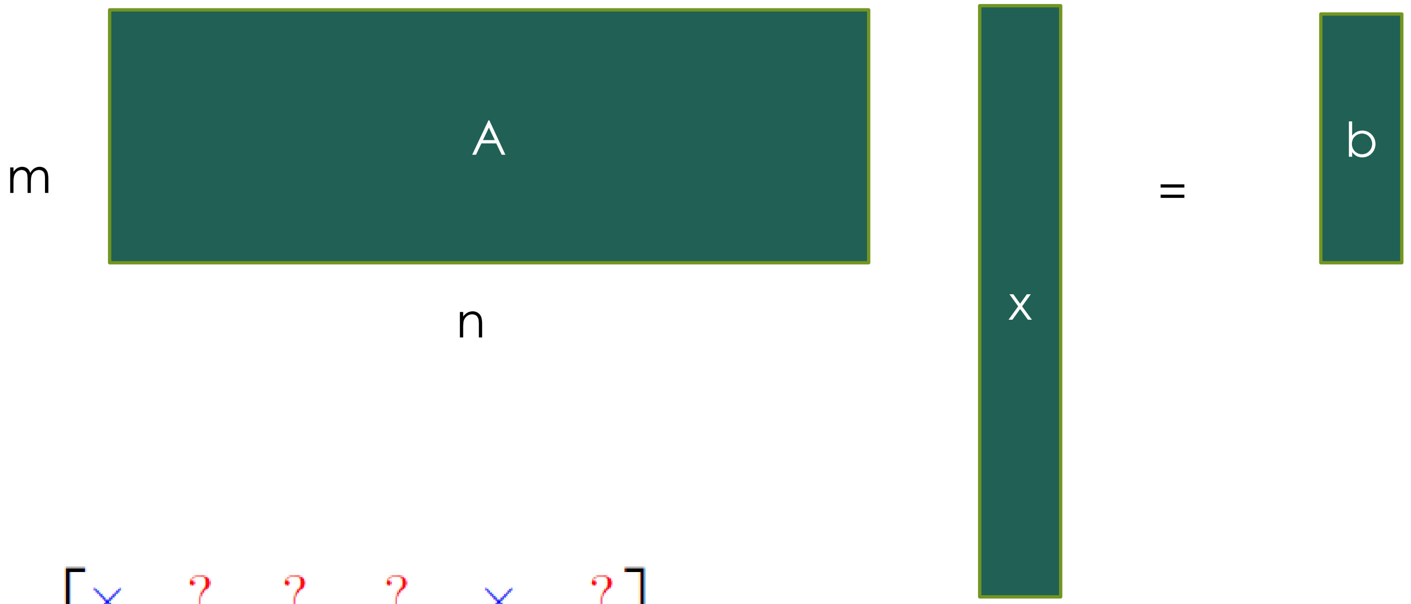
DS 503: Advanced Data Analytics

Lecture 6: Compressed Sensing

Dr. Gagan Raj Gupta

Based on talks of Emmanuel Candès

Under-determined systems of Linear Equations



$x \in R^n$
m<<n linear equations about x
$b \in R^m$, Ax =b
We would like to recover x

Arises in many fields of engineering
- Sparse regression in genetics
- Siesmology
- Remote Sensing
- MRI
- Cameras
- RF
- Collaborative Filtering

$$\begin{bmatrix} \times & ? & ? & ? & \times & ? \\ ? & ? & \times & \times & ? & ? \\ \times & ? & ? & \times & ? & ? \\ ? & ? & \times & ? & ? & \times \\ \times & ? & ? & ? & ? & ? \\ ? & ? & \times & \times & ? & ? \end{bmatrix}$$

If unknown is assumed to be
sparse (genomics example)
low-rank (Matrix Completion)
then one can often find solutions to these problems by convex optimization

A contemporary paradox

Raw: 15MB

JPEG: 150KB

- Massive data acquisition
- Most of the data is redundant and can be thrown away
- Seems enormously wasteful
- *"One can regard the possibility of digital compression as a failure of sensor design. If it is possible to compress measured data, one might argue that too many measurements were taken." – David Brady (Duke, ECE)*

What is compressive sensing?

- Possibility of compressed data acquisition protocols which directly acquire just the important information
- Incoherent/random measurements → compressed description
- Simultaneous signal acquisition and compression!
- All we need is to decompress...

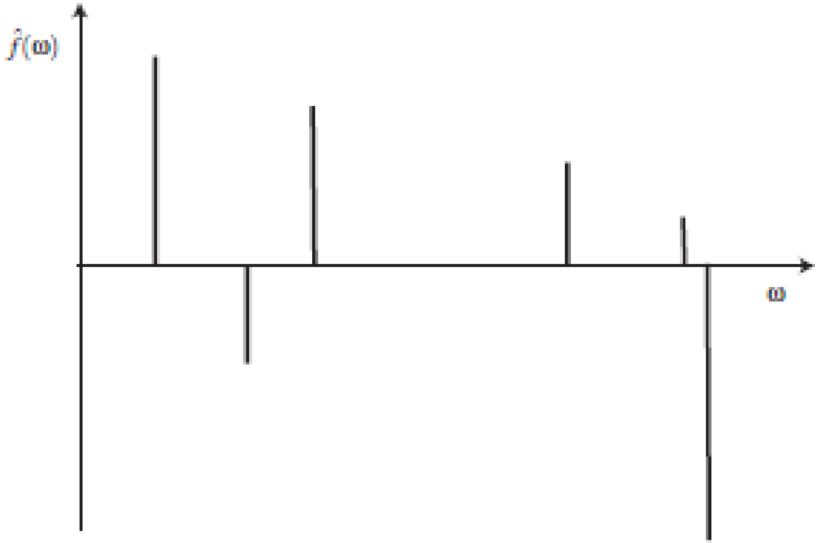
Three surprises:

