ELEG 867 - Compressive Sensing and Sparse Signal Representations

Gonzalo R. Arce
Depart. of Electrical and Computer Engineering
University of Delaware

Fall 2011

Outline

- Introduction and Motivation
- Vector Spaces and the Nyquist-Shannon Sampling Theorem
- Sparsity and the ℓ_1 Norm
- Sparse Signal Representation

Compressed Sensing encompasses exciting and surprising developments in signal processing resulting from sparse representations.

It is about the interplay between sparsity and signal recovery. Roots trace back to †

- Mathematics and harmonic analysis
- Physical sciences and geophysics
- Vision
- Optimization and computational tools

This course describes this fascinating topic and the tools needed in its applications.

†D. Donoho, "Scanning the Technology," Proceedings of the IEEE. Vol. 98, No. 6, June 2010

Shannon-Nyquist Sampling Theorem

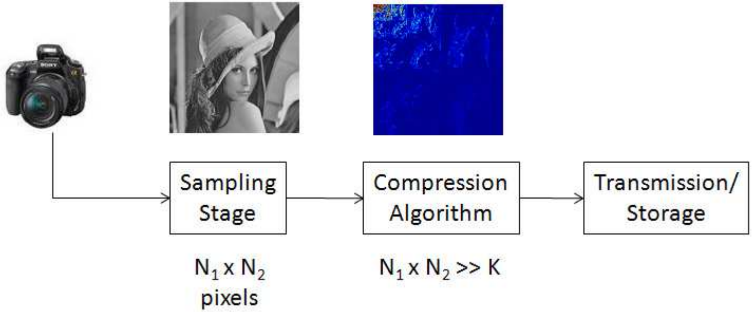
The Shannon-Nyquist Theorem: sampling frequency of an analog signal must be greater than twice the highest frequency of the signal in order to perfectly reconstruct the original signal from the sampled version.

Theorem

If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $\left(\frac{W}{2}\right)$ seconds apart.[†]

[†] C. E. Shannon. "Communication in the presence of noise." Proceedings of the IRE, Vol. 37, no.1, pp.10-21, Jan.1949.
H. Nyquist. "Certain topics in telegraph transmission theory." Trans. AIEE, vol.47, pp.617-644, Apr.1928.

- Traditional signal sampling and signal compression.



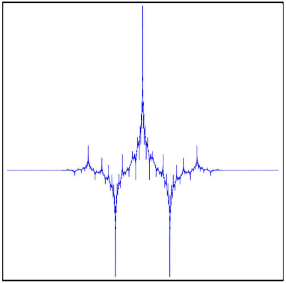
- Nyquist sampling rate gives exact reconstruction.

Pessimistic for some types of signals!

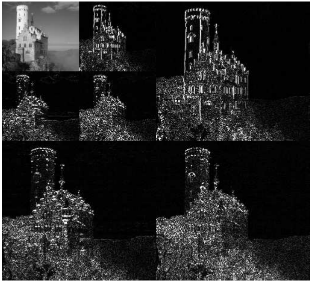
Sampling and Compression

- Transform data and keep important coefficients.

Original Image



Biorthogonal Spline Wavelet

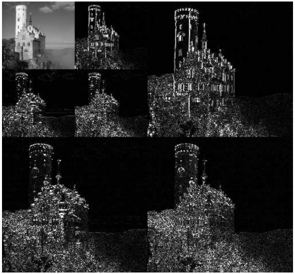


Wavelet Transform

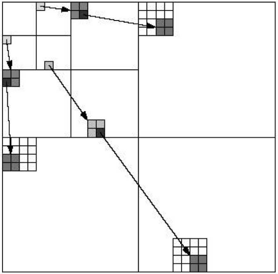
Sampling and Compression

- Lots of work to then throw away majority of data!.
  - e.g. JPEG 2000 Lossy Compression: A digital camera can take millions of pixels but the picture is encoded on a few hundred of kilobytes.

Original Image



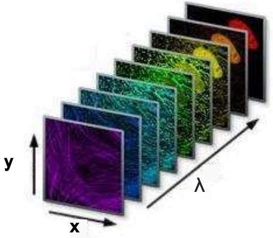
Wavelet Transform



Problem: Recent applications require a very large number of samples:

- Higher resolution in medical imaging devices, cameras, etc.
- Spectral imaging, confocal microscopy, radar arrays, etc.

Medical Imaging



Spectral Imaging

Sampling and Compressive Sensing

- Donoho †, Candès ‡, Romberg and Tao, discovered important results on the minimum number of data needed to reconstruct a signal
- Compressive Sensing (CS) unifies sensing and compression into a single task
- Minimum number of samples to reconstruct a signal depends on its *sparsity* rather than its *bandwidth*.

† D. Donoho. "Compressive Sensing". IEEE Trans. on Information Theory. Vol.52(2), pp.5406-5425, Dec.2006.
‡ E. Candès, J. Romberg and T. Tao. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information". IEEE Trans. on Information Theory. Vol.52(4), pp.1289-1306, Apr.2006.

Vector Spaces and the Nyquist-Shannon Sampling Theorem

Vector space: set of vectors H satisfying the following axioms:

- Associativity property: $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$.
- Commutativity property: $v_1 + v_2 = v_2 + v_1$.
- Identity element: $\exists 0 \in H$, such that $v + 0 = v, \forall v \in H$.
- Inverse element: $\forall v \in H$, then $\exists - v \in H$, such that $v + (-v) = 0$.
- Distribut. of scalar: s is a scalar, such that $s(v_1 + v_2) = sv_1 + sv_2$.
- Distribut. of scalar: s_1, s_2 are scalars, such that $(s_1 + s_2)v = s_1 v + s_2 v$.
- Associat. of scalars: s_1, s_2 are scalars, such that $s_1(s_2 v) = (s_1 s_2)v$.
- Identity element of product: $\exists$ a scalar 1, such that $1v = v$.

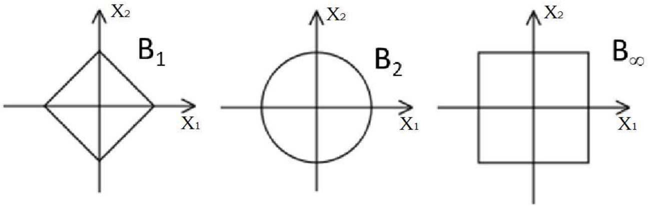
Norms: A norm $\|\cdot\|$ on the vector space H satisfies:

- $\forall x \in H$, $\|x\| \geq 0$, and $\|x\| = 0 \Leftrightarrow x = 0$.
- $\forall \alpha \in \mathbb{C}$, $\|\alpha x\| = |\alpha|\|x\|$. (Homogeneity).
- $\forall x, y \in H$, $\|x + y\| \leq \|x\| + \|y\|$. (Triangle inequality).

Examples of norms:

- H is the space $\mathbb{R}^n$, with norm $\|x\|_{\ell_p} = (\sum_{k=1}^{n} |x_k|^p)^{1/p}$, for $p \geq 1$.

