Introduction to compressed sensing

Yaniv Plan

September 8, 2015

- Seminal papers: (Candès, Romberg, Tao 2004), (Donoho 2004)
- For the past 9 years: $\sim$2 new papers per day

Motivating Problem: Magnetic Resonance Imaging (MRI)

"If Bryce took a single breath, the image would be blurred. That meant deepening the anesthesia enough to stop respiration. It would take a full two minutes for a standard MRI to capture the image, but if the anesthesiologists shut down Bryce's breathing for that long, his glitchy liver would be the least of his problems." [Wired Magazine 2010]

$+$

$=$

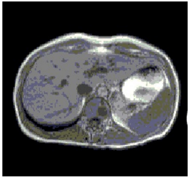


Figure: Standard MRI

Motivating Problem: Magnetic Resonance Imaging (MRI)

"If Bryce took a single breath, the image would be blurred. That meant deepening the anesthesia enough to stop respiration. It would take a full two minutes for a standard MRI to capture the image, but if the anesthesiologists shut down Bryce's breathing for that long, his glitchy liver would be the least of his problems." [Wired Magazine 2010]

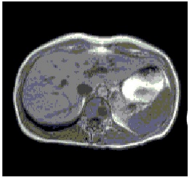
 $+$  $=$ 

Figure: Goal: reduce acquisition time

MRI: Singing protons!

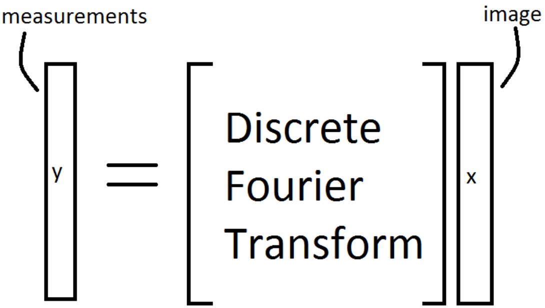


Figure: Standard MRI: Fully sampled DFT

- Image reconstruction: Invert the DFT

MRI: Singing protons!

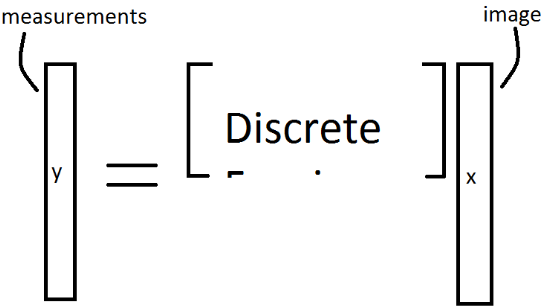


Figure: Goal: reconstruct image from part of the DFT

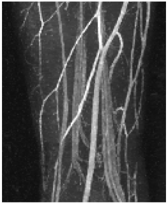
- Image reconstruction: Impossible?
- Is there any information that we have not used?
- Do images have structure?

Sparsity!

Sparse images

(a) Starry sky

(b) U. Mich.

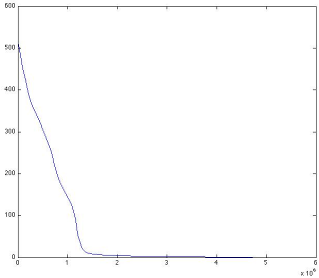


(c) Angiography (imaging blood veins)

- Are images generally sparse?

Signal sparsity

(a) Golden gate bridge



(b) Ordered wavelet coefficients

Images tend to be compressible, i.e. sparse in some basis.

- "Natural" images may be sparsified via the appropriate transform (wavelet, curvelet, shearlet...).
- Sparsity lives in audio signals, radar, statistical models, PDE solutions and much more.

Mathematics of sparsity

(Combining sparsity with undersampling)

Sparsity riddle

- Riddle: A sleeping dragon guards n locked treasure chests. One is filled with gold and the other $n-1$ are empty. You sneak into the dragon's lair carrying with you a large weighing machine. You wish to discover which chest holds the gold. How many times do you have to use the weighing machine?
- What if s of the chests contain gold, with $s \ll n$?

