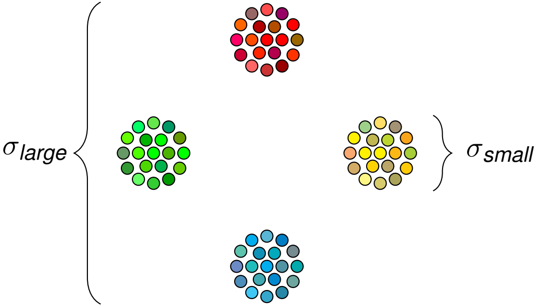
Fast RIP matrices with fewer rows

Jelani Nelson **Eric Price** Mary Wootters

Princeton MIT Michigan

2013-04-05



Outline

1 Introduction

- Compressive sensing
- Johnson Lindenstrauss Transforms
- Our result

Outline

1. Introduction
    - Compressive sensing
    - Johnson Lindenstrauss Transforms
    - Our result
2. Concentration of measure: a toolbox
    - Overview
    - Symmetrization
    - Gaussian Processes
    - Lipschitz Concentration

Outline

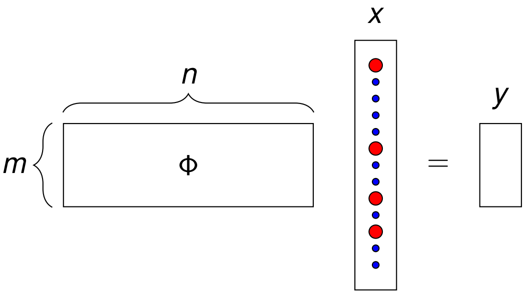
Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result
2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration
3. Proof
   - Overview
   - Covering Number
4. Conclusion

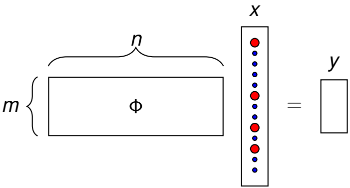
Outline

Compressive Sensing

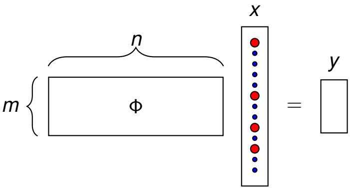
Given: A few linear measurements of an (approximately) k-sparse vector $x \in \mathbb{R}^n$.
Goal: Recover x (approximately).



Compressive Sensing Algorithms: Two Classes



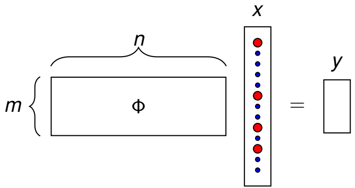
Compressive Sensing Algorithms: Two Classes



Structure-aware

Recovery algorithm
tied to matrix structure
(e.g. Count-Sketch)

Compressive Sensing Algorithms: Two Classes



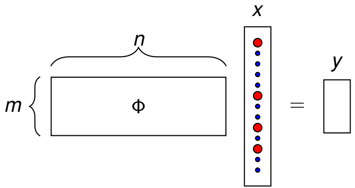
Structure-aware

Recovery algorithm
tied to matrix structure
(e.g. Count-Sketch)

Structure-oblivious

Recovery algorithms
just multiply by Φ, Φ^T
(e.g. L1 minimization)

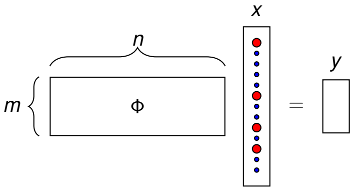
Compressive Sensing Algorithms: Two Classes



Structure-aware	Structure-oblivious
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)	Recovery algorithms just multiply by Φ, Φ^T (e.g. L1 minimization)

Often:
- Faster
- Sparse matrices
- Less robust

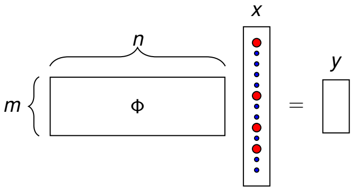
Compressive Sensing Algorithms: Two Classes



Structure-aware	Structure-oblivious
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)	Recovery algorithms just multiply by Φ, Φ^T (e.g. L1 minimization)

	Faster	Slower
Often:	Sparse matrices	Dense matrices
	Less robust	More robust

Compressive Sensing Algorithms: Two Classes



Structure-aware	Structure-oblivious
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)	Recovery algorithms just multiply by Φ, Φ^T (e.g. L1 minimization)

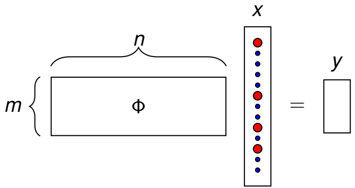
	Faster	Slower
Often:	Sparse matrices	Dense matrices
	Less robust	More robust

↑

Yesterday:

Fourier → sparse

Compressive Sensing Algorithms: Two Classes



Structure-aware	Structure-oblivious
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)	Recovery algorithms just multiply by Φ, Φ^T (e.g. L1 minimization)

	Faster	Slower
Often:	Sparse matrices	Dense matrices
	Less robust	More robust
	↑	↑
	Yesterday: Fourier → sparse	Today

Algorithms for compressive sensing

- Goal: recover approximately k-sparse x from $y = \Phi x$.

Algorithms for compressive sensing

- Goal: recover approximately k-sparse x from $y = \Phi x$.
- A lot of people use convex optimization:

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t. } \Phi x &= y \end{aligned}$$

Algorithms for compressive sensing

- Goal: recover approximately k-sparse x from $y = \Phi x$.
- A lot of people use convex optimization:

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t. } \Phi x &= y \end{aligned}$$

- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....

Algorithms for compressive sensing

- Goal: recover approximately k-sparse x from $y = \Phi x$.
- A lot of people use convex optimization:

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t. } \Phi x &= y \end{aligned}$$

- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....
- For all of these:
  - the time it takes to multiply by Φ or Φ^T is the bottleneck.

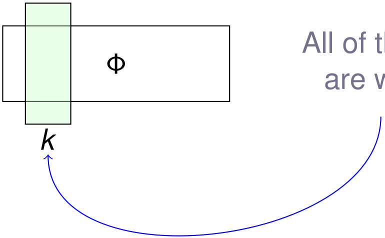
Algorithms for compressive sensing

- Goal: recover approximately k-sparse x from $y = \Phi x$.
- A lot of people use convex optimization:

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t. } \Phi x &= y \end{aligned}$$

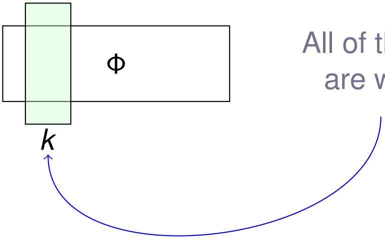
- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....
- For all of these:
  - the time it takes to multiply by Φ or Φ^T is the bottleneck.
  - the *Restricted Isometry Property* is a sufficient condition.

Restricted Isometry Property (RIP)



All of these submatrices are well conditioned.

Restricted Isometry Property (RIP)



All of these submatrices are well conditioned.

$$(1-\epsilon)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\epsilon)\|x\|_2^2$$

for all k-sparse $x \in \mathbb{R}^n$.

Goals

What properties should an RIP matrix have?

Goals

What properties should an RIP matrix have?

- Good compression: m small
  - Random Gaussian matrix: $\Theta(k \log(n/k))$ rows.

Goals

What properties should an RIP matrix have?

- Good compression: *m* small
  - Random Gaussian matrix: $\Theta(k \log n)$ rows.

* Talk will assume $n^{0.1} < k < n^{0.9}$, so $\log k \simeq \log n \simeq \log(n/k)$.

Goals

What properties should an RIP matrix have?

- Good compression: m small
  - Random Gaussian matrix: $\Theta(k \log n)$ rows.
- Fast multiplication:
  - Reconstruction dominated by $\log n$ multiplications by Φ, Φ^T.

* Talk will assume $n^{0.1} < k < n^{0.9}$, so $\log k \simeq \log n \simeq \log(n/k)$.

Goals

What properties should an RIP matrix have?

- Good compression: m small
    - Random Gaussian matrix: $\Theta(k \log n)$ rows.
- Fast multiplication:
    - Reconstruction dominated by $\log n$ multiplications by Φ, Φ^T.
    - Random Gaussian matrix: $\Theta(nk \log n)$ time.

* Talk will assume $n^{0.1} < k < n^{0.9}$, so $\log k \simeq \log n \simeq \log(n/k)$.

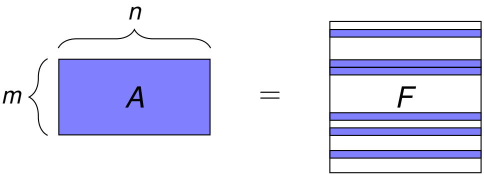
Goals

What properties should an RIP matrix have?

- Good compression: m small
  - Random Gaussian matrix: $\Theta(k \log n)$ rows.
- Fast multiplication:
  - Reconstruction dominated by $\log n$ multiplications by Φ, Φ^T.
  - Random Gaussian matrix: $\Theta(nk \log n)$ time.
- Goal: an RIP matrix with $O(n \log n)$ multiplication and small m.

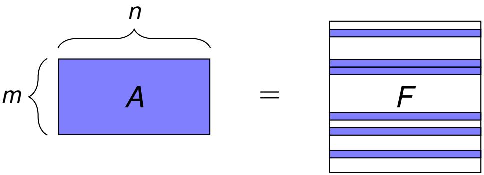
* Talk will assume $n^{0.1} < k < n^{0.9}$, so $\log k \simeq \log n \simeq \log(n/k)$.

An open question



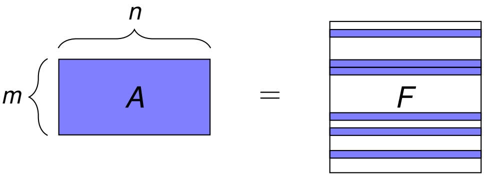
Let A contain random rows from a Fourier matrix.

An open question



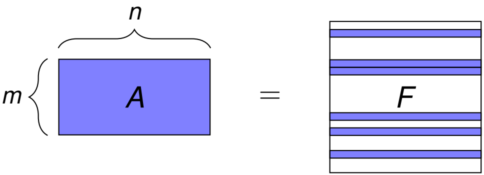
Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.

An open question



Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.
How many rows do you need to ensure that A has the RIP?

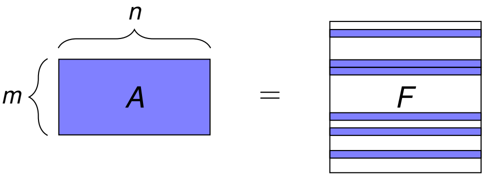
An open question



Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.
How many rows do you need to ensure that A has the RIP?

- $m = O(k \log^4 n)$ [CT06,RV08,CGV13].

An open question



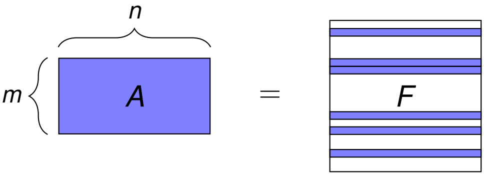
Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.
How many rows do you need to ensure that A has the RIP?

- $m = O(k \log^4 n)$ [CT06,RV08,CGV13].

Ideal:

- $m = O(k \log n)$.

An open question



Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.
How many rows do you need to ensure that A has the RIP?

- $m = O(k \log^4 n)$ [CT06,RV08,CGV13].

Ideal:

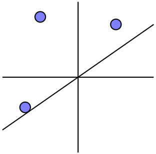
- $m = O(k \log n)$.

(Related: how about partial circulant matrices?)

- $m = O(k \log^4 n)$ [RRT12,KMR12].

Outline

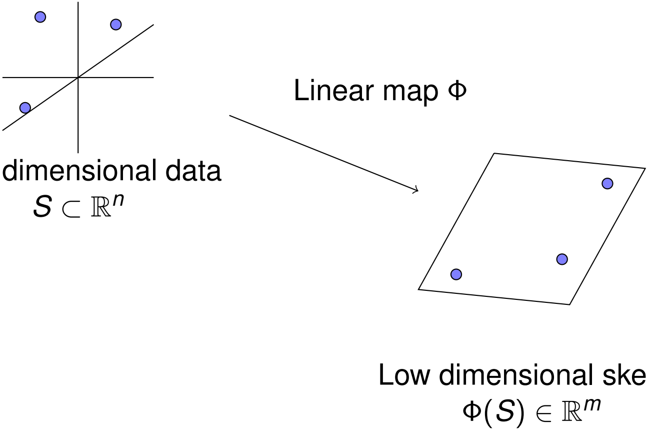
Another motivation: Johnson Lindenstrauss (JL) Transforms



High dimensional data

$S \subset \mathbb{R}^n$

Another motivation:
Johnson Lindenstrauss (JL) Transforms



Linear map Φ

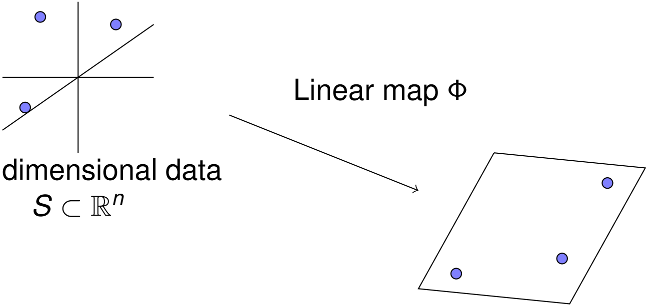
High dimensional data

$S \subset \mathbb{R}^n$

Low dimensional sketch

$\Phi(S) \in \mathbb{R}^m$

Another motivation: Johnson Lindenstrauss (JL) Transforms



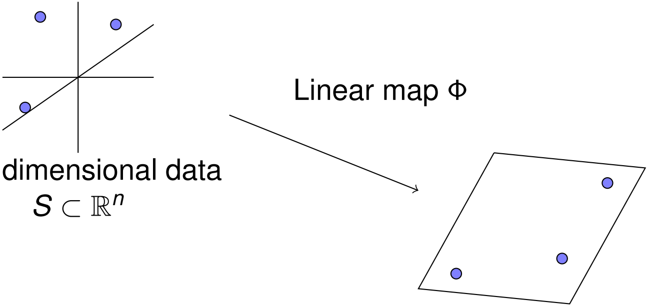
High dimensional data
$S \subset \mathbb{R}^n$

Linear map Φ

Low dimensional sketch
$\Phi(S) \in \mathbb{R}^m$

Φ preserves the geometry of S

Another motivation: Johnson Lindenstrauss (JL) Transforms



Linear map Φ

High dimensional data

$S \subset \mathbb{R}^n$

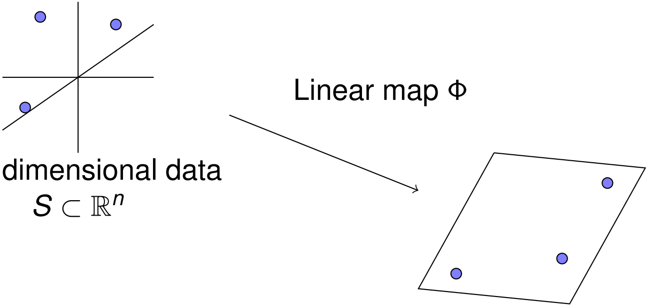
Low dimensional sketch

$\Phi(S) \in \mathbb{R}^m$

Φ preserves the geometry of S

$$(1-\epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\epsilon)\|x\|_2$$

Another motivation: Johnson Lindenstrauss (JL) Transforms



Linear map Φ

High dimensional data

$S \subset \mathbb{R}^n$

Low dimensional sketch

$\Phi(S) \in \mathbb{R}^m$

Φ preserves the geometry of S

$$(1-\epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\epsilon)\|x\|_2$$

$$\langle \Phi x, \Phi y\rangle = \langle x, y\rangle \pm \epsilon \|x\|_2 \|y\|_2$$

Johnson-Lindenstrauss Lemma

Theorem (variant of Johnson-Lindenstrauss '84)

Let $x \in \mathbb{R}^n$. A random Gaussian matrix Φ will have

$$(1-\epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\epsilon)\|x\|_2$$

with probability $1-\delta$, so long as

$$m \gtrsim \frac{1}{\epsilon^2}\log(1/\delta)$$

Johnson-Lindenstrauss Lemma

Theorem (variant of Johnson-Lindenstrauss '84)

Let $x \in \mathbb{R}^n$*. A random Gaussian matrix* Φ *will have*

$$(1-\epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\epsilon)\|x\|_2$$

with probability $1-\delta$*, so long as*

$$m \gtrsim \frac{1}{\epsilon^2}\log(1/\delta)$$

Set $\delta = 1/2^k$: embed 2^k points into $O(k)$ dimensions.

What do we want in a JL matrix?

What do we want in a JL matrix?

- Target dimension should be small (close to $\frac{1}{\epsilon^2}k$ for 2^k points).

What do we want in a JL matrix?

