Fast RIP matrices with fewer rows

Jelani Nelson  Eric Price  Mary Wootters

Princeton  MIT  Michigan

2013-04-05

\( \sigma_{\text{large}} \)  \( \sigma_{\text{small}} \)
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

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
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

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
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

\[ \Phi \]

\[ \begin{align*} m & \quad \Phi \quad n \\ x & \quad = \quad y \end{align*} \]
Compressive Sensing Algorithms: Two Classes

Structure-aware

Recovery algorithm tied to matrix structure (e.g. Count-Sketch)
Compressive Sensing Algorithms: Two Classes

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

Recovery algorithms just multiply by $\Phi, \Phi^T$
(e.g. L1 minimization)
Compressive Sensing Algorithms: Two Classes

Structure-aware
Recovery algorithm tied to matrix structure (e.g. Count-Sketch)
Faster
Often: Sparse matrices
Less robust

Structure-oblivious
Recovery algorithms just multiply by $\Phi, \Phi^T$
(e.g. L1 minimization)
Compressive Sensing Algorithms: Two Classes

\[ x = \Phi \Phi^T \]

Structure-aware
- Recovery algorithm tied to matrix structure (e.g. Count-Sketch)
- Faster
- Often: Sparse matrices
- Less robust

Structure-oblivious
- Recovery algorithms just multiply by $\Phi$, $\Phi^T$
- Slower
- Dense matrices
- More robust

Eric Price (MIT)

Fast RIP matrices with fewer rows
Compressive Sensing Algorithms: Two Classes

\[ x = \Phi m = y \]

Structure-aware

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

Faster

Often: Sparse matrices
Less robust

Structure-oblivious

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

Slower

Dense matrices
More robust

Yesterday:
Fourier \( \rightarrow \) sparse
Compressive Sensing Algorithms: Two Classes

\[ x = \Phi m = y \]

Structure-aware

- Recovery algorithm tied to matrix structure (e.g., Count-Sketch)
- Faster
- Often: Sparse matrices
- Less robust

Yesterday:
Fourier $\rightarrow$ sparse

Structure-oblivious

- Recovery algorithms just multiply by $\Phi, \Phi^T$
- Slower
- Dense matrices
- More robust

Today:

Fast RIP matrices with fewer rows
Algorithms for compressive sensing

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

- For all of these:
  - the time it takes to multiply by $\Phi$ or $\Phi^T$ is the bottleneck.
  - the Restricted Isometry Property is a sufficient condition.
Goal: recover approximately $k$-sparse $x$ from $y = \Phi x$.

A lot of people use convex optimization:

\[
\min \quad \|x\|_1 \\
\text{s.t.} \quad \Phi x = y
\]
Goal: recover approximately $k$-sparse $x$ from $y = \Phi x$.

A lot of people use convex optimization:

\[
\begin{align*}
\min & \quad \|x\|_1 \\
\text{s.t.} & \quad \Phi x = y
\end{align*}
\]

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:

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

- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....
- For all of these:
  - the time it takes to multiply by $\Phi$ or $\Phi^T$ is the bottleneck.
Algorithms for compressive sensing

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

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

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

All of these submatrices are well conditioned.

\[ \Phi \]

for all \( k \)-sparse \( x \in \mathbb{R}^n \).
Restricted Isometry Property (RIP)

\[ (1 - \epsilon) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon) \|x\|_2^2 \]

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.

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

Eric Price (MIT)  Fast RIP matrices with fewer rows  2013-04-05  8 / 52
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 $\log k \sim \log n \sim \log(n/k)$. 
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 \( \Phi, \Phi^T \).

* Talk will assume \( n^{0.1} < k < n^{0.9} \), so \( \log k \sim \log n \sim \log(n/k) \).
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 \( \Phi, \Phi^T \).
  - Random Gaussian matrix: \( \Theta(nk \log n) \) time.

* Talk will assume \( n^{0.1} < k < n^{0.9} \), so \( \log k \sim \log n \sim \log(n/k) \).
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 \( \Phi, \Phi^T \).
  - 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 \( \log k \sim \log n \sim \log(n/k) \).
An open question

Let \( A \) contain random rows from a Fourier matrix.

\[
m \begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} F \end{bmatrix}
\]

How many rows do you need to ensure that \( A \) has the RIP?

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

Ideal:

\[
m = O(k \log n) \quad [RRT12, KMR12].
\]

(Related: how about partial circulant matrices?)
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?

\[ m = O(k \log 4n) \quad \text{[CT06,RV08,CGV13]} \]

Ideal: \[ m = O(k \log n) \]

(Related: how about partial circulant matrices?)

\[ m = O(k \log 4n) \quad \text{[RRT12,KMR12]} \]
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?

Ideal: $m = O(k \log n)$. 
(Related: how about partial circulant matrices?) $m = O(k \log 4n)$ [RRT12,KMR12].
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:
- $m = O(k \log n)$. 

(Related: how about partial circulant matrices?)
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

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
Another motivation:
Johnson Lindenstrauss (JL) Transforms

High dimensional data
\[ S \subset \mathbb{R}^n \]
Another motivation: Johnson Lindenstrauss (JL) Transforms

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

Linear map $\Phi$

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 $\Phi$

$\Phi$ 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 \( \Phi \)

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

\( \Phi \) 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 \( \Phi \)

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

\( \Phi \) preserves the geometry of \( S \)

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

\[
\langle \Phi x, \Phi y \rangle = \langle x, y \rangle \pm \epsilon \|x\|_2 \|y\|_2
\]
Theorem (variant of Johnson-Lindenstrauss ’84)

Let \( x \in \mathbb{R}^n \). A random Gaussian matrix \( \Phi \) 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 $\Phi$ 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?

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

Gaussians

\[ O\left(\frac{\epsilon^2 k}{n}\right) \text{ multiplication time.} \]

Best way known for fast JL: by [Krahmer-Ward '11], RIP \( \Rightarrow \) JL.