In $\mathbb{R}^2$, set the unit ball $B_p = \{x : \|x\|_{\ell_p} = 1; p \geq 1\}$:



The unit ball is the set of all points (x_1, x_2) which satisfy the equations:

- $|x_1| + |x_2| = 1$, for B_1.
- $x_1^2 + x_2^2 = 1$, for B_2.
- $\max\{x_1, x_2\} = 1$, for B_∞.

In $\mathbb{R}^n$, $\|x\|_{\ell_1} = \sum_{k=1}^{n} |x_k|$ is a norm since it satisfies:

- $\forall x \in \mathbb{R}^n$, then $\|x\|_{\ell_1} = \sum_{k=1}^{n} |x_k| \geq 0$. Also, $\sum_{k=1}^{n} |x_k| = 0$, if and only if $x_k = 0, \forall k$.
- $\forall \alpha \in \mathbb{C}$, then $\|\alpha x\|_{\ell_1} = \sum_{k=1}^{n} |\alpha x_k| = |\alpha| \sum_{k=1}^{n} |x_k| = |\alpha| \|x\|_{\ell_1}$.
- $\forall x, y \in \mathbb{R}^n$, then

$$\begin{aligned} \|x + y\|_{\ell_1} &= \sum_{k=1}^{n} |x_k + y_k| \\ &\leq \sum_{k=1}^{n} (|x_k| + |y_k|); \text{ Convex Function} \\ &= \sum_{k=1}^{n} |x_k| + \sum_{k=1}^{n} |y_k| \\ &= \|x\|_{\ell_1} + \|y\|_{\ell_1}. \end{aligned}$$

In $\mathbb{R}^n$, $\|x\|_{\ell_p} = (\sum_{k=1}^{n} |x_k|^p)^{1/p}$, with $p = 0.5$, is *not* a norm:

- $\forall x \in \mathbb{R}^n$, then $\|x\|_{\ell_{0.5}} = (\sum_{k=1}^{n} |x_k|^{1/2})^2 \geq 0$. Also, $(\sum_{k=1}^{n} |x_k|^{0.5})^2 = 0$, if and only if $x_k = 0$, $\forall k$.
- $\forall \alpha \in \mathbb{C}$, then $\|\alpha x\|_{\ell_{0.5}} = (\sum_{k=1}^{n} |\alpha x_k|^{1/2})^2 = (\sum_{k=1}^{n} |\alpha|^{1/2}|x_k|^{1/2})^2 = (|\alpha|^{1/2} \sum_{k=1}^{n} |x_k|^{1/2})^2 = |\alpha|\|x\|_{\ell_{0.5}}$.
- $\forall x, y \in \mathbb{R}^n$, then

$$\begin{aligned}
\|x+y\|_{\ell_{0.5}} &= (\sum_{k=1}^{n} |x_k + y_k|^{1/2})^2 \\
&\geq (\sum_{k=1}^{n} |x_k|^{1/2} + \sum_{k=1}^{n} |y_k|^{1/2})^2 - 2\sum_{k=1}^{n} |x_k|^{1/2} \sum_{k=1}^{n} |y_k|^{1/2}; \\
&= (\sum_{k=1}^{n} |x_k|^{1/2})^2 + (\sum_{k=1}^{n} |y_k|^{1/2})^2 = \|x\|_{\ell_{0.5}} + \|y\|_{\ell_{0.5}}
\end{aligned}$$

(Triangle inequality is not satisfied)

Other Examples of Norms:

- Operator norm: H is the space of $m \times n$ matrices A
  $\|A\| = \sigma_{\max}(A)$ = maximum singular value of A.
- Frobenius norm: H is the space of $m \times n$ matrices A
  $\|A\|_F = (\sum_{i,j} A_{i,j}^2)^{1/2} = (\sum_k \sigma_k^2)^{1/2}$

Normed vector spaces: vector spaces H satisfying the norm properties.

Examples of normed vector spaces:

- $\ell_2(\mathbb{R})$ (also known as ℓ^2 or Euclidean space): the vector space $\mathbb{R}$ satisfying the properties of the ℓ_2-norm.
- $\ell_\infty(\mathbb{R})$: the vector space $\mathbb{R}$ satisfying the properties of the ℓ_∞-norm.

Inner Products

An inner product $< \cdot, \cdot >$ on H satisfies $\forall x, y, z \in H$ and $\alpha \in \mathbb{C}$:

- $< x, y > = < y, x >^*$
- $< \alpha x, y > = \alpha < x, y >$
- $< x + y, z > = < x, z > + < y, z >$
- $< x, x > \geq 0$, $< x, x > = 0 \Leftrightarrow x = 0$

A inner product operator induces a norm on H: $\sqrt{< x, x >} = \|x\|$.

In $\ell_2(\mathbb{R})$, for instance, the inner product is given by:

$$< x, y > = \int_{-\infty}^{\infty} x(t) y^*(t) dt. \tag{1}$$

$$< x, x > = \int_{-\infty}^{\infty} x(t) x^*(t) dt = \|x\|_{\ell_2}^2. \tag{2}$$

Hilbert Spaces

A vector space H that satisfies the inner product properties is known as Hilbert space.

Examples of Hilbert spaces:

- The Euclidean space $\mathbb{R}^n$ with the dot product as inner product: $< x,y >= \sum_{i=1}^{n} x_i y_i$.
- The space of real-valued, finite variance, zero-mean random variables: $< x,y >= E[xy]$.
- The space of $m \times n$ matrices with: $< A,B >_{tr} = \text{trace}(AB)$.

Definitions

- Orthogonality: two signals x, y are orthogonal if $< x, y >= 0$.
- Orthonormal basis: a basis of a vector space is orthonormal if their vectors are orthonormal.
- Orthonormal sequence: $\{\beta_n\}_{n\in\mathbb{Z}}$ is an orthonormal sequence if: $||\beta_n|| = 1, \forall n$, and $< \beta_n, \beta_m >= 0, \forall n \neq m$

Example:

- Fourier series: $\{\beta_n\}_{n\in\mathbb{Z}} = \{e^{j2\pi nt}\}_{n\in\mathbb{Z}}$ is an orthobasis for $\ell_2([0,1])$, since:
  - $||\beta_n||_{\ell_2} = 1$
  - $< \beta_n, \beta_m >= 0$

Definitions

- Cauchy-Schwarz Inequality: $| < x, y > | \leq \|x\| \|y\|$.
  - For the Euclidean space $H = \mathbb{R}^n$:
    $| < x, y > | = \sum_i x_i y_i \leq \sqrt{(\sum_i x_i^2)}\sqrt{(\sum_i y_i^2)} = \|x\|_{\ell_2} \|y\|_{\ell_2}$.
  - For the space of real-valued, finite variance, zero-mean random variables:$| < x, y > | = E[xy] \leq (E[x])(E[y]) = \|x\| \|y\|$.

Shannon-Nyquist Sampling Theorem

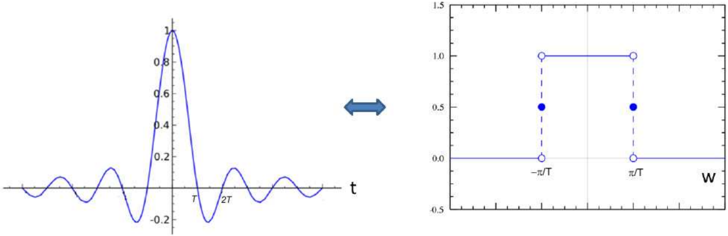
Sampling of a bandlimited signal.

