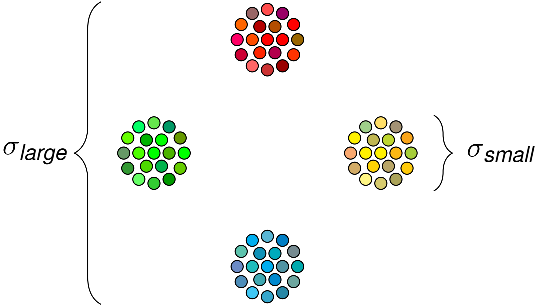
RIP of Subsampled Fourier Matrix

Based off Rudelson-Vershynin

Eric Price

2020-10-27



Outline

1. Introduction
   - Compressive sensing
   - Johnson Lindenstrauss Transforms

Outline

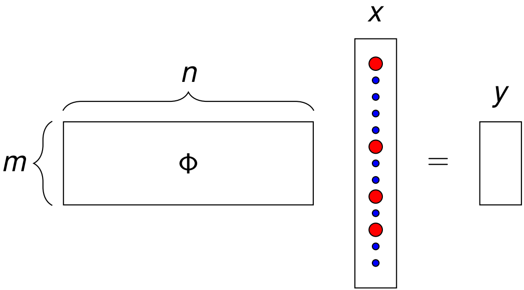
Outline

Outline

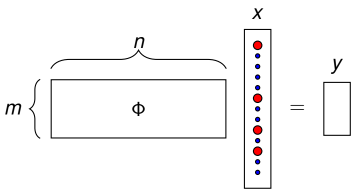
Outline

Compressive Sensing

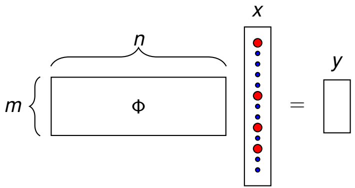
Given: A few linear measurements of an (approximately) k-sparse vector $x \in \mathbb{R}^n$.
Goal: Recover x (approximately).



Compressive Sensing Algorithms: Two Classes



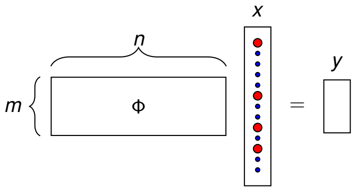
Compressive Sensing Algorithms: Two Classes



Structure-aware

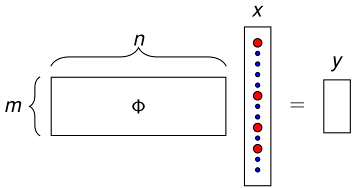
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)

Compressive Sensing Algorithms: Two Classes



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

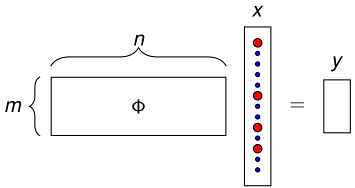
Compressive Sensing Algorithms: Two Classes



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

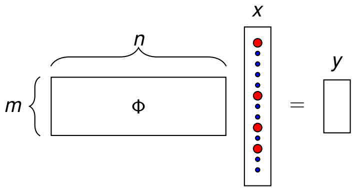
Compressive Sensing Algorithms: Two Classes



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

$\uparrow$
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.

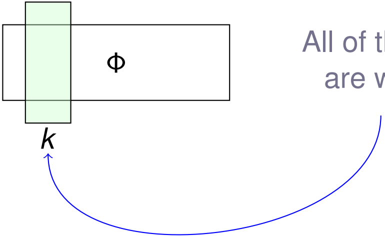
Algorithms for compressive sensing

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

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

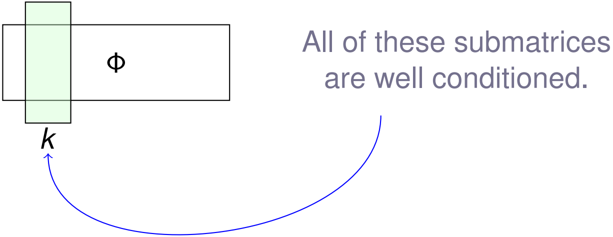
- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....
- For all of these:
  - the time it takes to multiply by Φ or Φ^T is the bottleneck.
  - the *Restricted Isometry Property* is a sufficient condition.

Restricted Isometry Property (RIP)



All of these submatrices
are well conditioned.

Restricted Isometry Property (RIP)



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

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 $\log k \simeq \log n \simeq \log(n/k)$.

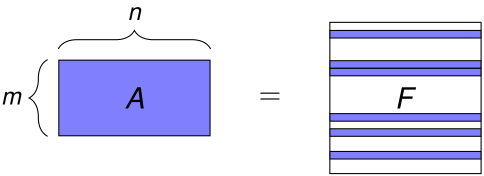
Goals

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.

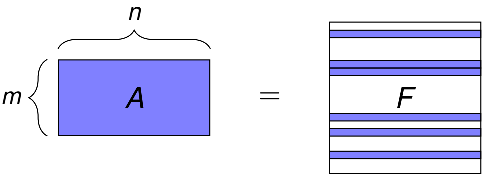
* Talk will assume $n^{0.1} < k < n^{0.9}$, so $\log k \simeq \log n \simeq \log(n/k)$.

An open question



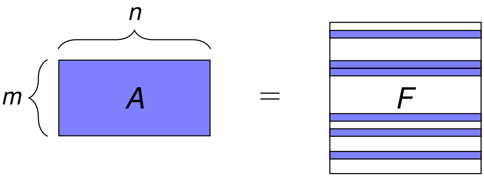
Let A contain random rows from a Fourier matrix.

An open question



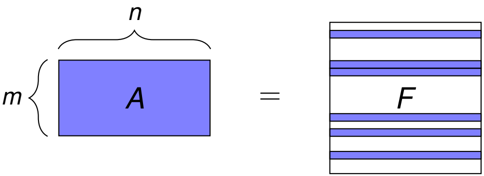
Let A contain random rows from a Fourier matrix.
You can multiply by A in $O(n \log n)$ time.

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?

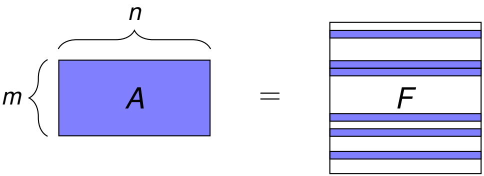
An open question



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?

- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].

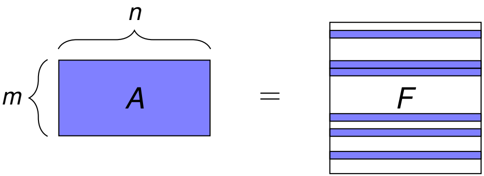
An open question



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?

- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].
    - Today: $m = O(k \log^4 n)$

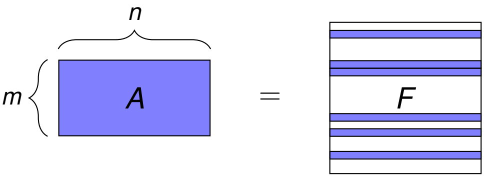
An open question



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?

- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].
  - Today: $m = O(k \log^4 n)$
  - Ideal: $m = O(k \log n)$

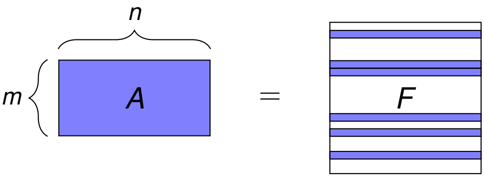
An open question



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?

- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].
  - Today: $m = O(k \log^4 n)$
  - Ideal: $m = O(k \log n)$
  - Subsampled *Hadamard* lower bound: $m = O(k \log n \log k)$ [BLLMR19].

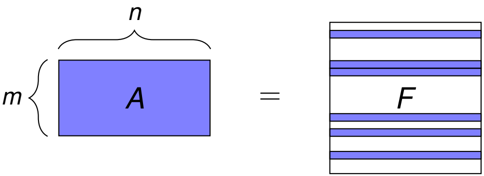
An open question



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?

- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].
  - Today: $m = O(k \log^4 n)$
  - Ideal: $m = O(k \log n)$
  - Subsampled *Hadamard* lower bound: $m = O(k \log n \log k)$ [BLLMR19].

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?

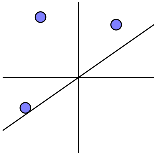
- $m = O(k \log n \log^2 k)$ [CT06,RV08,CGV13,B14,HV15].
    - Today: $m = O(k \log^4 n)$
    - Ideal: $m = O(k \log n)$
    - Subsampled *Hadamard* lower bound: $m = O(k \log n \log k)$ [BLLMR19].