Existing results: dimension \( O\left(\frac{\epsilon^2 k \log 4n}{n}\right) \).

\[ n \log n \text{ multiplication time.} \]

And by [BDDW '08], JL \( \Rightarrow \) RIP; so equivalent.

Round trip loses \( \log n \) factor in dimension.
How do we get a JL matrix?

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

Best way known for fast JL: by [Krahmer-Ward '11], RIP $\Rightarrow$ JL.

Existing results: dimension $O(\frac{1}{\epsilon^2} k \log^4 n)$.

And by [BDDW '08], JL $\Rightarrow$ RIP; so equivalent.

Round trip loses $\log n$ factor in dimension.
How do we get a JL matrix?

**Gaussians**
- Dimension $O\left(\frac{1}{\epsilon^2} k\right)$.
- $\frac{1}{\epsilon^2} nk$ 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.
  - Existing results: dimension $O(\frac{1}{\epsilon^2} k \log^4 n)$. 
How do we get a JL matrix?

- Gaussians
  - Dimension $O\left(\frac{1}{\epsilon^2} k\right)$.
  - $\frac{1}{\epsilon^2} nk$ multiplication time.
- Best way known for *fast JL*: by [Krahmer-Ward ’11], RIP $\Rightarrow$ JL.
  - Existing results: dimension $O\left(\frac{1}{\epsilon^2} k \log^4 n\right)$.
  - $n \log n$ multiplication time.
How do we get a JL matrix?

- Gaussians
  - Dimension $O\left(\frac{1}{\epsilon^2} k\right)$.
  - $\frac{1}{\epsilon^2} nk$ multiplication time.
- Best way known for fast JL: by [Krahmer-Ward ’11], RIP $\Rightarrow$ JL.
  - Existing results: dimension $O\left(\frac{1}{\epsilon^2} k \log^4 n\right)$.
  - $n \log n$ multiplication time.
- And by [BDDW ’08], JL $\Rightarrow$ 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 $m \sim k \log^3 n$, $B \sim \log^c n$, and $A$ is a random partial Fourier matrix, then $\Phi$ has the RIP with probability at least $2/3$.***
Generalization

Our approach is actually works for more general $A$:

\[
\text{Random sign flips} \quad \xrightarrow{\text{Subsampled Fourier}} \quad \text{Subsampled Fourier}
\]

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:
  \[
  \begin{align*}
  &\text{For all } A_i \\
  &\text{the RIP constant, although } \gg 1, \text{ is still controlled.}
  \end{align*}
  \]

Then $\Phi$ is a good RIP matrix:

- $\Phi$ has the RIP (whp) with $m = O(k \log n)$ rows.
- Time to multiply by $\Phi =$ time to multiply by $A + mB$.
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 $\Phi$ is a good RIP matrix:

- $\Phi$ has the RIP (whp) with $m = O(k \log^3 n)$ rows.
- Time to multiply by $\Phi = \text{time to multiply by } A + mB.$
### Results in the area

<table>
<thead>
<tr>
<th>Construction</th>
<th>Measurements $m$</th>
<th>Multiplication Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse JL matrices [KN12]</td>
<td>$\frac{1}{\epsilon^2} k \log n$</td>
<td>$\epsilon mn$</td>
</tr>
<tr>
<td>Partial Fourier [RV08,CGV13]</td>
<td>$\frac{1}{\epsilon^2} k \log^4 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Partial Circulant [KMR12]</td>
<td>$\frac{1}{\epsilon^2} k \log^4 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Our result: Hash of partial Fourier</td>
<td>$\frac{1}{\epsilon^2} k \log^3 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Our result: Hash of partial circulant</td>
<td>$\frac{1}{\epsilon^2} k \log^3 n$</td>
<td>$n \log n$</td>
</tr>
</tbody>
</table>

Iterated Fourier [AC06,AL09,AR13] $\frac{1}{\epsilon^2} k \log n \, ^\dagger$ $n \log n \, ^\dagger$

Requires $k \leq \frac{n}{2} - \delta$. This is the "easy" case:

Dimension: $k \log n$

Time: $k^2 \log^5 n$
### Results in the area

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

$\dagger$ Requires $k \leq n^{1/2-\delta}$. This is the “easy” case:
## Results in the area

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

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

**Dimension:**

\[
\begin{array}{c}
\text{n} \\
\downarrow \\
\text{k log}^4 n
\end{array}
\]

**Time:**

\[
\begin{array}{c}
n \\
\downarrow \\
\log n
\end{array}
\]

[RV08]
Results in the area

<table>
<thead>
<tr>
<th>Construction</th>
<th>Measurements $m$</th>
<th>Multiplication Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse JL matrices [KN12]</td>
<td>$\frac{1}{\epsilon^2} k \log n$</td>
<td>$\epsilon mn$</td>
</tr>
<tr>
<td>Partial Fourier [RV08,CGV13]</td>
<td>$\frac{1}{\epsilon^2} k \log^4 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Partial Circulant [KMR12]</td>
<td>$\frac{1}{\epsilon^2} k \log^4 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Our result: Hash of partial Fourier</td>
<td>$\frac{1}{\epsilon^2} k \log^3 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Our result: Hash of partial circulant</td>
<td>$\frac{1}{\epsilon^2} k \log^3 n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>Iterated Fourier [AC06,AL09,AR13]</td>
<td>$\frac{1}{\epsilon^2} k \log n$</td>
<td>$n \log n$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dimension:</th>
<th>$n$</th>
<th>$k \log^4 n$</th>
<th>$k \log n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>$n \log n$</td>
<td>$k^2 \log^5 n$</td>
<td></td>
</tr>
</tbody>
</table>

[RV08] Gaussian
Concentration of Measure

Let $\Sigma_k$ is unit-norm $k$-sparse vectors. We want to show for our distribution $\Phi$ on matrices that

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

Let $\Sigma_k$ is unit-norm $k$-sparse vectors. We want to show for our distribution $\Phi$ on matrices that

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

(Expectation of $\ast$) = $\ast$
Concentration of Measure

Let $\Sigma_k$ is unit-norm $k$-sparse vectors.
We want to show for our distribution $\Phi$ on matrices that

