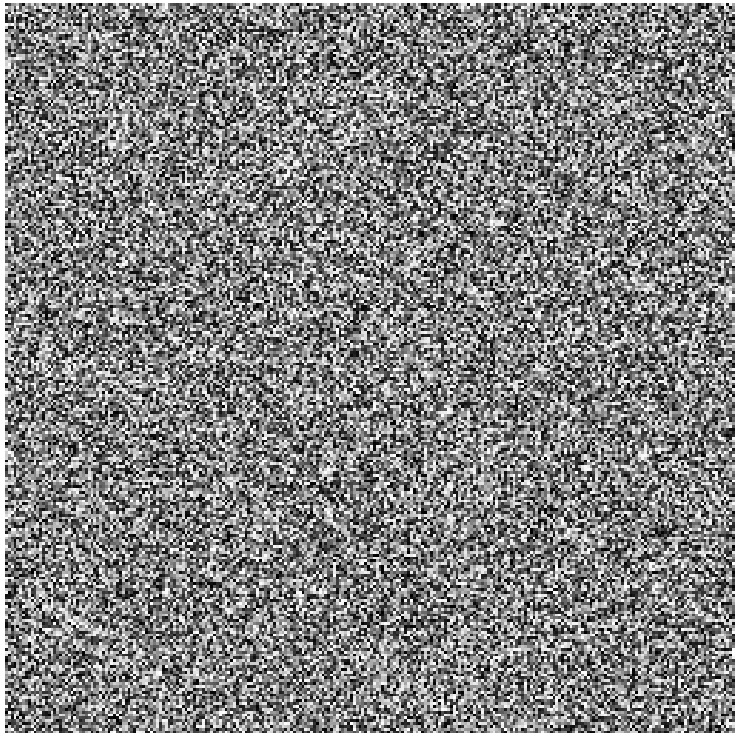
Some topics on Sparsity

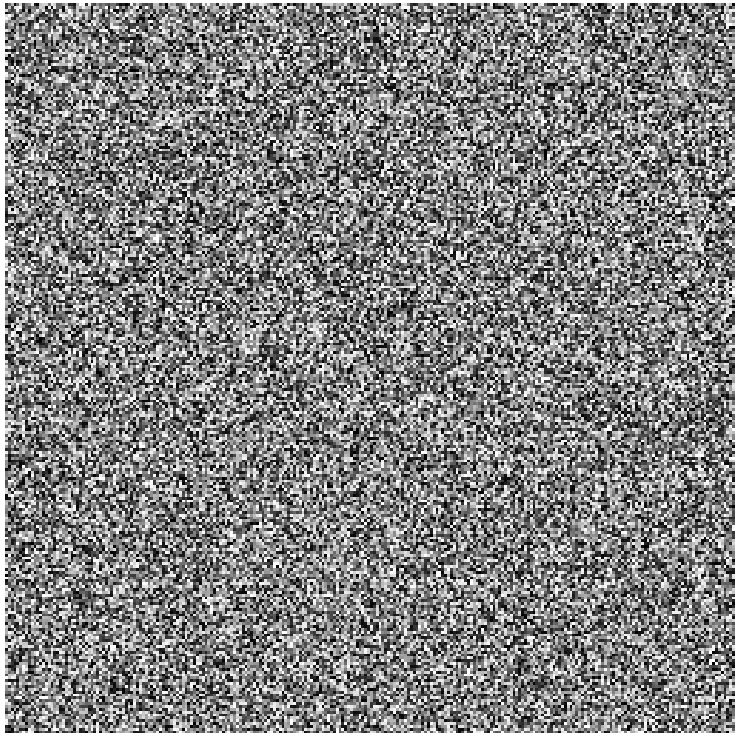
Sangnam Nam

August 12, 2010

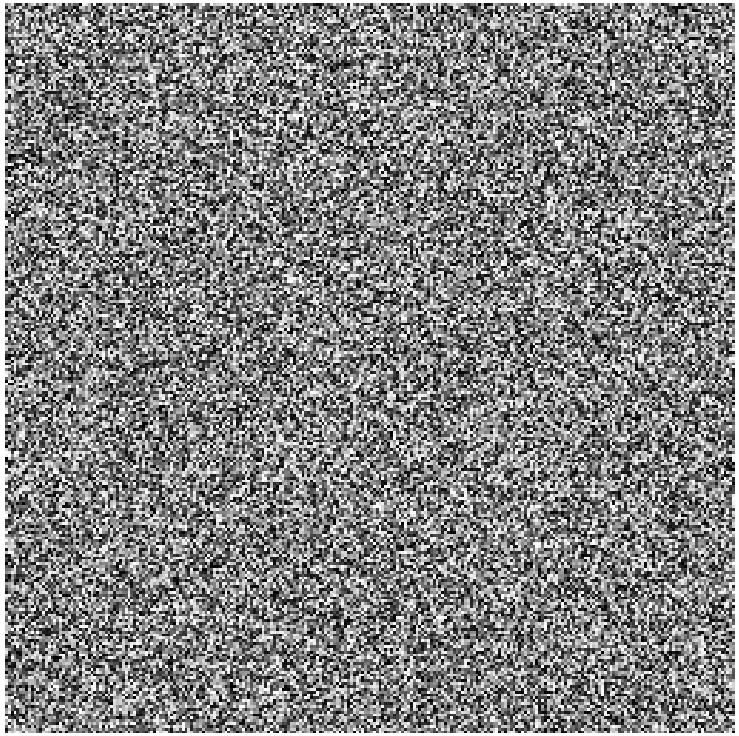
Why Sparsity?: randomly generated image



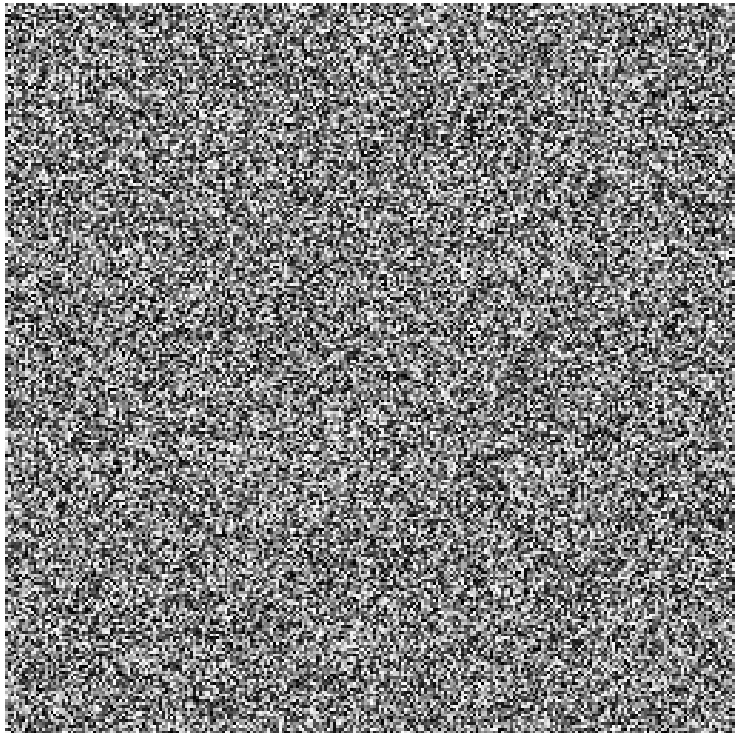
Why Sparsity?: randomly generated image



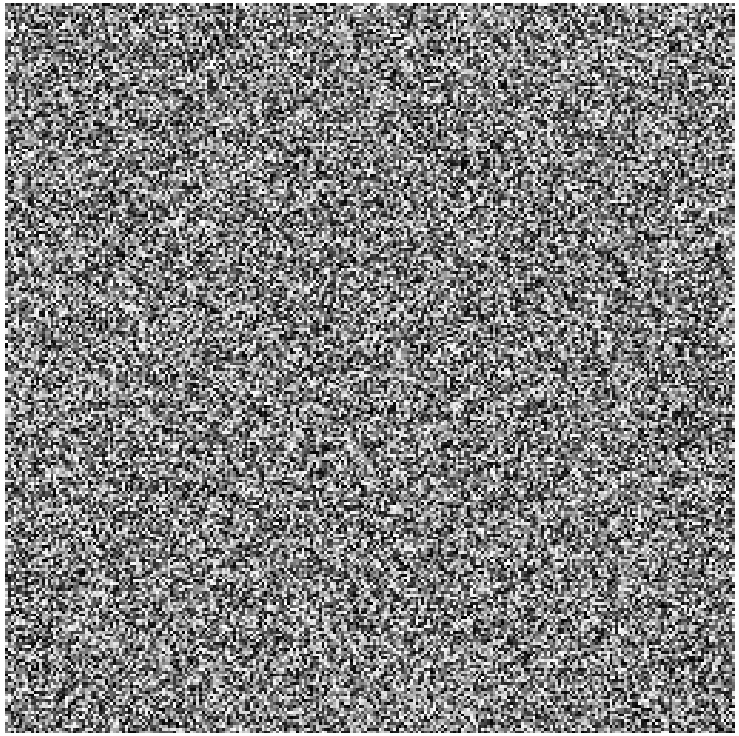
Why Sparsity?: randomly generated image



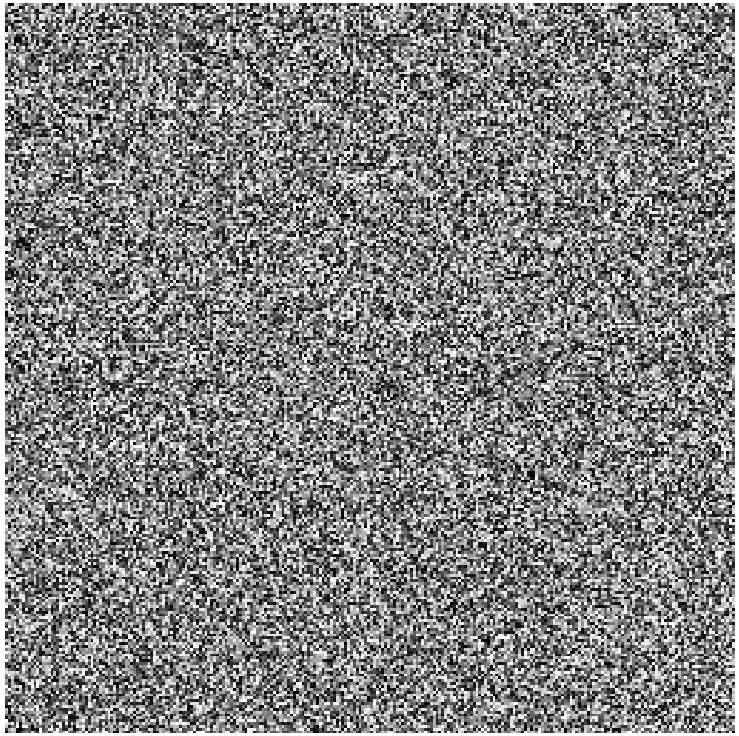
Why Sparsity?: randomly generated image



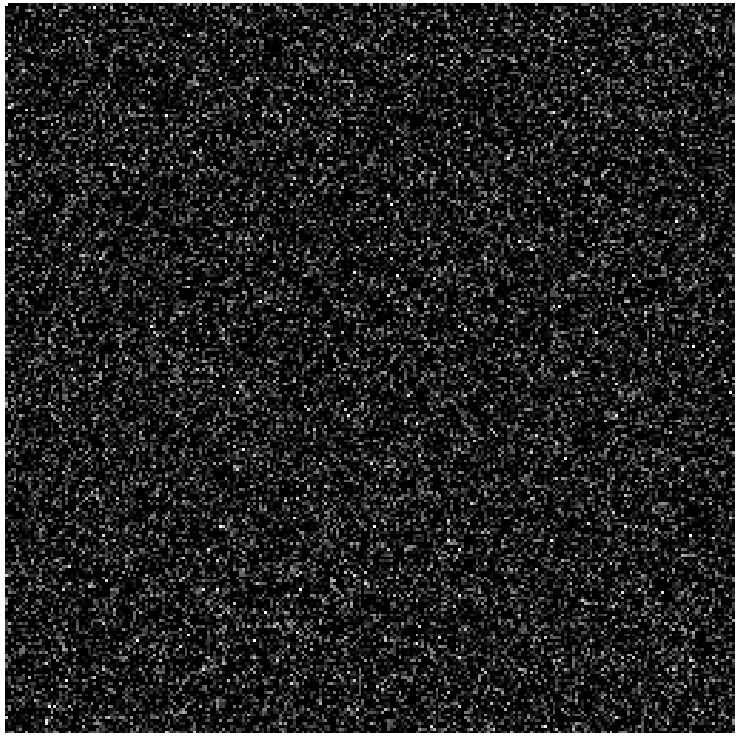
Why Sparsity?: randomly generated image



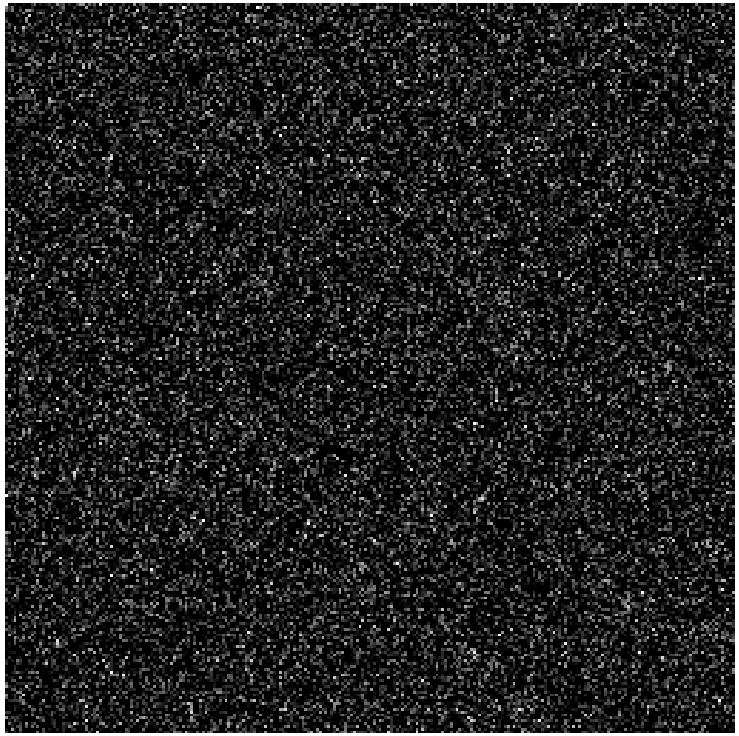
Why Sparsity?: randomly generated image



Why Sparsity?: DCT of randomly generated image

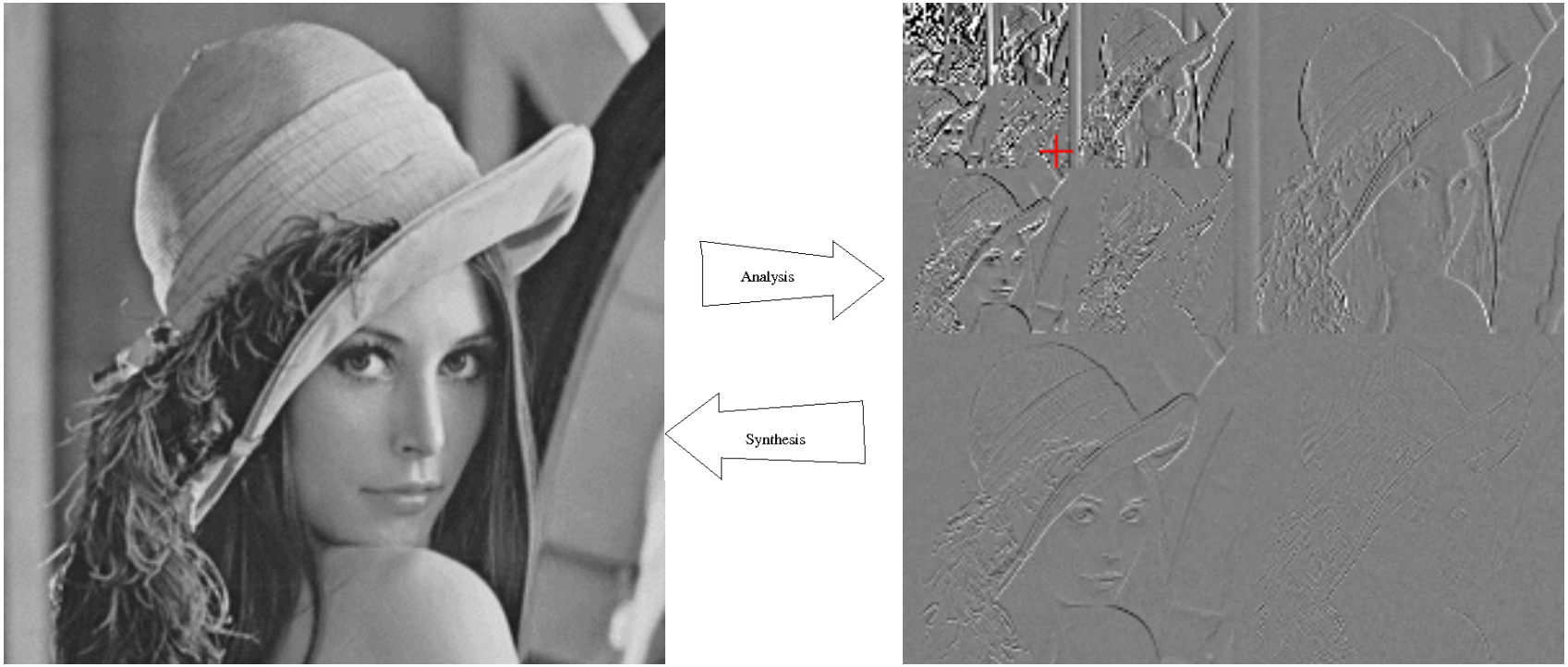


Why Sparsity?: DCT of randomly generated image

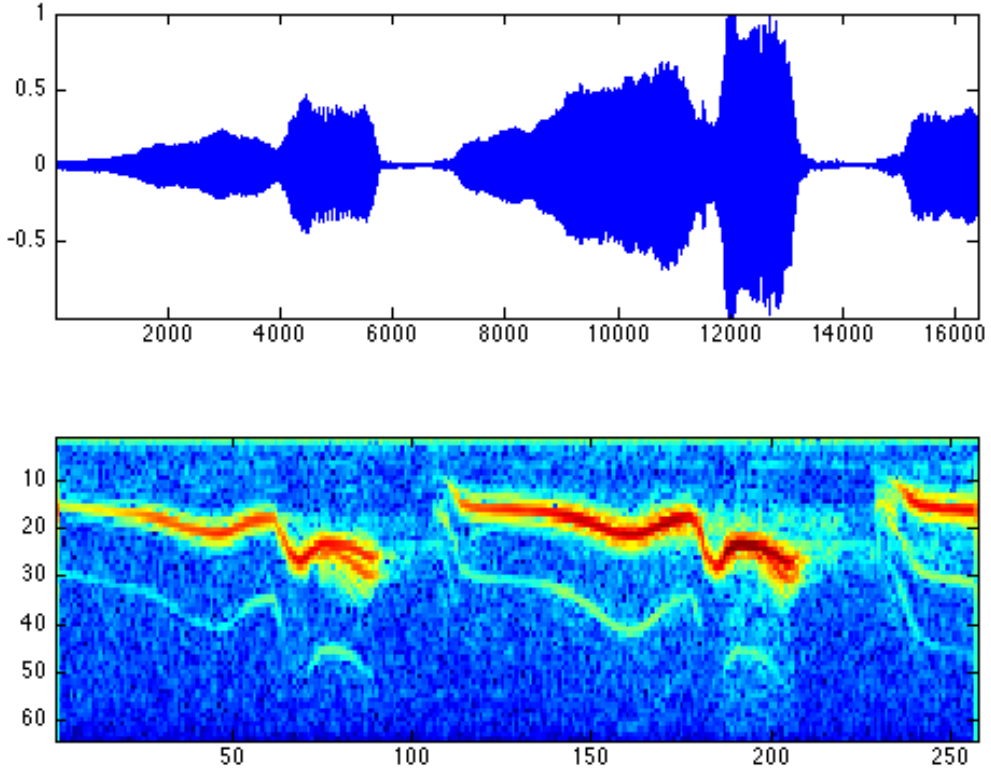


Compression: Image

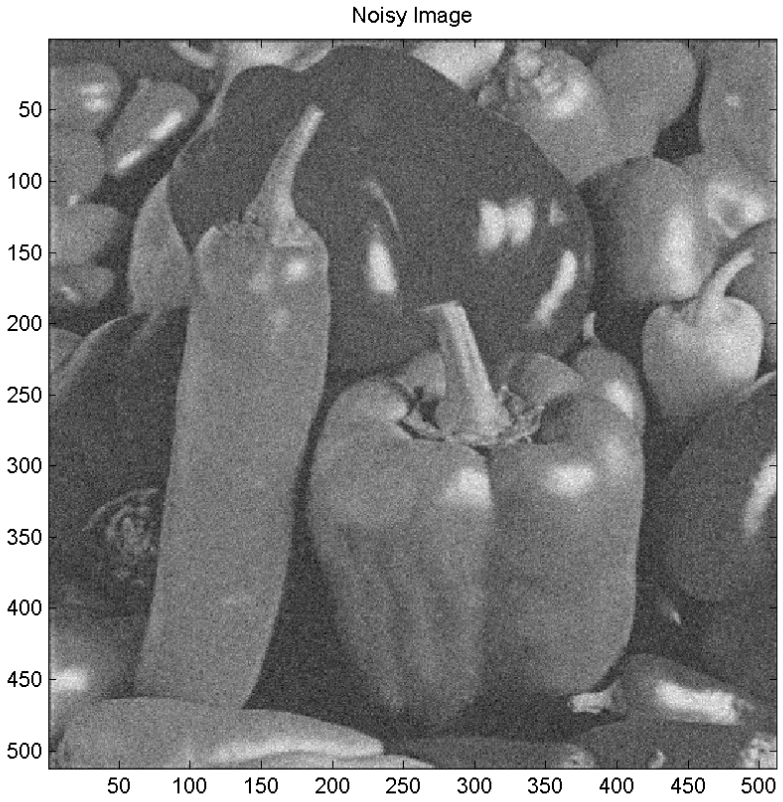
Compression: Image