(Related: how about partial circulant matrices?)

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

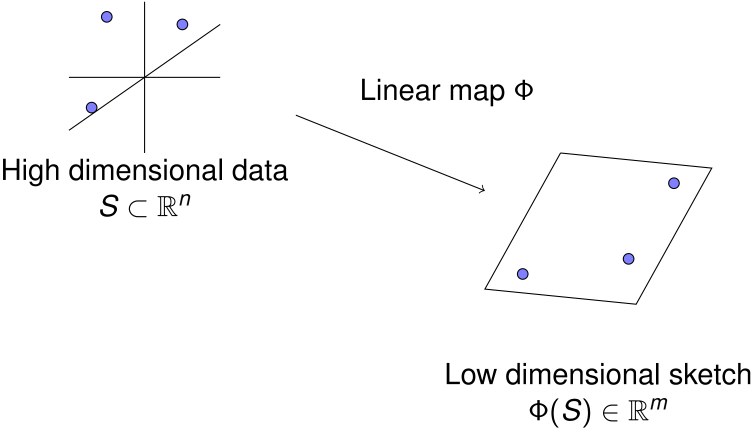
Outline

Another motivation: Johnson Lindenstrauss (JL) Transforms

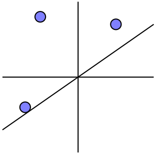


High dimensional data
$S \subset \mathbb{R}^n$

Another motivation: Johnson Lindenstrauss (JL) Transforms

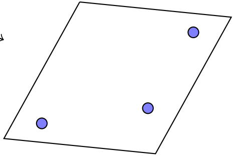


Another motivation:
Johnson Lindenstrauss (JL) Transforms



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

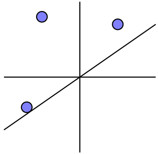
Linear map Φ



Φ preserves the geometry of S

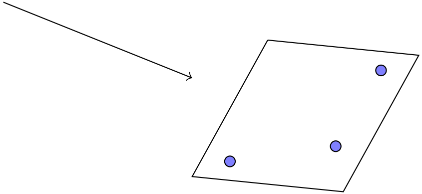
Low dimensional sketch
$\Phi(S) \in \mathbb{R}^m$

Another motivation: Johnson Lindenstrauss (JL) Transforms



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

Linear map Φ

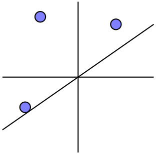


Low dimensional sketch
$\Phi(S) \in \mathbb{R}^m$

Φ preserves the geometry of S

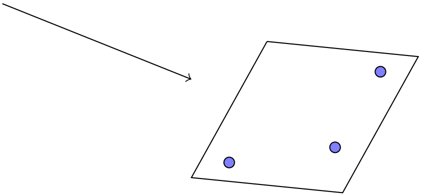
$$(1-\epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\epsilon)\|x\|_2$$

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

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

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

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 $\Rightarrow$ 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.
  - Known results: dimension $O(\frac{1}{\epsilon^2} k \log^3 n)$.

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 $\Rightarrow$ JL.
  - Known results: dimension $O(\frac{1}{\epsilon^2} k \log^3 n)$.
  - $n \log n$ 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 $\Rightarrow$ JL.
  - Known results: dimension $O(\frac{1}{\epsilon^2}k\log^3 n)$.
  - $n \log n$ multiplication time.
- And by [BDDW '08], JL $\Rightarrow$ RIP; so *equivalent*.[1]

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

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

Tools

Tools

Screwdriver

Tools

Screwdriver

Drill

Tools

Screwdriver

Bit sets

Drill

Tools

Screwdriver

Drill

Bit

Tools

Screwdriver

Drill

Bit

Tools

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

Tools

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

Hex shanks

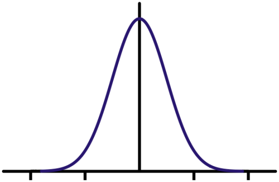
Common interface
for drill bits

Tools

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

Hex shanks

Common interface
for drill bits



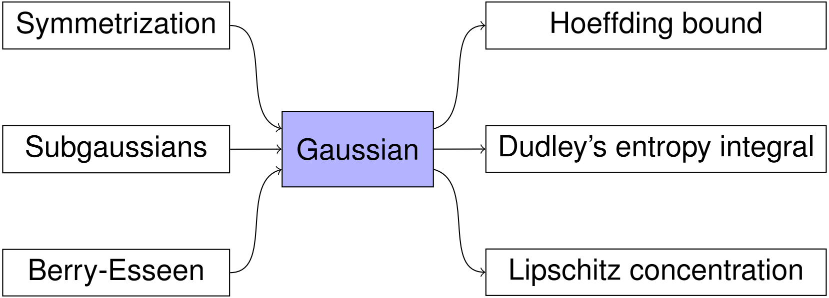
Gaussians

Common interface
for probability

A Probabilist's Toolbox

Convert to Gaussians

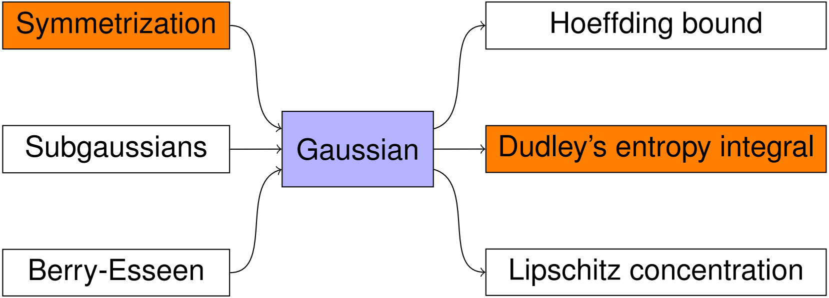
Gaussian concentration



A Probabilist's Toolbox

Convert to Gaussians

Gaussian concentration



Will prove: symmetrization and Dudley's entropy integral.

Outline

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 $\|\cdot\|$,

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

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 $X_1, \dots, X_t$ 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.

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

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

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

Proof.

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

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

Symmetrization

Lemma (Symmetrization)

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

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

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

Proof.

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

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

and apply the triangle inequality. $\square$

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.

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.

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 $X_1, \ldots, X_t$ 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 $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

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

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

Symmetrization

Lemma (Symmetrization)

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

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

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

Proof.

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

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

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

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

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

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

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

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.

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 $G_x - G_y$.

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

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

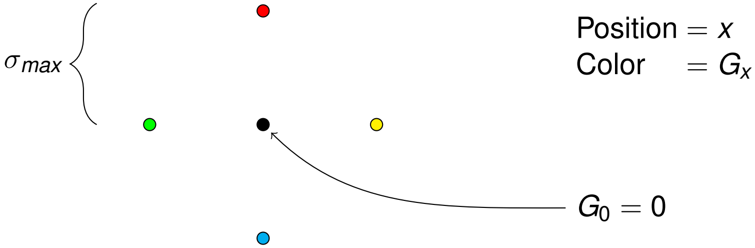
- Depends on the geometry of T.
- Distance: $\|x - y\|$ is standard deviation of $G_x - G_y$.
- In example: $\|(u, v) - (u', v')\| = \|uv^T - u'v'^T\|_F$.

Gaussian Processes

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

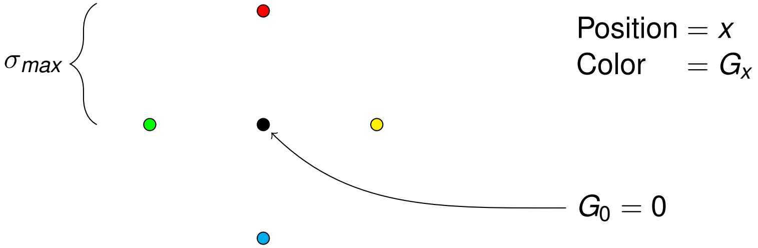
Gaussian Processes

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



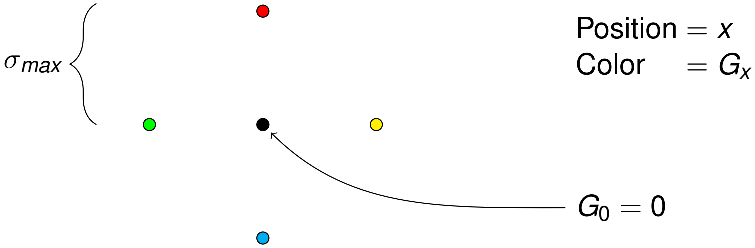
Gaussian Processes

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



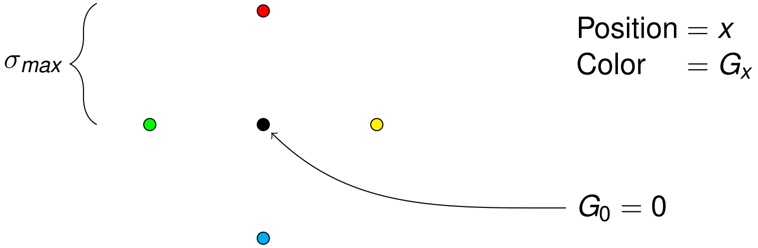
Gaussian Processes

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



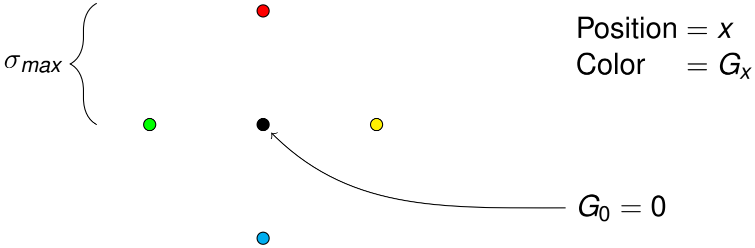
Gaussian Processes

- 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, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.