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

(Expectation of $\ast$) $= \ast$

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

Let $\Sigma_k$ is unit-norm $k$-sparse vectors.
We want to show for our distribution $\Phi$ on matrices that

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

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

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
- Drill
- Bit sets
Tools

Screwdriver

Bit sets

Drill

Bit
Tools

- Screwdriver
- Drill
- Bit sets
- 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

Symmetrization

Subgaussians

Berry-Esseen

Gaussian

Hoeffding bound

Dudley’s entropy integral

Lipschitz concentration

Will prove: symmetrization and Dudley’s entropy integral.
A Probabilist’s Toolbox

Convert to Gaussians

- Symmetrization
- Subgaussians
- Berry-Esseen

Gaussian concentration

- 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
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.
Lemma (Symmetrization)

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

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

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

$$\| A^T A - I \| = \| \sum A_i^T A_i - I \|.$$
**Lemma (Symmetrization)**

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.**
**Lemma (Symmetrization)**

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.

$$
\mathbb{E} \left[ \left\| \frac{1}{t} \sum_{i} X_i - \mathbb{E} \left[ \frac{1}{t} \sum_{i} X'_i \right] \right\| \right]
$$
**Symmetrization**

**Lemma (Symmetrization)**

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mu$. 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.

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

$$
= \mathbb{E}[\left\| \frac{1}{t} \sum s_i (X_i - X'_i) \right\|]
$$
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.

$$
\mathbb{E} \left[ \left\| \frac{1}{t} \sum_{i} X_i - \mathbb{E} \left[ \frac{1}{t} \sum_{i} X'_i \right] \right\| \right] \leq \mathbb{E} \left[ \left\| \frac{1}{t} \sum_{i} (X_i - X'_i) \right\| \right] = \mathbb{E} \left[ \left\| \frac{1}{t} \sum_{i} s_i (X_i - X'_i) \right\| \right]
$$

and apply the triangle inequality. \qed
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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.
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mu$. 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 $\mathbb{E}[|g_i|] \approx 0.8 > 2/3$. 
Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mathbb{E}[|g_i|] \approx .8 > 2/3$. 

\[
2 \mathbb{E}[\| \sum_i s_i X_i \|]
\]
**Lemma (Symmetrization)**

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mathbb{E}[|g_i|] \approx 0.8 > 2/3$.

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

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

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

Lemma (Symmetrization)

Suppose $X_1, \ldots, X_t$ are i.i.d. with mean $\mu$. 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 $\mathbb{E}[|g_i|] \approx .8 > 2/3$.

$$
2 \mathbb{E}[\| \sum s_i X_i \|] \leq 3 \mathbb{E}[\| \sum s_i \mathbb{E}[|g_i|] X_i \|] \\
\leq 3 \mathbb{E}[\| \sum s_i |g_i| X_i \|] \\
= 3 \mathbb{E}[\| \sum g_i X_i \|].
$$
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 $x \in T$. 

- 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 $G_{u,v} \sim \mathcal{N}(0, \|uv^T\|_2^2 F)$. 

- $E \|A\|_2^2 = E \sup_{u,v \in S_{m-1} \times S_{n-1}} u^T A v = E \sup_{u,v \in S_{m-1} \times S_{n-1}} G_{u,v}$ depends on the geometry of $T$. Distance: $\|x - y\|$ is standard deviation of $G_x - G_y$. In example: $\|((u,v) - (u',v'))\| = \|uv^T - u'v'^T\|_F$. 

Eric Price (MIT) 

Fast RIP matrices with fewer rows
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

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

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

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

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

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

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

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

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

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

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

- Depends on the geometry of $T$. 

$$\mathbb{E}\|A\|_2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T Av = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$
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 $G_{u,v} \sim N(0, \|uv^T\|_F^2)$.

$$\mathbb{E} \|A\|_2^2 = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} u^T A v = \mathbb{E} \sup_{u,v \in S^{m-1} \times S^{n-1}} G_{u,v}$$

- 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

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

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

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

Eric Price (MIT)

Fast RIP matrices with fewer rows

2013-04-05 27 / 52
Gaussian Processes

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

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

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

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

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

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

\[
\sigma_{\text{max}} \begin{cases} \\
\text{Position} = x \\
\text{Color} = G_x \\
G_0 = 0
\end{cases}
\]
Gaussian Processes

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

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

- **Ignoring geometry:**
  - \( \Pr[G_x > \sigma_{\text{max}} t] \leq e^{-t^2/2} \)
  - Union bound: with high probability, \( G_x \lesssim \sigma_{\text{max}} \sqrt{\log n} \).
  - \( \mathbb{E} \sup_{x \in T} G_x \lesssim \sigma_{\text{max}} \sqrt{\log n} \)

\[
\begin{align*}
\sigma_{\text{max}} & \quad \{ \text{Position} = x \} \\
\{ \sigma_{\text{small}} \} & \quad \{ \text{Color} = G_x \}
\end{align*}
\]

\( G_0 = 0 \)
Gaussian Processes

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

- \( \sigma_{\text{max}} \)
- \( \sigma_{\text{small}} \)
- Position = \( x \)
- Color = \( G_x \)
- \( G_0 = 0 \)
Gaussian Processes: chaining

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

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

$T$
Gaussian Processes: chaining

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

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

Why stop at two?

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

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

$\sigma_1$

$\sigma_2$

# 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?

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

# 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?

\[ \# \text{ balls necessary: } N(\sigma_2) \]

(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?

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

$\sigma_1$ $\sigma_2$

$T$

$\# \text{ balls necessary: } N(\sigma_2)$

(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?

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

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

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

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

$T$

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

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

$\sigma_r = \sigma_1 / 2^r$

$T$
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?

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

$\sigma_r = \sigma_1 / 2^r$
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

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

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*

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

*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 \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) 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.

---

Eric Price (MIT)
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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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.

$\Rightarrow$ the Johnson-Lindenstrauss lemma.
A Probabilist’s Toolbox (recap)

Convert to Gaussians

Symmetrization

Subgaussians

Berry-Esseen

Gaussian concentration

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
For $\Sigma_k$ 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.

