Bridging Matrix Recovery Gaps using Manifolds

Deanna Needell

Claremont McKenna College
Joint work with Y. C. Eldar [Technion], Y. Plan [Univ. of Michigan]

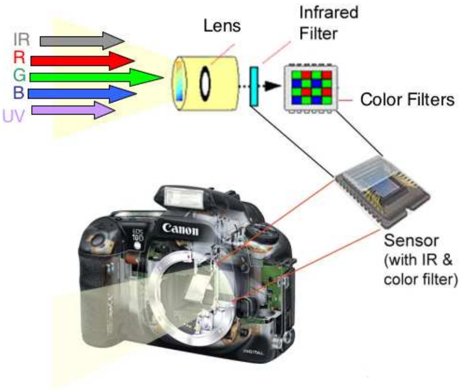
ANTC Seminar, Claremont, Jan. 2012

Outline

- Compressed Sensing (CS)
  - Applications
  - Mathematical Formulation
  - Best known results
- CS's sister: Matrix recovery
  - Applications
  - Mathematical Formulation
  - Best known results
- Comparison of the two problems
  - The question unanswered
  - Our answer
  - Proof via manifold theory

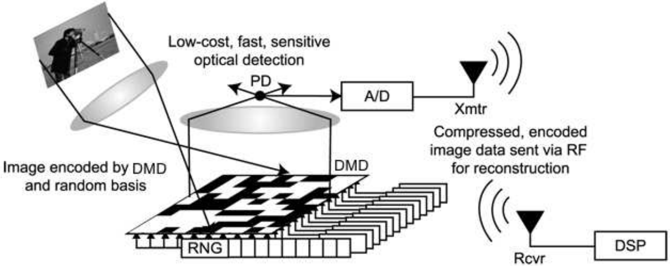
Digital Cameras

Today's digital cameras already "old school?"



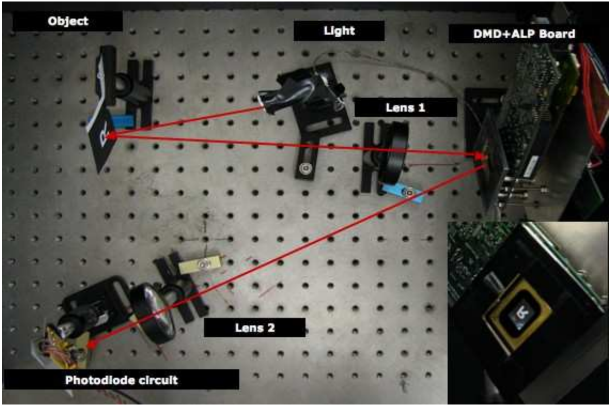
Digital Cameras

Save your nickels to buy the new digital camera?



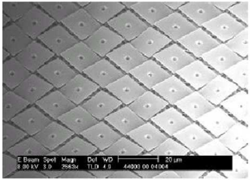
Digital Cameras

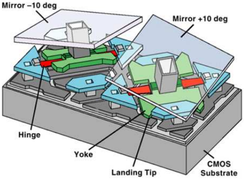
Save your nickels to buy the new digital camera?



Digital Cameras

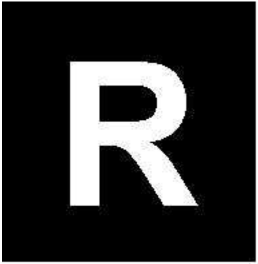
Save your nickels to buy the new digital camera?





Digital Cameras

Save your nickels to buy the new digital camera?



(Original)

(2%)

Digital Cameras

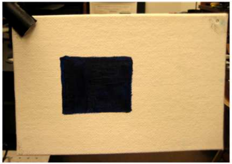
Save your nickels to buy the new digital camera?

(Original)

(10%)

Digital Cameras

Save your nickels to buy the new digital camera?



20%

5%

MRI

Feeling claustrophobic?

It'll only last a quick 45 minutes...

MRI

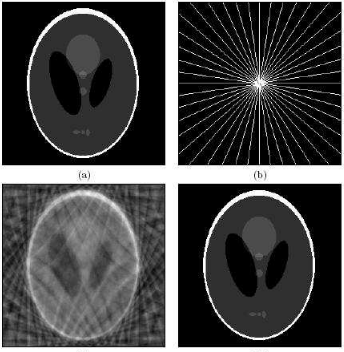
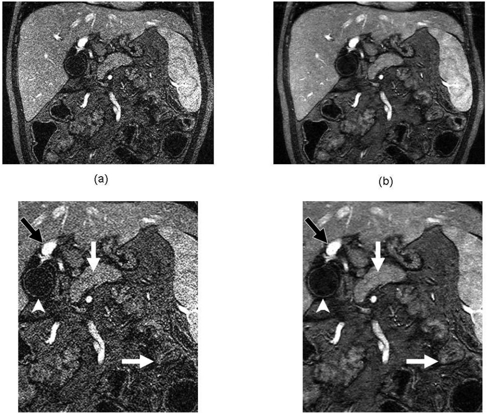


Figure 1: Example of a simple recovery problem. (a) The Logan-Shepp phantom test image. (b) Sampling domain Ω in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total variation, as in (1.1). The reconstruction is an exact replica of the image in (a).

