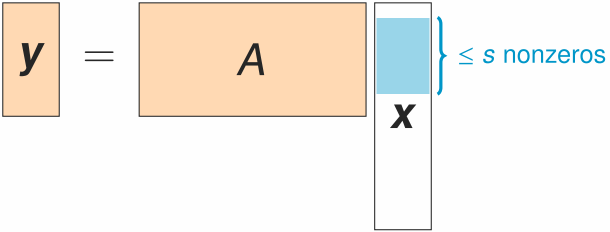
Nonconvex Compressed Sensing with the Sum-of-Squares Method

Tasuku Soma (Univ. Tokyo)

Joint work with:
Yuichi Yoshida (NII&PFI)

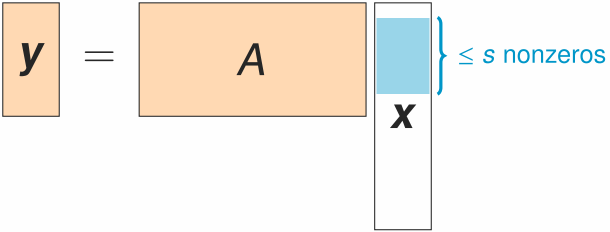
Compressed Sensing

Given: $A \in \mathbb{R}^{m \times n}$ $(m \ll n)$ and $\boldsymbol{y} = A\boldsymbol{x}$,
Task: Estimate the original **sparse** signal $\boldsymbol{x} \in \mathbb{R}^n$.



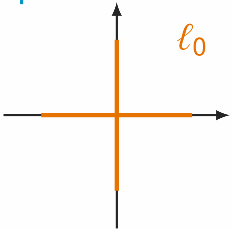
Compressed Sensing

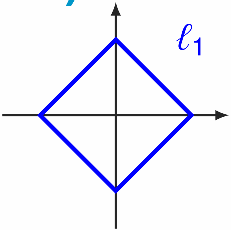
Given: $A \in \mathbb{R}^{m \times n}$ $(m \ll n)$ and $\boldsymbol{y} = A\boldsymbol{x}$,
Task: Estimate the original **sparse** signal $\boldsymbol{x} \in \mathbb{R}^n$.



Applications:
Image Processing, Statistics, Machine Learning...

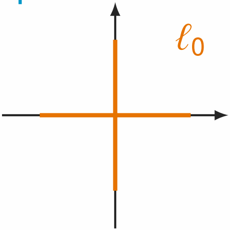
ℓ_1 Minimization (Basis Pursuit)

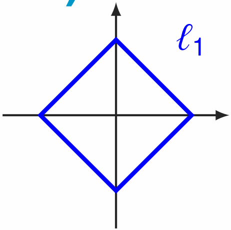




$$\min \quad \|\boldsymbol{z}\|_1 \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

ℓ_1 Minimization (Basis Pursuit)



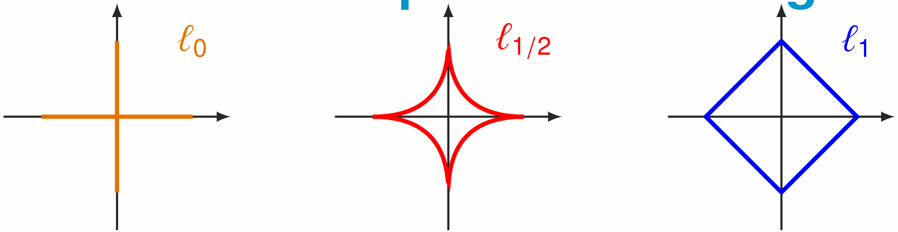


$$\min \quad \|\boldsymbol{z}\|_1 \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

- Convex relaxation for ℓ_0 minimization
- For a subgaussian A with $m = \Omega(s \log \frac{n}{s})$, ℓ_1 minimization reconstructs $\boldsymbol{x}$.
  [Candès-Romberg-Tao '06, Donoho '06]

s: sparsity of $\boldsymbol{x}$ (maybe unknown)

