Near Optimal Signal Recovery from Random Projections

Emmanuel Candès, California Institute of Technology

Multiscale Geometric Analysis in High Dimensions: Workshop # 2
IPAM, UCLA, October 2004

Collaborators: Justin Romberg (Caltech), Terence Tao (UCLA)

Recovery Problem

- Object $f \in \mathbf{R}^N$ we wish to reconstruct: digital signal, image; dataset.
- Can take linear measurements

$$y_k = \langle f, \psi_k \rangle, \qquad k = 1, 2, \ldots, K.$$

- How many measurements do we need to do recover f to within accuracy ϵ

$$\|f - f^\sharp\|_{\ell_2} \le \epsilon$$

for typical objects f taken from some class $f \in \mathcal{F} \subset \mathbf{R}^N$.

- Interested in practical reconstruction methods.

Agenda

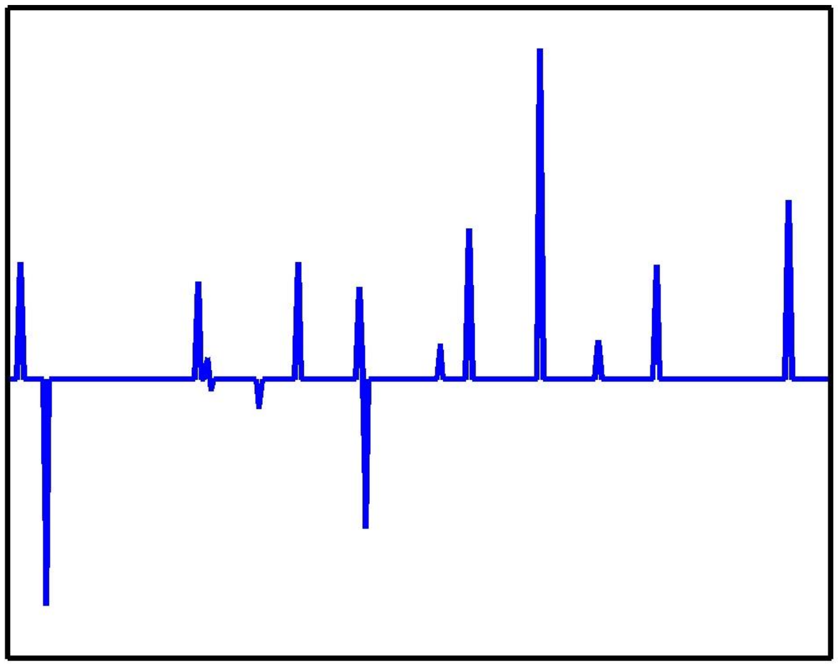
- Background: exact reconstruction of sparse signals
- Near-optimal reconstruction of compressible signals
- Uniform uncertainty principles
- Exact reconstruction principles
- Relationship with coding theory
- Numerical experiments

Sparse Signals

- Vector $f \in \mathbf{R}^N$; digital signal, coefficients of a digital signal/image, etc.)
- $|T|$ nonzero coordinates ($|T|$ spikes)

$$T := \{t,\ f(t) \neq 0\}$$

- Do not know the locations of the spikes
- Do not know the amplitude of the spikes



Recovery of Sparse Signals

- Sparse signal $\boldsymbol{f}$: $|\boldsymbol{T}|$ spikes
- Available information

$$\boldsymbol{y} = \boldsymbol{F}\,\boldsymbol{f},$$

  $\boldsymbol{F}$ is $\boldsymbol{K}$ by $\boldsymbol{N}$ with $\boldsymbol{K} << \boldsymbol{N}$

- Can we recover $\boldsymbol{f}$ from $\boldsymbol{K}$ measurements?

Fourier Ensemble

- Random set $\Omega \subset \{0, \dots, N-1\}$, $|\Omega| = K$.
- Random frequency measurements: observe $(Ff)_k = \hat{f}(k)$

$$\hat{f}(k) = \sum_{t=0}^{N-1} f(t) e^{-i2\pi kt/N}, \quad k \in \Omega$$

Exact Recovery from Random Frequency Samples

- Available information: $y_k = \hat{f}(k)$, Ω random and $|\Omega| = K$.
- To recover f, simply solve

$$(P_1) \qquad f^\sharp = \text{argmin}_{g \in \mathbf{R}^N} \|g\|_{\ell_1}, \quad \text{subject to} \quad Fg = Ff.$$

where

$$\|g\|_{\ell_1} := \sum_{t=0}^{N-1} |g(t)|.$$

Theorem 1 (C., Romberg, Tao) *Suppose*

$$|K| \geq \alpha \cdot |T| \cdot \log N.$$

Then the reconstruction is exact with prob. greater than $1 - O(N^{-\alpha\rho})$ *for some fixed* $\rho > 0$*:* $f^\sharp = f$. *(N.b.* $\rho \approx 1/29$ *works).*

Exact Recovery from Gaussian Measurements

- Gaussian random matrix

$$F(k,t) = X_{k,t}, \qquad X_{k,t} \ i.i.d. \ N(0,1)$$

- This will be called the *Gaussian ensemble*

Solve

$$(P_1) \qquad f^\sharp = \text{argmin}_{g \in \mathbf{R}^N} \|g\|_{\ell_1} \quad \text{subject to} \quad Fg = Ff.$$

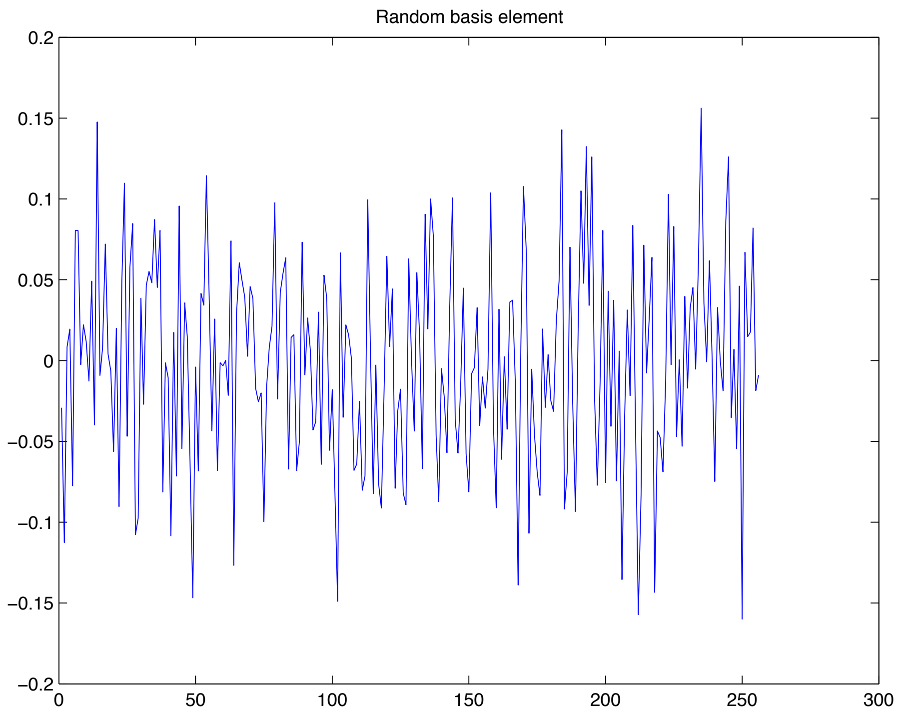
Theorem 2 (C., Tao) *Suppose*

$$|K| \geq \alpha \cdot |T| \cdot \log N.$$

Then the reconstruction is exact with prob. greater than $1 - O(N^{-\alpha\rho})$ *for some fixed* $\rho > 0$*:* $f^\sharp = f$*.*

Gaussian Random Measurements

$$y_k = \langle f, X \rangle, \quad X_t \ i.i.d. \ N(0,1)$$



Equivalence

- Combinatorial optimization problem

$$(P_0) \qquad \min_g \|g\|_{\ell_0} := \#\{t,\, g(t) \neq 0\}, \qquad Fg = Ff$$

- Convex optimization problem (LP)

$$(P_1) \qquad \min_g \|g\|_{\ell_1}, \qquad Fg = Ff$$

- Equivalence:

For $K \asymp |T| \log N$, the solutions to (P_0) and (P_1) are unique and are the same!

About the ℓ_1-norm

- Minimum ℓ_1-norm reconstruction in widespread use
- Santosa and Symes (1986) proposed this rule to reconstruct spike trains from incomplete data
- Connected with Total-Variation approaches, e.g. Rudin, Osher, Fatemi (1992)
- More recently, ℓ_1-minimization, *Basis Pursuit*, has been proposed as a convex alternative to the combinatorial norm ℓ_0. Chen, Donoho Saunders (1996)
- Relationships with uncertainty principles: Donoho & Huo (01), Gribonval & Nielsen (03), Tropp (03) and (04), Donoho & Elad (03)

min ℓ_1 as LP

$$\min \|x\|_{\ell_1} \quad \text{subject to} \quad Ax = b$$

- Reformulated as an LP (at least since the 50's).
- Split x into $x = x_+ - x_-$

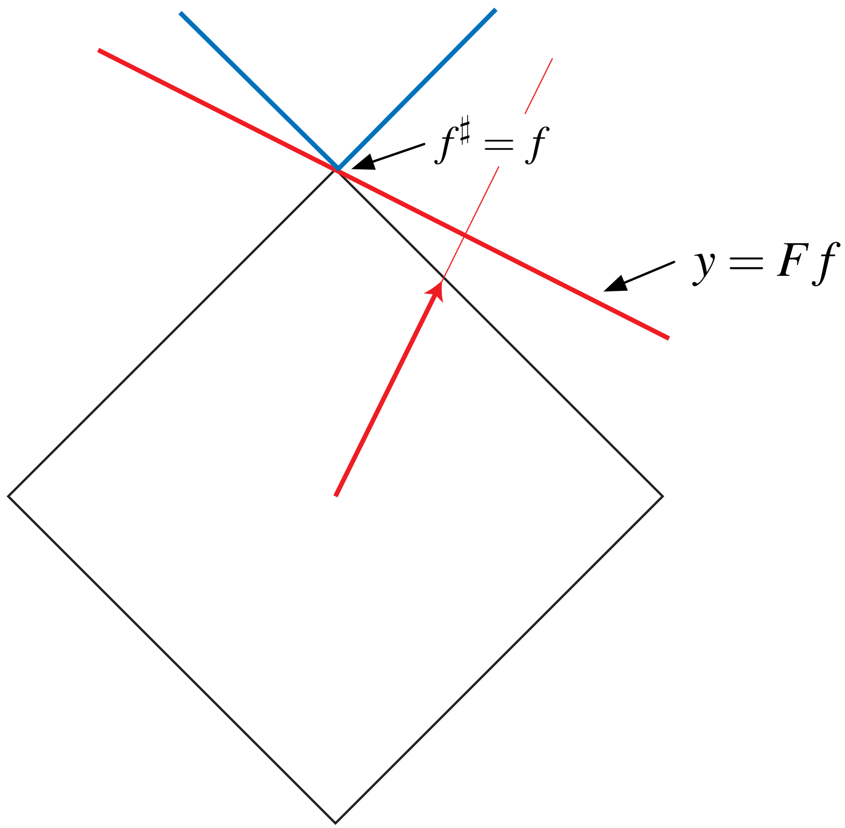
$$\min \mathbf{1}^T x_+ + \mathbf{1}^T x_- \quad \text{subject to} \quad \begin{cases} \begin{pmatrix} A & -A \end{pmatrix} \begin{pmatrix} x_+ \\ x_- \end{pmatrix} = b \\ x_+ \geq 0, x_- \geq 0 \end{cases}$$