Pediatric MRI



(a) (b)

(c) (d)

(a–d) Submillimeter near-isotropic–resolution contrast-enhanced T1-weighted MR images in 8-year-old boy. (a, c) Standard and (b, d) compressed sensing reconstruction images. (c, d) Zoomed images show improved delineation of the pancreatic duct (vertical arrow), bowel (horizontal arrow), and gallbladder wall (arrowhead), and equivalent definition of portal vein (black arrow) with L1 SPIR-iT reconstruction.

(Caffey Award : Faster Pediatric MRI Via Compressed Sensing - Shreyas Vasanawala et.al. (Stanford University))

Many more...

- Radar
- Error Correction
- Computational Biology (DNA Microarrays)
- Geophysical Data Analysis
- Data Mining, classification
- Neuroscience
- ...

The mathematical problem

1. Signal of interest $f \in \mathbb{R}^d$
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Measurements $y = Af$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & & A & & \end{bmatrix} \begin{bmatrix} f \\ \\ \end{bmatrix}$$

4. Problem: Reconstruct signal f from measurements y

The mathematical problem

1. Signal of interest $f \in \mathbb{R}^d$
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Measurements $y = Af$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

4. Problem: Reconstruct signal f from measurements y

The mathematical problem

1. Signal of interest $f \in \mathbb{R}^d$
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Measurements $y = Af$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & & A & & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

4. Problem: Reconstruct signal f from measurements y

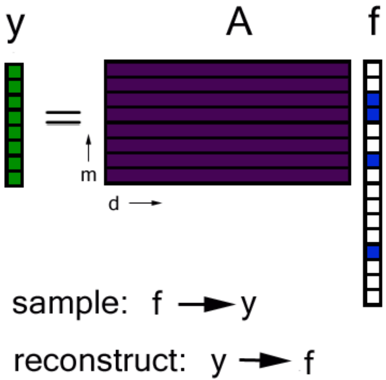
The mathematical problem

1. Signal of interest $f \in \mathbb{R}^d$
2. Measurement matrix $A : \mathbb{R}^d \to \mathbb{R}^m$.
3. Measurements $y = Af$.

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

4. Problem: Reconstruct signal f from measurements y

The mathematical problem



Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension d suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll d$.

In practice, we encounter compressible signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension d suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll d$.

In practice, we encounter compressible signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension d suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll d$.

In practice, we encounter compressible signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension d suggests.

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll d$.

In practice, we encounter compressible signals, and the measurements have noise. (Not in this talk.)

How do we actually reconstruct?

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

Important Questions

- What kind(s) of measurement matrices A?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

How do we actually reconstruct?

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & A & \quad \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

Important Questions

- What kind(s) of measurement matrices A?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

How do we actually reconstruct?

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

Important Questions

- What kind(s) of measurement matrices A?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

How do we actually reconstruct?

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad & A & \quad \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

Important Questions

- What kind(s) of measurement matrices A?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

How do we actually reconstruct?

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

Important Questions

- What kind(s) of measurement matrices A?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

ℓ_0-optimization

The First CS Theorem

Let A be one-to-one on s-sparse vectors and set:

$$\hat{f} = \operatorname*{argmin}_g \|g\|_0 \quad \text{such that} \quad Ag = y.$$

Then in the noiseless case, we have perfect recovery of all s-sparse signals: $\hat{f} = f$.

Proof:

Easy!

Moral of the story:

Theoretically, we need only $m = 2s$ measurements.

ℓ_0-optimization

The First CS Theorem

Let A be one-to-one on s-sparse vectors and set:

$$\hat{f} = \operatorname*{argmin}_{g} \|g\|_0 \quad \text{such that} \quad Ag = y.$$

Then in the noiseless case, we have perfect recovery of all s-sparse signals: $\hat{f} = f$.

Proof:

Easy!

Moral of the story:

Theoretically, we need only $m = 2s$ measurements.

ℓ_0-optimization

The First CS Theorem

Let A be one-to-one on s-sparse vectors and set:

$$\hat{f} = \operatorname*{argmin}_g \|g\|_0 \quad \text{such that} \quad Ag = y.$$

Then in the noiseless case, we have perfect recovery of all s-sparse signals: $\hat{f} = f$.