- Target dimension should be small (close to $\frac{1}{\epsilon^2}k$ for 2^k points).
- Fast multiplication.
  - Approximate numerical algebra problems (e.g., linear regression, low-rank approximation)
  - k-means clustering

How do we get a JL matrix?

How do we get a JL matrix?

- Gaussians
  - Dimension $O(\frac{1}{\epsilon^2}k)$.

How do we get a JL matrix?

- Gaussians
  - Dimension $O(\frac{1}{\epsilon^2}k)$.
  - $\frac{1}{\epsilon^2}nk$ multiplication time.

How do we get a JL matrix?

- Gaussians
  - Dimension $O(\frac{1}{\epsilon^2}k)$.
  - $\frac{1}{\epsilon^2}nk$ multiplication time.
- Best way known for *fast JL*: by [Krahmer-Ward '11], RIP ⇒ JL.

How do we get a JL matrix?

- Gaussians
  - Dimension $O(\frac{1}{\epsilon^2}k)$.
  - $\frac{1}{\epsilon^2}nk$ multiplication time.
- Best way known for *fast JL*: by [Krahmer-Ward '11], RIP $\Rightarrow$ JL.
  - Existing results: dimension $O(\frac{1}{\epsilon^2}k\log^4 n)$.

How do we get a JL matrix?

- Gaussians
    - Dimension $O(\frac{1}{\epsilon^2}k)$.
    - $\frac{1}{\epsilon^2}nk$ multiplication time.
- Best way known for *fast JL*: by [Krahmer-Ward ’11], RIP ⇒ JL.
    - Existing results: dimension $O(\frac{1}{\epsilon^2}k\log^4 n)$.
    - $n \log n$ multiplication time.

How do we get a JL matrix?

- Gaussians
  - Dimension $O(\frac{1}{\epsilon^2} k)$.
  - $\frac{1}{\epsilon^2} nk$ multiplication time.
- Best way known for *fast JL*: by [Krahmer-Ward '11], RIP ⇒ JL.
  - Existing results: dimension $O(\frac{1}{\epsilon^2} k \log^4 n)$.
  - $n \log n$ multiplication time.
- And by [BDDW '08], JL ⇒ RIP; so *equivalent*.[1]

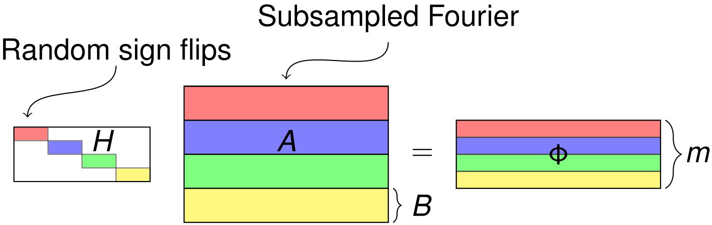
---

[1] Round trip loses $\log n$ factor in dimension

Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result
2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration
3. Proof
   - Overview
   - Covering Number
4. Conclusion

Our result: a fast RIP matrix with fewer rows



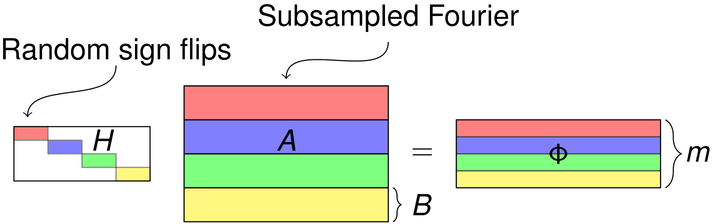
- New construction of fast RIP matrices: sparse times Fourier.
- $k \log^3 n$ rows and $n \log n$ multiplication time.

Theorem

If $m \simeq k \log^3 n$, $B \simeq \log^c n$, and A is a random partial Fourier matrix, then Φ has the RIP with probability at least $2/3$.

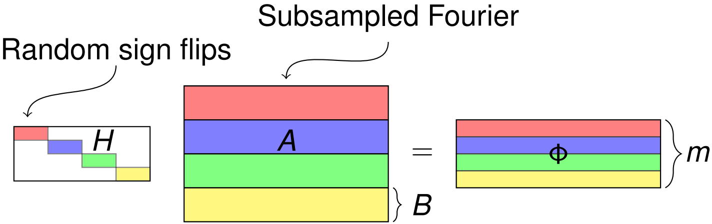
Generalization

Our approach is actually works for more general A:



Generalization

Our approach is actually works for more general A:

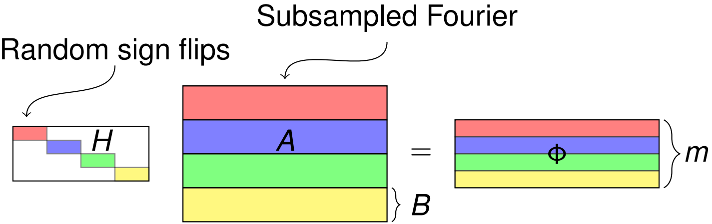


If A is a “decent” RIP matrix:

- A has RIP (whp), but too many (mB) rows.

Generalization

Our approach is actually works for more general A:

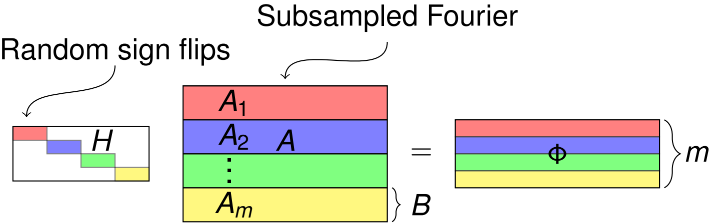


If A is a “decent” RIP matrix:

- A has RIP (whp), but too many (mB) rows.
- RIP-ness degrades “gracefully” as number of rows decreases:

Generalization

Our approach is actually works for more general A:

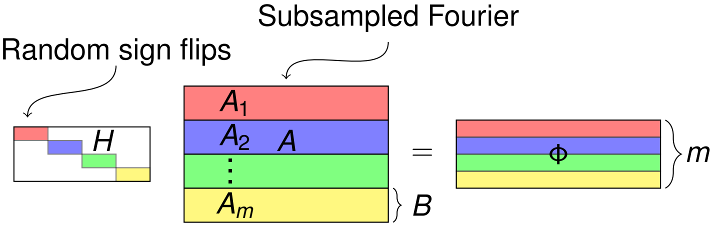


If A is a “decent” RIP matrix:

- A has RIP (whp), but too many (mB) rows.
- RIP-ness degrades “gracefully” as number of rows decreases:
  - For all A_i the RIP constant, although $\gg 1$, is still controlled.

Generalization

Our approach is actually works for more general A:



If A is a “decent” RIP matrix:

- A has RIP (whp), but too many (mB) rows.
- RIP-ness degrades “gracefully” as number of rows decreases:
    - For all A_i the RIP constant, although $\gg 1$, is still controlled.

Then Φ is a good RIP matrix:

- Φ has the RIP (whp) with $m = O(k \log^3 n)$ rows.
- Time to multiply by Φ = time to multiply by $A + mB$.

Results in the area

Construction	Measurements m	Multiplication Time
Sparse JL matrices [KN12]	$\frac{1}{\epsilon^2} k \log n$	ϵmn
Partial Fourier [RV08,CGV13]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Partial Circulant [KMR12]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Our result: Hash of partial Fourier	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Our result: Hash of partial circulant	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$

Results in the area

Construction	Measurements m	Multiplication Time
Sparse JL matrices [KN12]	$\frac{1}{\epsilon^2} k \log n$	ϵmn
Partial Fourier [RV08,CGV13]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Partial Circulant [KMR12]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Our result: Hash of partial Fourier	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Our result: Hash of partial circulant	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Iterated Fourier [AC06,AL09,AR13]	$\frac{1}{\epsilon^2} k \log n^\dagger$	$n \log n$

† Requires $k \leq n^{1/2-\delta}$. This is the "easy" case:

Results in the area

Construction	Measurements m	Multiplication Time
Sparse JL matrices [KN12]	$\frac{1}{\epsilon^2} k \log n$	ϵmn
Partial Fourier [RV08,CGV13]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Partial Circulant [KMR12]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Our result: Hash of partial Fourier	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Our result: Hash of partial circulant	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Iterated Fourier [AC06,AL09,AR13]	$\frac{1}{\epsilon^2} k \log n^\dagger$	$n \log n$

† Requires $k \leq n^{1/2-\delta}$. This is the "easy" case:

Dimension:	$n \xrightarrow{\hspace{3em}} k \log^4 n$
Time:	$n \log n$

[RV08]

Results in the area

Construction	Measurements m	Multiplication Time
Sparse JL matrices [KN12]	$\frac{1}{\epsilon^2} k \log n$	ϵmn
Partial Fourier [RV08,CGV13]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Partial Circulant [KMR12]	$\frac{1}{\epsilon^2} k \log^4 n$	$n \log n$
Our result: Hash of partial Fourier	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Our result: Hash of partial circulant	$\frac{1}{\epsilon^2} k \log^3 n$	$n \log n$
Iterated Fourier [AC06,AL09,AR13]	$\frac{1}{\epsilon^2} k \log n^\dagger$	$n \log n$

† Requires $k \leq n^{1/2-\delta}$. This is the "easy" case:

Dimension:	$n \longrightarrow k \log^4 n$	$\longrightarrow k \log n$
Time:	$n \log n$	$k^2 \log^5 n$
	[RV08]	Gaussian

Concentration of Measure

Let Σ_k is unit-norm k-sparse vectors.
We want to show for our distribution Φ on matrices that

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \epsilon,$$

Concentration of Measure

Let Σ_k is unit-norm k-sparse vectors.
We want to show for our distribution Φ on matrices that

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \epsilon,$$

(Expectation of $\star$) = $\star$

Concentration of Measure

Let Σ_k is unit-norm k-sparse vectors.
We want to show for our distribution Φ on matrices that

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \epsilon,$$

(Expectation of $\star$) = $\star$

Expected deviation of $\Phi^T \Phi$ from mean I_n, in a funny norm.

Concentration of Measure

Let Σ_k is unit-norm k-sparse vectors.
We want to show for our distribution Φ on matrices that

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \epsilon,$$

(Expectation of $\star$) = $\star$

Expected deviation of $\Phi^T \Phi$ from mean I_n, in a funny norm.

Probabilists have lots of tools to analyze this.

Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result
2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration
3. Proof
   - Overview
   - Covering Number
4. Conclusion

Tools

Tools

Screwdriver

Tools

Screwdriver

Drill

Tools

Screwdriver

Bit sets

Drill

Tools

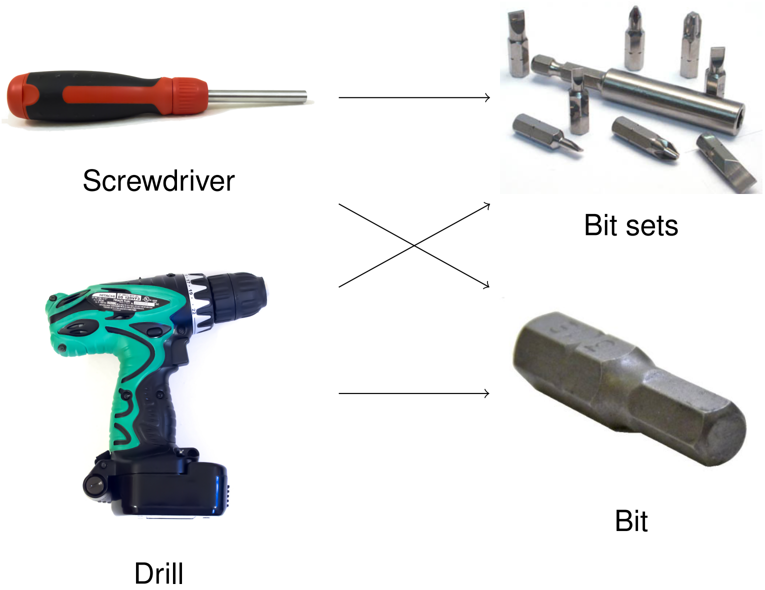
Screwdriver

Bit sets

Drill

Bit

Tools



Tools

Common interface: m drivers, n bits $\implies$ mn combinations.

Tools

Common interface: m drivers, n bits $\implies$ mn combinations.

Hex shanks

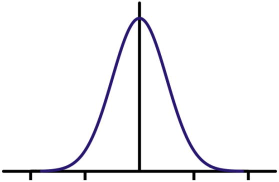
Common interface
for drill bits

Tools

Common interface: m drivers, n bits $\implies$ mn combinations.

Hex shanks

Common interface
for drill bits



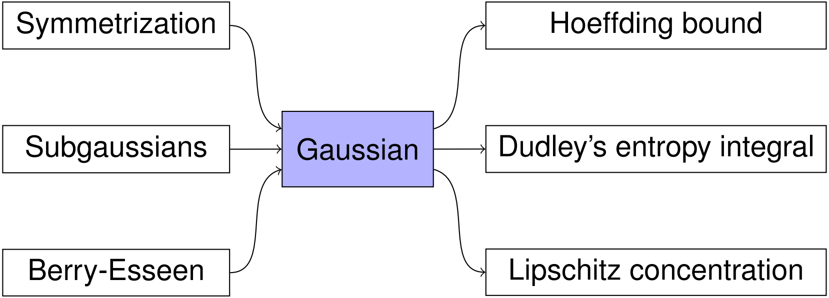
Gaussians

Common interface
for probability

A Probabilist's Toolbox

Convert to Gaussians

Gaussian concentration



Symmetrization

Subgaussians

Berry-Esseen

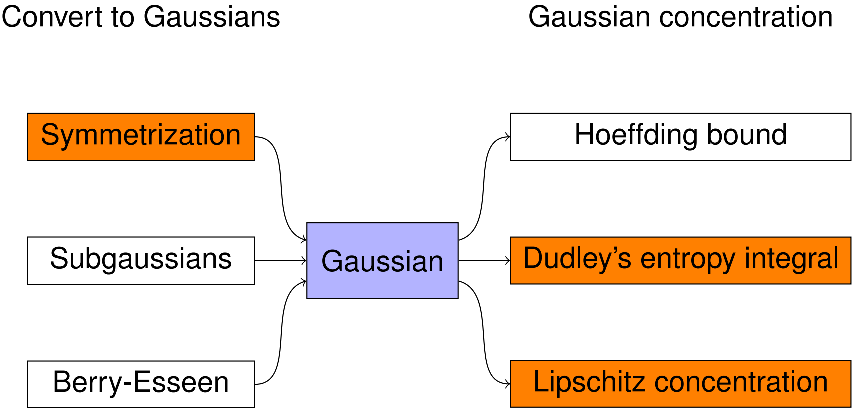
Gaussian

Hoeffding bound

Dudley's entropy integral

Lipschitz concentration

A Probabilist's Toolbox



Convert to Gaussians

Gaussian concentration

Symmetrization

Subgaussians

Berry-Esseen

Gaussian

Hoeffding bound

Dudley's entropy integral

Lipschitz concentration

Will prove: symmetrization and Dudley's entropy integral.

Outline

1. Introduction
    - Compressive sensing
    - Johnson Lindenstrauss Transforms
    - Our result
2. Concentration of measure: a toolbox
    - Overview
    - Symmetrization
    - Gaussian Processes
    - Lipschitz Concentration
3. Proof
    - Overview
    - Covering Number
4. Conclusion

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ *are i.i.d. with mean* μ*. For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently.*

How well does X concentrate about its mean?

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

How well does X concentrate about its mean?

Example (RIP)

For some norm $\|\cdot\|$, RIP constant of subsampled Fourier

$$\|A^T A - I\|$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \dots, X_t$ *are i.i.d. with mean* μ. *For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently.*

How well does X concentrate about its mean?

Example (RIP)

For some norm $\|\cdot\|$, RIP constant of subsampled Fourier

$$\|A^T A - I\| = \|\sum A_i^T A_i - I\|.$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Proof.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Proof.

Draw $X_1', \ldots, X_t'$ independently from the same distribution.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ *are i.i.d. with mean* μ*. For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently.*

Proof.

Draw $X'_1, \ldots, X'_t$ independently from the same distribution.

$$\mathbb{E}[\|\frac{1}{t}\sum X_i - \mathbb{E}[\frac{1}{t}\sum X'_i]\|]$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Proof.

Draw $X_1', \ldots, X_t'$ independently from the same distribution.

$$\mathbb{E}[\|\frac{1}{t}\sum X_i - \mathbb{E}[\frac{1}{t}\sum X_i']\|] \leq \mathbb{E}[\|\frac{1}{t}\sum(X_i - X_i')\|]$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Proof.

Draw $X_1', \ldots, X_t'$ independently from the same distribution.

$$\begin{aligned}\mathbb{E}[\|\frac{1}{t}\sum X_i - \mathbb{E}[\frac{1}{t}\sum X_i']\|] &\leq \mathbb{E}[\|\frac{1}{t}\sum (X_i - X_i')\|] \\ &= \mathbb{E}[\|\frac{1}{t}\sum s_i(X_i - X_i')\|]\end{aligned}$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently.