Reconstruction of Spike Trains from Fourier Samples

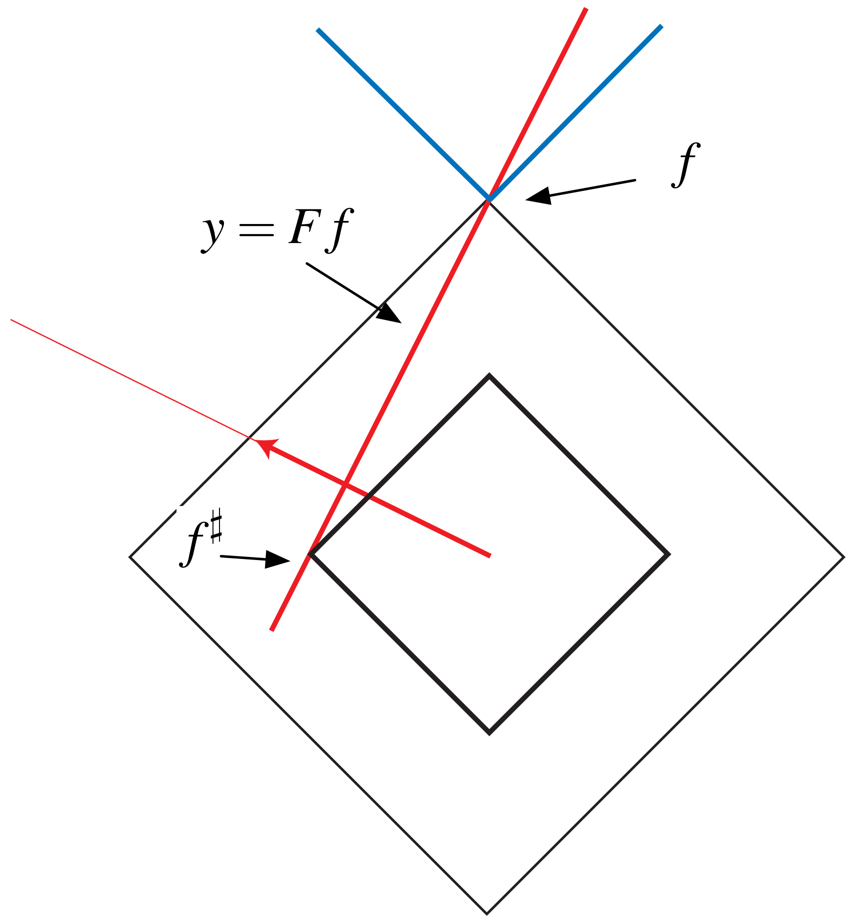
- Gilbert et al. (04)
- Santosa & Symes (86)
- Dobson & Santosa (96)
- Bresler & Feng (96)
- Vetterli et. al. (03)

Why Does This Work? Geometric Viewpoint

Suppose $\boldsymbol{f} \in \mathbf{R^2}$, $\boldsymbol{f} = \mathbf{(0, 1)}$.

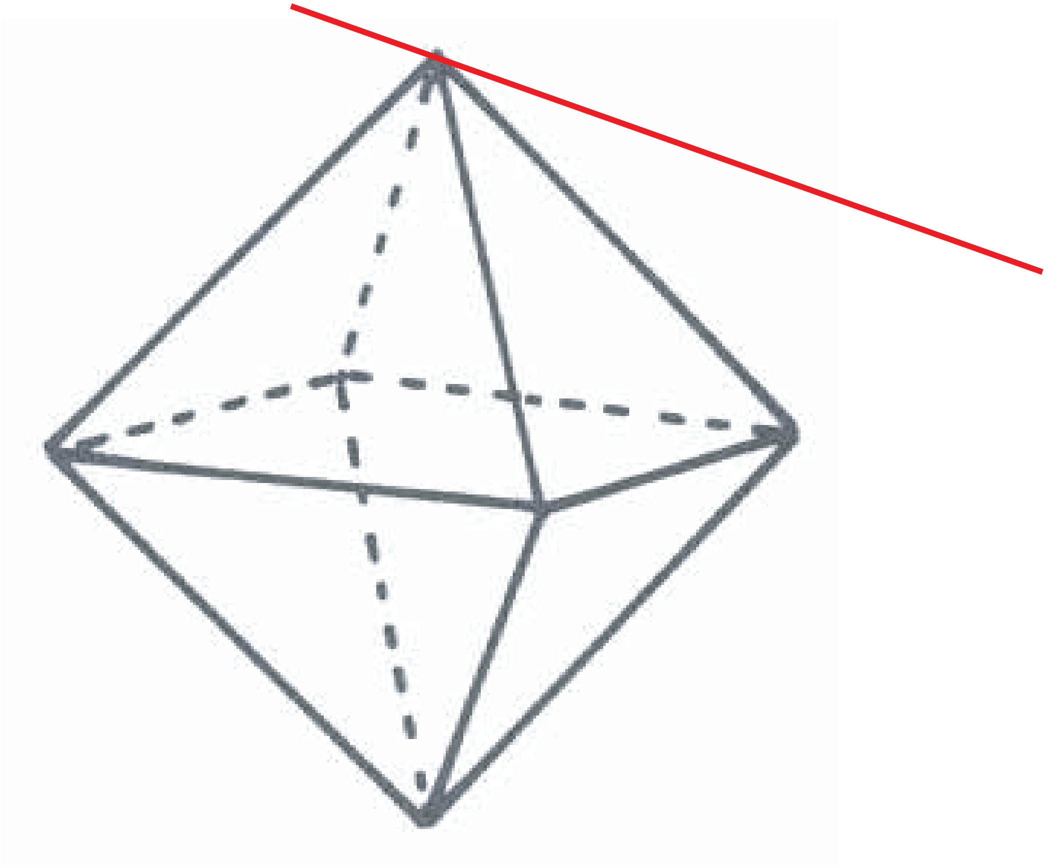


Exact



Miss

Higher Dimensions

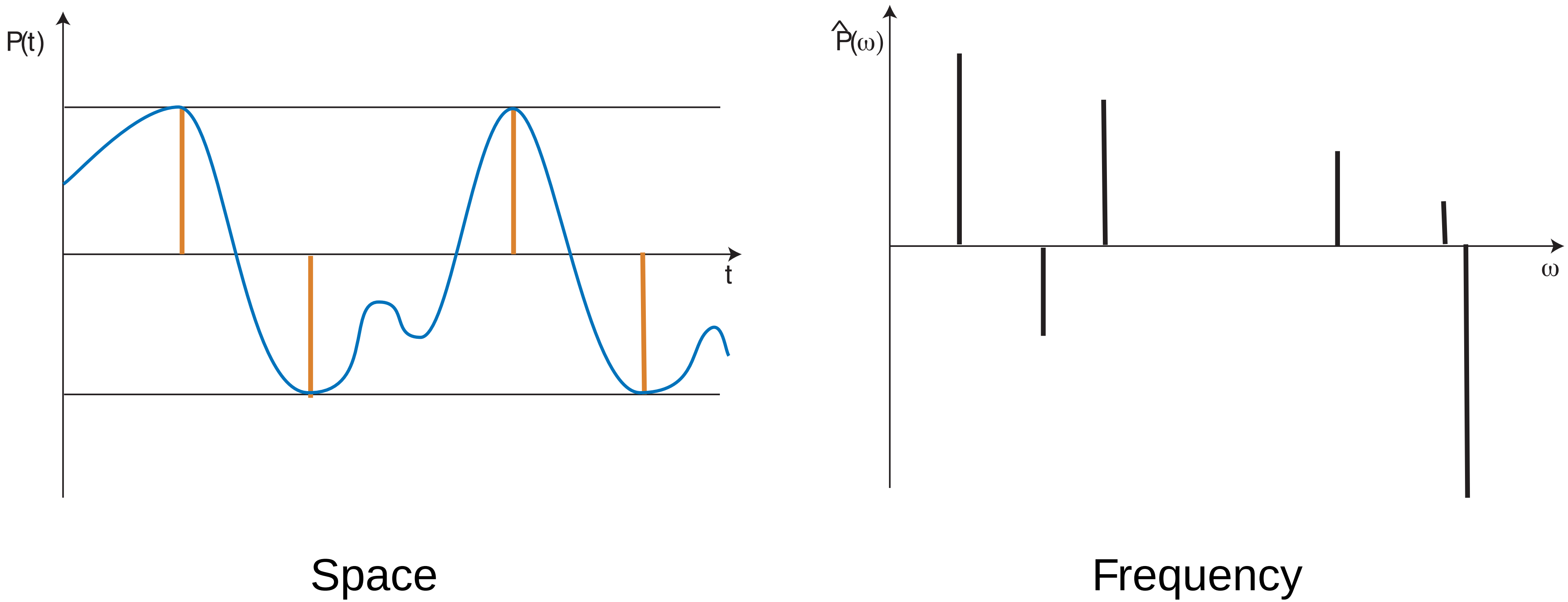


Duality in Linear/Convex Programming

- f unique solution 'if and only' if dual is feasible
- Dual is feasible if there is $P \in \mathbf{R}^N$
  - P is in the rowspace of F
  - P is a subgradient of $\|f\|_{\ell_1}$

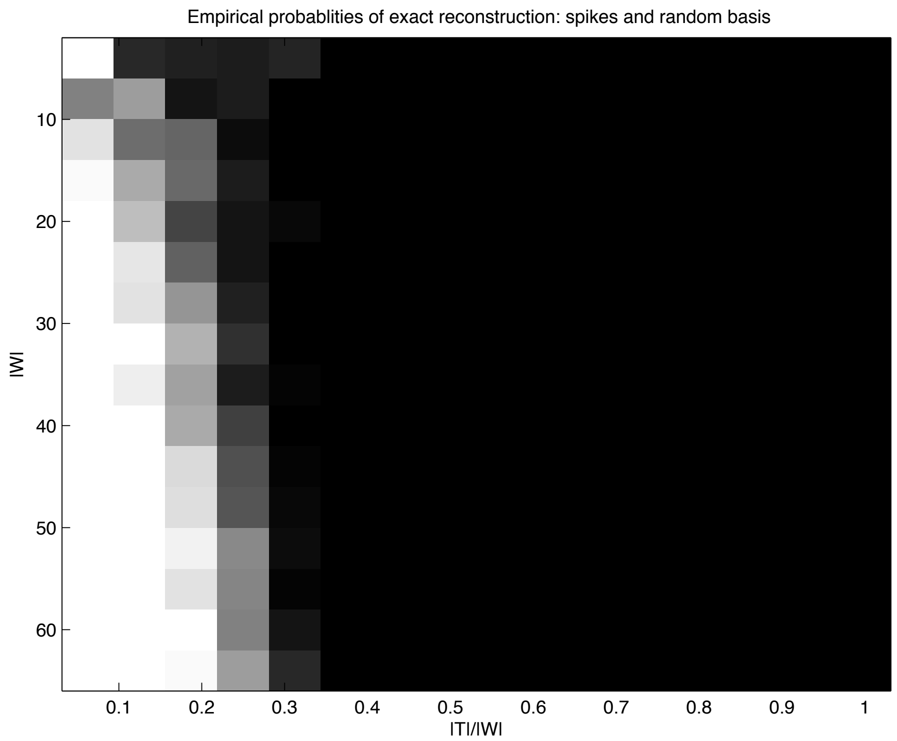
$$P \in \partial \|f\|_{\ell_1} \quad \Leftrightarrow \quad \begin{cases} P(t) = \text{sgn}(f(t)), & t \in T \\ |P(t)| < 1, & t \in T^c \end{cases}$$