Sparsity riddle

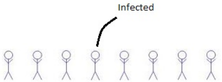
- Riddle: A sleeping dragon guards n locked treasure chests. One is filled with gold and the other $n-1$ are empty. You sneak into the dragon's lair carrying with you a large weighing machine. You wish to discover which chest holds the gold. How many times do you have to use the weighing machine?
  Ans: $\log n$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 100\text{ kg} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 100\text{ kg} \\ 100\text{ kg} \\ 0\text{ kg} \end{pmatrix}$$

- What if s of the chests contain gold, with $s \ll n$?

- Riddle: A sleeping dragon guards n locked treasure chests. One is filled with gold and the other $n-1$ are empty. You sneak into the dragon's lair carrying with you a large weighing machine. You wish to discover which chest holds the gold. How many times do you have to use the weighing machine?
- What if s of the chests contain gold, with $s \ll n$?

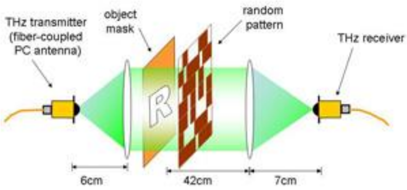
Group testing:



Sparsity riddle

- Riddle: A sleeping dragon guards n locked treasure chests. One is filled with gold and the other $n-1$ are empty. You sneak into the dragon's lair carrying with you a large weighing machine. You wish to discover which chest holds the gold. How many times do you have to use the weighing machine?
- What if s of the chests contain gold, with $s \ll n$?

Single pixel camera (first made at rice, now sold at InView)



Mathematical formalism

Measurements:

$$y_i = \langle a_i, x \rangle + z_i =$$

where z is a noise term, and x is sparse.

Mathematical formalism

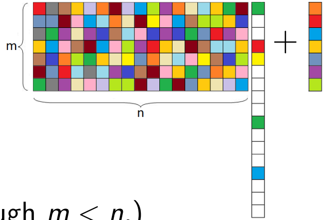
Measurements:

$$y_i = \langle a_i, x \rangle + z_i =$$

where z is a noise term, and x is sparse.

Matrix notation:

$$y = Ax + z =$$

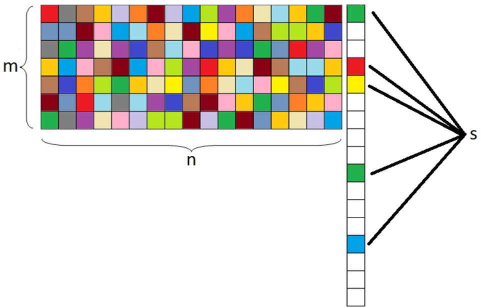


Goal: Estimate x. (Even though $m < n$.)

Noiseless lower bound

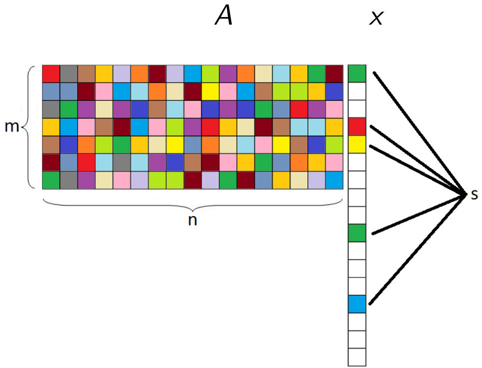
Question for the audience: What is a lower bound on the number of measurements m (rows of A) needed to recover x?

A $\qquad$ x



Noiseless lower bound

Question for the audience: What is a lower bound on the number of measurements m (rows of A) needed to recover x?



Answer: Recovery is impossible if there exists $x' \neq x$ such that $Ax = Ax'$ and $\|x'\|_0 \leq s$.
$\Rightarrow$ We need for $m \geq 2s$ for *uniform* recovery.

Can we reconstruct x with $O(s)$ measurements?

Can we reconstruct $\boldsymbol{x}$ with $O(s)$ measurements?

Reconstruction method:

$$\min \left\|\boldsymbol{x}'\right\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y} \tag{0.1}$$

- $\|\boldsymbol{x}'\|_0 :=$ number of non-zero entries of $\boldsymbol{x}'$.