Let $\hat{f}(w)$ be the Fourier transform of $f(t)$. Let the space of bandlimited signals be

$$B_{\pi/T} = \{f(t) \in \mathbb{R}^n \text{ s.t. } \hat{f}(w) = 0, \forall |w| > \pi/T\}.$$

Define

$$h_T(t) = \frac{\sqrt{T}\sin(\pi t/T)}{\pi t} \leftrightarrow \hat{h}(w) = \begin{cases} \sqrt{T} & ;\text{if } |w| \leq \pi/T \\ 0 & ;\text{if } |w| > \pi/T. \end{cases}$$



By the linear shift property of the Fourier series

$$h_T(t-nT) \leftrightarrow \sqrt{T}e^{jwnT}.$$

Using the Parseval theorem definition

- Parseval theorem: $\int_{-\infty}^{\infty} f(t)g^*(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(w)\hat{g}(w)dw$,

note that $h_T(t-nT)$ is an orthobasis for the bandlimited signals $f(t)$ in $B_{\pi/T}$:

$$\begin{aligned}\int_{-\infty}^{\infty} h_T(t)h(t-nT)dt &= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} Te^{jwnT}dw \\ &= \frac{1}{2j\pi n}e^{jwnT}\Big|_{-\pi/T}^{\pi/T} \\ &= \frac{1}{2j\pi n}(e^{j\pi n} - e^{-j\pi n}) \\ &= 0, \quad \forall n \in \mathbb{Z}.\end{aligned}$$

The signals $f(t)$ in $B_{\pi/T}$ can be expressed in terms of its orthobasis

$$f(t) = \sum_{n\in\mathbb{Z}} \langle f(t), h(t-nT)\rangle h(t-nT). \tag{3}$$

Using the inner product definition in (2) and the parseval theorem, the coefficients for the signal expansion in terms of its orthobasis are

$$\begin{aligned} \langle f(t), h(t-nT)\rangle &= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \hat{f}(w)\sqrt{T}e^{jwnT}dw \\ &= \sqrt{T}f(nT) \end{aligned} \tag{4}$$

Replacing (4) in (3), the signals $f(t)$ in $B_{\pi/T}$ can then be expressed in terms of a sequence

$$f(t) = \sqrt{T} \sum_{n \in \mathbb{Z}} f(nT) h(t - nT). \tag{5}$$

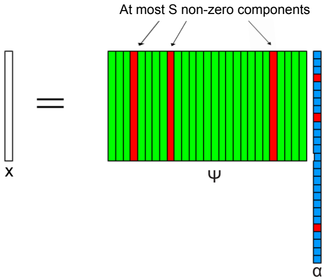
where, the coefficients $f(nT)$ of the sequence are samples of $f(t)$.

Nyquist-Shannon-Kotelnikov Theorem

If a signal $f(t)$ contains frequencies satisfying $|w| < \pi/T$, the signal is completely determined by series of points spaced T seconds apart.

Sparsity

- Signal sparsity critical to CS
- Plays roughly the same role in CS that bandwidth plays in Shannon-Nyquist theory
- A signal $x \in R^N$ is S-sparse on the basis Ψ if x can be represented by a linear combination of S vectors of Ψ as $x = \Psi\alpha$ with $S \ll N$



The ℓ_1 Norm and Sparsity

- The ℓ_0 norm is defined by: $\|x\|_0 = \#\{i : x(i) \neq 0\}$
  Sparsity of x is measured by its number of non-zero elements.

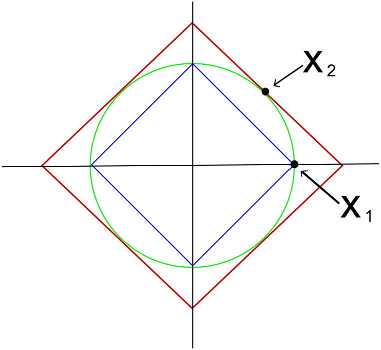
- The ℓ_1 norm is defined by: $\|x\|_1 = \sum_i |x(i)|$
  ℓ_1 norm has two key properties:
  - Robust data fitting
  - Sparsity inducing norm

- The ℓ_2 norm is defined by: $\|x\|_2 = (\sum_i |x(i)|^2)^{1/2}$
  ℓ_2 norm is not effective in measuring *sparsity* of x

Why ℓ_1 Norm Promotes Sparsity?

Given two N-dimensional signals:

- $x_1 = (1, 0, ..., 0) \rightarrow$ "Spike" signal
- $x_2 = (1/\sqrt{N}, 1/\sqrt{N}, ..., 1/\sqrt{N}) \rightarrow$ "Comb" signal



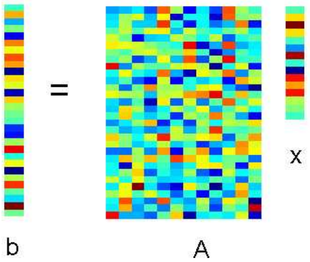
- x_1 and x_2 have the same ℓ_2 norm: $\|x_1\|_2 = 1$ and $\|x_2\|_2 = 1$.
- However, $\|x_1\|_1 = 1$ and $\|x_2\|_1 = \sqrt{N}$.

ℓ_1 Norm in Regression

- Linear regression is widely used in science and engineering.

$$\text{Given} \quad A \in R^{m \times n} \quad \text{and} \quad b \in R^m; \;\; m > n$$

$$\text{Find} \quad x \;\; \text{s.t.} \;\; b = Ax \quad \text{(overdetermined)}$$



ℓ_1 Norm Regression

Two approaches:

- Minimize the ℓ_2 norm of the residuals

$$\min_{x \in R^n} \|b - Ax\|_2$$

The ℓ_2 norm penalizes large residuals

- Minimizes the ℓ_1 norm of the residuals

$$\min_{x \in R^n} \|b - Ax\|_1$$

The ℓ_1 norm puts much more weight on small residuals

Matlab Code

- $\min_{x \in R^n} \|Ax - b\|_2$

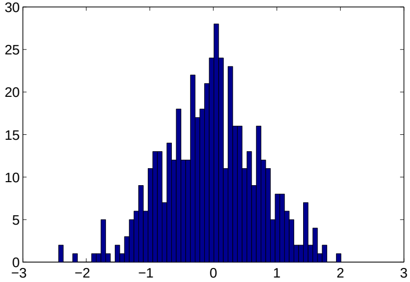
A=randn(500,150);
b=randn(500,1);
$x = (A' * A)^{(-1)} * A' * b$; Least Squares Solution

- $\min_{x \in R^n} \|Ax - b\|_1$

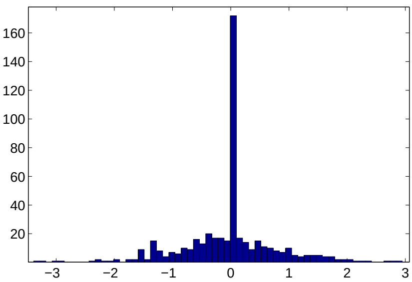
A=randn(500,150);
b=randn(500,1);
*X = medrec(b,A,max(A'*b),0,100,1e-5);*