Interpretation: Dual Feasibility with Freq. Samples



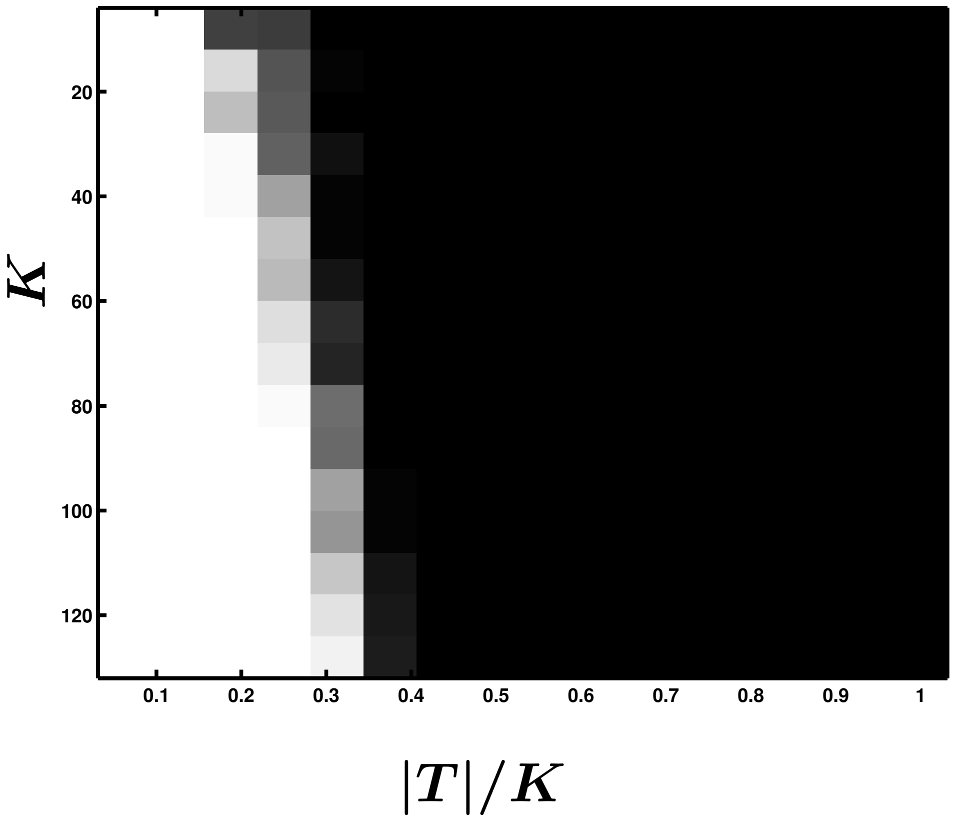
Numerical Results

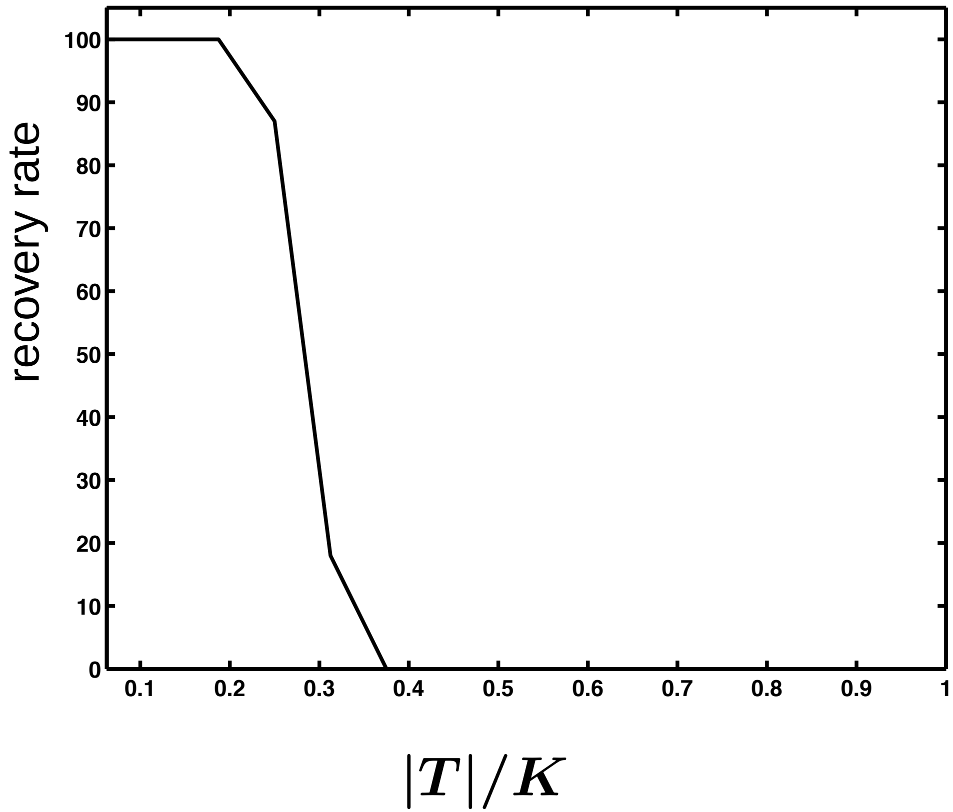
- Signal length $N = 256$
- Randomly place $|T|$ spikes, observe $K|$ Gaussian coefficients
- Measure $\%$ recovered perfectly
- White = always recovered, black = never recovered



Numerical Results

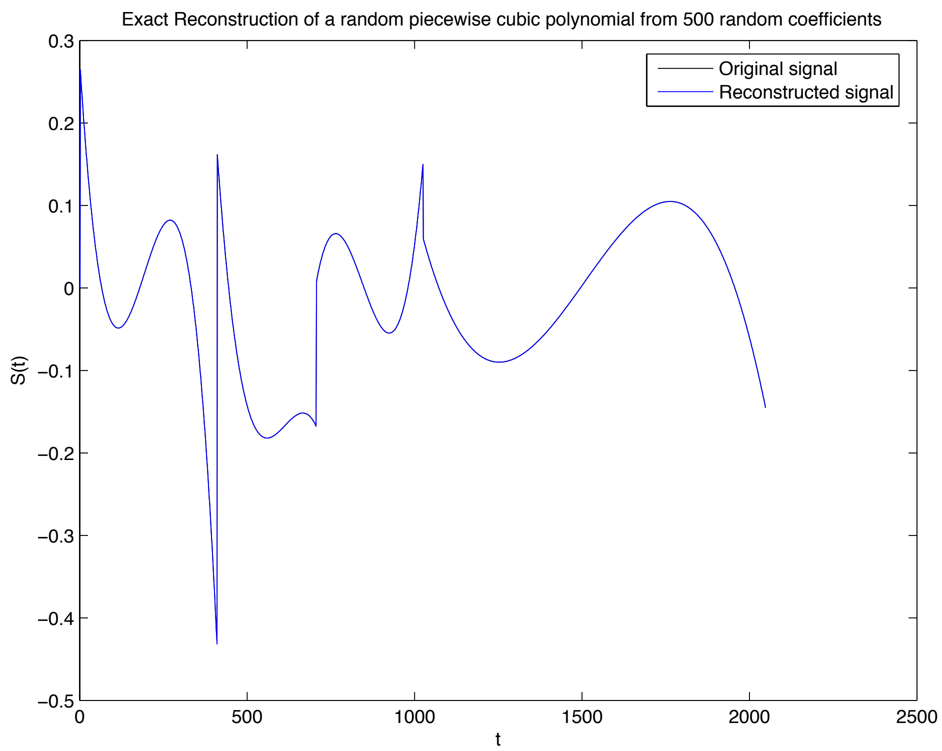
- Signal length $N = 1024$
- Randomly place $|T|$ spikes, observe K random frequencies
- Measure % recovered perfectly
- white = always recovered, black = never recovered





Reconstruction of Piecewise Polynomials, I

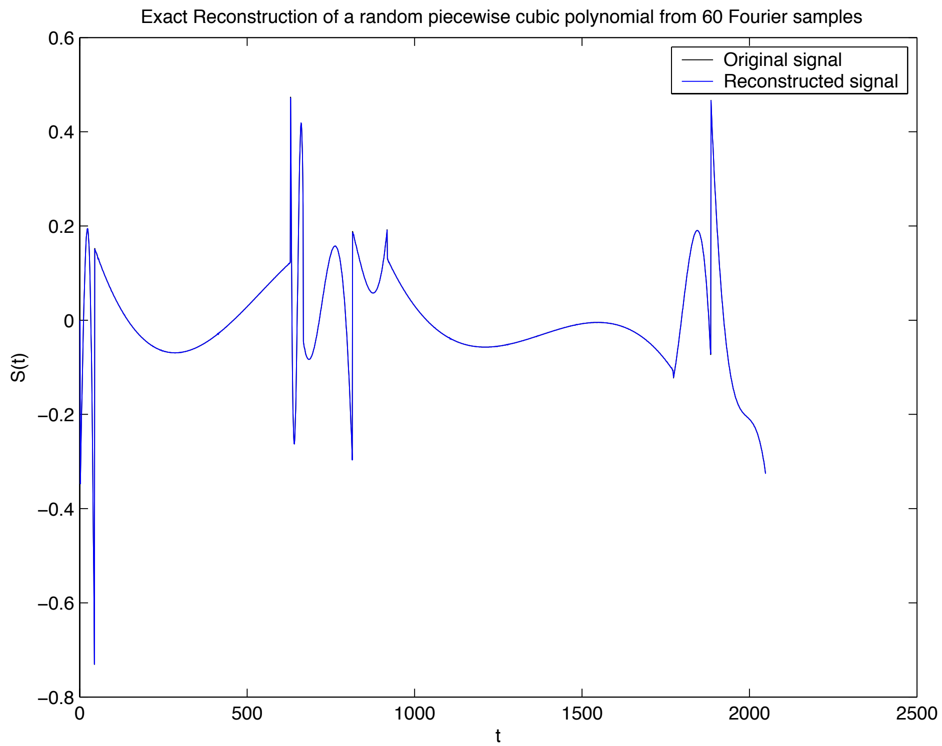
- Randomly select a few jump discontinuities
- Randomly select cubic polynomial in between jumps
- Observe about 500 random coefficients
- Minimize ℓ_1 norm of wavelet coefficients