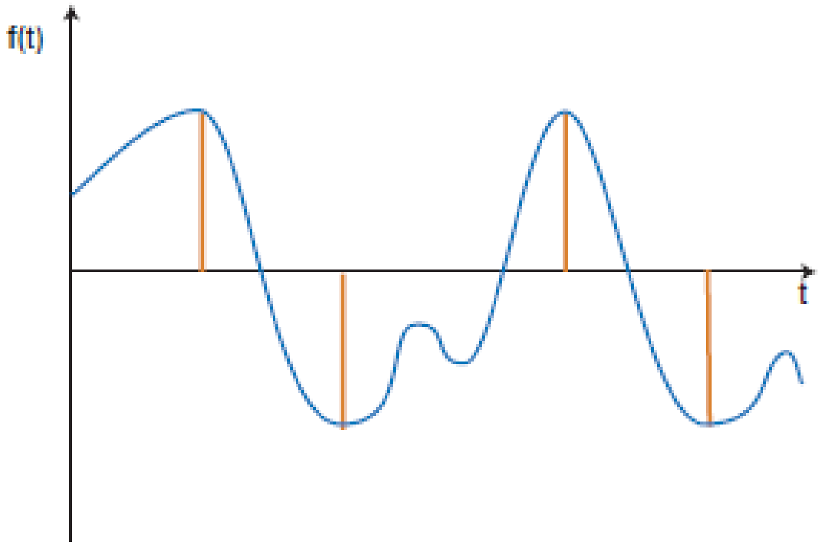
- Sensing is ultra efficient and nonadaptive
- Recovery is possible by tractable optimization
- Sensing/recovery is robust to noise (and other imperfections)

Analog to Digital Conversion

If `information bandwidth' less than total bandwidth, then should be able to

- sample below Nyquist without information loss
- recover missing samples by convex optimization





Changing the way design sensors (Image, Voice etc.)

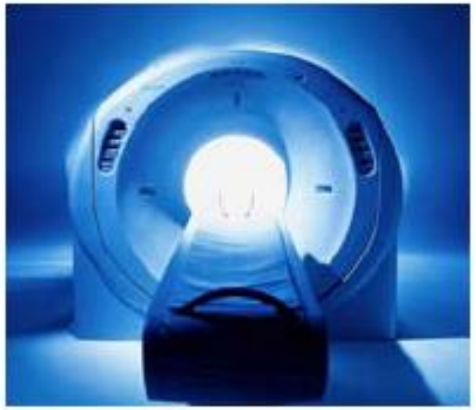
I → Sample →(N) M<<N Compress →(M) Transmit /Store →(M) Decompr ess →(N) Î

- Acquire/Sample (A-to-D converter, digital camera)
- Compress (signal dependent, nonlinear)
- Fundamental Question: Can we directly acquire just the useful part of the signal?
  - Huge Practical Implications: MRI (reduce radiation), Remote Sensing, Networks

Optical Systems

- Direct sampling: analog/digital photography, mid 19th century
- Indirect sampling: acquisition in a transformed domain, second half of 20th century; e.g. CT, MRI
- Compressive sampling: acquisition in an incoherent domain
  - Design incoherent analog sensors rather than usual pixels
  - Pay-off: need far fewer sensors

The first photograph?