Proof:

Easy!

Moral of the story:

Theoretically, we need only $m = 2s$ measurements.

Just relax: ℓ_1-optimization

Relaxation [Candès-Tao]

Let A satisfy the *Restricted Isometry Property* for $2s$-sparse vectors and set:

$$\hat{f} = \operatorname*{argmin}_g \|g\|_1 \quad \text{such that} \quad Ag = y.$$

Then in the noiseless case, we have perfect recovery of all s-sparse signals: $\hat{f} = f$.

Proof:

(Not so easy)

Just relax: ℓ_1-optimization

Relaxation [Candès-Tao]

Let A satisfy the *Restricted Isometry Property* for $2s$-sparse vectors and set:

$$\hat{f} = \underset{g}{\operatorname{argmin}} \|g\|_1 \quad \text{such that} \quad Ag = y.$$

Then in the noiseless case, we have perfect recovery of all s-sparse signals: $\hat{f} = f$.

Proof:

(Not so easy)

Restricted Isometry Property

- A satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

$$(1-\delta)\|f\|_2 \le \|Af\|_2 \le (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \le s.$$

- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

Moral of the story:

Practically, we need only $m = s \log d$ measurements.

Restricted Isometry Property

- A satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

$$(1-\delta)\|f\|_2 \le \|Af\|_2 \le (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \le s.$$

- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

Moral of the story:

Practically, we need only $m = s \log d$ measurements.

Restricted Isometry Property

- A satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

$$(1-\delta)\|f\|_2 \leq \|Af\|_2 \leq (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \leq s.$$

- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

Moral of the story:

Practically, we need only $m = s \log d$ measurements.

Restricted Isometry Property

- A satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

$$(1-\delta)\|f\|_2 \leq \|Af\|_2 \leq (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \leq s.$$

- Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

Moral of the story:

Practically, we need only $m = s \log d$ measurements.

The gap

Problem:	CS
Theoretical	$\min \|f\|_0$
Practical	$\min \|f\|_1$
m for Practical	$m \gtrsim s \log n$
m for Theoretical	$m \geq 2s$

The gap

Problem:	CS
Theoretical	$\min \|f\|_0$
Practical	$\min \|f\|_1$
m for Practical	$m \gtrsim s \log n$
m for Theoretical	$m \geq 2s$

The gap

Problem:	CS
Theoretical	$\min \|f\|_0$
Practical	$\min \|f\|_1$
m for Practical	$m \gtrsim s \log n$
m for Theoretical	$m \geq 2s$

The gap

Problem:	CS
Theoretical	$\min \|f\|_0$
Practical	$\min \|f\|_1$
m for Practical	$m \gtrsim s \log n$
m for Theoretical	$m \geq 2s$

The gap

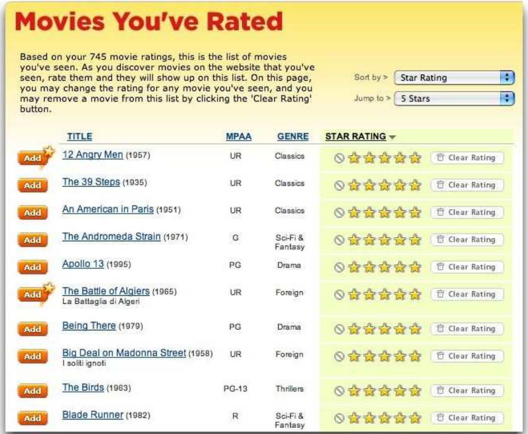
Problem:	CS
Theoretical	$\min \|f\|_0$
Practical	$\min \|f\|_1$
m for Practical	$m \gtrsim s \log n$
m for Theoretical	$m \geq 2s$

The Netflix problem

That check is worth how much??

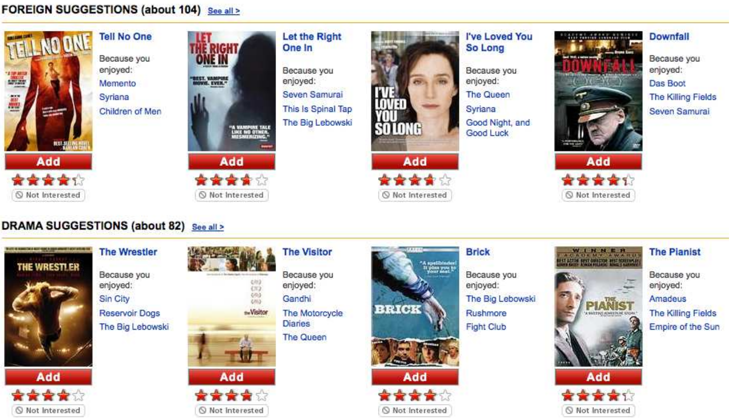
The Netflix problem

Tell us how you really feel...