Reconstructed signal

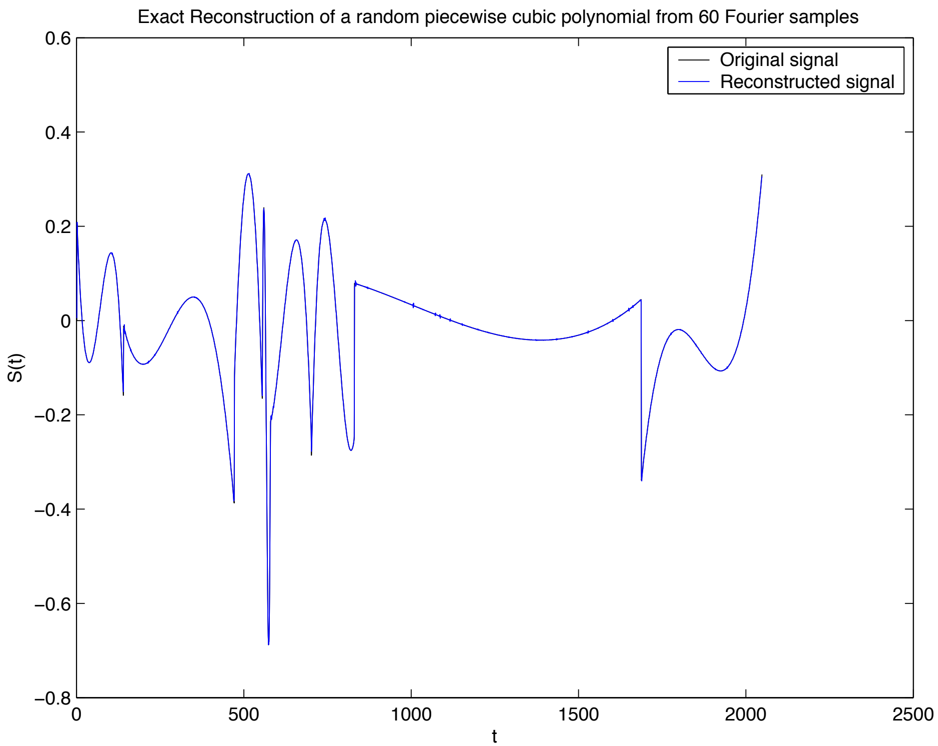
Reconstruction of Piecewise Polynomials,II

- Randomly select 8 jump discontinuities
- Randomly select cubic polynomial in between jumps
- Observe about 200 Fourier coefficients at random

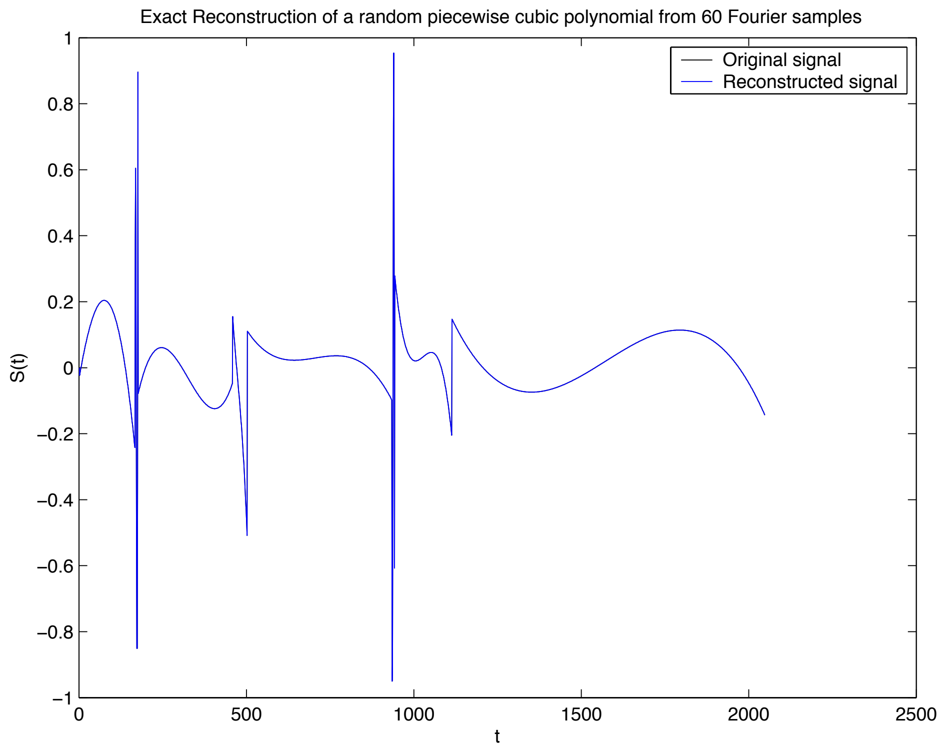


Reconstructed signal

Reconstruction of Piecewise Polynomials, III



Reconstructed signal



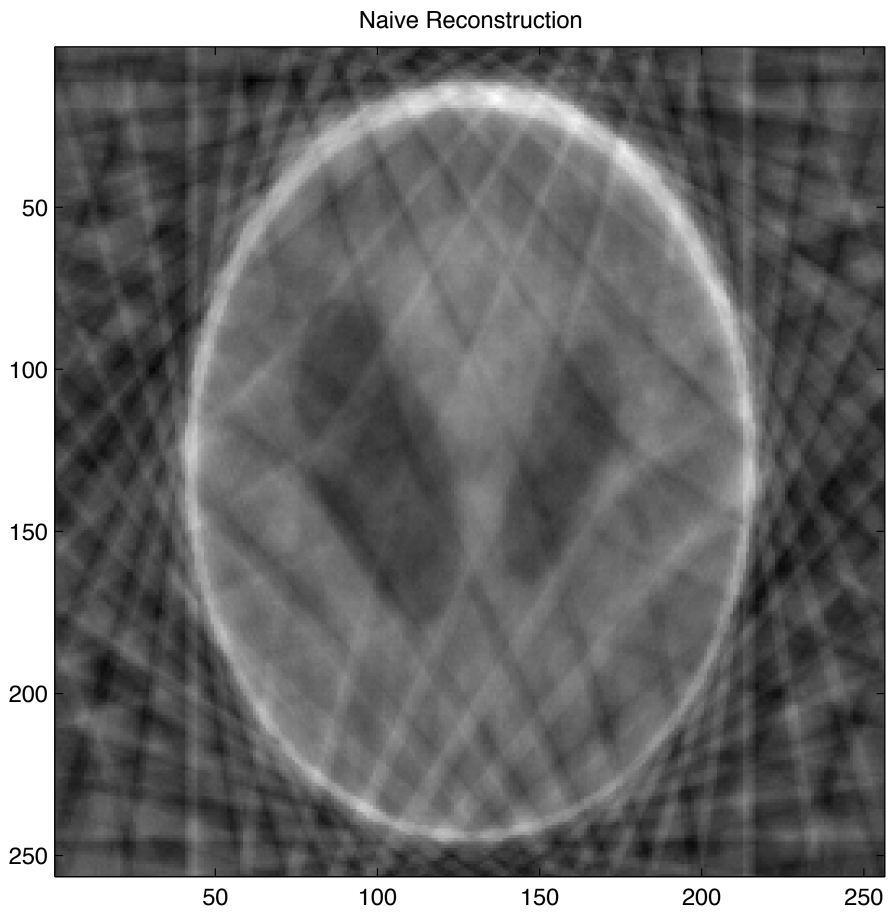
Reconstructed signal

About 200 Fourier coefficients only!

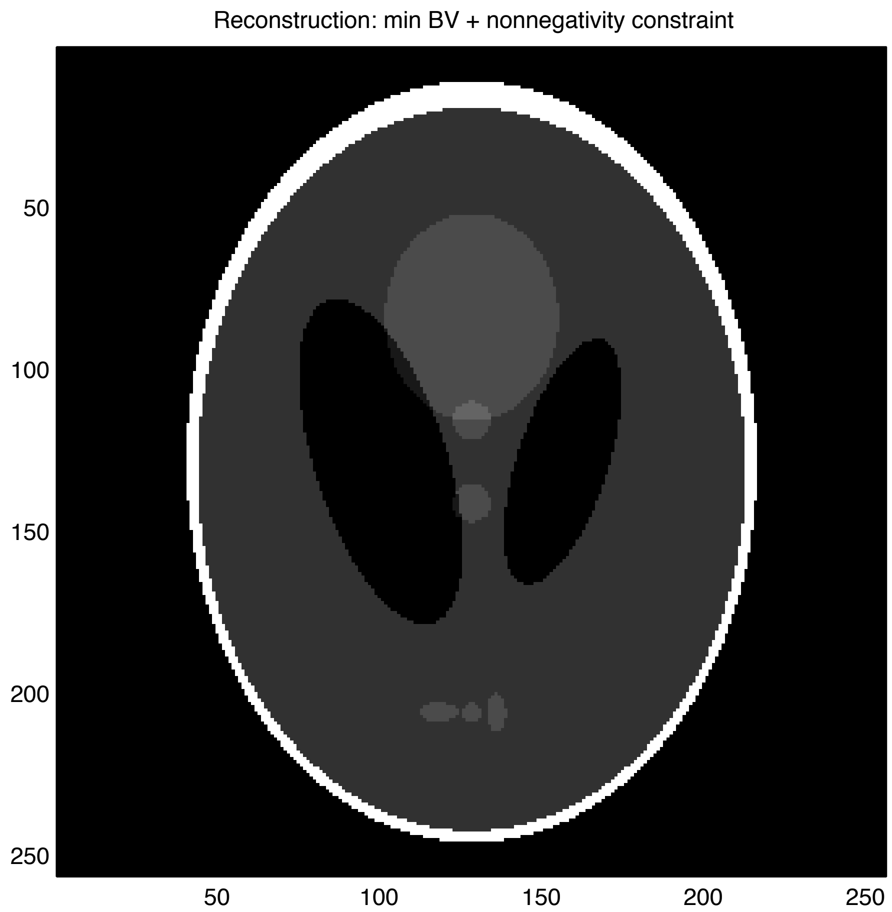
Minimum TV Reconstruction

Many extensions:

$$\min_{g} \|g\|_{TV} \quad \text{s.t.} \quad \hat{g}(\omega) = \hat{f}(\omega), \;\; \omega \in \Omega$$

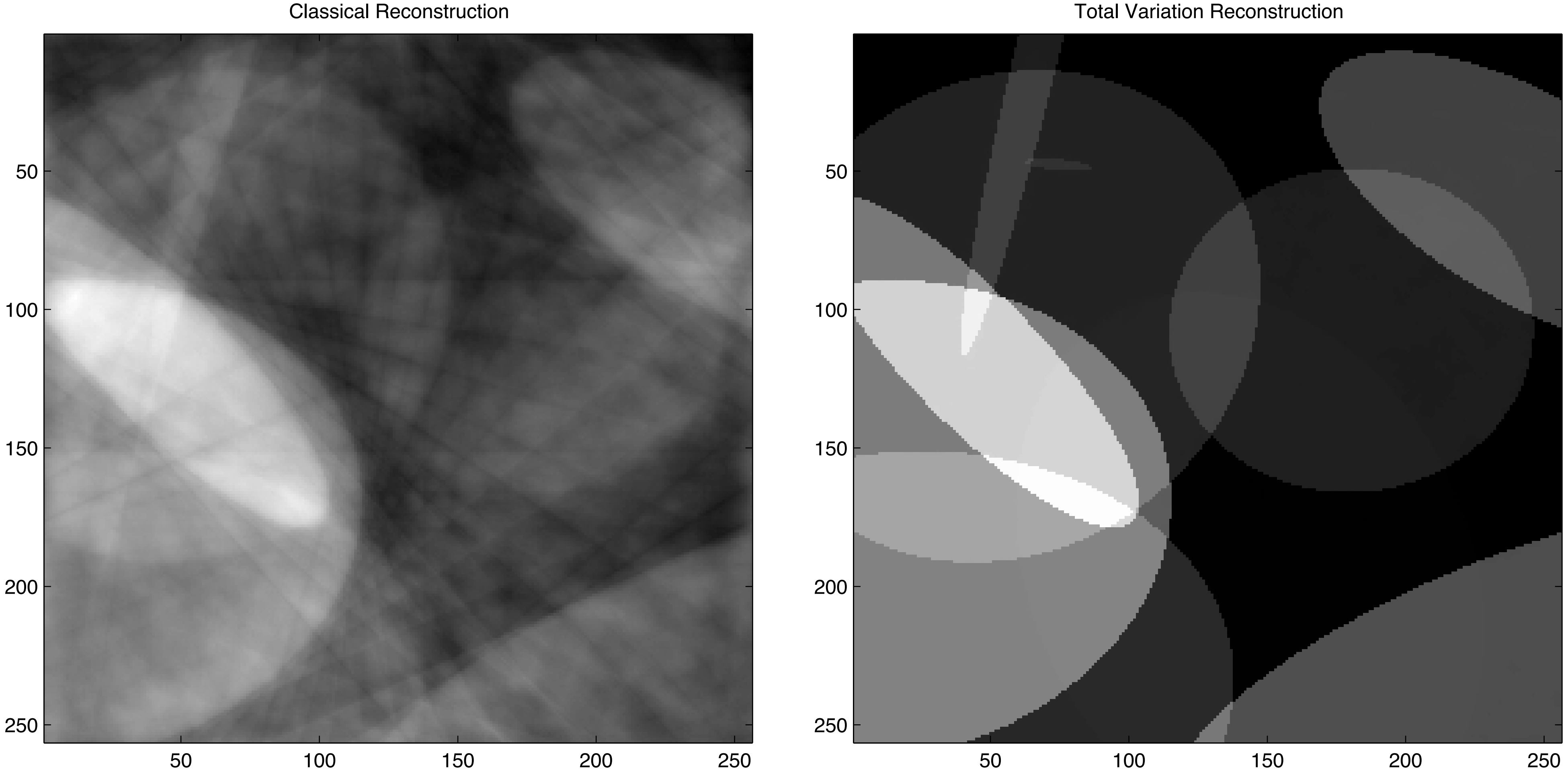


min ℓ_2



min TV – Exact!