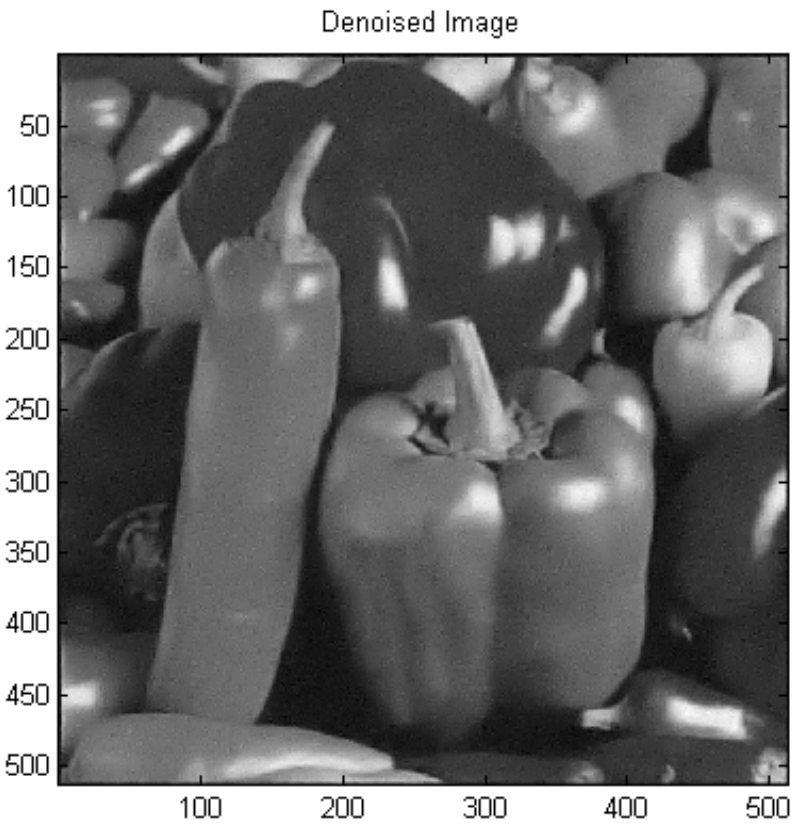
Compression: Audio



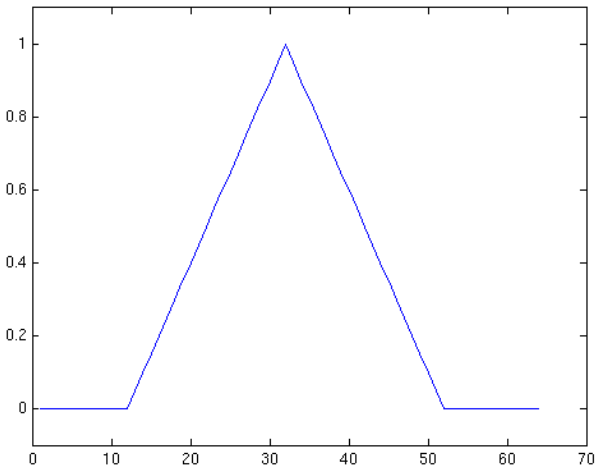
Denoising


Noisy Image
50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500

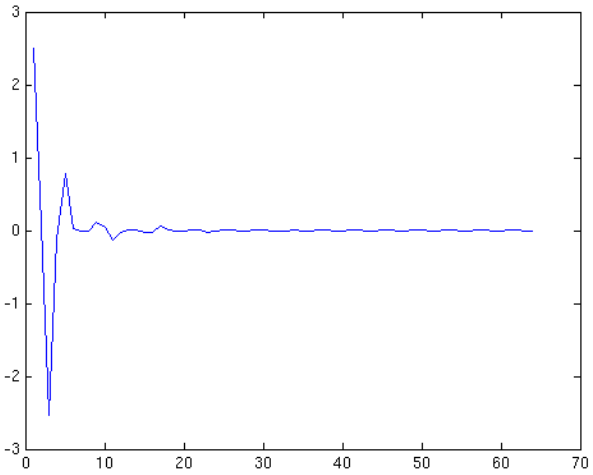
Denoising


Denoised Image
50
100
150
200
250
300
350
400
450
500
100
200
300
400
500

Denoising: Simple example

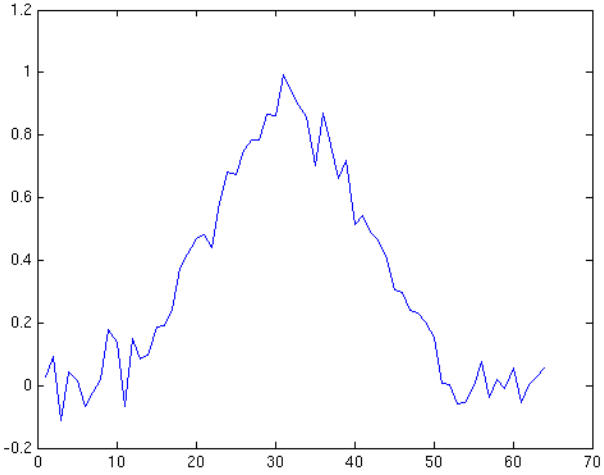


signal x

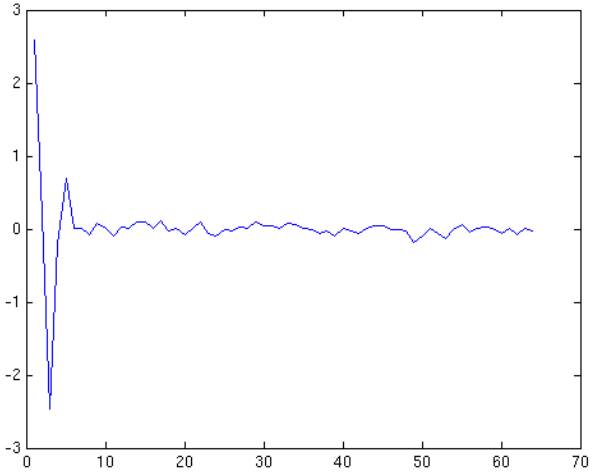


dct transform of x

Denoising: Simple example

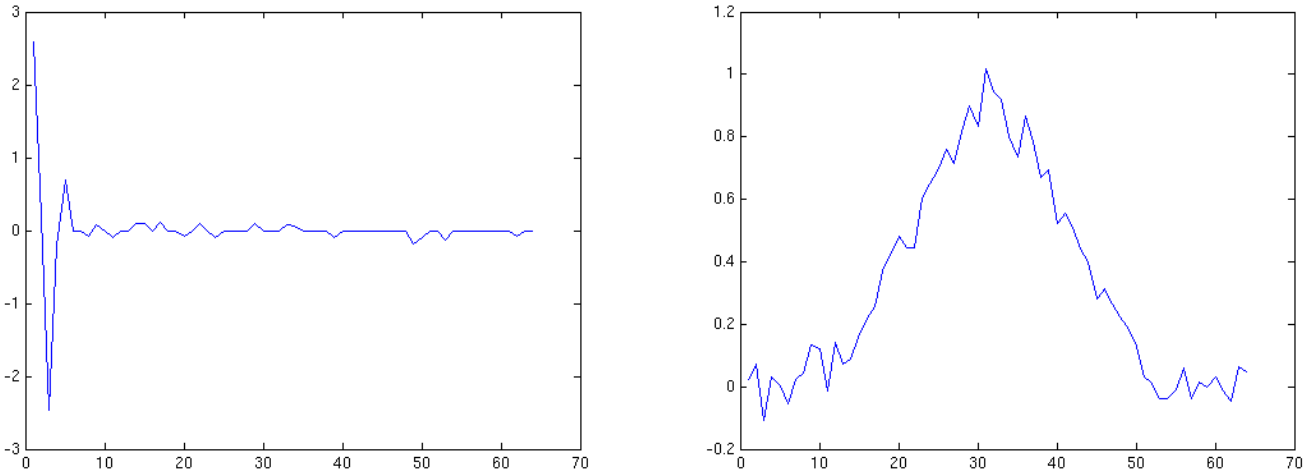


signal + noise



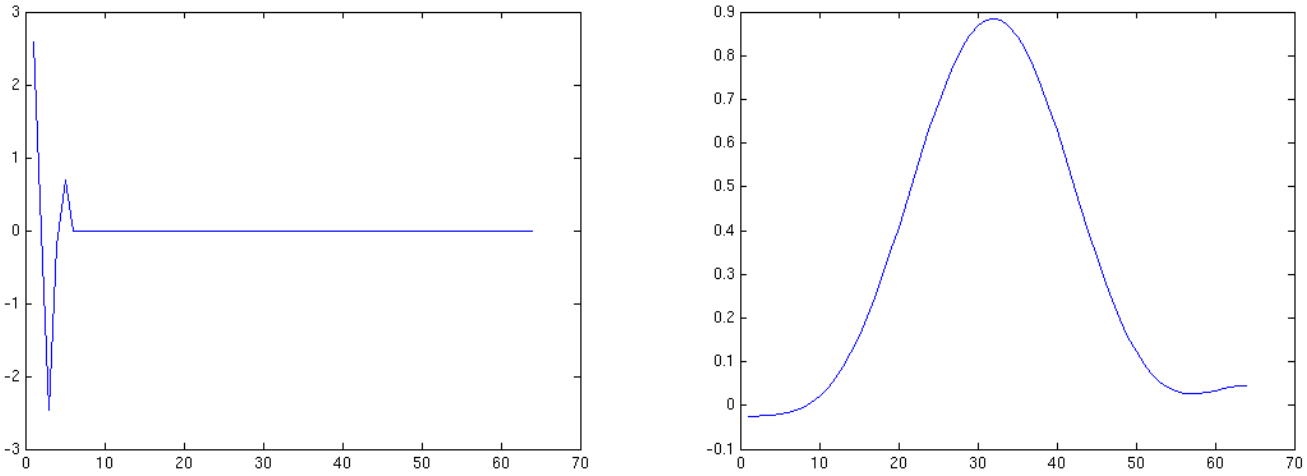
dct transform of (signal + noise)

Denoising: Simple example



Denoising with hard thresholding at 0.07

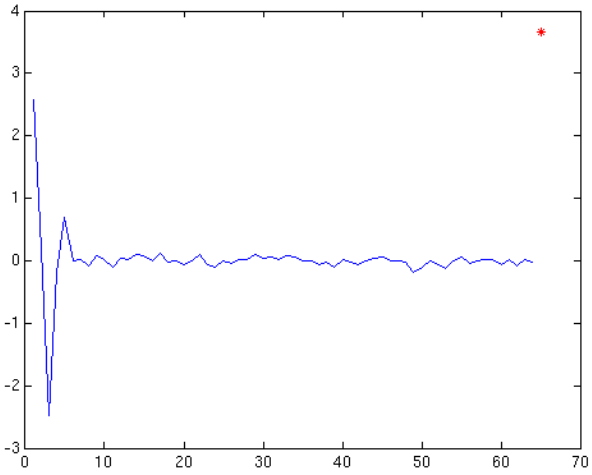
Denoising: Simple example



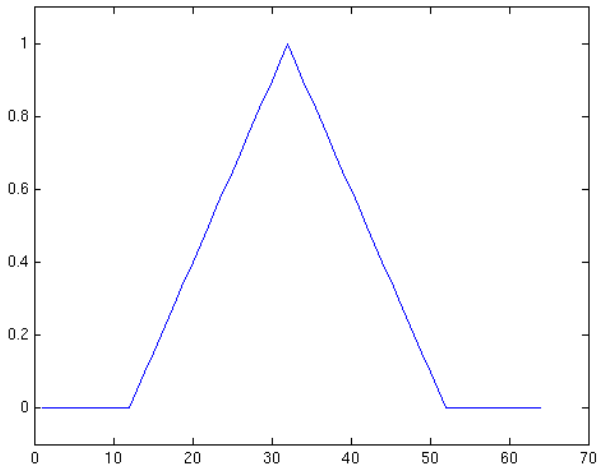
Denoising with hard thresholding at 0.18

Denoising: Simple example

With augmented dictionary:



Transform coefficients



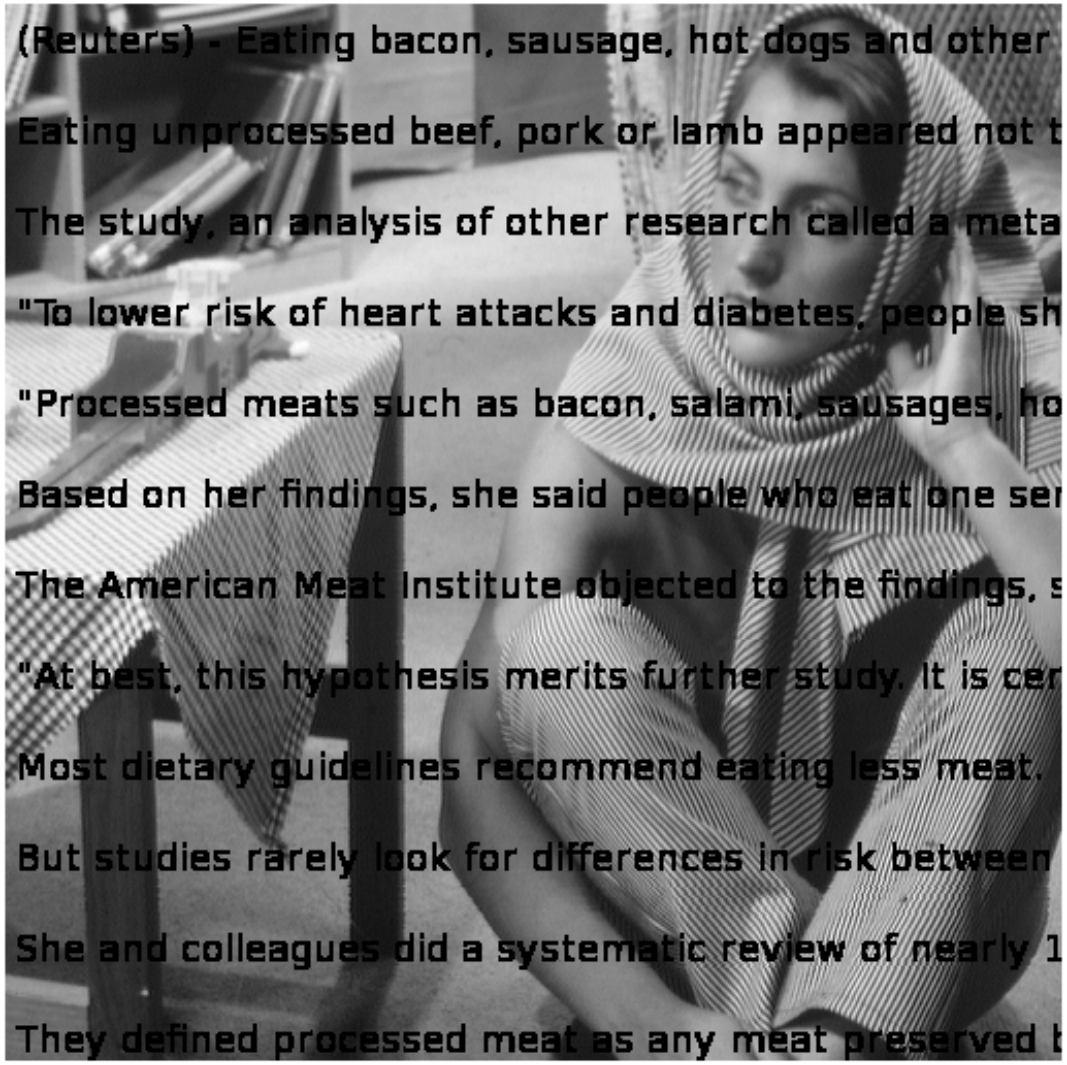
Denoised Result

Inpainting: Original Image

original image512x512=262144

Inpainting: letter

Given image512x512=262144



Inpainting: letter

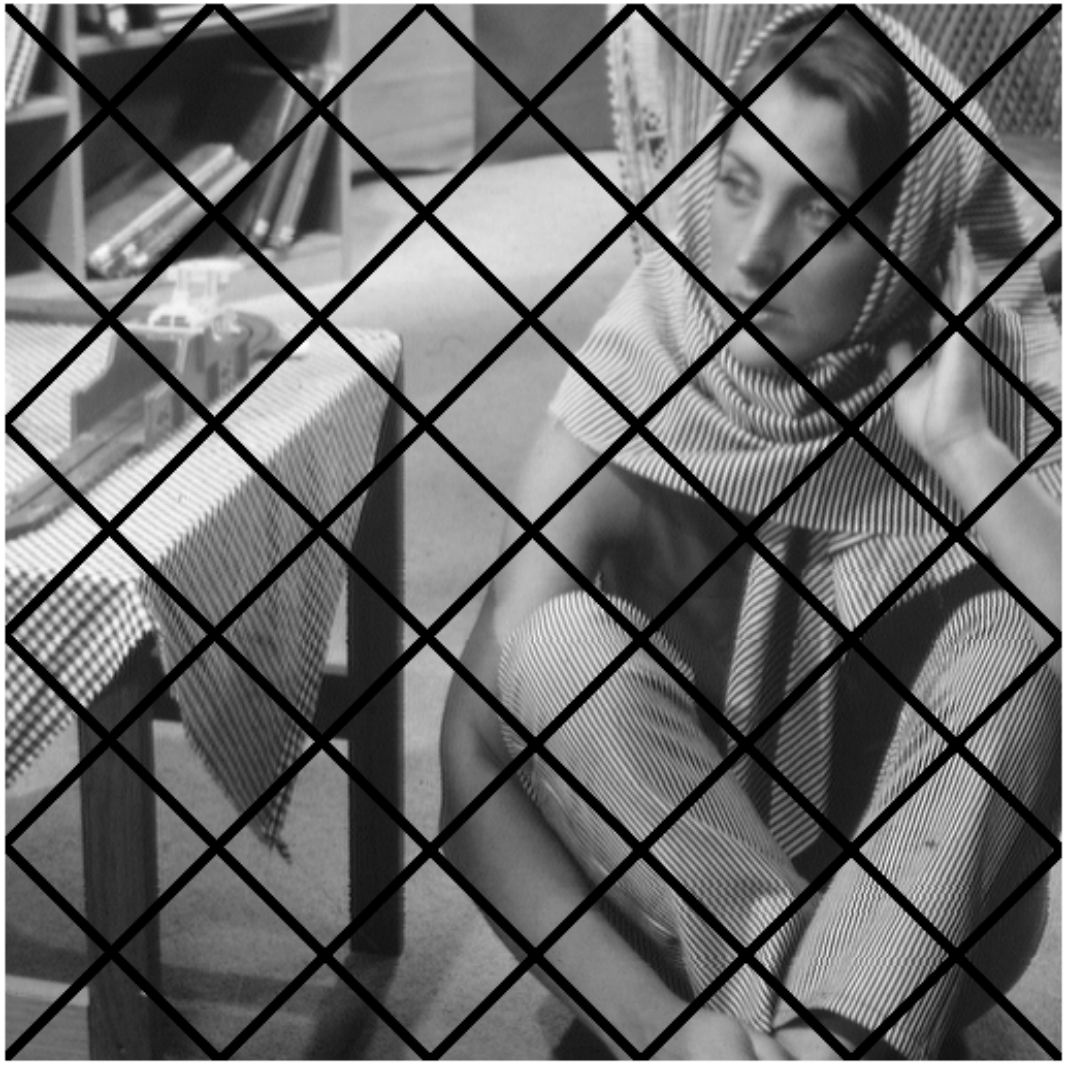
filled-in image: snr=25.2543

Inpainting: letter

reconstructed image: snr=34.3195

Inpainting: lines

Given image512x512=262144



Inpainting: lines

filled-in image: snr=23.5627

Inpainting: lines

reconstructed image: snr=30.8017

Inpainting: random

Given image512x512=262144,77.8% missing

Inpainting: random

filled-in image: snr=16.6297

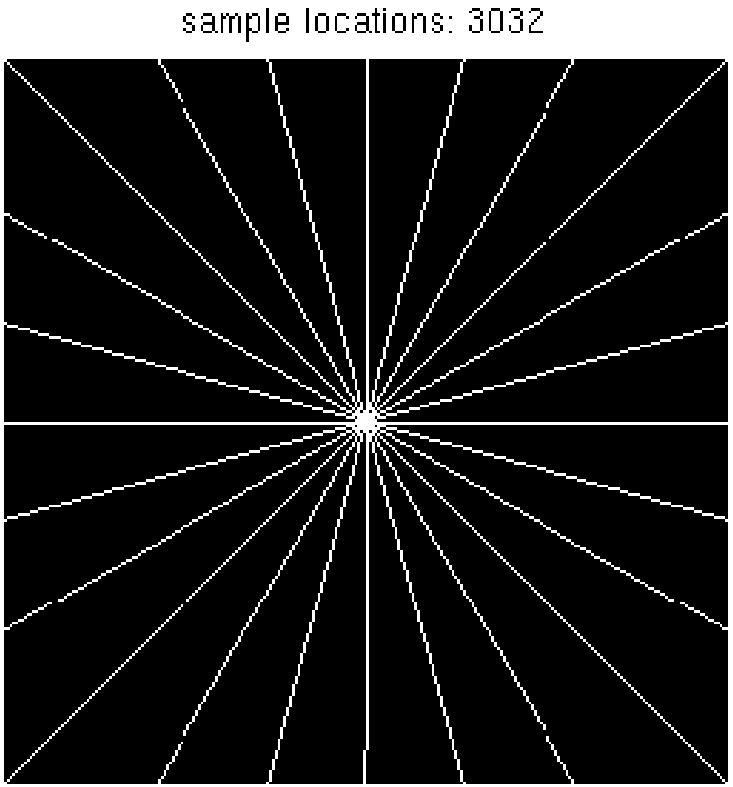
Inpainting: random

reconstructed image: snr=21.263

Compressed Sensing

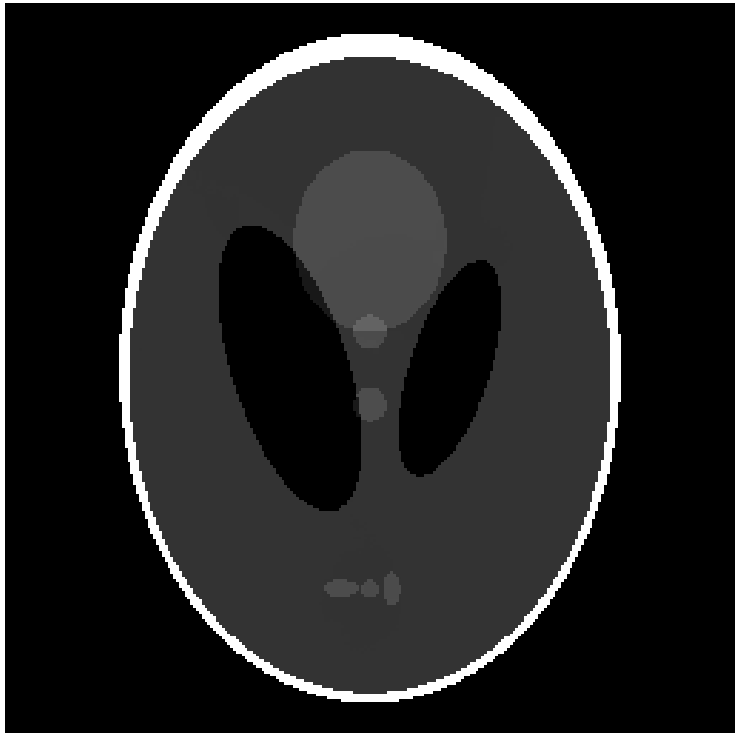
original image256x256=65536

Compressed Sensing



Compressed Sensing

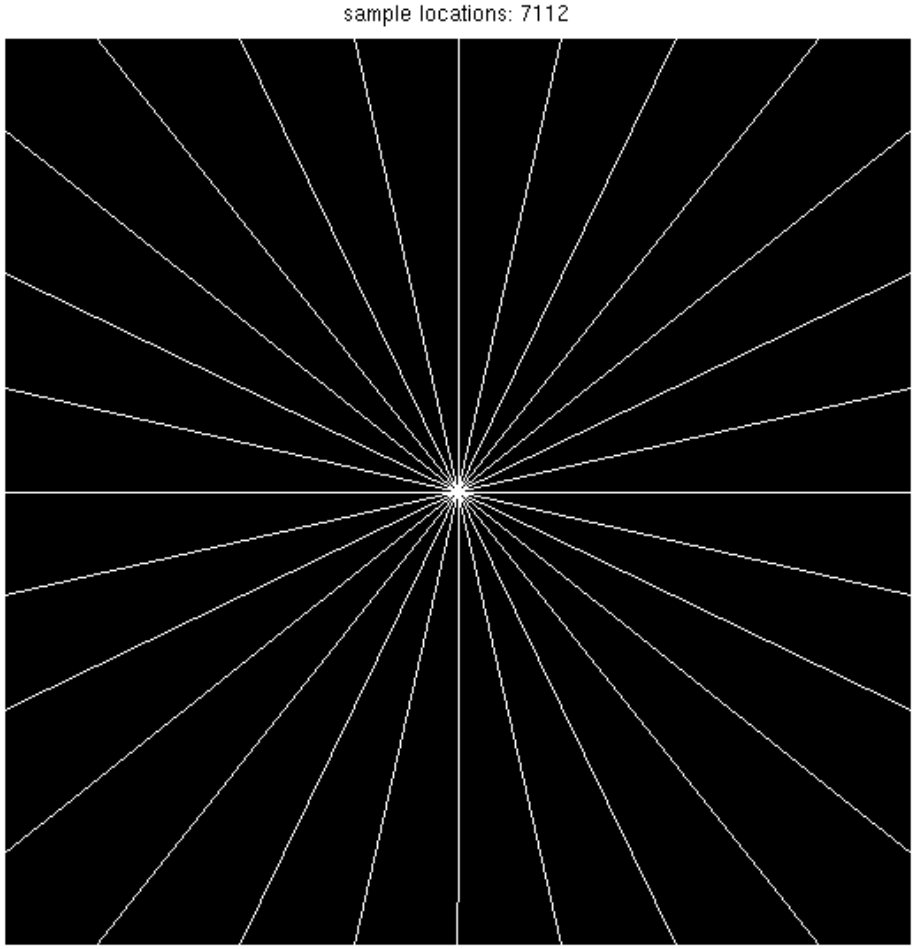
reconstructed image



Compressed Sensing

original image512x512=262144

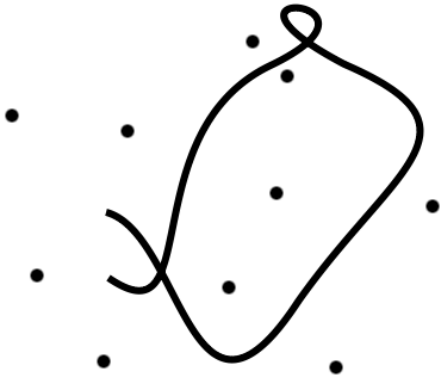
Compressed Sensing



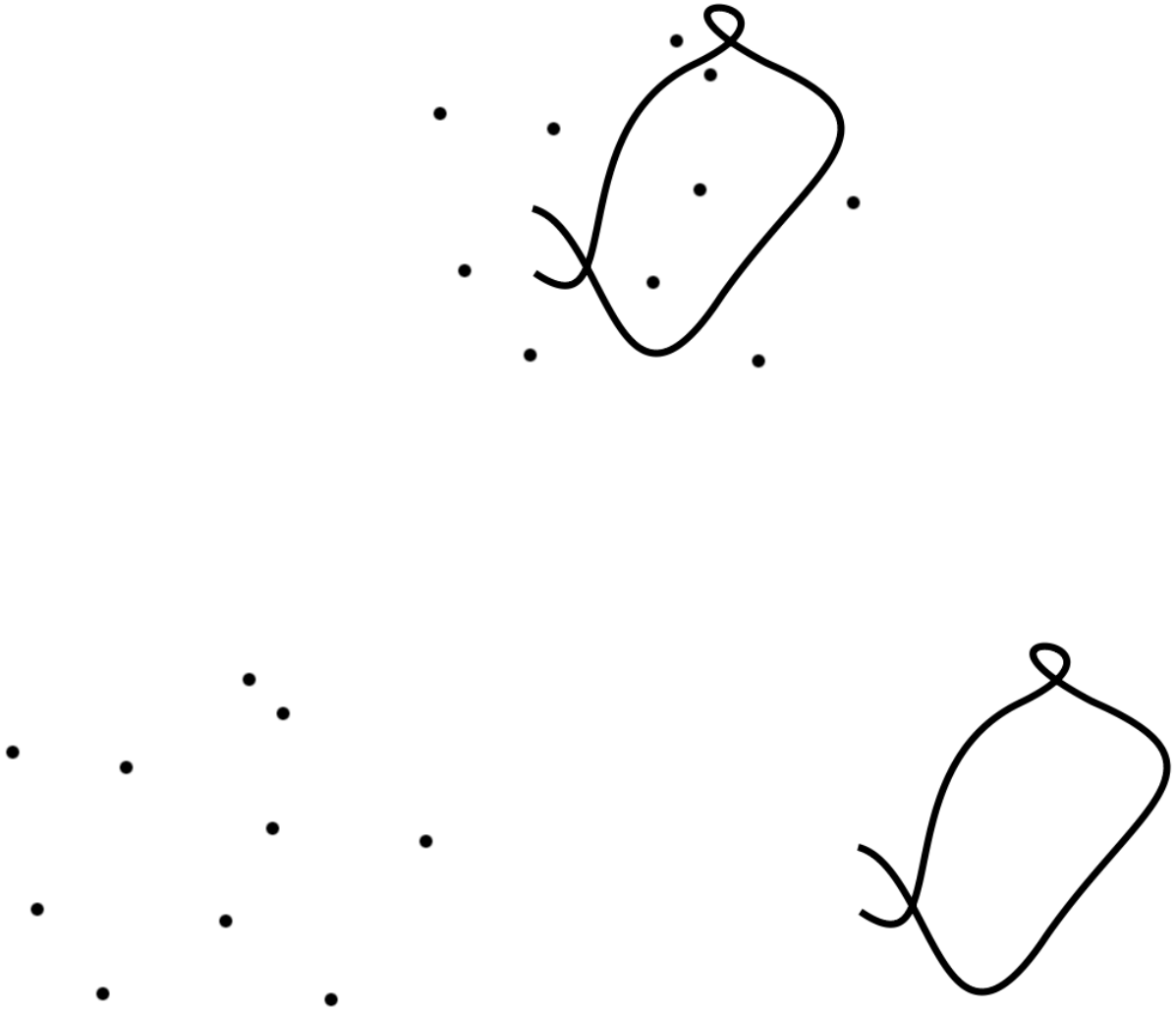
Compressed Sensing

reconstructed image

Separation



Separation



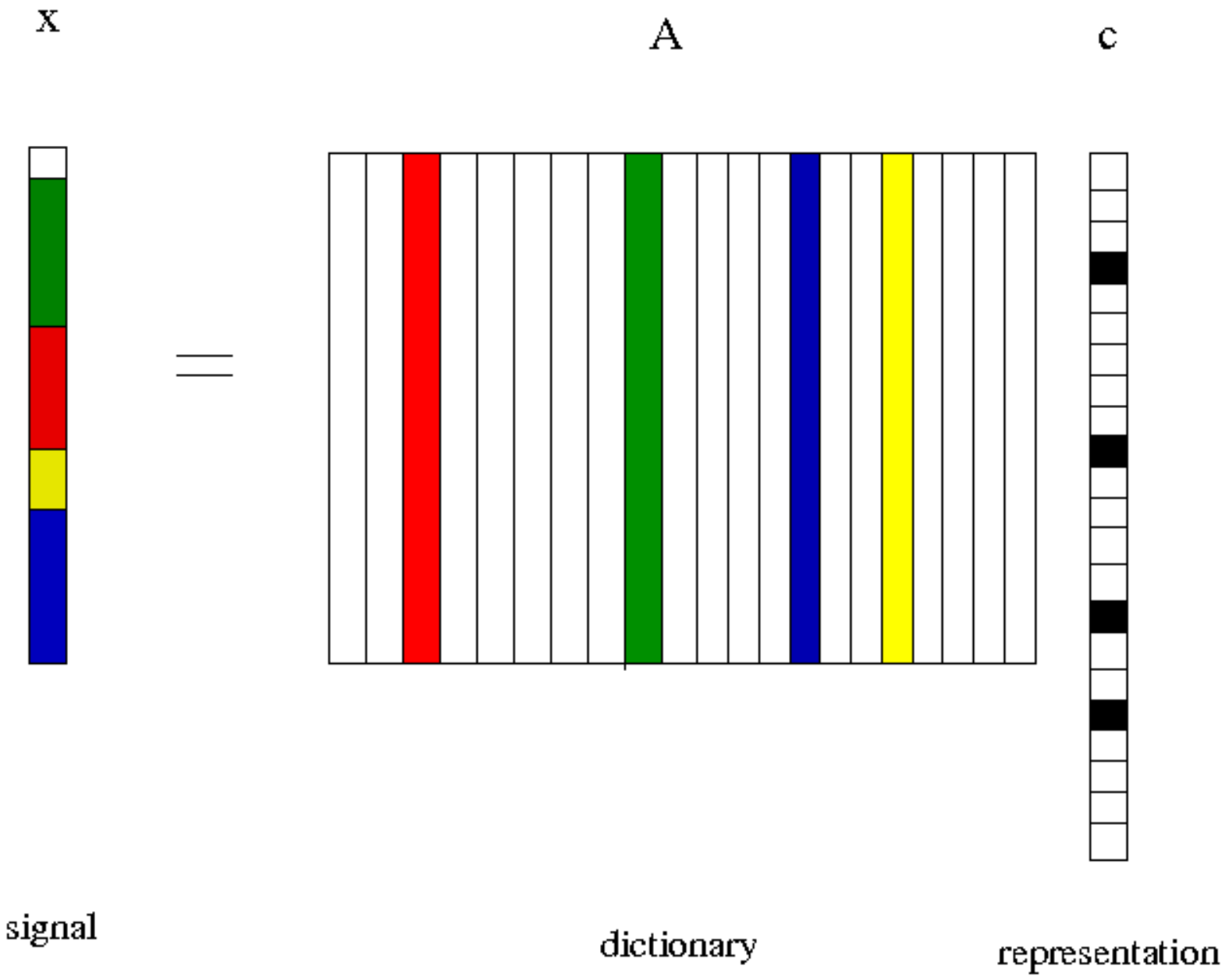
Super-resolution

Super-resolution

Notation/Terminology

- $y \in \mathbf{R}^d$: signal
- $\phi \in \Phi$: a column of $\Phi \in \mathbf{R}^{d \times K}$
- $\{\phi : \phi \in \Phi\}$ spans $\mathbf{R}^d \Rightarrow \Phi$: a dictionary for $\mathbf{R}^d$
- $\phi \in \Phi$: an atom
- $y = \Phi x \Rightarrow x$: a coefficient vector / representation

Sparse Model



ℓ_0-minimization

To find the sparsest representation of y:
we solve

