Compressed Sensing and Applications

Sibi Raj B Pillai

Ack: Students and Colleagues @ IIT Bombay

Outline

- Signal Recovery: Formulation
- Compressed Sensing: a technique for sparse signal recovery
- Connections to Coding
- ℓ_p Minimization/Decoding
- Random Measurements
- Examples and Applications
- Some Algorithms

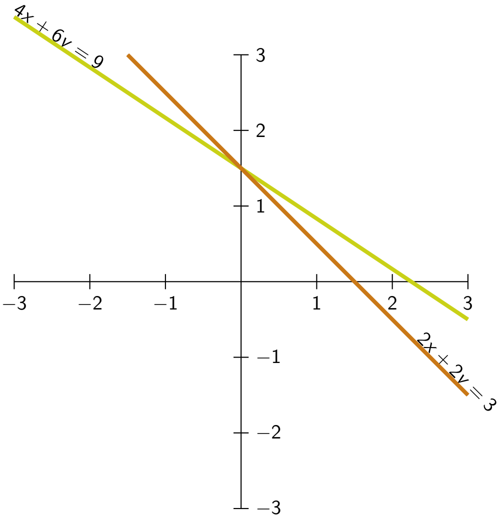
Solving Equations

$$4x + 6y = 9$$
$$2x + 2y = 3$$

$$\begin{bmatrix} 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\bar{x} = A^{-1} \; \bar{c}$$



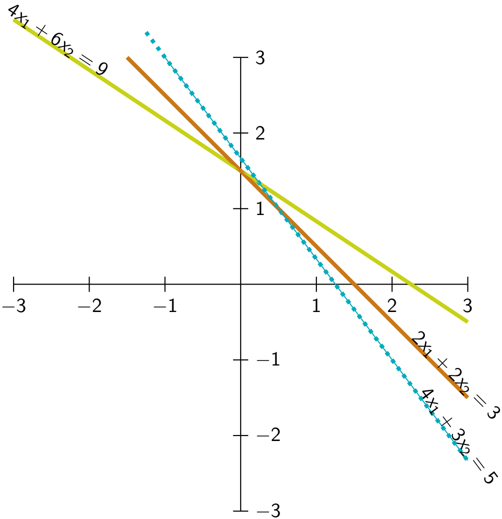
More Equations

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Estimated Solution

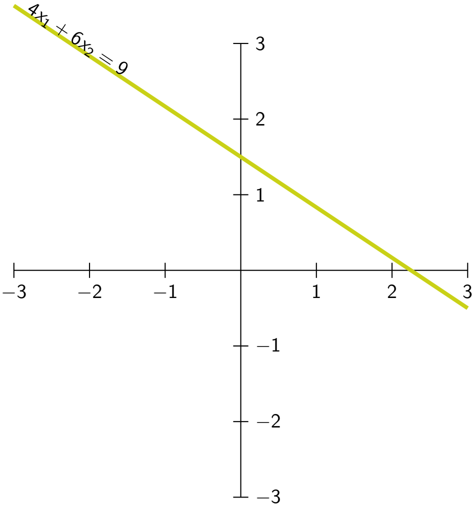
$$\hat{x} = \underset{\bar{x} \in \mathbb{R}^2}{\operatorname{argmin}} ||\bar{c} - A\bar{x}||^2$$

$$x_{LSE} = \left(A^T A\right)^{-1} A^T \bar{c}.$$



More Unknowns

$$4x_1 + 6x_2 = 9$$



- Many Solutions in general!
- Which one(s) do we need?

Linear Solvers

- We have M linear equations and N unknowns.

$$y_j = \sum_i a_{ji} x_i \quad , \; 0 \leq i \leq M - 1.$$

$$\underset{M\times 1}{\bar{y}} = \underset{M\times N}{A} \quad \underset{N\times 1}{\bar{x}}$$

- An **under-determined** set of equations, $M < N$.
- However, assume x to be **sparse** (a few *odd* values).
- More precisely, sparsity s represents the support of the signal $\bar{x}$.

System and Objectives

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1N} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{M1} & a_{M2} & \cdot & \cdot & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_N \end{bmatrix}$$

$$y_i = < \vec{a}_i, \bar{x} >$$

$$\bar{y} = \sum_j x_j \bar{a}_j$$

- Our aim is to find the sparse signal(s) $\hat{x}$ which satisfy the above.
- We need to design the **matrix** A, as well as a **recovery algorithm**.

Design Example 1

- $N = 8,\ M = 1,\ s = 1,\ x_i \in \{0, 1\}$:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = y_1$$

- Solution:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

If $y_1 = j$, declare $x_j = 1$ and all others are zero.

Design Example 2

- $N = 8,\ M = 2,\ s = 1,\ x_i \in \{0, 1, 2 \cdots\}$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

- Solution:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

Set $j = \frac{y_2}{y_1}$ and declare $x_j = y_1$; all others are zero.