Other Phantoms



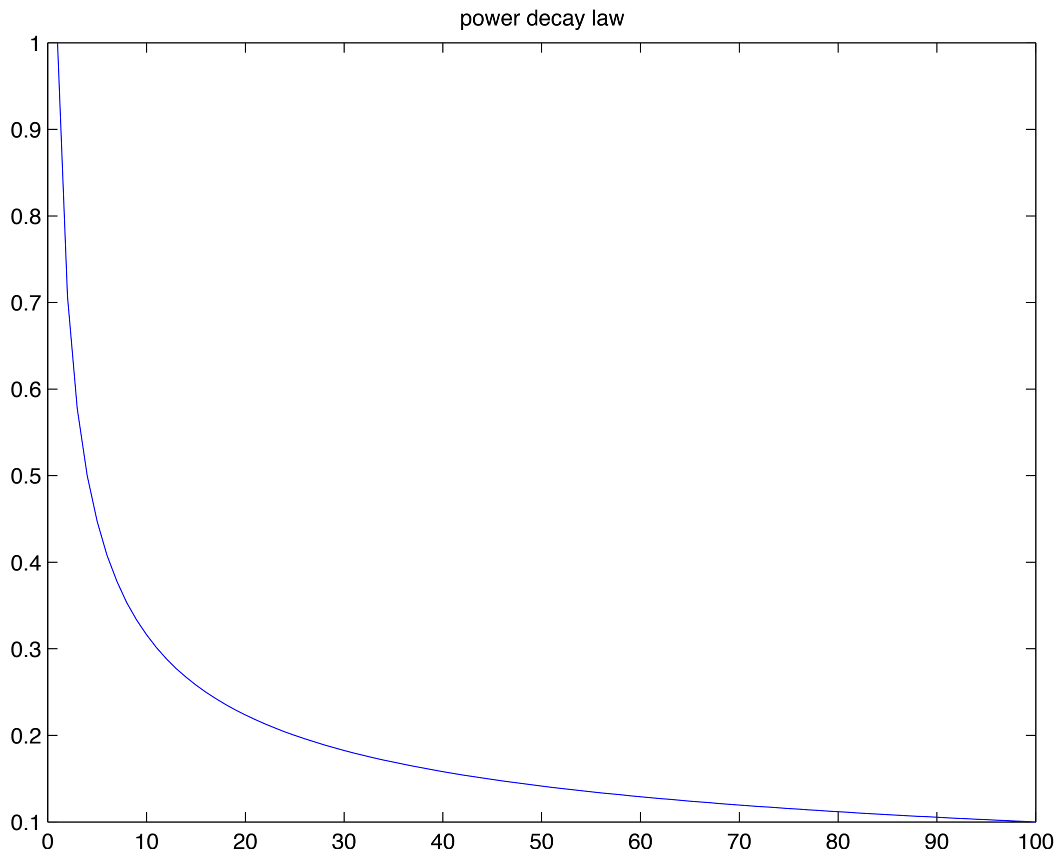
min $\boldsymbol{\ell_2}$ min TV – Exact!

Compressible Signals

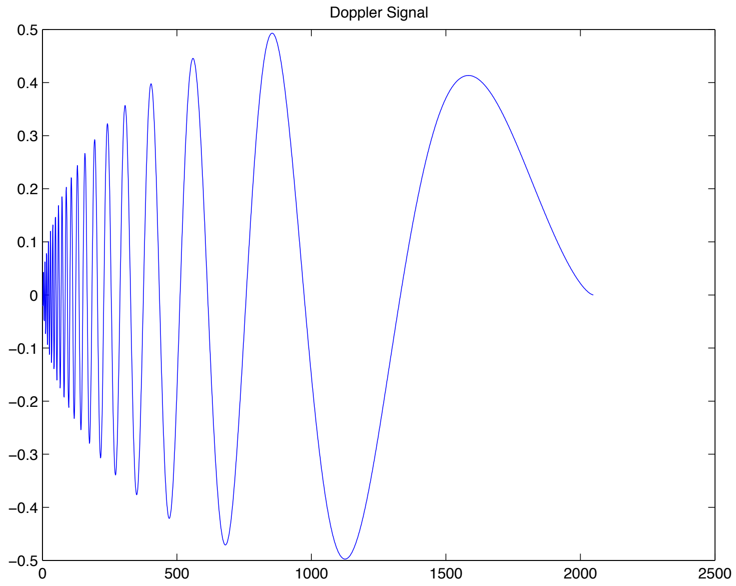
- In real life, signals are not sparse but most of them are compressible
- Compressible signals: rearrange the entries in decreasing order $|f|^2_{(1)} \geq |f|^2_{(2)} \geq \cdots \geq |f|^2_{(N)}$

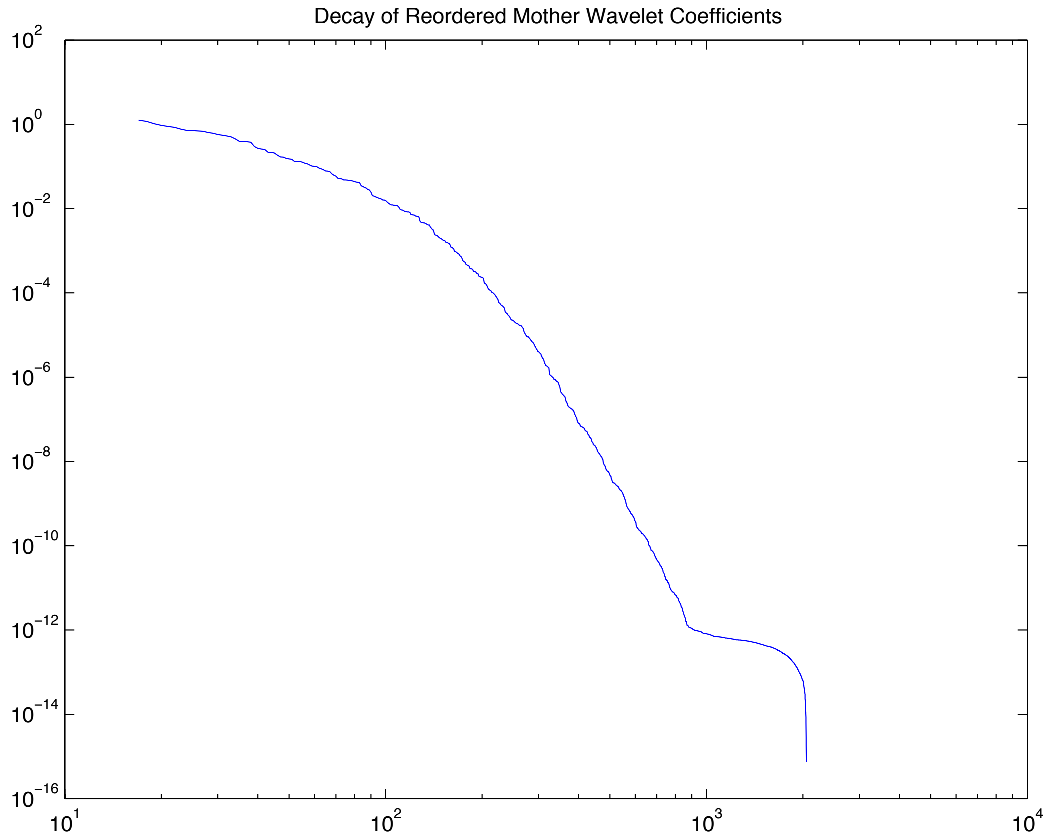
$$\mathcal{F}_p(C) = \{f : |f|_{(n)} \leq C n^{-1/p}, \forall n\}$$

- This is what makes transform coders work (sparse coding)