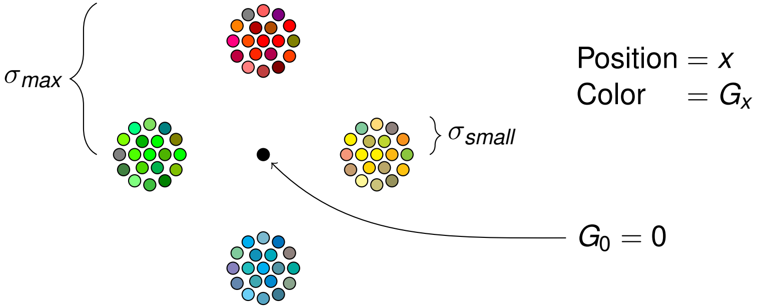
Gaussian Processes

- 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, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.
  - $\mathbb{E}\sup_{x \in T} G_x \lesssim \sigma_{max}\sqrt{\log n}$



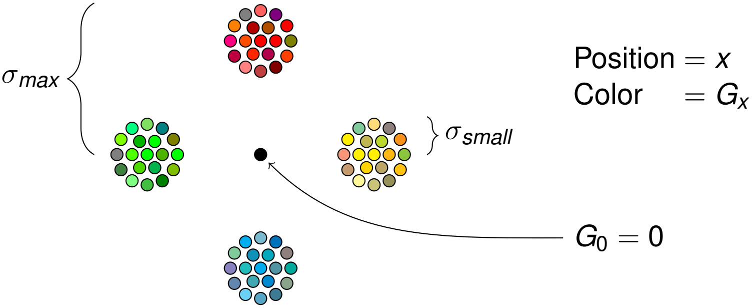
Gaussian Processes

- 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, $G_x \lesssim \sigma_{max}\sqrt{\log n}$.
  - $\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_{max}\sqrt{\log n}$



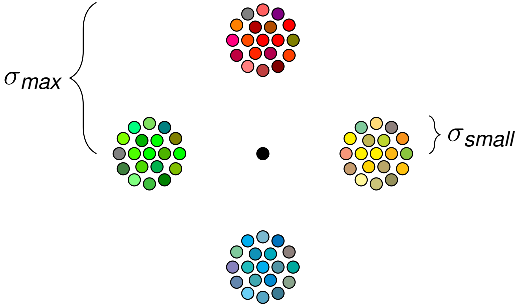
Gaussian Processes

- 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: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.



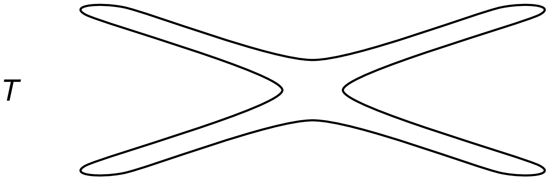
Gaussian Processes: chaining

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



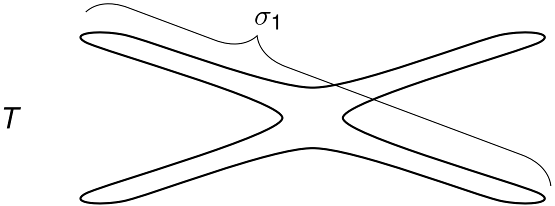
Gaussian Processes: chaining

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



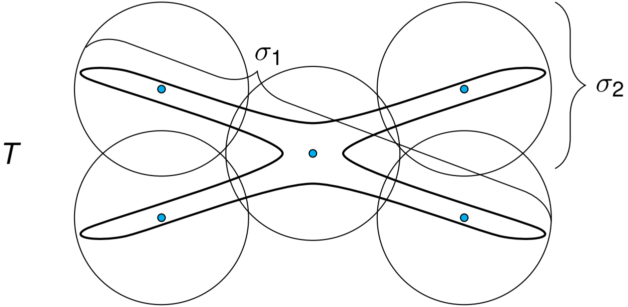
Gaussian Processes: chaining

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



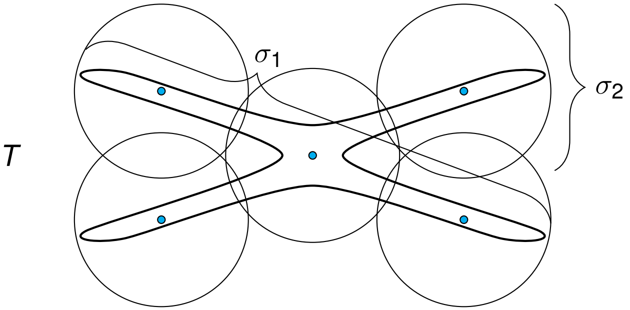
Gaussian Processes: chaining

- 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: $\sigma_{max}\sqrt{\log 4} + \sigma_{small}\sqrt{\log n}$.

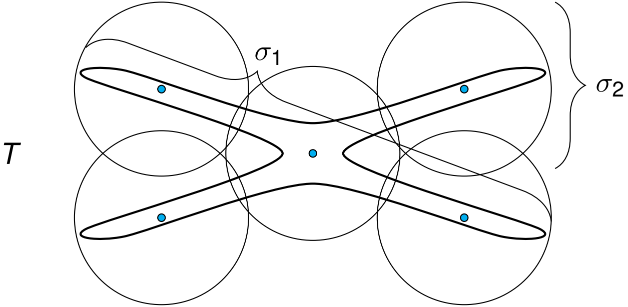


balls necessary: $N(\sigma_2)$

(*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: $\sigma_1 \sqrt{\log N(\sigma_2)} + \sigma_2 \sqrt{\log n}$.

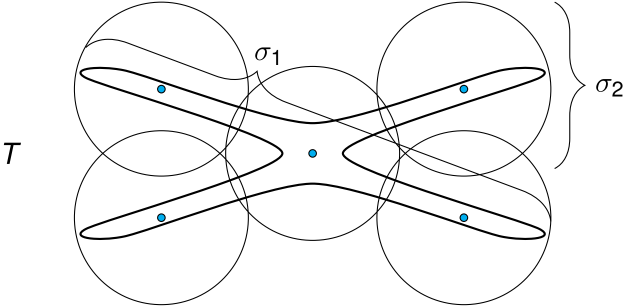


balls necessary: $N(\sigma_2)$

(*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: $N(\sigma_2)$

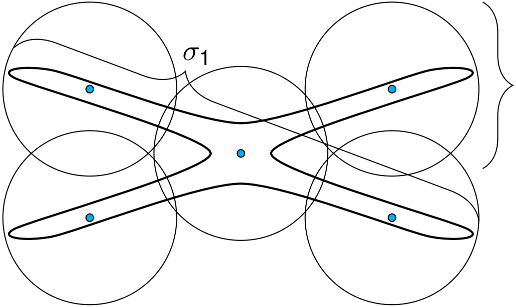
(*covering number* depends on $T, \|\cdot\|$)

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?

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

T



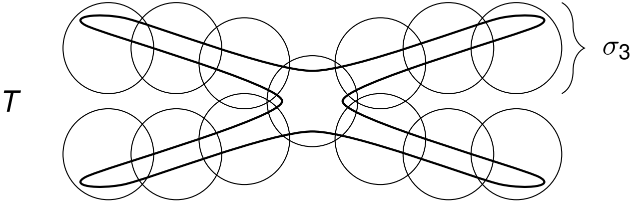
balls necessary: $N(\sigma_2)$

(*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?