Design Example 3

- $N = 8,\ M = 4,\ s = 2,\ x_i \in \{0, 1, \cdot \cdot 9\}$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2^0 & 3^0 & 4^0 & 5^0 & 6^0 & 7^0 & 8^0 \\ 1 & 2^1 & 3^1 & 4^1 & 5^1 & 6^1 & 7^1 & 8^1 \\ 1 & 2^2 & 3^2 & 4^2 & 5^2 & 6^2 & 7^2 & 8^2 \\ 1 & 2^3 & 3^3 & 4^3 & 5^3 & 6^3 & 7^3 & 8^3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

- Challenge:

Quick Recap

- A set of equations

$$\underset{M\times 1}{\bar{y}} = \underset{M\times N}{A} \underset{N\times 1}{\bar{x}}$$

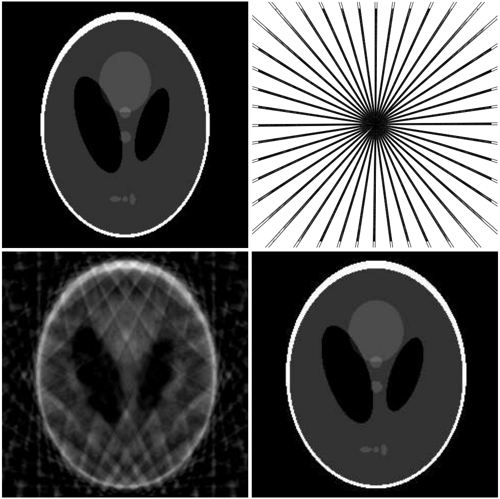
- We wish to get back $x \in \mathbb{R}^N$ from this M $m$easurements.
- This is under-determined, little hope of getting x back in general.
- However, we can design the system to recover all sparse inputs x.
- In particular, if $\text{supp}(x) \leq s$, it has to be recovered.
- **Goal:** Design the matrix A and a recovery strategy.

Compressed Sensing: Overview

Compressed Sensing

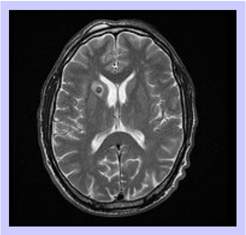
OLEPFLnd, Switzerland, Oct 2006

Magnetic Resonance Imaging

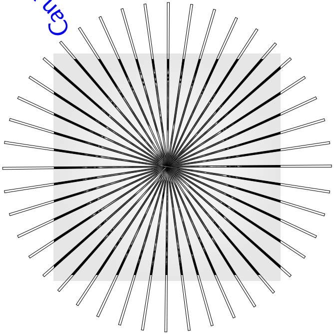


Candes Romberg Tao' 2006

CS Magic?



Can we get back the original

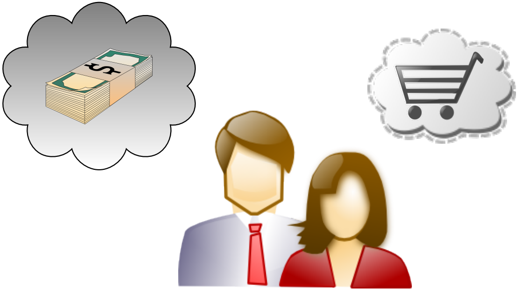


- Surely you are joking Mr. Xxxxman.
- Frequency domain seems to do wonders here.

Family Affairs

Is there a way to weed out the junk at the source itself, while collecting most of the essential stuff.?

ala Candes, Romberg, Tao



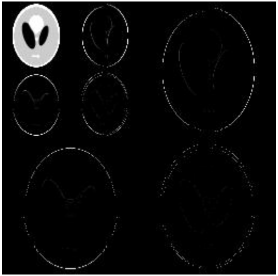
Moral: There are **adversarial** signals for which *CompressedSensing* may fail

Fourier Incoherence

- How can we get back our MRI from 10% of the measurements.

;

$\xrightarrow[\text{Transform}]{\text{Haar}}$

;

- These images have inherent sparsity, or they are **compressible**.
- Time-sparseness and frequency acquisition have a close connection.

Coding Connections

Fourier Transform

- Discrete Fourier Transform

$$X[k] = \sum_{n} x_n \alpha_n^k$$

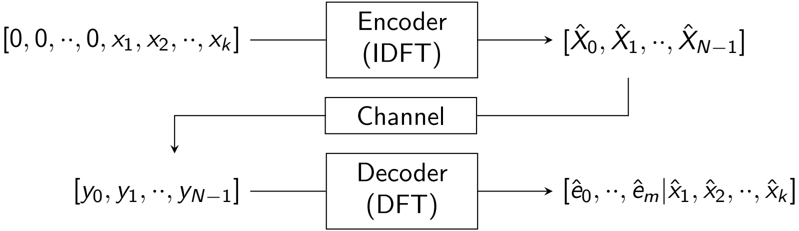
$$\alpha = e^{-j\frac{2\pi}{N}} \;,\; \alpha_n = \alpha^n$$

$$\begin{array}{cc}
\begin{array}{c} s = 1 \\ \\ s = 2 \\ \\ \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & \alpha_1^1 & \alpha_2^1 & \alpha_3^1 & \alpha_4^1 & \alpha_5^1 & \alpha_6^1 & \alpha_7^1 \\
1 & \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 & \alpha_7^2 \\
1 & \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \alpha_4^3 & \alpha_5^3 & \alpha_6^3 & \alpha_7^3 \\
1 & \alpha_1^4 & \alpha_2^4 & \alpha_3^4 & \alpha_4^4 & \alpha_5^4 & \alpha_6^4 & \alpha_7^4 \\
1 & \alpha_1^5 & \alpha_2^5 & \alpha_3^5 & \alpha_4^5 & \alpha_5^5 & \alpha_6^5 & \alpha_7^5 \\
1 & \alpha_1^6 & \alpha_2^6 & \alpha_3^6 & \alpha_4^6 & \alpha_5^6 & \alpha_6^6 & \alpha_7^6 \\
1 & \alpha_1^7 & \alpha_2^7 & \alpha_3^7 & \alpha_4^7 & \alpha_5^7 & \alpha_6^7 & \alpha_7^7
\end{bmatrix}
\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix}
=
\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix}
\end{array}$$

Coding Connections

- Communication Channel

- Channel Coding



- If there are no errors, $\hat{e}_i = 0$ and $\hat{x}_i = x_i$, $\forall i$.