ℓ_1 Norm Regression

$m = 500$, $n = 150$. $A = randn(m, n)$ and $b = randn(m, 1)$



ℓ_2 Residuals

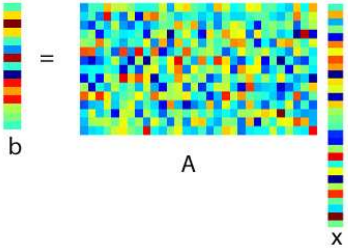


ℓ_1 Residuals

ℓ_1 Norm in Regression

Given $A \in R^{m \times n}$ and $b \in R^m$; $m < n$

Find x s.t. $b = Ax$ (underdetermined)



ℓ_1 Norm Regression

Two approaches:

- Minimize the ℓ_2 norm of x

$$\min_{x \in R^n} \|x\|_2 \quad \text{subject to} \quad Ax = b$$

- Minimize the ℓ_1 norm of x

$$\min_{x \in R^n} \|x\|_1 \quad \text{subject to} \quad Ax = b$$

Matlab Code

- $\min_{x \in R^n} \|x\|_2 \quad$ subject to $\quad Ax = b$

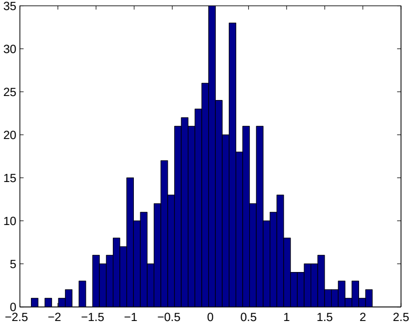
```
A=randn(150,500);
b=randn(150,1);
C=eye(150,500);
d=zeros(150,1);
X=lsqlin(C,d,[],[],A,b);
```

- In general:
  $\min_{x \in R^n} f(x) \quad$ subject to $\quad Ax = b$

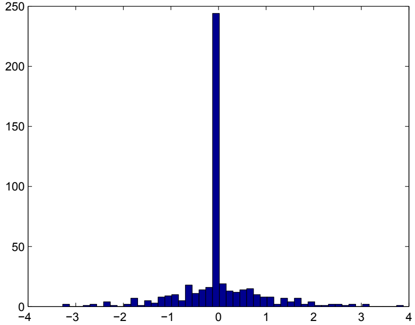
```
X= fmincon(@(x) f(x),zeros(500,1),[],[],A,b,[],[],options);
```

where $f(x)$ is a convex function.

ℓ_1 Norm Regression



ℓ_2 Solution



ℓ_1 Solution

ℓ_1 Norm Regression

Consider N observation pairs (x_i, b_i) modeled in a linear fashion

$$b_i = Ax_i + c + U_i, \qquad i = 1, 2, \ldots, N \tag{6}$$

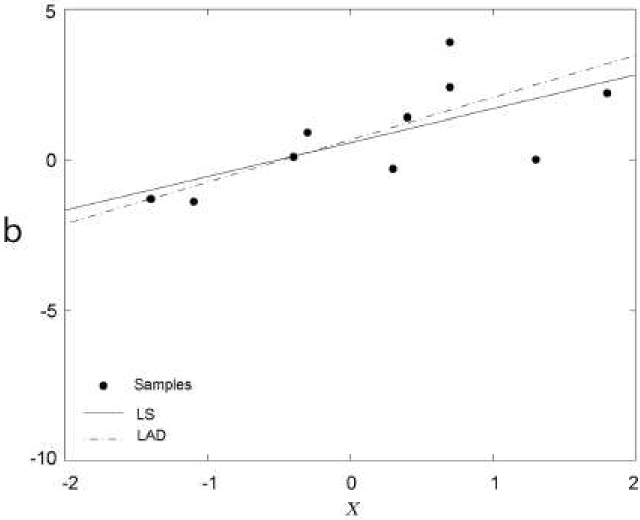
A: Unknown slope of the fitting line.

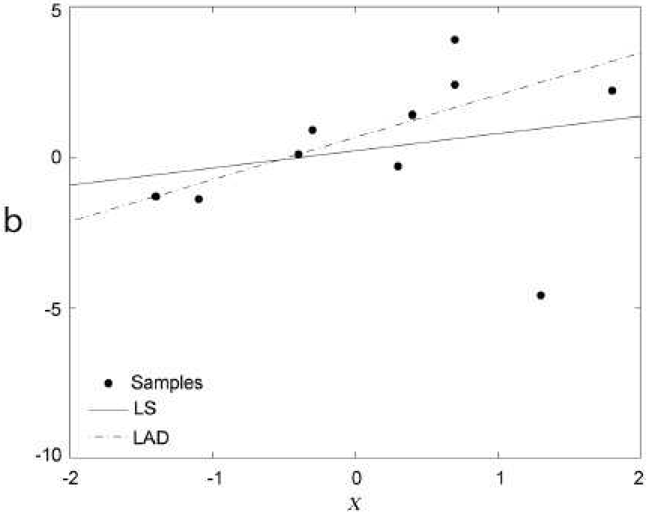
c: Intercept.

U_i: Unobservable errors

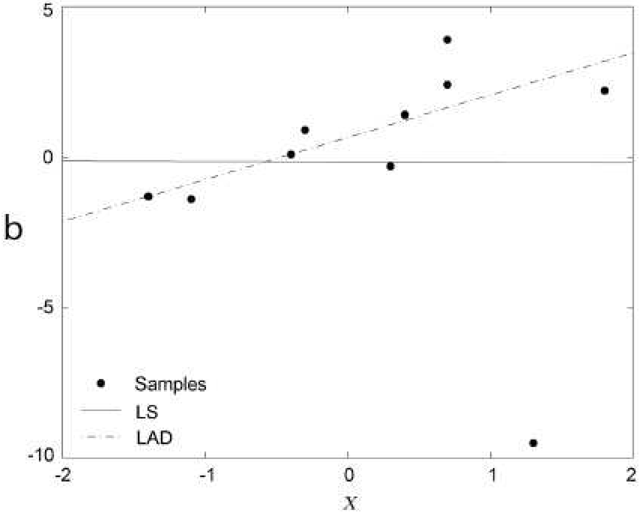
The Least Absolute Deviation regression is