Proof.

Draw $X_1', \ldots, X_t'$ independently from the same distribution.

$$\begin{aligned}\mathbb{E}[\|\frac{1}{t}\sum X_i - \mathbb{E}[\frac{1}{t}\sum X_i']\|] &\leq \mathbb{E}[\|\frac{1}{t}\sum(X_i - X_i')\|] \\ &= \mathbb{E}[\|\frac{1}{t}\sum s_i(X_i - X_i')\|]\end{aligned}$$

and apply the triangle inequality. □

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ *are i.i.d. with mean* μ. *For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently.*

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \dots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently and $g_i \sim N(0, 1)$ independently.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently and $g_i \sim N(0, 1)$ independently.

Proof.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \dots, X_t$ *are i.i.d. with mean* μ. *For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently and* $g_i \sim N(0,1)$ *independently.*

Proof.

We have $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently and $g_i \sim N(0,1)$ independently.

Proof.

We have $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

$$2\,\mathbb{E}[\|\sum s_i X_i\|]$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently and $g_i \sim N(0,1)$ independently.

Proof.

We have $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

$$2\,\mathbb{E}[\|\sum s_i X_i\|] \leq 3\,\mathbb{E}[\|\sum s_i \mathbb{E}[|g_i|] X_i\|]$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ *are i.i.d. with mean* μ. *For any norm* $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ *independently and* $g_i \sim N(0,1)$ *independently.*

Proof.

We have $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

$$\begin{aligned} 2\,\mathbb{E}[\|\sum s_i X_i\|] &\leq 3\,\mathbb{E}[\|\sum s_i \mathbf{\mathbb{E}[|g_i|]} X_i\|] \\ &\leq 3\,\mathbb{E}[\|\sum s_i |\mathbf{g_i}| X_i\|] \end{aligned}$$

Symmetrization

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean μ. For any norm $\|\cdot\|$,

$$\mathbb{E}\left[\left\|\frac{1}{t}\sum_i X_i - \mu\right\|\right] \leq 2\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i s_i X_i\right\|\right] \leq 3\,\mathbb{E}\left[\left\|\frac{1}{t}\sum_i g_i X_i\right\|\right]$$

where $s_i \in \{\pm 1\}$ independently and $g_i \sim N(0,1)$ independently.

Proof.

We have $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

$$\begin{aligned} 2\,\mathbb{E}[\|\sum s_i X_i\|] &\leq 3\,\mathbb{E}[\|\sum s_i \mathbf{\mathbb{E}[|g_i|]} X_i\|] \\ &\leq 3\,\mathbb{E}[\|\sum s_i \mathbf{|g_i|} X_i\|] \\ &= 3\,\mathbb{E}[\|\sum \mathbf{g_i} X_i\|]. \end{aligned}$$

□

Outline

1. Introduction
    - Compressive sensing
    - Johnson Lindenstrauss Transforms
    - Our result
2. Concentration of measure: a toolbox
    - Overview
    - Symmetrization
    - Gaussian Processes
    - Lipschitz Concentration
3. Proof
    - Overview
    - Covering Number
4. Conclusion

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle uv^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle uv^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.
- Standard problem: $\mathbb{E} \sup_{x \in T} G_x$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle uv^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.
- Standard problem: $\mathbb{E} \sup_{x \in T} G_x$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle uv^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

- Depends on the geometry of T.

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.
- Standard problem: $\mathbb{E} \sup_{x \in T} G_x$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle u v^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

- Depends on the geometry of T.
- Distance: $\|x - y\|$ is standard deviation of $G_x - G_y$.

Gaussian Processes

- Gaussian process G_x: a Gaussian at each point $x \in T$.
- Standard problem: $\mathbb{E} \sup_{x \in T} G_x$.

Example (Maximum singular value of random Gaussian matrix)

Let A be a random $m \times n$ Gaussian matrix. For any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, define

$$G_{u,v} := u^T A v = \langle uv^T, A \rangle.$$

Then $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

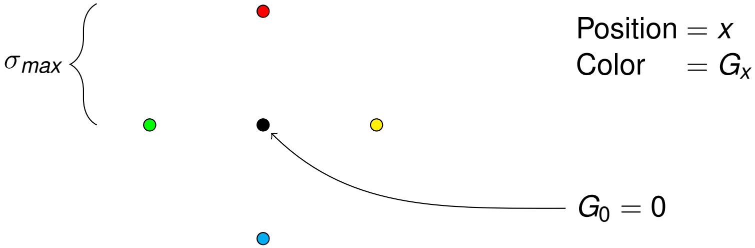
- Depends on the geometry of T.
- Distance: $\|x - y\|$ is standard deviation of $G_x - G_y$.
- In example: $\|(u, v) - (u', v')\| = \|uv^T - u'v'^T\|_F$.

Gaussian Processes

- Goal: $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x - y\|^2)$.

Gaussian Processes

- Goal: $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x - y\|^2)$.



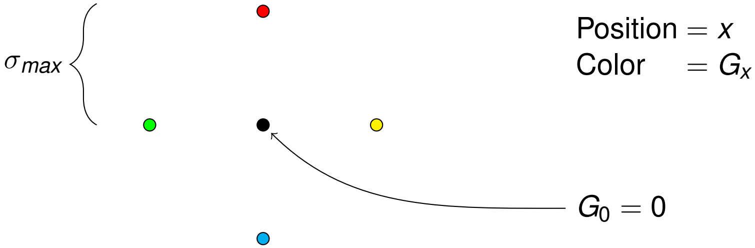
σ_{max}

Position $= x$
Color $= G_x$

$G_0 = 0$

Gaussian Processes

- Goal: $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x-y\|^2)$.
- Ignoring geometry:



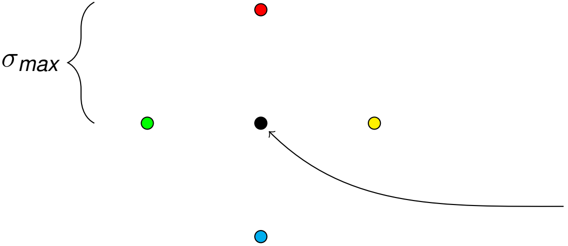
σ_{max}

Position $= x$
Color $= G_x$

$G_0 = 0$

Gaussian Processes

- Goal: $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x-y\|^2)$.
- Ignoring geometry:
  - $\Pr[G_x > \sigma_{max} t] \leq e^{-t^2/2}$

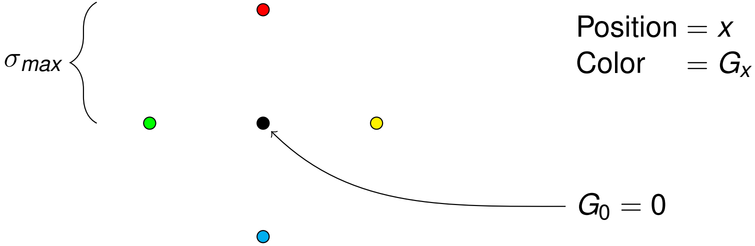


Position $= x$
Color $= G_x$

$G_0 = 0$

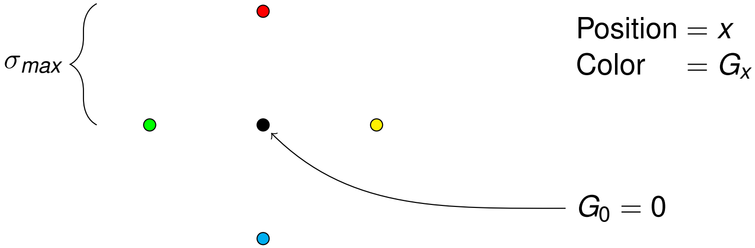
Gaussian Processes

- Goal: $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x - y\|^2)$.
- Ignoring geometry:
  - $\Pr[G_x > \sigma_{max} t] \le e^{-t^2/2}$
  - Union bound: with high probability, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.



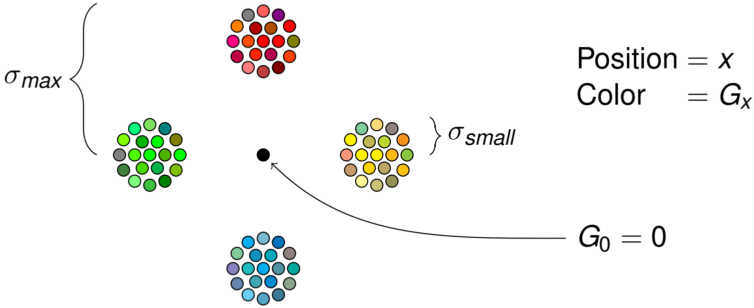
Gaussian Processes

- Goal: $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x-y\|^2)$.
- Ignoring geometry:
  - $\Pr[G_x > \sigma_{max} t] \leq e^{-t^2/2}$
  - Union bound: with high probability, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.
  - $\mathbb{E}\sup_{x \in T} G_x \lesssim \sigma_{max}\sqrt{\log n}$



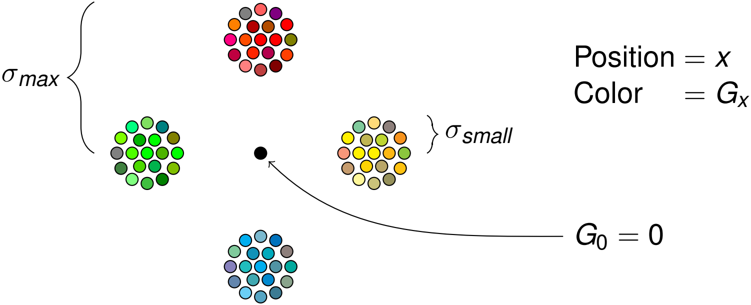
Gaussian Processes

- Goal: $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x-y\|^2)$.
- Ignoring geometry:
  - $\Pr[G_x > \sigma_{max} t] \leq e^{-t^2/2}$
  - Union bound: with high probability, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.
  - $\mathbb{E}\sup_{x \in T} G_x \lesssim \sigma_{max}\sqrt{\log n}$



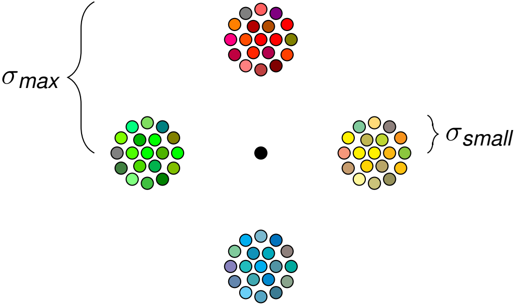
Gaussian Processes

- Goal: $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y \sim N(0, \|x - y\|^2)$.
- Ignoring geometry:
  - $\Pr[G_x > \sigma_{max} t] \leq e^{-t^2/2}$
  - Union bound: with high probability, $G_x \lesssim \sigma_{max} \sqrt{\log n}$.
  - $\mathbb{E} \sup_{x \in T} G_x \lesssim \sigma_{max} \sqrt{\log n}$
- Two levels: $\sigma_{max} \sqrt{\log 4} + \sigma_{small} \sqrt{\log n}$.



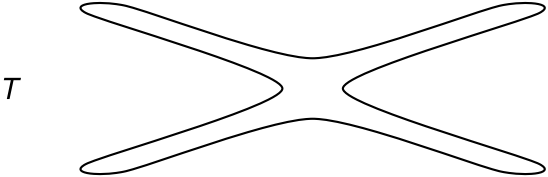
Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.



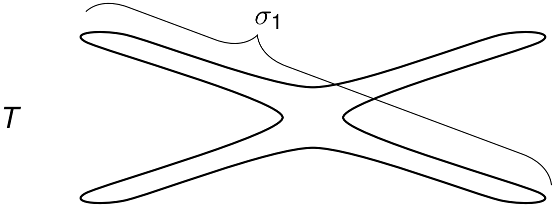
Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.



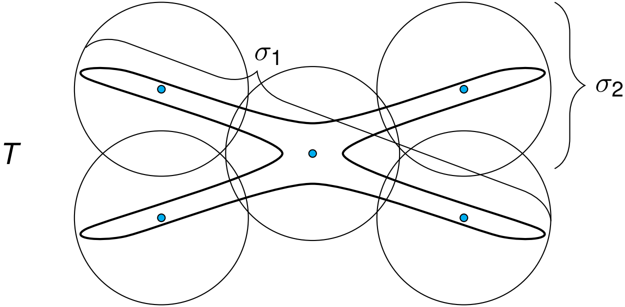
Gaussian Processes: chaining

- Bound $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.



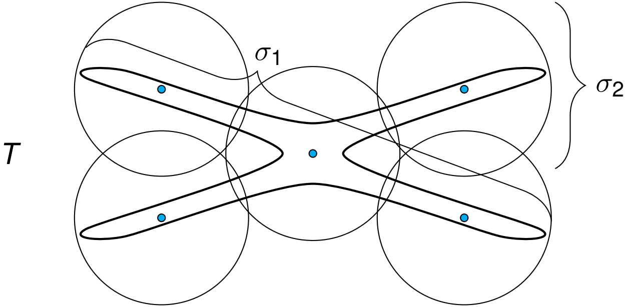
Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.



Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.

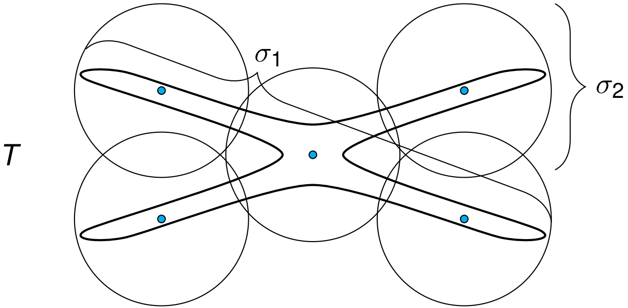


balls necessary:
$N(\sigma_2)$

(*covering number* depends on $T, \|\cdot\|$)

Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.

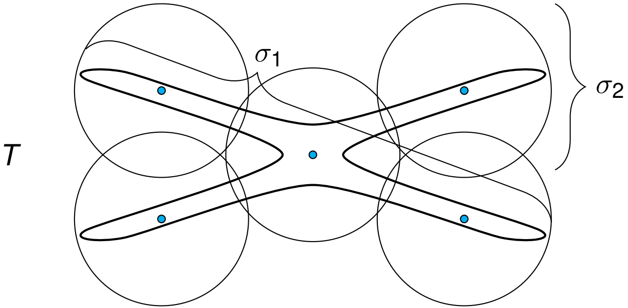


\# balls necessary: $N(\sigma_2)$

(*covering number* depends on $T, \|\cdot\|$)

Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_1 \sqrt{\log N(\sigma_2)} + \sigma_2 \sqrt{\log n}$.
- Why stop at two?



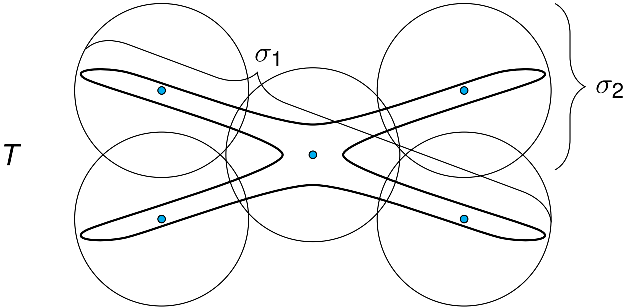
balls necessary:
$N(\sigma_2)$

(*covering number* depends on T, $\|\cdot\|$)

Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x\in T} G_x$, where $G_x - G_y$ has variance $\|x-y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

$$\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} +$$



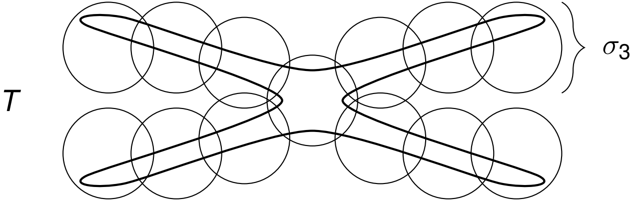
balls necessary: $N(\sigma_2)$

(*covering number* depends on T, $\|\cdot\|$)

Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x\in T} G_x$, where $G_x - G_y$ has variance $\|x-y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

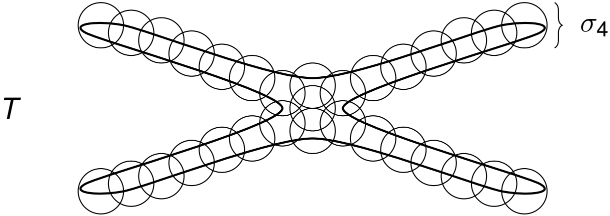
$$\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log N(\sigma_3)} +$$



Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x\in T} G_x$, where $G_x - G_y$ has variance $\|x-y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

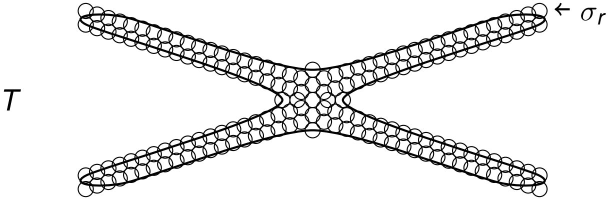
$$\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log N(\sigma_3)} + \sigma_3\sqrt{\log N(\sigma_4)} +$$



Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

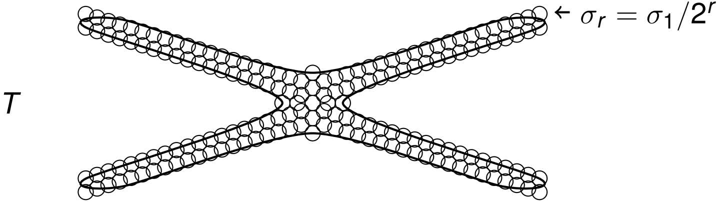
$$\mathbb{E}\sup_{x \in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log N(\sigma_3)} + \sigma_3\sqrt{\log N(\sigma_4)} + \cdots$$



Gaussian Processes: chaining

- Bound $\mathbb{E} \sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_1 \sqrt{\log N(\sigma_2)} + \sigma_2 \sqrt{\log n}$.
- Why stop at two?