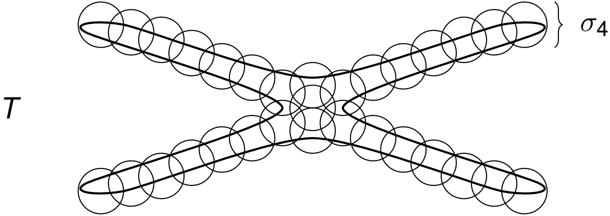
$$\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log N(\sigma_3)} +$$



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?

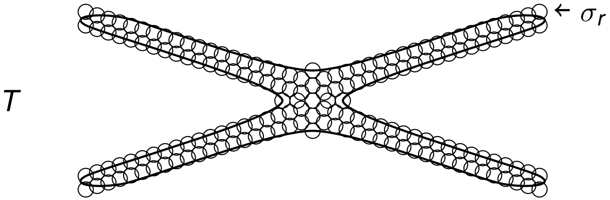
$$\mathbb{E}\sup_{x\in T} G_x \lesssim \sigma_1\sqrt{\log N(\sigma_2)} + \sigma_2\sqrt{\log N(\sigma_3)} + \sigma_3\sqrt{\log N(\sigma_4)} +$$



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?

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

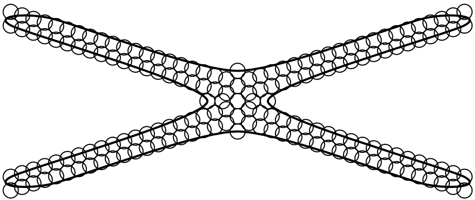


Gaussian Processes: chaining

- 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?

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

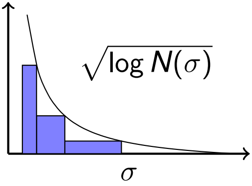
T

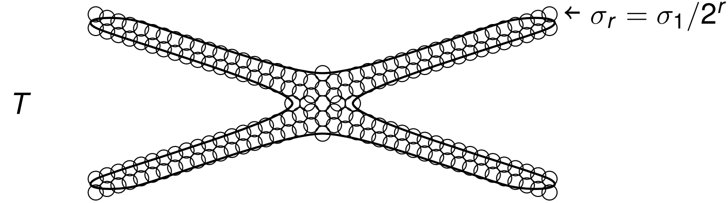


Gaussian Processes: chaining

- 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?

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

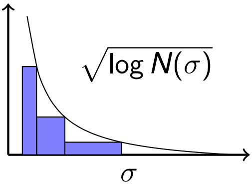




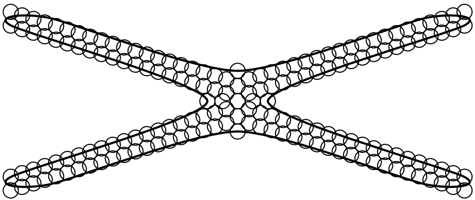
Gaussian Processes: chaining

- 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?

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



T



$\leftarrow \sigma_r = \sigma_1/2^r$

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

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

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

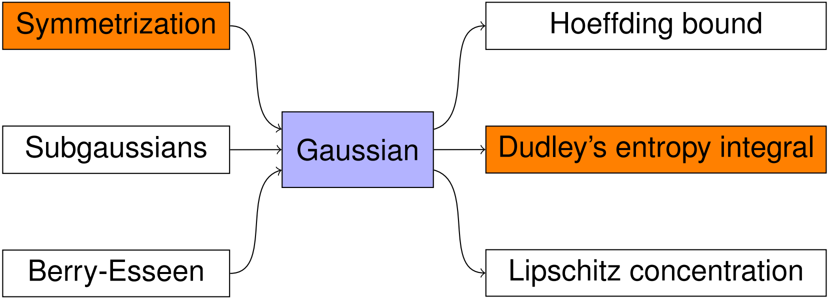
$$\gamma_2(T, \|\cdot\|) := \mathbb{E} \sup_{x \in T} G_x \lesssim \int_0^\infty \sqrt{\log N(T, \|\cdot\|, u)} du$$

- Bound a random variable using geometry.

A Probabilist's Toolbox (recap)

Convert to Gaussians

Gaussian concentration



Outline

Goal

Let $\Omega \subset [n]$ have each $i \in [n]$ independently with probability m/n. Let

$$A = \frac{1}{\sqrt{m}} F_\Omega$$

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

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

(Expectation of $\color{red}{*}$) = $\color{blue}{*}$

Proof outline: Rudelson-Vershynin

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

$$\mathbb{E} \sup_{x} x^T(A^T A - \mathrm{I})x$$

Expected
sup deviation

Proof outline: Rudelson-Vershynin

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

$\mathbb{E} \sup$
$x^T(A^T A - \mathrm{I})x$

Expected sup deviation

Symmetrization

Proof outline: Rudelson-Vershynin

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

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

↓

Symmetrization

↓

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

γ_2 : supremum of Gaussian process

Σ_k : k-sparse unit vectors

$\|\cdot\|$: a norm that depends on A

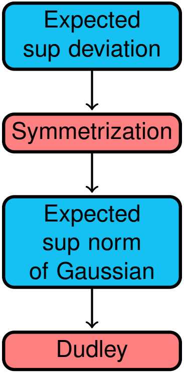
(specified in a few slides)

Proof outline: Rudelson-Vershynin

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

$\mathbb{E} \sup x^T(A^T A - \mathrm{I})x$

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



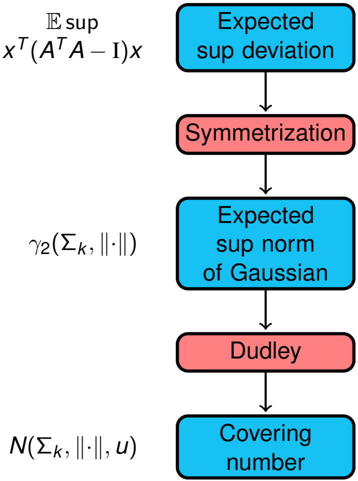
γ_2 : supremum of Gaussian process

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.



γ_2 : supremum of Gaussian process

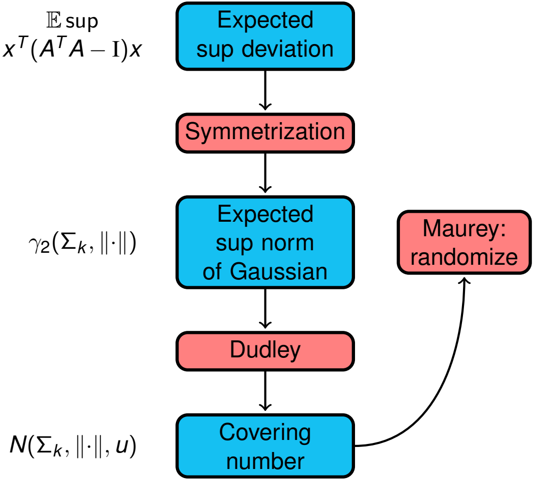
Σ_k : k-sparse unit vectors

$\|\cdot\|$: a norm that depends on A

(specified in a few slides)