$$F_1(A, c) = \sum_{i=1}^{N} |b_i - Ax_i - c|, \tag{7}$$

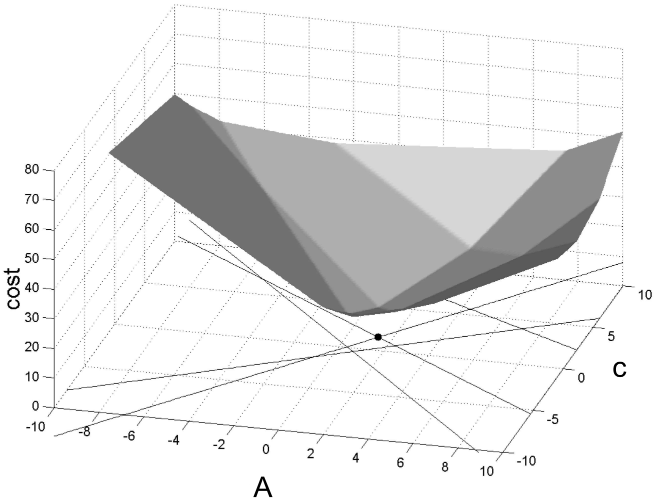

5
0
-5
-10
b
-2
-1
0
1
2
X
Samples
LS
LAD



5
0
-5
-10
b
-2
-1
0
1
2
x
Samples
LS
LAD


5
0
-5
-10
b
-2
-1
0
1
2
X
Samples
LS
LAD

$\sum_{i=1}^{N} |b_i - Ax_i - c|$ $\qquad$ $c = -x_i A + b_i$



ℓ_1 Norm in Estimation

Location Estimate in Gaussian Noise

Let $x_1, x_2, \cdots, x_N$, i.i.d. Gaussian with a constant but unknown mean β. The Maximum Likelihood estimate of location is the value $\hat{\beta}$ which maximizes the likelihood function

$$
\begin{aligned}
f(x_1, x_2, \cdots, x_N; \beta) &= \prod_{i=1}^{N} f(x_i - \beta) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\beta)^2/2\sigma^2} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} e^{-\sum_{i=1}^{N}(x_i-\beta)^2/2\sigma^2}.
\end{aligned}
\tag{8}
$$

ℓ_1 Norm in Estimation

The ML estimate $\hat{\beta}$ minimizes the least squares sum

$$\hat{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} (x_i - \beta)^2. \tag{9}$$

Results in the sample mean

$$\hat{\beta}_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{10}$$

ℓ_1 Norm in Estimation

Location Estimate in Generalized Gaussian Noise

If the $x's$ obey a generalized Gaussian distribution, the ML estimate of location is

$$\begin{aligned} f(x_1, x_2, \cdots, x_N; \beta) &= \prod_{i=1}^{N} f_\gamma(x_i - \beta) \\ &= \prod_{i=1}^{N} C\, e^{-|x_i-\beta|^\gamma/\sigma} \\ &= C^N e^{-\sum_{i=1}^{N}|x_i-\beta|^\gamma/\sigma}, \end{aligned} \tag{11}$$

where C is a normalizing constant, and γ is the dispersion parameter.

ℓ_1 Norm in Estimation

Maximizing the likelihood function is equivalent to

$$\tilde{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} |x_i - \beta|^{\gamma}.$$

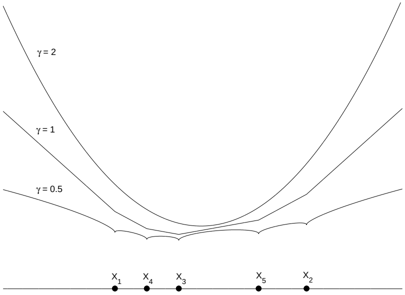


Figure: Cost function for $\gamma = 0.5, 1$, and 2.

ℓ_1 Norm in Estimation

For N odd there is an integer k, such that the slopes over the intervals $(x_{(k-1)}, x_{(k)}]$ and $(x_{(k)}, x_{(k+1)}]$, are negative and positive, respectively.

$$\begin{aligned}
\hat{\beta}_{ML} &= \arg\min_{\beta} \sum_{i=1}^{N} |x_i - \beta| \\
&= \begin{cases} x_{\left(\frac{N+1}{2}\right)} & N \; odd \\ \left(x_{\left(\frac{N}{2}\right)}, x_{\left(\frac{N}{2}\right)}\right] & N \; even \end{cases} \\
&= \text{MEDIAN}(x_1, x_2, \cdots, x_N). \qquad (12)
\end{aligned}$$

ℓ_1 Norm Regression

ML Estimate of Location for Generalized Gaussian

Here the samples have a common location parameter β, but different scale parameter σ_i. The ML estimate of location is

$$G_p(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i^p} |x_i - \beta|^p. \tag{13}$$

For the Gaussian distribution ($p = 2$), the ML estimate reduces to

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (x_i - \beta)^2 = \frac{\sum_{i=1}^{N} W_i \cdot x_i}{\sum_{i=1}^{N} W_i} \tag{14}$$

where $W_i = 1/\sigma_i^2 > 0$.

For the Laplacian distribution ($p = 1$), the ML estimate minimizes

$$G_1(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i} |x_i - \beta|. \tag{15}$$

where $W_i \triangleq 1/\sigma_i > 0$. $G_1(\beta)$ is piecewise linear and convex. The weighted median output is defined as

$$\begin{aligned} Y(n) &= \arg\min_{\beta} \sum_{i=1}^{N} W_i |x_i - \beta| \\ &= \text{MEDIAN}[W_1 \Diamond x_1(n),\ W_2 \Diamond x_2(n), \cdots,\ W_N \Diamond x_N(n)] \end{aligned}$$

where $W_i > 0$ and $\Diamond$ is the replication operator defined as

$$W_i \Diamond x_i = \overbrace{x_i,\ x_i, \cdots,\ x_i}^{w_i \text{ times}}.$$

ℓ_1 Norm Regression

Next, consider N observation pairs (x_i, b_i)

$$b_i = Ax_i + c + U_i, \qquad i = 1, 2, \ldots, N \tag{16}$$

A: Unknown slope of the fitting line.

c: Intercept.

U_i: Unobservable errors

The L_1 or Least Absolute Deviation (LAD) regression is

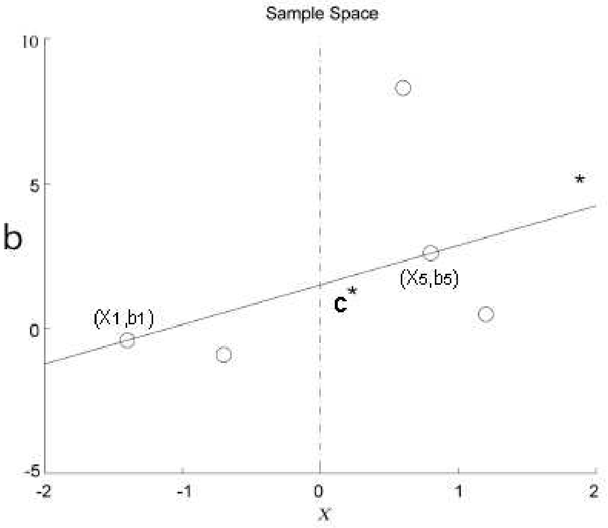
$$F_1(A, c) = \sum_{i=1}^{N} |b_i - Ax_i - c|, \tag{17}$$