$$\min_x \|x\|_0 \text{ subject to } y = Mx$$

ℓ_0-minimization

To find the sparsest representation of y:
we solve

$$\min_x \|x\|_0 \text{ subject to } y = Mx$$

How do we know we have *the* solution?

Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.

Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.
- Let *Spark* of $\Phi := \min_{x \in ker(\Phi), x \neq 0} \|x\|_0$.

Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.
- Let *Spark* of $\Phi := \min_{x \in ker(\Phi), x \neq 0} \|x\|_0$.
- Let $y = Mx$, $\|x\|_0 \leq (\text{Spark of } \Phi)/2$

Uniqueness of Sparse Solutions

- Let $\|x\|_0 := \#\{i : x_i \neq 0\}$.
- Let *Spark* of $\Phi := \min_{x \in ker(\Phi), x \neq 0} \|x\|_0$.
- Let $y = Mx$, $\|x\|_0 \leq (\text{Spark of } \Phi)/2$

$\Rightarrow x$ is unique solution of $\min_z \|z\|_0$ subject to $y = Mz$.

ℓ_0-minimization

Good, we (kind of) know when we have the solution.

ℓ_0-minimization

Good, we (kind of) know when we have the solution.
Unfortunately, solving ℓ_0-minimization is NP-hard.

ℓ_0-minimization

Good, we (kind of) know when we have the solution.
Unfortunately, solving ℓ_0-minimization is NP-hard.
What to do?

Alternatives

- Greedy Algorithm/Matching Pursuit
- Convex Optimization/Basis Pursuit

Matching Pursuit

Matching Pursuit [Mallat & Zhang]

1. Given signal y, dictionary Φ.
2. Set $k = 0$, $\hat{y}_0 = 0$, $r_0 = y$.
3. Find $\phi := \arg\min_{\phi \in \Phi} \langle \phi, r_k \rangle$.
4. Update $\hat{y}_{k+1} := \hat{y}_k + \langle r_k, \phi \rangle \phi$, $r_{k+1} := r_k - \langle r_k, \phi \rangle \phi$.