CT scanner

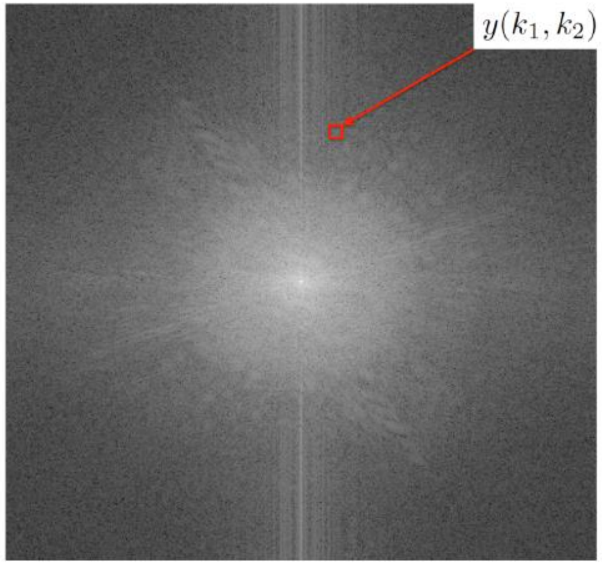
MRI: Magnetic Resonance Imaging

MR scanner

MR image

Image from K. Pauly, G. Gold, RAD220

MRI Acquisition Process



$$y(k_1, k_2) = \iint f(x_1, x_2) e^{-i2\pi(k_1 x_1 + k_2 x_2)} \, dx_1 dx_2$$

Fourier transform for frequency k1 and k2

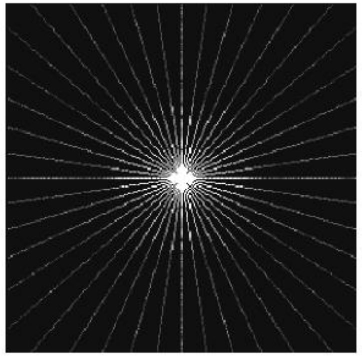
$$f(x_1, x_2) \approx \sum \sum y(k_1, k_2) e^{i2\pi(k_1 x_1 + k_2 x_2)}$$

- (Powerful) magnetic field aligns nuclear magnetization of (usually) hydrogen atoms in water in the body
- RF fields systematically alter the alignment of this magnetization -> hydrogen nuclei produce a rotating magnetic field detectable by the scanner
- Make excitation strength space dependent
- Goal is to recover proton density

Discovery (Logan-Shepp Test Image)!

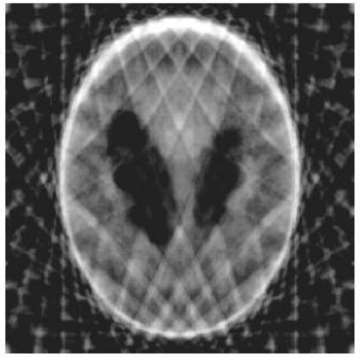
A surprising experiment

Fourier transform



highly subsampled

classical reconstruction



compressed sensing reconstruction

Algorithm:

$$\min \sum_{x_1, x_2} ||\nabla f(x_1, x_2)|| \text{ subj. to data constraints}$$

C., Romberg and Tao ('04)

https://en.wikipedia.org/wiki/Shepp%E2%80%93Logan_phantom

Collaborative Filtering Example

3	2	4	2	1
4	2	6	4	2
3	1	5	4	2
3	1	4	3	1
1	0	3	3	2

Ground truth

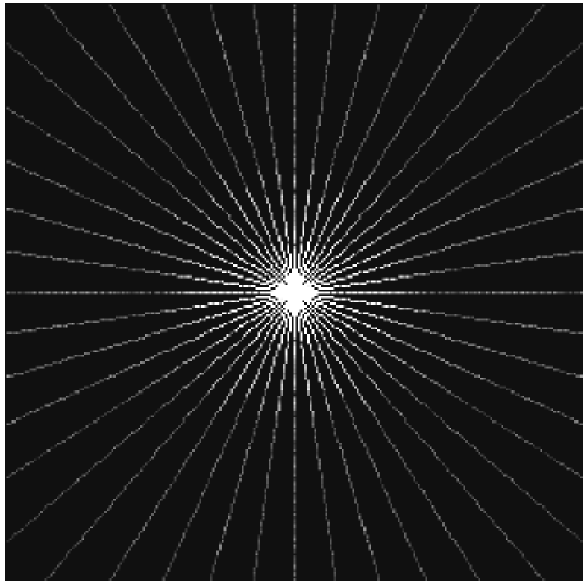
	2		2	
		6		
3	1		4	
		4		1
	0			

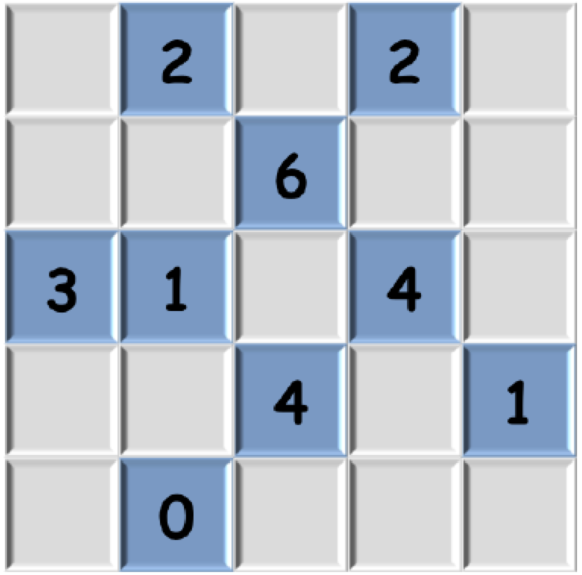
Observed samples

3	2	4	2	1
4	2	6	4	2
3	1	5	4	2
3	1	4	3	1
1	0	3	3	2

Estimate via nuclear norm min

Common Theme





- Underdetermined system of linear equations about $x \in \mathbb{R}^n, \mathbb{C}^n$