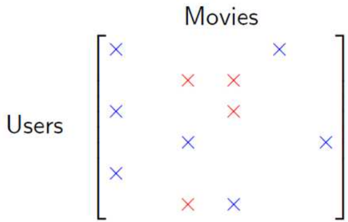
The Netflix problem

And we'll tell you how you really feel...



Collaborative Filtering

We can use other people's preferences too, but still...

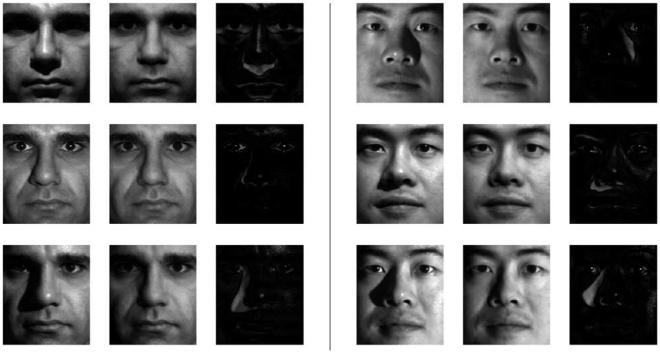


Surveillance

Separation of foreground and background!

Computer Vision

Removing shadow and lighting effects!



Removing corruptions

And now enjoy the film...

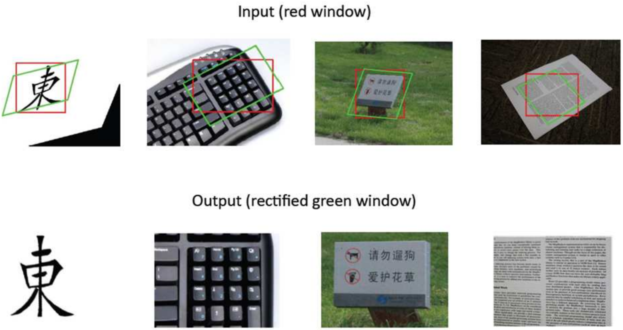
Original D

Corruptions

Repaired A

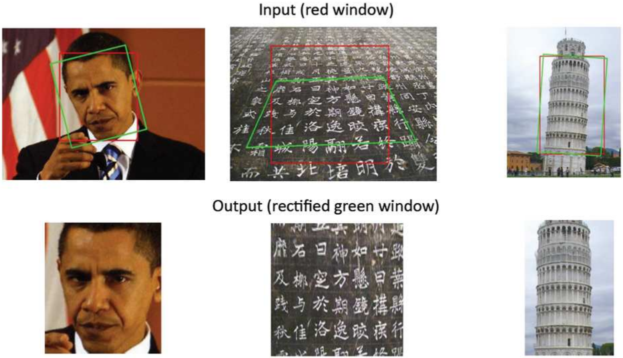
Tilt [Candès et.al.]

For humans and computers who have trouble reading sideways...



Tilt [Candès et.al.]

Fixing the leaning tower without any digging!



The mathematical problem

1. Signal of interest $X \in \mathbb{R}^{n \times n}$
2. Linear measurement operator $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$.
3. Measurements $y = \mathcal{A}(X)$ of the form:

   $$(\mathcal{A}(X))_i = \langle A_i, X \rangle = \text{trace}(A_i^* X) \text{ for } A_i \in \mathbb{R}^{n \times n}$$

4. Problem: Reconstruct signal X from measurements y

The mathematical problem

1. Signal of interest $X \in \mathbb{R}^{n \times n}$
2. Linear measurement operator $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$.
3. Measurements $y = \mathcal{A}(X)$ of the form:

   $$(\mathcal{A}(X))_i = \langle A_i, X \rangle = \mathrm{trace}(A_i^* X) \text{ for } A_i \in \mathbb{R}^{n \times n}$$

4. Problem: Reconstruct signal X from measurements y

The mathematical problem

1. Signal of interest $X \in \mathbb{R}^{n \times n}$
2. Linear measurement operator $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$.
3. Measurements $y = \mathcal{A}(X)$ of the form:

   $$(\mathcal{A}(X))_i = \langle A_i, X \rangle = \text{trace}(A_i^* X) \text{ for } A_i \in \mathbb{R}^{n \times n}$$

4. Problem: Reconstruct signal X from measurements y

The mathematical problem

1. Signal of interest $X \in \mathbb{R}^{n \times n}$
2. Linear measurement operator $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$.
3. Measurements $y = \mathcal{A}(X)$ of the form:

   $$(\mathcal{A}(X))_i = \langle A_i, X \rangle = \text{trace}(A_i^* X) \text{ for } A_i \in \mathbb{R}^{n \times n}$$

4. Problem: Reconstruct signal X from measurements y

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension $n \times n$ suggests.