What are conditions on $\boldsymbol{A}$ that allow reconstruction of $\boldsymbol{x}$?

Bad: Let A_i be the i-th column of $\boldsymbol{A}$ and suppose $A_1 = A_2$. Then $Ae_1 = Ae_2 \Rightarrow$ reconstruction of sparse vectors is impossible.

Incoherence condition

Bad: Let A_i be the i-th column of $\boldsymbol{A}$ and suppose $A_1 = A_2$. Then $Ae_1 = Ae_2 \Rightarrow$ reconstruction of sparse vectors is impossible.

Definition (Incoherence)

Suppose (w.l.o.g.) that $\boldsymbol{A}$ is column normalized. Define

$$\mu(A) := \max_{i \neq j} \langle A_i, A_j \rangle$$

Good: μ is small.

- Predates compressed sensing.

Theory before compressed sensing

$$\min \left\| \boldsymbol{x}' \right\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y} \tag{0.2}$$

Theorem (Donoho, Elad 2003)

Suppose that $\mu \cdot s < 1/2$. Then $\hat{\boldsymbol{x}} = \boldsymbol{x}$.

- Sparsity + incoherence leads to exact reconstruction of $\boldsymbol{x}$.

Challenge 1:
Challenge 2:

Theory before compressed sensing

$$\min \left\| \boldsymbol{x}' \right\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y} \tag{0.2}$$

Theorem (Donoho, Elad 2003)

Suppose that $\mu \cdot s < 1/2$. Then $\hat{\boldsymbol{x}} = \boldsymbol{x}$.

- Sparsity + incoherence leads to exact reconstruction of $\boldsymbol{x}$.

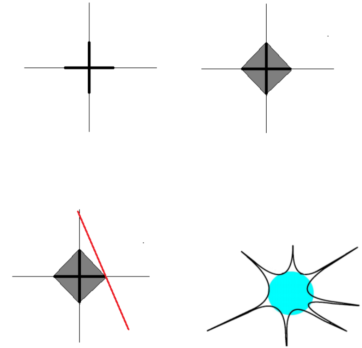
Challenge 1: Welch bound: Whenever $m < n/2$, $\mu > 1/\sqrt{2m}$.
$\Rightarrow$ the theorem requires $m \geq 2s^2$.
Challenge 2: Sparsity minimization takes exponential time. One needs to check each of the 2^m sparsity patterns.

ℓ_1 minimization (Chen, Donoho, Saunders 1999)

$$\min \left\|\boldsymbol{x}'\right\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y}. \tag{0.3}$$

ℓ_1 minimization (Chen, Donoho, Saunders 1999)

$$\min \left\| \boldsymbol{x}' \right\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y}. \tag{0.3}$$



(a) Subspace hits corner of ℓ_1 ball

(b) ℓ_1 ball in high dimension

Theory before compressed sensing

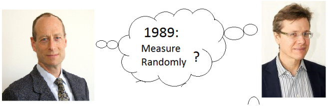
$$\min \left\| \boldsymbol{x}' \right\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y} \tag{0.4}$$

Theorem (Donoho, Elad 2003)

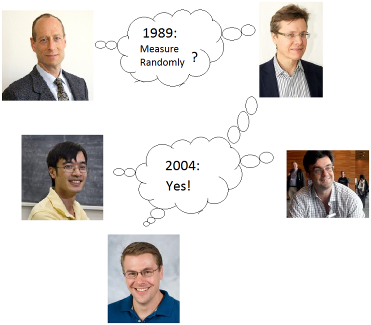
Suppose that $\mu \cdot s < 1/2$. Then $\hat{\boldsymbol{x}} = \boldsymbol{x}$.

Challenge 1: Welch bound: Whenever $m < n/2$, $\mu > 1/\sqrt{2m}$.
$\Rightarrow$ the theorem requires $m \geq 2s^2$.