5. Increment k.
6. If $\hat{y}_k$ uses a specified number of atoms or r_k is smaller than a specified error, stop. Otherwise, go to step 3.

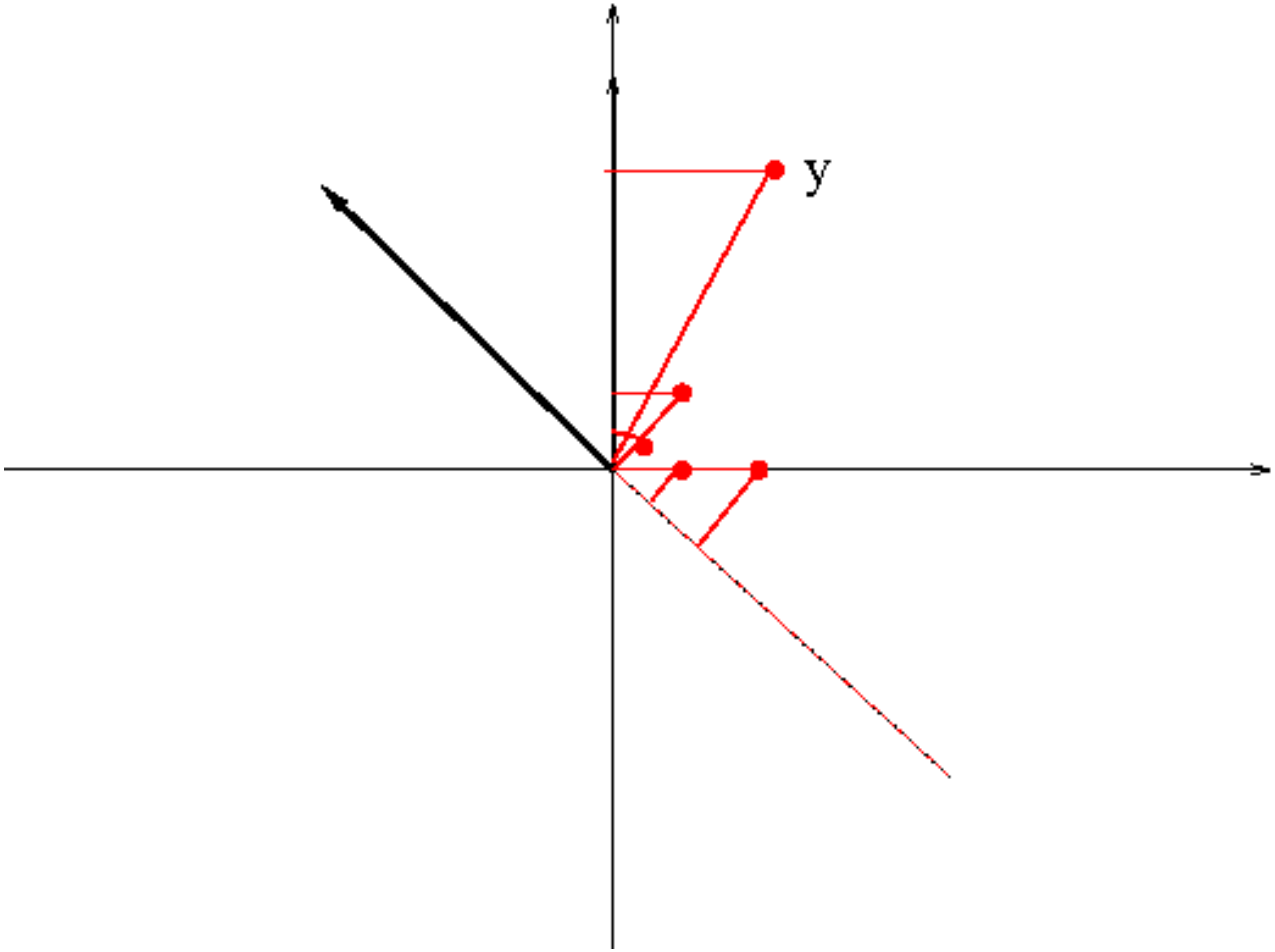
Energy Preservation

Energy Preservation (Pythagoras theorem)

$$\|r_k\|_2^2 = \|r_{k+1}\|_2^2 + |\langle r_k, \phi \rangle|^2$$

Drawback of Matching Pursuit?

MP can select an atom more than once.



Orthogonal Matching Pursuit

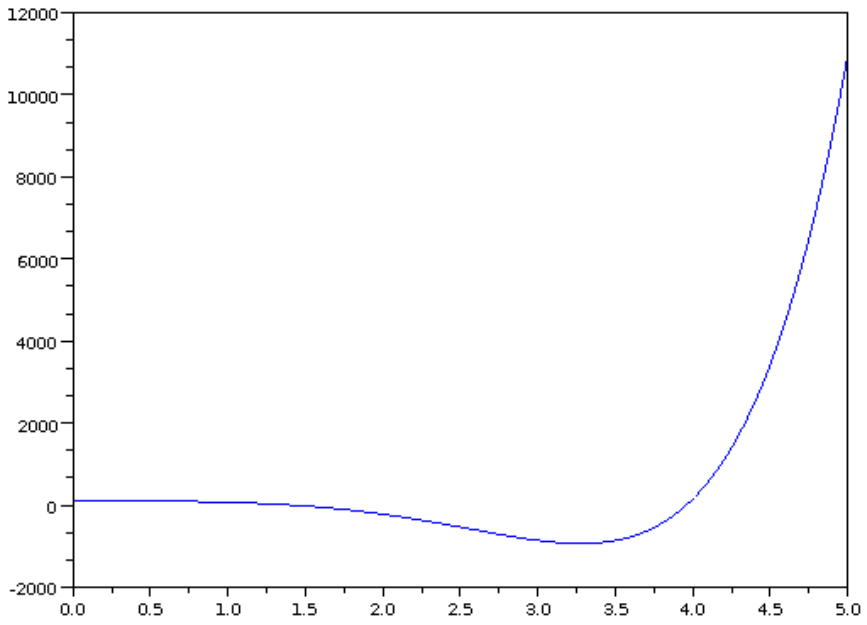
Orthogonal Matching Pursuit (OMP)

1. Given signal y, dictionary Φ.
2. Set $k = 0$, $\Lambda = \{\}$, $\hat{y}_0 = 0$, $r_0 = y$.
3. Find $i_k := \arg\min_i \langle \phi_i, r_k \rangle$ and set $\Lambda := \Lambda \cup \{i_k\}$.
4. Compute $\Delta\hat{y} :=$ (best ℓ_2-approximation from span of Φ_Λ to r_k), update $\hat{y}_{k+1} := \hat{y}_k + \Delta\hat{y}$, and $r_{k+1} := r_k - \Delta\hat{y}$.
5. Increment k.
6. If $\hat{y}_k$ uses a specified number of atoms or r_k is smaller than a specified error, stop. Otherwise, go to step 3.

Variety of Matching Pursuits

- Morphological Component Analysis [MCA, Bobin et al]
- Stagewise OMP [Donoho et al]
- CoSAMP [Needell & Tropp]
- ROMP [Needell & Vershynin]
- Iterative Hard Thresholding [Blumensath & Davies]

Sampling Issue



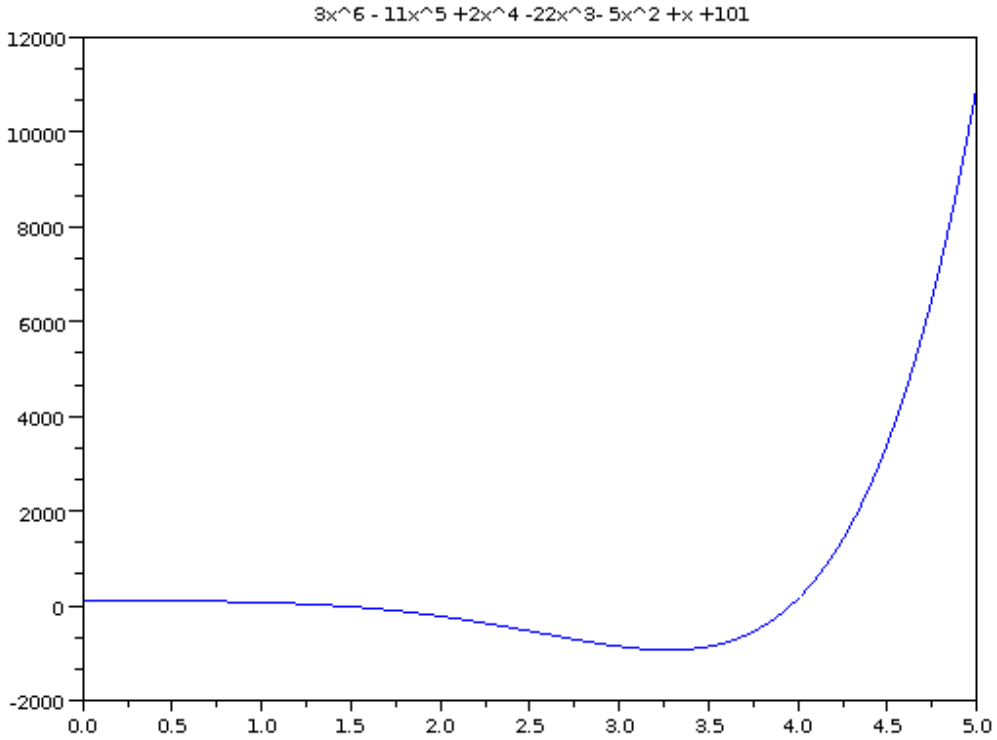
To find good approximation of $f : [0, 5] \to \mathbf{R}$, one may sample the function values

1. Uniformly at N locations for some large N, or
2. Adaptively sample–more locations near 5 in the figure– at some large number N locations,

then, use linear interpolation, cubic spline approximation, etc.