Assume X is low-rank: $\text{rank}(X) \leq r$. In practice, we encounter approximately low-rank signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension $n \times n$ suggests.

Assume X is low-rank: $\text{rank}(X) \le r$. In practice, we encounter approximately low-rank signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension $n \times n$ suggests.

Assume X is low-rank: $\text{rank}(X) \leq r$. In practice, we encounter approximately low-rank signals, and the measurements have noise. (Not in this talk.)

Wait, isn't this impossible?

Without further assumptions, this problem is ill-posed.

Why will this work?

Most signals of interest contain far less information than their dimension $n \times n$ suggests.

Assume X is low-rank: $\text{rank}(X) \leq r$. In practice, we encounter approximately low-rank signals, and the measurements have noise. (Not in this talk.)

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

How do we actually reconstruct?

Important Questions

- What kind(s) of linear operators $\mathcal{A}$?
- How many measurements needed?
- Are the guarantees *uniform*?
- Is algorithm *stable*?
- Fast runtime?

Critical Connection

A matrix X is low-rank if and only if its vector $\sigma(X)$ of singular values is sparse!

Rank optimization

ℓ_0-minimization

$$\hat{f} = \underset{g}{\operatorname{argmin}} \|g\|_0 \quad \text{such that} \quad Ag = y.$$

Rank-minimization

$$\hat{X} = \underset{M}{\operatorname{argmin}} \|\sigma(M)\|_0 = \underset{M}{\operatorname{argmin}} \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

Rank optimization

ℓ_0-minimization

$$\hat{f} = \operatorname*{argmin}_g \|g\|_0 \quad \text{such that} \quad Ag = y.$$

Rank-minimization

$$\hat{X} = \operatorname*{argmin}_M \|\sigma(M)\|_0 = \operatorname*{argmin}_M \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

Relaxed optimization

ℓ_1-minimization

$$\hat{f} = \operatorname*{argmin}_g \|g\|_1 \quad \text{such that} \quad Ag = y.$$

Nuclear norm minimization

$$\hat{X} = \operatorname*{argmin}_M \|\sigma(M)\|_1 \quad \text{such that} \quad \mathcal{A}(M) = y$$

Nuclear norm

$$\|M\|_* = \|\sigma(M)\|_1 = \operatorname{trace}(\sqrt{M^*M})$$

Relaxed optimization

ℓ_1-minimization

$$\hat{f} = \operatorname*{argmin}_g \|g\|_1 \quad \text{such that} \quad Ag = y.$$

Nuclear norm minimization

$$\hat{X} = \operatorname*{argmin}_M \|\sigma(M)\|_1 \quad \text{such that} \quad \mathcal{A}(M) = y$$

Nuclear norm

$$\|M\|_* = \|\sigma(M)\|_1 = \text{trace}(\sqrt{M^* M})$$

Relaxed optimization

ℓ_1-minimization

$$\hat{f} = \underset{g}{\operatorname{argmin}} \|g\|_1 \quad \text{such that} \quad Ag = y.$$

Nuclear norm minimization

$$\hat{X} = \underset{M}{\operatorname{argmin}} \|\sigma(M)\|_1 \quad \text{such that} \quad \mathcal{A}(M) = y$$

Nuclear norm

$$\|M\|_* = \|\sigma(M)\|_1 = \operatorname{trace}(\sqrt{M^* M})$$

Nuclear norm optimization

Theorem [Oymak-Hassibi]

Let $\mathcal{A}$ be a Gaussian linear operator and set,

$$\hat{X} = \underset{M}{\text{argmin}} \|M\|_* \quad \text{such that} \quad \mathcal{A}(M) = y,$$

where $\|M\|_* = \text{trace}(\sqrt{M^*M}) = \|\sigma(M)\|_1$.
Then in the noiseless case, to guarantee perfect recovery of any rank-r matrix X, we need only $m = 16nr$ measurements.

Moral of the story:

Practically, we need only $m = 16nr$ measurements.

Nuclear norm optimization

Theorem [Oymak-Hassibi]

Let $\mathcal{A}$ be a Gaussian linear operator and set,

$$\hat{X} = \underset{M}{\text{argmin}} \|M\|_* \quad \text{such that} \quad \mathcal{A}(M) = y,$$

where $\|M\|_* = \text{trace}(\sqrt{M^* M}) = \|\sigma(M)\|_1$.
Then in the noiseless case, to guarantee perfect recovery of any rank-r matrix X, we need only $m = 16nr$ measurements.