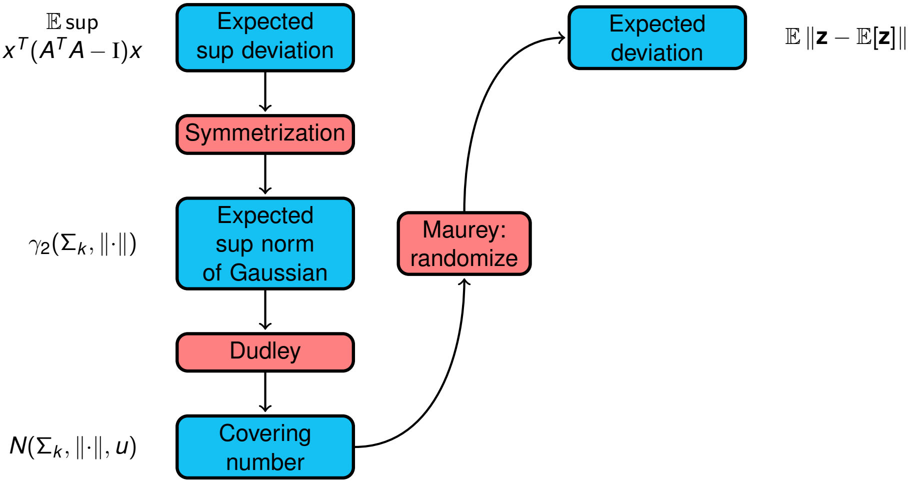
Proof outline: Rudelson-Vershynin

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



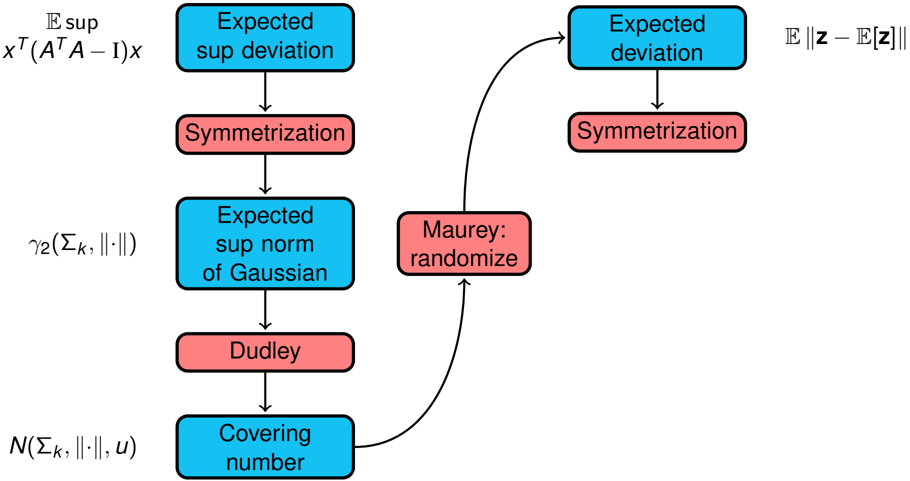
Proof outline: Rudelson-Vershynin

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



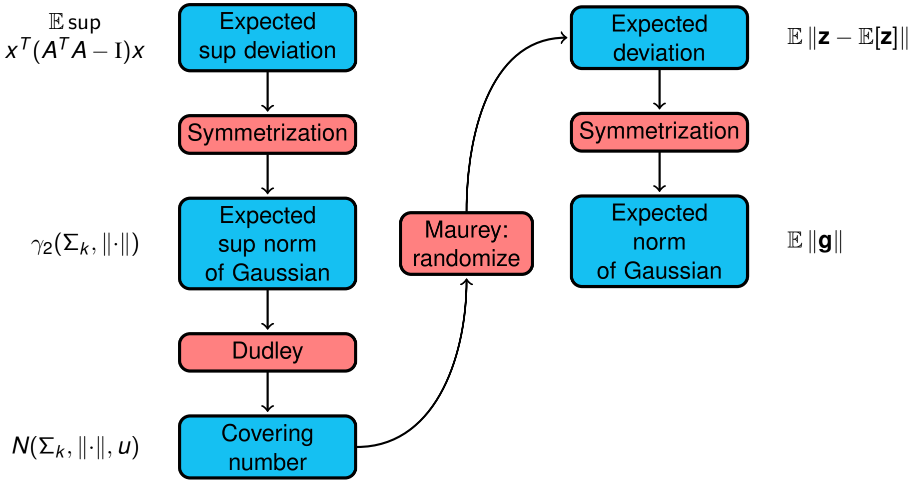
Proof outline: Rudelson-Vershynin

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



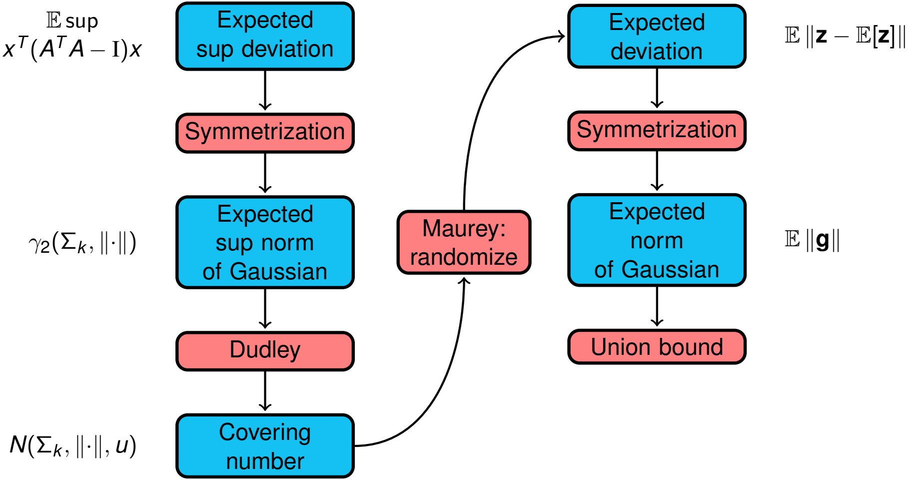
Proof outline: Rudelson-Vershynin

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



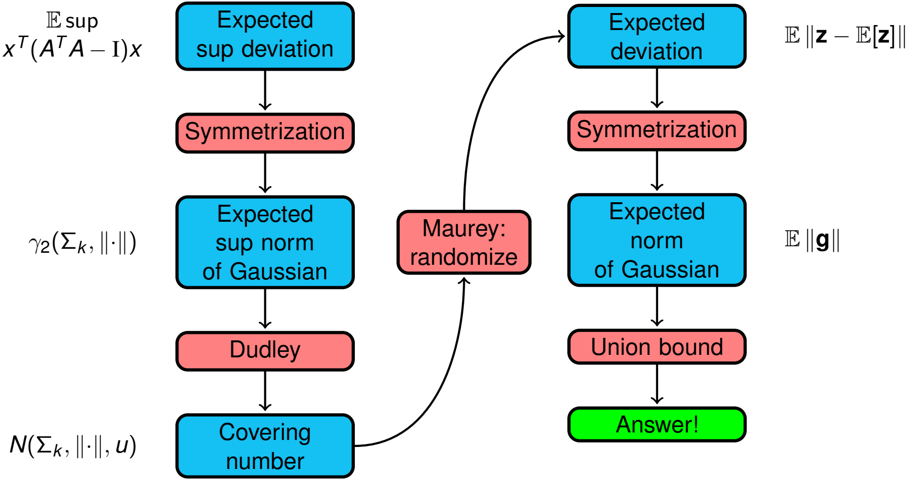
Proof outline: Rudelson-Vershynin

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



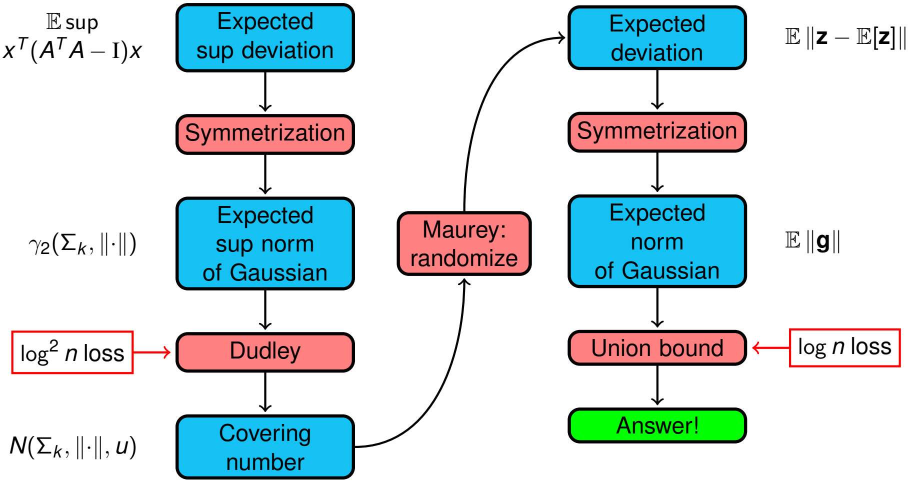
Proof outline: Rudelson-Vershynin

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



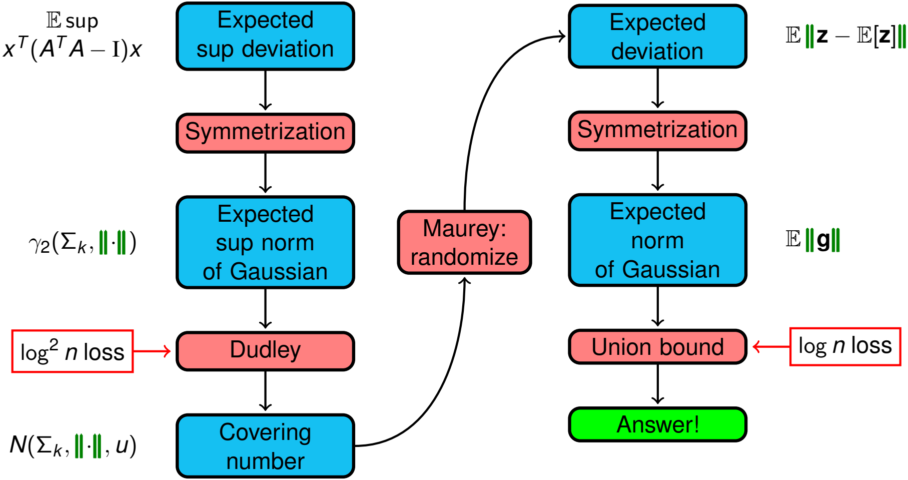
Proof outline: Rudelson-Vershynin

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



Proof outline: Rudelson-Vershynin

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



Setup

Let $\delta_i = 1_{i\in\Omega}$. Then $\Pr[\delta_i] = m/n$, independently for all i.

