Sparse Representations in High Dimensional Geometry: What’s the Excitement?

Bradley K. Alpert

NIST

A Few Questions

- Why collect so much data from sensors, then immediately compress it?

A Few Questions

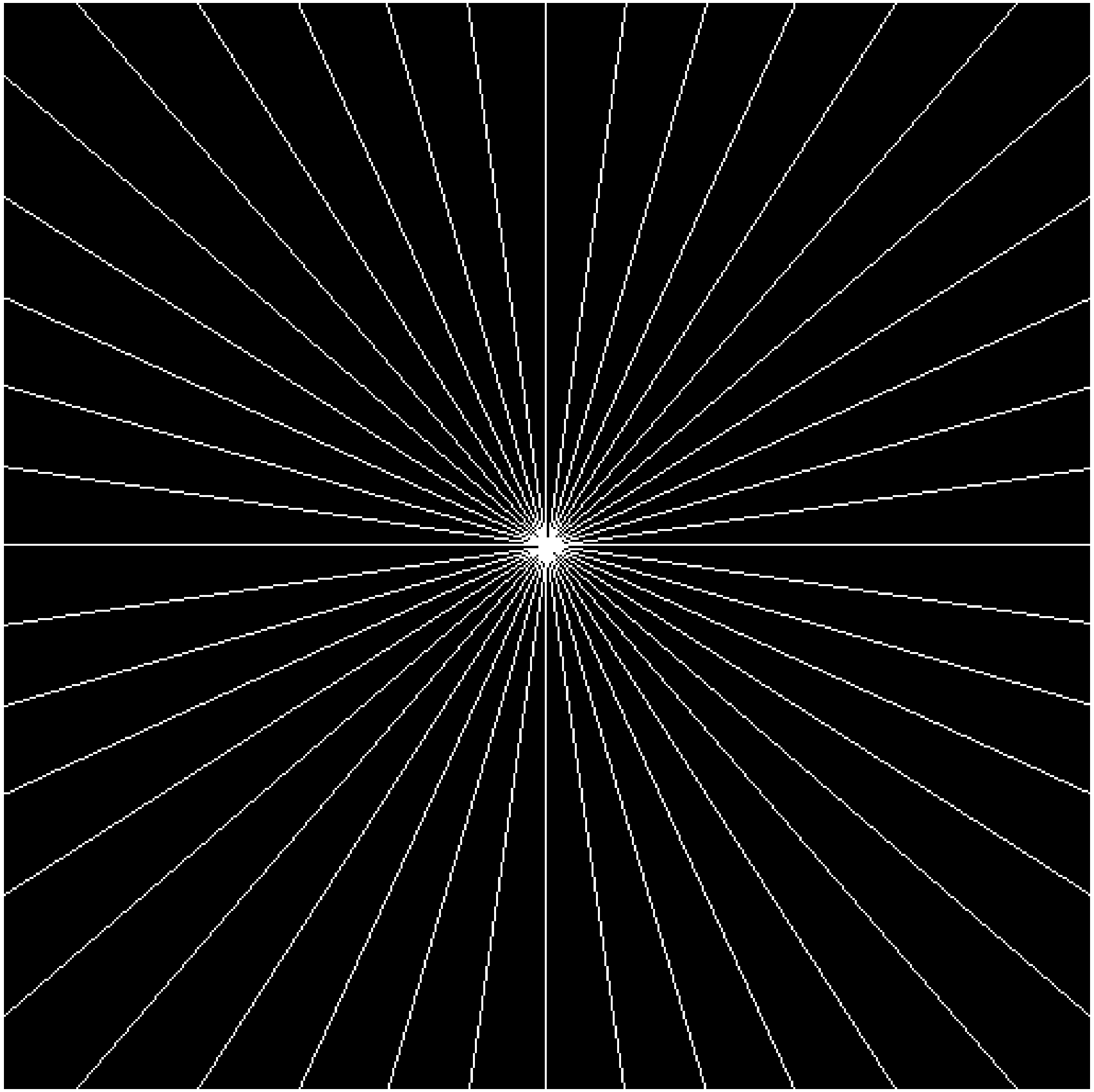
- Why collect so much data from sensors, then immediately compress it?
- Is full image reconstruction possible from incomplete Fourier data?

A Few Questions

- Why collect so much data from sensors, then immediately compress it?
- Is full image reconstruction possible from incomplete Fourier data?
- How can knowledge of sparsity be exploited?