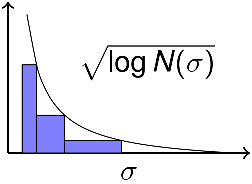
$$\mathbb{E} \sup_{x \in T} G_x \lesssim \sum_{r=0}^{\infty} \frac{\sigma_1}{2^r} \sqrt{\log N\left(\frac{\sigma_1}{2^{r+1}}\right)}$$

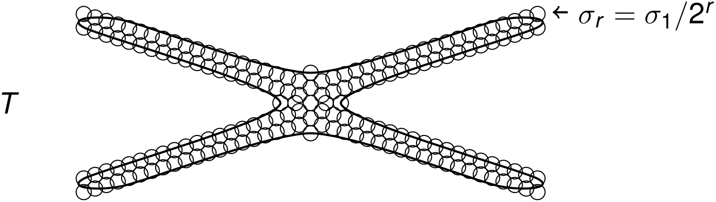


Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x \in T} G_x$, where $G_x - G_y$ has variance $\|x - y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

$$\mathbb{E}\sup_{x \in T} G_x \lesssim \sum_{r=0}^{\infty} \frac{\sigma_1}{2^r}\sqrt{\log N\left(\frac{\sigma_1}{2^{r+1}}\right)}$$

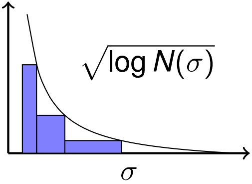


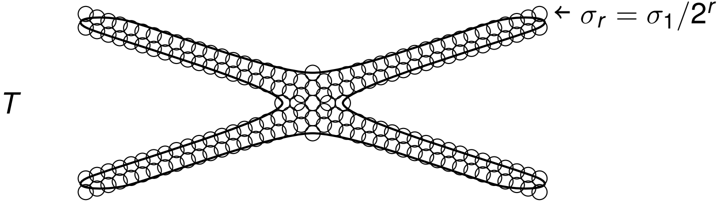


Gaussian Processes: chaining

- Bound $\mathbb{E}\sup_{x\in T} G_x$, where $G_x - G_y$ has variance $\|x-y\|^2$.
- Two levels: $\sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log n}$.
- Why stop at two?

$$\mathbb{E}\sup_{x\in T} G_x \lesssim \int_0^\infty \sqrt{\log N(\sigma)}d\sigma$$





Gaussian Processes

Dudley's Entropy Integral, Talagrand's generic chaining

Theorem (Dudley's Entropy Integral)

Define the norm $\|\cdot\|$ *of a Gaussian process G by*

$$\|x - y\| = \textit{standard deviation of } (G_x - G_y).$$

Then

$$\mathbb{E} \sup_{x \in T} G_x \lesssim \int_0^\infty \sqrt{\log N(T, \|\cdot\|, u)} du$$

Gaussian Processes

Dudley's Entropy Integral, Talagrand's generic chaining

Theorem (Dudley's Entropy Integral)

Define the norm $\|\cdot\|$ of a Gaussian process G by

$$\|x - y\| = \textit{standard deviation of } (G_x - G_y).$$

Then

$$\gamma_2(T, \|\cdot\|) := \mathbb{E} \sup_{x \in T} G_x \lesssim \int_0^\infty \sqrt{\log N(T, \|\cdot\|, u)} du$$

Gaussian Processes

Dudley's Entropy Integral, Talagrand's generic chaining

Theorem (Dudley's Entropy Integral)

Define the norm $\|\cdot\|$ of a Gaussian process G by

$$\|x - y\| = \textit{standard deviation of } (G_x - G_y).$$

Then

$$\gamma_2(T, \|\cdot\|) := \mathbb{E} \sup_{x \in T} G_x \lesssim \int_0^\infty \sqrt{\log N(T, \|\cdot\|, u)} du$$

- Bound a random variable using geometry.

Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result
2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration
3. Proof
   - Overview
   - Covering Number
4. Conclusion

Lipschitz Concentration of Gaussians

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is C-Lipschitz and $g \sim N(0, I_n)$, then for any $t > 0$,

$$\Pr[f(g) > \mathbb{E}[f(g)] + Ct] \le e^{-\Omega(t^2)}.$$

- f concentrates as well as *individual* Gaussians.
- Can replace f with $-f$ to get lower tail bound.

Lipschitz Concentration of Gaussians

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is C-Lipschitz and $g \sim N(0, I_n)$, then for any $t > 0$,

$$\Pr[f(g) > \mathbb{E}[f(g)] + Ct] \leq e^{-\Omega(t^2)}.$$

- f concentrates as well as *individual* Gaussians.
- Can replace f with $-f$ to get lower tail bound.

Example

If $g \sim N(0, I_n)$, then with probability $1 - \delta$,

$$\|g\|_2 \leq \sqrt{n} + O(\sqrt{\log(1/\delta)}).$$

Lipschitz Concentration of Gaussians

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is C-Lipschitz and $g \sim N(0, I_n)$, then for any $t > 0$,

$$\Pr[f(g) > \mathbb{E}[f(g)] + Ct] \leq e^{-\Omega(t^2)}.$$

- f concentrates as well as *individual* Gaussians.
- Can replace f with $-f$ to get lower tail bound.

Example

If $g \sim N(0, I_n)$, then with probability $1 - \delta$,

$$\|g\|_2 \leq \sqrt{n} + O(\sqrt{\log(1/\delta)}).$$

For $n = O(1/\epsilon^2 \log(1/\delta))$, this is $1 \pm \epsilon$ approximation.

Lipschitz Concentration of Gaussians

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is C-Lipschitz and $g \sim N(0, I_n)$, then for any $t > 0$,

$$\Pr[f(g) > \mathbb{E}[f(g)] + Ct] \leq e^{-\Omega(t^2)}.$$

- f concentrates as well as *individual* Gaussians.
- Can replace f with $-f$ to get lower tail bound.

Example

If $g \sim N(0, I_n)$, then with probability $1 - \delta$,

$$\|g\|_2 \leq \sqrt{n} + O(\sqrt{\log(1/\delta)}).$$

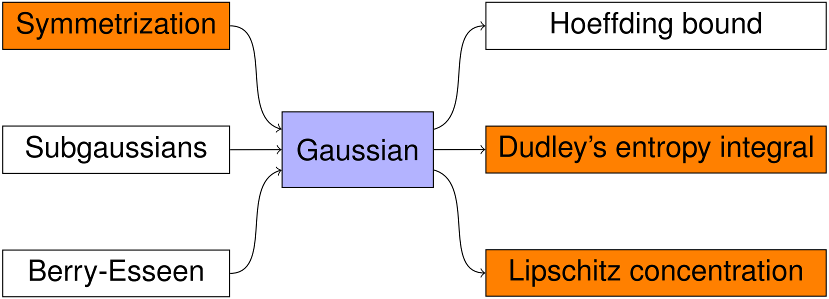
For $n = O(1/\epsilon^2 \log(1/\delta))$, this is $1 \pm \epsilon$ approximation.

$\implies$ the Johnson-Lindenstrauss lemma.

A Probabilist's Toolbox (recap)

Convert to Gaussians

Gaussian concentration



Symmetrization

Subgaussians

Berry-Esseen

Gaussian

Hoeffding bound

Dudley's entropy integral

Lipschitz concentration

Outline

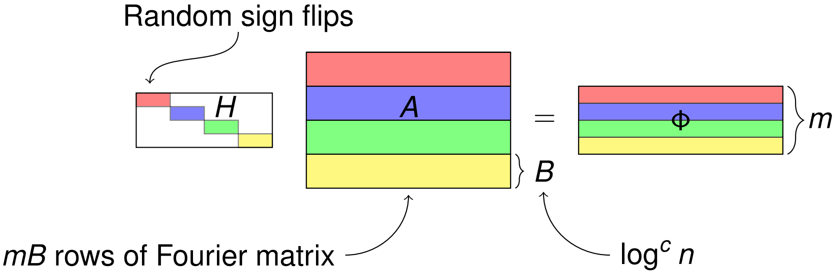
1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result

2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration

3. Proof
   - Overview
   - Covering Number

4. Conclusion

Goal



For Σ_k denoting unit-norm k-sparse vectors, we want

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \epsilon,$$

(Expectation of $*$) = $*$

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

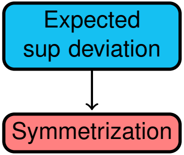
$\mathbb{E} \sup$
$\|A^T A - \mathrm{I}\|$

Expected
sup deviation

Proof outline: Rudelson-Vershynin

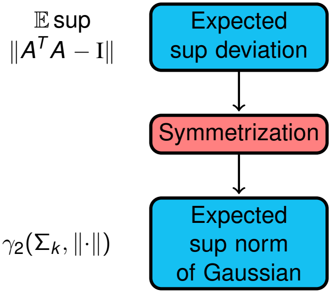
Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

$\mathbb{E}\sup$
$\|A^T A - \mathrm{I}\|$



Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



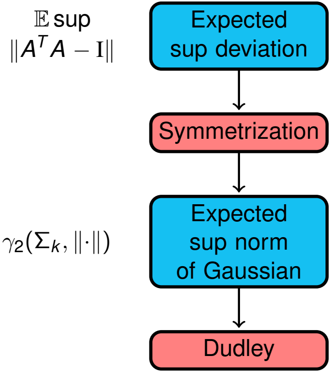
γ_2 : supremum of Gaussian process

Σ_k : k-sparse unit vectors

$\|\cdot\|$: a norm that depends on A (specified in a few slides)

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E} \sup$
$\|A^T A - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

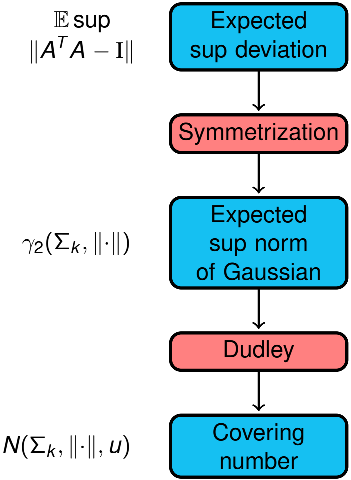
γ_2 : supremum of Gaussian process

Σ_k : k-sparse unit vectors

$\|\cdot\|$: a norm that depends on A
(specified in a few slides)

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E} \sup \|A^T A - \mathrm{I}\|$ — Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$ — Expected sup norm of Gaussian

Dudley

$N(\Sigma_k, \|\cdot\|, u)$ — Covering number

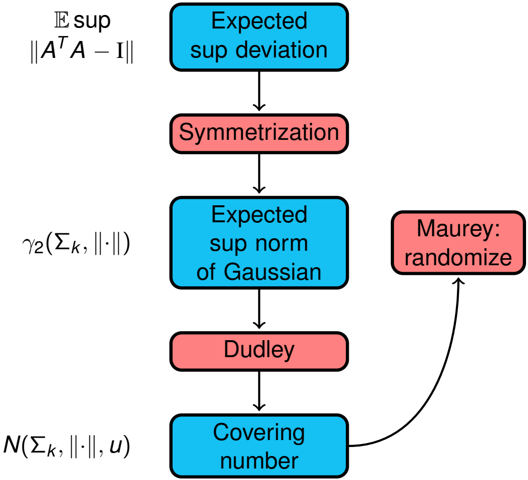
γ_2 : supremum of Gaussian process

Σ_k : k-sparse unit vectors

$\|\cdot\|$: a norm that depends on A (specified in a few slides)

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E} \sup \|A^T A - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Maurey: randomize

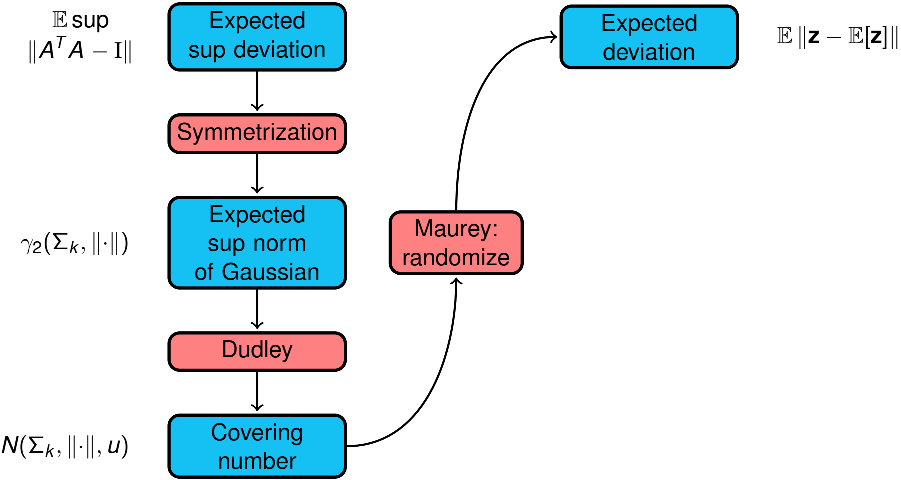
Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E} \sup \|A^T A - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

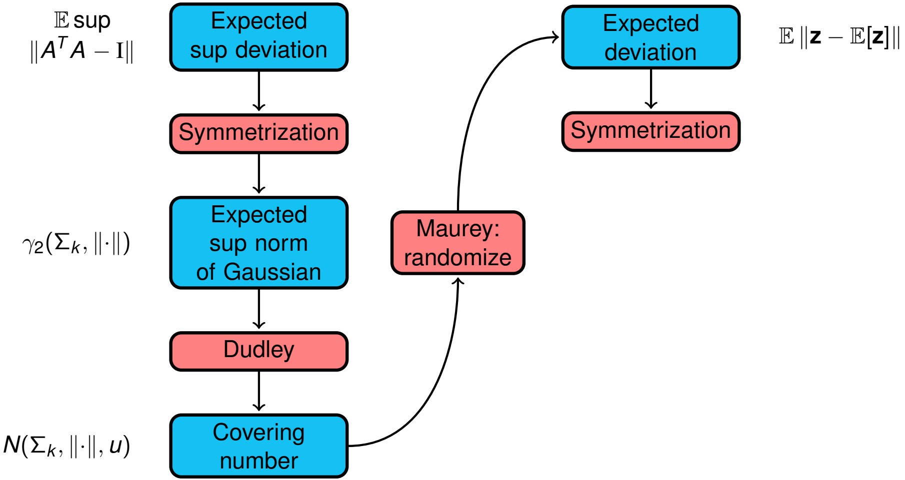
Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k\log^4 n)$ rows.