Nonconvex Compressed Sensing

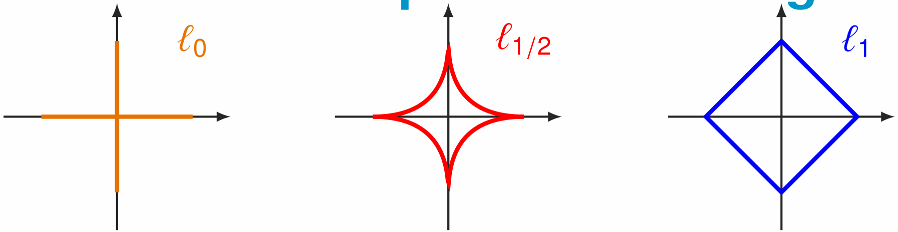


ℓ_q minimization ($0 < q \leq 1$):

[Laska-Davenport-Baraniuk '09,Cherian-Sra-Papanikolopoulos '11]

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

Nonconvex Compressed Sensing



ℓ_q minimization ($0 < q \leq 1$):

[Laska-Davenport-Baraniuk ’09,Cherian-Sra-Papanikolopoulos ’11]

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

- Requires fewer samples than ℓ_1 minimization
- Recovers arbitrary sparse signals as $q \to 0$
- **Nonconvex** Optimization!

Stable Signal Recovery

$\boldsymbol{x}$ needs not to be sparse but close to a sparse signal.

Stable Signal Recovery

$\boldsymbol{x}$ needs not to be sparse but close to a sparse signal.

$A \in \mathbb{R}^{m\times n}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ are **ℓ_p-stable recovery**

$$\iff \|\Delta(A\boldsymbol{x}) - \boldsymbol{x}\|_p \leq O(\sigma_s(\boldsymbol{x})_p)$$

for any $\boldsymbol{x} \in \mathbb{R}^n$.

Stable Signal Recovery

$\boldsymbol{x}$ needs not to be sparse but close to a sparse signal.

$A \in \mathbb{R}^{m \times n}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ are **ℓ_p-stable recovery**

$$\iff \|\Delta(A\boldsymbol{x}) - \boldsymbol{x}\|_p \le O(\sigma_s(\boldsymbol{x})_p)$$

for any $\boldsymbol{x} \in \mathbb{R}^n$.

ℓ_p distance to s-sparse vector

Stable Signal Recovery

$\boldsymbol{x}$ needs not to be sparse but close to a sparse signal.

$A \in \mathbb{R}^{m \times n}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ are **ℓ_p-stable recovery**

$$\iff \|\Delta(A\boldsymbol{x}) - \boldsymbol{x}\|_p \leq O(\sigma_s(\boldsymbol{x})_p)$$

for any $\boldsymbol{x} \in \mathbb{R}^n$.

ℓ_p distance to s-sparse vector

- Gaussian matrix with $m = \Omega(s \log \frac{n}{s})$ and ℓ_1 minimization are ℓ_1-stable [Candès-Romberg-Tao '06,Candès '08]
- Gaussian A and ℓ_q minimization are ℓ_q-stable $(0 < q \leq 1)$ [Cohen-Dahmen-DeVore '09]
- **Smaller q yields better bound when noise is sparse.**

Our Result

$A_{ij} \sim \{\pm 1/\sqrt{m}\}$

Theorem

For $\|\mathbf{x}\|_\infty \leq 1$ *and fixed* $q = 2^{-k}$*, there exist* $A \in \mathbb{R}^{m \times n}$ *and a* ***polytime*** *algorithm* $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ *s.t.*

$$\|\Delta(A\mathbf{x}) - \mathbf{x}\|_q \leq O(\sigma_s(\mathbf{x})_q) + \varepsilon,$$

provided that $m = \Omega(s^{2/q} \log n)$*.*

- (Nearly) ℓ_q-stable recovery
- #samples >> $O(s \log(n/s))$ (**Sample Complexity Trade Off**)
- Use of **SoS Method** and **Ellipsoid Method**

High Level Picture

Naive Idea: Reduce ℓ_q minimization to **polynomial optimization**

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

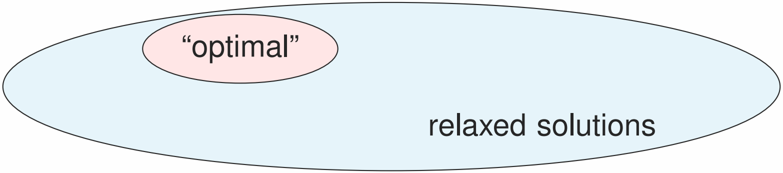
→ Does the **SoS method** find an "optimal solution"?