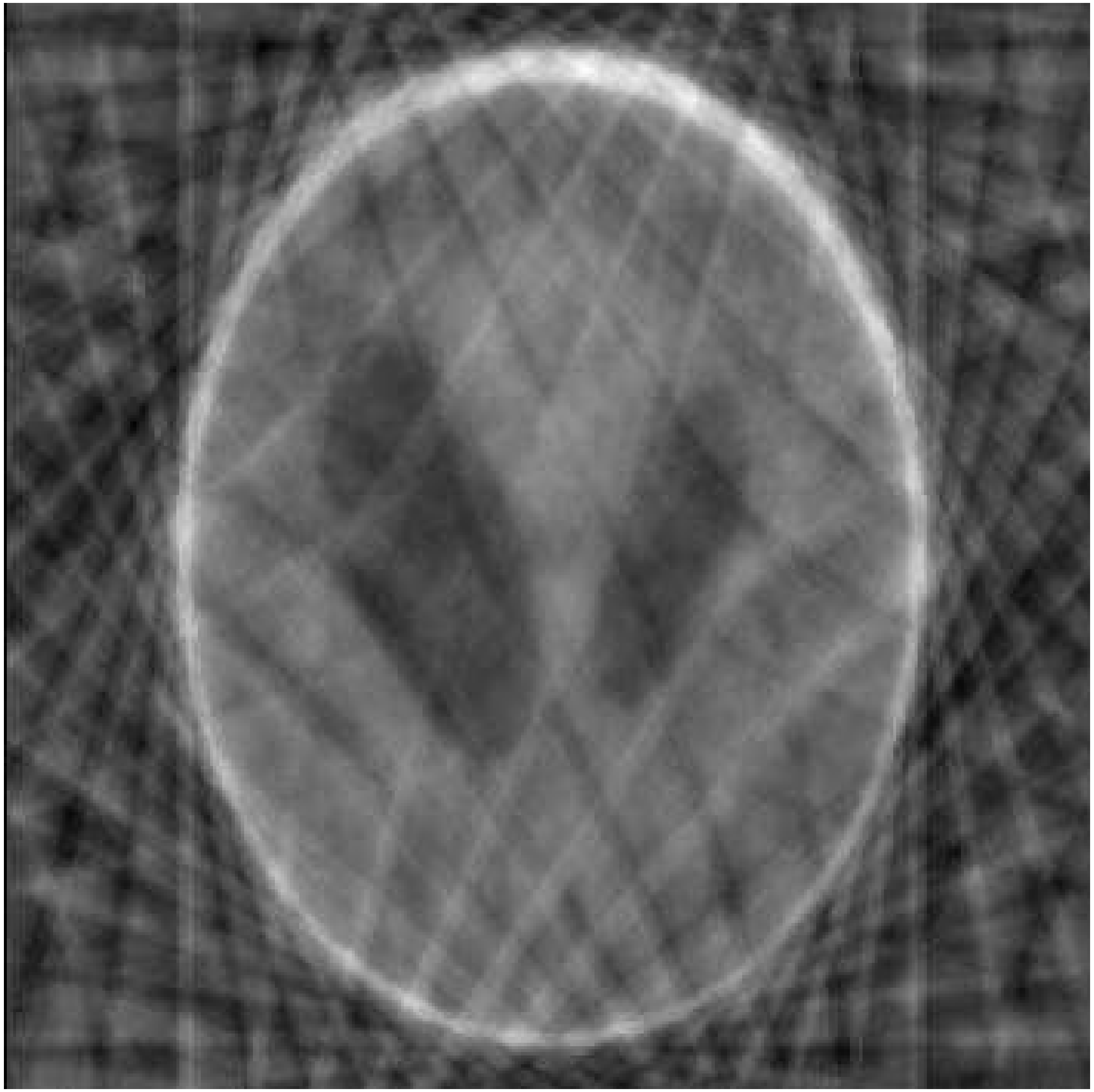
Shepp-Logan Phantom

Samples from Discrete FT



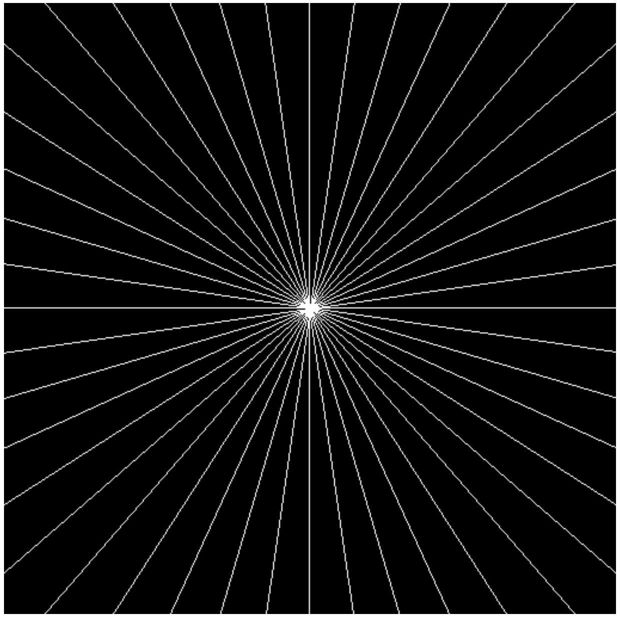
Candes, Romberg, Tao (2004)

Direct Reconstruction

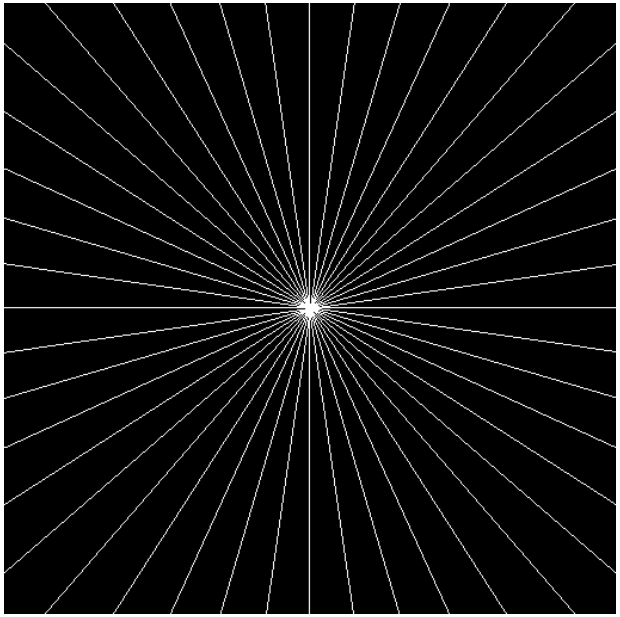


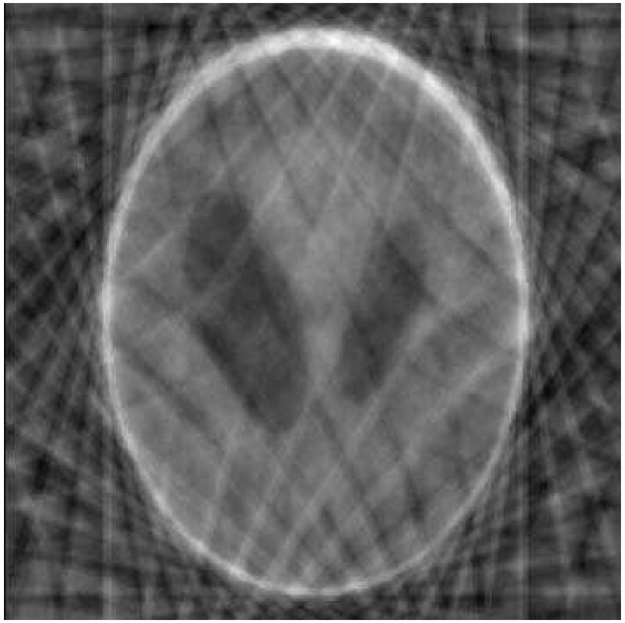
“Smart” Reconstruction Exact

“Smart” Reconstruction Exact

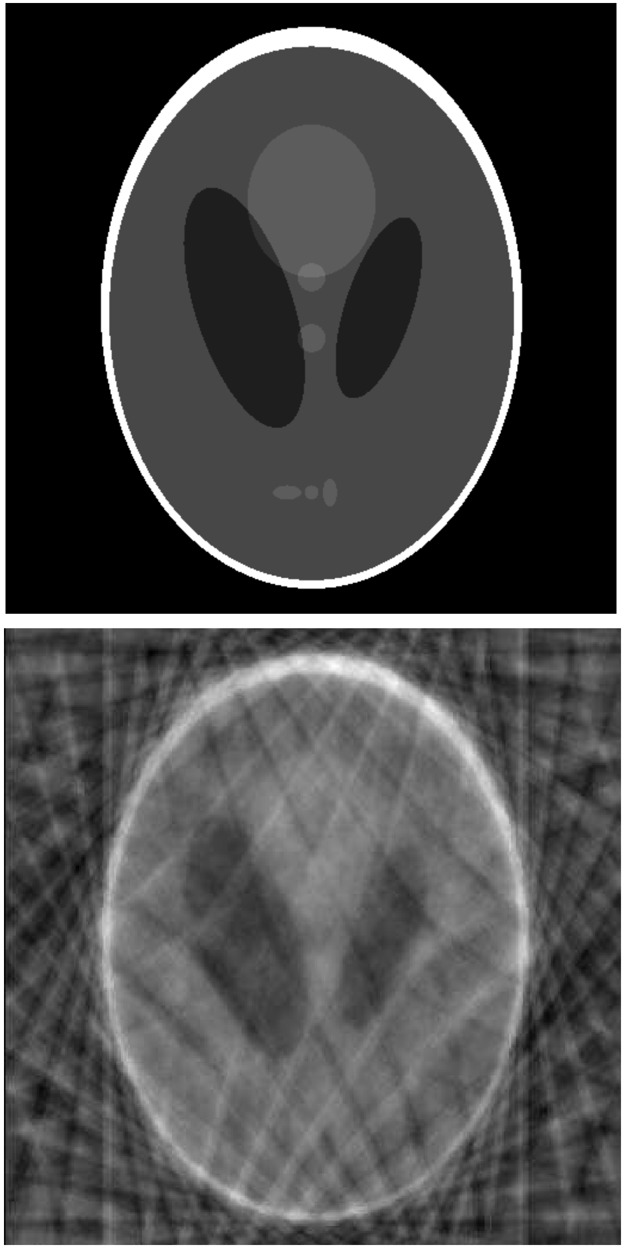


“Smart” Reconstruction Exact





“Smart” Reconstruction Exact



Smart Reconstruction?

Minimize

$$\int_{\mathrm{Image}} |\nabla f(\mathbf{x})| d\mathbf{x}$$

subject to

$$\hat{f}(\boldsymbol{\xi}) = \hat{f}_0(\boldsymbol{\xi}), \qquad \boldsymbol{\xi} \in \mathcal{D}$$

where $\hat{f}_0$ is the discrete Fourier data and $\mathcal{D}$ is the (restricted) domain of knowledge.

More Generally...

Suppose we are given

- An underdetermined problem
- Having a sparse solution

More Generally...

Suppose we are given

- An underdetermined problem
- Having a sparse solution

Can we obtain

- the unique sparse solution?
- a good approximate solution, subject to noise?