$\mathbb{E}\sup \|A^TA - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

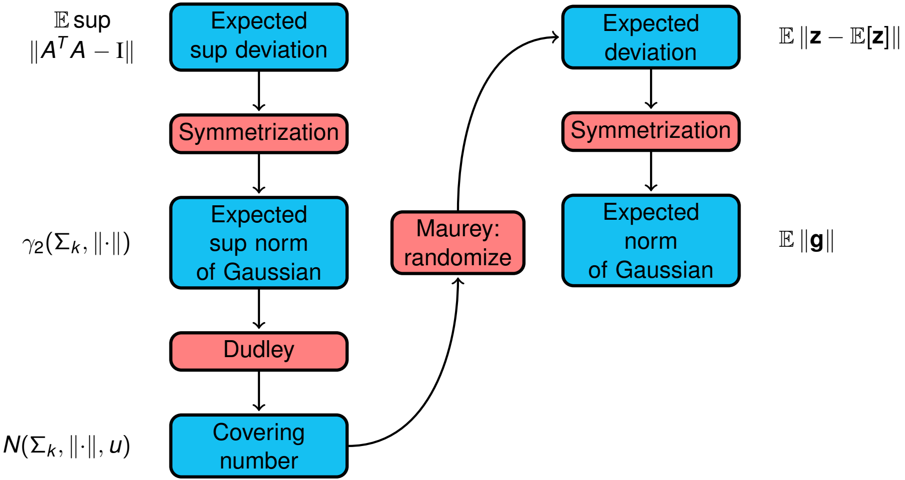
Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Symmetrization

Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E}\sup \|A^T A - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

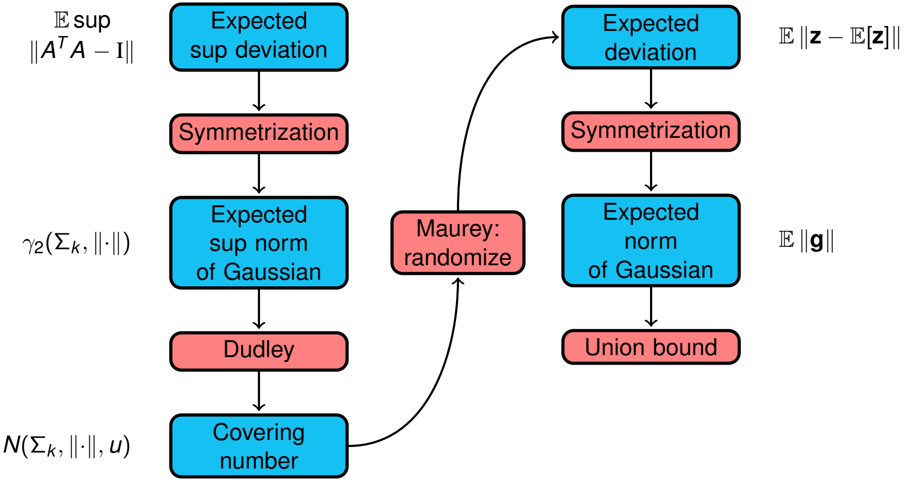
Symmetrization

Expected norm of Gaussian

$\mathbb{E}\|\mathbf{g}\|$

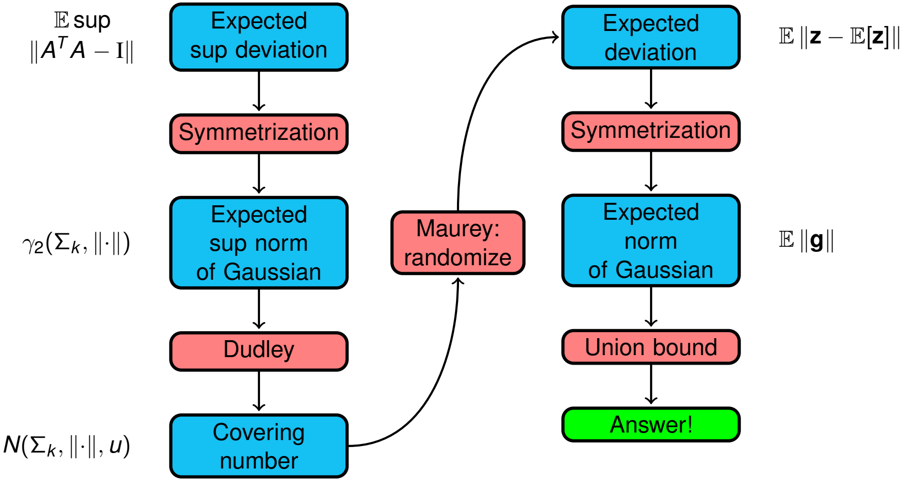
Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



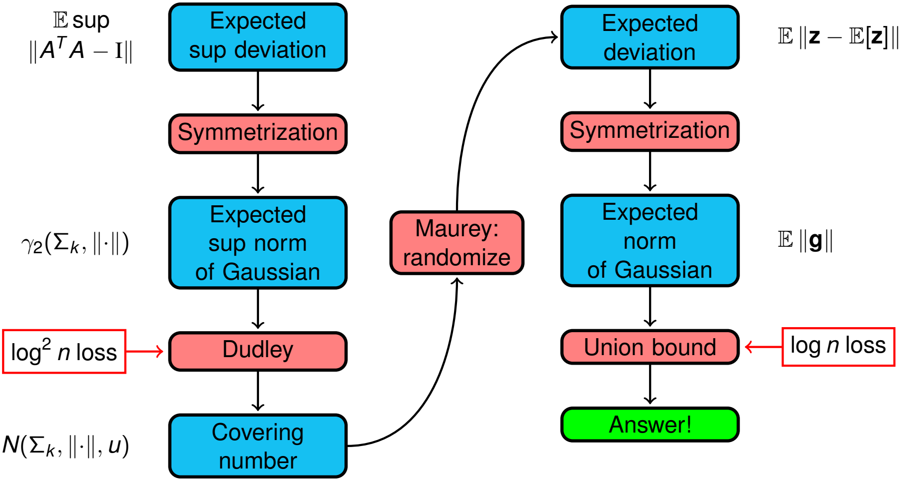
Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



Proof outline: Rudelson-Vershynin

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.



$\mathbb{E}\sup \|A^T A - \mathrm{I}\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

log^2 *n* loss

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Symmetrization

Expected norm of Gaussian

$\mathbb{E}\|\mathbf{g}\|$

Union bound

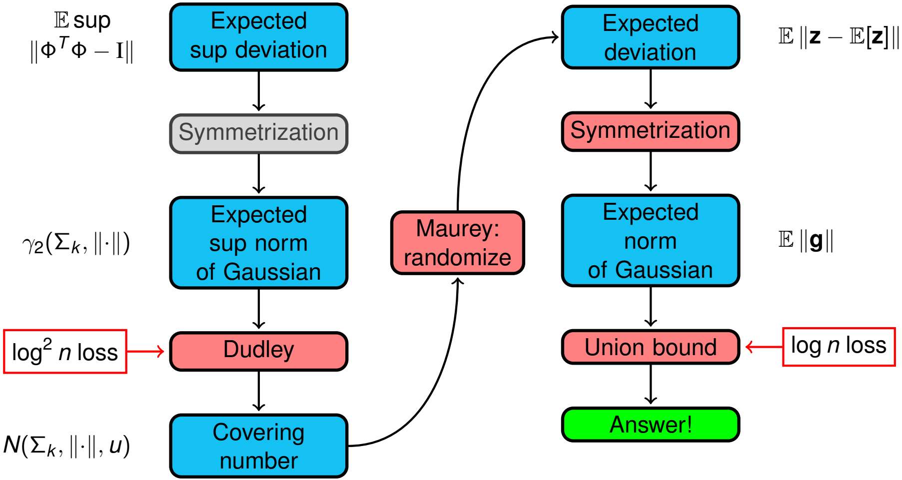
log *n* loss

Answer!

Proof outline

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

Nelson-P-Wootters: sparse times Fourier, $O(k \log^3 n)$ rows.



$\mathbb{E} \sup \|\Phi^T\Phi - I\|$

Expected sup deviation

Symmetrization

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

$\log^2 n$ loss

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Expected deviation

Symmetrization

$\mathbb{E}\|\mathbf{g}\|$

Expected norm of Gaussian

Union bound

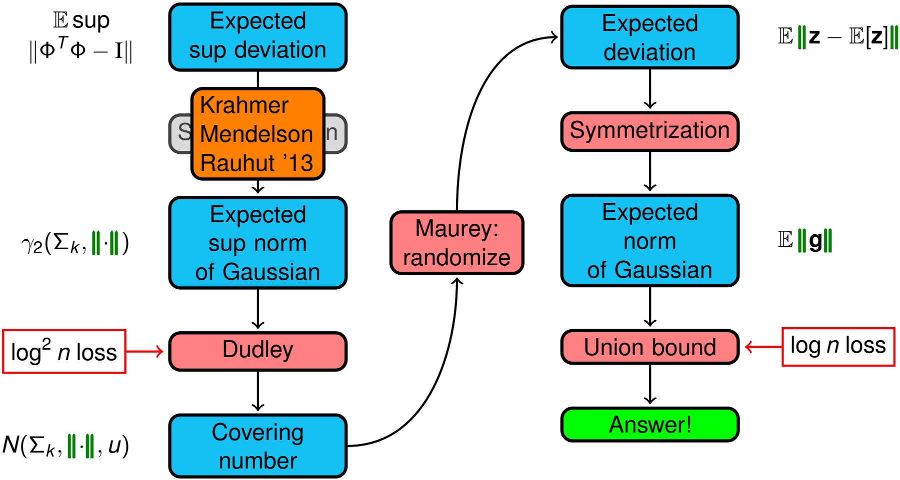
$\log n$ loss

Answer!

Proof outline

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

Nelson-P-Wootters: sparse times Fourier, $O(k \log^3 n)$ rows.



$\mathbb{E} \sup \|\Phi^T \Phi - I\|$

Expected sup deviation

S Krahmer Mendelson Rauhut '13 n

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

log^2 n loss

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Expected deviation

Symmetrization

$\mathbb{E}\|\mathbf{g}\|$

Expected norm of Gaussian

log n loss

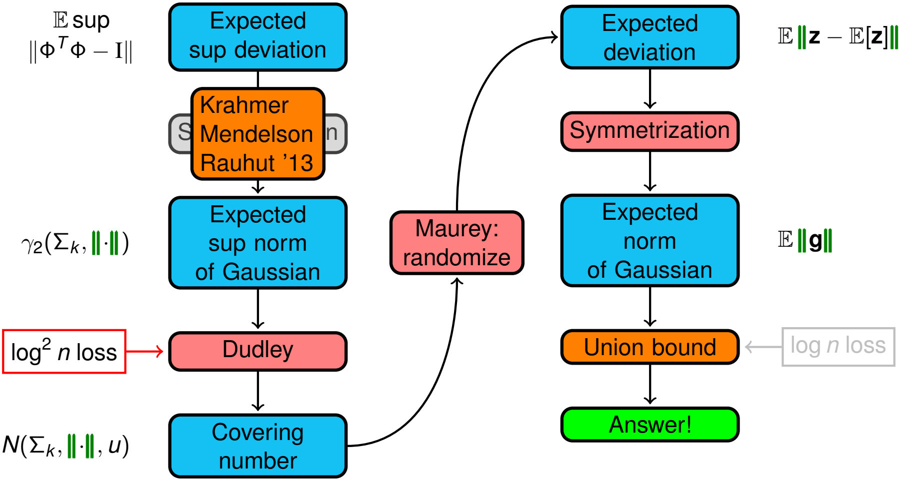
Union bound

Answer!

Proof outline

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

Nelson-P-Wootters: sparse times Fourier, $O(k \log^3 n)$ rows.



$\mathbb{E} \sup \|\Phi^T \Phi - I\|$

Expected sup deviation

S n

Krahmer Mendelson Rauhut '13

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

log² n loss

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Expected deviation

Symmetrization

$\mathbb{E}\|\mathbf{g}\|$

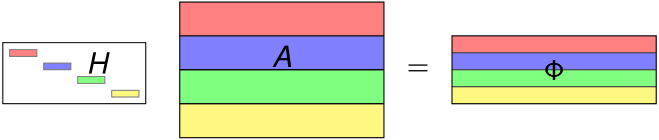
Expected norm of Gaussian

Union bound

log n loss

Answer!

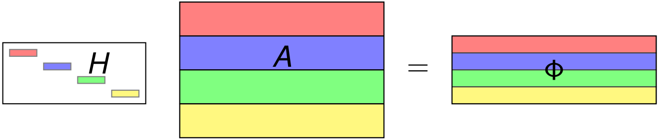
Proof part I: triangle inequality



$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right|$$

$$\leq \mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|Ax\|_2^2 \right| + \mathbb{E} \sup_{x \in \Sigma_k} \left| \|Ax\|_2^2 - \|x\|_2^2 \right|$$

Proof part I: triangle inequality



$$
\begin{aligned}
&\mathbb{E}\sup_{x\in\Sigma_k}\left|\|\Phi x\|_2^2-\|x\|_2^2\right| \\
&\quad\le \mathbb{E}\sup_{x\in\Sigma_k}\left|\|\Phi x\|_2^2-\|Ax\|_2^2\right|+\mathbb{E}\sup_{x\in\Sigma_k}\left|\|Ax\|_2^2-\|x\|_2^2\right| \\
&\quad= \mathbb{E}\sup_{x\in\Sigma_k}\left|\|X_As\|_2^2-\mathbb{E}_s\|X_As\|_2^2\right|+(\text{RIP constant of }A),
\end{aligned}
$$

where X_A is some matrix depending x and A, and s is the vector of random sign flips used in H.

Proof part I: triangle inequality

$$\mathbb{E}\sup_{x\in\Sigma_k}\left|\|X_A s\|_2^2 - \mathbb{E}_s\|X_A s\|_2^2\right| + (\text{RIP constant of } A)$$

Proof part I: triangle inequality

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|X_A s\|_2^2 - \mathbb{E}_s \|X_A s\|_2^2 \right| + (\text{RIP constant of } A)$$

By assumption, this is small.
(Recall *A* has extra rows)

Proof part I: triangle inequality

$$\mathbb{E}\sup_{x\in\Sigma_k}\left|\|X_A s\|_2^2 - \mathbb{E}_s\|X_A s\|_2^2\right| + (\text{RIP constant of } A)$$

By assumption, this is small.
(Recall *A* has extra rows)

This is a *Rademacher Chaos Process.*
We have to do some work to show that it is small.

Proof part II: probability and geometry

By [KMR12] and some manipulation, can bound the Rademacher chaos using

$$\gamma_2(\Sigma_k, \| \cdot \|_A)$$

Some norm induced by A

The supremum of a Gaussian process over Σ_k with norm $\|\cdot\|_A$

Dudley's entropy integral: can estimate this by bounding the *covering number* $N(\Sigma_k, \|\cdot\|_A, u)$.

Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

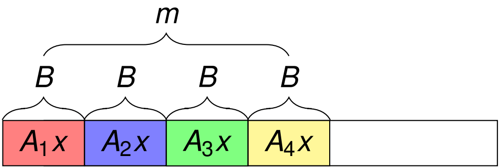
for the norm $\|x\|_A$:

Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

for the norm $\|x\|_A$:

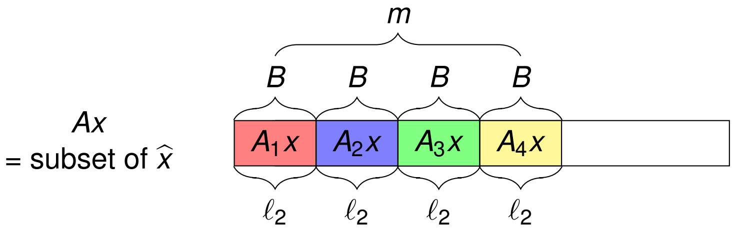
Ax
$= \text{subset of } \widehat{x}$



Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

for the norm $\|x\|_A$:

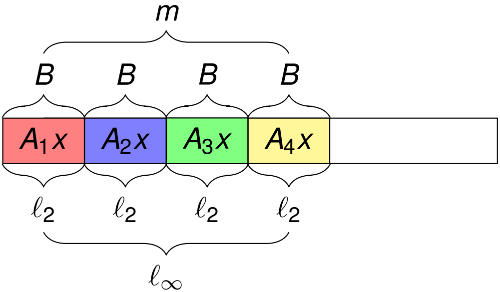


Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

for the norm $\|x\|_A$:

Ax
$= \text{subset of } \widehat{x}$

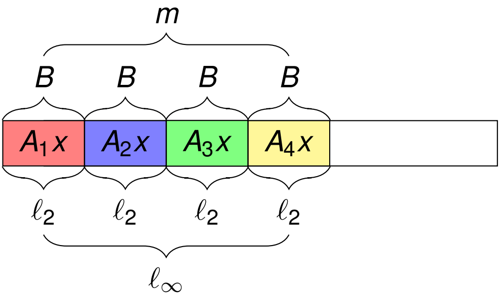


Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

for the norm $\|x\|_A$:

Ax
$= \text{subset of } \widehat{x}$



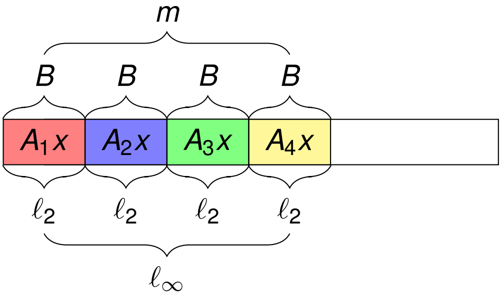
$$\|x\|_A = \max_{i \in [m]} \|A_i x\|_2.$$

Definition of the Norm

$$N(\Sigma_k, \|\cdot\|_A, u)$$

for the norm $\|x\|_A$:

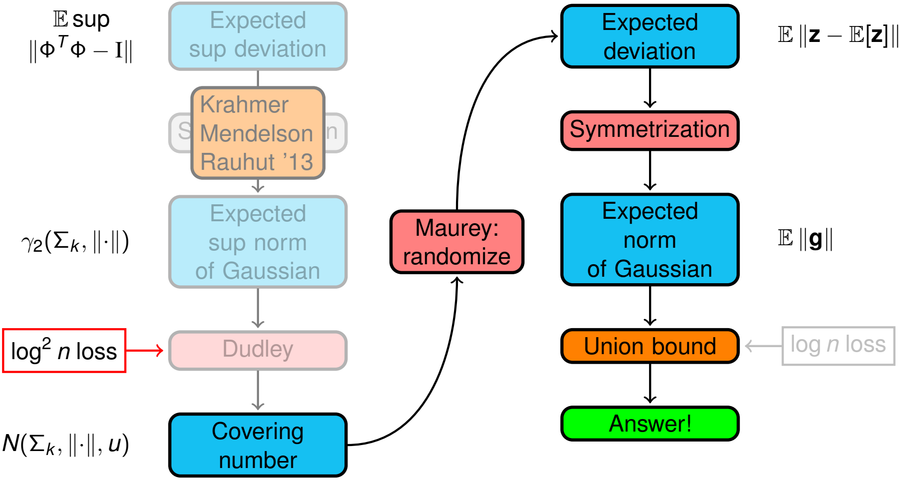
Ax
= subset of $\widehat{x}$



$$\|x\|_A = \max_{i \in [m]} \|A_i x\|_2.$$

Rudelson-Vershynin: estimates $N(\Sigma_k, \|\cdot\|_A, u)$ when $B = 1$.