Compressed sensing ~2004



Compressed sensing $\sim$2004



Theorem (Candès, Romberg, Tao 2004, Donoho 2004)

$m = O(s\log(n))$ random Fourier coefficients are sufficient to reconstruct an s-sparse vector by ℓ_1 minimization.

- Ignoring logarithmic terms, $O(s)$ measurements are sufficient. $\log(n)$ is the 'price of not knowing the support'.
- Sets off an explosion of work still going today.

Back to MRI

Do the conditions of the theorem match MRI applications?

Back to MRI

Do the conditions of the theorem match MRI applications?
"If Bryce took a single breath, the image would be blurred. That meant deepening the anesthesia enough to stop respiration. It would take a full two minutes for a standard MRI to capture the image, but if the anesthesiologists shut down Bryce's breathing for that long, his glitchy liver would be the least of his problems."
[Wired Magazine 2010]
Compressed sensing + Vasanawala + Lustig = "Vasanawala and Lustig needed only 40 seconds to gather enough data to produce a crystal-clear image of Bryces liver [...] good enough for Vasanawala to see the blockages in both bile ducts. An interventional radiologist snaked a wire into each duct, gently clearing the blockages and installing tiny tubes that allowed the bile to drain properly. And with that a bit of math and a bit of medicine Bryces lab test results headed back to normal."

Theory of compressed sensing.

- Original theorem relied on constructing a *dual certificate* and analyzing the geometry of the ℓ_1 ball (challenging).
- Over time, simpler methods were developed.

Restricted isometry property (RIP) [Candès and Tao 2005]

Definition (Restricted isometry property)

We say that $\boldsymbol{A}$ satisfies the restricted isometry property of order s if there exists $\delta_s \in (0,1)$ such that

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{Ax}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{x}\|_2^2$$

for all $\boldsymbol{x}$ satisfying $\|\boldsymbol{x}\|_0 \leq s$.

- Sparse signals are far from the null space of $\boldsymbol{A}$.
- Many kinds of *random matrices* satisfy the RIP with high probability as long as $m \geq O(s \text{ polylog}(n))$. No deterministic constructions are known!
- An appeal to *geometric functional analysis* or *probability in high dimensions*.

Implications of RIP

- $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.
-
$$\hat{\boldsymbol{x}} := \arg\min \left\|\boldsymbol{x}'\right\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y}$$

Theorem (Candès, Romberg, Tao 2005)

Suppose $\|\boldsymbol{x}\|_0 \leq s$ *and that* $\boldsymbol{A}$ *satisfies the restricted isometry property of order* $2s$ *with* $\delta_{2s} \leq \sqrt{2} - 1$*. Then*

$$\hat{\boldsymbol{x}} = \boldsymbol{x}.$$

- Robust to noise and approximate sparsity.
- Many other *greedy* reconstruction techniques are also exact under the RIP.

Implications of RIP

- $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.
- $$\hat{\boldsymbol{x}} := \arg\min \left\|\boldsymbol{x}'\right\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}' = \boldsymbol{y}$$

Theorem (Candès, Romberg, Tao 2005)

Suppose $\|\boldsymbol{x}\|_0 \leq s$ and that $\boldsymbol{A}$ satisfies the restricted isometry property of order $2s$ with $\delta_{2s} \leq \sqrt{2} - 1$. Then

$$\hat{\boldsymbol{x}} = \boldsymbol{x}.$$

Proof sketch:

1. Note: $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{A}(\hat{\boldsymbol{x}} - \boldsymbol{x}) = 0.$
2. By the RIP, $\|\boldsymbol{A}(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_2 \geq c\,\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2$. (Takes a bit of work.)

Thus $0 = \|\boldsymbol{A}(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_2 \geq c\,\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2$
$\Rightarrow \hat{\boldsymbol{x}} - \boldsymbol{x} = 0 \quad \blacksquare$.

Restricted Isometry Property (RIP)

RIP theory (Candès, Tao 2005, Vershynin, Rudelson 2006):

- If $\boldsymbol{A}$ is a Gaussian matrix, then we need $m = O(s\log(n/s))$ measurements.
- If $\boldsymbol{A}$ is a subsampled discrete Fourier transform, we need $m = O(s\log^4(n))$ measurements.
- Open problem: Prove the RIP for the DFT with $m = O(s\log n)$ (solution would solve the Λ_1 problem).