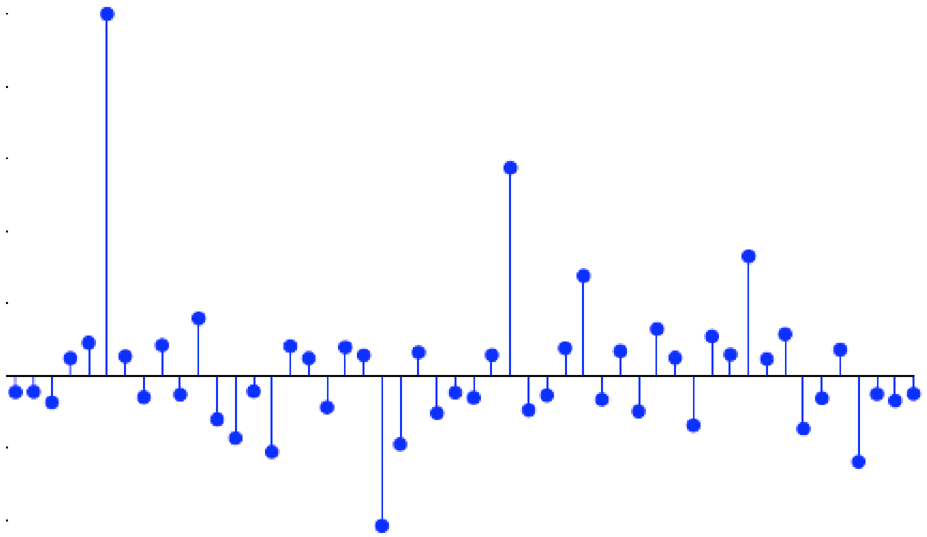
$$y_k = \langle a_k, x \rangle, \quad k = 1, \ldots, m, \qquad m \ll n$$

- Convex programming returns the correct solution

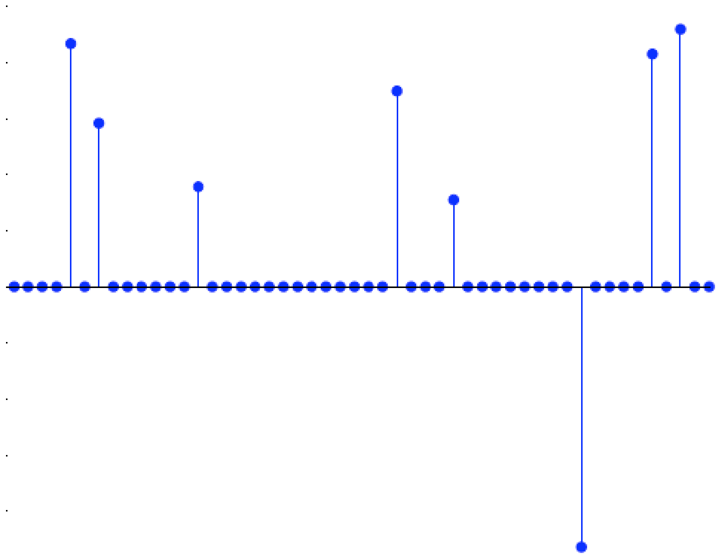
Recipe for success

1. Look for structured solutions (sparse signal or low rank matrix)
2. Recovery via convex programming (use L1 instead of L0 norm)
3. Incoherence in measurements

Sparsity



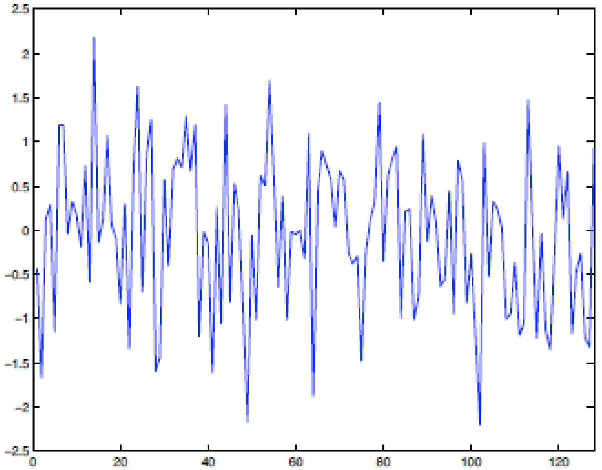
Nearly Sparse Signal



Sparse Signal

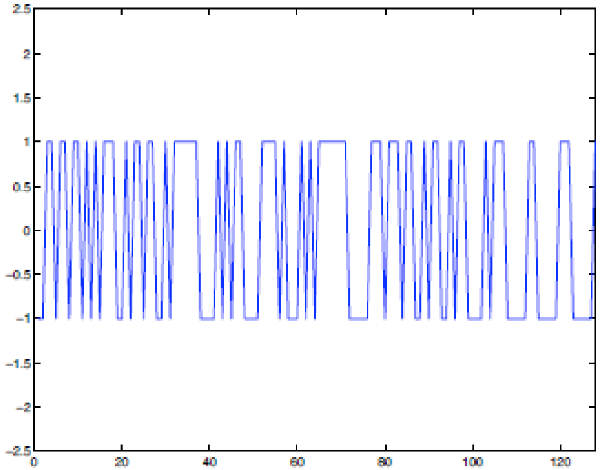
- x: signal coefficients in our convenient representation
- Collect information by measuring largest components of x
- Question: How to choose the basis and measure when the positions are not known in advance?
  - How to reconstruct the signal?

