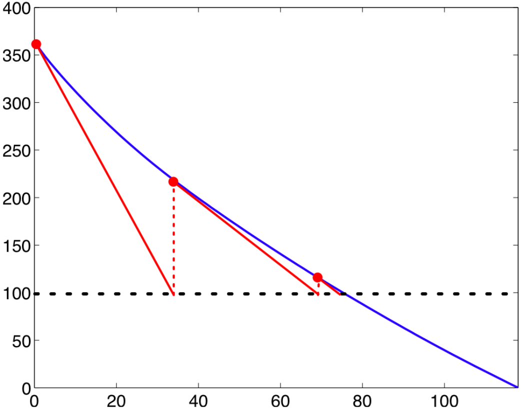
Use duality gap to monitor iterations

$$\delta_k := r^T r - b^T r + \tau_k \lambda$$

Convergence

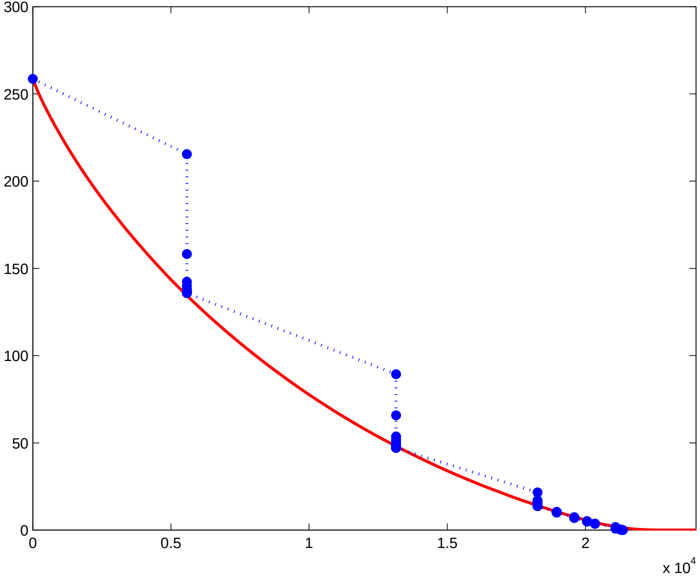
$$|\tau_{k+1} - \tau_\sigma| \leq \gamma\delta_k + \eta_k|\tau_k - \tau_\sigma|^2$$

with $\eta_k \to 0$



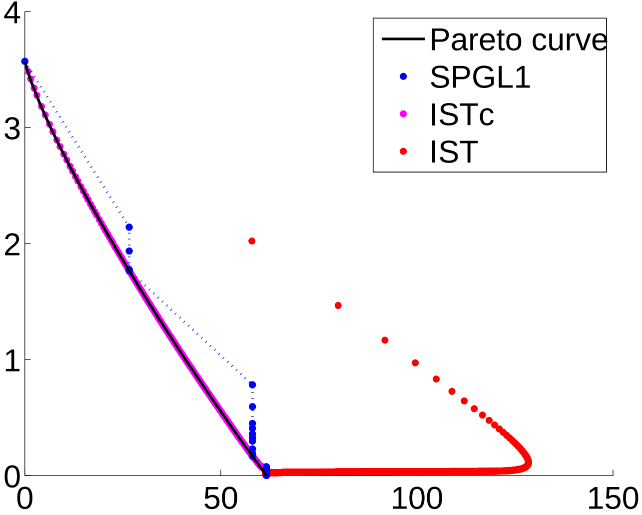
Root Finding for the Seismic Image

$$\text{minimize} \quad \|x\|_1 \quad \text{subj to} \quad \|RC^H x - b\|_2 \le \sigma \ (= 0)$$



Root Finding vs Usual Suspects

$$\text{minimize} \quad \|x\|_1 \quad \text{subj to} \quad \|Ax - b\|_2 \le \sigma \ (= 0)$$

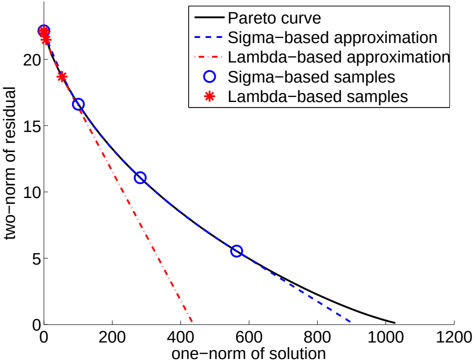


[Hennenfent, van den Berg, F., Herrmann '08]