Model Problem

Let $\mathcal{F} : \mathbb{C}^n \to \mathbb{C}^n$ be the discrete Fourier transform

$$\hat{f}_j = \sum_{k=0}^{n-1} f_k \mathrm{e}^{2\pi \mathrm{i} jk/n}, \qquad j = 0, \ldots, n-1.$$

Consider $T = \mathrm{supp}(f) \subset \mathbb{Z}^n$ with $|T| < n$ and $\Omega \subset \mathbb{Z}^n$. We define $\mathcal{F}_{T\to\Omega} : \mathbb{C}^{|T|} \to \mathbb{C}^{|\Omega|}$ to be the restricted transform

$$\hat{f}_j = \sum_{k\in T} f_k \mathrm{e}^{2\pi \mathrm{i} jk/n}, \qquad j \in \Omega.$$

Model Problem

Candes, Romberg, Tao studied when

$$f = \langle f_0, \ldots, f_{n-1} \rangle$$

with T=supp(f), can be recovered from

$$\hat{f}|_\Omega = \left\{ \hat{f}_j \mid j \in \Omega \subset \mathbb{Z}^n \right\}.$$

Solution is unique if n is prime and $|T| \leq \frac{1}{2}|\Omega|$.

Model Problem

Why? For n prime, $\mathcal{F}_{T\to\Omega}$ is

- Injective (one-to-one) if $|T| \leq |\Omega|$
- Surjective (onto) if $|T| \geq |\Omega|$
- Bijective if $|T| = |\Omega|$

Model Problem

Why? For n prime, $\mathcal{F}_{T\to\Omega}$ is

- Injective (one-to-one) if $|T| \le |\Omega|$
- Surjective (onto) if $|T| \ge |\Omega|$
- Bijective if $|T| = |\Omega|$

Suppose f, g are such that

- $\hat{f}|_\Omega = \hat{g}|_\Omega$ and
- $|\text{supp}(f)| \le \frac{1}{2}|\Omega|$ and $|\text{supp}(g)| \le \frac{1}{2}|\Omega|$.

Then $|\text{supp}(f-g)| \le |\Omega|$, hence $\mathcal{F}_{\text{supp}(f-g)\to\Omega}$ is injective and therefore $f - g = 0$.

Model Problem

In addition, the convex optimization problem

$$\text{minimize } \|f\|_1 = \sum_{k=0}^{n-1} |f_k|$$

$$\text{subject to } \hat{f}|_\Omega = \mathcal{F}_{\mathbb{Z}^n \to \Omega} f$$

yields the unique solution.

What if n is not prime?

Model Problem

For non-prime n, subgroups of $\mathbb{Z}^n$ spoil uniqueness.

For example, let $n = k^2$ and $|T| = k$ with $f_{jk} = 1$ for $j = 0, \ldots, k-1$. Then $\hat{f} = k \cdot f$, so $\hat{f}$ vanishes on sets Ω with $|\Omega|$ as large as $n - k$.

Solution:

- randomly choose Ω and
- settle for high probability of uniqueness.

Model Problem

Theorem (Candes, Romberg, Tao) Let $f \in \mathbb{C}^n$ be supported on an unknown set T and choose $\Omega \subset \mathbb{Z}^n$ of size $|\Omega|$ uniformly at random. For a given accuracy parameter m, if

$$|T| \leq C_m \cdot (\log n)^{-1} \cdot |\Omega|$$

then with probability at least $1 - O(n^{-m})$, the solution to the convex optimization problem

$$\min ||g||_1 \quad \text{s.t.} \;\; \hat{f}|_\Omega = \mathcal{F}_{\mathbb{Z}^n \to \Omega} g$$

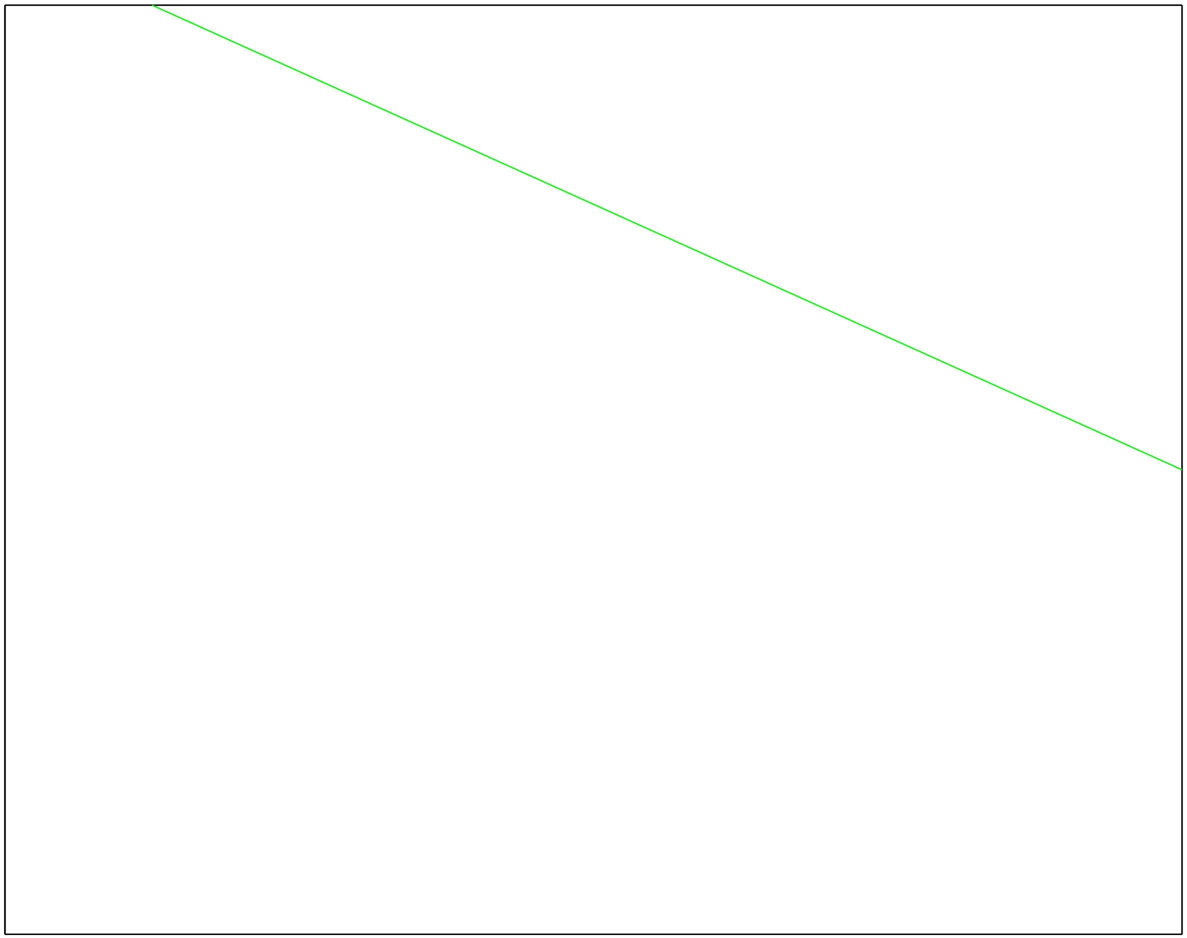
is unique and $g = f$.