$$Ax = \frac{1}{\sqrt{m}} \sum_{i=1}^{n} \delta_i F_i x.$$

where $F_{ij} = e^{2\pi\sqrt{-1}ij/n}$. We would like to analyze the RIP constant

$$R_\Omega := \sup_{x\in\Sigma_k} |x^T A^T A x - 1|.$$

Now, for any fixed x,

$$\mathop{\mathbb{E}}_{\Omega}[x^T A^T A x] = \frac{1}{n} x^T F^T F x = \|x\|_2^2$$

and hence

$$\mathop{\mathbb{E}}_{\Omega}[R_\Omega] = \mathop{\mathbb{E}}_{\Omega} \sup_{x\in\Sigma_k} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| = \mathop{\mathbb{E}}_{\Omega} \sup_{x\in\Sigma_k} \left| \|Ax\|_2^2 - \mathop{\mathbb{E}}_{\Omega}[\|Ax\|_2^2] \right|$$

Proof part 1: symmetrization

$$\begin{aligned}
\mathbb{E}[R_\Omega] &= \mathop{\mathbb{E}}_{\Omega} \sup_{x \in \Sigma_k} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \\
&= \frac{1}{m} \mathop{\mathbb{E}}_{\delta} \sup_{x \in \Sigma_k} \left| \sum_{i=1}^n \delta_i \langle F_i, x\rangle^2 - \mathbb{E}[\sum_{i=1}^n \delta_i \langle F_i, x\rangle^2] \right| \\
&\leq \frac{1}{m} 3 \mathop{\mathbb{E}}_{\delta, g} \sup_{x \in \Sigma_k} \left| \sum_{i=1}^n g_i \delta_i \langle F_i, x\rangle^2 \right| \\
&\leq \frac{1}{m} 3 \mathop{\mathbb{E}}_{\Omega} \mathop{\mathbb{E}}_{g} \sup_{x \in \Sigma_k} \left| \sum_{i \in \Omega} g_i \langle F_i, x\rangle^2 \right| .
\end{aligned}$$

The Gaussian Process

So by symmetrization,

$$m\,\mathbb{E}[R_\Omega] \lesssim \mathbb{E}_{\Omega}\,\mathbb{E}_{g} \sup_{x\in\Sigma_k} \left|\sum_{i\in\Omega} g_i\langle F_i, x\rangle^2\right|.$$

Now fix Ω, and define the Gaussian process

$$G_x = \sum_{i\in\Omega} g_i\langle F_i, x\rangle^2.$$

which induces the norm

$$\|x-y\|_G^2 = \mathbb{E}[(G_x - G_y)^2] = \sum_{i\in\Omega}(\langle F_i, x\rangle^2 - \langle F_i, y\rangle^2)^2$$

so that

$$m\,\mathbb{E}[R] \lesssim \mathbb{E}_{\Omega}\,\mathbb{E} \sup_{x\in\Sigma_k} G_x =: \mathbb{E}_{\Omega}\,\gamma_2(\Sigma_k, \|\cdot\|_G)$$

$$\leq \mathbb{E}_{\Omega} \int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_G, u)du}$$

by Dudley's entropy integral.

Simplifying the norm

$$\begin{aligned}
\|x-y\|_G^2 &= \sum_{i\in\Omega}(\langle F_i, x\rangle^2 - \langle F_i, y\rangle^2)^2 \\
&= \sum_{i\in\Omega}(\langle F_i, x+y\rangle \cdot \langle F_i, x-y\rangle)^2 \\
&\leq (\sum_{i\in\Omega}\langle F_i, x+y\rangle^2)\max_{i\in\Omega}\langle F_i, x-y\rangle^2 \\
&\leq (4\sup_{x'\in\Sigma_k}\sum_{i\in\Omega}\langle F_i, x'\rangle^2)\max_{i\in[n]}\langle F_i, x-y\rangle^2 \\
&\leq 4m(1+R_\Omega)\|F(x-y)\|_\infty^2.
\end{aligned}$$

If we define $\|x\|_F := \|Fx\|_\infty$, this means

$$\begin{aligned}
m\,\mathbb{E}[R_\Omega] &\lesssim \mathop{\mathbb{E}}_{\Omega}\int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u/\sqrt{1+R_\Omega})du} \\
&\leq \mathop{\mathbb{E}}_{\Omega}\sqrt{m(1+R_\Omega)}\int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u)du}
\end{aligned}$$

New goal

Have:

$$m\,\mathbb{E}[R_\Omega] \lesssim \left(\mathop{\mathbb{E}}_{\Omega}\sqrt{m(1+R_\Omega)}\right)\int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u)du}$$

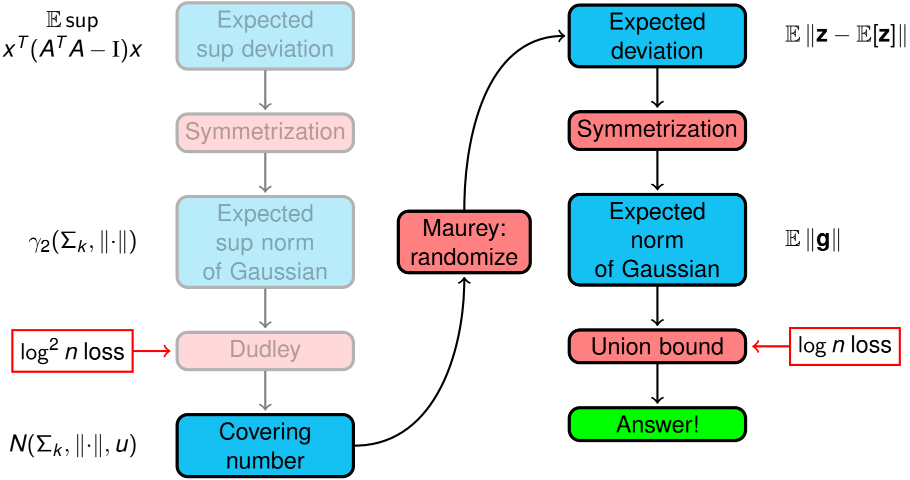
Will show:

$$\int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u)du} \lesssim \epsilon\sqrt{m}$$

for $\epsilon < 1$. This implies that $\mathbb{E}[R_\Omega] \lesssim \epsilon\,\mathbb{E}[\sqrt{1+R_\Omega}]$, and hence

$$\mathbb{E}[R_\Omega] \lesssim \epsilon.$$

Progress

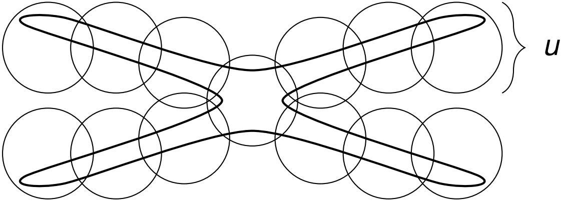


$\mathbb{E}\sup_x x^T(A^TA - I)x$
Expected sup deviation
Symmetrization
$\gamma_2(\Sigma_k, \|\cdot\|)$
Expected sup norm of Gaussian
$\log^2 n$ loss
Dudley
$N(\Sigma_k, \|\cdot\|, u)$
Covering number
Maurey: randomize
Expected deviation
$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$
Symmetrization
Expected norm of Gaussian
$\mathbb{E}\|\mathbf{g}\|$
Union bound
$\log n$ loss
Answer!

Outline

Covering Number Bound

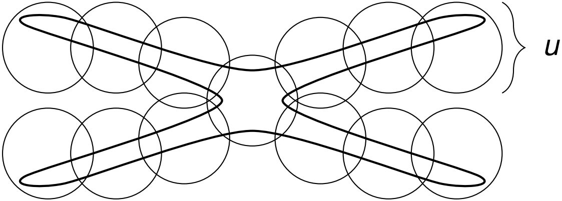
$$N(\Sigma_k, \|\cdot\|_F, u)$$



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

Covering Number Bound

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



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

Covering number bound

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

Covering number bound

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

- Simpler to imagine: what about ℓ_2?

