Near optimal recovery of arbitrary signals from highly incomplete measurements

Albert Cohen

Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie
Paris

3-10-2008

A classical result in signal processing : Shannon's theorem

Let f be a band-limited function of one variable, i.e. such that $\mathrm{Supp}(\hat{f}) \subset [-F, F]$.

Then f can be exactly reconstructed from its samples $f(nT)$, $n \in \mathbb{Z}$ with $T = \pi/F$.

Many generalizations : several variables, local averages instead of exact values...

A classical result in signal processing : Shannon's theorem

Let f be a band-limited function of one variable, i.e. such that $\text{Supp}(\hat{f}) \subset [-F, F]$.

Then f can be exactly reconstructed from its samples $f(nT)$, $n \in \mathbb{Z}$ with $T = \pi/F$.

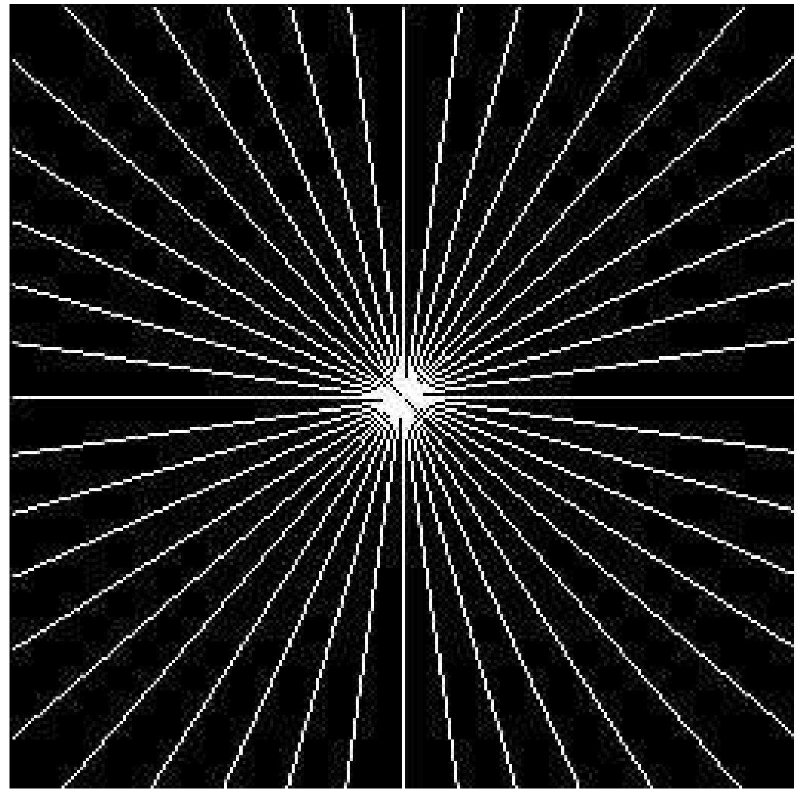
Many generalizations : several variables, local averages instead of exact values...

Problem : most interesting signals are not band-limited, typically due to sharp transitions, such as edge discontinuities in images.

Can we still hope to recover such signals with only a few measurements ?

An instructive test case : 2D tomography (Candes-Romberg-Tao)

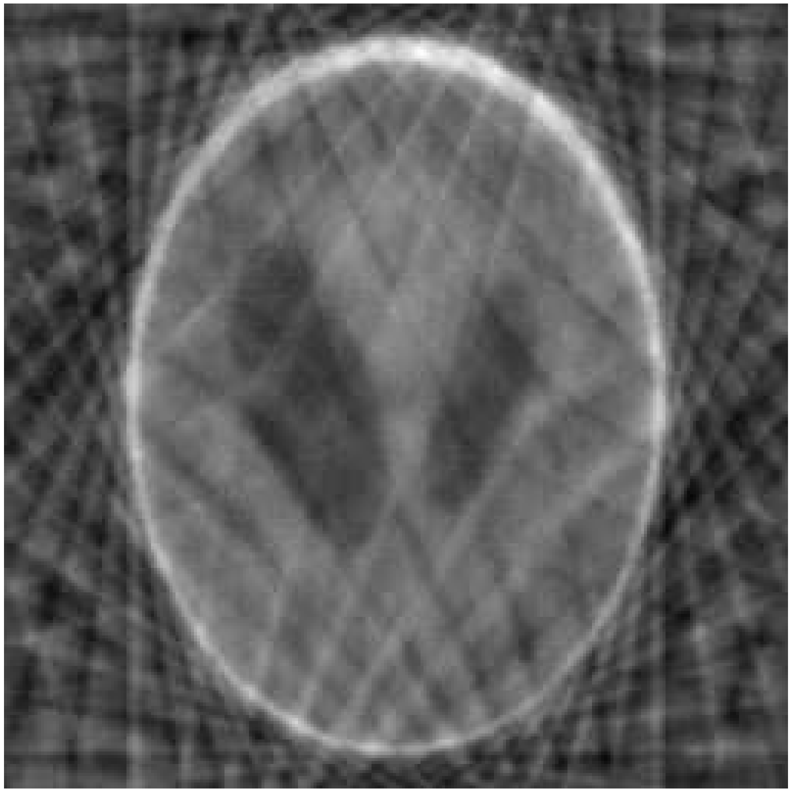
The Radon transform captures partial Fourier information.