Error Correction

- In the presence of errors, say $N = 8$ and at most two errors,

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \alpha_1^1 & \alpha_2^1 & \alpha_3^1 & \alpha_4^1 & \alpha_5^1 & \alpha_6^1 & \alpha_7^1 \\ 1 & \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 & \alpha_7^2 \\ 1 & \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \alpha_4^3 & \alpha_5^3 & \alpha_6^3 & \alpha_7^3 \end{bmatrix} \begin{bmatrix} e_0 \\ e_1 \\ \cdot \\ \cdot \\ e_N \end{bmatrix} = \begin{bmatrix} \hat{e}_0 \\ \hat{e}_1 \\ \cdot \\ \cdot \\ \hat{e}_m \end{bmatrix}$$

- We can solve this to find the error locations and values.
- Once the error vector is known, simply subtract it from $\bar{y}$.
- This is the principle of Reed-Solomon codes, used in CDs/Memorys.

How it works

- Consider the first $m = 2s$ rows of the Fourier matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \alpha_1^1 & \alpha_2^1 & \alpha_3^1 & \alpha_4^1 & \alpha_5^1 & \alpha_6^1 & \alpha_7^1 \\ 1 & \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 & \alpha_7^2 \\ 1 & \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \alpha_4^3 & \alpha_5^3 & \alpha_6^3 & \alpha_7^3 \end{bmatrix} \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} \hat{e}_0 \\ \hat{e}_1 \\ \vdots \\ \hat{e}_m \end{bmatrix}$$

- Pick **any** $2s$ columns from this restricted matrix.

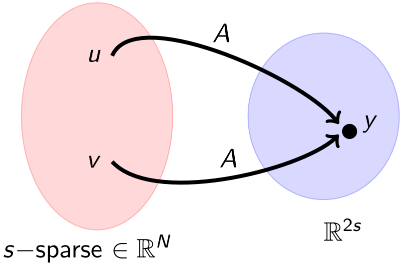
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ \alpha_k^1 & \alpha_l^1 & \alpha_m^1 & \alpha_n^1 \\ \alpha_k^2 & \alpha_l^2 & \alpha_m^2 & \alpha_n^2 \\ \alpha_k^3 & \alpha_l^3 & \alpha_m^3 & \alpha_n^3 \end{bmatrix}$$

- The determinant is non-zero if $\alpha_i \neq \alpha_j \Rightarrow$ cols. linearly independent.

Unique Solution

- Suppose $\exists u, v \in \mathbb{R}^N$ s.t. $\text{supp}(u) \leq s$, $\text{supp}(v) \leq s$ and,

$$Au = y \text{ and } Av = y$$



"A cannot distinguish between u and v"

This will never happen!

- Indeed, $Au = Av \Rightarrow A(u - v) = 0$ and we know $u - v$ is $2s-$sparse.

- But any $2s$ columns of A are linearly independent $\Rightarrow A(u - v) \neq 0$.

Key Property

Any $2s$ columns of the measurement matrix A has to be linearly independent for detecting all $s-$sparse signals from $y = Ax$.

- We will say $A \in \mathcal{P}$, if A has the above property (future use).
- *Naive* Decoding: Brute force search . (i.e. ℓ_0- **minimization**)
- There is only one sparse x such that $Ax = y$, if $A \in \mathcal{P}$.

Fourier Inverse

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdot & \cdot & 1 \\ 1 & \alpha_1^1 & \alpha_2^1 & \alpha_3^1 & \alpha_4^1 & \alpha_5^1 & \alpha_6^1 & \alpha_7^1 & \cdot & \cdot & \alpha_{N'}^1 \\ 1 & \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 & \alpha_7^2 & \cdot & \cdot & \alpha_{N'}^2 \\ 1 & \alpha_1^3 & \alpha_2^3 & \alpha_3^3 & \alpha_4^3 & \alpha_5^3 & \alpha_6^3 & \alpha_7^3 & \cdot & \cdot & \alpha_{N'}^3 \\ 1 & \alpha_1^4 & \alpha_2^4 & \alpha_3^4 & \alpha_4^4 & \alpha_5^4 & \alpha_6^4 & \alpha_7^4 & \cdot & \cdot & \alpha_{N'}^4 \\ 1 & \alpha_1^5 & \alpha_2^5 & \alpha_3^5 & \alpha_4^5 & \alpha_5^5 & \alpha_6^5 & \alpha_7^5 & \cdot & \cdot & \alpha_{N'}^5 \\ 1 & \alpha_1^6 & \alpha_2^6 & \alpha_3^6 & \alpha_4^6 & \alpha_5^6 & \alpha_6^6 & \alpha_7^6 & \cdot & \cdot & \alpha_{N'}^6 \\ 1 & \alpha_1^7 & \alpha_2^7 & \alpha_3^7 & \alpha_4^7 & \alpha_5^7 & \alpha_6^7 & \alpha_7^7 & \cdot & \cdot & \alpha_{N'}^7 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \alpha_1^{N'} & \alpha_2^{N'} & \alpha_3^{N'} & \alpha_4^{N'} & \alpha_5^{N'} & \alpha_6^{N'} & \alpha_7^{N'} & \cdot & \cdot & \alpha_{N'}^{N'} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ \cdot \\ \cdot \\ x_{N'} \end{bmatrix} = \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ \cdot \\ \cdot \\ X_{N'} \end{bmatrix}$$