Covering number bound

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

- 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

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

- 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

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

- 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

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

- 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

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

- 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

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

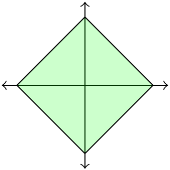
- Simpler to imagine: what about ℓ_2?
- How many ℓ_2 balls of radius u required to cover B_1?

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

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

Covering Number Bound

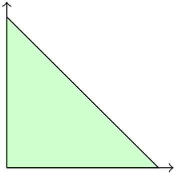
Maurey's empirical method



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

Covering Number Bound

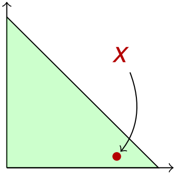
Maurey's empirical method



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

Covering Number Bound

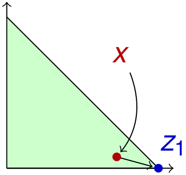
Maurey's empirical method



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

Covering Number Bound

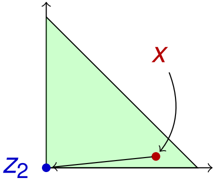
Maurey's empirical method



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

Covering Number Bound

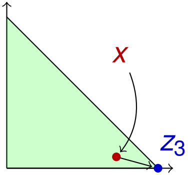
Maurey's empirical method



- 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.

Covering Number Bound

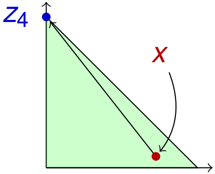
Maurey's empirical method



- 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

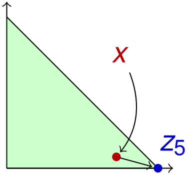
Maurey's empirical method



- 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.

Covering Number Bound

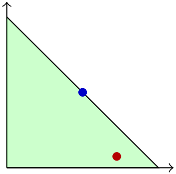
Maurey's empirical method



- 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

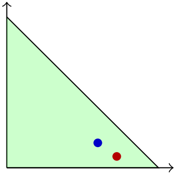
Maurey's empirical method



- 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

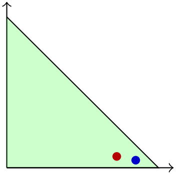
Maurey's empirical method



- 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

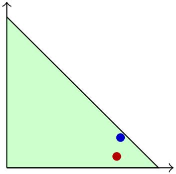
Maurey's empirical method



- 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

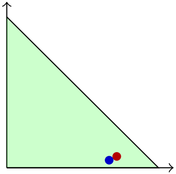
Maurey's empirical method



- 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

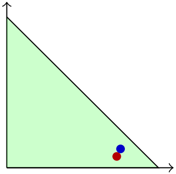
Maurey's empirical method



- 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

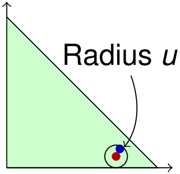
Maurey's empirical method



- 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

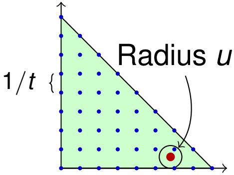


- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

Covering Number Bound

Maurey's empirical method



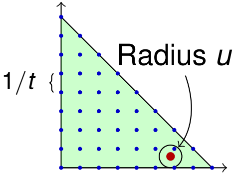
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

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

Covering Number Bound

Maurey's empirical method



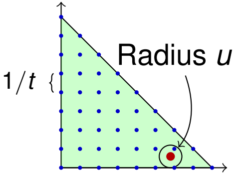
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t} \sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

- 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



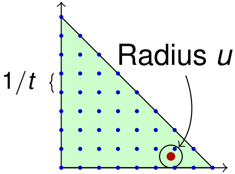
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

- 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



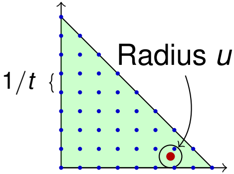
- How many balls of radius u required to cover B_1^+?
- Consider any $x \in B_1^+$.
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

- 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

Maurey's empirical method



Will show: $\mathbb{E}[\|\mathbf{z} - x\|_F] \leq \sqrt{1/t}$

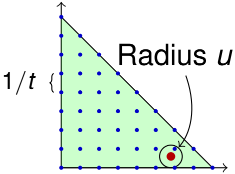
- Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of x to simplex.
- The sample mean $\mathbf{z} = \frac{1}{t}\sum z_i$ converges to x as $t \to \infty$.
- Let t be large enough that, regardless of x,

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

- 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



Will show: $\mathbb{E}[\|\mathbf{z} - x\|_F] \leq \sqrt{1/t} \implies N(T, \|\cdot\|_F, u) \leq n^{1/u^2}$

- 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,

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

- 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

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

Covering Number Bound

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

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

Covering Number Bound

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

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

Covering Number Bound

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

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

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: $\mathbb{E}[\|\mathbf{z} - x\|_F] \lesssim \sqrt{\log n/t}$.
- Symmetrize!

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

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.

- For each i, $(Fg)_i \sim N(0, 1)$.

Covering Number Bound

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

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

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.

- For each i, $(Fg)_i \sim N(0, 1)$.
- Hence $\|\mathbf{g}\|_F = \|F\mathbf{g}\|_\infty \lesssim \sqrt{\log n}$ with high probability (& in expectation).

Covering Number Bound

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

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

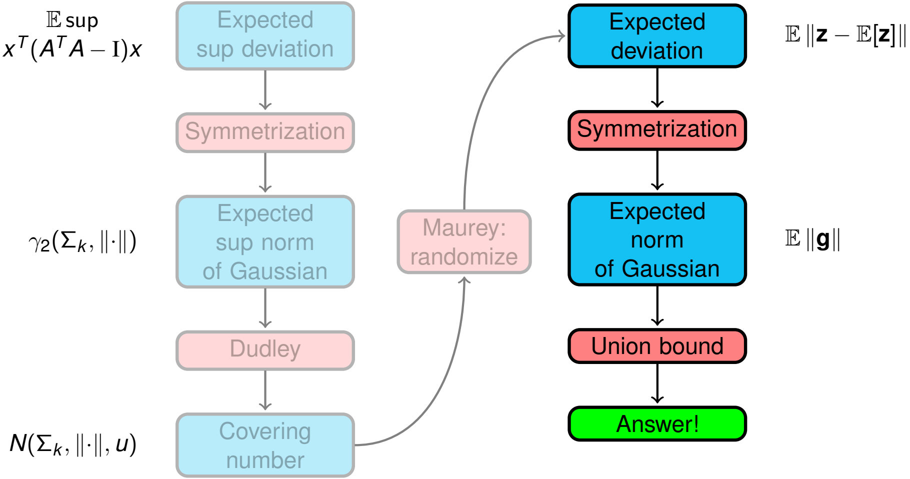
where $\mathbf{g} \in \mathbb{R}^n$ has

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

independently in each coordinate.

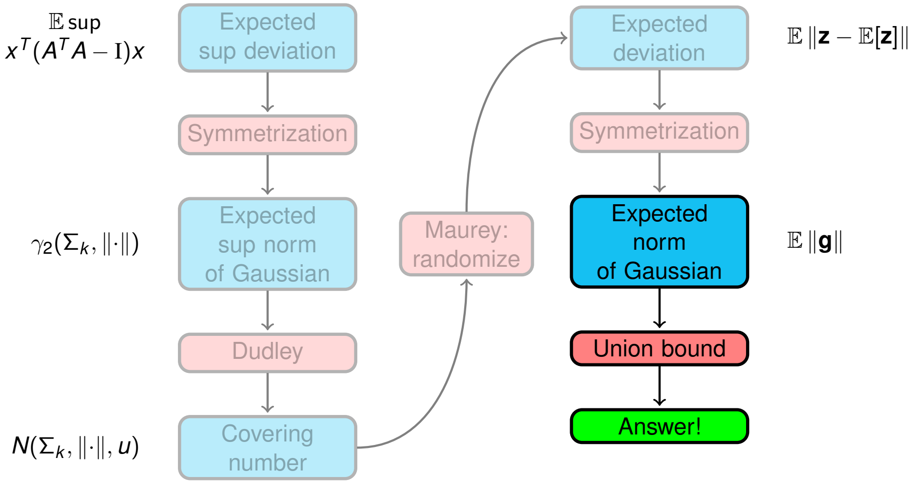
- For each i, $(Fg)_i \sim N(0, 1)$.
- Hence $\|\mathbf{g}\|_F = \|F\mathbf{g}\|_\infty \lesssim \sqrt{\log n}$ with high probability (& in expectation).
- Thus $t = \log n/u^2$ suffices, for $N(B_1, \|\cdot\|_F, u) \le n^{O(\log n)/u^2}$.