High Level Picture

Naive Idea: Reduce ℓ_q minimization to **polynomial optimization**

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

→ Does the **SoS method** find an "optimal solution"? ✕No

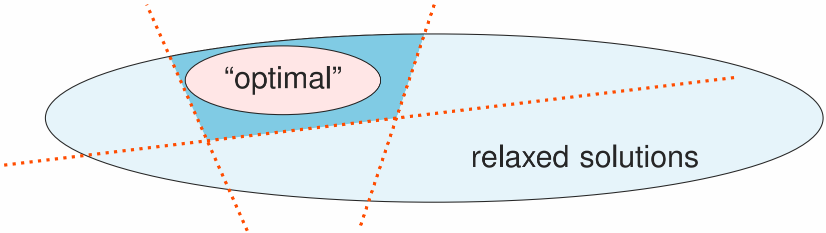


High Level Picture

Naive Idea: Reduce ℓ_q minimization to **polynomial optimization**

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

$\rightarrow$ Does the **SoS method** find an "optimal solution"? ✕No



Idea: Add cuts to the SoS method

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{s.t.} \quad A\boldsymbol{z} = \boldsymbol{y}, \ \textbf{Additional Constraints}$$

SoS Method [Lasserre '06, Parrilo '00, Nesterov '00, Shor '87]

Polynomial Optimization: $f, g_1, \dots, g_i \in \mathbb{R}[\boldsymbol{z}]$: polynomials

$$\begin{aligned} \min_{\boldsymbol{z}} \quad & f(\boldsymbol{z}) \\ \text{sub. to} \quad & g_i(\boldsymbol{z}) = 0 \quad (i = 1, \dots, m) \end{aligned}$$

SoS Method [Lasserre '06, Parrilo '00, Nesterov '00, Shor '87]

Polynomial Optimization: $f, g_1, \ldots, g_i \in \mathbb{R}[\boldsymbol{z}]$: polynomials

$$\begin{aligned} \min_{\boldsymbol{z}} \quad & f(\boldsymbol{z}) \\ \text{sub. to} \quad & g_i(\boldsymbol{z}) = 0 \quad (i = 1, \ldots, m) \end{aligned}$$

SoS Relaxation (of degree d):

$$\begin{aligned} \min_{\widetilde{\mathbf{E}}} \quad & \widetilde{\mathbf{E}}[f(\boldsymbol{z})] \\ \text{sub. to} \quad & \widetilde{\mathbf{E}} : \mathbb{R}[\boldsymbol{z}]_d \to \mathbb{R}, \text{linear operator "pseudoexpectation"} \\ & \widetilde{\mathbf{E}}[1] = 1 \\ & \widetilde{\mathbf{E}}[p(\boldsymbol{z})^2] \geq 0 \qquad (p \in \mathbb{R}[\boldsymbol{z}] : \deg(p) \leq d/2) \\ & \widetilde{\mathbf{E}}[g_i(\boldsymbol{z})p(\boldsymbol{z})] = 0 \quad (p \in \mathbb{R}[\boldsymbol{z}] : \deg(g_i p) \leq d, i = 1, \ldots, m) \end{aligned}$$

Facts on SoS Method

- The SoS Relaxation (of degree d) reduces to **Semidefinite Programming (SDP)** with $n^{O(d)}$-size matrix.

Facts on SoS Method

- The SoS Relaxation (of degree d) reduces to **Semidefinite Programming (SDP)** with $n^{O(d)}$-size matrix.

- Dual View: **SoS Proof System**
  Any (low-degree) "proof" in SoS proof system yields an algorithm via the SoS method.

Facts on SoS Method

- The SoS Relaxation (of degree d) reduces to **Semidefinite Programming (SDP)** with $n^{O(d)}$-size matrix.