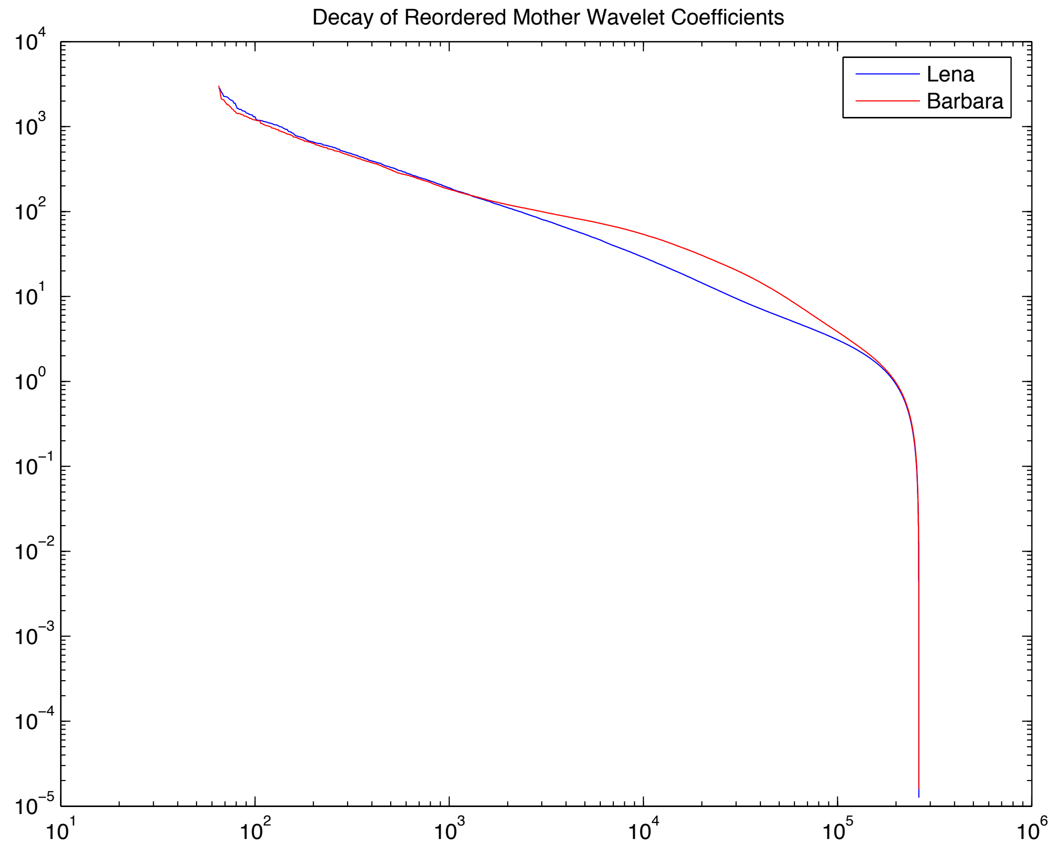


Compressible Signals I: Wavelets in 1D

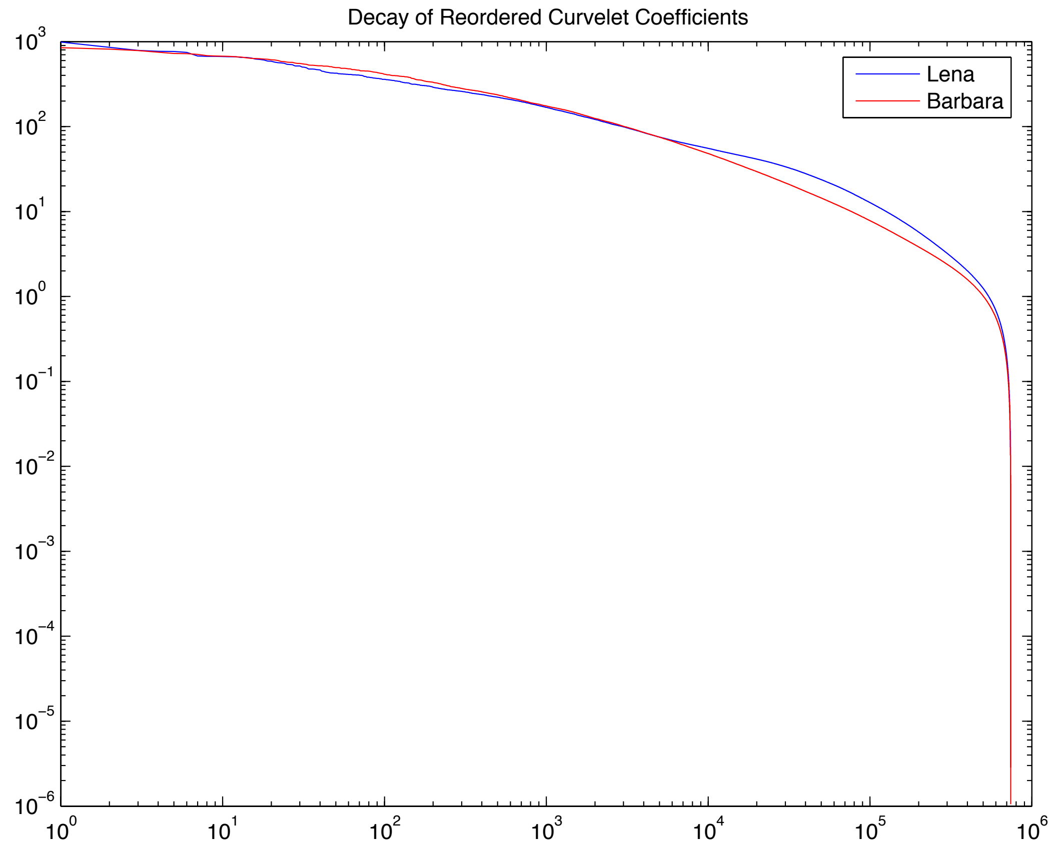




Compressible Signals II: Wavelets in 2D



Compressible Signals II: Curvelets



Examples of Compressible Signals

- *Smooth signals.* Continuous-time object has s bounded derivatives, then nth largest entry of the wavelet or Fourier coefficient sequence

$$|f|_{(n)} \leq \begin{cases} C \cdot n^{-s-1/2} & \text{1 dimension} \\ C \cdot n^{-s/d-1/2} & \text{d dimension} \end{cases}$$

- *Signals with bounded variations.* In 2 dimensions, the BV norm of a continuous time object is approximately

$$\|f\|_{BV} \approx \|\nabla f\|_{L_1}$$

  In the wavelet domain

$$|\theta(f)|_{(n)} \leq C \cdot n^{-1}.$$

- Many other examples: e.g. Gabor atoms and certain classes of oscillatory signals, curvelets and images with edges, etc.

Nonlinear Approximation of Compressible Signals

- $f \in \mathcal{F}_p(C)$, $|f|_{(n)} \leq C \cdot n^{-1/p}$
- Keep K-largest entries in $f \to f_K$

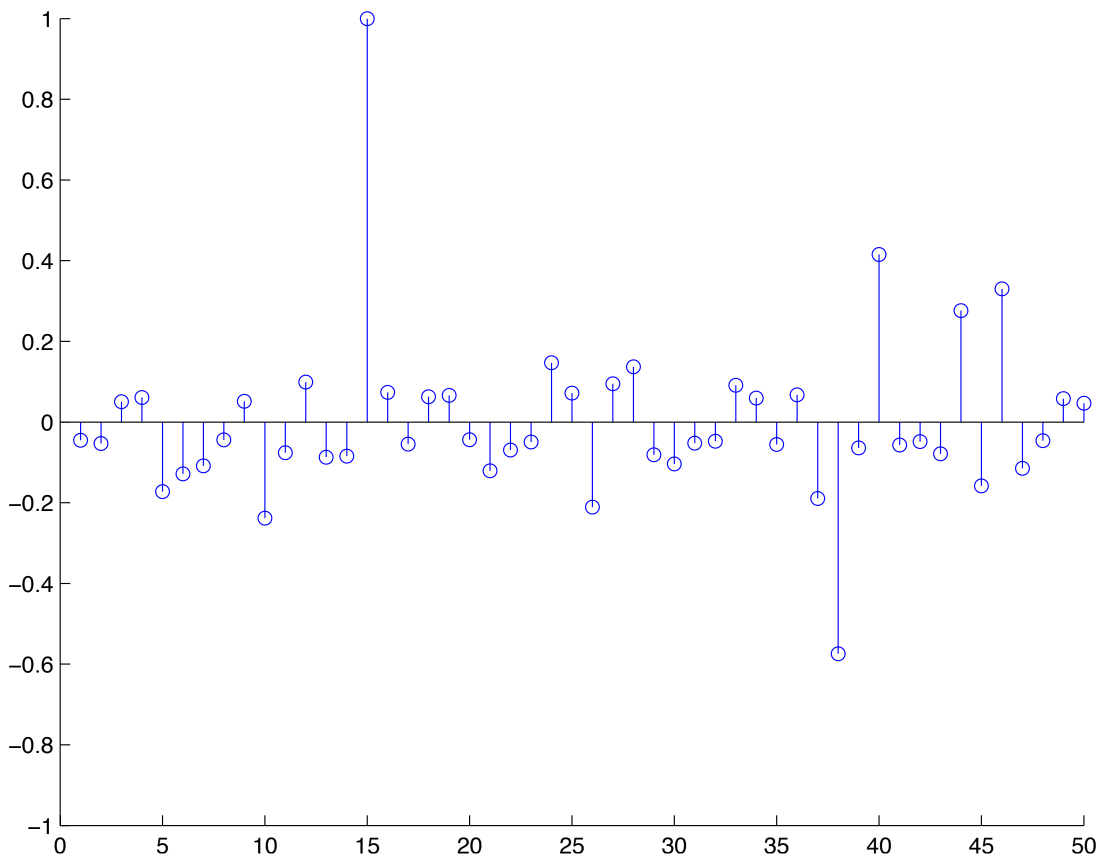
$$\|f - f_K\| \leq C \cdot K^{-r}, \quad r = 1/p - 1/2.$$

- E.g. $p = 1$, $\|f - f_K\| \leq C \cdot K^{-1/2}$.

Recovery of Compressible Signals

- How many measurements to recover f to within precision $\epsilon = K^{-r}$.
- Intuition: at least K, probably many more.

Where Are the Largest Coefficients?



Near Optimal Recovery of Compressible Signals

- Select K Gaussian random vectors (X_k), $k = 1, \ldots, K$

$$X_k \sim N(0, I_N)$$

- Observe $y_k = \langle f, X_k \rangle$
- Reconstruct by solving (P_1); minimize the ℓ_1-norm subject to constraints.

Theorem 3 (C., Tao) *Suppose that $f \in \mathcal{F}_p(C)$ for $0 < p \leq 1$ or $\|f\|_{\ell_1}$ for $p = 1$. Then with overwhelming probability,*

$$\|f^\# - f\|_2 = C \cdot (K/\log N)^{-r}.$$

See also Donoho (2004)

Big Surprise

Want to know an object up to an error ϵ; e.g. an object whose wavelet coefficients are sparse.

- *Strategy 1:* Oracle tells exactly (or you collect all N wavelet coefficients) which K coefficients are large and measure those

$$\|f - f_K\| \asymp \epsilon$$

- *Strategy 2:* Collect $K \log N$ random coefficients and reconstruct using ℓ_1.

Surprising claim

- Same performance but with only $K \log N$ coefficients!
- Performance is achieved by solving an LP.

Optimality

- Can you do with fewer than $K \log N$ for accuracy K^{-r}?
- Simple answer: NO (at least in the range $K << N$)

Optimality: Example

$f \in B_1 := \{f, \|f\|_{\ell_1} \leq 1\}$

- *Entropy numbers*: for a given set $\mathcal{F} \subset \mathrm{R}^N$, we $N(\mathcal{F}, r)$ is the smallest number of Euclidean balls of radius r which cover $\mathcal{F}$

$$e_k = \inf\{r > 0 : N(\mathcal{F}, r) \leq 2^{k-1}\}.$$

  Interpretation in coding theory: to encode a signal from $\mathcal{F}$ to within precision e_k, one would need at least k bits.

- Entropy estimates (Schütt (1984), Kühn (2001))

$$e_k \asymp \left(\frac{\log(N/k+1)}{k}\right)^{1/2}, \quad \log N \leq k \leq N.$$

  To encode an object f in the ℓ_1-ball to within precision $1/\sqrt{K}$ one would need to spend at least $O(K \log(N/K))$ bits. For $K \asymp N^\beta$, $\beta < 1$, $O(K \log N)$ bits.