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

Expected sup deviation
Proof outline: Rudelson-Vershynin

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

\[ \mathbb{E} \sup_{\|A^T A - I\|} \]

Expected sup deviation

\[ \text{Symmetrization} \]
Proof outline: Rudelson-Vershynin

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

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

Symmetrization

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

$\gamma_2: \sup$ supremum of Gaussian process

$\Sigma_k: k$-sparse unit vectors

$\|\cdot\|: \text{a norm that depends on } A$  
(specified in a few slides)
Proof outline: Rudelson-Vershynin
Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

\[ \mathbb{E} \sup_{||A^T A - I||} \]

Expected sup deviation

Symmetrization

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

Expected sup norm of Gaussian

Dudley

$\gamma_2$ : supremum of Gaussian process

$\Sigma_k$ : $k$-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.

- Expected sup deviation
- Symmetrization
- $\gamma_2$ : supremum of Gaussian process
- $\Sigma_k$ : $k$-sparse unit vectors
- $\| \cdot \|$ : a norm that depends on $A$
  (specified in a few slides)

Expected sup deviation

Symmetrization

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

Expected sup norm of Gaussian

Dudley

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

Covering number
Proof outline: Rudelson-Vershynin

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

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

Expected sup deviation

Symmetrization

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

Expected sup norm of Gaussian

Dudley

Maurey: randomize

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

Covering number

Eric Price (MIT) Fast RIP matrices with fewer rows 2013-04-05 36 / 52
Proof outline: Rudelson-Vershynin
Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

\[ \mathbb{E} \sup_{\|A^T A - I\|} \]

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

\[ \mathbb{E} \| z - \mathbb{E}[z] \| \]
Proof outline: Rudelson-Vershynin

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

\begin{align*}
\mathbb{E} \sup_{\|A^T A - I\|} \\
\gamma_2(\Sigma_k, \|\cdot\|) \\
N(\Sigma_k, \|\cdot\|, u)
\end{align*}

Expected sup deviation

Symmetrization

Expected sup norm of Gaussian

Dudley

Covering number

Expected deviation

Symmetrization

Maurey: randomize

$\mathbb{E} \|z - \mathbb{E}[z]\|$
Proof outline: Rudelson-Vershynin

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

- $\mathbb{E} \sup_{\|A^T A - I\|}$
  - Expected sup deviation
  - Symmetrization
  - $\gamma_2(\Sigma_k, \|\cdot\|)$
    - Expected sup norm of Gaussian
    - Dudley
    - $N(\Sigma_k, \|\cdot\|, u)$
  - Maurey: randomize

- $\mathbb{E} \|z - \mathbb{E}[z]\|$
Proof outline: Rudelson-Vershynin

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

\[ \mathbb{E} \sup_{\|A^T A - I\|} \]

Expected sup deviation

Symmetrization

Expected sup norm of Gaussian

Dudley

\[ \gamma_2(\Sigma_k, \| \cdot \|) \]

Maurey: randomize

Expected norm of Gaussian

Union bound

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

\[ \mathbb{E} \| g \| \]

\[ N(\Sigma_k, \| \cdot \|, u) \]

Covering number

\[ \log \frac{2}{\log n} \text{ loss} \]
Proof outline: Rudelson-Vershynin
Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.

Expected sup deviation
\[ \mathbb{E} \sup \| A^T A - I \| \]

Symmetrization

Expected sup norm of Gaussian
\[ \gamma_2(\Sigma_k, \| \cdot \|) \]

Dudley

Covering number
\[ N(\Sigma_k, \| \cdot \|, u) \]

Expected deviation
\[ \mathbb{E} \| z - \mathbb{E}[z] \| \]

Symmetrization

Expected norm of Gaussian
\[ \mathbb{E} \| g \| \]

Union bound

Answer!
Proof outline: Rudelson-Vershynin
Rudelson-Vershynin: subsampled Fourier, \(O(k \log^4 n)\) rows.

\[ \mathbb{E} \sup_{\|A^T A - I\|} \]

Symmetrization

Expected sup norm of Gaussian

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

Dudley

\(\log^2 n\) loss

\[ N(\Sigma_k, \|\cdot\|, u) \]

Covering number

Maurey: randomize

Expected deviation

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

Expected deviation

Symmetrization

Expected norm of Gaussian

\[ \mathbb{E} \|g\| \]

Union bound

Answer!

\[ \log n\] loss
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
Dudley
$log^2 n$ loss
$N(\Sigma_k, \| \cdot \|, u)$
Covering number

$\mathbb{E} \sup \| \Phi^T \Phi - I \|$
Expected deviation
Symmetrization
Maurey: randomize
Expected norm of Gaussian
$\mathbb{E} \| g \|$
Union bound
Answer!
$log n$ loss

Eric Price (MIT)
Fast RIP matrices with fewer rows
2013-04-05 36 / 52
**Proof outline**

Rudelson-Vershynin: subsampled Fourier, $O(k \log^4 n)$ rows.
Nelson-P-Wootters: sparse times Fourier, $O(k \log^3 n)$ rows.

\[\mathbb{E} \sup_{\Phi^T \Phi - I} \| \Phi^T \Phi - I \|\]

E. Krahmer, R. Mendelson, H. Rauhut ‘13

\[\gamma_2(\Sigma_k, \| \cdot \|)\]

\[\log^2 n \text{ loss}\]

Expected sup norm of Gaussian

\[N(\Sigma_k,\| \cdot \|, u)\]

Dudley

Expected deviation

Expected sup deviation

Symmetrization

Maurey: randomize

Expected norm of Gaussian

Union bound

Answer!

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

\[\mathbb{E} \| g \|\]

\[\log n \text{ loss}\]
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

Krahmer
Mendelson
Rauhut ’13

Expected sup norm of Gaussian

\[\gamma_2(\Sigma_k, \|\cdot\|)\]

log loss

Expected deviation

Symmetrization

Expected norm of Gaussian

Maurey: randomize

Expected deviation

Union bound

Answer!