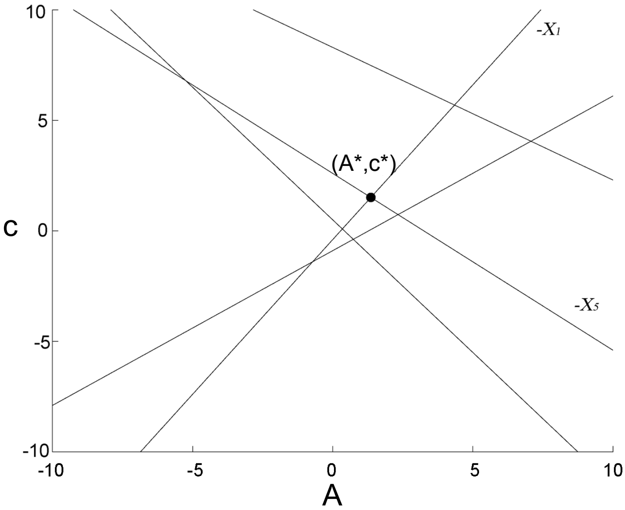
Sample space: $b_i = Ax_i + c$

1. Each sample pair (x_i, b_i) represents a point on the plane
2. The solution is a line with slope A^* and intercept c^*.
3. If this line goes through some sample pair (x_i, b_i), then the equation $b_i = A^* x_i + c^*$ is satisfied

Parameter space: $c = -x_i A + b_i$

1. The solution (A^*, b^*) is a point.
2. The sample pair (x_i, b_i) defines a line with slope $-x_i$ and intercept b_i.
3. When $c^* = -x_i A^* + b_i$ holds, it can be inferred that the point (A^*, c^*) is on the line defined by $(-x_i, b_i)$


Sample Space
10
5
0
-5
b
(X1,b1)
(X5,b5)
c*
*
-2
-1
0
1
2
X


10
5
0
-5
-10
c
-10
-5
0
5
10
A
-X1
-X5
(A*,c*)

Set $A = A_0$, the objective function now becomes a one-parameter function of c

$$F(c) = \sum_{i=1}^{N} |\underbrace{b_i - A_0 x_i}_{\text{Observations}} - c|. \tag{18}$$

The parameter c^* is the Maximum Likelihood estimator of location for c. It can be obtained by

$$c^* = \text{MED}(b_i - A_0 x_i) \mid_{i=1}^{N}. \tag{19}$$

Set $c = c_0$, the objective function reduces to

$$\begin{aligned} F(a) &= \sum_{i=1}^{N} |b_i - c_0 - Ax_i| \\ &= \sum_{i=1}^{N} |x_i| \left| \frac{b_i - c_0}{x_i} - A \right|. \end{aligned} \tag{20}$$

The parameter A^* can be seen as the ML estimator of location for A, and can be calculated as the *weighted median*,

$$A^* = \text{MED}\left(|x_i| \diamond \frac{b_i - c_0}{x_i} \right) \Big|_{i=1}^{N}, \tag{21}$$

A simple and intuitive way of solving the LAD regression problem is:

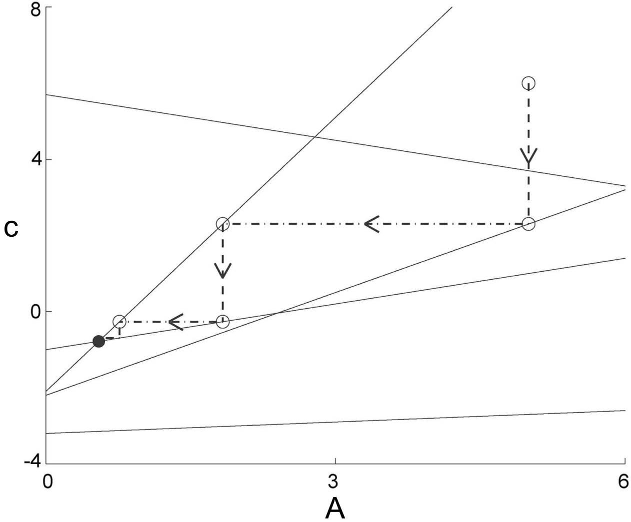
1. Set $k = 0$. Find an initial value A_0 for A, such as the Least Squares (LS) solution.
2. Set $k = k + 1$ and obtain a new estimate of c for a fixed A_{k-1} using

$$c_k = \text{MED}(b_i - A_{k-1}x_i) \mid_{i=1}^{N}.$$

3. Obtain a new estimate of A for a fixed c_k using

$$A_k = \text{MED}\left(|x_i| \diamond \frac{b_i - c_k}{x_i} \right) \Bigg|_{i=1}^{N}.$$

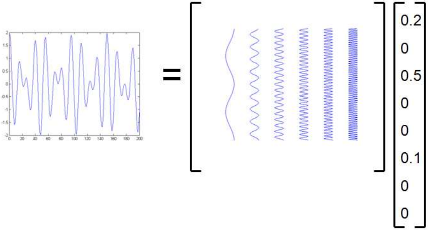
4. Once A_k and c_k do not deviate from A_{k-1} and c_{k-1} within a tolerance range, end the iteration. Otherwise, go back to step 2).


8
4
0
-4
c
0
3
6
A

Signal Representation

- A sparse signal $x \in R^N$ can be represented by a linear combination of basis of an orthogonal representation matrix Ψ

$$x(t) = \sum_i \alpha_i \psi_i(t)$$



Sparse Signal Representation

Active development for effective signal representation in the 90's

- Fourier
- Wavelet
- Curvelet

There is no universal best representation

- Best representation $=$ sparsest

Wavelets

A wavelet is a ”small wave” with finite energy that allows the analysis of transient, or time-varying phenomena.

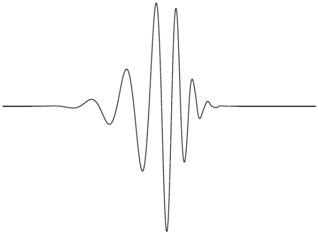


Figure: Daubechies (D20) Wavelet example

A signal $x(t)$ can be represented in terms of its wavelet coefficients as

$$x(t) = \sum_{j\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} \langle x, \Psi_{j,n}\rangle \Psi_{j,n}(t)$$

where:

- $\Psi_{j,n}$ are the wavelets that form an orthogonal basis.
- $\langle x, \Psi_{j,n}\rangle$ are the wavelet coefficients.

Wavelets are vectors of a orthogonal basis formed by shifting and dilating a *mother wavelet*, $\Psi(t)$:

$$\Psi_{j,n}(t) = 2^{-j/2}\Psi(2^{-j}t - n), \quad \forall j, n \in \mathbb{Z}$$

where j is the scale parameter and n is the location parameter.

Examples of wavelet expansion functions are:

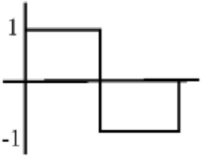


Figure: Haar wavelet

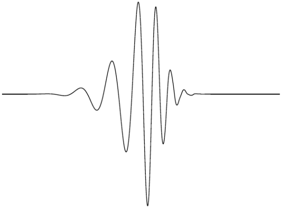


Figure: Daubechies wavelet

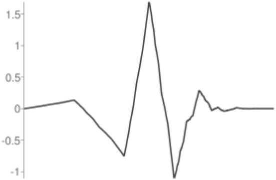


Figure: Symlet wavelet

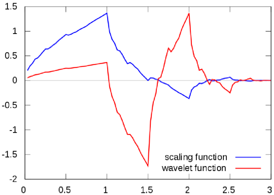
Daubechies Wavelet

- Daubechies Wavelets are continuous and smooth wavelets.
- The *mother wavelet* is defined by means of a *scaling function*.
- A daubechies wavelet $\Psi(t)$ has $p-1$ vanishing moments if:

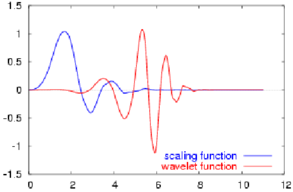
$$\int_{-\infty}^{\infty} t^k \Psi(t) dt = 0; \quad \text{for} \quad 0 \leq k < p.$$

- The smoothness of the scaling and wavelet functions increase as the number of vanishing moments increases.