Gelfand n-width (Optimality)

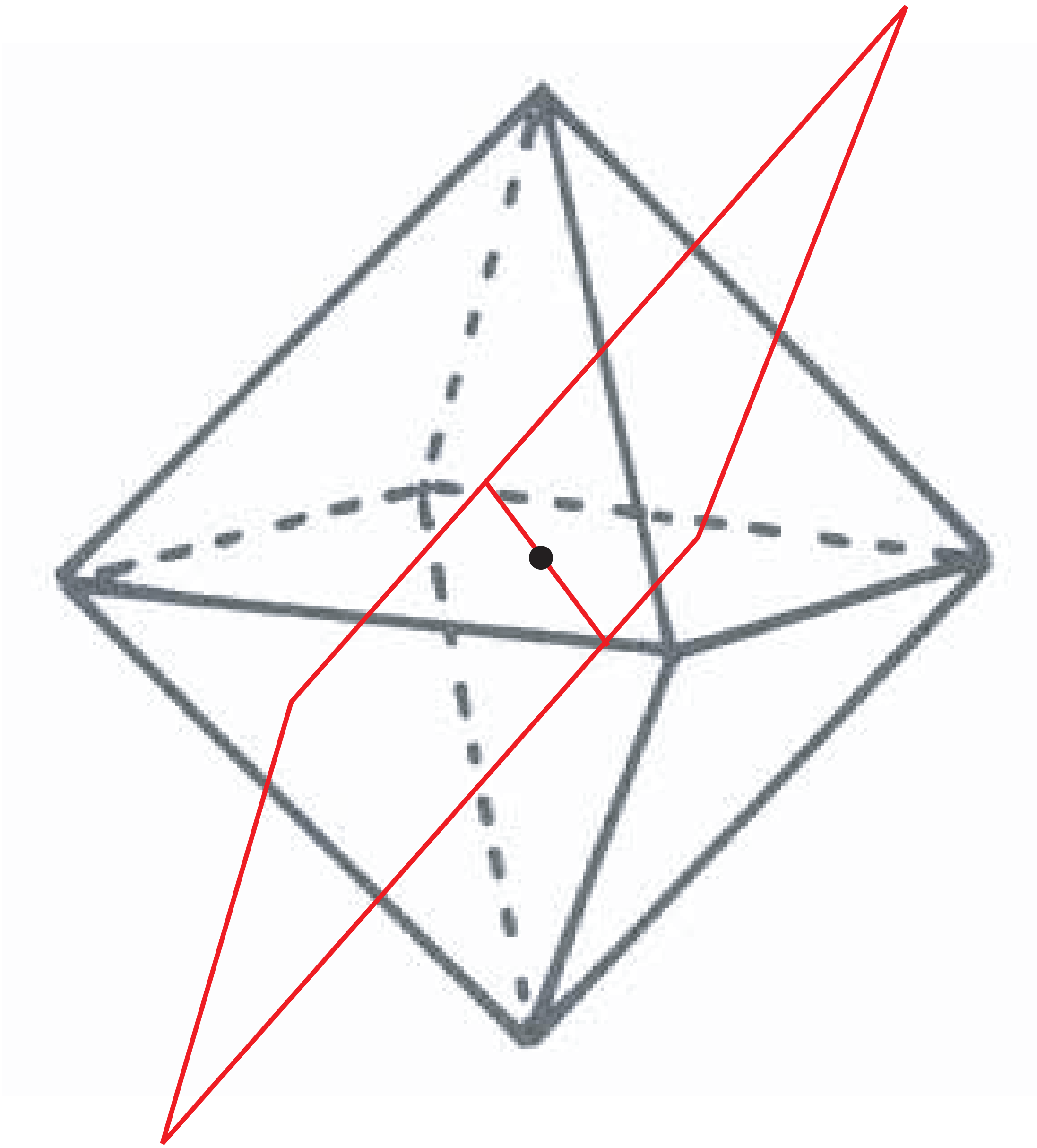
- k measurements Ff; this sets the constraint that f live on an affine space $f_0 + S$ where S is a linear subspace of co-dimension less or equal to k.
- *The data available for the problem cannot distinguish any object belonging to that plane.* For our problem, the data cannot distinguish between any two points in the intersection $B_1 \cap f_0 + S$. Therefore, any reconstruction procedure $f^*(y)$ based upon $y = F_\Omega f$ would obey

$$\sup_{f \in \mathcal{F}} \|f - f^*\| \geq \frac{\text{diam}(B_1 \cap S)}{2}.$$

- The *Gelfand* numbers of a set $\mathcal{F}$ are defined as

$$c_k = \inf_S \{\sup_{f \in \mathcal{F}} \|f_{|S}\| : \text{codim}(S) < k\},$$

Gelfand Width



Gelfand and entropy numbers (Optimality)

- Gelfand numbers dominate the entropy numbers (Carl, 1981)

$$\left(\frac{\log N/k}{k}\right)^{1/2} \asymp e_k \lessapprox c_k$$

- Therefore, for error $1/k$

$$k \log N/k \lessapprox \#\text{meas.}$$

- Similar argument for $\mathcal{F}_p$

Something Special about Gaussian Measurements?

- Works with the other measurement ensembles
- *Binary ensemble*: $F(k, t) = \pm 1$ with prob. $1/2$

$$\|f^\sharp - f\|_2 = C \cdot (K/\log N)^{-r}.$$

- *Fourier ensemble*:

$$\|f^\sharp - f\|_2 = C \cdot (K/\log^3 N)^{-r}.$$

Axiomatization I: Uniform Uncertainty Principle (UUP)

A measurement matrix F obeys the UUP with oversampling factor λ if for *all* subsets T such that

$$|T| \leq \alpha \cdot K/\lambda,$$

the matrix F_T obtained by extracting T columns obeys the bounds

$$\frac{1}{2} \cdot K/N \leq \lambda_{\min}(F_T^* F_T) \leq \lambda_{\max}(F_T^* F_T) \leq \frac{3}{2} \cdot K/N$$

This must be true w.p. at least $1 - O(N^{-\rho/\alpha})$

UUP: Interpretation

W. Heisenberg

- Suppose $\boldsymbol{F}$ is the randomly sampled DFT, $\boldsymbol{\Omega}$ set of random frequencies.
- Signal $\boldsymbol{f}$ with support $\boldsymbol{T}$ obeying

$$|\boldsymbol{T}| \leq \boldsymbol{\alpha K / \lambda}$$

- UUP says that with overwhelming probability

$$\|\hat{\boldsymbol{f}}_{|\boldsymbol{\Omega}}\| \leq \sqrt{\boldsymbol{3K/2N}} \|\boldsymbol{f}\|$$

- No concentration is possible unless $\boldsymbol{K} \asymp \boldsymbol{N}$
- "Uniform" because must hold for all such $\boldsymbol{T}$'s

Axiomatization II: Exact Reconstruction Principle (ERP)

A measurement matrix $\boldsymbol{F}$ obeys the ERP with oversampling factor $\boldsymbol{\lambda}$ if for each fixed subset $\boldsymbol{T}$

$$|\boldsymbol{T}| \leq \boldsymbol{\alpha} \cdot \boldsymbol{K}/\boldsymbol{\lambda},$$

and each 'sign' vector $\boldsymbol{\sigma}$ defined on $\boldsymbol{T}$, $|\boldsymbol{\sigma}(\boldsymbol{t})| = \mathbf{1}$, there exists $\boldsymbol{P} \in \mathbf{R}^{\boldsymbol{N}}$ s.t.

(i) $\boldsymbol{P}$ is in the row space of $\boldsymbol{F}$

(ii) $\boldsymbol{P}(\boldsymbol{t}) = \boldsymbol{\sigma}(\boldsymbol{t})$, for all $\boldsymbol{t} \in \boldsymbol{T}$;

(iii) and $|\boldsymbol{P}(\boldsymbol{t})| \leq \frac{1}{2}$ for all $\boldsymbol{t} \in \boldsymbol{T}^{\boldsymbol{c}}$

Interpretation: Gives exact reconstruction for sparse signals

Near-optimal Recovery Theorem [C., Tao]

- Measurement ensemble obeys UUP with oversampling factor $\boldsymbol{\lambda_1}$
- Measurement ensemble obeys ERP with oversampling factor $\boldsymbol{\lambda_2}$
- Object $\boldsymbol{f} \in \boldsymbol{\mathcal{F}_p(C)}$.

$$\|\boldsymbol{f} - \boldsymbol{f}^\sharp\| \leq \boldsymbol{C} \cdot (\boldsymbol{K}/\boldsymbol{\lambda})^{-\boldsymbol{r}}, \qquad \boldsymbol{\lambda} = \mathbf{max}(\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}).$$

UUP for the Gaussian Ensemble

- $F(k,t) = X_{k,t}/\sqrt{N}$
- Singular values of random Gaussian matrices

$$(1-\sqrt{c}) \cdot \sqrt{\frac{K}{N}} \lessapprox \sigma_{\min}(F_T) \leq \sigma_{\max}(F_T) \lessapprox (1+\sqrt{c}) \cdot \sqrt{\frac{K}{N}}$$

  with overwhelming probability (exceeding $1 - e^{-\beta K}$).

  Reference, S. J. Szarek, Condition numbers of random matrices, *J. Complexity* (1991),

  See also Ledoux (2001), Johnstone (2002), El-Karoui (2004)

- Marchenko-Pastur law: $c = |T|/K$.
- Union bound give result for all T provided

$$|T| \leq \gamma \cdot K/(\log N/K).$$

ERP for the Gaussian Ensemble

$$P = F^* F_T (F_T^* F_T)^{-1} \text{sgn}(f) := F^* V.$$

- P is in the row space of F
- P agrees with sgn(f) on T

$$P_{|T} = F_T^* F_T (F_T^* F_T)^{-1} \text{sgn}(f) = \text{sgn}(f)$$

- On the complement of T: $P_{|T^c} = F_{T^c}^* V$.

$$P(t) \frac{1}{\sqrt{N}} \sum_{k=1}^{K} X_{k,t} V_k.$$

- By independence of $F_{T^c}^*$ and V, conditional distribution

$$\mathcal{L}(P(t)|V) \sim N(0, \|V\|^2/N)$$

- With overwhelming probability (UUP)

$$\|V\| \leq \|F_T(F_T^* F_T)^{-1}\| \cdot \|\text{sgn}(f)\| \leq \sqrt{6N/K} \cdot \sqrt{|T|}$$

so that

$$\mathcal{L}(P(t)|V) \sim N(0, \sigma^2). \quad \sigma^2 \leq 6|T|/K$$

- In conlusion: for $t \in T^c$

$$P(P(t) > 1/2) \leq e^{-\beta K/|T|}.$$

- ERP holds if

$$|K| \geq \alpha \cdot |T| \cdot \log N.$$

Universal Codes

Want to compress sparse signals

- *Encoder*. To encode a discrete signal $\boldsymbol{f}$, the encoder simply calculates the coefficients $\boldsymbol{y_k = \langle f, X_k \rangle}$ and quantizes the vector $\boldsymbol{y}$.
- *Decoder*. The decoder then receives the quantized values and reconstructs a signal by solving the linear program $(\boldsymbol{P_1})$.

Conjecture Asymptotically nearly achieves the information theoretic limit.