Progress



$\mathbb{E}\sup$ $\|\Phi^T\Phi - \mathrm{I}\|$

Expected sup deviation

Krahmer Mendelson Rauhut '13

Expected sup norm of Gaussian

$\gamma_2(\Sigma_k, \|\cdot\|)$

log^2 *n* loss

Dudley

Covering number

$N(\Sigma_k, \|\cdot\|, u)$

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

Symmetrization

Expected norm of Gaussian

$\mathbb{E}\|\mathbf{g}\|$

Union bound

log *n* loss

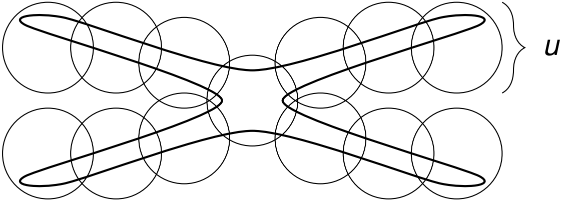
Answer!

Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms
   - Our result
2. Concentration of measure: a toolbox
   - Overview
   - Symmetrization
   - Gaussian Processes
   - Lipschitz Concentration
3. Proof
   - Overview
   - Covering Number
4. Conclusion

Covering Number Bound

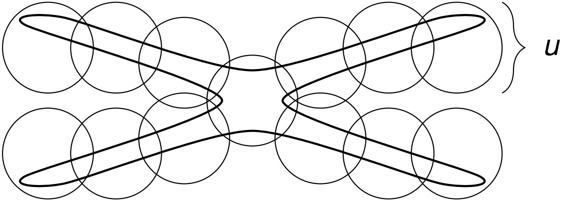
$$N(\Sigma_k, \|\cdot\|_A, u)$$



$$\Sigma_k \quad = \{k\text{-sparse } x \mid \|x\|_2 \leq 1\}$$

Covering Number Bound

$$N(\Sigma_k, \|\cdot\|_A, u) \leq N(B_1, \|\cdot\|_A, u/\sqrt{k})$$



$$\begin{aligned} \Sigma_k &= \{k\text{-sparse } x \mid \|x\|_2 \leq 1\} \\ \subset \sqrt{k}B_1 &= \{x \mid \|x\|_1 \leq \sqrt{k}\} \end{aligned}$$

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_2, u)$$

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_2, u) \lesssim \begin{cases} (1/u)^{O(n)} & \text{by an easy volume argument} \\ \end{cases}$$

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_2, u) \lesssim \begin{cases} (1/u)^{O(n)} & \text{by an easy volume argument} \\ n^{O(1/u^2)} & \text{trickier; next few slides} \end{cases}$$

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_2, u) \lesssim \begin{cases} (1/u)^{O(n)} & \text{by an easy volume argument} \\ n^{O(1/u^2)} & \text{trickier; next few slides} \end{cases}$$

- Latter bound is better when $u \gg 1/\sqrt{n}$.

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_2, u) \lesssim \begin{cases} (1/u)^{O(n)} & \text{by an easy volume argument} \\ n^{O(1/u^2)} & \text{trickier; next few slides} \end{cases}$$

- Latter bound is better when $u \gg 1/\sqrt{n}$.
- *Maurey's empirical method*: generalizes to arbitrary norms

Covering number bound

$$N(B_1, \|\cdot\|_A, u)$$

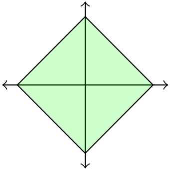
- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

$$N(B_1, \|\cdot\|_A, u) \lesssim \begin{cases} (\sqrt{B}/u)^{O(n)} & \text{by an easy volume argument} \\ n^{O(B/u^2)} & \text{trickier; next few slides} \end{cases}$$

- Latter bound is better when $u \gg 1/\sqrt{n}$.
- *Maurey's empirical method*: generalizes to arbitrary norms

Covering Number Bound

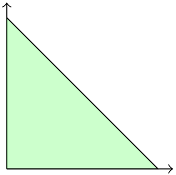
Maurey's empirical method



- How many balls of radius u required to cover B_1 ?

Covering Number Bound

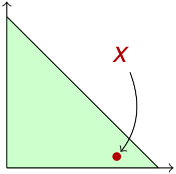
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?

Covering Number Bound

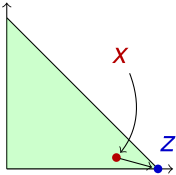
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.

Covering Number Bound

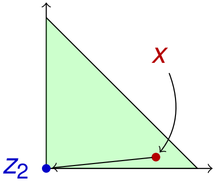
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \dots, z_t$ be i.i.d. randomized roundings of x to simplex.

Covering Number Bound

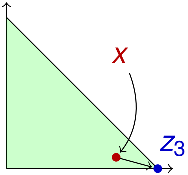
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.

Covering Number Bound

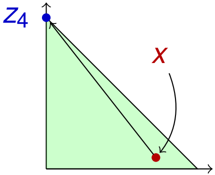
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \dots, z_t$ be i.i.d. randomized roundings of x to simplex.

Covering Number Bound

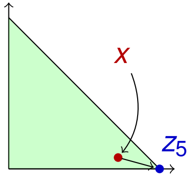
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \dots, z_t$ be i.i.d. randomized roundings of x to simplex.

Covering Number Bound

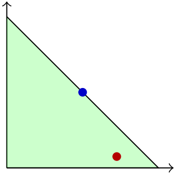
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.

Covering Number Bound

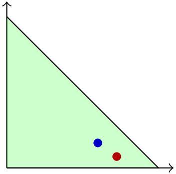
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

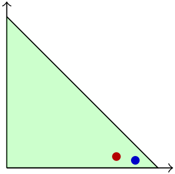
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

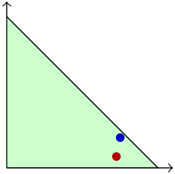
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

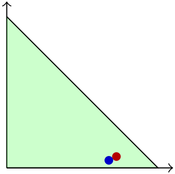
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

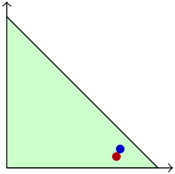
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

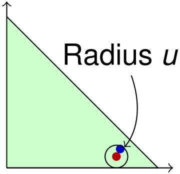
Maurey's empirical method



- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.

Covering Number Bound

Maurey's empirical method

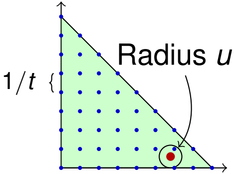


- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \dots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - x\|] \leq u.$$

Covering Number Bound

Maurey's empirical method



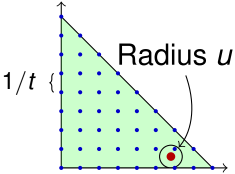
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - x\|] \leq u.$$

- All x lie within u of at least one possible $\mathbf{z}$.

Covering Number Bound

Maurey's empirical method



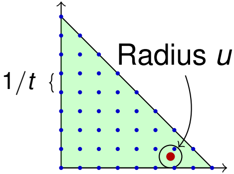
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - x\|] \leq u.$$

- All x lie within u of at least one possible $\mathbf{z}$.
  - Then $N(B_1, \|\cdot\|, u) \leq$ number of $\mathbf{z}$

Covering Number Bound

Maurey's empirical method



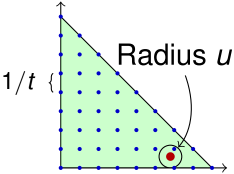
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - x\|] \leq u.$$

- All x lie within u of at least one possible $\mathbf{z}$.
  - Then $N(B_1, \|\cdot\|, u) \leq$ number of $\mathbf{z} \leq (n+1)^t$.
  - Only $(n+1)^t$ possible tuples $(z_1, \ldots, z_t) \implies \mathbf{z}$.

Covering Number Bound

Maurey's empirical method



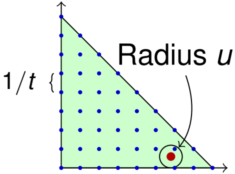
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - \boldsymbol{x}\|] \le \boldsymbol{u}.$$

- All x lie within u of at least one possible $\mathbf{z}$.
    - Then $N(B_1, \|\cdot\|, u) \le$ number of $\mathbf{z} \le (n+1)^t$.
    - Only $(n+1)^t$ possible tuples $(z_1, \ldots, z_t) \implies \mathbf{z}$.

Covering Number Bound

Maurey's empirical method



Will show: $\mathbb{E}[\|\mathbf{z} - x\|_A] \leq \sqrt{B/t}$

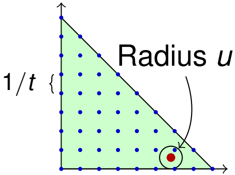
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - \mathbf{x}\|] \leq \mathbf{u}.$$

- All x lie within u of at least one possible $\mathbf{z}$.
  - Then $N(B_1, \|\cdot\|, u) \leq$ number of $\mathbf{z} \leq (n+1)^t$.
  - Only $(n+1)^t$ possible tuples $(z_1, \ldots, z_t) \implies \mathbf{z}$.

Covering Number Bound

Maurey's empirical method



Will show: $\mathbb{E}[\|\mathbf{z} - x\|_A] \leq \sqrt{B/t} \implies N(T, \|\cdot\|_A, u) \leq n^{B/u^2}$

- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

$$\mathbb{E}[\|\mathbf{z} - \boldsymbol{x}\|] \leq \boldsymbol{u}.$$

- All x lie within u of at least one possible $\mathbf{z}$.
  - Then $N(B_1, \|\cdot\|, u) \leq$ number of $\mathbf{z} \leq (n+1)^t$.
  - Only $(n+1)^t$ possible tuples $(z_1, \ldots, z_t) \implies \mathbf{z}$.

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A]$$

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A] \lesssim \mathbb{E}[\|\frac{1}{t}\sum g_i z_i\|_A]$$

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\begin{aligned}\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A] &\lesssim \mathbb{E}[\|\frac{1}{t}\sum g_i z_i\|_A] \\ &=: \frac{1}{\sqrt{t}}\mathbb{E}[\|\mathbf{g}\|_A]\end{aligned}$$

where $\mathbf{g} \in \mathbb{R}^n$ has

$$\mathbf{g}_j \sim N(0, \frac{\text{number of } z_i \text{ at } e_j}{t})$$

independently in each coordinate.

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A] \lesssim \mathbb{E}[\|\frac{1}{t}\sum g_i z_i\|_A]$$
$$=: \frac{1}{\sqrt{t}}\mathbb{E}[\|\mathbf{g}\|_A]$$

where $\mathbf{g} \in \mathbb{R}^n$ has

$$\mathbf{g}_j \sim N(0, \frac{\text{number of } z_i \text{ at } e_j}{t})$$

independently in each coordinate.

- Hence $\mathbb{E}[\|\mathbf{g}\|_2^2] = (\text{fraction of } z_i \text{ that are nonzero}) \leq 1$.

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A] \lesssim \mathbb{E}[\|\frac{1}{t}\sum g_i z_i\|_A]$$
$$=: \frac{1}{\sqrt{t}}\mathbb{E}[\|\mathbf{g}\|_A]$$

where $\mathbf{g} \in \mathbb{R}^n$ has

$$\mathbf{g}_j \sim N(0, \frac{\text{number of } z_i \text{ at } e_j}{t})$$

independently in each coordinate.

- Hence $\mathbb{E}[\|\mathbf{g}\|_2^2] =$ (fraction of z_i that are nonzero) ≤ 1.
- Goal: $\mathbb{E}[\|\mathbf{g}\|_A] \leq \sqrt{B}$.

Covering Number Bound

- Goal: $\mathbb{E}[\|\mathbf{z} - x\|_A] \lesssim \sqrt{B/t}$.
- Symmetrize!

$$\mathbb{E}[\|\frac{1}{t}\sum z_i - x\|_A] \lesssim \mathbb{E}[\|\frac{1}{t}\sum g_i z_i\|_A]$$
$$=: \frac{1}{\sqrt{t}}\mathbb{E}[\|\mathbf{g}\|_A]$$

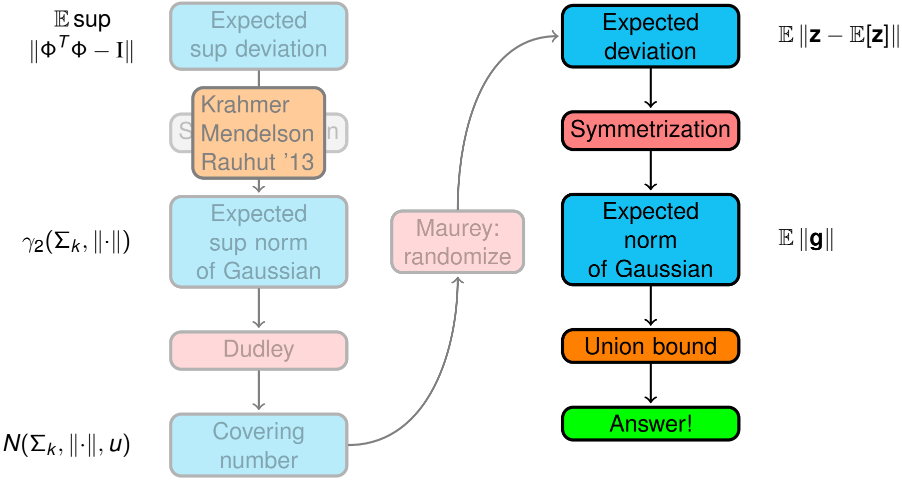
where $\mathbf{g} \in \mathbb{R}^n$ has

$$\mathbf{g}_j \sim N(0, \frac{\text{number of } z_i \text{ at } e_j}{t})$$

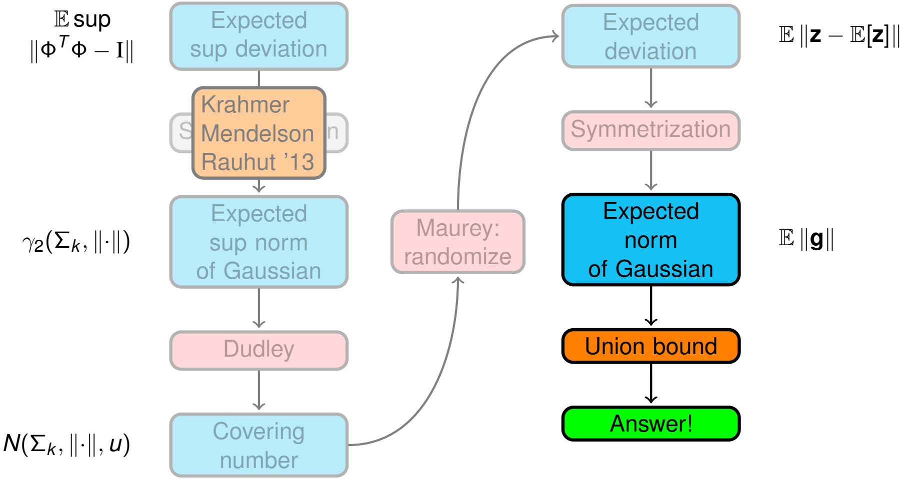
independently in each coordinate.

- Hence $\mathbb{E}[\|\mathbf{g}\|_2^2] =$ (fraction of z_i that are nonzero) ≤ 1.
- Goal: $\mathbb{E}[\|\mathbf{g}\|_A] \leq \sqrt{B}$.
- (Note: $\mathbb{E}[\|\mathbf{g}\|_2] \leq 1 \implies N(B_1, \ell_2, u) \leq n^{1/u^2}$.)

Progress


$\mathbb{E}\sup$
$\|\Phi^T\Phi - \mathrm{I}\|$
Expected sup deviation
S
n
Krahmer Mendelson Rauhut '13
$\gamma_2(\Sigma_k, \|\cdot\|)$
Expected sup norm of Gaussian
Dudley
$N(\Sigma_k, \|\cdot\|, u)$
Covering number
Maurey: randomize
Expected deviation
$\mathbb{E}\,\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$
Symmetrization
Expected norm of Gaussian
$\mathbb{E}\,\|\mathbf{g}\|$
Union bound
Answer!