Intuition

L_1-norm minimization with hyperplane constraint

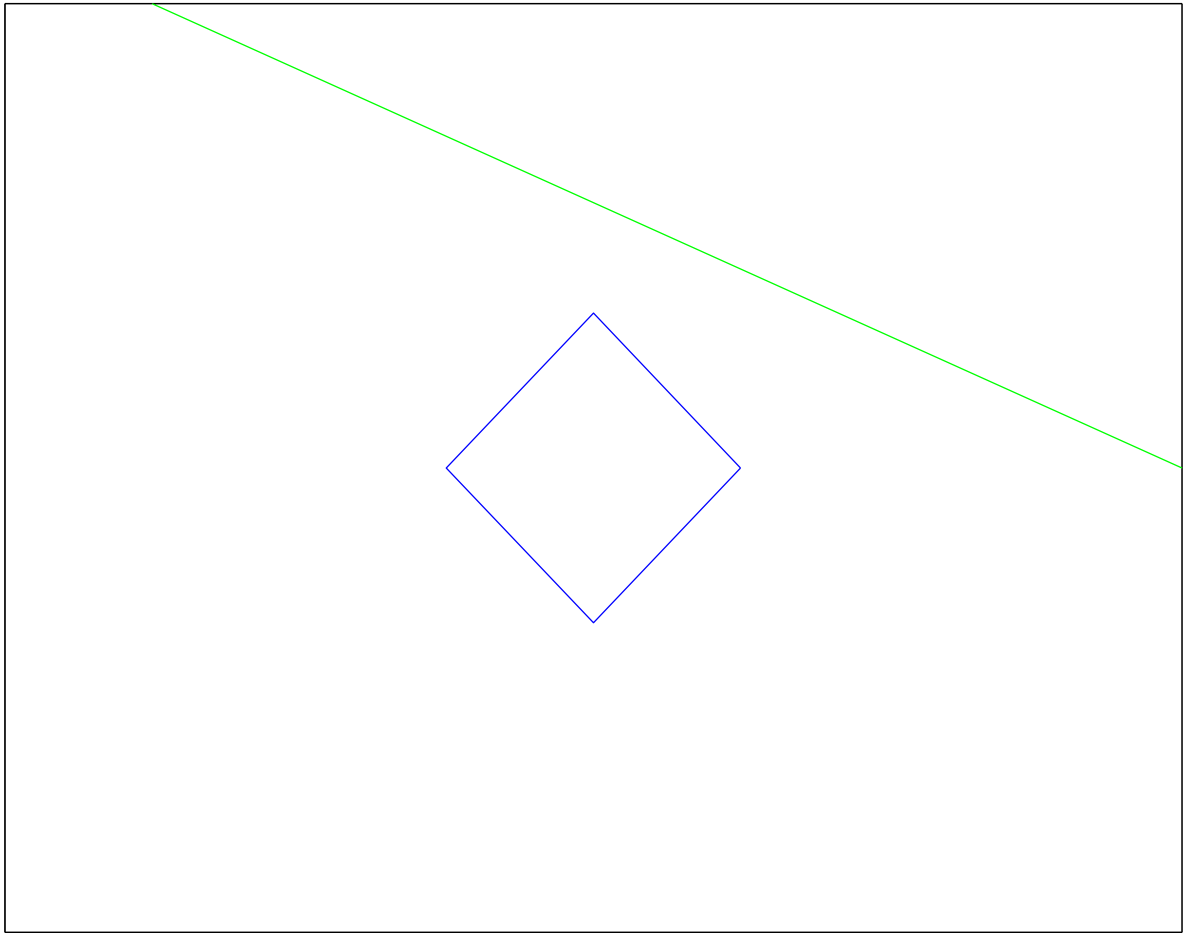
Intuition

L_1-norm minimization with hyperplane constraint



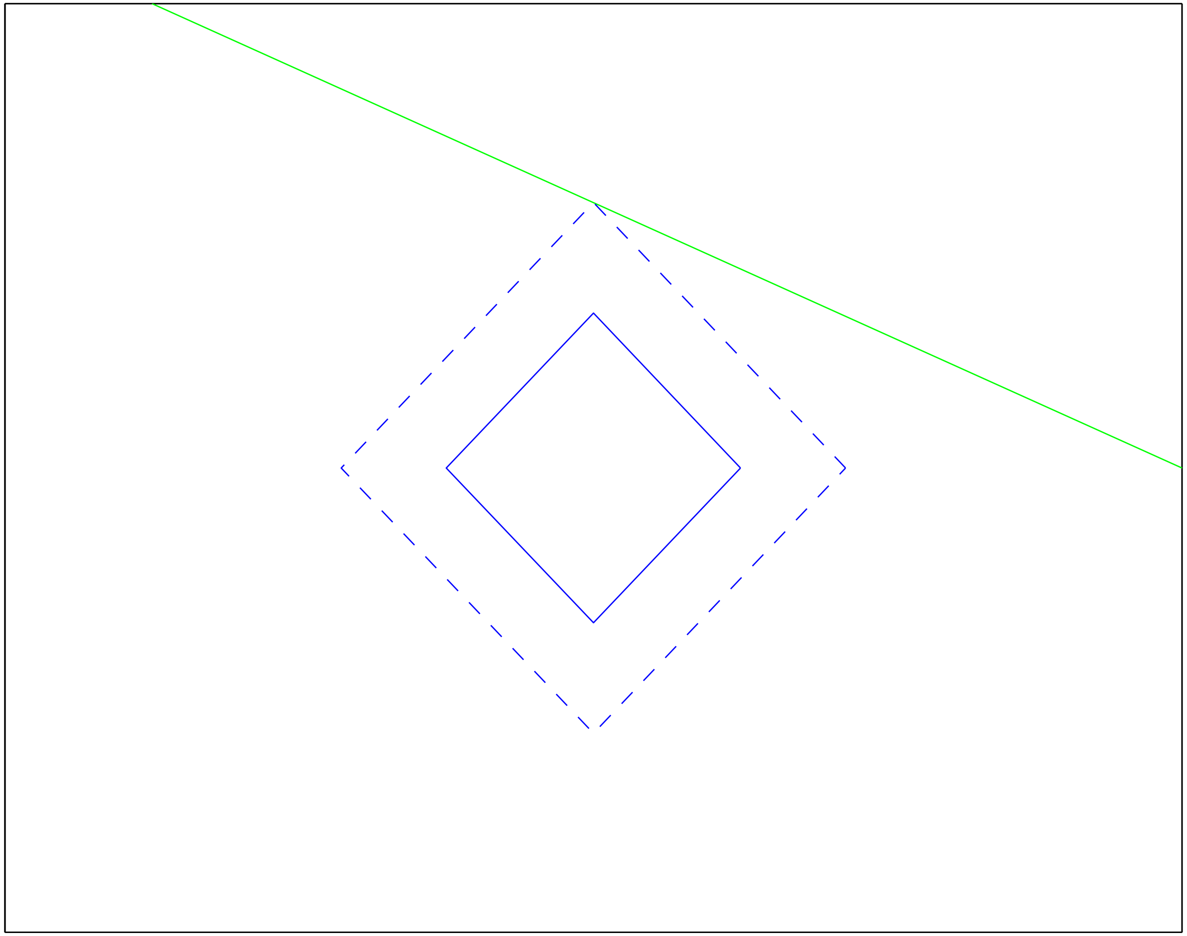
Intuition

L_1-norm minimization with hyperplane constraint



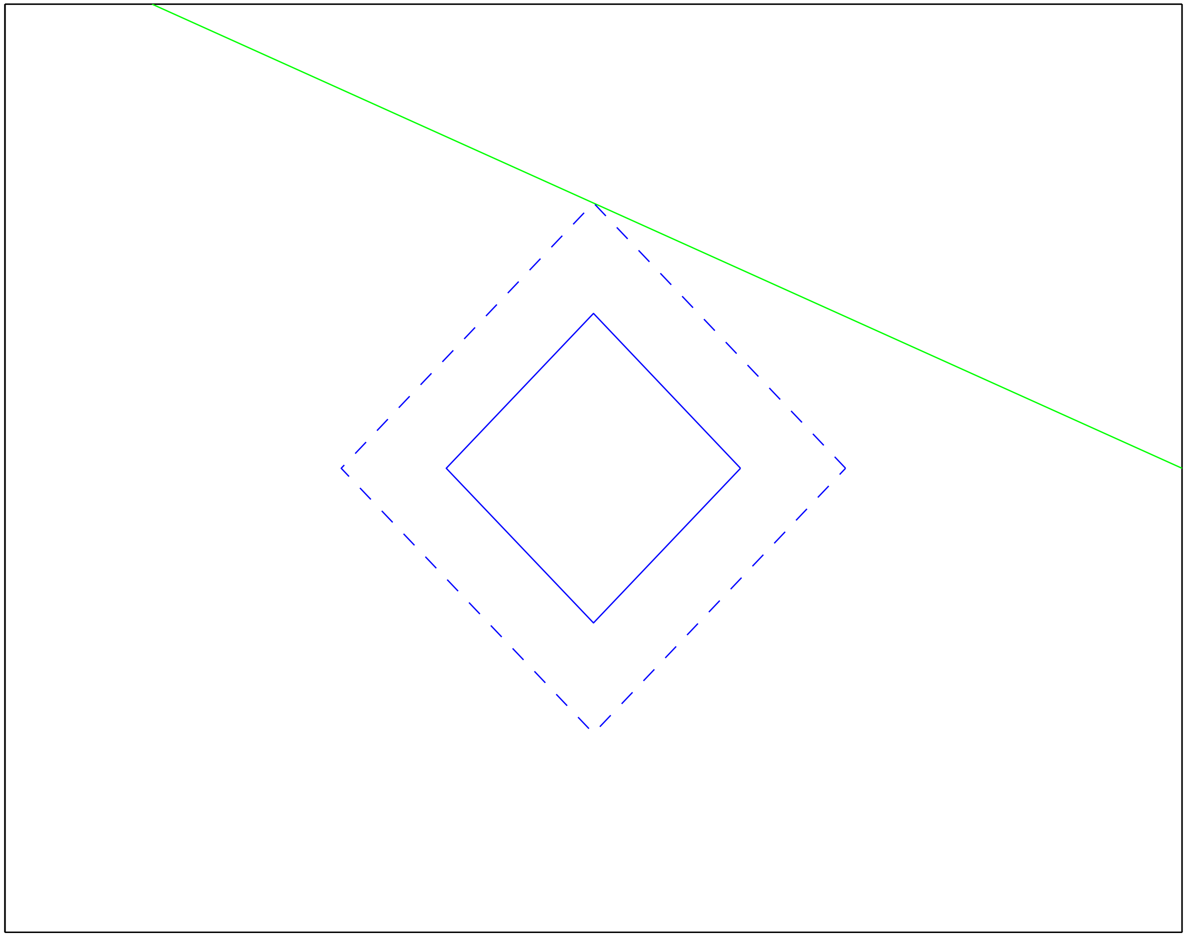
Intuition

L_1-norm minimization with hyperplane constraint



Intuition

L_1-norm minimization with hyperplane constraint



Contrast this with L_2-norm minimization

More General Problem

Recovery of x from

- Underdetermined $|\Omega| \times n$ system $Ax = b$
- Minimization of $\|x\|_1 = \sum_i |x_i|$

What properties must A have?

What relation must hold between three sizes

$$|T| = \|x\|_0, \quad |\Omega| = \dim(b), \quad n = \dim(x)?$$

More General Problem

If the columns of A are chosen uniformly at random from the unit sphere of dimension $\dim(b)$ and

$$\|x\|_0 \leq C \cdot [\log \dim(x)]^{-1} \cdot \dim(b),$$

where C depends weakly on the probability of correctness, then the solution of the same convex optimization problem is x.

Proof Tools

Restricted Isometry Property

- Sparse subsets of columns of $k \times n$-matrix A must be approximately orthogonal

For each $\mathcal{M} \subset \{1, \ldots, n\}$, let $A[\mathcal{M}]$ be the $k \times |\mathcal{M}|$ submatrix of A consisting of columns indexed by $\mathcal{M}$.

Define δ_s as the smallest number obeying

$$(1 - \delta_s) \|x\|^2 \leq \|A[\mathcal{M}]x\|^2 \leq (1 + \delta_s) \|x\|^2$$

for all subsets $\mathcal{M}$ with $|\mathcal{M}| \leq s$ and all vectors x.

Proof Tools

Restricted Isometry Property

Related to δ_s there is $\gamma_{s,s'}$, which is the smallest number such that

$$|(A[\mathcal{M}]x) \cdot (A[\mathcal{M}']x')| \leq \gamma_{s,s'} \|x\| \, \|x'\|$$

holds for all **disjoint** sets $\mathcal{M}, \mathcal{M}' \subset \{1, \ldots n\}$ of size not exceeding s and s', respectively, and all vectors x and x'.

Dantzig Selector

Candes, Tao studied underdetermined, noisy problem

Number of parameters n, number of measurements k, possibly $n \gg k$.