Incoherent (random sensing)



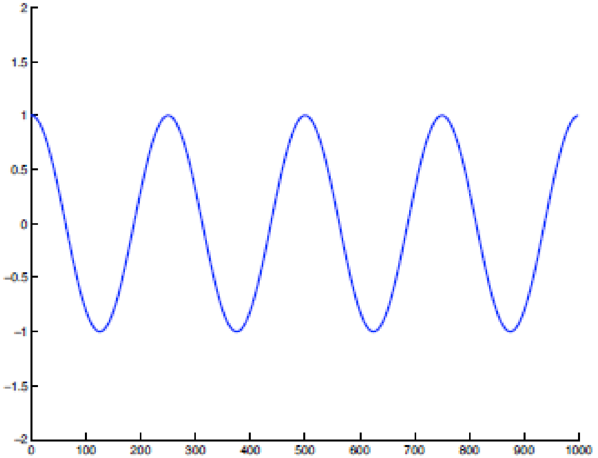
random waveform $N(0,1)$

a_k i.i.d. N(0,1) (white noise)



random waveform ± 1

a_k i.i.d. 1

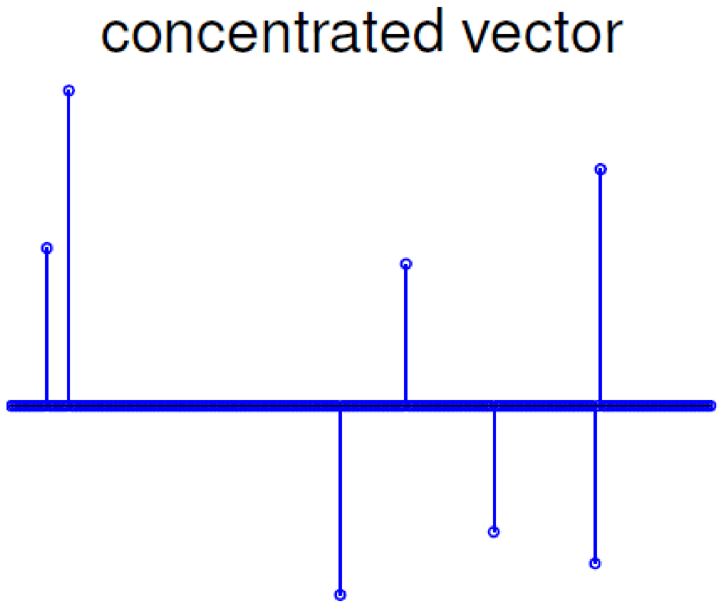


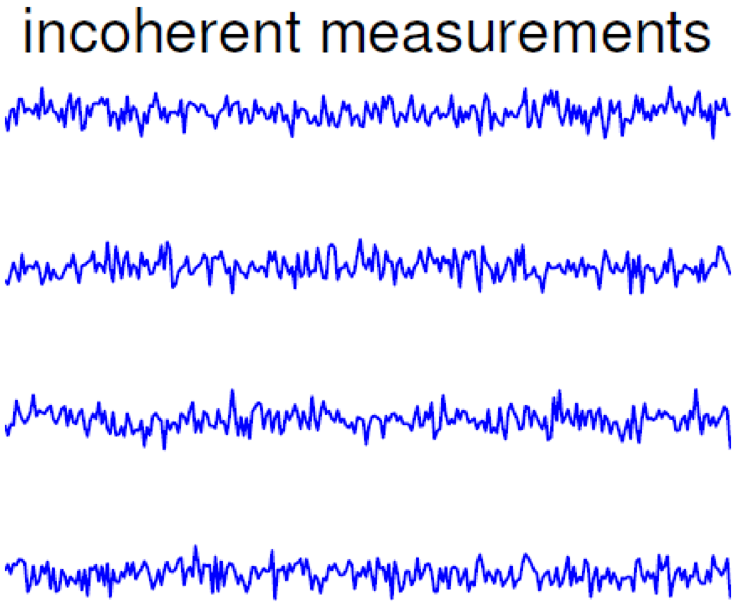
random sinusoid

a_k = exp(i2ω_kt) with i.i.d. frequencies ω_k

- $y =< a_k, x >, k = 1,2,3, \ldots, m$: These are m measurements using inner products with x
- Want sensing waveforms as spread out/“incoherent" as possible
- Span of $\{a_k\}$ should be as random as possible (general orientation)
  - $a_k \sim F\ (i.i.d.)$ {Come from the same distribution}
  - $E[a_k a_k^*] = \mathrm{I}$ and $\mathrm{a_k}$ spread out {a*: Is the complex conjugate transpose of a}

Incoherence





- Signal is local, measurements are global
- Each measurement picks up a little information about each component
- Triangulate significant components from measurements
- Formalization: Relies on uncertainty principles between sparsity basis and incoherent measurements