Information Theoretic Limit: Example

- Want to encode the unit-ℓ_1 ball: $f \in \mathbf{R}^N : \sum_t |f(t)| \leq 1$.
- Want to achieve distortion D

$$\|f - f^\sharp\|^2 \leq D$$

- How many bits? Lower bounded by entropy of the unit-ℓ_1 ball:

$$\# \text{ bits } \geq C \cdot D \cdot (\log(N/D) + 1)$$

- Same as number of measured coefficients

Robustness

- Say with K coefficients

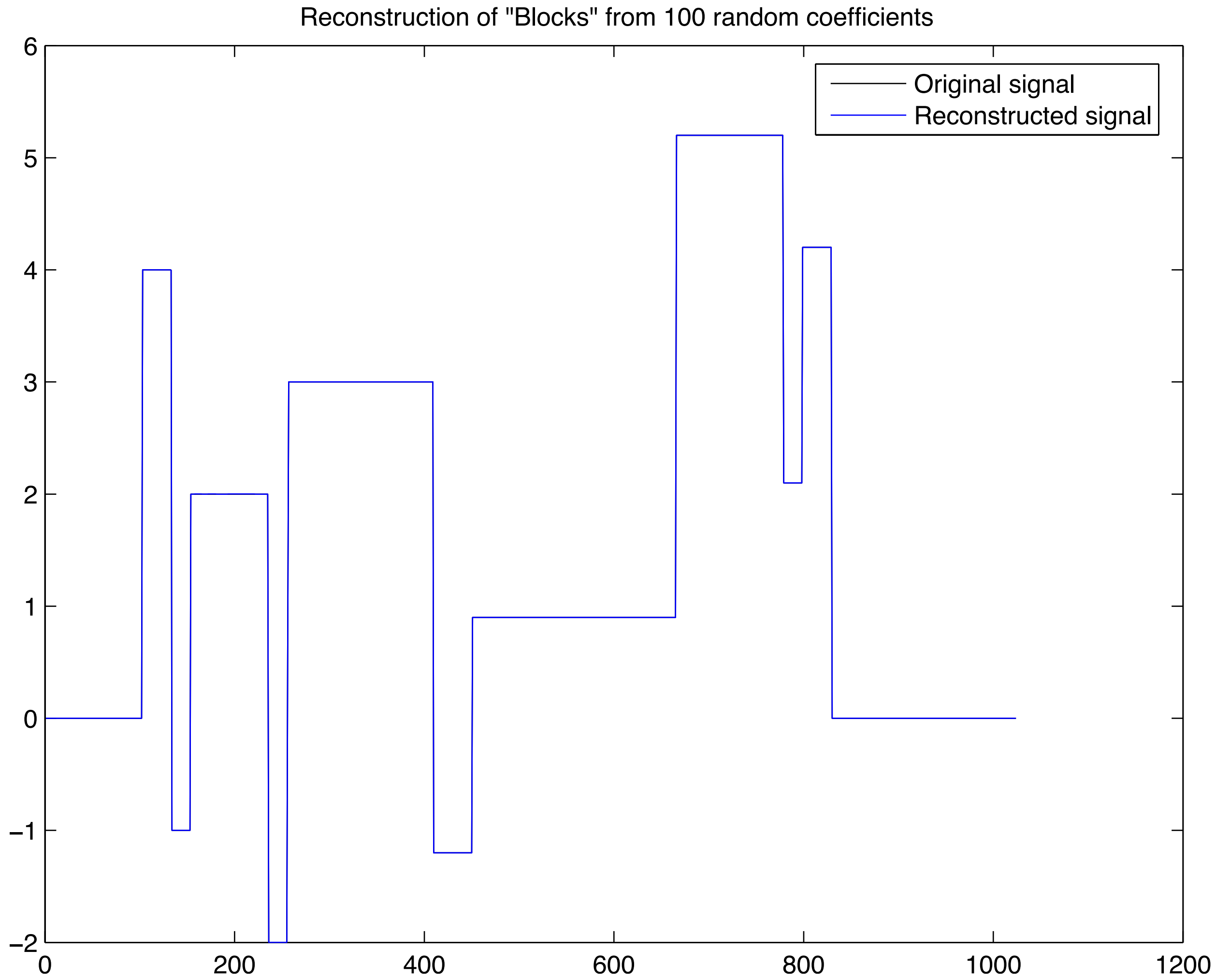
$$\|f - f^\sharp\|^2 \asymp 1/K$$

- Say we loose half of the bits (packet loss). How bad is the reconstruction?

$$\|f - f^\sharp_{50\%}\|^2 \asymp 2/K$$

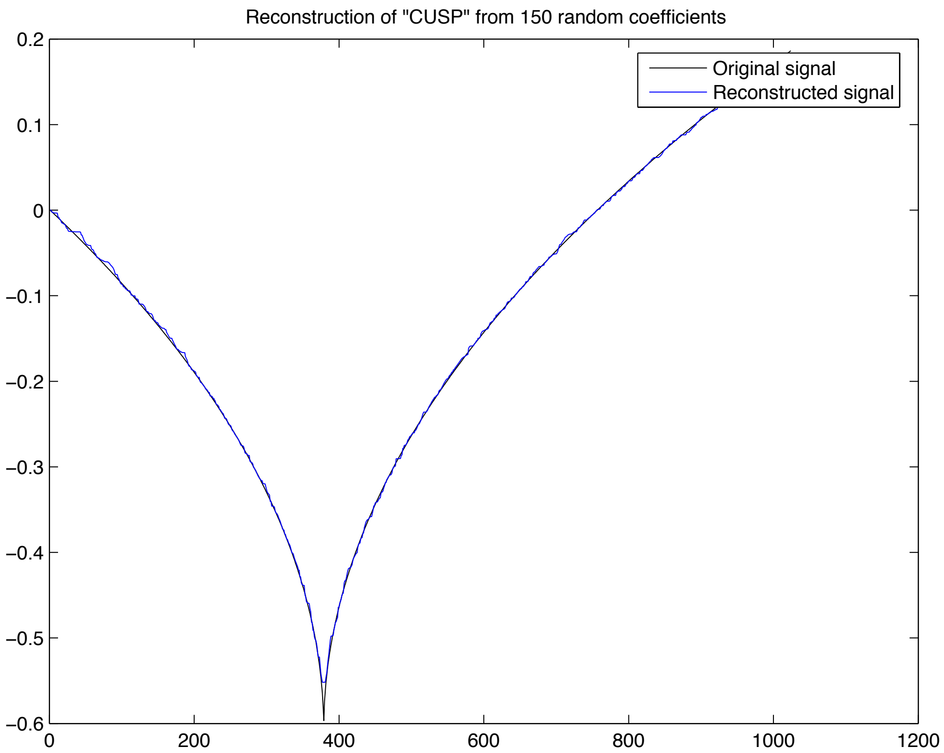
- Democratic!

Reconstruction from 100 Random Coefficients

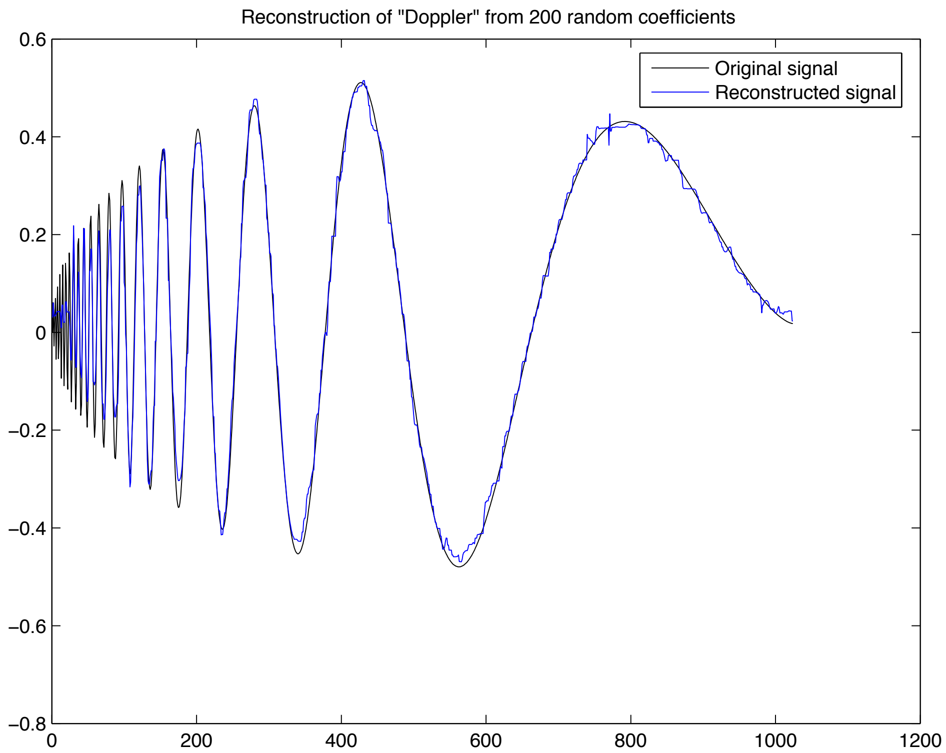


'Compressed' signal

Reconstruction from Random Coefficients

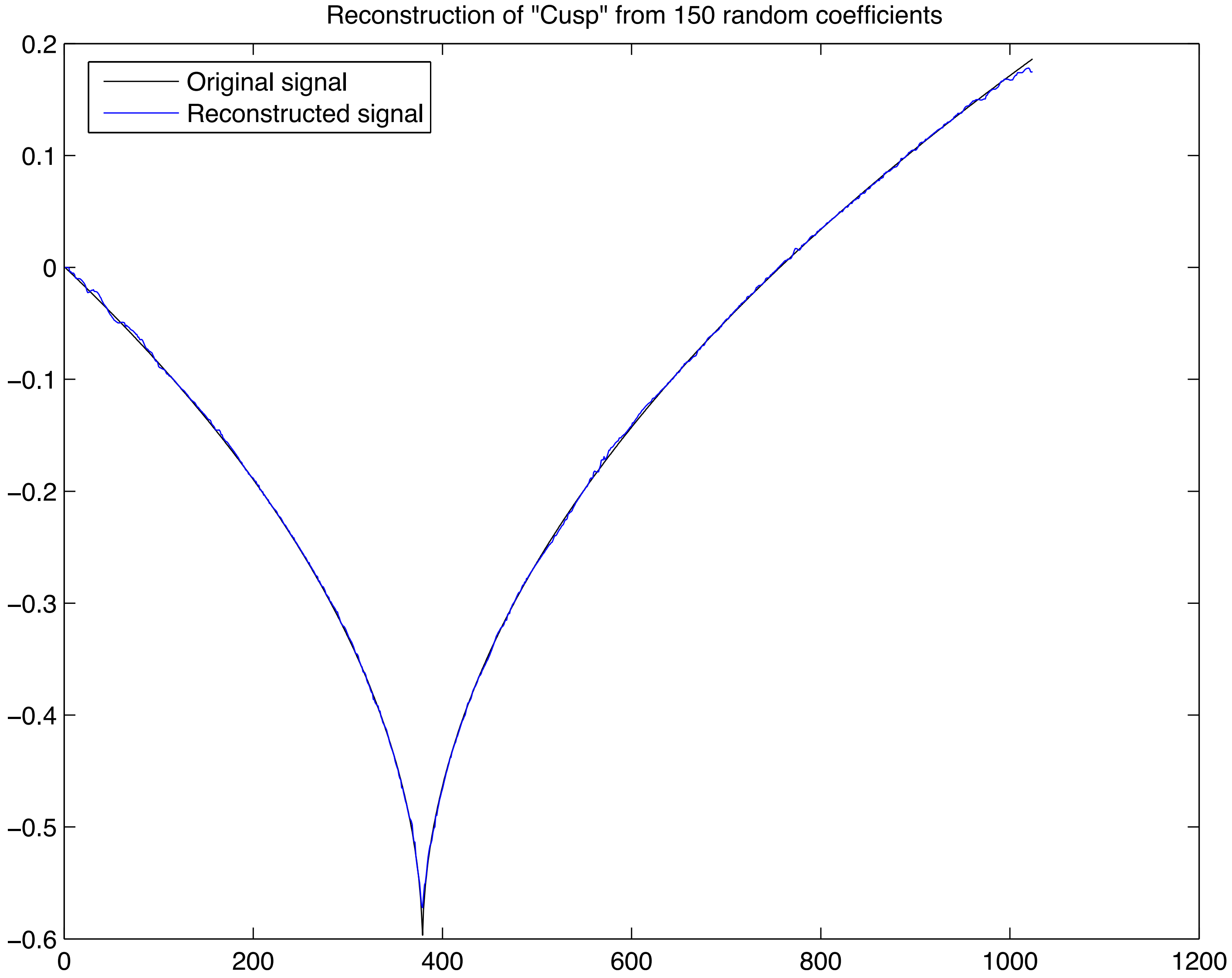


Cusp from 150 coeffs.



Doppler from 200 coeffs.

Reconstruction from Random Coefficients (Method II)



Summary

- Possible to reconstruct a compressible signal from a few measurements only
- Need to randomize measurements
- Need to solve an LP
- This strategy is nearly optimal
- Many applications