Progress



$\mathbb{E}$ sup $\|\Phi^T\Phi - I\|$

$\gamma_2(\Sigma_k, \|\cdot\|)$

$N(\Sigma_k, \|\cdot\|, u)$

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$

$\mathbb{E}\|\mathbf{g}\|$

Expected sup deviation

Krahmer Mendelson Rauhut ’13

Expected sup norm of Gaussian

Dudley

Covering number

Maurey: randomize

Expected deviation

Symmetrization

Expected norm of Gaussian

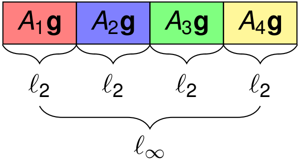
Union bound

Answer!

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



$A_1\mathbf{g}$ $A_2\mathbf{g}$ $A_3\mathbf{g}$ $A_4\mathbf{g}$

ℓ_2 ℓ_2 ℓ_2 ℓ_2

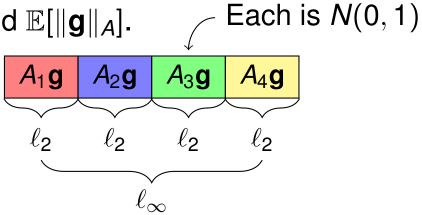
ℓ_∞

$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
= subset of $\widehat{\mathbf{g}}$



$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

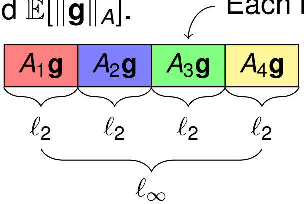
- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$

= subset of $\widehat{\mathbf{g}}$


Each is $N(0,1)$

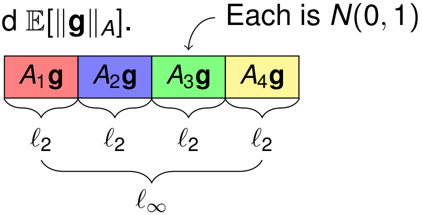
$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0, 1)$.
- Rudelson-Vershynin: $B = 1$

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



Each is $N(0, 1)$

$A_1\mathbf{g}$ $A_2\mathbf{g}$ $A_3\mathbf{g}$ $A_4\mathbf{g}$

ℓ_2 ℓ_2 ℓ_2 ℓ_2

ℓ_∞

$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

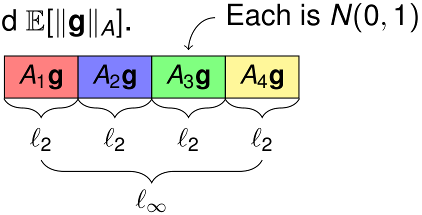
- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0, 1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$

$= \text{subset of } \widehat{\mathbf{g}}$



Each is $N(0,1)$

$A_1\mathbf{g}$ $A_2\mathbf{g}$ $A_3\mathbf{g}$ $A_4\mathbf{g}$

ℓ_2 ℓ_2 ℓ_2 ℓ_2

ℓ_∞

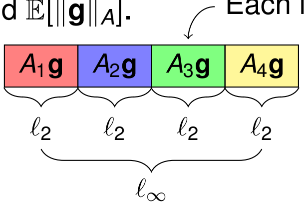
$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the log n loss in their bound.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



Each is $N(0,1)$

$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the $\log n$ loss in their bound.
- If the $\widehat{\mathbf{g}}_i$ were independent:

$$\|A_i\mathbf{g}\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

$\uparrow$

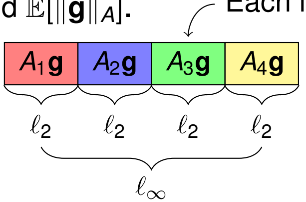
Lipschitz concentration
(just like $\sqrt{n} + \sqrt{\log(1/\delta)}$ in tutorial)

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

Each is $N(0,1)$

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the $\log n$ loss in their bound.
- If the $\widehat{\mathbf{g}}_i$ were independent:

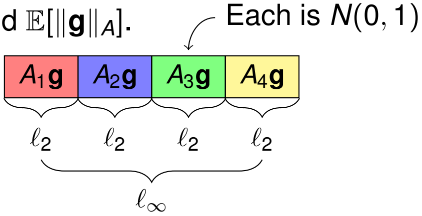
$$\|A_i\mathbf{g}\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

- Would get $\mathbb{E}[\|\mathbf{g}\|_A] \lesssim \sqrt{B}$

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



Each is $N(0,1)$

$A_1\mathbf{g}$ $A_2\mathbf{g}$ $A_3\mathbf{g}$ $A_4\mathbf{g}$

ℓ_2 ℓ_2 ℓ_2 ℓ_2

ℓ_∞

$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the log n loss in their bound.
- If the $\widehat{\mathbf{g}}_i$ were independent:

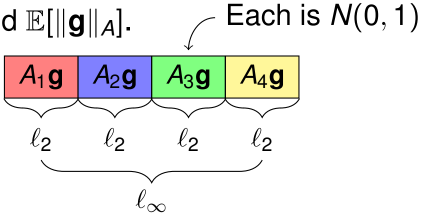
$$\|A_i\mathbf{g}\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

- Would get $\mathbb{E}[\|\mathbf{g}\|_A] \lesssim \sqrt{B}$ so union bound just loses a constant.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the $\log n$ loss in their bound.
- If the $\widehat{\mathbf{g}}_i$ were independent:

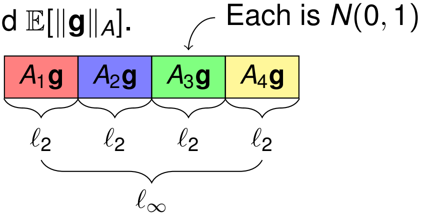
$$\|A_i\mathbf{g}\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

- Would get $\mathbb{E}[\|\mathbf{g}\|_A] \lesssim \sqrt{B}$ so union bound just loses a constant.
- They're not independent... but the A_i satisfy "very weak" RIP.

Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

$A\mathbf{g}$
$= \text{subset of } \widehat{\mathbf{g}}$



Each is $N(0,1)$

$$\underbrace{\underbrace{A_1\mathbf{g}}_{\ell_2}\;\underbrace{A_2\mathbf{g}}_{\ell_2}\;\underbrace{A_3\mathbf{g}}_{\ell_2}\;\underbrace{A_4\mathbf{g}}_{\ell_2}}_{\ell_\infty}$$

$\mathbf{g} \in \mathbb{R}^n$ has Gaussian coordinates, k-sparse, total variance 1.

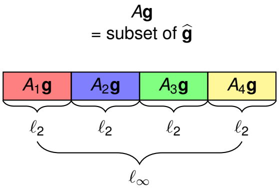
- Each coordinate $\widehat{\mathbf{g}}_j = F_j\mathbf{g} \sim N(0,1)$.
- Rudelson-Vershynin: $B = 1$
  - Naive: $\|\widehat{\mathbf{g}}\|_\infty \lesssim \sqrt{\log n}$.
  - Gives the $\log n$ loss in their bound.
- If the $\widehat{\mathbf{g}}_i$ were independent:

$$\|A_i\mathbf{g}\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

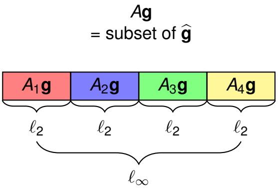
- Would get $\mathbb{E}[\|\mathbf{g}\|_A] \lesssim \sqrt{B}$ so union bound just loses a constant.
- They're not independent... but the A_i satisfy "very weak" RIP.
  - Bound $\|A_i\mathbf{g}\|_2$ using $\|\mathbf{g}\|_2$, which has independent entries.

Bounding the norm (by example)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.



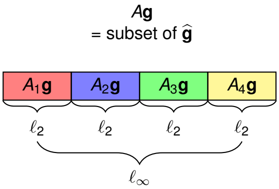
Bounding the norm (by example)



Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.

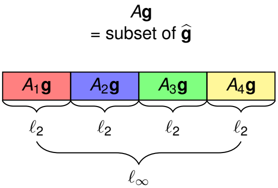
Bounding the norm (by example)



Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.

Bounding the norm (by example)



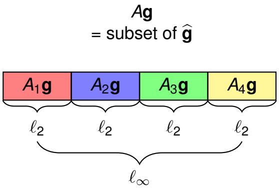
Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \dots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

(the "large *t*" limit).

Bounding the norm (by example)



$A\mathbf{g}$

$= \text{subset of } \widehat{\mathbf{g}}$

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

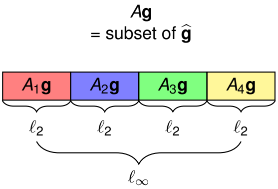
- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

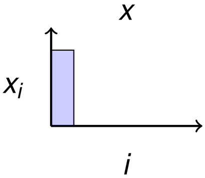
  (the "large t" limit).
- Two examples:

Bounding the norm (by example)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.



- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \dots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume
$$\mathbf{g}_i \sim N(0, x_i)$$
(the “large *t*” limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.



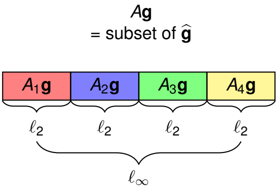
Bounding the norm (by example)

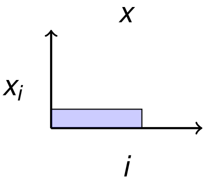
Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \dots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

  (the “large t” limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.





Bounding the norm (by example)

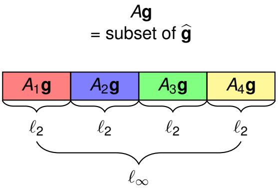
Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

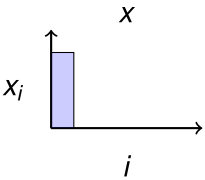
- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

  (the “large t” limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.

Example (1):





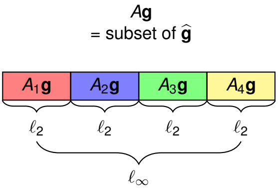
Bounding the norm (by example)

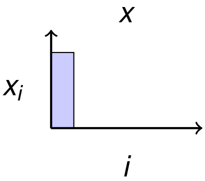
Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

  (the "large t" limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.

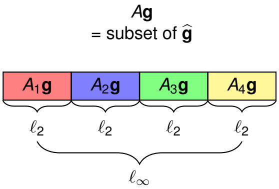




Example (1):

- $\|x\|_1 \ll 1$.

Bounding the norm (by example)

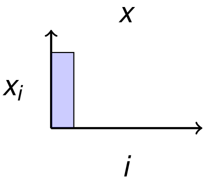


Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

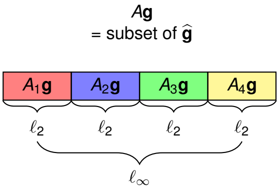
  (the "large t" limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.



Example (1):

- $\|x\|_1 \ll 1$.
- Fourier transform $\widehat{\mathbf{g}}$ is Gaussian with variance $1/\sqrt{k}$.

Bounding the norm (by example)

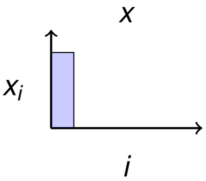


Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

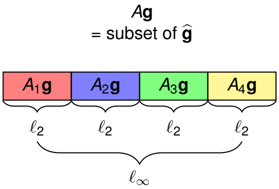
  (the "large t" limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.



Example (1):

- $\|x\|_1 \ll 1$.
- Fourier transform $\widehat{\mathbf{g}}$ is Gaussian with variance $1/\sqrt{k}$.
- $\|\widehat{\mathbf{g}}\|_\infty \lesssim k^{-1/4} \cdot \sqrt{\log n} \ll 1$.

Bounding the norm (by example)

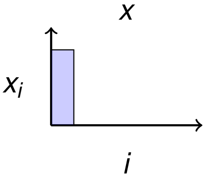


Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \ldots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

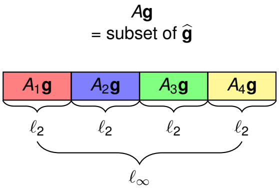
  (the “large t” limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.



Example (1):

- $\|x\|_1 \ll 1$.
- Fourier transform $\widehat{\mathbf{g}}$ is Gaussian with variance $1/\sqrt{k}$.
- $\|\widehat{\mathbf{g}}\|_\infty \lesssim k^{-1/4} \cdot \sqrt{\log n} \ll 1$.
- Hence $\|A_i\mathbf{g}\|_2 \ll \sqrt{B}$ for all i.

Bounding the norm (by example)

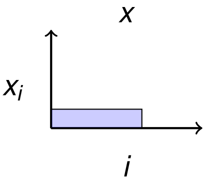


Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \longrightarrow z_1, \dots, z_t \longrightarrow \mathbf{g}$.
- The full proof is complicated.
- Will assume

$$\mathbf{g}_i \sim N(0, x_i)$$

  (the "large t" limit).
- Two examples:
  1. Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  2. Very spread out: $x_1 = \cdots = x_k = 1/k$.



Example (1):

- $\|x\|_1 \ll 1$.
- Fourier transform $\widehat{\mathbf{g}}$ is Gaussian with variance $1/\sqrt{k}$.
- $\|\widehat{\mathbf{g}}\|_\infty \lesssim k^{-1/4} \cdot \sqrt{\log n} \ll 1$.
- Hence $\|A_i\mathbf{g}\|_2 \ll \sqrt{B}$ for all i.

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i\mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

$$\underbrace{\boxed{A_i\mathbf{g}}}_{\ell_2}$$

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i \mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

$$\underbrace{\boxed{A_i \mathbf{g}}}_{\ell_2}$$

- Then with high probability,

$$\|A_i \mathbf{g}\|_2 \lesssim \sqrt{B} + C\sqrt{\log n}$$

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i\mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

$$\underbrace{\boxed{A_i\mathbf{g}}}_{\ell_2}$$

- Then with high probability,

$$\|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} + C\sqrt{\log n}$$

- Naive bound:

$$\|A_i\|_{RIP} \leq \|A_i\|_F = \sqrt{Bk} \implies C \leq \sqrt{B}$$

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i\mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

$$\underbrace{\boxed{A_i\mathbf{g}}}_{\ell_2}$$

- Then with high probability,

$$\|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} + C\sqrt{\log n}$$

- Naive bound:

$$\|A_i\|_{RIP} \leq \|A_i\|_F = \sqrt{Bk} \implies C \leq \sqrt{B}$$

- "Very weak" RIP bound, with $B = \log^c n$:

$$\|A_i\|_{RIP} \lesssim \|A_i\|_F / \sqrt{\log n}$$

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i\mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

$$\underbrace{\boxed{A_i\mathbf{g}}}_{\ell_2}$$

- Then with high probability,

$$\|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} + C\sqrt{\log n}$$

- Naive bound:

$$\|A_i\|_{RIP} \leq \|A_i\|_F = \sqrt{Bk} \implies C \leq \sqrt{B}$$

- “Very weak” RIP bound, with $B = \log^c n$:

$$\|A_i\|_{RIP} \lesssim \|A_i\|_F / \sqrt{\log n}$$

- So

$$C \lesssim \sqrt{B/\log n} \implies \|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} \quad \text{w.h.p}$$

Bounding the norm (example 2)

Just want to bound $\mathbb{E}[\|\mathbf{g}\|_A]$ when $\mathbf{g}_i \sim N(0, 1/k)$ for $i \in [k]$.

$$\underbrace{\boxed{A_i\mathbf{g}}}_{\ell_2}$$

- $\|A_i\mathbf{g}\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

- Then with high probability,

$$\|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} + C\sqrt{\log n}$$

- Naive bound:

$$\|A_i\|_{RIP} \leq \|A_i\|_F = \sqrt{Bk} \implies C \leq \sqrt{B}$$

- "Very weak" RIP bound, with $B = \log^c n$:

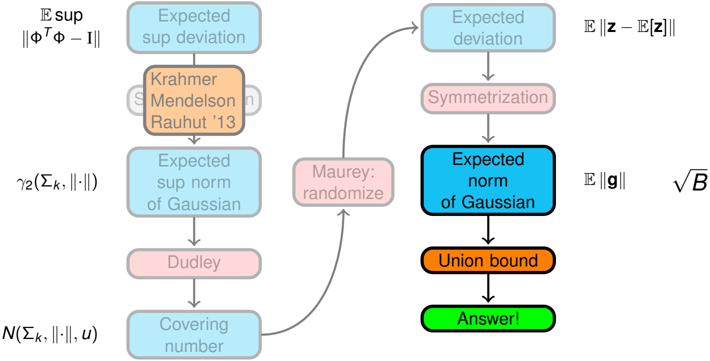
$$\|A_i\|_{RIP} \lesssim \|A_i\|_F / \sqrt{\log n}$$

- So

$$C \lesssim \sqrt{B/\log n} \implies \|A_i\mathbf{g}\|_2 \lesssim \sqrt{B} \quad \text{w.h.p}$$

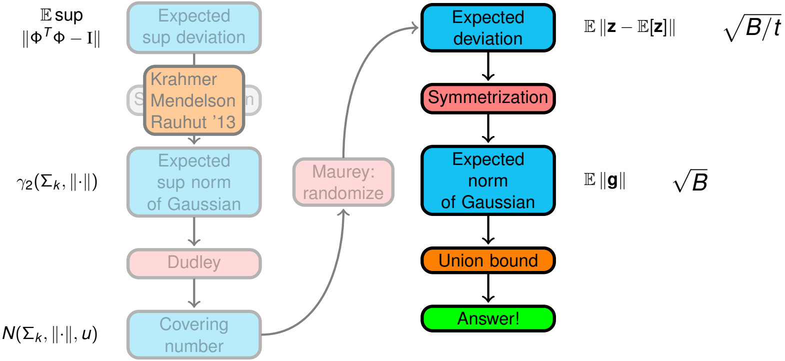
- So $\mathbb{E}\|\mathbf{g}\|_A = \max\|A_i\mathbf{g}\|_2 \lesssim \sqrt{B}$.