What if it is a polynomial of degree 6?

What we know it is a polynomial of degree 6? One needs only 7 different samples! Approximation is perfect!



What if it is a sum of 6 monomials?

What if it is a sum of 6 monomials of degree less than 200? How many samples do we need?

Think about Economy!

Digital Camera

1. High resolution photo (e.g. size 1024x1024) requires large number of samples (e.g. 1024x1024 samples).
2. Image is (immediately) compressed to smaller and *lossy* JPEG file.
3. Not a big deal for digital cameras?

Shannon Sampling / Nyquist Rate

Nyquist rate $\Rightarrow$ 44.1 KHz sampling rate

Shannon Sampling / Nyquist Rate

$$\begin{aligned} \text{Nyquist rate} &\Rightarrow 44.1 \text{ KHz sampling rate} \\ &\Rightarrow \text{large audio file} \end{aligned}$$

Shannon Sampling / Nyquist Rate

$$\begin{aligned}\text{Nyquist rate} &\Rightarrow \text{44.1 KHz sampling rate}\\ &\Rightarrow \text{large audio file}\\ &\Rightarrow \text{Compress to } \textit{lossy} \text{ MP3 file.}\end{aligned}$$

Fewer Samples

- Can we sample less to begin with?

Fewer Samples

- Can we sample less to begin with?
- How do we recover/approximate the original data?

Fewer Samples

- Can we sample less to begin with?
- How do we recover/approximate the original data?
- Key idea to exploit: Sparsity

ℓ_1-minimization

Relax

$$\min_x \|x\|_0 \text{ subject to } y = \Phi x$$

Solve

$$\min_x \|x\|_1 \text{ subject to } y = \Phi x$$

ℓ_1-minimization

Relax

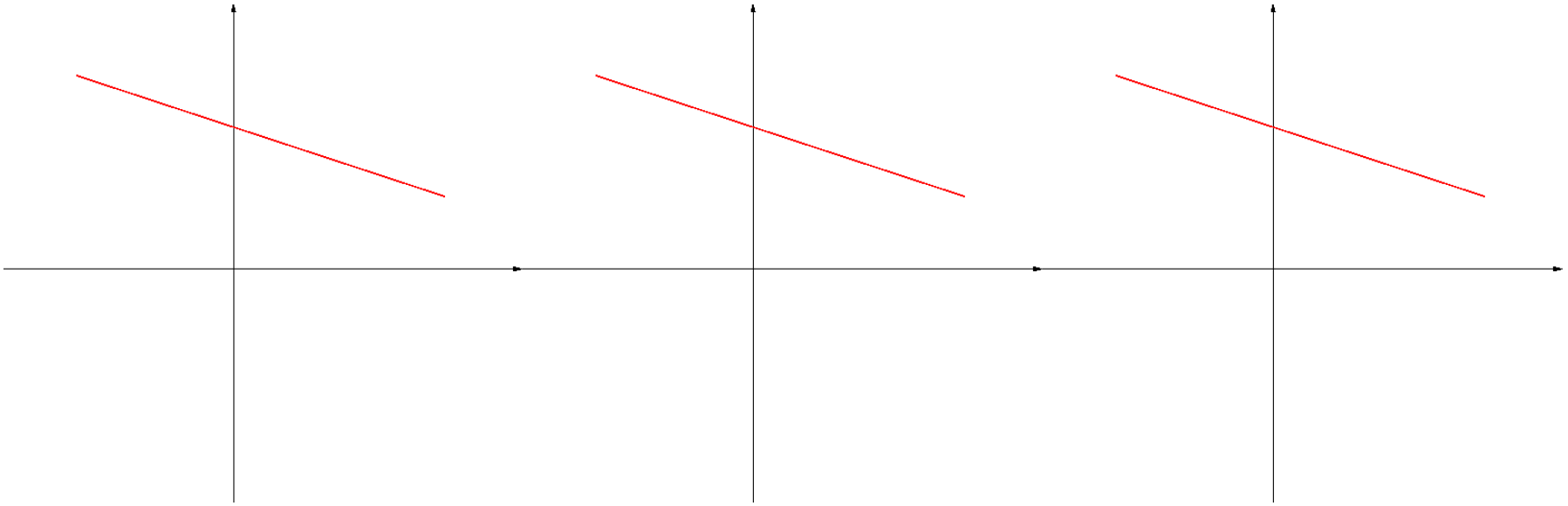
$$\min_x \|x\|_0 \text{ subject to } y = \Phi x$$

Solve

$$\min_x \|x\|_1 \text{ subject to } y = \Phi x$$

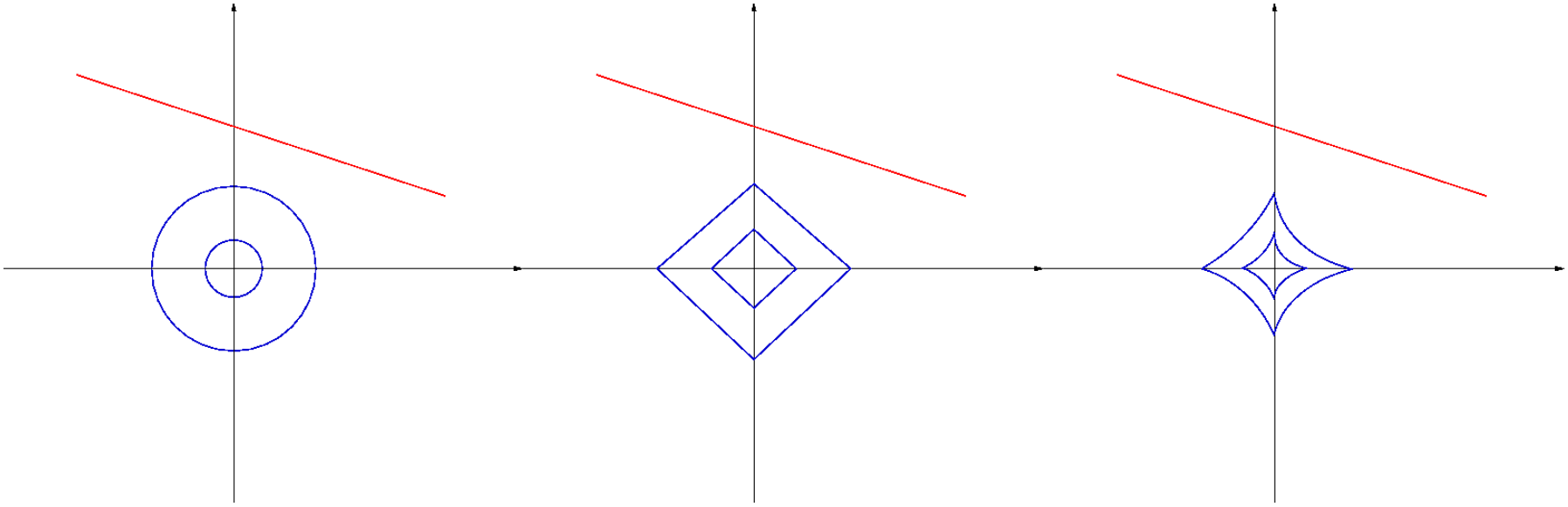
Convex problem. Can be recast into Linear Programming.

Why does ℓ_1-minimization work?



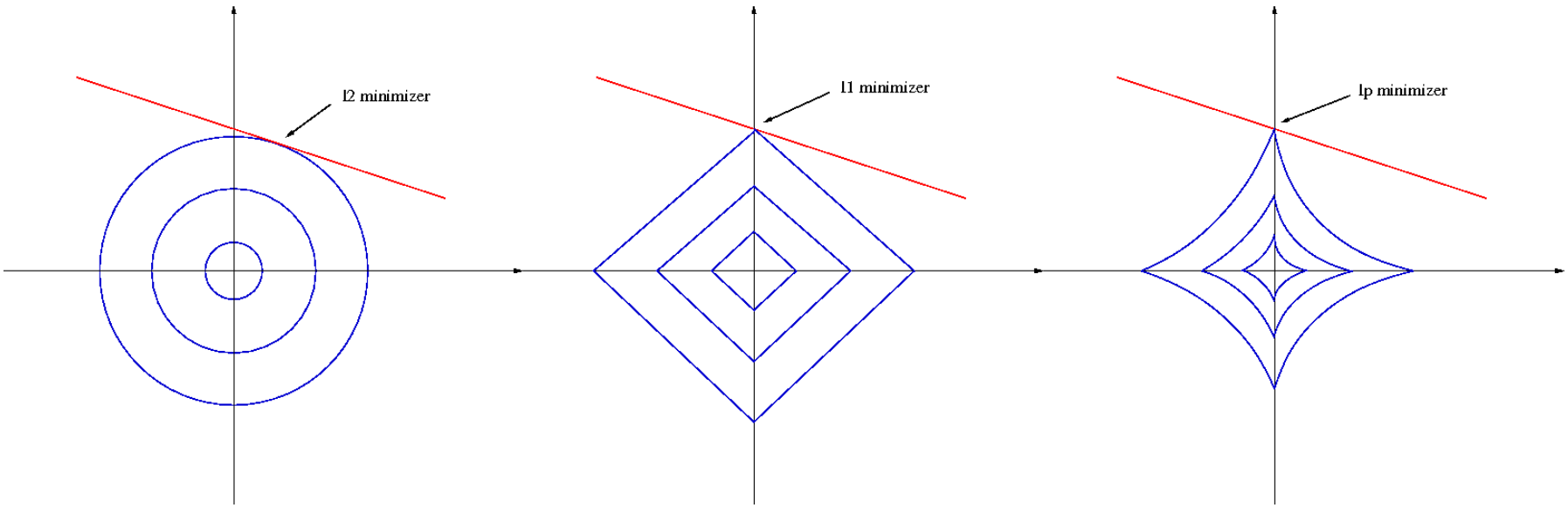
Feasible solutions

Why does ℓ_1-minimization work?



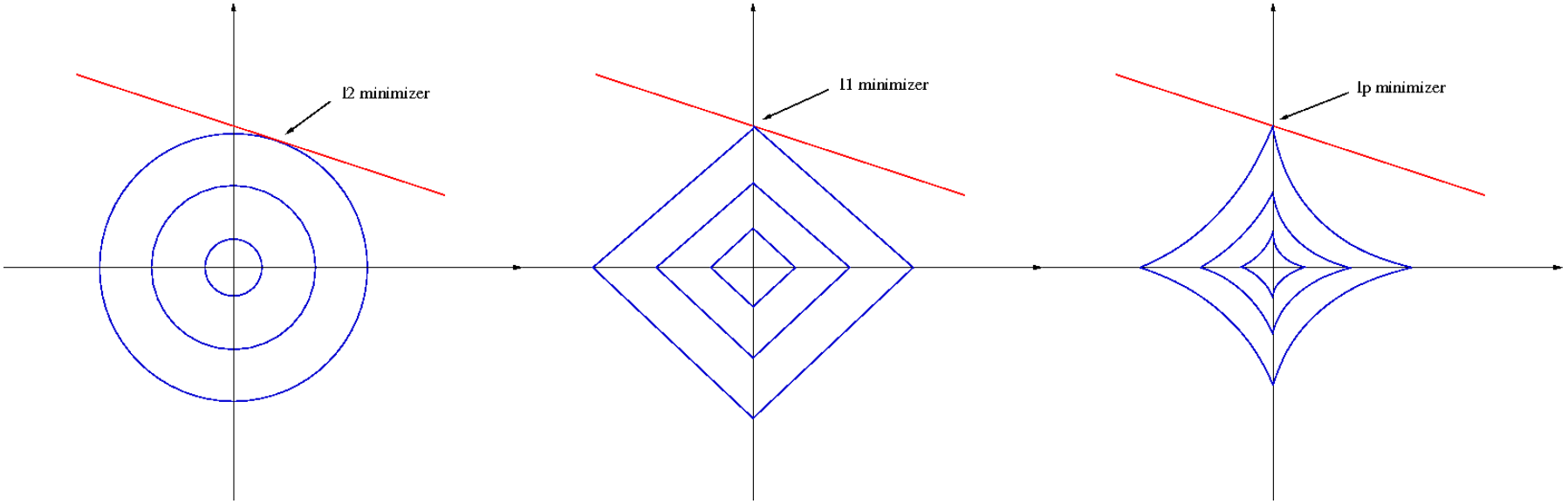
ℓ_2-balls $\qquad$ ℓ_1-balls $\qquad$ ℓ_p-balls, $0 < p < 1$

Why does ℓ_1-minimization work?



ℓ_2-minimizer ℓ_1-minimizer ℓ_p-minimizer, $0 < p < 1$

Why does ℓ_1-minimization work?



ℓ_2-minimizer ℓ_1-minimizer ℓ_p-minimizer, $0 < p < 1$

When can we guarantee successful recovery?

Null Space Property

Notation:
$N(\Phi) :=$ (Kernel / Null Space of Φ)
$\Lambda :=$ (Support of x^*)
$\bar{\Lambda} := \Lambda^c$

Null Space Property

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^* \tag{A}$$

ℓ_1-minimization (A) recovers x^* if and only if

$$|\langle z, sign(x^*)\rangle| < \|z_{\bar{\Lambda}}\|_1$$

for all $z \in N(\Phi)$.