Left : the Logan-Shep phantom test image

Right : position of the observed Fourier coefficients (white)

Two different reconstructions



Left : put the unknown coefficient to zero (minimum ℓ^2 norm) and reconstruct the partial Fourier serie $\Rightarrow$ oscillation artifacts.

Right : adjust the unknown coefficients so to minimize the total variation of the image $|f|_{TV} = \int |\nabla f|$ $\Rightarrow$ exact reconstruction !

Related work

1. Pioneering ideas in Logan's PhD thesis (1965) and seminal paper by Donoho-Stark (1989).

2. Use of TV criterion in image processing: foundational paper by Osher-Rudin-Fatemi (1992)

3. Reconstruction of piecewise smooth functions from low frequency information (since 1990's): Gottlieb-Tadmor, Vetterli-Dragotti.

4. Randomized algorithms for approximation and reconstruction (since 1990's): Strauss, Gilbert-Tropp, Mutukrishnan...

5. Compressed sensing (since 2000's): Candes-Romberg-Tao, Elad, Donoho-Tsaig-Tanner, Baraniuk-Wakin, Vershunyin-Rudelson...

6. High-dimensional geometry (since 1970's): Kashin, Tomczak-Jaegermann, Pajor-Mendelson...

The compressed sensing problem

- $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ unknown object of high dimension $n >> 1$.
- We observe $y = \Phi x \in \mathbb{R}^m$, m linear measurements of x, $m << n$.
- Φ is the $(m \times n)$ encoding matrix.

The compressed sensing problem

- $x = (x_1, \cdots, x_n) \in \mathrm{I\!R}^n$ unknown object of high dimension $n >> 1$.
- We observe $y = \Phi x \in \mathrm{I\!R}^m$, m linear measurements of x, $m << n$.
- Φ is the $(m \times n)$ encoding matrix.

Goal: recover from y an approximation $\Delta(y)$ of x which is as accurate as possible in some sense.

- $\Delta : \mathrm{I\!R}^m \to \mathrm{I\!R}^n$ is the "decoder" (in general non-linear).

The compressed sensing problem

- $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ unknown object of high dimension $n >> 1$.

- We observe $y = \Phi x \in \mathbb{R}^m$, m linear measurements of x, $m << n$.

- Φ is the $(m \times n)$ encoding matrix.

Goal: recover from y an approximation $\Delta(y)$ of x which is as accurate as possible in some sense.

- $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the "decoder" (in general non-linear).

Severely ill posed inverse problem: Φ is non-injective.

Admissible solutions:. $\mathcal{F}(y) := \{z \text{ s.t. } \Phi z = y\} = x + \mathcal{N}$ with $\mathcal{N} = \mathcal{N}(\Phi)$ the null space of Φ. Note that $\dim(\mathcal{N}) \geq n - m$.

The compressed sensing problem

- $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ unknown object of high dimension $n >> 1$.

- We observe $y = \Phi x \in \mathbb{R}^m$, m linear measurements of x, $m << n$.

- Φ is the $(m \times n)$ encoding matrix.

Goal: recover from y an approximation $\Delta(y)$ of x which is as accurate as possible in some sense.

- $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the "decoder" (in general non-linear).

Severely ill posed inverse problem: Φ is non-injective.

Admissible solutions:. $\mathcal{F}(y) := \{z \text{ s.t. } \Phi z = y\} = x + \mathcal{N}$ with $\mathcal{N} = \mathcal{N}(\Phi)$ the null space of Φ. Note that $\dim(\mathcal{N}) \geq n - m$.