Example of Foundational Result

- Classical viewpoint
  - Measure everything (all the pixels, all the coefficients)
  - Keep d largest coefficients in the sparse representation: distortion is $||x - x_d||$
- Compressed sensing viewpoint
  - Take m random measurements: $y_k =< x, a_k >$
  - Reconstruct by linear programming: $(||x||_{l_1} = \sum_i |x_i|)$

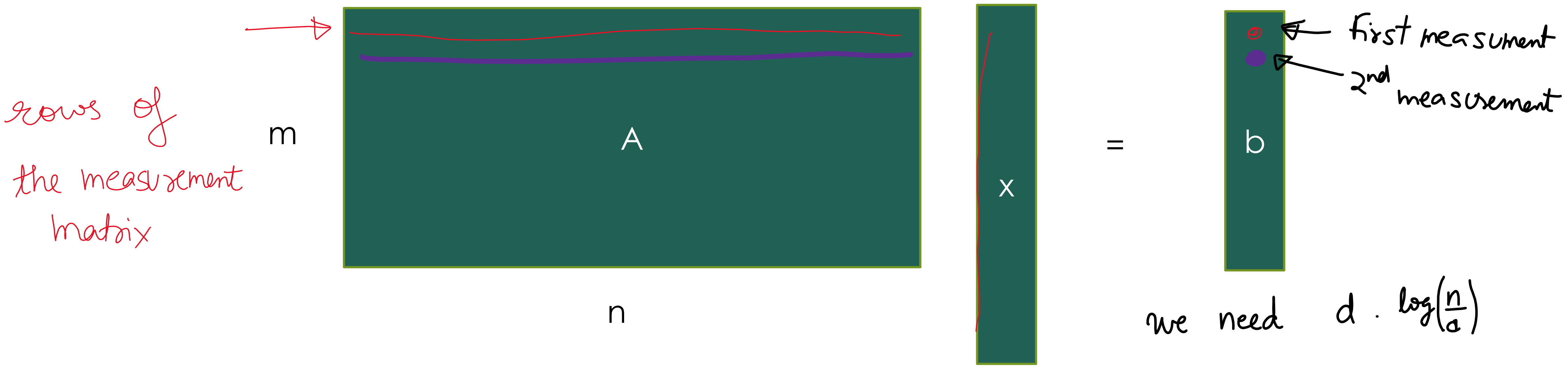
$$\boxed{x^* = \arg\min ||\check{x}||_{l_1} \text{ subject to } y_k =< \check{x}, a_k > \text{ , k = 1,2,...,m}}$$

- Among all the objects consistent with data, pick *min* l_1

Same performance as the classical method with about $m = d \log\left(\frac{n}{d}\right)$ measurements:

$$\boxed{||x^* - x||_{l_2} \leq ||x - x_d||_{l_2}}$$

Sparsest solutions of Linear equations



- Find a sparsest solution of linear system:

$$(P_0) \qquad \boxed{\min\{||x||_0 : Ax = b, x \in R^n\}}$$

Where $||x||_0$ = number of non zeros of x and $A \in R^{mxn}$ with m<n

The solution of P_0 is in general not unique and this problem is NP Hard.

Basis Pursuit

- Let's try to solve this in L_1 norm

$$(P_1) \qquad \boxed{\min\{\|x\|_1 : Ax = b, x \in R^n\}}$$

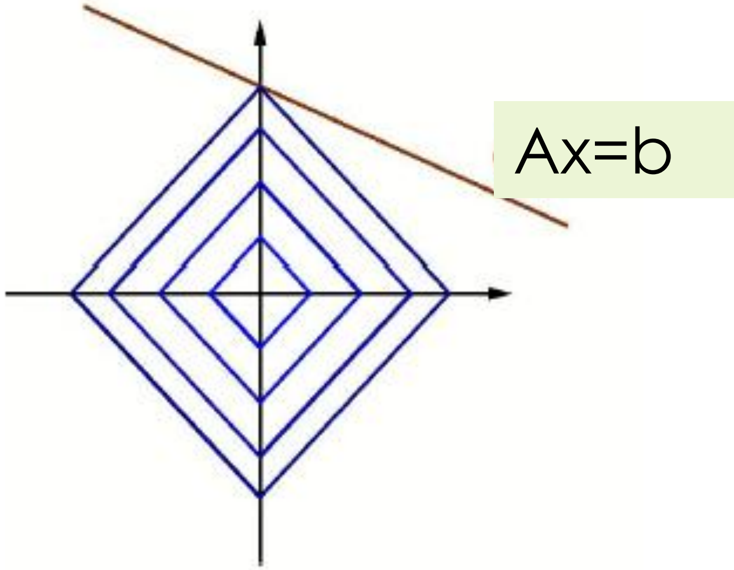
Where $\|x\|_1 = \sum_i |x_i|$ and $A \in R^{mxn}$ with m<n