Null Space Property

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^* \tag{A}$$

ℓ_1-minimization (A) recovers x^* if and only if

$$|\langle z, sign(x^*)\rangle| < \|z_{\bar{\Lambda}}\|_1$$

ℓ_1-minimization (A) recovers every x^* with support Λ if

$$\|z_\Lambda\|_1 < \|z_{\bar{\Lambda}}\|_1$$

for all $z \in N(\Phi)$.

Null Space Property

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = \Phi x^* \tag{A}$$

ℓ_1-minimization (A) recovers x^* if and only if

$$|\langle z, sign(x^*)\rangle| < \|z_{\bar{\Lambda}}\|_1$$

ℓ_1-minimization (A) recovers every x^* with support Λ if

$$\|z_\Lambda\|_1 < \|z_{\bar{\Lambda}}\|_1$$

ℓ_1-minimization (A) recovers every k-sparse x^* if

$$\|z_\Lambda\|_1 < \|z_{\bar{\Lambda}}\|_1$$

for all $z \in N(\Phi)$ and Λ with $\#\Lambda \leq k$.

Uniqueness for ℓ_1 using NSP

If x^* is k-sparse, then

$$\begin{aligned} \|x^* + z\|_1 &\geq \|x^*\|_1 - \|z_\Lambda\|_1 + \|z_{\bar{\Lambda}}\|_1 \\ &\geq \|x^*\|_1 . \end{aligned}$$

I.e., x^* is the unique ℓ_1-minimizer.

Coherence

Coherence $M(\Phi) := \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$.

Coherence

Coherence $M(\Phi) := \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$.

[Gribonval,Nielsen]

Theorem

If $\|x^*\|_0 \leq \frac{1}{2}\left(1 + \frac{1}{M(\Phi)}\right)$ *is a solution of* $y = Mx$*, then* x^* *is the unique minimizer of the* ℓ_1*-problem.* x^* *is also the unique minimizer of the* ℓ_0*-problem!*

Derivation of ℓ_1-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$

Derivation of ℓ_1-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

Derivation of ℓ_1-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

$$(1 + M(\Phi))|z_j| \leq M(\Phi) \|z\|_1$$

Derivation of ℓ_1-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

$$(1 + M(\Phi))|z_j| \leq M(\Phi) \|z\|_1$$

$$(1 + M(\Phi)) \|z_\Lambda\|_1 \leq (\#\Lambda) M(\Phi) \|z\|_1$$

Derivation of ℓ_1-guaranteeing sparsity level

For $z \in N(\Phi)$,

$$\phi_j z_j = -\sum_{i \neq j} \phi_i z_i.$$

$$|z_j| \leq M(\Phi) \sum_{i \neq j} |z_i|$$

$$(1 + M(\Phi))|z_j| \leq M(\Phi) \|z\|_1$$

$$(1 + M(\Phi)) \|z_\Lambda\|_1 \leq (\#\Lambda) M(\Phi) \|z\|_1$$

$\Rightarrow \|z_\Lambda\|_1 \leq \frac{\|z\|_1}{2}$ if

$$\#\Lambda \leq \frac{1}{2}(1 + \frac{1}{M(\Phi)}).$$

Coherence gives suboptimal estimates

For $\Phi \in \mathbf{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.

Coherence gives suboptimal estimates

For $\Phi \in \mathbf{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.

We can recover x^* if x^* is $\frac{1}{2}\left(1 + \sqrt{O(d)}\right) = O(\sqrt{d})$ sparse.

Coherence gives suboptimal estimates

For $\Phi \in \mathbf{R}^{d \times K}$, we may suppose $M(\Phi) = O\left(\frac{1}{\sqrt{d}}\right)$.

We can recover x^* if x^* is $\frac{1}{2}\left(1 + \sqrt{O(d)}\right) = O(\sqrt{d})$ sparse.

To put it differently, in order to recover k-sparse x^*, $d = O(k^2)$ measurements are desired.

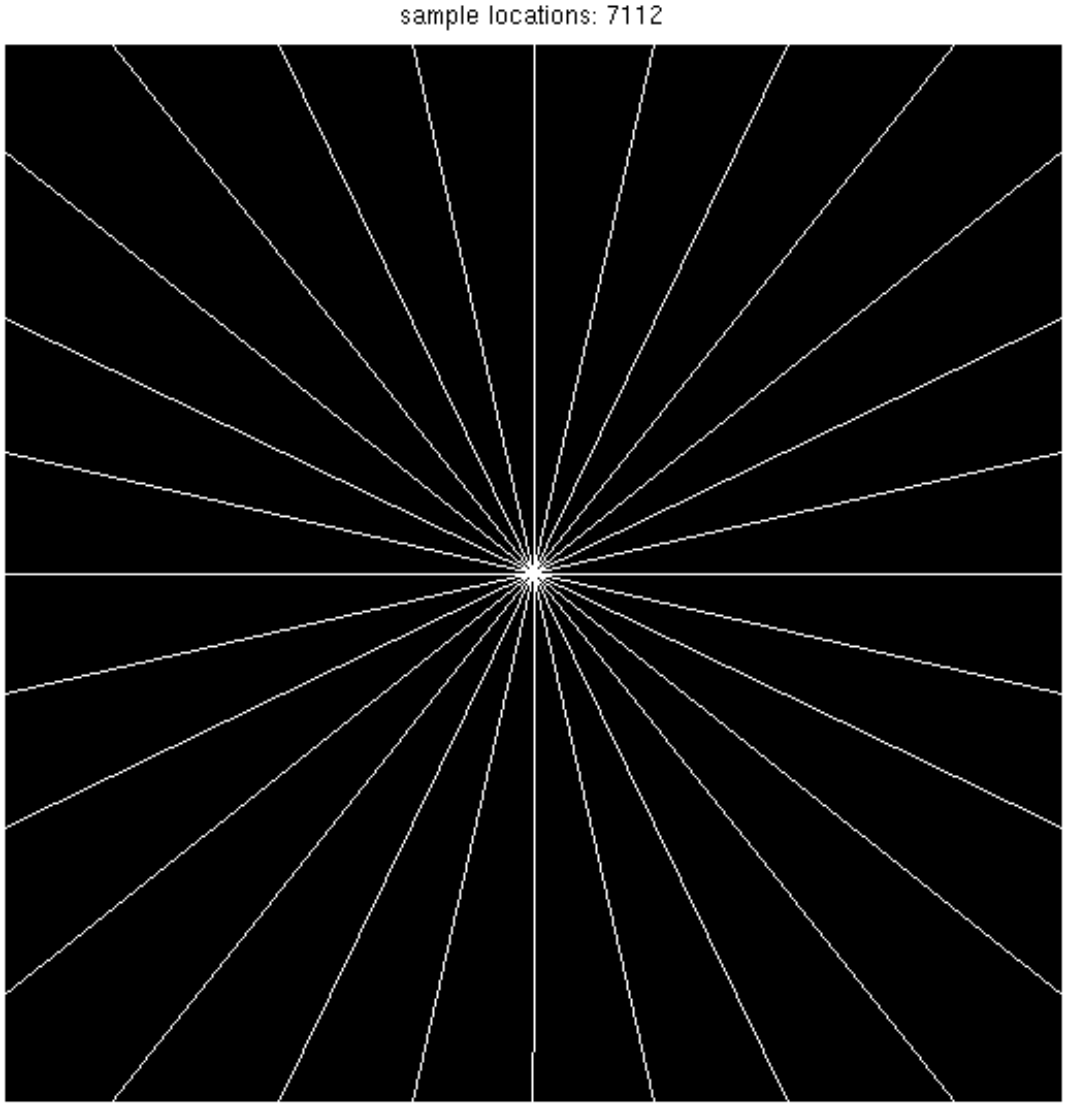
Recovery of Logan-Shepp phantom using Fourier Transform Samples

An Image to recover:

original image512x512=262144

Recovery of Logan-Shepp phantom using Fourier Transform Samples

Available samples in Fourier Domain (2.7% (4.3%)):



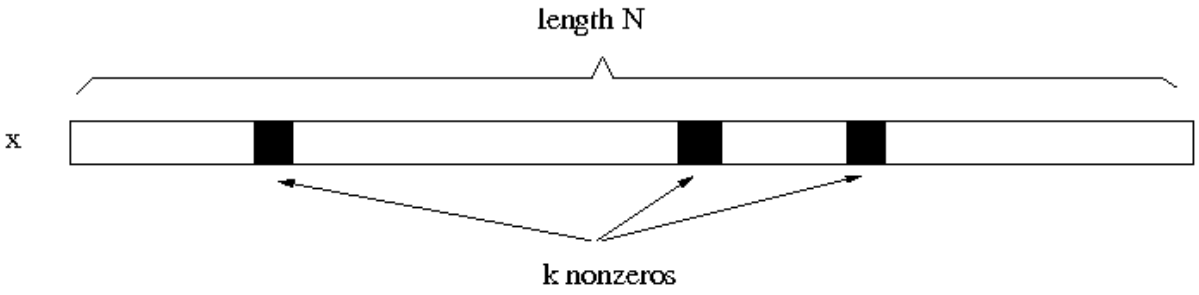
Recovery of Logan-Shepp phantom using Fourier Transform Samples

Reconstructed image—perfect!:

reconstructed image

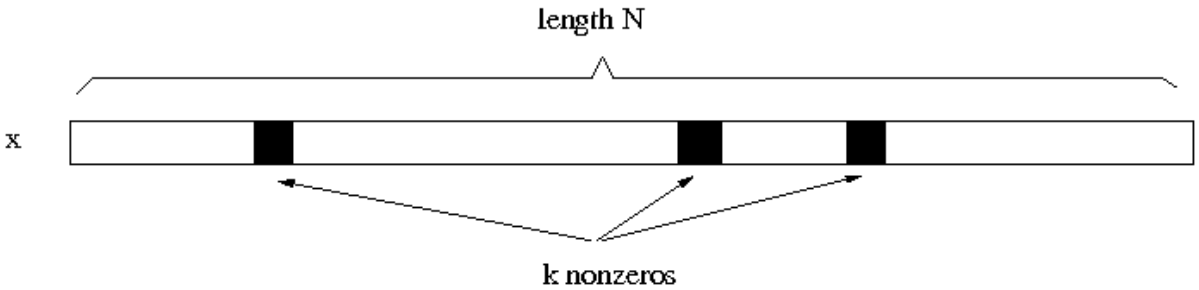
How many samples do we need?

To recover



How many samples do we need?

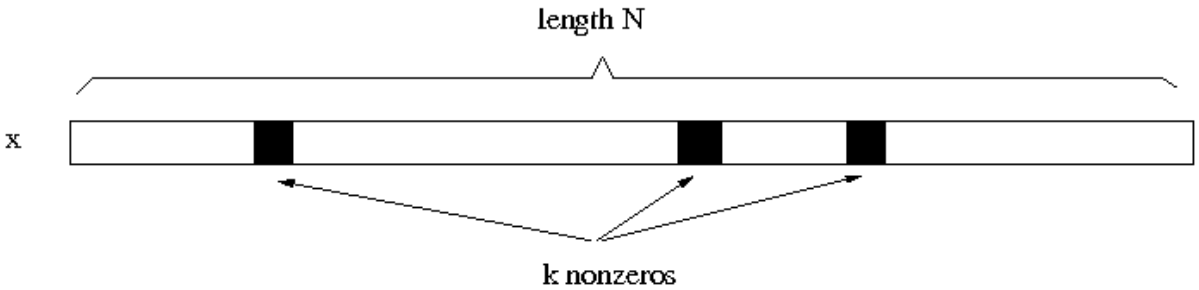
To recover



we (obviously) need $k \leq m \leq N$ samples.

How many samples do we need?

To recover



we (obviously) need $k \leq m \leq N$ samples.
Need to identify k nonzero positions out of N locations $\Rightarrow$
$\log_2 \binom{N}{k} \approx k \log \frac{N}{k}$ measurements

Random Sampling of Fourier Transform

[Candès, Romberg, Tao]

Theorem

Let $N = n^2$, f be an $n \times n$ real-valued image supported on T. Let Ω be a subset of $\{1, \ldots, n\}^2$ chosen unformly at random. If

$$\#\Omega \geq C(\#T) \log N,$$

then with high probability $(1 - O(N^{-M}))$, the minimizer to

$$\min_g \|g\|_1 \quad s.t. \quad \hat{g}|_\Omega = \hat{f}|_\Omega$$

is unique and is equal to f.

Restricted Isometry Property

For $s \in \mathbf{N}$, A matrix Φ satisfies the *Restricted Isometry Property* with the isometry constant δ_s if

$$(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2$$

for all s-sparse signal x.

Perfect Recovery condition in terms of RIP constant

[Candès]

Theorem

Assume that $\delta_{2s} < \sqrt{2} - 1$. *Then, the solution* x^* *to*

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \ \ s.t. \ \ M\tilde{x} = Mx$$

satisfies

$$\|x^* - x\|_2 \leq C_0 s^{-1/2} \|x - x_s\|_1$$

where C_0 *is a constant.*

Remark: The sensing matrix M is non-adaptive, but the ℓ_1-minimization with M recovers every s-sparse signal.

RIP and NSP

[Candès]

Theorem

If h is in the nullspace of Φ, then

$$\|h_T\|_1 \leq \rho \|h_{T^c}\|_1 , \quad \rho := \sqrt{2}\delta_{2s}(1-\delta_{2s})^{-1}$$

for every T with $\#T = s$.

Examples of Measurement Matrices with good RIP

- Random matrices with i.i.d. entries. If entries of $\Phi \in \mathbf{R}^{m \times d}$ are drawn from i.i.d. Gaussian with mean 0 and variance $1/m$, then with overwhelming probability the RIC δ_s is less than $\sqrt{2} - 1$ when $s \leq Cm/\log(d/m)$.
- Fourier ensemble. $\Phi \in \mathbf{R}^{m \times d}$ is constructed by sampling random m rows of the discrete Fourier transform.
- General orthogonal measurement ensembles.