Progress



$\mathbb{E} \sup x^T(A^TA - \mathrm{I})x$
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
$\mathbb{E}\|\mathbf{g}\|$
Union bound
Answer!

Progress


$\mathbb{E}\sup x^T(A^TA - \mathrm{I})x$
Expected sup deviation
Symmetrization
$\gamma_2(\Sigma_k, \|\cdot\|)$
Expected sup norm of Gaussian
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
Answer!

$$\log N(B_1, \|\cdot\|_F, u) \lesssim \frac{1}{u^2} \log^2 n$$
$$\log N(\Sigma_k, \|\cdot\|_F, u) \lesssim \frac{k}{u^2} \log^2 n$$

And hence

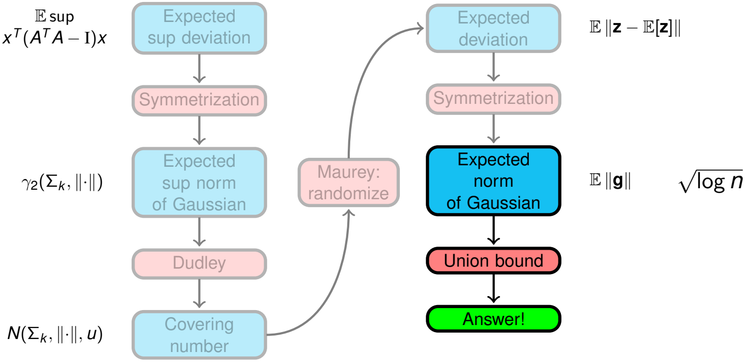
$$\int_{1/n^{100}}^{n} \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u)} du \lesssim \sqrt{k \log^4 n} \leq \epsilon \sqrt{m}$$

if $m \geq \frac{1}{\epsilon^2} k \log^4 n$, which is what we needed.
[Note: Small u are negligible by the volume argument:

$$\int_0^{1/n^{100}} \sqrt{\log N(\Sigma_k, \|\cdot\|_F, u)} du \lesssim \sqrt{k \log^2 n} \int_0^{1/n^{100}} n \log(1/u) du \ll 1/n^{97}$$

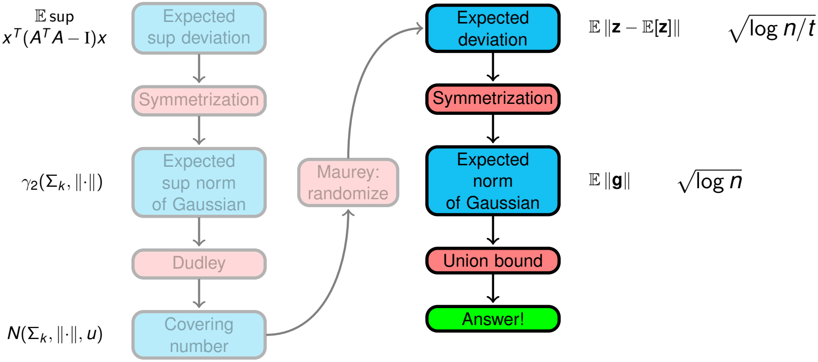
]

Unrolling everything



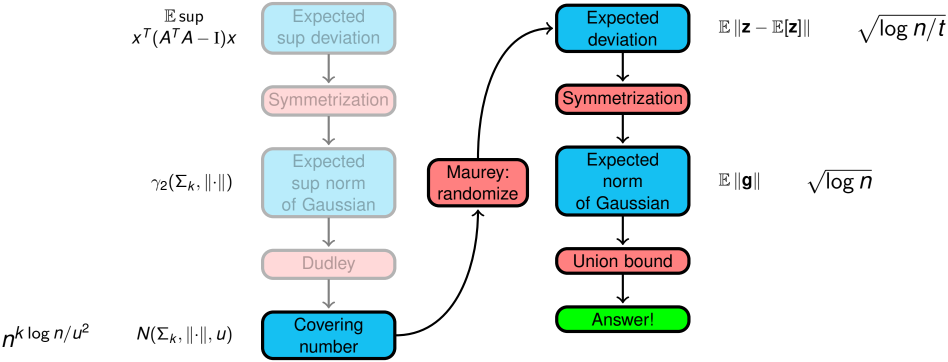
Union bound of $[n]$ uses $\log n$ factor

Unrolling everything



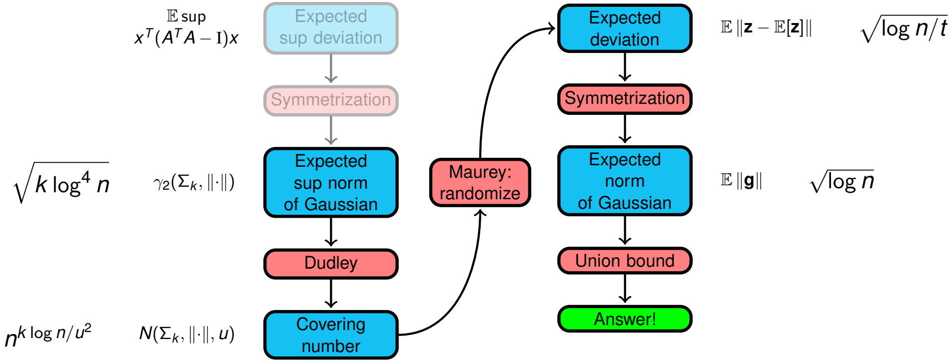
Sample mean $\mathbf{z}$ expects to lie within u of $\mathbf{x}$ for $t \geq \log n/u^2$

Unrolling everything



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

Unrolling everything



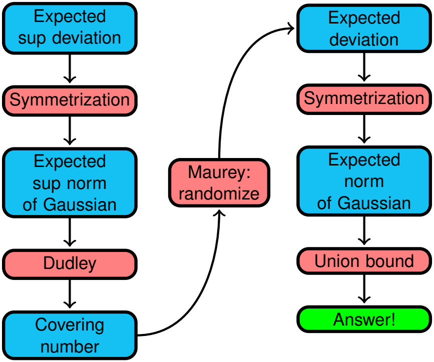
Entropy integral is $\sqrt{\frac{k\log^4 n}{m}}$

Unrolling everything

$\sqrt{\frac{k \log^4 n}{m}}$ $\quad \mathbb{E} \sup x^T(A^TA - \mathrm{I})x$

$\sqrt{k \log^4 n}$ $\quad \gamma_2(\Sigma_k, \|\cdot\|)$

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



$\mathbb{E}\|\mathbf{z} - \mathbb{E}[\mathbf{z}]\|$ $\quad \sqrt{\log n/t}$

$\mathbb{E}\|\mathbf{g}\|$ $\quad \sqrt{\log n}$

$$\text{RIP constant } \epsilon \lesssim \sqrt{\frac{k \log^4 n}{m}}$$

Summary

- Symmetrization and covering numbers are very general tools!

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

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

- So the entropy integral gives

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

- 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
$$N(B_1, \ell_2, u) \le n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \le \log^{3/2} n.$$
- 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
$$N(B_1, \ell_2, u) \le n^{1/u^2}$$
- So the entropy integral gives
$$\gamma_2(B_1, \ell_2) \le \log^{3/2} n.$$
- 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 $\sup d(x, A_i) \le \sigma_1 / 2^i$.

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

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

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

Covering Number Bound

Maurey's empirical method

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

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

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

- 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

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

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

- 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

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

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

- 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,

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

  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.

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

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.

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

- 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 $z_1, \ldots, z_t$ symmetrized to g.

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

- 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.

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

- 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.

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

- 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.

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

- 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 $(\log n)/k$

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.

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

- 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.

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.

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

- 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 $z_1, \ldots, z_t$ 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 $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2$

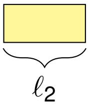
Bounding the norm in our case (part 2)

- 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

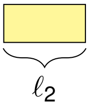
$$C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty$$



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

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



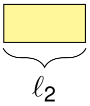
- Naive bound:

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

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

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



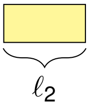
- Naive bound:

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

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:

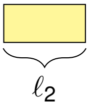
$$C \lesssim \|A_i\|_F \cdot \sqrt{\|x\|_\infty / t} \leq \sqrt{Bk} \cdot \sqrt{\log n/(kt)} = \sqrt{B \log n / t}$$

- "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:

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

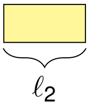
- "Very weak" RIP bound: for some $B = \log^c n$,

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

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 $z_1, \ldots, z_t$ 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

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

$$\underbrace{\boxed{\qquad}}_{\ell_2}$$

- 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

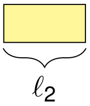
$$C \lesssim \sqrt{B/(t \log n)}$$

- 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:

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

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

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

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