Important features : (i) we make some a-priori assumptions on x (ii) we get to choose Φ

The compressed sensing problem

- $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ unknown object of high dimension $n >> 1$.

- We observe $y = \Phi x \in \mathbb{R}^m$, m linear measurements of x, $m << n$.

- Φ is the $(m \times n)$ encoding matrix.

Goal: recover from y an approximation $\Delta(y)$ of x which is as accurate as possible in some sense.

- $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the "decoder" (in general non-linear).

Severely ill posed inverse problem: Φ is non-injective.

Admissible solutions:. $\mathcal{F}(y) := \{z \text{ s.t. } \Phi z = y\} = x + \mathcal{N}$ with $\mathcal{N} = \mathcal{N}(\Phi)$ the null space of Φ. Note that $\dim(\mathcal{N}) \geq n - m$.

Important features : (i) we make some a-priori assumptions on x (ii) we get to choose Φ

Noisy version: observe $y = \Phi x + e$ with $\|e\|_{\ell^2} \leq \varepsilon$.

Sparse signals

The vector x is k-sparse if at most k of its coordinates are non-zero.

$$\Sigma_k = \{x \in \mathbb{R}^{\mathrm{n}} \; ; \; \|\mathrm{z}\|_{\ell^0} := |\mathrm{Supp}(\mathrm{x})| \leq \mathrm{k}\}$$

Example of sparse signal

Digital Image 512x512

Wavelet Decomposition

Multiscale decompositions of natural images into wavelet bases are quasi-sparse: a few numerically significant coefficients concentrate most of the energy and information.

Sparse signals

The vector x is k-sparse if at most k of its coordinates are non-zero.

$$\Sigma_k = \{x \in \mathbb{R}^n \; ; \; \|z\|_{\ell^0} := |\mathrm{Supp}(x)| \leq k\}$$

Sparse up to a change of basis : $y = \Phi x = \Phi M \tilde{x} = \tilde{\Phi} \tilde{x}$ with $\tilde{x} \in \Sigma_k$.

Question: how many linear measurements are needed to characterize any k-sparse signal ?

Sparse signals

The vector x is k-sparse if at most k of its coordinates are non-zero.

$$\Sigma_k = \{x \in \mathbb{R}^n \; ; \; \|z\|_{\ell^0} := |\mathrm{Supp}(x)| \leq k\}$$

Sparse up to a change of basis : $y = \Phi x = \Phi M \tilde{x} = \tilde{\Phi} \tilde{x}$ with $\tilde{x} \in \Sigma_k$.

Question: how many linear measurements are needed to characterize any k-sparse signal ?

Answer: $m = 2k$.

Theorem: the following properties are equivalent

(i) if $x, x' \in \Sigma_k$ and $\Phi x = \Phi x'$, then $x = x'$.

(ii) there exists a decoder Δ such that $\Delta(\Phi x) = x$ for all $x \in \Sigma_k$

(iii) $\mathcal{N} \cap \Sigma_{2k} = \{0\}$

(iv) any $2k$ columns picked in Φ are independent.

These properties can be fullfilled by matrices Φ with $m = 2k$ rows.

Example: take $a_1 < a_2 \cdots < a_n$ and define $\Phi = (a_j^i)$ for $i = 1, \cdots 2k$.

These properties can be fullfilled by matrices Φ with $m = 2k$ rows.

Example: take $a_1 < a_2 \cdots < a_n$ and define $\Phi = (a_j^i)$ for $i = 1, \cdots 2k$.

Decoding: define $\Delta(y) = \mathrm{Argmin}_{\mathrm{z} \in \Sigma_\mathrm{k}} \|y - \Phi z\|_{\ell_2}^2$

Equivalently: for each $T \subset \{1, \cdots, n\}$ with $|T| = k$, we consider the matrix Φ_T consisting of the columns of Φ with indices in T and solve the normal equation

$$z_T := (\Phi_T^t \Phi_T)^{-1} \Phi_T^t y,$$

then minimize $\|y - \Phi_T z_T\|_{\ell^2}$ over all T.

Alternate approach : sparsest solution $\Delta(y) = \mathrm{Argmin}_{\Phi z = y} \|z\|_{\ell^0}$.

These properties can be fullfilled by matrices Φ with $m = 2k$ rows.

Example: take $a_1 < a_2 \cdots < a_n$ and define $\Phi = (a_j^i)$ for $i = 1, \cdots 2k$.

