Compressive (or Compressed) Sensing
"Detección Comprimida"

Javier Peña
Carnegie Mellon University

UN Encuentro de Matemáticas
Universidad Nacional, July 2012

Materials available at
`http://andrew.cmu.edu/user/jfp/UNencuentro`

Compressive Sensing, Lecture 1

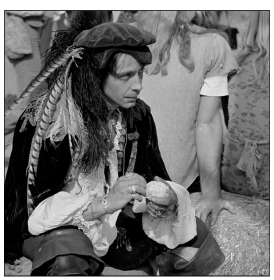
Consider the following two pictures

One of these is a raw jpg 2.7MB file. The other one is a compressed jpg 300KB version.

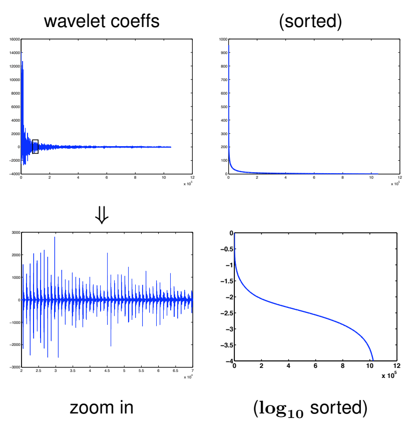
Can you tell them apart?

Wavelets and images

Compressibility corresponds to sparsity in a suitable basis.

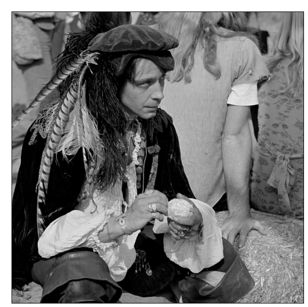


1 megapixel image



Compression

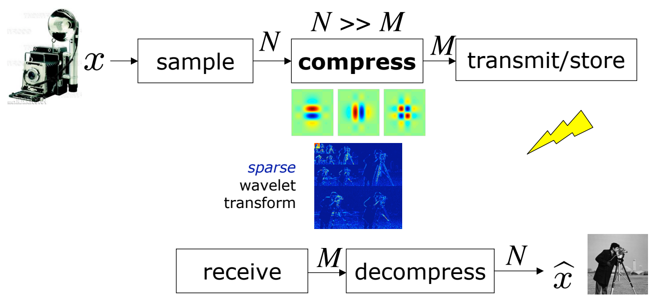
- Take original 1 megapixel image
- Compute all 1 million wavelet coefficients
- Keep only the 25K largest and set the others to zero
- Invert the wavelet transform



original image

25K approximation

Conventional signal acquisition paradigm



Question

If the signal is compressible, can it be acquired efficiently?

Compressive sensing

- Term coined by Donoho
- Body of theory and algorithms for sparse signal acquisition and recovery
- Seminal papers:
  - Candès, Romberg and Tao (2006)
  - Candès and Tao (2006)
  - Donoho (2006)
- Hot area of research spanning information theory, signal processing, statistics, mathematics, etc.
- Applications where measurements are
  - slow or costly (MRI)
  - missing or wasteful
  - beyond other capabilities such as memory

Plan

- Introduction to compressive sensing, undetermined systems of equations, ℓ_1 minimization
- Main theoretical and computational techniques
- Matrix completion, undetermined linear matrix equations, nuclear norm minimization.

Undetermined systems of equations

Problem

Recover a signal $\bar{x} \in \mathbb{R}^n$ from $m \ll n$ linear measurements

$$b_k = \langle a_k, \bar{x} \rangle, \ k = 1, \ldots, m \rightsquigarrow b = A\bar{x}$$

- In general this is impossible.
- Suppose we know that $\bar{x}$ is sparse. Does that help?

Example

Suppose only one component of $\bar{x}$ is different from zero.
Can we get by with fewer than n measurements?

Possible approach to recover sparse $\bar{x}$

Take $m \ll n$ measurements $b = A\bar{x}$ and then solve

$$\min \; \|x\|_0 \quad Ax = b$$

Here $\|\cdot\|_0$ stands for the ℓ_0 quasi-norm: $\|x\|_0 = |\{i : x_i \neq 0\}|$.

Relevant questions

- Does this work (provided we take enough measurements)?
- Suppose $\bar{x}$ is k-sparse, i.e., $\|\bar{x}\|_0 = k$. How many measurements suffice?
- How hard is it to solve the above ℓ_0-minimization problem?