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

\[\mathbb{E} \|g\|\]

\[\log^2 n\]

Dudley

Covering number

\[N(\Sigma_k, \|\cdot\|, u)\]

\[\log n\]
Proof part I: triangle inequality

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

\[ \leq E \sup_{x \in \Sigma_k} \left| \| \Phi x \|_2^2 - \| Ax \|_2^2 \right| + E \sup_{x \in \Sigma_k} \left| \| Ax \|_2^2 - \| x \|_2^2 \right| \]
Proof part I: triangle inequality

\[
\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| \\
\leq \mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|Ax\|_2^2 \right| + \mathbb{E} \sup_{x \in \Sigma_k} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \\
= \mathbb{E} \sup_{x \in \Sigma_k} \left| \|X_A s\|_2^2 - \mathbb{E}_s \|X_A s\|_2^2 \right| + \text{(RIP constant of } A\text{)},
\]

where \(X_A\) is some matrix depending on \(x\) and \(A\), and \(s\) is the vector of random sign flips used in \(H\).
Proof part I: triangle inequality

\[ \mathbb{E} \sup_{x \in \Sigma_k} \| X_A s \|_2^2 - \mathbb{E}_s \| X_A s \|_2^2 \] + (RIP constant of A)
Proof part I: triangle inequality

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

By assumption, this is small.
(Recall A has extra rows)
Proof part I: triangle inequality

\[ E \sup_{x \in \Sigma_k} \| X_A s \|_2^2 - E_s \| X_A s \|_2^2 \] 

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 $\Sigma_k$ 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 \( \| x \|_A \):

\[ Ax = \text{subset of } \hat{x} \]

\[ \| x \|_A = \max_{i \in [m]} \| A_i x \|_2. \]
Definition of the Norm

\[ N(\Sigma_k, \| \cdot \|_A, u) \]

for the norm \( \| x \|_A \):

\[ Ax = \text{subset of } \hat{x} \]

\[ \ell_2 \]

Eric Price (MIT) Fast RIP matrices with fewer rows 2013-04-05 40 / 52
Definition of the Norm

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

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

$N(\Sigma_k, \|\cdot\|_A, u)$
Definition of the Norm

\[ N(\Sigma_k, \| \cdot \|_A, u) \]

for the norm \( \| x \|_A \):

\[ \| x \|_A = \max_{i \in [m]} \| A_i x \|_2. \]
Definition of the Norm

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

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

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

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

\[ \mathbb{E} \sup_{\|\Phi^T \Phi - I\|} \]

Expected sup deviation

Krahmer Mendelson Rauhut '13

\[ \gamma_2(\Sigma_k, \|\cdot\|) \]

Expected sup norm of Gaussian

Maurey: randomize

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

Expected deviation

Symmetrization

\[ \mathbb{E} \|g\| \]

Expected norm of Gaussian

Union bound

Answer!

\[ \log^2 n \text{ loss} \]

\[ N(\Sigma_k, \|\cdot\|, u) \]

Covering number

\[ \log n \text{ loss} \]

Dudley

Union bound

Answer!

Eric Price (MIT)

Fast RIP matrices with fewer rows

2013-04-05 41 / 52
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

\[ N(\Sigma_k, \| \cdot \|_A, u) \]

\[ \Sigma_k = \{ k\text{-sparse } x \mid \| x \|_2 \leq 1 \} \]
Covering Number Bound

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

\[ \Sigma_k = \{ k\text{-sparse } x \mid \| x \|_2 \leq 1 \} \]
\[ \subset \sqrt{k}B_1 = \{ x \mid \| x \|_1 \leq \sqrt{k} \} \]
Covering number bound

\[ N(B_1, \| \cdot \|_A, u) \]
Covering number bound

\[ N(B_1, \| \cdot \|_A, u) \]

- Simpler to imagine: what about \( \ell_2 \)?
Covering number bound

\[ N(B_1, \| \cdot \|_A, u) \]

- Simpler to imagine: what about \( \ell_2 \)?
- How many \( \ell_2 \) 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 \( \ell_2 \)?
- How many \( \ell_2 \) balls of radius \( u \) required to cover \( B_1 \)?

\[ N(B_1, \| \cdot \|_2, u) \lesssim \left\{ \frac{1}{u} \right\}^{O(n)} \]

by an easy volume argument
Covering number bound

\[ N(B_1, \| \cdot \|_A, u) \]

- Simpler to imagine: what about \( \ell_2 \)?
- How many \( \ell_2 \) 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 \( \ell_2 \)?
- How many \( \ell_2 \) 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 \( \ell_2 \)?
- How many \( \ell_2 \) 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 \( \ell_2 \)?
- How many \( \ell_2 \) balls of radius \( u \) required to cover \( B_1 \)?

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

- Latter bound is better when \( u \gg 1/\sqrt{n} \).
- **Maurey’s empirical method**: generalizes to arbitrary norms
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 $x \in B^+_1$.

Let $z_1, \ldots, z_t$ be i.i.d. randomized roundings of $x$ to simplex.

The sample mean $\bar{z} = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$.

Let $t$ be large enough that, regardless of $x$,

$$E[\|z - x\|] \leq u.$$ 

All $x$ lie within $u$ of at least one possible $z$.
How many balls of radius $u$ required to cover $B_1^+$?

Consider any $x \in B_1^+$. 

\[ \text{Let } z_1, \ldots, z_t \text{ be i.i.d. randomized roundings of } x \text{ to simplex.} \]

\[ \text{The sample mean } z = \frac{1}{t} \sum z_i \text{ converges to } x \text{ as } t \to \infty. \]

\[ \text{Let } t \text{ be large enough that, regardless of } x, \]

\[ E[\|z - x\|] \leq u. \]

\[ \text{All } x \text{ lie within } u \text{ of at least one possible } z. \]

\[ \text{Then } N(B_1^+, \|\cdot\|, u) \leq \text{number of } z \leq (n+1)t. \]

\[ \text{Only } (n+1)t \text{ possible tuples } (z_1, \ldots, z_t) \Rightarrow z. \]
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.
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.
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.
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.
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.
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 $t \to \infty$. 
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 \( z = \frac{1}{t} \sum z_i \) converges to \( x \) as \( t \to \infty \).
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$. 