Q: Is the RIP requirement too strong?

A: RIP gives a *uniform* result, i.e., based on $\boldsymbol{A}$ an adversary can pick $\boldsymbol{x}$, but whatever sparse $\boldsymbol{x}$ she picks, it will be reconstructed. Can we weaken our assumptions if the signal is not adversarial?

A general, RIPless theory (Candès, P. 2011)

We assume $\boldsymbol{a}_i \stackrel{\text{dist}}{=} \boldsymbol{a} \sim F$ and require:

- **Isotropy property:**
$$\mathbb{E}\, \boldsymbol{a}\boldsymbol{a}^* = \mathrm{Id}\,.$$
- **Incoherence property:**
$$\mu := \max_{1 \le i \le n} \langle \boldsymbol{a}, \boldsymbol{e}_i \rangle^2 < \infty \tag{0.5}$$

Note:

- The isotropy condition prevents A from being rank deficient when enough samples are taken. ($\lim_{m\to\infty} \frac{1}{m}\boldsymbol{A}^*\boldsymbol{A} = \mathrm{Id} \quad$ a.s.)
- Regardless of F, $\mu \ge 1$. Low μ implies strong results.

Bad example: Subsampled identity matrix

$\boldsymbol{a} \sim \sqrt{n} \cdot \mathrm{Uniform}(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_n)$.

- **Isotropy property:**

$$\mathbb{E}\, \boldsymbol{a}\boldsymbol{a}^* = \frac{1}{n} \sum_{i=1}^{n} n e_i e_i^* = \mathrm{Id}\,.$$

- **Incoherence property:** $\mu := \sqrt{n}\, \langle e_1, \boldsymbol{e}_1 \rangle^2 = n \gg 1$

$$\boldsymbol{y} = \text{subset of the entries of } \boldsymbol{x}$$

$\Rightarrow$ Reconstruction of sparse $\boldsymbol{x}$ is impossible.

Measurement matrices that fit the model

- **Sensing vectors with independent components (with mean zero and variance 1).**
  - Gaussian ($\mu = 6 \log n$)
  - Binary ($\mu = 1$)
- **Subsampled orthogonal/unitary transforms.**
  - Fourier ($\mu = 1$)
- **Subsampled convolutions.** (The Fourier coefficients of the convolution vector must have magnitude (nearly) 1 to ensure (near) isotropy.)
- **Subsampled tight or continuous frames.**
  - Fourier sampled from continuous frequency spectrum.
- **Statistical linear model.**
  - Samples chosen independently from a population.

Theorem

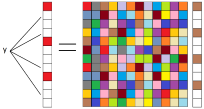
Theorem

Let x be an arbitrary fixed vector of length n with $\|x\|_0 = s$. Then with high probability $\hat{\boldsymbol{x}} = \boldsymbol{x}$ provided that $m \geq C \cdot \mu(F) \cdot s \cdot \log n$.

- Robust to noise and inexact sparsity.
- Prior stability results [Rudelson and Vershynin, 2008] for subsampled Fourier required $m \geq Cs \cdot \log^4(n)$ (using the RIP).

Near optimality

- **(Near) minimal number of measurements:** Take a_i to be a subsampled row from a complex DFT with $n = s \cdot t$. Take x to be the s-sparse dirac comb $\Rightarrow$ Fx is t-sparse.



It follows that $\langle a_i, x \rangle = 0$ for $i = 1, 2, \ldots, m$ with probability at least $1/n$ as long as $m \lesssim s \log n$, in which case no method can successfully recover x. (Note $\mu = 1$.)
This can be generalized to show that we need $m \gtrsim s\mu \log n$ when $\mu > 1$.

Statistical variable selection

Linear model in statistics

Example:

- $y_i :=$ lifespan of person i.
- $\boldsymbol{a}_i =$(exercise, weight, relationships, drugs, diet, ...)