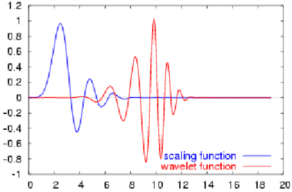
Examples of Daubechies wavelets:



(a)



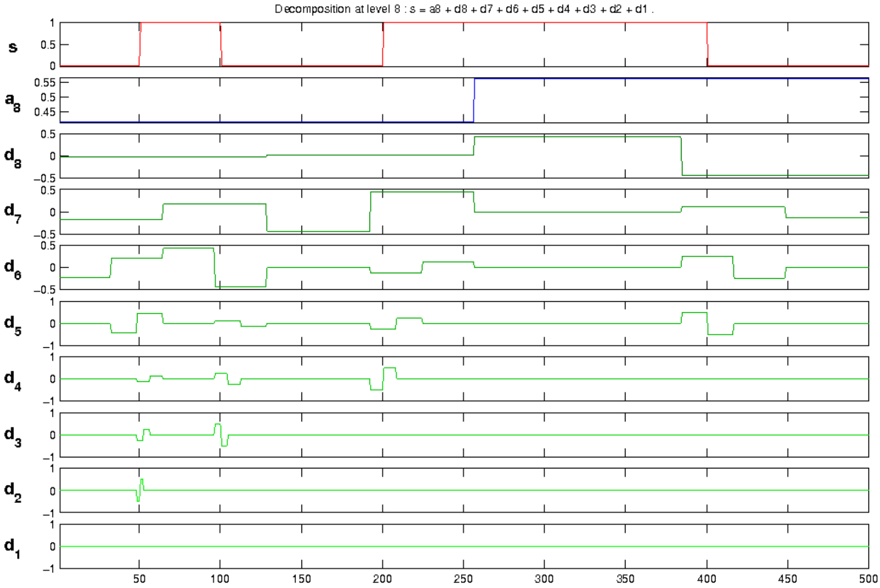
(b)



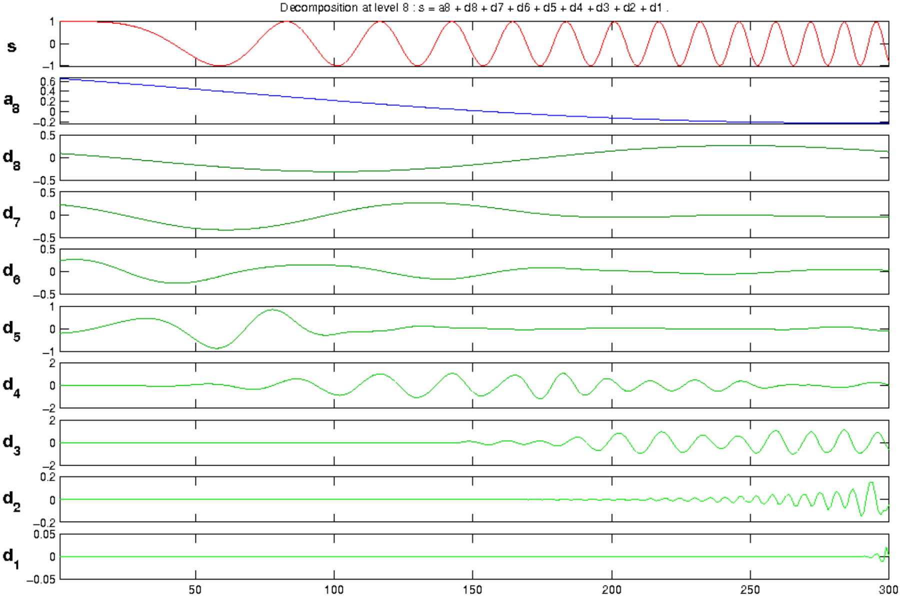
(c)

- (a) Daubechies scaling and wavelet functions with 2 vanishing moments.
- (b) Daubechies scaling and wavelet functions with 6 vanishing moments.
- (c) Daubechies scaling and wavelet functions with 10 vanishing moments.