- For a given N, any $2s$ consecutive rows can be used as A.
- If N is prime, arbitrary $2s$ rows suffice Tao'2004.

ℓ_p minimization

Fourier Incoherence *Tao'04*

Consider a discrete-time signal $x \in \mathbb{C}^N$ and a set of frequencies Ω. Is it possible to reconstruct x from the partial knowledge of its Fourier coefficients on the set Ω ?.

Let N be a **prime** integer.

Suppose x is supported on $T \subset \mathbb{Z}_N$

Consider a set of frequencies $\Omega \subset \mathbb{Z}_N$, and $X_k, \forall k \in \Omega$.

$$\boxed{2|T| < |\Omega| \Rightarrow x \text{ can be reconstructed from } X_k|_\Omega}$$

ℓ_0 Minimization

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1N} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{M1} & a_{M2} & \cdot & \cdot & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_N \end{bmatrix}$$

- Assume the solution $\hat{x}$ is unique, and $supp(\hat{x}) \leq s$.
- For every $\binom{N}{s}$ columns, say $\underset{M \times s}{B}$, check if $\exists\ u : Bu = y$.
- Can be very hard, if A is not very structured.
- **Drawback:** Severe limitation on the choice of matrix A.

3 Drawbacks

- **Issue 1**
  $2s$ Fourier measurements suffice if N is prime or successive coefficients are available.

- **Issue 2**
  Fourier measurements are good only for a selected class of signals.

- **Issue 3**
  Even with in a '*working class*', it seems prohibitively complex to isolate the actual solution (recovery).

Extensions: *CandesTao'05 '06*

- Any $M = c\ 2s \log \frac{N}{s}$ measurements suffice to recover $s-$sparse signals *most of the times.*

- We can replace the elements of A by random entries from an appropriate distribution, example: Gaussian, ± 1 so on.

- Furthermore, and most importantly there exists **linear programs** which effectively recover sparse signals from the M measurements.

Compressed Sensing: *CandesRombergTao'06, Donoho'06*

Consider a discrete-time signal $x \in \mathbb{C}^N$ and a **randomly** chosen set of frequencies Ω. Is it possible to reconstruct x from the partial knowledge of its Fourier coefficients X_k, on the set Ω ?.

Suppose x is the superposition of $|T|$ spikes

$$x_n = \sum_{m \in T} x_m \delta(n-m)$$

and

$$2|T| \leq C_M.(\log N)^{-1}.|\Omega|$$

then with probability $> 1 - O(N^{-M})$ the answer is YES!

$$x = \underset{u}{\text{argmin}} \quad ||u||_{\ell_1}\,, \qquad \text{s.t.} \quad U_k = X_k \text{ for all } k \in \Omega$$

Why ℓ_1 Minimization?

A first attempt: *minimum energy* solution or ℓ_2 minimum

$$\min \; x^\dagger x \;\; \text{s.t.} \;\; Ax = y$$

Using Lagrange cost function,

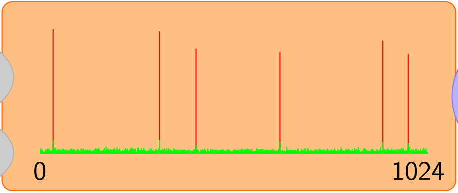
$$J(x) = ||x||_{\ell_2}^2 + \lambda^\dagger (y - Ax)$$

Pre-mulitplying with Moore-Penrose pseudo-inverse is the minimizer

$$x = A^\dagger (AA^\dagger)^{-1} y$$

ℓ_0-norm for min. supp

NP-hard to solve!

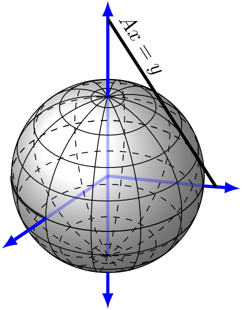


ℓ_1 the convex answer!

Why l_2 fails?

$$\min_{x \in \mathbb{R}^N} \; ||x||_{\ell_2}$$

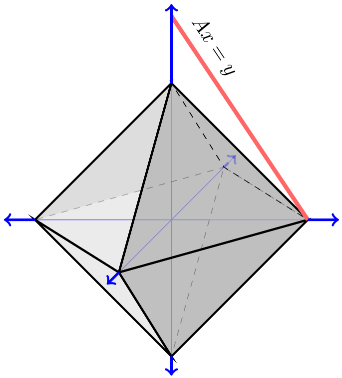
$$\text{s. t. } Ax = y$$



- $||x||_{\ell_2} = r$: sphere of radius r.
- Dual View:
  Find the *maximal* sphere
  touching the half-space.