Decoding: define $\Delta(y) = \mathrm{Argmin}_{z \in \Sigma_k} \|y - \Phi z\|_{\ell_2}^2$

Equivalently: for each $T \subset \{1, \cdots, n\}$ with $|T| = k$, we consider the matrix Φ_T consisting of the columns of Φ with indices in T and solve the normal equation

$$z_T := (\Phi_T^t \Phi_T)^{-1} \Phi_T^t y,$$

then minimize $\|y - \Phi_T z_T\|_{\ell^2}$ over all T.

Alternate approach : sparsest solution $\Delta(y) = \mathrm{Argmin}_{\Phi z = y} \|z\|_{\ell^0}$.

Two main objections to these approaches:

- computationally irrealistic, since one needs to solve $\binom{n}{k}$ systems.

- Φ_T non-singular but might be very ill-conditionned for some T's.

The restricted isometry property

Introduced in the work by Candes-Romberg-Tao and Donoho.

The encoding matrix Φ satisfies the restricted isometry property (RIP) at order k with parameter $\delta \in (0,1)$ if the following hold: for all $T \subset \{1, \cdots, n\}$ with $|T| \leq k$ one has

$$(1-\delta)I \leq \Phi_T^t \Phi_T \leq (1+\delta)I$$

The restricted isometry property

Introduced in the work by Candes-Romberg-Tao and Donoho.

The encoding matrix Φ satisfies the restricted isometry property (RIP) at order k with parameter $\delta \in (0,1)$ if the following hold: for all $T \subset \{1, \cdots, n\}$ with $|T| \le k$ one has

$$(1-\delta)I \le \Phi_T^t \Phi_T \le (1+\delta)I$$

Theorem (Candes-Romberg-Tao): if Φ satisfies RIP at order $3k$ with $\delta \le 1/3$, and if we define the decoder

$$\Delta(y) := \text{Argmin}_{\Phi z = y} \|z\|_{\ell^1},$$

then $\Delta(\Phi x) = x$ for any $x \in \Sigma_k$. This decoder is implementable in $\mathcal{O}(n^3)$ complexity. Robustness: $\|x - \Delta(\Phi x + e)\|_{\ell^2} \le C\|e\|_{\ell^2}$.

Interpretation: convex relaxation of the ℓ^0 minimization, the minimum ℓ^1 solution of $\Phi z = y$ is also the sparsest one.

The price to pay

- RIP at order k achieved by matrices Φ with $m \sim k \log(n/k)$ rows.

- Existing constructions of such Φ are all based on probabilities, i.e. generate random matrices $\Phi = \Phi(\omega)$ such that RIP holds with high probability. Examples:

1. Entries $\Phi_{i,j}$ are i.i.d. Bernoulli variables $\frac{\pm 1}{\sqrt{m}}$.

2. Entries $\Phi_{i,j}$ are i.i.d. Gaussian variables $\mathcal{N}(0, 1/m)$.

3. m rows picked independently in the $n \times n$ Fourier matrix.

In all these examples, it can be shown that for any $\delta > 0$, there exists $c > 0$ such that when $m \geq ck \log(n/k)$ (in example 3 $m \geq ck(\log n)^5$) then Φ satisfies RIP of order k with parameter δ, with probability $1 - \varepsilon(m)$ where $\varepsilon(m)$ decreases fastly to 0 as m grows.

The price to pay

- RIP at order k achieved by matrices Φ with $m \sim k \log(n/k)$ rows.

- Existing constructions of such Φ are all based on probabilities, i.e. generate random matrices $\Phi = \Phi(\omega)$ such that RIP holds with high probability. Examples:

1. Entries $\Phi_{i,j}$ are i.i.d. Bernoulli variables $\frac{\pm 1}{\sqrt{m}}$.

2. Entries $\Phi_{i,j}$ are i.i.d. Gaussian variables $\mathcal{N}(0, 1/m)$.

3. m rows picked independently in the $n \times n$ Fourier matrix.

In all these examples, it can be shown that for any $\delta > 0$, there exists $c > 0$ such that when $m \geq ck \log(n/k)$ (in example 3 $m \geq ck(\log n)^5$) then Φ satisfies RIP of order k with parameter δ, with probability $1 - \varepsilon(m)$ where $\varepsilon(m)$ decreases fastly to 0 as m grows.

Important issues

1. Derandomize

2. Very fast algorithms

3. Non-sparse signals

Compressed sensing web site: rice university dsp group.