ℓ_0 versus ℓ_1-optimization

$$\min \; \|x\|_0 \quad Ax = b$$

computationally hard

$$\min \; \|x\|_1 \quad Ax = b$$

computationally tractable (linear program)

- ℓ_1 norm is the convex envelope of the ℓ_0 quasi-norm.
- The ℓ_1 minimization problem is a *convex relaxation* of the ℓ_0 minimization problem.
- In many cases the above ℓ_1 minimization problem yields the same solution as the ℓ_0 problem.

A bit of history of ℓ_1 optimization

ℓ_1 minimization often finds the "right" answer

- Seismology
- Lasso regression
- Bandlimited deconvolution
- Total variation (TV) denoising
- Basis pursuit

Acquiring a sparse signal

Suppose $\bar{x} \in \mathbb{R}^n$ is s-sparse.

- Take m random and nonadaptive measurements

$$b_k = \langle a_k, \bar{x} \rangle, \; k = 1, \ldots, m \; \rightsquigarrow b = A\bar{x}$$

- Try to reconstruct $\bar{x}$ via ℓ_1 minimization.

First fundamental result

If $m \gtrsim s \cdot \log n$ and the a_k are suitably chosen, then the recovery is exact.

Nonadaptive sensing of compressible signals

Classical approach

- Measure the full signal $\bar{x}$ (all coefficients)
- Store s largest coefficients
- Distortion $\|\bar{x} - \bar{x}_s\|_2$

Compressive sensing

- Take m random measurements
- Reconstruct $\hat{x}$ via ℓ_1 minimization

Second fundamental result

If $m \gtrsim s \cdot \log n$ then

$$\|\hat{x} - \bar{x}\|_2 \lesssim \|\bar{x} - \bar{x}_s\|_2$$

Optimality of compressive sensing

It is not possible to do better

- with fewer measurements
- with other reconstruction algorithms

Key features of compressive sensing

- Obtain compressible signals from few sensors
- Sensing is nonadaptive: no knowledge about the signal
- Simple acquisition followed by ℓ_1 decoder

Next: formal statements

- Probabilistic approach
  - Isotropy and incoherence
  - Incoherent sampling theorems
- Deterministic approach
  - Restricted isometry property
  - Signal recovery theorem
- Robustness to noise
- Optimality

Probabilistic approach: random sensing

Acquire $\bar{x} \in \mathbb{R}^n$ by measuring

$$b_k = \langle a_k, \bar{x} \rangle, \ k = 1, \ldots, m \ \rightsquigarrow b = A\bar{x}$$

where a_k are iid F for some distribution F in $\mathbb{C}^n$.

Two key properties

- Isotropy: $\mathbb{E}(aa^*) = I$
- Coherence measure $\mu(F)$: smallest number such that

$$\max_{i=1,\ldots,n} |\langle a, e_i \rangle|^2 \leq \mu(F) \quad \text{with high probability}$$

We want low coherence $\mu(F)$.

Coherence

Observe

- $\mathbb{E}(aa^*) = I$ implies $\mu(F) \geq 1$.
- We would like $\mu(F)$ to be as close as possible to 1 (incoherent sensing).

Examples of isotropic incoherent sensing

- Gaussian sensing
- Binary sensing
- Partial Fourier transform

Notation: for $a \in \mathbb{C}^n$ and $i = 1, \ldots, n$

$$a[i] := \langle a, e_i \rangle.$$

Isotropic incoherent sensing

Gaussian sensing

$a \sim \mathcal{N}(0, I)$, that is, $a[1], \ldots, a[n]$ are iid $\mathcal{N}(0, 1)$.

In this case $\mu(F) = \log n$.

Binary sensing

$a[1], \ldots, a[n]$ are iid with distribution $\mathbb{P}(a[i] = \pm 1) = 1/2$.

In this case $\mu(F) = 1$.

Partial discrete Fourier transform

- Select k uniformly at random in $\{0, 1, \ldots, n-1\}$
- Set $a[t] := e^{i2\pi kt/n}, \;\; t = 0, 1 \ldots, n-1$.

In this case $\mu(F) = 1$.

An example of coherent sensing

Sample random components of x

- Select j uniformly at random in $\{1, \ldots, n\}$
- Set $a = \sqrt{n} e_j$.