Assume measurements $y = X\beta + z$, with

- $\beta \in \mathbb{R}^n$ parameters of interest
- $z \in \mathbb{R}^k$ noise i.i.d. N$(0, \sigma^2)$

They obtain approximation $\tilde{\beta}$ as solution

of linear program.

Dantzig Selector

The linear program

$$\text{minimize } \|\tilde{\beta}\|_1 \quad \text{subject to}$$

$$\left\| X^t (y - X\tilde{\beta}) \right\|_\infty \leq \left(1 - t^{-1}\right) \sqrt{2 \log n} \cdot \sigma$$

yields solution $\tilde{\beta}$ that with very high probability satisfies

$$\|\tilde{\beta} - \beta\|^2 \leq C^2 \cdot 2 \log n \cdot \left(\sigma^2 + \sum_i \min\left({\beta_i}^2, \sigma^2\right) \right).$$

Compare with omitting the factor $2 \log n$ if the nonzero locations of β were known in advance.

Orthogonal Matching Pursuit

Tropp, Gilbert adapted a greedy algorithm to recover x from knowledge of $A = (a_1, a_2, \ldots, a_n)$ and b, where column vectors $a_1, \ldots, a_n$ have unit norm.

- Choose column j_1 to maximize $a_{j_1} \cdot b$
- Let residual $r_1 = b - (a_{j_1} \cdot b)\, a_{j_1}$
- Repeat, for steps $s = 2, \ldots, \|x\|_0$, picking column j_s to maximize $a_{j_s} \cdot r_{s-1}$
- New residual r_s is obtained by removing from b the orthogonal projection of b on the columns chosen so far

Orthogonal Matching Pursuit

Claims: compared to $\|x\|_1$ minimization

- Faster
- Similar convergence properties

Distinctions will be understood through numerical experience

Caveats for C-R-T Shepp-Logan

Candes, Romberg, Tao interpretation of Shepp-Logan

- Blurs distinction between DFT and FT
  - In tomography and MRI, true Fourier data are collected
  - For discontinuous functions, DFT and FT very different
  - FT requires a continuous, rather than discrete (pixel) representation of image function
- Works for piecewise constant, not piecewise smooth, images

Recovery of piecewise smooth images from truncated Fourier data remains an open problem

Numerical Experiments

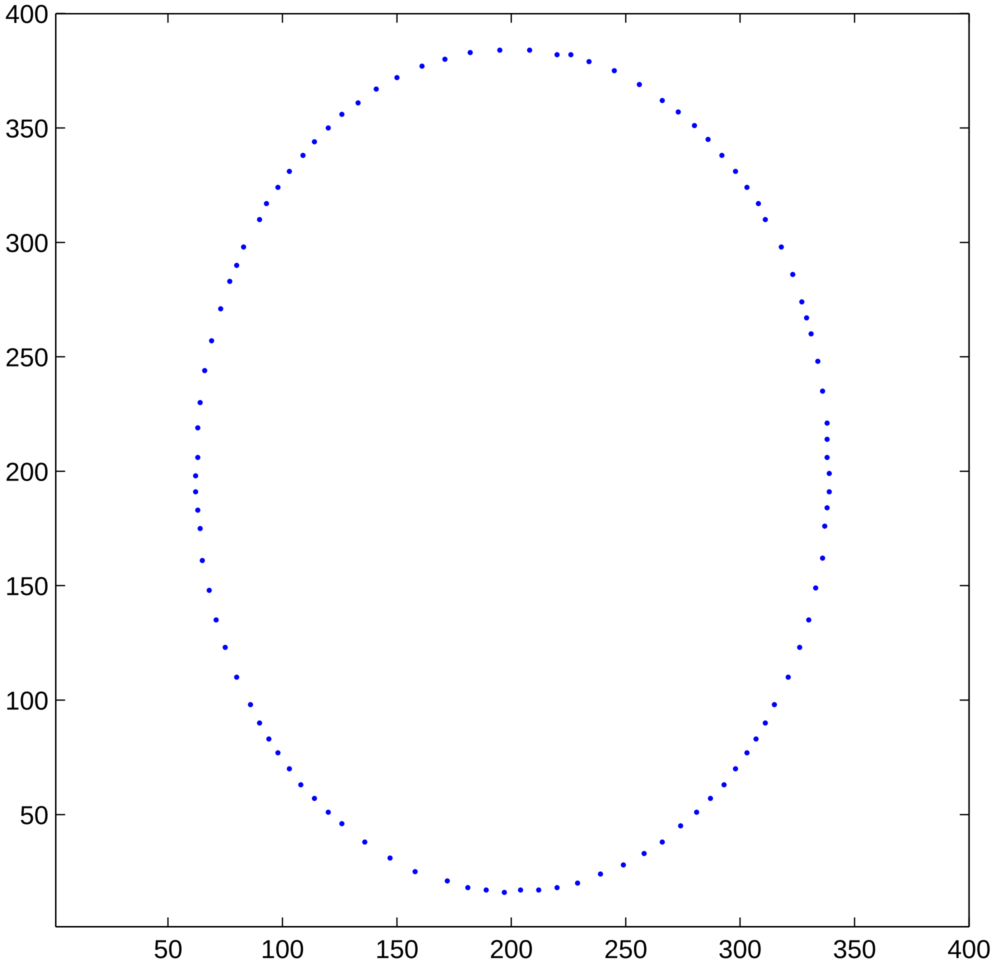
Collaboration with Yu Chen (Courant Institue, NYU)