- Dual View: **SoS Proof System**
  Any (low-degree) "proof" in SoS proof system yields an algorithm via the SoS method.

- Very Powerful Tool in Computer Science:
  - Subexponential Alg. for UG [Arora-Barak-Steurer'10]
  - Planted Sparse Vector [Barak-Kelner-Steurer'14]
  - Sparse PCA [Ma-Wigderson'14]

Outline

ℓ_q-stability proof

A is a Rademacher matrix

↓

A has small coherence

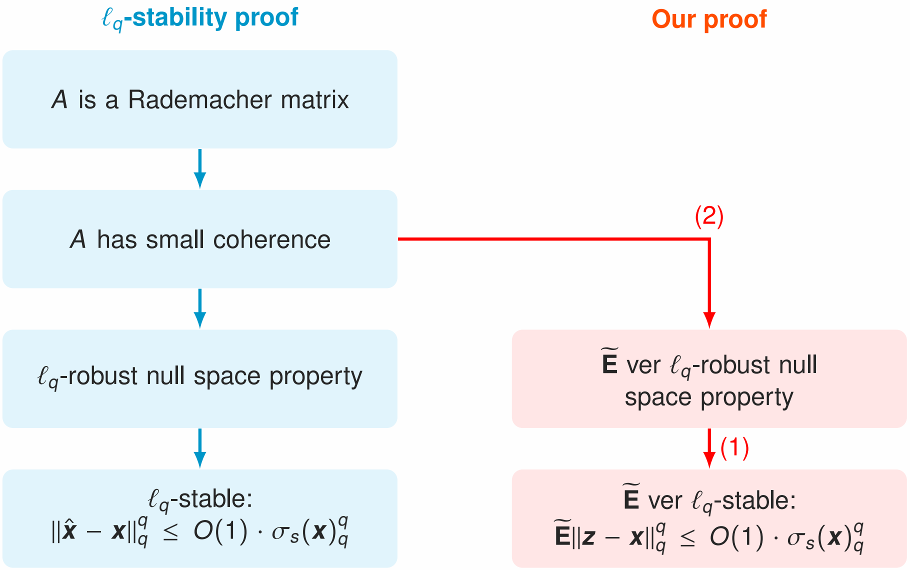
↓

ℓ_q-robust null space property

↓

ℓ_q-stable:
$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(1) \cdot \sigma_s(\boldsymbol{x})_q^q$

Outline


ℓ_q-stability proof
Our proof
A is a Rademacher matrix
A has small coherence
(2)
ℓ_q-robust null space property
$\widetilde{\mathbf{E}}$ ver ℓ_q-robust null space property
(1)
ℓ_q-stable: $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(1) \cdot \sigma_s(\boldsymbol{x})_q^q$
$\widetilde{\mathbf{E}}$ ver ℓ_q-stable: $\widetilde{\mathbf{E}}\|\boldsymbol{z} - \boldsymbol{x}\|_q^q \leq O(1) \cdot \sigma_s(\boldsymbol{x})_q^q$

Basic Idea

Formulate ℓ_q minimization as polynomial optimization:

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

Note: $|z(i)|^q$ is not a polynomial, but representable by **lifting**;

$$\left(\sqrt{|z(i)|}\right)^4 = z(i)^2, \quad \sqrt{|z(i)|} \geq 0$$

Basic Idea

Formulate ℓ_q minimization as polynomial optimization:

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

Note: $|z(i)|^q$ is not a polynomial, but representable by **lifting**;

$$\left(\sqrt{|z(i)|}\right)^4 = z(i)^2, \quad \sqrt{|z(i)|} \geq 0$$

✗Solutions of SoS method do not satisfy triangle inequalities:

$$\widetilde{\mathbf{E}}\|\boldsymbol{z}+\boldsymbol{x}\|_q^q \not\leq \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q + \|\boldsymbol{x}\|_q^q$$

Basic Idea

Formulate ℓ_q minimization as polynomial optimization:

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{sub. to} \quad A\boldsymbol{z} = \boldsymbol{y}$$

Note: $|z(i)|^q$ is not a polynomial, but representable by **lifting**;

$$\left(\sqrt{|z(i)|}\right)^4 = z(i)^2, \quad \sqrt{|z(i)|} \geq 0$$

✕Solutions of SoS method do not satisfy triangle inequalities:

$$\widetilde{\mathbf{E}}\|\boldsymbol{z} + \boldsymbol{x}\|_q^q \not\leq \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q + \|\boldsymbol{x}\|_q^q$$

Add Valid Constraints!