Eric Price (MIT)
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$. 
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$. 
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$. 
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$.

Let $t$ be large enough that, regardless of $x$,

$$\mathbb{E}[\|z - x\|] \leq u.$$
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 \( z = \frac{1}{t} \sum_{i} z_i \) converges to \( x \) as \( t \to \infty \).

Let \( t \) be large enough that, regardless of \( x \),

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

All \( x \) lie within \( u \) of at least one possible \( z \).
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$.

Let $t$ be large enough that, regardless of $x$,

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

All $x$ lie within $u$ of at least one possible $z$.

- Then $N(B_1, \|\cdot\|, u) \leq$ number of $z$
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 $z = \frac{1}{t} \sum z_i$ converges to $x$ as $t \to \infty$.

Let $t$ be large enough that, regardless of $x$,

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

All $x$ lie within $u$ of at least one possible $z$.

- Then $N(B_1, \| \cdot \|, u) \leq \text{number of } z \leq (n + 1)^t$.
- Only $(n + 1)^t$ possible tuples $(z_1, \ldots, z_t) \Rightarrow z$. 
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 $\mathbf{z} = \frac{1}{t} \sum z_j$ 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 $N(B_1, \| \cdot \|, u) \leq \text{number of } \mathbf{z} \leq (n + 1)^t$.
- Only $(n + 1)^t$ possible tuples $(z_1, \ldots, z_t) \Rightarrow \mathbf{z}$. 
Will show: \( \mathbb{E}[\|z - x\|_A] \leq \sqrt{B/t} \)

- Let \( z_1, \ldots, z_t \) be i.i.d. randomized roundings of \( x \) to simplex.
- The sample mean \( z = \frac{1}{t} \sum z_i \) converges to \( x \) as \( t \to \infty \).
- Let \( t \) be large enough that, regardless of \( x \),
  \[
  \mathbb{E}[\|z - x\|] \leq u.
  \]
- All \( x \) lie within \( u \) of at least one possible \( z \).
  - Then \( N(B_1, \|\cdot\|, u) \leq \) number of \( z \leq (n + 1)^t \).
  - Only \( (n + 1)^t \) possible tuples \((z_1, \ldots, z_t) \Rightarrow z\).
Covering Number Bound
Maurey’s empirical method

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

- Let \( z_1, \ldots, z_t \) be i.i.d. randomized roundings of \( x \) to simplex.
- The sample mean \( z = \frac{1}{t} \sum z_i \) converges to \( x \) as \( t \to \infty \).
- Let \( t \) be large enough that, regardless of \( x \),
  \[
  \mathbb{E}[\|z - x\|] \leq u.
  \]

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

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

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

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

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

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

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

\[
g_j \sim \mathcal{N}(0, \frac{\text{number of } z_i \text{ at } e_j}{t})
\]

independently in each coordinate.
Covering Number Bound

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

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

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

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

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

(Note: $\mathbb{E}[\| \mathbf{g} \|_2^2] \leq 1 \Rightarrow \mathcal{N}(B_1, \ell_2, u)$.

Eric Price (MIT)  
Fast RIP matrices with fewer rows  
2013-04-05  46 / 52
Covering Number Bound

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

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

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

$$
g_j \sim \mathcal{N}(0, \frac{\text{number of } z_i \text{ at } e_j}{t})
$$

independently in each coordinate.

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

- Goal: \( \mathbb{E}[(\|z - x\|_A)^2] \lesssim \sqrt{B/t} \).
- Symmetrize!

\[
\mathbb{E}[(\|1/t \sum z_i - x\|_A)^2] \lesssim \mathbb{E}[(\|1/t \sum g_i z_i\|_A)^2]
\]

\[
= : \frac{1}{\sqrt{t}} \mathbb{E}[(\|g\|_A)^2]
\]

where \( g \in \mathbb{R}^n \) has

\[g_j \sim \mathcal{N}(0, \frac{\text{number of } z_i \text{ at } e_j}{t})\]

independently in each coordinate.

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

\[ \mathbb{E} \sup_{\|\Phi^T \Phi - I\|} \]

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

Symmetrization

Expected norm of Gaussian

Union bound

Answer!

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

\[ \mathbb{E} \|g\| \]
Progress

\[ \mathbb{E} \sup_{\|\Phi^T\Phi - I\|} \]

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

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

Symmetrization

\[ \mathbb{E} \|g\| \]

Union bound

Answer!

Eric Price (MIT)

Fast RIP matrices with fewer rows

2013-04-05 47 / 52
Bounding the norm (intuition)

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

$Ag$ = subset of $\hat{g}$

$g \in \mathbb{R}^n$ has Gaussian coordinates, $k$-sparse, total variance 1.
Bounding the norm (intuition)

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

$Ag$ = subset of $\hat{g}$

$g \in \mathbb{R}^n$ has Gaussian coordinates, $k$-sparse, total variance 1.
- Each coordinate $\hat{g}_j = F_j g \sim N(0, 1)$. 

Each is $N(0, 1)$

$A_1g$ $A_2g$ $A_3g$ $A_4g$

$\ell_2$ $\ell_2$ $\ell_2$ $\ell_2$

$\ell_\infty$

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

- Just want to bound $\mathbb{E}[\|g\|_A]$.
- Each $g \in \mathbb{R}^n$ has Gaussian coordinates, $k$-sparse, total variance 1.
- Each coordinate $\hat{g}_j = F_jg \sim N(0, 1)$.
- Rudelson-Vershynin: $B = 1$
Bounding the norm (intuition)

- Just want to bound $E[\|g\|_A]$.
- Each is $N(0, 1)$

$Ag$ = subset of $\hat{g}$

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

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

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

$Ag$ = subset of $\hat{g}$

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

Each coordinate $\hat{g}_j = F_jg \sim N(0, 1)$.