- Assume that, aside from discontinuities, available Fourier data are complete
- Recall that discontinuities in function (resp. derivative) along curves can be represented by double (resp. single) layer potential
- Single and double layer densities can be discretized as monopoles and dipoles

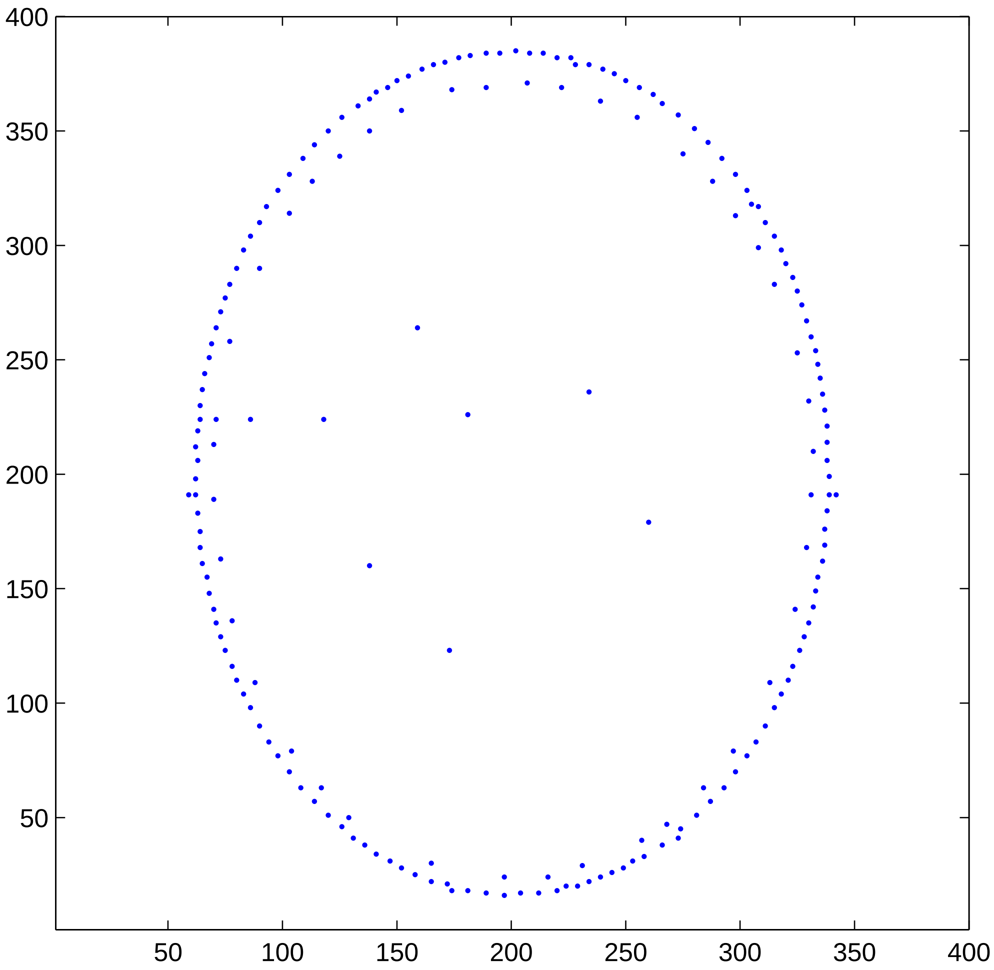
Orthogonal matching pursuit used in attempt to recover dipoles