L1 succeeds

$$\min ||x||_{\ell_1}$$
$$\text{s .t. } Ax = y$$



- $||x||_{\ell_1} = c$, is a polyhedron.
- Find the biggest such polyhedron touching $Ax = y$.
- Key: We should avoid sparse signals competing with each other.

Random Measurement Matrices

Sparse Competition

- If we choose $A \in \mathcal{P}$, then every s-sparse vector at the input gives rise to a unique set of measurements.
- In this case, ℓ_0 minimization will *theoretically* succeed, perhaps with unmanageable computational complexity.
- In order that ℓ_1 minimization succeeds, we need more properties than $A \in \mathcal{P}$.
- Nevertheless, let us first search for matrices which are in $\mathcal{P}$ (i.e, any $2s$ columns are linearly independent)

Common sense is Random

- How will you generate a tall $M \times k$ full rank matrix B.
- Assume that under the choice of iid selection from some distribution,
$$P(\text{rank}(B) = k) \geq 1 - \exp(-cM).$$
- If we generate our measurement matrix A using this distribution,
$$\begin{aligned} P(A \notin \mathcal{P}) &\leq \binom{N}{2s} \exp(-cM) \\ &\leq \left(\frac{Ne}{2s}\right)^{2s} \exp(-cM) \end{aligned}$$
- $$M = O(s \log \frac{N}{s})$$
may guarantee that A is in $\mathcal{P}$.

Random Full Rank *Rudelson-Vershynin'08-10*

- $$P(\text{rank}(B) = k) \geq P(\sigma_{min}(B) \geq \epsilon), \epsilon > 0.$$

  where $\sigma_{min}(B)$ is the smallest eigen value of $\sqrt{B^{\dagger}B}$.

- Non-asymptotic analysis of tall random matrices,

  $$\sigma_{min}(B) \geq \sqrt{M} - C\sqrt{k} - d$$

  with probability exceeding $1 - 2\exp(-cd^2)$, if the elements are taken independently from any sub-Gaussian distribution.

- **Restricted Isometry**

  $$(1 - \delta_s)||x||_{\ell_2} \leq ||Ax||_{\ell_2} \leq (1 + \delta_s)||x||_{\ell_2}$$

  under suitable scaling of the matrix A for all $s-$sparse signals.

Moving from Fourier

- Fourier measurements tackle time-sparsity.
  - however, images are sparse in wavelet basis.
- Natural extensions can cover all sorts of sparsity.

$$\begin{aligned} \bar{f} &= \sum_{i \in T} \alpha_i \, \bar{\psi}_i, \quad \text{where } |T| << N \\ &= \mathbf{\Psi} x, \ \text{where } x \text{ is sparse }. \end{aligned}$$

- The measurements now become

$$y = \mathbf{\Phi} \mathbf{\Psi} x$$

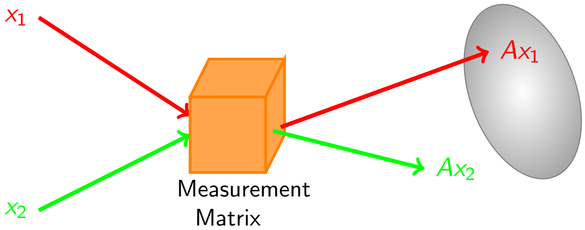
- Random ϕ provides a 'universality' to the encoding process, thus we can **rip** any sparse signal.

Restricted Isometry Property (RIP) *CandesTao'05*

Exact reconstruction by ℓ_1 minimization is guaranteed by the measurement matrix A satisfying the following property.

RIP of order $2s$

"Any $2s$ columns of A are nearly orthogonal"



"$2s$ RIP $\Rightarrow$ Pairwise distances are nearly *preserved* for s-sparse signals"

RIP and ℓ_1

- Assume v was the true signal and $y = Av$.
- ℓ_1 minimization will succeed if,

$$||v||_{\ell_1} \leq ||u||_{\ell_1} + \lambda^T(y - Au) \text{ , } \forall u \in \mathbb{R}^N$$

for some $\lambda \in \mathbb{R}^M$ (Lagrange multiplier)

- Denote
  - J : the set of column indices of A.
  - T : the support set of v (note that $T \subset J$).

$$Au = \sum_{j \in J} u_j \bar{a}_j \qquad ; \qquad Av = \sum_{j \in T} v_j \bar{a}_j$$

where $\bar{a}_j$ denotes the j^{th} column of A.