Unrolling everything



Union bound just loses a constant factor

Unrolling everything



$\mathbb{E}\sup \|\Phi^T\Phi - I\|$

Expected sup deviation

Krahmer Mendelson Rauhut '13

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

$N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$ $\sqrt{B/t}$

Symmetrization

Expected norm of Gaussian

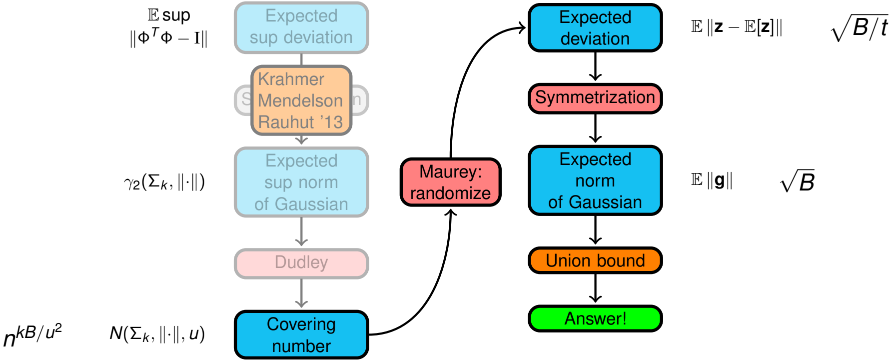
$\mathbb{E}\|\mathbf{g}\|$ $\sqrt{B}$

Union bound

Answer!

Sample mean $\mathbf{z}$ expects to lie within u of $\mathbf{x}$ for $t \geq B/u^2$

Unrolling everything



$\mathbb{E}\sup \|\Phi^T\Phi - I\|$

Expected sup deviation

Krahmer Mendelson Rauhut '13

$\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

n^{kB/u^2} $N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$ $\sqrt{B/t}$

Symmetrization

Expected norm of Gaussian

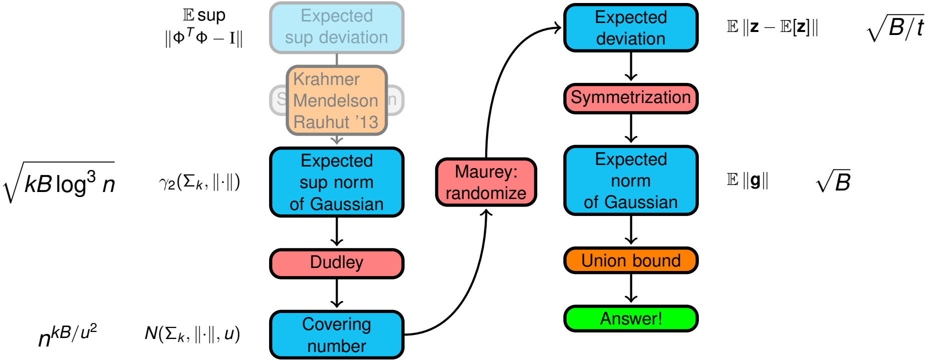
$\mathbb{E}\|\mathbf{g}\|$ $\sqrt{B}$

Union bound

Answer!

Covering number of B_1 is $(n+1)^{B/u^2}$

Unrolling everything



$\mathbb{E}\sup \|\Phi^T\Phi - I\|$

Expected sup deviation

Krahmer Mendelson Rauhut '13

$\sqrt{kB\log^3 n}$ $\gamma_2(\Sigma_k, \|\cdot\|)$

Expected sup norm of Gaussian

Dudley

n^{kB/u^2} $N(\Sigma_k, \|\cdot\|, u)$

Covering number

Maurey: randomize

Expected deviation

$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$ $\sqrt{B/t}$

Symmetrization

Expected norm of Gaussian

$\mathbb{E}\|\mathbf{g}\|$ $\sqrt{B}$

Union bound

Answer!

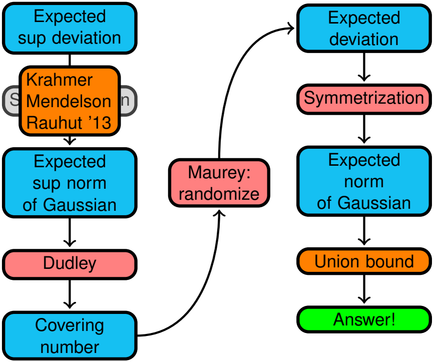
Entropy integral is $\sqrt{kB\log^3 n}$

Unrolling everything

$\sqrt{\frac{k \log^3 n}{m}}$ $\mathbb{E}\sup \|\Phi^T\Phi - \mathrm{I}\|$

$\sqrt{kB \log^3 n}$ $\gamma_2(\Sigma_k, \|\cdot\|)$

n^{kB/u^2} $N(\Sigma_k, \|\cdot\|, u)$



$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$ $\sqrt{B/t}$

$\mathbb{E}\|\mathbf{g}\|$ $\sqrt{B}$

$$\text{RIP constant } \epsilon \lesssim \sqrt{\frac{k \log^3 n}{m}}$$

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
  - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
  - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

  - Generic chaining: tight but harder to use. [Fernique, Talagrand]

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
  - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

  - Generic chaining: tight but harder to use. [Fernique, Talagrand]
- For JL, is the extra $\log^3 n$ necessary?
  - [Krahmer-Ward] only needs "model-based" RIP; could save $\log n$.

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
  - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

  - Generic chaining: tight but harder to use. [Fernique, Talagrand]
- For JL, is the extra $\log^3 n$ necessary?
  - [Krahmer-Ward] only needs "model-based" RIP; could save $\log n$.
- Any other constructions?

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
  - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

  - Generic chaining: tight but harder to use. [Fernique, Talagrand]
- For JL, is the extra $\log^3 n$ necessary?
  - [Krahmer-Ward] only needs "model-based" RIP; could save $\log n$.
- Any other constructions?
- Lots of cool techniques in the field; can we use them elsewhere?

Summary and Open Questions

- We get fast RIP matrices with $O(k \log^3 n)$ rows.
- Is the extra $\log^2 n$ necessary?
    - Loss seems to be from Dudley's entropy integral:

$$\sup \sum \leq \sum \sup$$

    - Generic chaining: tight but harder to use. [Fernique, Talagrand]
- For JL, is the extra $\log^3 n$ necessary?
    - [Krahmer-Ward] only needs "model-based" RIP; could save $\log n$.
- Any other constructions?
- Lots of cool techniques in the field; can we use them elsewhere?

Thanks!

Thoughts on loss

- Recall that

$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$

Thoughts on loss

- Recall that

$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$

- So the entropy integral gives

$$\gamma_2(B_1, \ell_2) \leq \log^{3/2} n.$$

Thoughts on loss

- Recall that
$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \leq \log^{3/2} n.$$
- But the associated Gaussian process is just:
$$\gamma_2(B_1, \ell_2) = \mathbb{E} \sup_{x \in B_1} \langle g, x \rangle$$

for $g \sim N(0, I_n)$. (Check: $\mathbb{E}[(\langle g, x\rangle - \langle g, y\rangle)^2] = \|x - y\|_2^2$)

Thoughts on loss

- Recall that
$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \leq \log^{3/2} n.$$
- But the associated Gaussian process is just:
$$\gamma_2(B_1, \ell_2) = \mathbb{E} \sup_{x \in B_1} \langle g, x \rangle$$
for $g \sim N(0, I_n)$. (Check: $\mathbb{E}[(\langle g, x \rangle - \langle g, y \rangle)^2] = \|x - y\|_2^2$)
- We can compute this directly:
$$\mathbb{E} \sup_{x \in B_1} \langle g, x \rangle = \|g\|_\infty = \sqrt{\log n}.$$

Thoughts on loss

- Recall that
$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \leq \log^{3/2} n.$$
- But the associated Gaussian process is just:
$$\gamma_2(B_1, \ell_2) = \mathbb{E} \sup_{x \in B_1} \langle g, x \rangle$$
for $g \sim N(0, I_n)$. (Check: $\mathbb{E}[(\langle g, x\rangle - \langle g, y\rangle)^2] = \|x - y\|_2^2$)
- We can compute this directly:
$$\mathbb{E} \sup_{x \in B_1} \langle g, x \rangle = \|g\|_\infty = \sqrt{\log n}.$$
- Generic chaining: there exists a partition $A_1, A_2, \ldots$ such that
$$\gamma_2 \simeq \sup_x \sum \sqrt{\log |A_{i+1}|} d(x, A_i)$$

Thoughts on loss

- Recall that
$$N(B_1, \ell_2, u) \leq n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \leq \log^{3/2} n.$$
- But the associated Gaussian process is just:
$$\gamma_2(B_1, \ell_2) = \mathbb{E} \sup_{x \in B_1} \langle g, x \rangle$$
for $g \sim N(0, I_n)$. (Check: $\mathbb{E}[(\langle g, x \rangle - \langle g, y \rangle)^2] = \|x - y\|_2^2$)
- We can compute this directly:
$$\mathbb{E} \sup_{x \in B_1} \langle g, x \rangle = \|g\|_\infty = \sqrt{\log n}.$$
- Generic chaining: there exists a partition $A_1, A_2, \ldots$ such that
$$\gamma_2 \simeq \sup_x \sum \sqrt{\log |A_{i+1}|} d(x, A_i)$$
- Dudley: choose A_i so $\sup d(x, A_i) \leq \sigma_1 / 2^i$.

Covering Number Bound

Maurey's empirical method

- Answer is n^t, where t is such that

$$E := \mathbb{E}[\|\frac{1}{t}\sum z_i - x\|] \leq u.$$

Covering Number Bound

Maurey's empirical method

- Answer is n^t, where t is such that

$$E := \mathbb{E}[\|\frac{1}{t}\sum z_i - \overset{\mathbb{E}[\frac{1}{t}\sum z_i]}{\overset{\downarrow}{x}}\|] \leq u.$$

Covering Number Bound

Maurey's empirical method

- Answer is n^t, where t is such that

$$E := \mathbb{E}[\|\frac{1}{t}\sum z_i - \overset{\mathbb{E}[\frac{1}{t}\sum z_i]}{x}\|] \leq u.$$

- Symmetrize:

$$E \lesssim \frac{1}{t}\mathbb{E}[\|\sum g_i z_i\|]$$

for $g_i \sim N(0,1)$ i.i.d.

Covering Number Bound

Maurey's empirical method

- Answer is n^t, where t is such that

$$E := \mathbb{E}[\|\frac{1}{t}\sum z_i - \overset{\mathbb{E}[\frac{1}{t}\sum z_i]}{x}\|] \leq u.$$

- Symmetrize:

$$E \lesssim \frac{1}{t}\mathbb{E}[\|\sum g_i z_i\|]$$

  for $g_i \sim N(0,1)$ i.i.d.
- Then $g := \sum g_i z_i$ is an independent Gaussian in each coordinate.

Covering Number Bound

Maurey's empirical method

- Answer is n^t, where t is such that

$$E := \mathbb{E}[\|\frac{1}{t}\sum z_i - \overset{\mathbb{E}[\frac{1}{t}\sum z_i]}{\overset{\downarrow}{x}}\|] \leq u.$$

- Symmetrize:

$$E \lesssim \frac{1}{t}\mathbb{E}[\|\sum g_i z_i\|]$$

  for $g_i \sim N(0,1)$ i.i.d.
- Then $g := \sum g_i z_i$ is an independent Gaussian in each coordinate.
- In ℓ_2,

$$\frac{1}{t}\mathbb{E}[\|g\|_2] \leq \frac{1}{t}\mathbb{E}[\|g\|_2^2]^{1/2} = \frac{\sqrt{\text{number nonzero } z_i}}{t} \leq \frac{1}{\sqrt{t}}.$$

  giving an $n^{O(1/u^2)}$ bound.

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \leq \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \dots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \leq \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1$$

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \leq \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1$$

  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{large}\|_1$ if spread out.

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into "large" and "small" coordinates.

$$\mathcal{G}(x) \leq \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1$$

  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{large}\|_1$ if spread out.
  - $k/(\log^2 n)$ of value $(\log n)/k$

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \le \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1 \le 1/\log n.$$

  - Given $\|x\|_2^2 \le 1/k$, maximal $\|x_{large}\|_1$ if spread out.
  - $k/(\log^2 n)$ of value $(\log n)/k$

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \le \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1 \le 1/\log n.$$

  - Given $\|x\|_2^2 \le 1/k$, maximal $\|x_{large}\|_1$ if spread out.
  - $k/(\log^2 n)$ of value $(\log n)/k$
  - Absorbs the loss from union bound.

Bounding the norm in our case (part 1)

- $x \in \Sigma_k/\sqrt{k} \subset B_1$ rounded to $z_1, \ldots, z_t$ symmetrized to g.

$$\mathcal{G}(x) = \mathbb{E}_{z,g}\|g\|_A$$

- First: split x into “large” and “small” coordinates.

$$\mathcal{G}(x) \leq \mathcal{G}(x_{large}) + \mathcal{G}(x_{small})$$

- x_{large}: Locations where $x_i > (\log n)/k$
  - Bound:

$$\|x_{large}\|_1 \leq 1/\log n.$$

  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{large}\|_1$ if spread out.
  - $k/(\log^2 n)$ of value $(\log n)/k$
  - Absorbs the loss from union bound.
- So can focus on $\|x\|_\infty < (\log n)/k$.

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \dots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2$

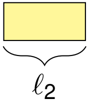
Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

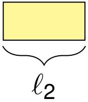
$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



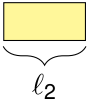
- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t}$$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\# z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



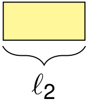
- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n / t}$$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



- Naive bound:

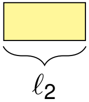
$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n/t}$$

- “Very weak” RIP bound:

$$\|A_i\|_{RIP} \lesssim \log^4 n(\sqrt{B} + \sqrt{k})$$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor



$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$

- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty/t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n/t}$$

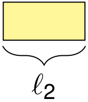
- "Very weak" RIP bound: for some $B = \log^c n$,

$$\|A_i\|_{RIP} \lesssim \log^4 n(\sqrt{B} + \sqrt{k}) \leq \|A_i\|_F / \log n.$$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n/t}$$

- "Very weak" RIP bound: for some $B = \log^c n$,

$$\|A_i\|_{RIP} \lesssim \log^4 n(\sqrt{B} + \sqrt{k}) \leq \|A_i\|_F / \log n.$$

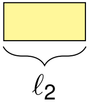
- Gives

$$C \lesssim \sqrt{B/(t \log n)}$$

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \dots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n / t}$$

- “Very weak” RIP bound: for some $B = \log^c n$,

$$\|A_i\|_{RIP} \lesssim \log^4 n(\sqrt{B} + \sqrt{k}) \leq \|A_i\|_F / \log n.$$

- Gives

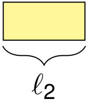
$$C \lesssim \sqrt{B/(t \log n)}$$

- So with high probability, $\|A_i g\|_2 \lesssim \sqrt{B/t} + C\sqrt{\log n} \lesssim \sqrt{B/t}$.

Bounding the norm in our case (part 2)

- k-sparse x rounded to $z_1, \ldots, z_t$ symmetrized to g.
- $\|x\|_\infty < (\log n)/k$
- $g_i \sim N(0, \sigma_i^2)$ for $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_i g\|_2$ is C-Lipschitz with factor

$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



- Naive bound:

$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty/t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n/t}$$

- “Very weak” RIP bound: for some $B = \log^c n$,

$$\|A_i\|_{RIP} \lesssim \log^4 n(\sqrt{B} + \sqrt{k}) \leq \|A_i\|_F / \log n.$$

- Gives

$$C \lesssim \sqrt{B/(t \log n)}$$

- So with high probability, $\|A_i g\|_2 \lesssim \sqrt{B/t} + C\sqrt{\log n} \lesssim \sqrt{B/t}$.
- So $\mathbb{E}\|g\|_A = \max \|A_i g\|_2 \lesssim \sqrt{B/t}$.