Numerical Experiments

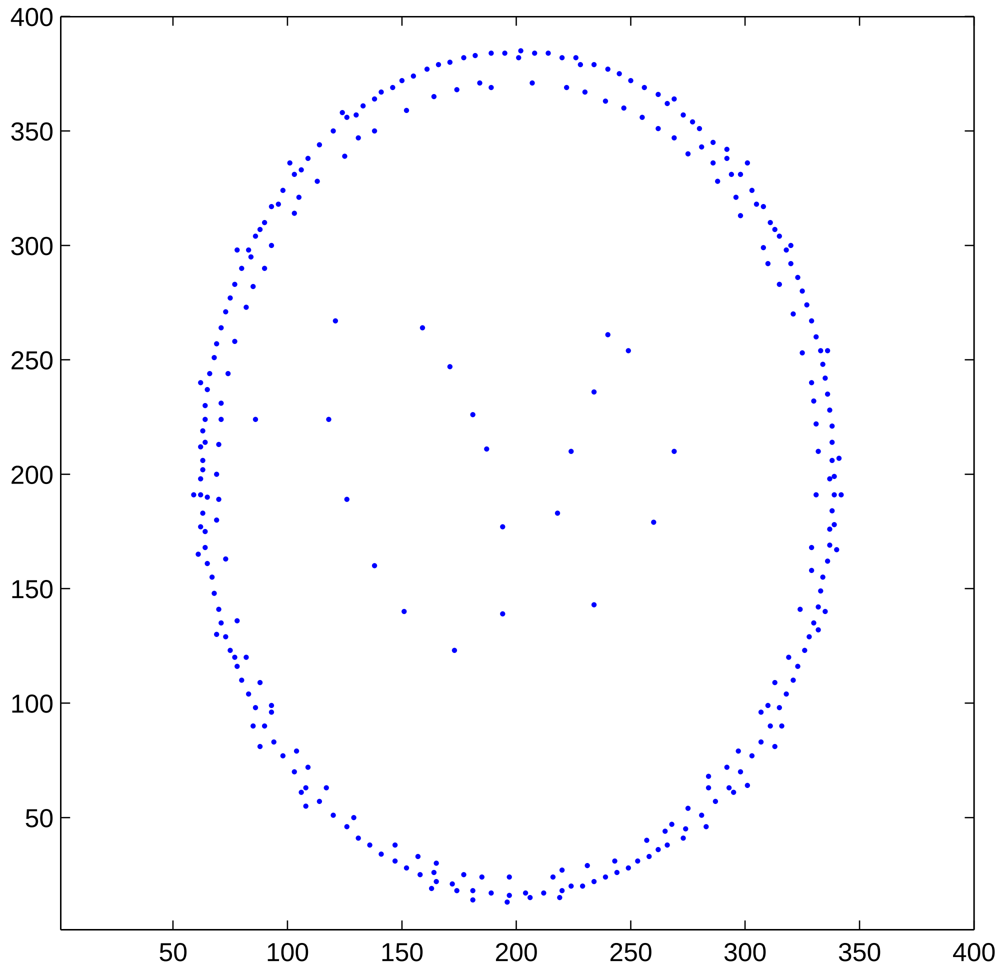
Numerical Experiments



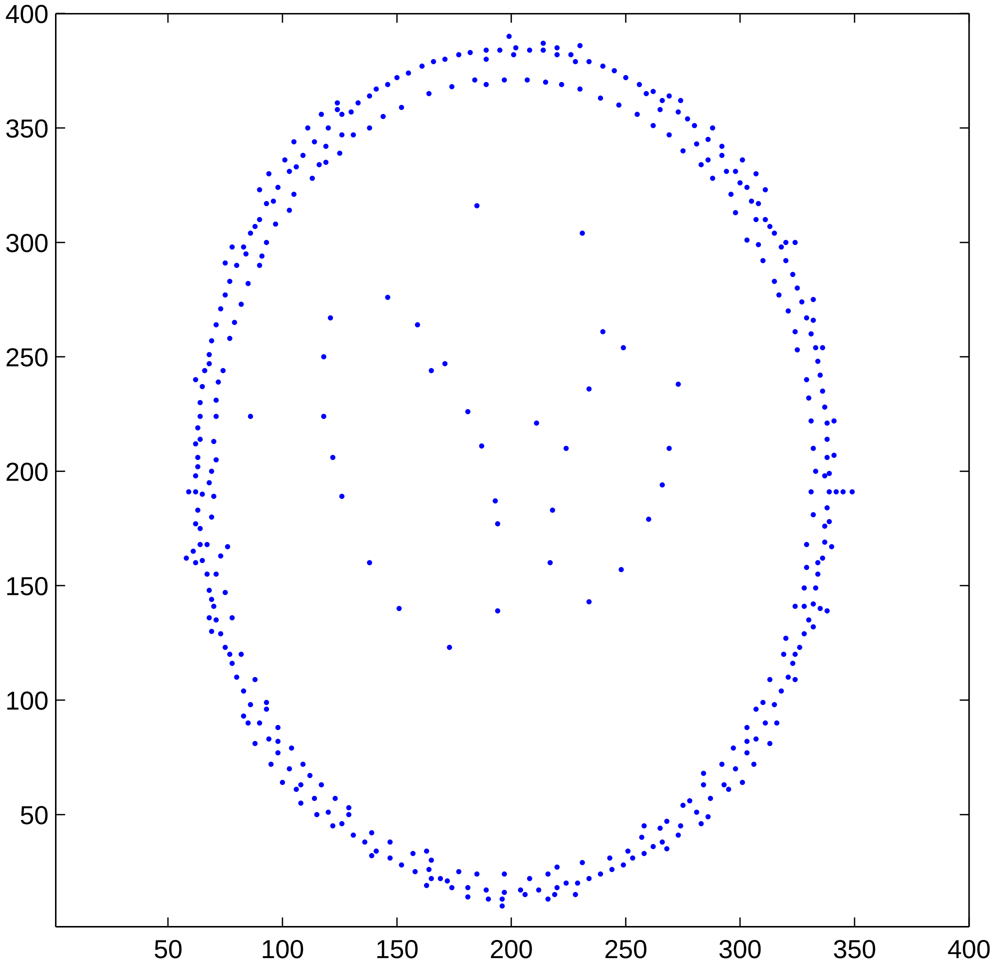
Numerical Experiments



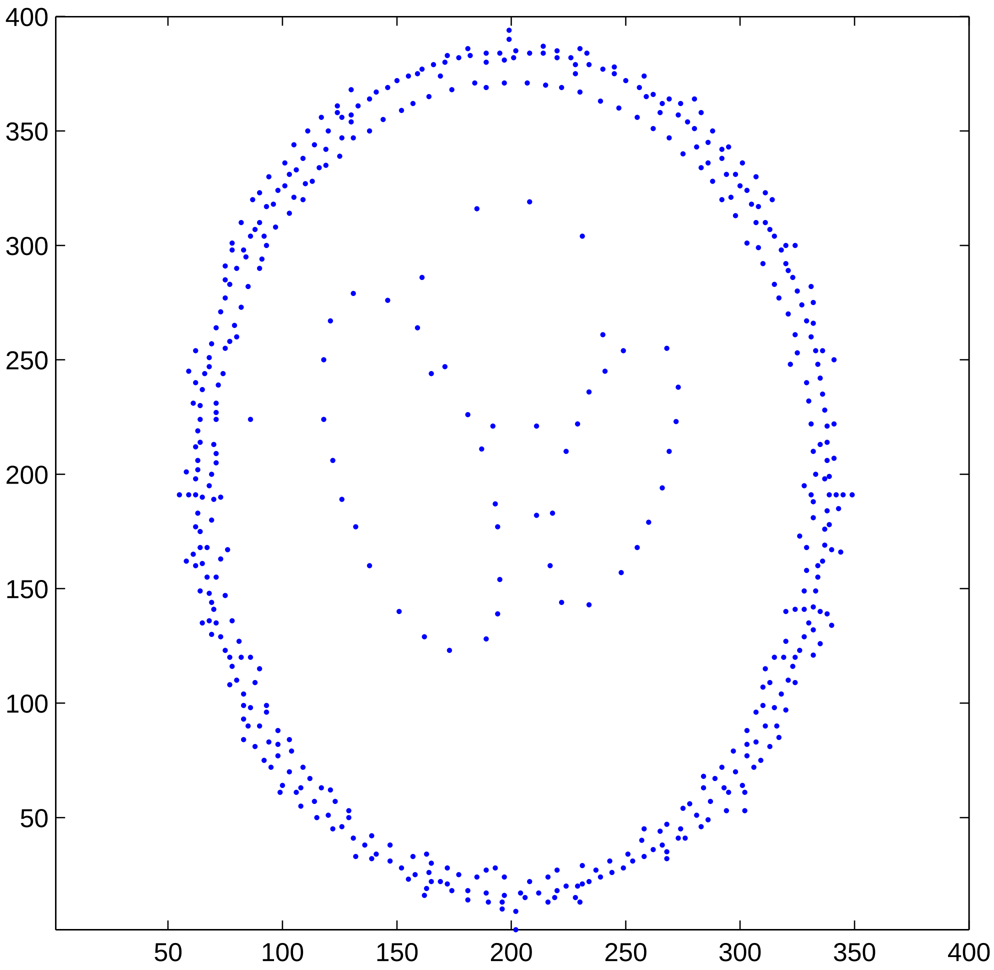
Numerical Experiments



Numerical Experiments



Numerical Experiments

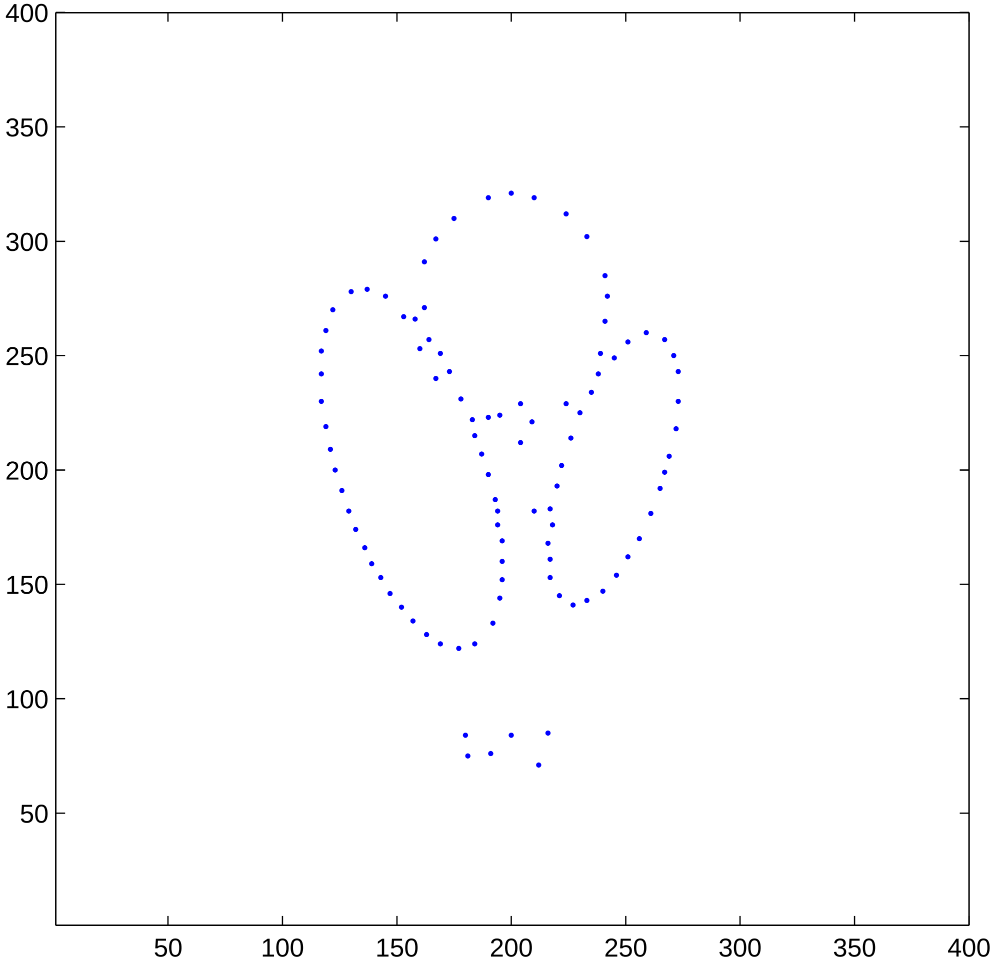


Numerical Experiments

Shepp-Logan without outer two ellipses

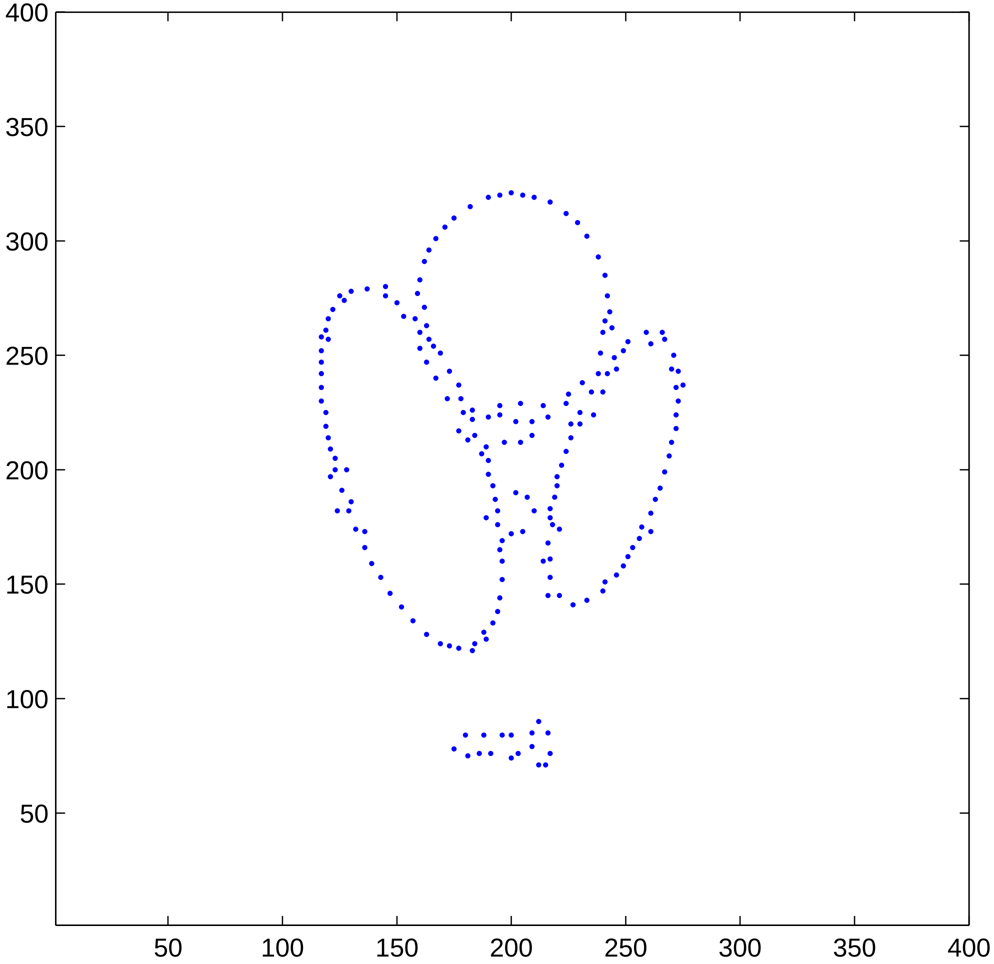
Numerical Experiments

Shepp-Logan without outer two ellipses



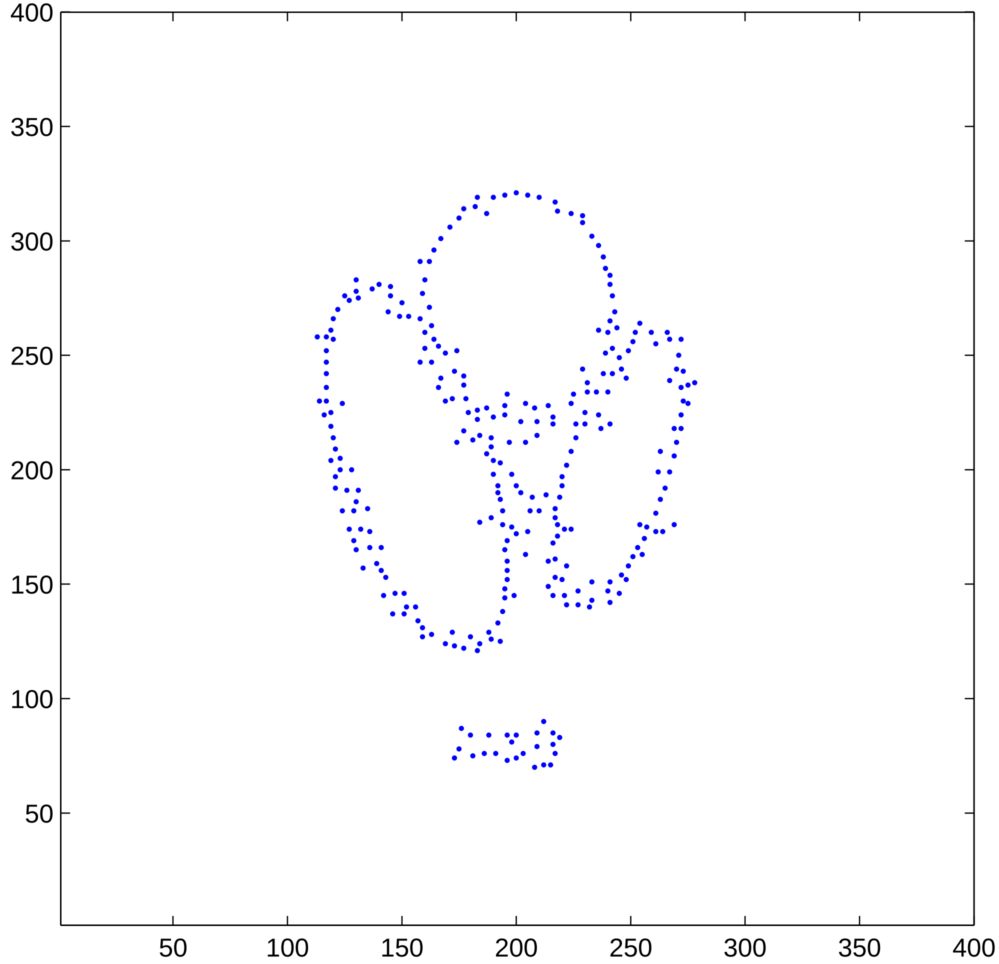
Numerical Experiments

Shepp-Logan without outer two ellipses



Numerical Experiments

Shepp-Logan without outer two ellipses



Numerical Experiments

Prompted by unmet need to recover piecewise smooth images from truncated Fourier data

- Shows that truncated Fourier data supplies excellent, recoverable position information
- Suggests
  - Dipoles should not be restricted to grid
  - Asking how to “connect the dots”
- More experimentation needed
- More innovation in representation needed

Summary

L_1-norm minimization, subject to the constraint of an underdetermined linear system

- Can be highly effective at sparse recovery
- Plausible workhorse, due to advances in interior point methods
- Still costly enough to prompt alternatives

The bigger questions about sparse recovery remain open

- Dantzig selector suggests great opportunity in sparse estimation problems
- What about estimation in nonlinear setting?

Sparse Recovery

Additional information

- Search Google for “L1 magic”
- Site http://www.acm.caltech.edu/l1magic contains links, preprints