Linear model:

$$y_i = \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle + \boldsymbol{z}_i \quad i = 1, 2, \ldots, m.$$

Sparsity: Only a small number of *covariates* are significant.
Challenge: $\boldsymbol{a}_i$ usually does not satisfy the probabilistic assumptions needed for compressed sensing.

Sparsity with deterministic measurements

- What if $\boldsymbol{A}$ is non-random?
  No RIP. Adversarial $\boldsymbol{x}$ may be in the null space of $\boldsymbol{A}$.
- Can we say something about most signals $\boldsymbol{x}$?
- Solution: Put a prior distribution on $\boldsymbol{x}$. Prove that $\boldsymbol{x}$ is well reconstructed with high probability.

Sparsity with deterministic measurements

- What if $\boldsymbol{A}$ is non-random?
  No RIP. Adversarial $\boldsymbol{x}$ may be in the null space of $\boldsymbol{A}$.
- Can we say something about most signals $\boldsymbol{x}$?
- Solution: Put a prior distribution on $\boldsymbol{x}$. Prove that $\boldsymbol{x}$ is well reconstructed with high probability.
- **Prior:** The non-zero entries of $\boldsymbol{x}$ are chosen at random, and the signs of the non-zero entries of $\boldsymbol{x}$ are random.

Support recovery

- $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{z}$
- $\boldsymbol{z}_i \sim N(0, \sigma^2)$
- LASSO:

$$\hat{\boldsymbol{x}} := \arg\min \left\|\boldsymbol{A}\boldsymbol{x}' - \boldsymbol{y}\right\|_2^2 + 2\sigma\sqrt{\log n}\left\|\boldsymbol{x}'\right\|_1$$

Theorem

Let T be the support of $\boldsymbol{x}$ and suppose $m \geq |T| \log n$. Suppose that

$$\textbf{Assumption 1:}\ \min_{i \in T} |x_i| > 8\sigma\sqrt{2\log n}$$

$$\textbf{Assumption 2:}\ \mu(\boldsymbol{A}) \lesssim \frac{1}{\log n} \quad \textit{and} \quad ||A|| \lesssim \sqrt{\frac{n}{m}}.$$

Then with high probability $\hat{\boldsymbol{x}}$ has the same support as $\boldsymbol{x}$.

Support recovery

Theorem

Let T be the support of $\boldsymbol{x}$ and suppose $m \geq |T| \log n$. Suppose that

$$\textbf{Assumption 1:} \quad \min_{i \in T} |x_i| > 8\sigma\sqrt{2 \log n}$$

$$\textbf{Assumption 2:} \quad \mu(\boldsymbol{A}) \lesssim \frac{1}{\log n} \quad \textit{and} \quad ||A|| \lesssim \sqrt{\frac{n}{m}}.$$

Then with high probability $\hat{\boldsymbol{x}}$ has the same support as $\boldsymbol{x}$.

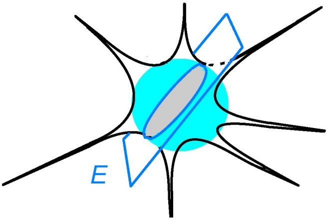
- There is also a near-optimal bound on $\|\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{A}\boldsymbol{x}\|_2$.
- Problem: An optimal bound on $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2$ without **Assumption 1**. (A sub-optimal bound is known (Dossal, Tesson 2012)).

Beyond compressed sensing

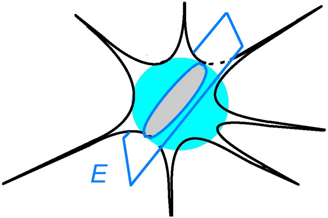
- **Theme 1: Subsampling.** The amount of information appears too small to reconstruct the signal.
- **Theme 2: Low dimensionality.** The signal resides in a set with small dimension in comparison to the ambient dimension.

General theory for Gaussian measurements

- Signal structure: $\boldsymbol{x} \in K \subset \mathbb{R}^n$.
- $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.