ℓ_1 Solution

$$
\begin{aligned}
&||u||_{\ell_1} + \lambda^T(y - Au) \\
&\qquad = ||u||_{\ell_1} + \lambda^T(\sum_{j\in T} v_j \bar{a}_j) - \lambda^T(\sum_{j\in J} u_j \bar{a}_j) \\
&\qquad = ||u||_{\ell_1} + \sum_{j\in T} v_j < \lambda, \bar{a}_j > - \sum_{j\in J} u_j < \lambda, \bar{a}_j >
\end{aligned}
$$

If $< \lambda, \bar{a}_j >= \text{sign}(v_j)$, $\forall j \in T$, then

$$
\begin{aligned}
&||u||_{\ell_1} + \lambda^T(y - Au) \\
&\qquad = ||u||_{\ell_1} + ||v||_{\ell_1} - \sum_{j\in T} u_j \text{sign}(v_j) - \sum_{j\notin T} u_j < \lambda, \bar{a}_j > \\
&\qquad = ||v||_{\ell_1} + \sum_{j\in T} u_j(\text{sign}(u_j) - \text{sign}(v_j)) \\
&\qquad\qquad + \sum_{j\notin T} u_j(\text{sign}(u_j) - < \lambda, \bar{a}_j >)
\end{aligned}
$$

Existence of $\lambda \in \mathbb{R}^M$

- If there exists a $\lambda \in \mathbb{R}^M$ such that,
  - $< \lambda, \bar{a}_j > \quad = \text{sign}(v_j), \forall j \in T$
  - $| < \lambda, \bar{a}_j > | < 1, \forall j \notin T$

  then ℓ_1 minimization will indeed find the vector v.

- Loosely, this existence is guaranteed for any $s-$ sparse vector, by having a RIP of order $2s$ on the matrix A.

RIP guarantees exact recovery of $S-$ sparse signals and recovers the $S-$ largest entry of *compressible* vectors *[CandesWakin'08]*.

There is no probability, it is *deterministic*

This is more powerful and general than Fourier measurements.

How to find measurement matrices obeying RIP

Measurements obeying RIP

The following measurement matrices,

(i) Gaussian $\mathcal{N}(0, \frac{1}{m})$ iid entries

(ii) Bernoulli $(\frac{1}{2})$ iid entries

(iii) Columns uniformly at random on unit sphere

satisfy RIP with overwhelming probability, if

$$M \geq C.S \log \frac{N}{S}$$

Examples and Applications

Reconstruction

By a Linear Program

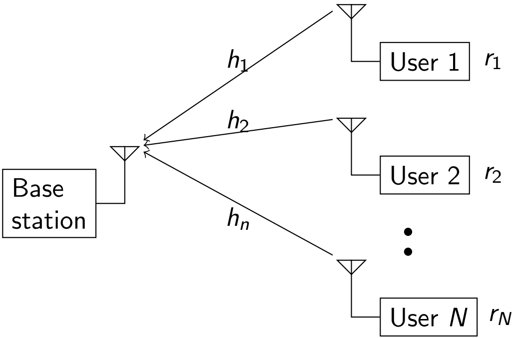
$$\min ||x||_{\ell_1} \text{ such that } Ax = y$$

Off the shelf efficient solvers are available

In presence of error, LASSO algortihm

$$\min ||x||_{\ell_1} \text{ such that } ||Ax - y||_{\ell_2} \leq \epsilon$$

CS as CDMA channel



$$r_j = h_j x + w_j$$

$$\bar{y} = A\bar{h} + \bar{z}$$

Each user sends a **short** normalized pseudo-noise sequence in a CDMA scheme.

Simulation Study I

$\sigma^2 = 0.01,\ SNR \approx 0dB$ (per user above noise)

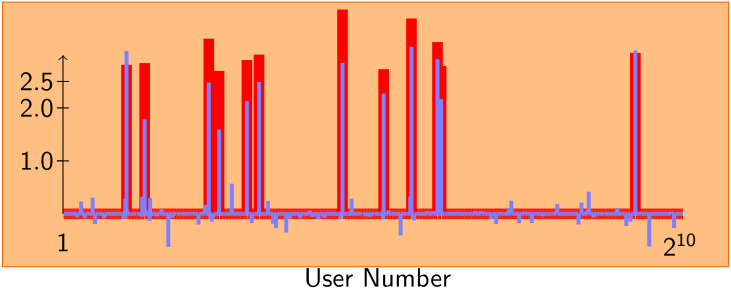


Figure: Original Signal and Reconstruction from 100 measurements

Simulation Study II

$\sigma^2 = 10^{-4}, \; SNR \approx 20dB$ (per user above noise)