In this case

$$\mathbb{E}(aa^*) = I$$

and

$$\max_{i=1,\ldots,n} |a[i]|^2 = n.$$

For this type of sensing, how many samples do we need to recover a 1-sparse vector with high probability?

Incoherent sampling theorems

Compressive sampling approach

measure

$$b_k = \langle a_k, \bar{x} \rangle, \; k = 1, \ldots, m \;\rightsquigarrow b = A\bar{x}$$

perform ℓ_1 recovery

$$\hat{x} := \underset{x}{\operatorname{argmin}} \{ \|x\|_1 : Ax = b \}$$

Theorem (Candès & Plan)

Assume $\bar{x}$ is s-sparse. Then recovery is exact with probability at least $1 - 5/n - e^{-\beta}$ provided that

$$m \geq C_0 \cdot (1 + \beta) \cdot \mu(F) \cdot s \cdot \log n$$

for some constant C_0.

Related predecessors:

Theorem (Candès & Tao)

Assume $\bar{x}$ is s-sparse and sample m Fourier coefficients selected at random. Then recovery is exact with probability at least $1 - \mathcal{O}(n^{-\beta})$ provided that

$$m \geq C_\beta \cdot s \cdot \log n$$

for some constant C_β that depends only on the desired accuracy β.

This result is optimal: any reliable recovery method would require at least $s \cdot \log n$ samples.

Theorem (Candès & Tao)

Assume $\bar{x} \in \mathbb{R}^n$ is s-sparse, n is prime, and we sample m Fourier coefficients. Then $\bar{x}$ can be reconstructed from the m samples if $m \geq 2s$.

Sampling of non-sparse signals

Given $x \in \mathbb{R}^n$ and $s < n$ define

$$x_s := \operatorname*{argmin}_{\|z\|_0 \leq k} \|z - x\|_2$$

Theorem (Candès & Plan)

Let $x \in \mathbb{R}^n$, $\beta > 0$ and $\bar{s}$ be such that $m \geq C_\beta \cdot \bar{s} \cdot \log n$. Then with probability at least $1 - 6/n - 6e^{-\beta}$ the ℓ_1 solution $\hat{x}$ satisfies

$$\|\hat{x} - x\|_1 \leq \min_{1 \leq s \leq \bar{s}} C \cdot (1 + \alpha) \cdot \|x - x_s\|_1$$

for some constant C and

$$\alpha = \sqrt{\frac{(1+\beta) s \mu(F) \log n \log m \log^2 s}{m}}.$$

Deterministic approach: restricted isometry

Restricted isometry property (RIP)

Given $A \in \mathbb{R}^{m \times n}$ and $k \in \{1, \ldots, m\}$, the k-isometry constant δ_k is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2$$

for all k-sparse $x \in \mathbb{R}^n$.

If $\delta_k < 1$, we say that A satisfies the RIP with constant δ_k.

Signal recovery with a RIP matrix

Compressive sampling approach

measure

$$b_k = \langle a_k, \bar{x} \rangle, \; k = 1, \ldots, m \; \rightsquigarrow b = A\bar{x}$$

perform ℓ_1 recovery

$$\hat{x} := \operatorname*{argmin}_x \{\|x\|_1 : Ax = b\}$$

Theorem (Candès, Romberg, Tao)

Assume $\delta_{2s} \leq \sqrt{2} - 1$. Then the solution $\hat{x}$ satisfies

$$\|\hat{x} - \bar{x}\|_1 \leq C \cdot \|\bar{x} - \bar{x}_s\|_1$$

and

$$\|\hat{x} - \bar{x}\|_2 \leq C \cdot \frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}$$

for some constant C.

Matrices that satisfy RIP

With high probability an $m \times n$ matrix A satisfies the RIP in the following cases:

- $m \gtrsim s \log(n/s)$ and A is Gaussian
- $m \gtrsim s \log(n/s)$ and A is binary
- $m \gtrsim s \log^4 n$ and A is partial DFT
- $m \gtrsim \mu(F) s \log^4 n$ and rows of A are iid F.