Questions

1. 'Random dictionaries' are good for compressed sensing. Are there any deterministic dictionaries that are as good as random ones?
2. Any alternative to RIP?
3. What happens when the model is not exactly sparse?
4. What happens when there is noise?

Stable and Robust Recovery

We observe

$$y = \Phi x + \mathbf{n}.$$

We reconstruct x as the solution to

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \quad \text{s.t.} \|y - \Phi\tilde{x}\|_2 \leq \epsilon.$$

Stable and Robust Recovery

[Candès]

Theorem

Assume that $\delta_{2s} < \sqrt{2} - 1$ *and* $\|\mathbf{n}\|_2 \leq \epsilon$. *Then, the solution* x^* *satisfies*

$$\|x^* - x\|_2 \leq C_0 s^{-1/2} \|x - x_s\|_1 + C_1 \epsilon$$

where C_0 *and* C_1 *are some constants.*

Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \|Mx - y\|_2^2 + \lambda \|x\|_1$$

for some $\lambda > 0$.

Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \|Mx - y\|_2^2 + \lambda \|x\|_1$$

for some $\lambda > 0$.

$\lambda \to 0$?

Global Optimization Approach for Noisy Problem

We solve

$$\min_x \frac{1}{2} \|Mx - y\|_2^2 + \lambda \|x\|_1$$

for some $\lambda > 0$.

$\lambda \to 0$?

$\lambda \to \infty$?

Compressed Sensing

Paper: Compressed Sensing by David Donoho

$$\|\theta - \theta_N\|_2 \leq C(N+1)^{1/2-1/p}, \quad N = 0, 1, \ldots.$$

We Need Good Dictionaries

Techniques/Algorithms of sparse modeling begin with some dictionary Φ that provides sparse representations of the class of signals of interest. What are those Φ's?

Good Dictionaries

- Smooth functions $\Rightarrow$ Fourier transform
- Smooth functions with point singularities $\Rightarrow$ Wavelets
- Singularities along smooth curves $\Rightarrow$ Curvelets, Shearlets, etc.
- ...

PCA / SVD?

We want to design/learn dictionaries from data.

PCA / SVD?

We want to design/learn dictionaries from data.

Principal Component Analysis $\Rightarrow$ Not suitable for non-Gaussian mixtures

PCA / SVD?

We want to design/learn dictionaries from data.

Principal Component Analysis $\Rightarrow$ Not suitable for non-Gaussian mixtures

[Olshausen & Field]
Used Sparsity Promoting regularization to learn Dictionary
$\Rightarrow$ Localized, oriented, bandpass receptive fields emerge

Dictionary Learning via ℓ_1-criterion

If Φ were a good dictionary for a signal x, then we can recover x by solving

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \quad \text{s.t.} \quad y = \Phi\tilde{x}.$$

Dictionary Learning via ℓ_1-criterion

If Φ were a good dictionary for a signal x, then we can recover x by solving

$$\min_{\tilde{x}} \|\tilde{x}\|_1 \quad \text{s.t.} \quad y = \Phi\tilde{x}.$$

$\Rightarrow$ Given a data set $Y := [y_1, \ldots, y_N]$, we may find a good dictionary Φ by solving

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X.$$

Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because

$$\alpha\Phi\alpha^{-1}X = \Phi X$$

for any $\alpha \in \mathbf{R}$.

Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because

$$\alpha \Phi \alpha^{-1} X = \Phi X$$

  for any $\alpha \in \mathbf{R}$.
- The problem

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

  is not convex.

Some issues

- We can decrease $\|X\|_1$ as small as we wish by scaling because

$$\alpha\Phi\alpha^{-1}X = \Phi X$$

  for any $\alpha \in \mathbf{R}$.
- The problem

$$\min_{\Phi,X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

  is not convex.

The first issue can be dealt with by requiring that each column of Φ to be of unit length. We will write $\Phi \in \mathcal{U}(d, K)$ to mean that $\Phi \in \mathbf{R}^{d\times K}$ is such a dictionary.

Dictionary Identification

General Question: Given a data set Y, we learned a dictionary Φ via some learning method. Is Φ the 'right' one?

Dictionary Identification

Suppose that the training data Y is generated by

$$Y = \Phi_0 X_0.$$

Under what condition on X_0 (and Φ_0), can we be sure that the minimization problem

$$\min_{\Phi, X} \|X\|_1 \quad \text{s.t.} \quad Y = \Phi X$$

will recover the pair Φ_0 and X_0?

Necessary and Sufficient Condition

[Gribonval & Schnass]

Theorem

Suppose that $K \geq d$. (Φ_0, X_0) is a strict local minimum of the problem

$$\min_{\Phi, X} \|X\|_1 \quad s.t. \quad \Phi X = \Phi_0 X_0, \Phi \in \mathcal{U}(d, K)$$

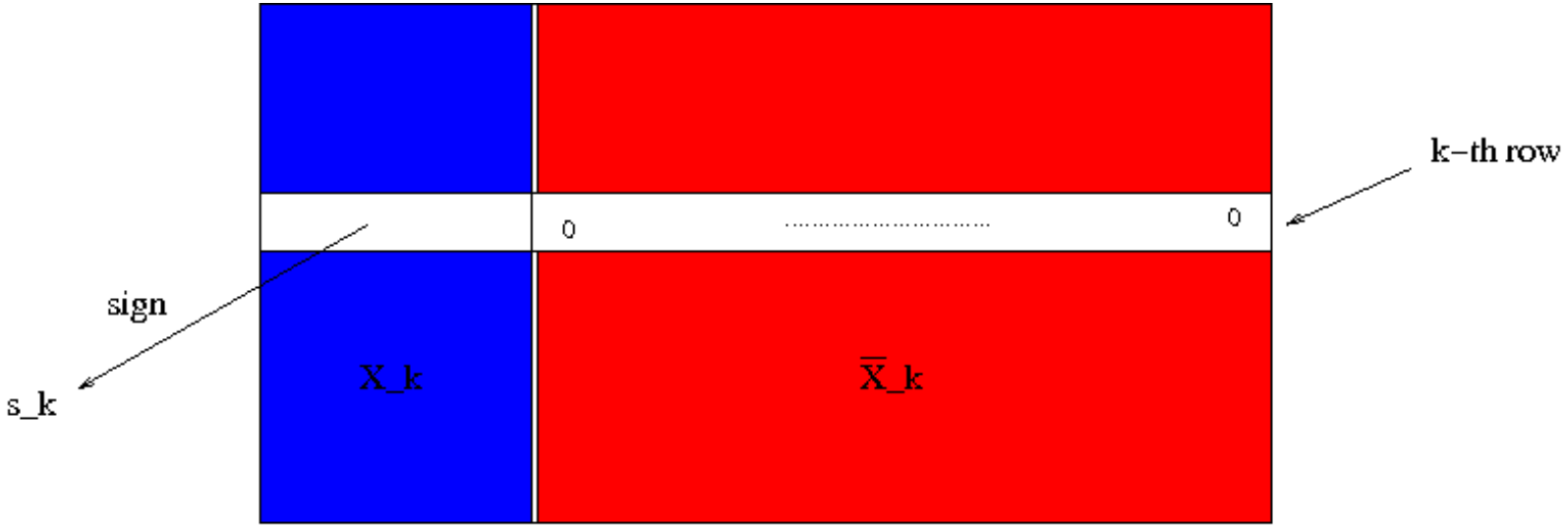
if and only if

$$|\langle CX_0 + V, sign(X_0)\rangle| < \|(CX_0 + V)_{\bar{\Lambda}}\|_1$$

for every $CX_0 + V \neq 0$ with $\mathrm{diag}(\Phi_0^* \Phi_0 C) = 0$ and $V \in N(\Phi_0)$.

Sufficient Condition for Basis

Notation:



Coefficient matrix X 0

Sufficient Condition for Basis

Theorem

(Φ_0, X_0) is a strict local minimum of the problem

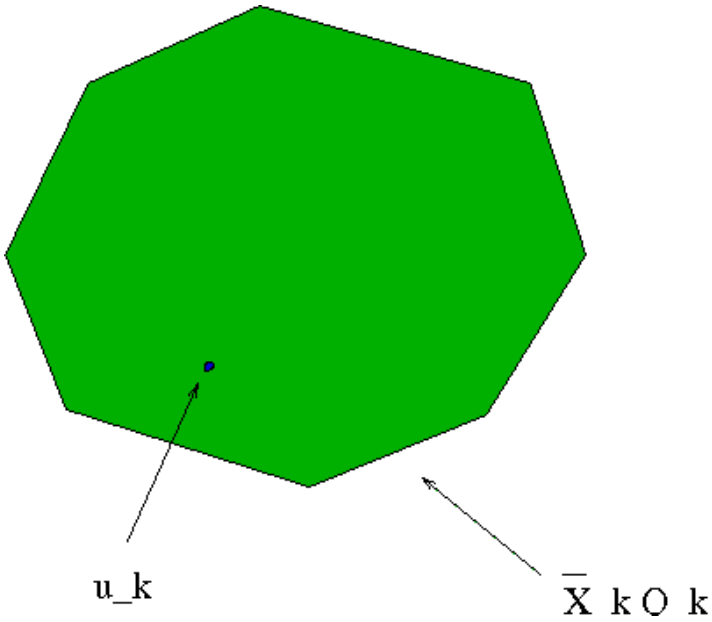
$$\min_{\Phi, X} \|X\|_1 \quad s.t. \quad \Phi X = \Phi_0 X_0, \Phi \in U(d, d)$$

if for every $k = 1, \ldots, d$, there exists $d_k \in \mathbf{R}^d$ with $\|d_k\|_\infty < 1$ such that

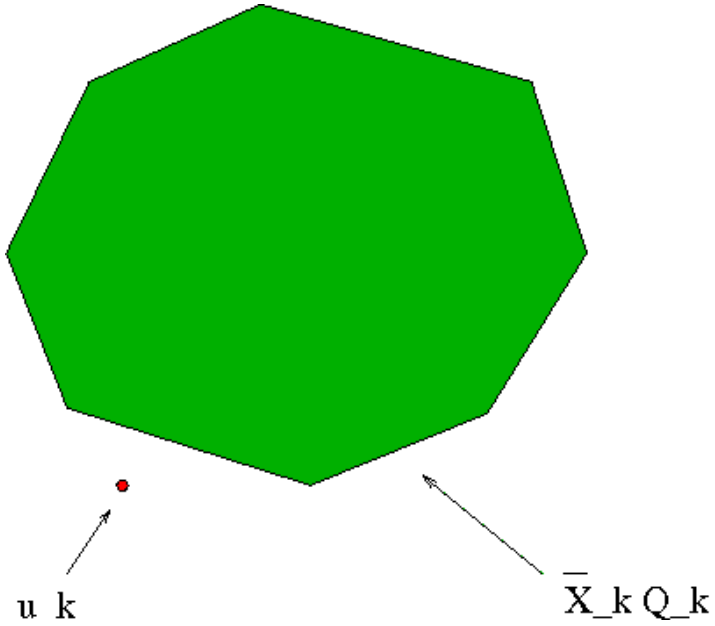
$$\bar{X}_k d_k = X_k s_k - \operatorname{diag}\left(\left(\left\|x^i\right\|\right)_i\right) m_k$$

where m_k is the k-th column of $\Phi_0^* \Phi_0$.

Geometric Interpretation



Local minimum



Not local minimum

$$u_k := X_k s_k - \operatorname{diag}\left(\left(\left\|x^i\right\|\right)_i\right) m_k, \quad Q_k := [-1, 1]^{r_k}.$$

Bernoulli-Gaussian Model

X_0 follows the Bernoulli-Gaussian Model with parameter $0 < p < 1$ if each entry of X_0 is the product of independent Bernoulli(p) random variable ξ_{ij} and Normal random variable g_{ij}, i.e.,

$$x_{ij} = \xi_{ij} g_{ij}.$$

Concentration of Measure Phenomena

If X_0 follows the Bernoulli-Gaussian distribution with parameter p, then with high probability $\bar{X}_k Q_k$ contains the ℓ_2-ball of radius

$$\alpha \approx Np(1-p)\sqrt{\frac{2}{\pi}},$$

Concentration of Measure Phenomena

If X_0 follows the Bernoulli-Gaussian distribution with parameter p, then with high probability $\bar{X}_k Q_k$ contains the ℓ_2-ball of radius

$$\alpha \approx Np(1-p)\sqrt{\frac{2}{\pi}},$$

$X_k s_k$ is contained in the ℓ_2-ball of radius

$$\beta \approx \sqrt{NK}p,$$

Concentration of Measure Phenomena

If X_0 follows the Bernoulli-Gaussian distribution with parameter p, then with high probability $\bar{X}_k Q_k$ contains the ℓ_2-ball of radius

$$\alpha \approx Np(1-p)\sqrt{\frac{2}{\pi}},$$

$X_k s_k$ is contained in the ℓ_2-ball of radius

$$\beta \approx \sqrt{NK}p,$$

and $\operatorname{diag}\left((\|x^i\|)_i\right) m_k$ is contained in the ℓ_2-ball of radius

$$\gamma \approx Np\sqrt{\frac{2}{\pi}}.$$

Sufficient Condition under Bernoulli-Gaussian Model

If the coherence $M(\Phi_0)$ of Φ_0 is less than $1 - p$ then for large enough N, Φ_0 is locally identifiable.

Sufficient Condition under Bernoulli-Gaussian Model

If the coherence $M(\Phi_0)$ of Φ_0 is less than $1 - p$ then for large enough N, Φ_0 is locally identifiable.

Surprise: One needs only

$$N \geq CK \log K.$$

Some dictionary learning methods

- K-SVD [Elad et al.]
- ISI [Gowreesunker]
- ...