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{s.t.} \quad A\boldsymbol{z} = \boldsymbol{y}, \ \boxed{\textbf{Valid Constraints}}$$

Triangle Inequalities

$$\|\boldsymbol{z} + \boldsymbol{x}\|_q^q \le \|\boldsymbol{z}\|_q^q + \|\boldsymbol{x}\|_q^q$$

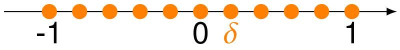
We have to add $|z(i) + x(i)|^q$, but do not know $x(i)$.

Triangle Inequalities

$$\|\boldsymbol{z} + \boldsymbol{x}\|_q^q \leq \|\boldsymbol{z}\|_q^q + \|\boldsymbol{x}\|_q^q$$

We have to add $|z(i) + x(i)|^q$, but do not know $x(i)$.

Idea: Using **Grid**
L: set of multiples of δ in $[-1, 1]$.



- new variable for $|z(i) - b|^q$ ($b \in L$)
- triangle inequalities for
  $|z(i) - b|^q$, $|z(i) - b'|^q$, and $|b - b'|^q$ ($b, b' \in L$)

Robust ℓ_q Minimization

Instead of $\boldsymbol{x}$, we will find $\boldsymbol{x}^L \in L^n$ closest to $\boldsymbol{x}$.

Robust ℓ_q Minimization

$\eta = \sigma_{\max}(A)\sqrt{s}\delta$

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{s.t.} \quad \|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 \leq \eta^2$$

Robust ℓ_q Minimization

Instead of $\boldsymbol{x}$, we will find $\boldsymbol{x}^L \in L^n$ closest to $\boldsymbol{x}$.

Robust ℓ_q Minimization

$\eta = \sigma_{\max}(A)\sqrt{s}\delta$

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{s.t.} \quad \|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 \leq \eta^2$$

ℓ_q Robust Null Space Property

$$\|\boldsymbol{v}_S\|_q^q \leq \rho\|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau\|A\boldsymbol{v}\|_2^q$$

for any $\boldsymbol{v}$ and $S \subseteq [n]$ with $|S| \leq s$.

Robust ℓ_q Minimization

Instead of $\boldsymbol{x}$, we will find $\boldsymbol{x}^L \in L^n$ closest to $\boldsymbol{x}$.

Robust ℓ_q Minimization

$\eta = \sigma_{\max}(A)\sqrt{s}\delta$

$$\min \quad \|\boldsymbol{z}\|_q^q \qquad \text{s.t.} \quad \|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 \leq \eta^2$$

ℓ_q Pseudo Robust Null Space Property (ℓ_q-PRNSP)

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q \leq \rho\,\widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}$$

for any $\boldsymbol{v} = \boldsymbol{z} - \boldsymbol{b}$ ($\boldsymbol{b} \in L^n$) and $S \subseteq [n]$ with $|S| \leq s$.

(1) PRNSP $\Longrightarrow$ Stable Recovery

Theorem

If $\widetilde{\mathbf{E}}$ satisfies ℓ_q-PRNSP, then

$$\widetilde{\mathbf{E}}\|\boldsymbol{z} - \boldsymbol{x}^L\|_q^q \leq \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q,$$

where $\boldsymbol{x}^L$ is the closest vector in L^n to $\boldsymbol{x}$.