Sampling with noise

- Measurements in real life are generally noisy
- More appropriate model

$$b = A\bar{x} + z$$

noise term $z \sim \mathcal{N}(0, \sigma^2 I)$

- Assume all columns of A have Euclidean norm equal to one.
- Modify ℓ_1 minimization to account for noise

Lasso

$$\min \ \frac{1}{2}\|b - Ax\|_2^2 + \lambda \cdot \sigma \cdot \|x\|_1$$

Dantzig selector

$$\begin{array}{ll} \min & \|x\|_1 \\ & \|A^*(b - Ax)\|_\infty \le \lambda \cdot \sigma \end{array}$$

Noise aware recovery (random sensing)

Theorem (Candès & Plan)

Let $\bar{x} \in \mathbb{R}^n$, $\beta > 0$ and $\bar{s}$ be such that $m \ge C_\beta \cdot \bar{s} \cdot \log n$. Then with probability at least $1 - 6/n - 6e^{-\beta}$ the solution $\hat{x}$ to the Lasso or the Dantzig selector with $\lambda = 10\sqrt{\log n}$ satisfies

$$\|\hat{x} - \bar{x}\|_1 \le \min_{1 \le s \le \bar{s}} C \cdot (1 + \alpha^2) \cdot \left(\|\bar{x} - \bar{x}_s\|_1 + s\sigma\sqrt{\frac{\log n}{m}} \right)$$

for some constant C.

Optimality of Compressive Sensing

Back to Gaussian sensing

m Gaussian measurements and ℓ_1 decoding:

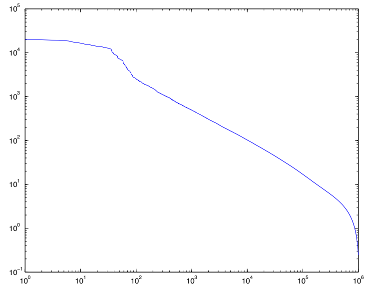
$$\|\hat{x} - x\|_2 \lesssim \frac{\|x - x_s\|_1}{\sqrt{s}}, \ s \approx \frac{m}{\log(n/m)}.$$

Question

Can we do better with other measurements or other algorithms?

Signals with power law

image



sorted wavelet coefficients

Power law decay: $|x|_{(1)} \geq |x|_{(2)} \geq \cdots \geq |x|_{(n)}$

$$|x|_{(k)} \leq \frac{C}{k^p}$$

Model

ℓ_p ball $\mathcal{B}_p := \{x : \|x\|_p \leq 1\}$.
Discuss case $p = 1$ but same discussion applies to $0 \leq p \leq 1$.

Recovery of ℓ_1 ball $\mathcal{B}$

Gaussian sensing

- Suppose unknown vector is in $\mathcal{B}$
- Take m Gaussian measurements

$$\|\hat{x} - x\|_2 \lesssim \sqrt{\frac{\log(n/m) + 1}{m}}.$$

Ideal sensing

Best we can hope from m linear measurements:

$$E_m(\mathcal{B}) = \inf_{D,E} \sup_{x \in \mathcal{B}} \|x - D(F(x))\|.$$

Gelfand widths

Theorem (Donoho)

$$d_m(\mathcal{B}) \leq E_m(\mathcal{B}) \leq C \cdot d_m(\mathcal{B}),$$

where $d_m(\mathcal{B})$ is the m-width of $\mathcal{B}$:

$$d_m(\mathcal{B}) := \inf_V \left\{ \sup_{x \in \mathcal{B}} \|P_V x\|_2 : codim(V) < m \right\}$$

Theorem (Kashin, Garnaev-Gluskin)

For ℓ_1 ball

$$C_1 \cdot \sqrt{\frac{\log(n/m) + 1}{m}} \leq d_m(\mathcal{B}) \leq C_2 \cdot \sqrt{\frac{\log(n/m) + 1}{m}}.$$

Compressive sensing achieves the limits of performance.

References for today's material

- Slides for this minicourse:
  `http://andrew.cmu.edu/user/jfp/UNencuentro`
- E. Candès & Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory,* vol 57, no. 11, pp. 7235–7254, Nov. 2011.
- E. Candès, J. Romberg, & T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory,* vol 52, no. 2, pp. 489–509, Feb. 2006.
- E. Candès & T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory,* vol 51, no. 12, pp. 4203–4215, Dec. 2005.
- D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory,* vol 52, no. 4, pp. 1289–1306, April 2006.
- Papers on compressive sensing: `http://dsp.rice.edu/cs`