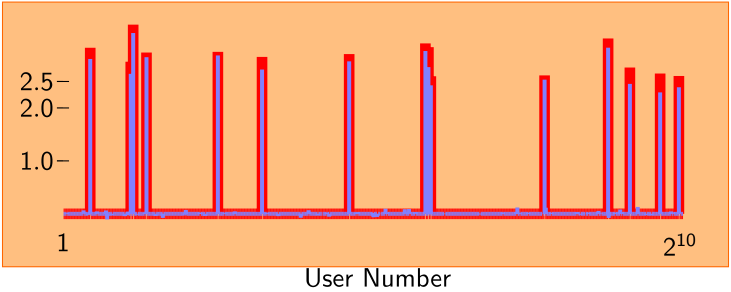


Figure: Original Signal and Reconstruction from 100 measurements

CS vs Matched Filtering

In MF, each user matches y with its signature

$$\sigma^2 = 10^{-4},\ SNR \approx 20dB \text{ (per user above noise)}$$

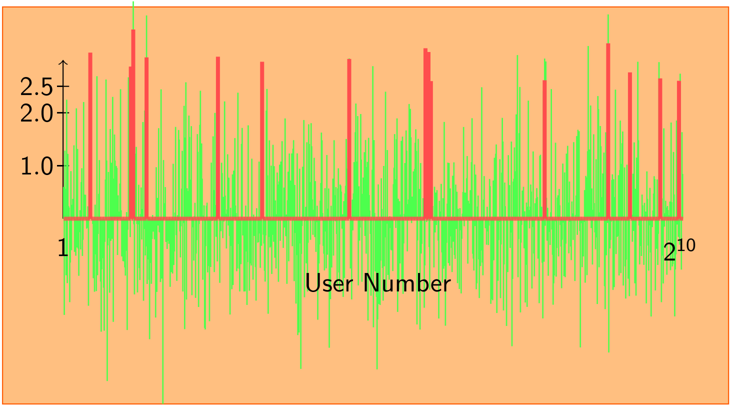
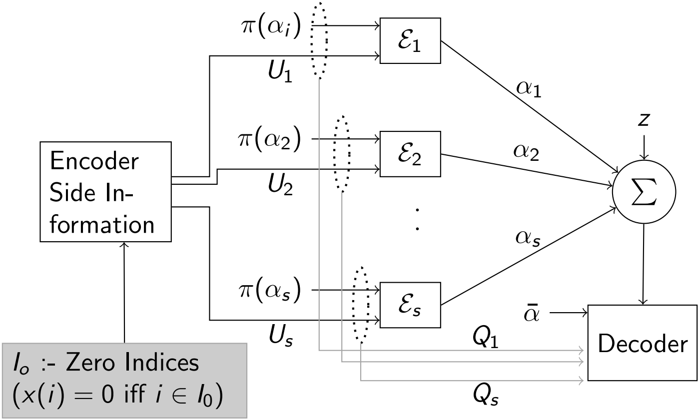


Figure: Reconstructions from 100 measurements for MF and CS

Converse



Consider CS as a CDMA scheme with the non-zero values as transmitters

Bernoulli Measurements

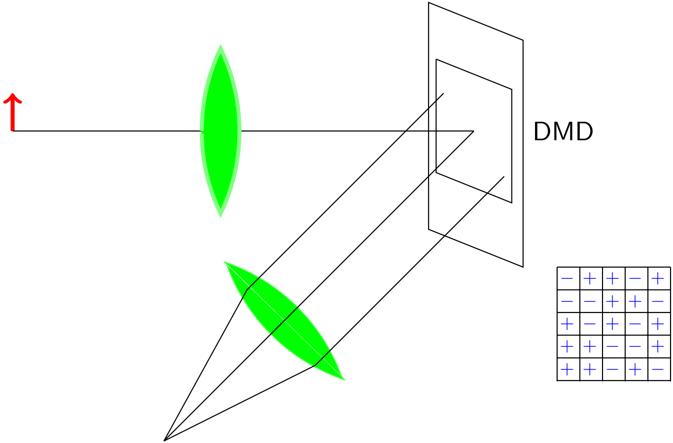
Theorem

If the measurement matrix is chosen by Bernoulli($\frac{1}{2}$) on $\{\pm 1\}$,

$$m \geq \frac{2s \log_2 \frac{n}{s}}{\log_2 \pi e s/2}$$

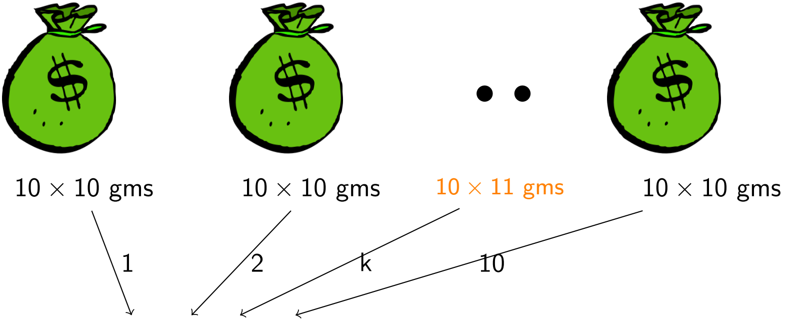
for P_{error} to go vanishingly small.

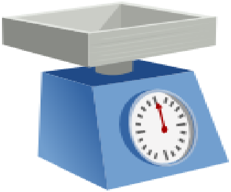
Single Pixel Camera



- Can change imaging concepts in the non-visible spectrum.

A Small *Puzzle*





$$\text{Bag Number} = W - 550$$

$$W = \sum a_i x_i, \text{ where } a_i = i, x_i \in \{10, 11\}$$

1) One Sparsity needs *one* measurement
2) Measurements should be **incoherent** to sparsity

Compressed Sensing vs Group Testing

N soldiers return after a war. It is compulsory to have their blood tested before reunion with family. Blood Tests are very costly, so is the time spent waiting to meet the families. We need a strategy (a sequence of tests) with objectives,

- find the persons having a particular disease A.
- the minimum number of tests (and/or time).

- Huge savings by **grouping** and **testing**
- Compress sensing is analogous to group-testing
  - a number of tests in parallel
  - all the results available simultaneously.
- An exercise of **picking** the **odd** man/men (signals) out.