The solution of P_1 is possible in polynomial time by linear programming using simplex

$$\min \sum_i y_i$$

s.t. Ax=b and

$$-y_i \le x_i \le y_i$$



Faster algorithms have been developed for solving this problem

Sparse Recovery and Mutual Incoherence

Mutual incoherence (column vectors of A):

$$M(A) = \max_{i \neq j} |a_i^* a_j|$$

Where $A = [a_1 \ \ldots a_n] \in R^{mxn}$ and $||a_i||_2 = 1$

Suppose that for the sparsest solution x^* we have

$$||x^*||_0 < \frac{\sqrt{2} - \frac{1}{2}}{M(A)}$$

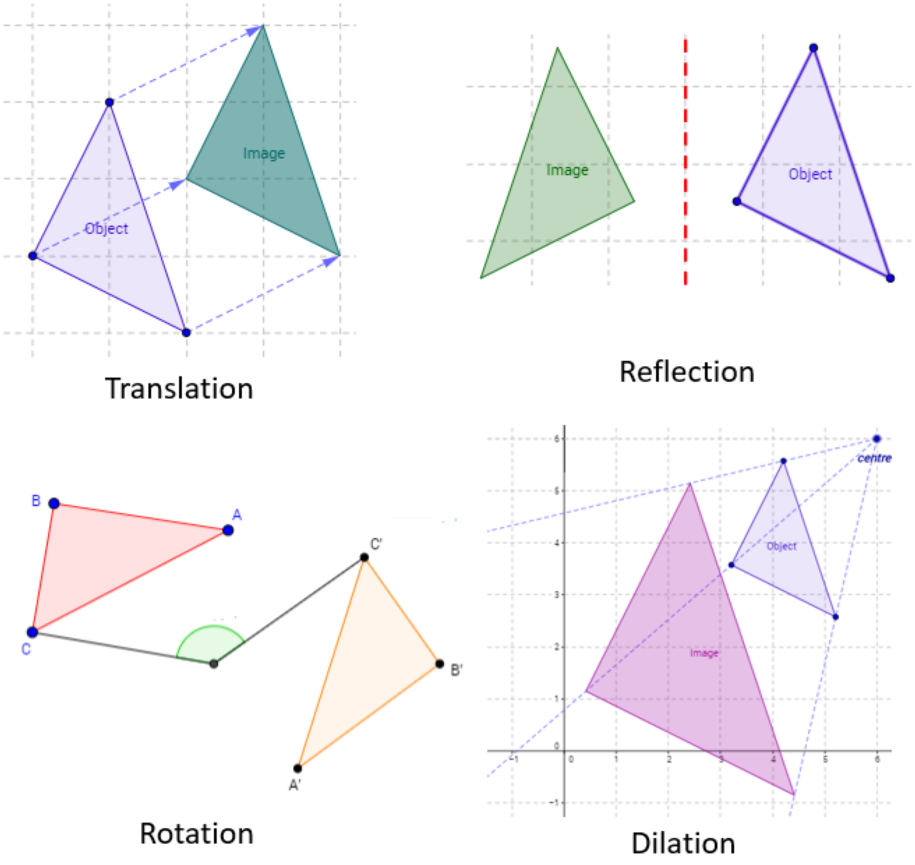
Then the solution of P_1 is equal to the solution of P_0, w.h.p.

Small value of M(A) will guarantee sparse solutions (perfect recovery)

Isometric mappings: Distance Preserving transformations

- Recall ISOMAP
- We want to maintain the pair-wise distance between points
- An isometry is a <u>transformation</u> which maps elements to the same or another metric space such that the distance between the image elements in the new metric space is equal to the distance between the elements in the original metric space.