Rudelson-Vershynin: $B = 1$

- Naive: $\|\hat{g}\|_\infty \lesssim \sqrt{\log n}$.
- Gives the log $n$ loss in their bound.
Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|g\|_A]$.
  
  $Ag$ = subset of $\hat{g}$

  $\begin{align*}
  A_1g & \quad \quad A_2g \\
  \ell_2 & \quad \ell_2 \\
  A_3g & \quad \quad A_4g \\
  \ell_2 & \quad \ell_2 \\
  \ell_\infty & \\
  \end{align*}$

- Each is $N(0, 1)$

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

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

- If the $\hat{g}_i$ were independent:

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

  $\uparrow$

  Lipschitz concentration

  (just like $\sqrt{n} + \sqrt{\log(1/\delta)}$ in tutorial)
Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|g\|_A]$.
- Each is $N(0, 1)$

$Ag$ = subset of $\hat{g}$

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

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

- If the $\hat{g}_i$ were independent:
  $$\|A_i g\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \quad \text{w.h.p.}$$

- Would get $\mathbb{E}[\|g\|_A] \lesssim \sqrt{B}$
Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|g\|_A]$.
- Each is $N(0, 1)$

$$Ag = \text{subset of } \hat{g}$$

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

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

- If the $\hat{g}_i$ were independent:
  $$\|A_ig\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \text{ w.h.p.}$$

- Would get $\mathbb{E}[\|g\|_A] \lesssim \sqrt{B}$ so union bound just loses a constant.
Bounding the norm (intuition)

- Just want to bound $\mathbb{E}[\|g\|_A]$.
  - Each is $N(0, 1)$

$Ag$ = subset of $\hat{g}$

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

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

- If the $\hat{g}_i$ were independent:
  \[ \|A_i g\|_2 \leq \sqrt{B} + O(\sqrt{\log n}) \text{ w.h.p.} \]

- Would get $\mathbb{E}[\|g\|_A] \lesssim \sqrt{B}$ 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 $\mathbb{E}[\|g\|_A]$.
  - Each is $N(0, 1)$

$Ag$ = subset of $\hat{g}$

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

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

If the $\hat{g}_i$ were independent:

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

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

- They're not independent... but the $A_i$ satisfy “very weak” RIP.
  - Bound $\|A_i g\|_2$ using $\|g\|_2$, which has independent entries.
Bounding the norm (by example)

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

$A_g = \text{subset of } \hat{g}$

Example (1):

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

Fourier transform $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.

$\|\hat{g}\|_{\ell_\infty} \lesssim k^{-1/4} \cdot \sqrt{\log n} \ll 1.$

Hence $\|A_i g\|_2 \ll \sqrt{B}$ for all $i$. 

---

Eric Price (MIT)
Fast RIP matrices with fewer rows
2013-04-05 49 / 52
Bounding the norm (by example)

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

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

![Diagram of $A_g$ subset of $\hat{g}$](image)
Bounding the norm (by example)

Just want to bound \( \mathbb{E}[\|g\|_A] \).

- Recall: \( x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g \).
- The full proof is complicated.
Bounding the norm (by example)

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

- Recall: $x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim \mathcal{N}(0, x_i)$ (the “large $t$” limit).

\begin{align*}
A_1g &\quad \ell_2 \\
A_2g &\quad \ell_2 \\
A_3g &\quad \ell_2 \\
A_4g &\quad \ell_\infty
\end{align*}
Bounding the norm (by example)

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

- Recall: $x \in \Sigma_k / \sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim \mathcal{N}(0, x_i)$ (the “large $t$” limit).
- Two examples:
Bounding the norm (by example)

Just want to bound \( \mathbb{E}[\|g\|_A] \).

- Recall: \( x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g \).
- The full proof is complicated.
- Will assume
  \[ g_i \sim N(0, x_i) \]
  (the “large \( t \)” limit).
- Two examples:
  1. Very concentrated: \( x_1 = 1/\sqrt{k}, \text{rest is zero.} \)
Bounding the norm (by example)

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

- Recall: $x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume
  
  $$g_i \sim \mathcal{N}(0, x_i)$$

  (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 $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.

$$\|\hat{g}\|_\infty \ll k^{-1/4} \cdot \sqrt{\log n} \ll 1.$$ 

Hence $\|A_i g\|_2 \ll \sqrt{B}$ for all $i$. 

Eric Price (MIT)

Fast RIP matrices with fewer rows

2013-04-05 49/52
Bounding the norm (by example)

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

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

  \[ g_i \sim \mathcal{N}(0, x_i) \]

  (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 $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.

\[ \|\hat{g}\|_\infty \ll k^{-1/4} \cdot \sqrt{\log n} \ll 1. \]

Hence $\|A_i g\|_2 \ll \sqrt{B}$ for all $i$. 

\[ Ag = \text{subset of } \hat{g} \]

\[ A_1g \overset{\ell_2}{\longrightarrow} A_2g \overset{\ell_2}{\longrightarrow} A_3g \overset{\ell_2}{\longrightarrow} A_4g \overset{\ell_\infty}{\longrightarrow} g \]

\[ x \]

\[ x_i \]

\[ i \]
Bounding the norm (by example)

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

- Recall: $x \in \Sigma_k / \sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim N(0, x_i)$ (the “large $t$” limit).
- Two examples:
  - Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  - 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 $\mathbb{E}[\|g\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \to z_1, \ldots, z_t \to g$.
- The full proof is complicated.
- Will assume $g_i \sim N(0, x_i)$ (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 $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.
Bounding the norm (by example)

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

- Recall: $x \in \Sigma_{k}/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim N(0, x_i)$ (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 $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.
- $\|\hat{g}\|_\infty \lesssim k^{-1/4} \cdot \sqrt{\log n} \ll 1$. 

$\text{Ag}$

= subset of $\hat{g}$

$A_1g$ $A_2g$ $A_3g$ $A_4g$

$\ell_2$ $\ell_2$ $\ell_2$ $\ell_2$

$\ell_\infty$

$X$

$x_i$

$i$
Bounding the norm (by example)

Just want to bound $E[\|g\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim N(0, x_i)$ (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 $\hat{g}$ is Gaussian with variance $1/\sqrt{k}$.
- $\|\hat{g}\|_\infty \ll k^{-1/4} \cdot \sqrt{\log n} \ll 1$.
- Hence $\|A_i g\|_2 \ll \sqrt{B}$ for all $i$. 
Bounding the norm (by example)

Just want to bound $E[\|g\|_A]$.