Recovery Algorithms

Recovery Algorithms

- Greedy Approaches
  - Orthogonal Matching Pursuit (*Tropp*)
  - CoSamp and other variants
- LP Based methods
  - Basis Pursuit (*Donoho*)
  - LASSO (QP) and gradient based methods
- Iterative algorithms
  - Iterative Hard Thresholding

Orthogonal Matching Pursuit (OMP)

- Greedy search to find x from $Ax = y$.
- Let $S \subset J$ be a set of column indices.
- A_S denotes matrix formed by those columns in S.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1l} & & a_{1k} & a_{1N} \\ a_{21} & a_{22} & a_{23} & & a_{2l} & & a_{2k} & a_{2N} \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\ a_{M1} & a_{M2} & a_{M3} & & a_{Ml} & & a_{Mk} & a_{MN} \end{bmatrix}$$

$$S[i+1] = \quad S[i] \quad \bigcup \quad j^*[i]$$

$$j^*[i] = \underset{j \notin S[i]}{\operatorname{argmax}} |< \bar{a}_j, r[i] >|$$

- $r[i]$ is the residual error in the best fit of y using $A_{S[i]}$.

OMP Performance

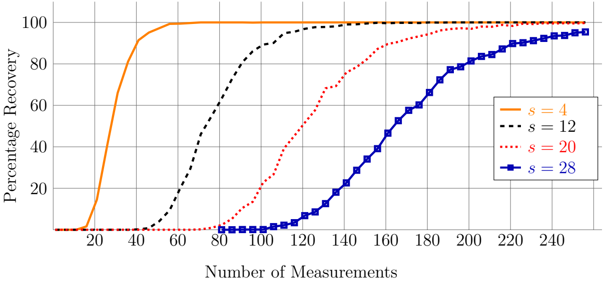


Figure: Percentage of 1000 input signals correctly recovered as a function of the number M of measurements for different sparsity levels s in dimension $N = 256$.

Pattern Recognition

- Consider an image data-base with N images,

$$\{I_1, I_2, \cdot\cdot, I_N\}, I_i \in \mathbb{R}^K$$

- Given a query image q, find the *closest allies*.

$$\begin{bmatrix} I_{11} & I_{21} & I_{31} & I_{l1} & I_{k1} & I_{N1} \\ I_{12} & I_{22} & I_{32} & I_{l2} & I_{k2} & I_{N2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ I_{1K} & I_{2K} & I_{3K} & I_{lK} & I_{kK} & I_{NK} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_K \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ q_K \end{bmatrix}$$

$$S[i+1] = \quad S[i] \quad \bigcup \quad j^*[i]$$

- Each column here is a (nearly) sparse vector in some basis.

Searching Random Projections

- Generate a lower dimensional projection by premultiplying with a random A.

$$\underset{M\times K}{A}\ \underset{K\times N}{D_I}\ \underset{K\times 1}{x} = \underset{M\times K}{A}\ \underset{K\times 1}{q}$$

  where D_I is the database matrix.

- Now run OMP to resolve x from the measurements $y = Aq$.

- The system complexity is greatly reduced.

Conclusion

The usual **acquire**(sense) and then compress paradigm is not cost effective.

Compressed Sensing alleviates this problem by doing **compression** at **sensing** itself

We studied some simple connections between CS and coding theory

Random acquisition schemes give a universality to the acquisition schemes

Applications and Extensions are emerging at rapid pace, a good area for research.

Source

Most related material (including software, applications, documents, opportunities) available at the dedicated website http://dsp.rice.edu/cs

IEEE Signal Processing Magazine, March 2008 was on the theme 'Compressed Sensing', and covered several applications and algorithms.

Terrence Tao's Blog has lot of useful information on compressed sensing and complexity.

References I

(CandesTao06) E. Candes, J. Romberg, T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. on Info Theory*, pp. 489 - 509, Feb 2006.

(Donoho06) David L. Donoho, "Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289 - 1306, April 2006.

(CandesTao05), E. Candes and T. Tao, "Decoding by LP", IEEE Trans on Info Theory, Dec 2005.

(Tao04), " An uncertainty principle for cyclic groups..", arxiv: math.CA0308286

(Baranuik07)R. Baranuik, "Compressive sensing", Lecture Notes in *IEEE Signal Processing Magazine*, Volume 24, July 2007.

(Baranuik08)R. Baranuik, M. Davenport, R. DeVore, M.B. Wakin, "A simple proof of the restricted isometry property for random matrices", *Constructive Approximation*, 28(3), pp. 253-263, December 2008.

(Tropp07)Joel A. Tropp, Anna C. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit", *IEEE Trans. on Information Theory*, 53(12), pp. 4655-4666, December 2007.

Reference II

(CandesWakin08) E. Candes and J.Wakin, “Introduction to CS”, IEEE Signal Processing, March 2008

(NeedelTrop08)D. Needell, J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples”. (Preprint 2008, at http://dsp.rice.edu/cs)

(Nowak07) M. Figuerado, R. Nowak, S. Wright,“GPSR: Application to CS and other inverse problems”, IEEE Journal of Selected Areas in Signal Processing, 2007.

(Wainwright07) M Wainwright, “Information theoretic bounds on Sparsity recovery”, ISIT 2007.

(Tune09) P. Tune, S. R. Bhaskaran, S.V.Hanly, “Number of measurements in Sparse Signal Recovery”, ISIT 2009.

(Vershynin10) R. Vershynin, “Nonasymptotic analysis of random matrices”, Lecturenotes, available at http://arxiv.org.