Transformations



Sparse Recovery and RIP

Restricted Isometry Property of Order k

Let δ_k be the smallest number such that

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

for all k-sparse vectors $x \in R^n$ where $A = [a_1 \ \ldots a_n] \in R^{mxn}$

Theorem

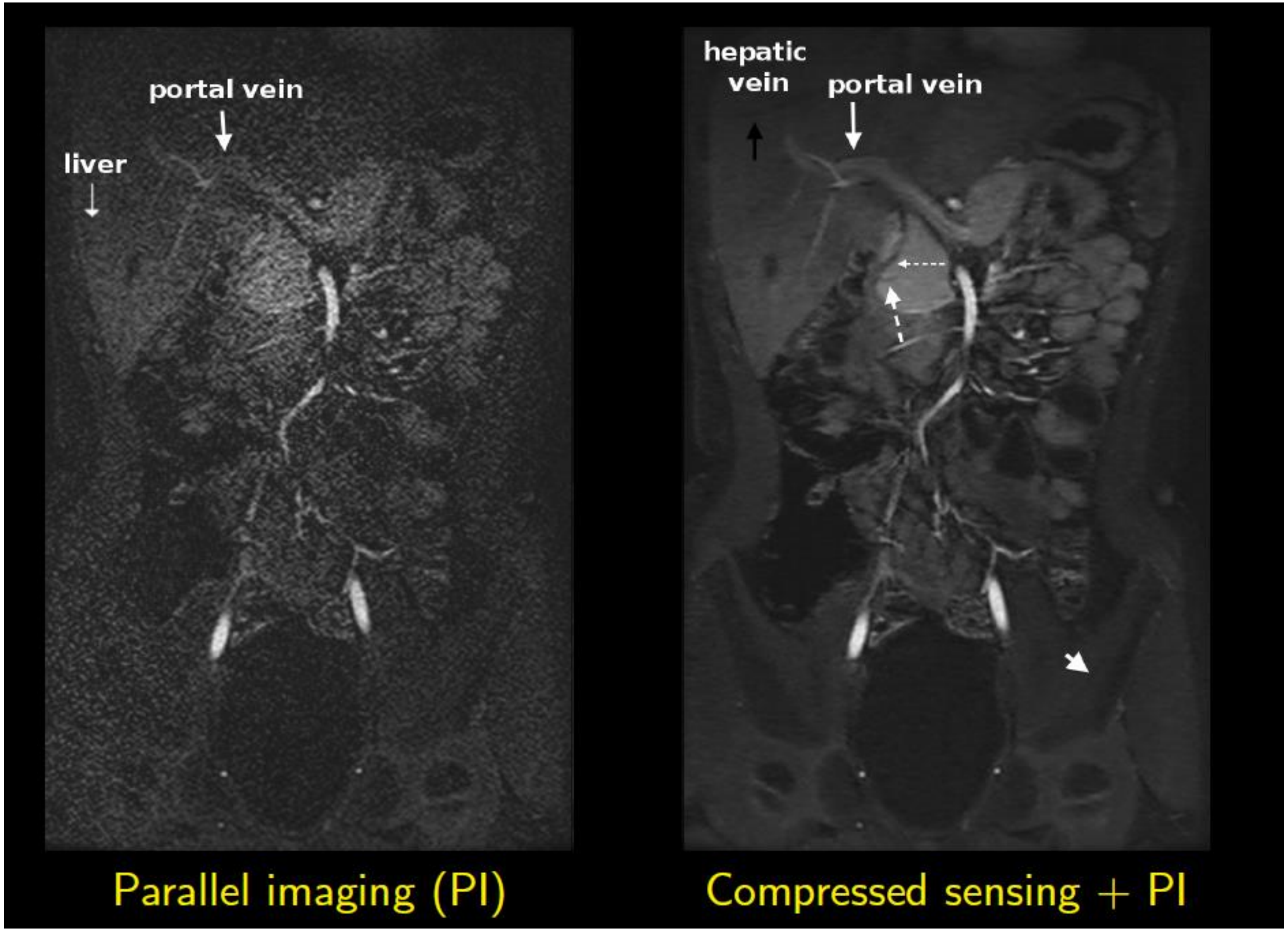
If $\delta_{2k} < \sqrt{2} - 1$, then for all k-sparse vectors x such that Ax=b, the solution of P_1 is equal to the solution of P_0.

Approximate Recovery and RIP

- In case the measurements are noisy (which almost always will be the case)
- We solve for $\boxed{\min\{||x||_1 : ||Ax - b||_2 < \epsilon \quad , x \in R^n\}}$
- Again, if A satisfies RIP for 2k sparse vectors, the above solution would be very close to the exact solution
- We can construct such matrices (RIP) using random Gaussian vectors with unit variance coordinates
- Pick $\boldsymbol{k}$ Gaussian vectors with unit variance; $\boldsymbol{u}_1, \dots, \boldsymbol{u}_{\boldsymbol{k}}$ i.i.d: $\boldsymbol{u}_i \sim N_D(0^D, I)$

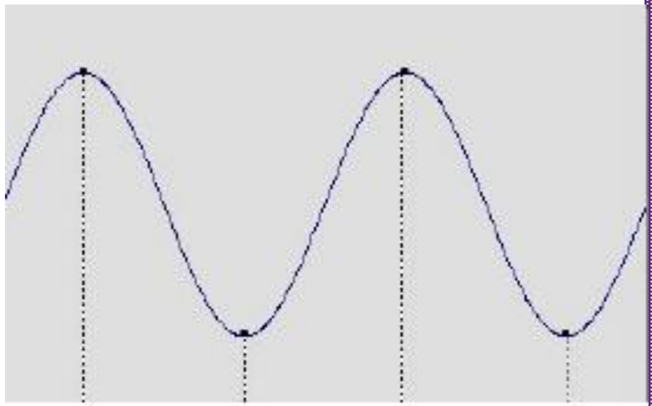
$$f(\boldsymbol{v}) = (\langle \boldsymbol{u}_1, \boldsymbol{v} \rangle, \dots, \langle \boldsymbol{u}_{\boldsymbol{k}}, \boldsymbol{v} \rangle)$$

6 year old boy’s abdomen



Summary

- Can obtain super-resolved signals from just a few sensors
  - Apparently breaking the Shannon-Nyquist Theorem
    - Requires sampling at twice the highest frequency (f)
    - If f is high, then it is difficult to build circuits to sample at the desired rate
- Sensing is nonadaptive: no effort to understand the signal
  - Sample is a linear functional applied to the signal ($F(x)=a^Tx$)
- Simple acquisition process followed by numerical optimization
- When will it work? Two conditions:
  - Sparsity: In some domain (choice of basis functions)
  - Incoherence (applied through Isometric property)



Applications
Analog to Digital
RF Receivers
Cameras
Medical Imaging