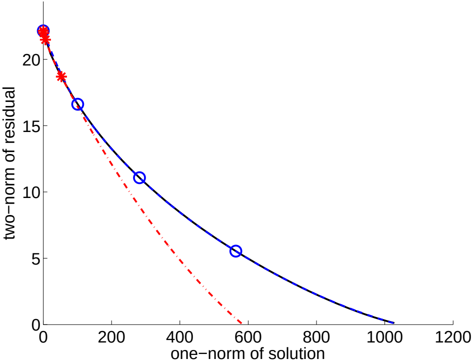
Inter/Extra-polating the Pareto Curve

Sparco problem: `srcsep1`

linear extrapolation



quadratic extrapolation



EVALUATING

ϕ and ϕ'

Projected Gradient

$$\text{Evaluate } \phi(\tau) \Longrightarrow \boxed{\text{minimize} \quad f(x) \quad \text{subj to} \quad x \in \mathcal{C}}$$

Projected gradient path

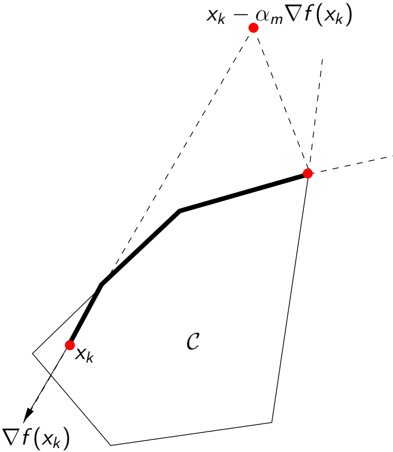
- $x_k(\alpha) = P[x_k + \alpha\Delta x], \qquad \Delta x = -g_k$

Each iteration

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

Properties

- $x_k \to x^*$
- Large changes to active set possible
- Finite active-set identification



Projection onto One-Norm Ball

$$P_\tau[c] \iff \boxed{\underset{x}{\text{minimize}} \ \|c - x\|_2 \quad \text{subj to} \quad \|x\|_1 \le \tau}$$

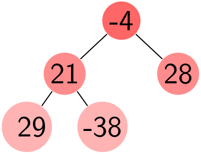
Stages

- Reduce all components c_i equally by $\Delta c := \|c\|_1 - \tau$
- Do not let components c_i change sign

Example: $\tau = 20$

c	x	$\Longrightarrow$	c_1	x_1	$\Longrightarrow$	c_2	x^*
28	8		28	4		28	3
29	9		29	5		29	4
-38	-18		-38	-14		-38	-13
21	1		21	-3			
-4	16						
120	52		116	26		95	20

- $n = 5,\ \Delta c = 20$
- $n = 4,\ \Delta c = 24$
- $n = 3,\ \Delta c = 25$
- Done!



Algorithm cost

- Maintain elements in min-abs-val heap
- $\mathcal{O}(\mathbf{n \log n})$ operations worst case

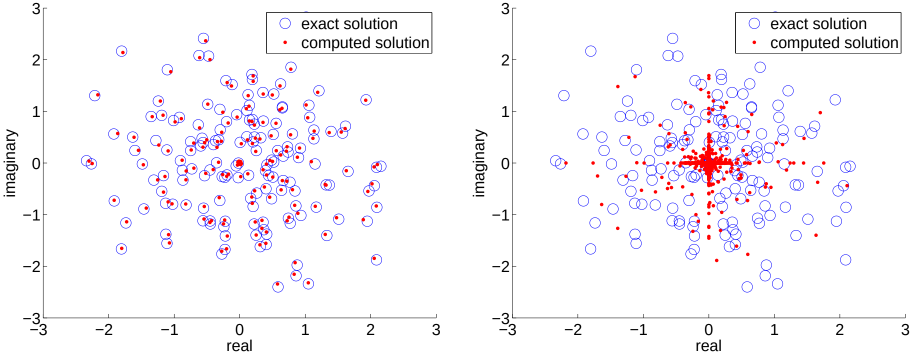
Projection onto Complex One-Norm Ball

$$\mathcal{P}_\tau[c] \quad \Longleftrightarrow \quad \boxed{\underset{z\in\mathbb{C}^n}{\text{minimize}} \;\; \|c-z\|_2 \quad \text{subj to} \quad \|z\|_1 \le \tau}$$