Moral of the story:

Practically, we need only $m = 16nr$ measurements.

Nuclear norm optimization

Theorem [Oymak-Hassibi]

Let $\mathcal{A}$ be a Gaussian linear operator and set,

$$\hat{X} = \underset{M}{\text{argmin}} \|M\|_* \quad \text{such that} \quad \mathcal{A}(M) = y,$$

where $\|M\|_* = \text{trace}(\sqrt{M^*M}) = \|\sigma(M)\|_1$.
Then in the noiseless case, to guarantee perfect recovery of any rank-r matrix X, we need only $m = 16nr$ measurements.

Moral of the story:

Practically, we need only $m = 16nr$ measurements.

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	??

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	??

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	??

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	??

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	??

The missing gap

The unanswered question

How many measurements m are needed to guarantee exact recovery of a rank-r matrix X via the rank minimization method?

$$\hat{X} = \operatorname*{argmin}_{M} \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

Answering the question

Theorem [Eldar-N-Plan]

Let $r \leq n/2$. When $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is a Gaussian operator with $m \geq 4nr - 4r^2$, any rank-r (or less) matrix X is exactly recovered via rank minimization:

$$\hat{X} = \operatorname*{argmin}_M \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \text{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

The gaps

Problem:	CS	MR
Theoretical	$\min \lVert f \rVert_0$	$\min \text{rank}(X)$
Practical	$\min \lVert f \rVert_1$	$\min \lVert X \rVert_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

Proof Sketch

$$\hat{X} = \operatorname*{argmin}_{M} \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee null($\mathcal{A}$) is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

Proof Sketch

$$\hat{X} = \operatorname*{argmin}_{M} \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n \times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee $\operatorname{null}(\mathcal{A})$ is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

Proof Sketch

$$\hat{X} = \underset{M}{\operatorname{argmin}}\, \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n \times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee null($\mathcal{A}$) is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

Proof Sketch

$$\hat{X} = \operatorname*{argmin}_M \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n \times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee null($\mathcal{A}$) is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

Proof Sketch

$$\hat{X} = \operatorname*{argmin}_M \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee $\operatorname{null}(\mathcal{A})$ is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

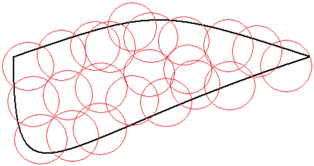
Proof Sketch

$$\hat{X} = \operatorname*{argmin}_{M} \operatorname{rank}(M) \quad \text{such that} \quad \mathcal{A}(M) = y$$

- Success of rank minimization is equivalent to asking that no rank-$2r$ or less matrix resides in the kernel of $\mathcal{A}$.
- Set $\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \operatorname{rank}(X) = 2r, \|X\|_F = 1\}$.
- $\mathcal{R}$ is a smooth manifold of $4nr - 4r^2 - 1$ dimensions.
- Step 1: Compute how large m must be to guarantee null$(\mathcal{A})$ is disjoint from $\mathcal{R}$.
- Step 2: Repeat for smaller values of the rank.

Covering Numbers

- For a set B, norm $\|\cdot\|$, and value ε, we define $N(B, \|\cdot\|, \varepsilon)$ to be the smallest number of $\|\cdot\|$-balls of radius ε whose union contains B.

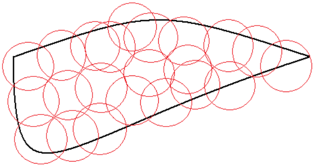


- The covering itself is called an ε-net.
- Euclidean covering numbers are well-known:

$$N(B_2^d, \|\cdot\|_2, \varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^d$$

Covering Numbers

- For a set B, norm $\|\cdot\|$, and value ε, we define $N(B, \|\cdot\|, \varepsilon)$ to be the smallest number of $\|\cdot\|$-balls of radius ε whose union contains B.

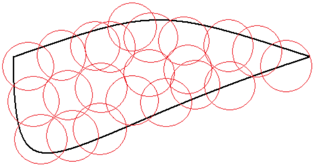


- The covering itself is called an ε-net.
- Euclidean covering numbers are well-known:

$$N(B_2^d, \|\cdot\|_2, \varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^d$$

Covering Numbers

- For a set B, norm $\|\cdot\|$, and value ε, we define $N(B, \|\cdot\|, \varepsilon)$ to be the smallest number of $\|\cdot\|$-balls of radius ε whose union contains B.