General theory for Gaussian measurements



Theorem (M^* estimate: Milman 1981, Pajor, Tomczak 1985, Mendelson, Pajor, Tomczak 2007)

Consider a random subspace $\mathbb{E}$ in $\mathbb{R}^n$ of codimension m. Then with high probability

$$diam(K \cap E) \lesssim \frac{w(K)}{\sqrt{m}}.$$

- $w(K)$ is the *mean width* of K.

Matrix completion

$$\begin{pmatrix} M_{1,1} & M_{1,2} & M_{1,3} & M_{1,4} \\ M_{2,1} & M_{2,2} & M_{2,3} & M_{2,4} \\ M_{3,1} & M_{3,2} & M_{3,3} & M_{3,4} \\ M_{4,1} & M_{4,2} & M_{4,3} & M_{4,4} \end{pmatrix}, \quad M_{i,j} = \text{How much user } i \text{ likes movie } j$$

Matrix completion

$$\begin{pmatrix} ? & M_{1,2} & ? & M_{1,4} \\ M_{2,1} & ? & M_{2,3} & M_{2,4} \\ M_{3,1} & ? & M_{3,3} & M_{3,4} \\ ? & M_{4,2} & ? & M_{4,4} \end{pmatrix}, \quad M_{i,j} = \text{How much user } i \text{ likes movie } j$$

Matrix completion

$$\begin{pmatrix} ? & M_{1,2} & ? & M_{1,4} \\ M_{2,1} & ? & M_{2,3} & M_{2,4} \\ M_{3,1} & ? & M_{3,3} & M_{3,4} \\ ? & M_{4,2} & ? & M_{4,4} \end{pmatrix}, \quad M_{i,j} = \text{How much user } i \text{ likes movie } j$$

Rank-1 model:

- $a_j =$ Amount of action in movie j
- $x_i =$ How much user i likes action

$$M_{i,j} = x_i \cdot a_j$$

Matrix completion

$$\begin{pmatrix} ? & M_{1,2} & ? & M_{1,4} \\ M_{2,1} & ? & M_{2,3} & M_{2,4} \\ M_{3,1} & ? & M_{3,3} & M_{3,4} \\ ? & M_{4,2} & ? & M_{4,4} \end{pmatrix}, \quad M_{i,j} = \text{How much user } i \text{ likes movie } j$$

Rank-1 model:

- a_j = Amount of action in movie j
- x_i = How much user i likes action

$$M_{i,j} = x_i \cdot a_j$$

Rank-2 model:

- b_j = Amount of comedy in movie j
- y_i = How much user i likes comedy

$$M_{i,j} = x_i \cdot a_j + y_i \cdot b_j$$

Example 2: Sparse binary data

Example: Predicting heart attacks (10 year study).

- $y_i = \begin{cases} 1 & \text{if patient } i \text{ has heart attack in 10 year window} \\ -1 & \text{otherwise} \end{cases}$
- $\boldsymbol{a}_i =$ (cholesterol, weight, BMI, blood-test results, exercise habits...).

Sparsity: Only a few factors are significant.
Which ones?

Riddle answer

Riddle: A sleeping dragon guards n locked treasure chests. s chests are filled with gold and the other $n-s$ are empty. You sneak into the dragon's lair carrying with you a large weighing machine. You wish to discover which chest holds the gold. How many times do you have to use the weighing machine?

Ans: Make $m = O(s\log(n/s))$ measurements of the weight of a random subset of chests. Reconstruct all weights using ℓ_1 minimization.

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 100\text{ kg} \\ 0 \\ 100\text{ kg} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 200\text{ kg} \\ 100\text{ kg} \\ 0\text{ kg} \end{pmatrix}$$

Riddle answer

Ans: Make $m = O(s \log(n/s))$ measurements of the weight of a random subset of chests. Reconstruct all weights using ℓ_1 minimization.

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 100 \text{ kg} \\ 0 \\ 100 \text{ kg} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 200 \text{ kg} \\ 100 \text{ kg} \\ 0 \text{ kg} \end{pmatrix}$$

Q: Does the above matrix satisfy *any of* the assumptions of the theory we discussed?

A: No, but we can *precondition* the matrix by projecting onto the space orthogonal to the all ones vector.

Thank you!

yanivplan.com