Proof Idea:

A proof of stability only needs:

- ℓ_q^q triangle inequalities for $\boldsymbol{z} - \boldsymbol{x}^L$, $\boldsymbol{x}$ and $\boldsymbol{z} + \boldsymbol{x}^L$
- ℓ_2 triangle inequality

Rounding

Extract an actual vector $\hat{\boldsymbol{x}}$ from a pseudoexpectation $\widetilde{\mathbf{E}}$.

$$\hat{x}(i) := \operatorname*{argmin}_{b \in L} \widetilde{\mathbf{E}} |z(i) - b|^q \qquad (i = 1, \ldots, n)$$

Theorem

If $\widetilde{\mathbf{E}}$ satisfies PRNSP,

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^L\|_q^q \leq 2 \left[\frac{2(1+\rho)}{1-\rho} \sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho} \eta^q \right]$$

Outline

ℓ_q-stability proof

A is a Rademacher matrix

A has small coherence

ℓ_q-robust null space property

ℓ_q-stable:
$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(1) \cdot \sigma_s(\boldsymbol{x})_q^q$

Our proof

(2)

$\widetilde{\mathbf{E}}$ ver ℓ_q-robust null space property

(1)

$\widetilde{\mathbf{E}}$ ver ℓ_q-stable:
$\widetilde{\mathbf{E}}\|\boldsymbol{z} - \boldsymbol{x}\|_q^q \leq O(1) \cdot \sigma_s(\boldsymbol{x})_q^q$

Imposing PRNSP

How can we obtain $\widetilde{\mathbf{E}}$ satisfying PRNSP?

Idea: Follow known proofs for robust NSP!

- From Restricted Isometry Property (RIP) [Candès '08]
- **From Coherence** [Gribonval-Nielsen '03, Donoho-Elad '03]
- From Lossless Expander [Berinde et al. '08]

Coherence

The **coherence** of a matrix $A = [\boldsymbol{a}_1 \dots \boldsymbol{a}_n]$ is

$$\mu = \max_{i \neq j} \frac{|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|}{||\boldsymbol{a}_i||_2 ||\boldsymbol{a}_j||_2}$$

Facts:

- If $\mu^q < \frac{1}{2s}$, ℓ_q Robust NSP holds.
- If A is a Rademacher matrix with $m = O(s^{2/q} \log n)$, then $\mu^q < \frac{1}{2s}$ w.h.p.

Small Coherence $\Longrightarrow$ PRNSP

Issue: Naive import needs **exponentially many** variables and constraints!

Lemma

If A is a Rademacher matrix,

- *additinal variables are polynomially many*
- *additional constraints have a* ***separation oracle***

Thus ellipsoid methods find $\widetilde{\mathbf{E}}$ with PRNSP.

Our Result

$A_{ij} \sim \{\pm 1/\sqrt{m}\}$

Theorem

For $\|\mathbf{x}\|_\infty \leq 1$ and fixed $q = 2^{-k}$, there exist $A \in \mathbb{R}^{m \times n}$ and a ***polytime*** *algorithm $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ s.t.*

$$\|\Delta(A\mathbf{x}) - \mathbf{x}\|_q \leq O(\sigma_s(\mathbf{x})_q) + \varepsilon,$$

provided that $m = \Omega(s^{2/q} \log n)$.

- (Nearly) ℓ_q-stable recovery
- #samples >> $O(s \log(n/s))$ (**Sample Complexity Trade Off**)
- Use of **SoS Method** and **Ellipsoid Method**

Putting Things Together

Using a Rademacher matrix yields PRNSP:

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q \leq O(1) \cdot \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + O(s) \cdot \left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}$$

Putting Things Together

Using a Rademacher matrix yields PRNSP:

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q \leq O(1) \cdot \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + O(s) \cdot \left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}$$

This guarantees:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(\sigma_s(\boldsymbol{x}^L)_q^q) + O(s) \cdot \eta^q$$

Putting Things Together

Using a Rademacher matrix yields PRNSP:

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q \leq O(1) \cdot \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + O(s) \cdot \left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}$$

This guarantees:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(\sigma_s(\boldsymbol{x}^L)_q^q) + O(s) \cdot \eta^q$$

Theorem

If we take δ small, then the rounded vector $\hat{\boldsymbol{x}}$ satisfies

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_q^q \leq O(\sigma_s(\boldsymbol{x})_q^q) + \varepsilon.$$

(pf) $\eta = \sigma_{\max}(A)\sqrt{s}\delta$ and $\sigma_{\max}(A) = \widetilde{O}(n/m)$ □