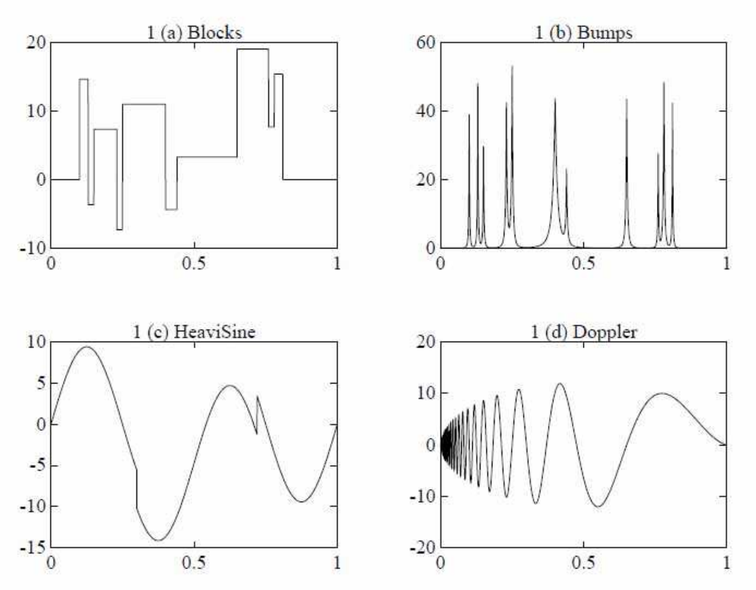
Examples of Wavelet decompositions



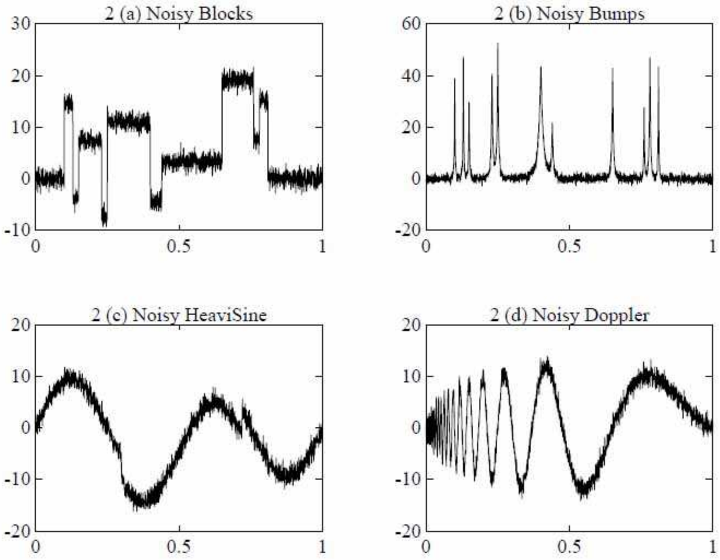
Examples of Wavelet decompositions



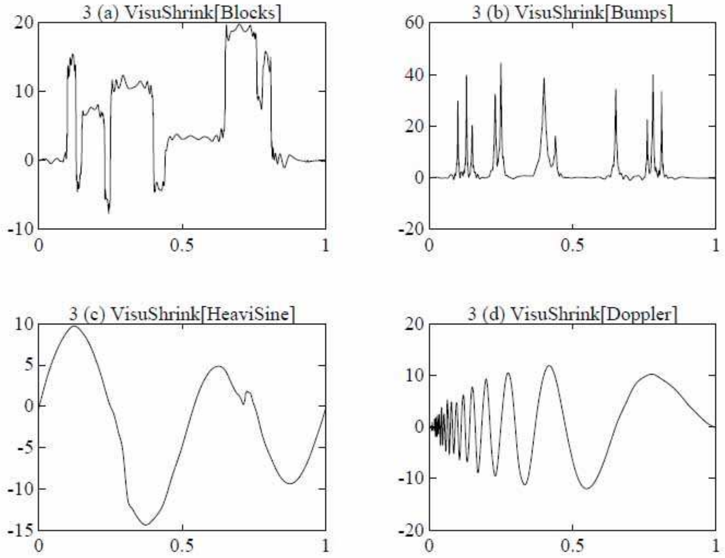
Other examples: original signals



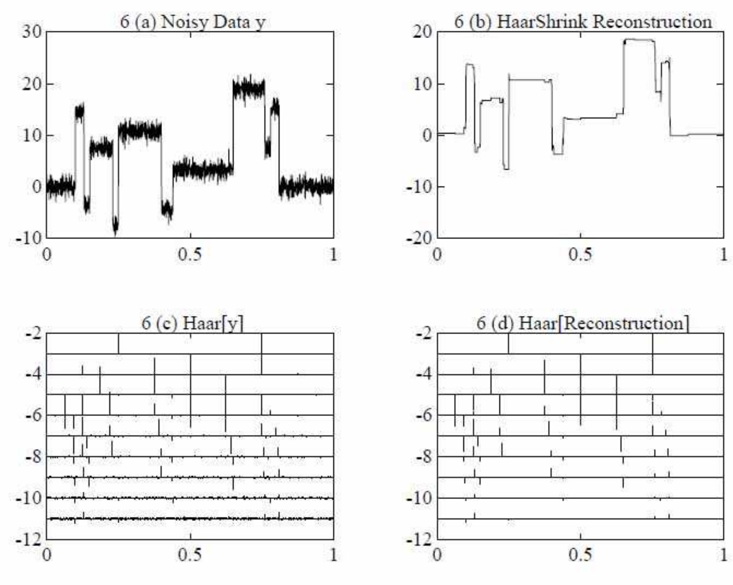
Noisy signals



Denoising using wavelet approximation



Denoising using wavelet approximation



Sampling and Compression

JPEG

JPEG2000

JPEG

JPEG2000

Sparse Signal Representation

- Different representations are best for different applications.
  - Fourier Dictionary $\rightarrow$ For oscillatory phenomena
  - Wavelet Dictionary $\rightarrow$ For images with isolated singularities
  - Curvelet Dictionary $\rightarrow$ For images with contours and edges

This motivates overcomplete signal representation ‡

‡ S. Mallat and Z. Zhang. ”Matching Pursuit in a Time-Frequency Dictionary”. IEEE Trans. on Signal Proc. Vol.41, pp.3397-3415, 1993.

Sparse Signal Representation

Overcomplete dictionary representation

- Different bases merged into a combined dictionary

$$\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_N]$$

- Representation of x in an overcomplete dictionary

$$x = \sum_i \alpha_i \psi_i, \quad \text{with the sparsest} \quad \alpha$$

Basis Pursuit (BP)

Basis Pursuit $\rightarrow$ find the sparsest approximation of x

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \;\; x = \Psi\alpha$$

where $\|\alpha\|_1 = \sum_i |\alpha_i|$.

- BP decomposes a signal into a superposition of dictionary elements having the smallest ℓ_1-norm among all such decompositions.

† D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inform. Theory, 47:2845-2862, 2001.

Compressible Signals

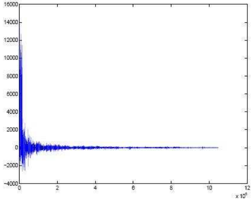
In most applications

- Signals are not perfectly sparse, but only a few coefficients concentrate most of the energy.
- Most of the transform coefficients are negligible.
- Compressible signals can be approximated by a S-sparse signal:
  - There is a transform vector α_S with only S terms such that $\|\alpha_S - \alpha\|_2$ is small.

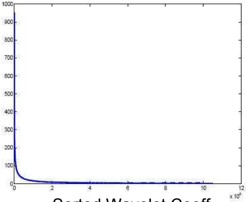
Compressible Signals

- Wavelet coefficients of natural scenes exhibit the $(1/n)$-decay[†].

1 Megapixel Image



Wavelet Coefficients

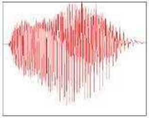


Sorted Wavelet Coeff.

† E. J. Candès and J. Romberg "Sparsity and Incoherence in Compressive Sampling." Inverse Problems. vol.23, pp.969-985. 2006.

Examples of Compressible Signals

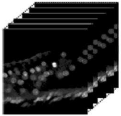
- Bat echolocation



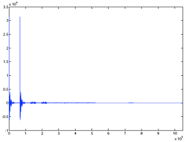
Time Signal

Time-Frequency Representation

- Confocal microscopy

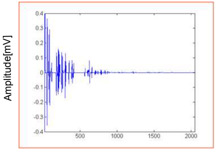


3D Image

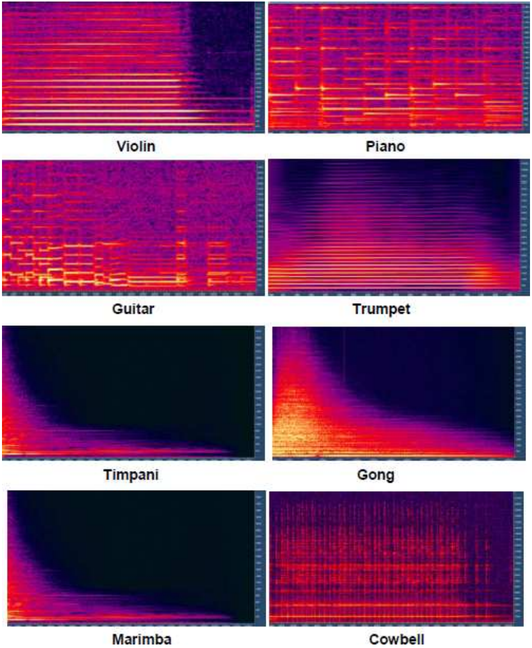


3D Wavelet Coefficients

- Ultra wideband signaling



Time[ps]


Violin
Piano
Guitar
Trumpet
Timpani
Gong
Marimba
Cowbell