- The covering itself is called an ε-net.
- Euclidean covering numbers are well-known:

$$N(B_2^d, \|\cdot\|_2, \varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^d$$

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- Since $\mathcal{R}$ is a smooth manifold, there is a countable partition $\{\mathcal{V}_i\}$ of closed sets with C^1-diffeomorphisms $\phi_i : \mathcal{V}_i \to B_2^d$.
- Fix i. ϕ^{-1} is Lipschitz: $\|\phi^{-1}(x) - \phi^{-1}(y)\|_F \leq L\|x - y\|_2$.
- Let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d of size at most $\left(\frac{3L}{\varepsilon}\right)^d$.
- Then $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$.

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- Since $\mathcal{R}$ is a smooth manifold, there is a countable partition $\{\mathcal{V}_i\}$ of closed sets with C^1-diffeomorphisms $\phi_i : \mathcal{V}_i \to B_2^d$.
- Fix i. ϕ^{-1} is Lipschitz: $\|\phi^{-1}(x) - \phi^{-1}(y)\|_F \leq L\|x - y\|_2$.
- Let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d of size at most $\left(\frac{3L}{\varepsilon}\right)^d$.
- Then $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$.

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n \times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- Since $\mathcal{R}$ is a smooth manifold, there is a countable partition $\{\mathcal{V}_i\}$ of closed sets with C^1-diffeomorphisms $\phi_i : \mathcal{V}_i \to B_2^d$.
- Fix i. ϕ^{-1} is Lipschitz: $\|\phi^{-1}(x) - \phi^{-1}(y)\|_F \leq L\|x - y\|_2$.
- Let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d of size at most $\left(\frac{3L}{\varepsilon}\right)^d$.
- Then $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$.

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- Since $\mathcal{R}$ is a smooth manifold, there is a countable partition $\{\mathcal{V}_i\}$ of closed sets with C^1-diffeomorphisms $\phi_i : \mathcal{V}_i \to B_2^d$.
- Fix i. ϕ^{-1} is Lipschitz: $\|\phi^{-1}(x) - \phi^{-1}(y)\|_F \leq L\|x - y\|_2$.
- Let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d of size at most $\left(\frac{3L}{\varepsilon}\right)^d$.
- Then $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$.

Proof Sketch

$$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\},\ d = 4nr - 4r^2 - 1$$

- Since $\mathcal{R}$ is a smooth manifold, there is a countable partition $\{\mathcal{V}_i\}$ of closed sets with C^1-diffeomorphisms $\phi_i : \mathcal{V}_i \to B_2^d$.
- Fix i. ϕ^{-1} is Lipschitz: $\|\phi^{-1}(x) - \phi^{-1}(y)\|_F \leq L\|x - y\|_2$.
- Let $\overline{B_2^d}$ be an (ε/L)-net for B_2^d of size at most $\left(\frac{3L}{\varepsilon}\right)^d$.
- Then $\overline{\mathcal{V}}$ defined by $\overline{\mathcal{V}} = \phi^{-1}(\overline{B_2^d})$ is an ε-net for $\mathcal{V}$.

Proof Sketch

- Since $\overline{\mathcal{V}}$ is an ε-net,

$$\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty \geq \min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty - \varepsilon \cdot \|\mathcal{A}\|_{F\to\infty}.$$

- Therefore,

$$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \leq \mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty \leq \varepsilon \log(1/\varepsilon)\right)$$
$$\leq \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty - \varepsilon \cdot \|\mathcal{A}\|_{F\to\infty} \leq \varepsilon \log(1/\varepsilon)\right).$$

Proof Sketch

- Since $\overline{\mathcal{V}}$ is an ε-net,

$$\inf_{X \in \mathcal{V}} \|\mathcal{A}(X)\|_\infty \geq \min_{\overline{X} \in \overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty - \varepsilon \cdot \|\mathcal{A}\|_{F \to \infty}.$$

- Therefore,

$$\mathbb{P}\Big(\inf_{X \in \mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0 \Big) \leq \mathbb{P}\left(\inf_{X \in \mathcal{V}} \|\mathcal{A}(X)\|_\infty \leq \varepsilon \log(1/\varepsilon) \right)$$
$$\leq \mathbb{P}\left(\min_{\overline{X} \in \overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty - \varepsilon \cdot \|\mathcal{A}\|_{F \to \infty} \leq \varepsilon \log(1/\varepsilon) \right).$$

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}, \; d = 4nr - 4r^2 - 1$

- $$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \lesssim \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right)$$
- $\leq \left(\frac{3L}{\varepsilon}\right)^d \cdot \prod_{i=1}^m \left(\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right)\right)$, (where z_i is an entry of $\mathcal{A}(\overline{X})$)
- $\lesssim \varepsilon^{m-d} \cdot (\log(1/\varepsilon))^m$
- So we choose $m = d + 1$ and take $\varepsilon \to 0$!