1. Compute vector of moduli: $r \leftarrow (\|c_1\|_2 \;\; \dots \;\; \|c_n\|_2)^T$
2. (Real) one-norm projection: $\bar{r} \leftarrow P_\tau[r]$
3. Scale components of c by $\bar{r}$: $\mathcal{P}_\tau[c] \leftarrow \begin{cases} c_i(\bar{r}_i/r_i) & \text{if } r_i > 0 \\ 0 & \text{otherwise} \end{cases}$

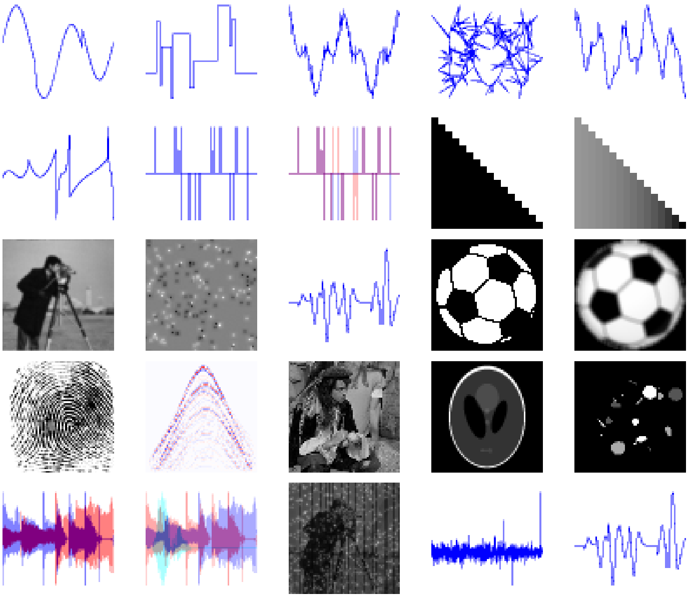
You Can't Fake It

$$\underset{x,y}{\text{minimize}} \quad \|x + iy\|_1 \qquad\qquad \underset{x,y}{\text{minimize}} \quad \|x\|_1 + \|y\|_1$$



$$\|z\|_1 = \sum_j \sqrt{x_j^2 + y_j^2} \leq \sum_j |x_j| + |y_j| = \|x\|_1 + \|y\|_1$$

SPARCO: Sparse Reconstruction Test Suite



LOOKING AHEAD

Looking Ahead

Nonlinear misfit measures

- minimize $\|x\|_1$ subj to $f(x) \leq \sigma$

Decomposition algorithms

- $A = [A_1 \; A_2 \; \ldots \; A_k]$ and $x = (x_1, x_2, \ldots, x_k)$
- Optimize over $x_1, \ldots, x_k$ seperately

Dantzig Selector

[Candes & Tao '07]

- $$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & \|A^T(Ax - b)\|_\infty \leq \lambda \end{array}$$

Thanks!

References

- E. van den Berg and M. P. Friedlander, *Probing the Pareto frontier for basis pursuit solutions*, UBC CS TR-2008-01, January 2008
- M. P. Friedlander and M. A. Saunders, *Discussion: The Dantzig selector: Statistical estimation when p is much larger than n* Annals of Statistics, 35(6):2385-2391, December 2007
- SPGL1: A solver for large-scale sparse reconstruction Available at `www.cs.ubc.ca/labs/scl/spgl1`
- SPARCO: A testing environment for sparse reconstruction Available at `www.cs.ubc.ca/labs/scl/sparco`

Research support

This presentation was carried out as part of the SINBAD project with financial support, secured through ITF, from the following organizations: BG, BP, Chevron, ExxonMobil, and Shell. SINBAD is part of the collaborative research & development (CRD) grant number 334810-05 funded by the Natural Science and Engineering Research Council (NSERC).