Proof Sketch

$$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}, \; d = 4nr - 4r^2 - 1$$

- $$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \lesssim \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty \le 2\varepsilon \log(1/\varepsilon)\right)$$
- $\le \left(\frac{3L}{\varepsilon}\right)^d \cdot \prod_{i=1}^m \left(\mathbb{P}\left(|z_i| \le 2\varepsilon\log(1/\varepsilon)\right)\right)$, (where z_i is an entry of $\mathcal{A}(\overline{X})$)
- $\lesssim \varepsilon^{m-d} \cdot (\log(1/\varepsilon))^m$
- So we choose $m = d + 1$ and take $\varepsilon \to 0$!

Proof Sketch

$$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}, \; d = 4nr - 4r^2 - 1$$

- $$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \lesssim \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right)$$
- $\leq \left(\frac{3L}{\varepsilon}\right)^d \cdot \prod_{i=1}^m \left(\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right)\right)$, (where z_i is an entry of $\mathcal{A}(\overline{X})$)
- $\lesssim \varepsilon^{m-d} \cdot (\log(1/\varepsilon))^m$
- So we choose $m = d + 1$ and take $\varepsilon \to 0$!

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- $$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \lesssim \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty \leq 2\varepsilon\log(1/\varepsilon)\right)$$
- $\leq \left(\frac{3L}{\varepsilon}\right)^d \cdot \prod_{i=1}^m \left(\mathbb{P}\left(|z_i| \leq 2\varepsilon\log(1/\varepsilon)\right)\right)$, (where z_i is an entry of $\mathcal{A}(\overline{X})$)
- $\lesssim \varepsilon^{m-d} \cdot (\log(1/\varepsilon))^m$
- So we choose $m = d + 1$ and take $\varepsilon \to 0$!

Proof Sketch

$\mathcal{R} = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = 2r, \|X\|_F = 1\}$, $d = 4nr - 4r^2 - 1$

- $$\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) \lesssim \mathbb{P}\left(\min_{\overline{X}\in\overline{\mathcal{V}}} \|\mathcal{A}(\overline{X})\|_\infty \le 2\varepsilon\log(1/\varepsilon)\right)$$
- $\le \left(\frac{3L}{\varepsilon}\right)^d \cdot \prod_{i=1}^m \left(\mathbb{P}\left(|z_i| \le 2\varepsilon\log(1/\varepsilon)\right)\right)$, (where z_i is an entry of $\mathcal{A}(\overline{X})$)
- $\lesssim \varepsilon^{m-d} \cdot (\log(1/\varepsilon))^m$
- So we choose $m = d + 1$ and take $\varepsilon \to 0$!

Proof Sketch

- Therefore, $\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$ when $m = d + 1 = 4nr - 4r^2$.
- Apply a union bound over all (countably many) $\mathcal{V}_i$: $\mathbb{P}\left(\inf_{X\in\mathcal{R}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$
- Applying this same argument for all ranks less than $2r$ proves our result.

Proof Sketch

- Therefore, $\mathbb{P}\left(\inf_{X \in \mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$ when $m = d + 1 = 4nr - 4r^2$.
- Apply a union bound over all (countably many) $\mathcal{V}_i$: $\mathbb{P}\left(\inf_{X \in \mathcal{R}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$
- Applying this same argument for all ranks less than $2r$ proves our result.

Proof Sketch

- Therefore, $\mathbb{P}\left(\inf_{X\in\mathcal{V}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$ when $m = d + 1 = 4nr - 4r^2$.
- Apply a union bound over all (countably many) $\mathcal{V}_i$: $\mathbb{P}\left(\inf_{X\in\mathcal{R}} \|\mathcal{A}(X)\|_\infty = 0\right) = 0$
- Applying this same argument for all ranks less than $2r$ proves our result.

The gaps

Problem:	CS	MR
Theoretical	$\min \|f\|_0$	$\min \operatorname{rank}(X)$
Practical	$\min \|f\|_1$	$\min \|X\|_*$
m for Practical	$m \gtrsim s \log n$	$m \geq 16nr$
m for Theoretical	$m \geq 2s$	$4nr - 4r^2$

Thank you!

E-mail:

- dneedell@cmc.edu

Web:

- www.cmc.edu/pages/faculty/DNeedell

References:

- E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 59(8):12071223, 2006.
- Oymak, S. and Hassibi, B., Tight Recovery Thresholds and Robustness Analysis for Matrix Rank Minimization, Submitted.
- Y. C. Eldar, D. Needell and Y. Plan. Uniqueness Conditions For Low-Rank Matrix Recovery. Submitted.