- Recall: $x \in \Sigma_k/\sqrt{k} \rightarrow z_1, \ldots, z_t \rightarrow g$.
- The full proof is complicated.
- Will assume $g_i \sim N(0, x_i)$ (the “large $t$” limit).
- Two examples:
  - Very concentrated: $x_1 = 1/\sqrt{k}$, rest is zero.
  - Very spread out: $x_1 = \cdots = x_k = 1/k$.

Example (1):

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

$A g$

\begin{align*}
A_1 g & \quad \ell_2 \\
A_2 g & \quad \ell_2 \\
A_3 g & \quad \ell_2 \\
A_4 g & \quad \ell_\infty \\
\end{align*}

$\hat{g}$

$\Sigma_k/\sqrt{k}$

$z_1, \ldots, z_t$

$g$

$x_1, \ldots, x_k$

$\|\hat{g}\|_\infty$

$k^{-1/4} \cdot \sqrt{\log n}$

$\sqrt{B}$
Bounding the norm (example 2)

Just want to bound \( \mathbb{E}[\|g\|_A] \) when \( g_i \sim N(0, 1/k) \) for \( i \in [k] \).
Bounding the norm (example 2)

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

- \( \|A_i g\|_2 \) is \( C \)-Lipschitz with factor

\[
C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}
\]
Bounding the norm (example 2)

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

- $\|A_i g\|_2$ is $C$-Lipschitz with factor
  $$C = \|A_i\|_{RIP} \cdot 1/\sqrt{k}$$

- Then with high probability,
  $$\|A_i g\|_2 \leq \sqrt{B} + C \sqrt{\log n}$$
Bounding the norm (example 2)

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

- $\|A_i g\|_2$ is $C$-Lipschitz with factor
  \[ C = \|A_i\|_{RIP} \cdot 1/\sqrt{k} \]

- Then with high probability,
  \[ \|A_i g\|_2 \lesssim \sqrt{B} + C \sqrt{\log n} \]

- 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 $\mathbb{E}[\|g\|_A]$ when $g_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i g\|_2$ is $C$-Lipschitz with factor
  \[ C = \|A_i\|_{RIP} \cdot 1/\sqrt{k} \]

- Then with high probability,
  \[ \|A_i g\|_2 \lesssim \sqrt{B} + C \sqrt{\log n} \]

- 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 $\mathbb{E}[\|g\|_A]$ when $g_i \sim N(0, 1/k)$ for $i \in [k]$.

- $\|A_i g\|_2$ is $C$-Lipschitz with factor
  \[
  C = \|A_i\|_{\text{RIP}} \cdot 1/\sqrt{k}
  \]

- Then with high probability,
  \[
  \|A_i g\|_2 \lesssim \sqrt{B} + C \sqrt{\log n}
  \]

- Naive bound:
  \[
  \|A_i\|_{\text{RIP}} \leq \|A_i\|_F = \sqrt{Bk} \implies C \leq \sqrt{B}
  \]

- “Very weak” RIP bound, with $B = \log^c n$:
  \[
  \|A_i\|_{\text{RIP}} \lesssim \|A_i\|_F / \sqrt{\log n}
  \]

- So
  \[
  C \lesssim \sqrt{B/ \log n} \implies \|A_i g\|_2 \lesssim \sqrt{B} \quad \text{w.h.p}
  \]
Bounding the norm (example 2)

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

- $\|A_i g\|_2$ is $C$-Lipschitz with factor
  \[ C = \|A_i\|_{RIP} \cdot 1/\sqrt{k} \]

- Then with high probability,
  \[ \|A_i g\|_2 \lesssim \sqrt{B} + C \sqrt{\log n} \]

- 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
  \[ C \lesssim \sqrt{B/\log n} \implies \|A_i g\|_2 \lesssim \sqrt{B} \text{ w.h.p} \]

- So
  \[ \mathbb{E}\|g\|_A = \max \|A_i g\|_2 \lesssim \sqrt{B}. \]
Union bound just loses a constant factor
Unrolling everything

\[ \mathbb{E} \sup_{\|\Phi^T \Phi - I\|} \]

\[ \gamma_2(\Sigma_k, \|\cdot\|) \]

\[ N(\Sigma_k, \|\cdot\|, u) \]

\[ \text{Expected sup deviation} \]

\[ \text{Expected sup norm of Gaussian} \]

\[ \text{Dudley} \]

\[ \text{Covering number} \]

\[ \text{Krahmer Mendelson Rauhut '13} \]

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

\[ \sqrt{B/t} \]

\[ \mathbb{E} \|g\| \]

\[ \sqrt{B} \]

\[ \text{Expected deviation} \]

\[ \text{Symmetrization} \]

\[ \text{Expected norm of Gaussian} \]

\[ \text{Union bound} \]

\[ \text{Answer!} \]

Sample mean \( z \) expects to lie within \( u \) of \( x \) for \( t \geq B/u^2 \)
Unrolling everything

\[ \mathbb{E} \sup_{\|\Phi^T\Phi - I\|} \]

\[ \gamma_2(\Sigma_k, \|\cdot\|) \]

\[ n^{kB/u^2} \]

\[ N(\Sigma_k, \|\cdot\|, u) \]

**Expected sup deviation**

Krahmer, Mendelson, Rauhut '13

Expected sup norm of Gaussian

Dudley

Maurey: randomize

Expected norm of Gaussian

Symmetrization

Union bound

Answer!

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

\[ \sqrt{\frac{B}{t}} \]

\[ \mathbb{E} \|g\| \]

\[ \sqrt{B} \]

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

\[ \mathbb{E} \sup_{\| \Phi^T \Phi - I \|} \]

\[ \gamma_2(\Sigma_k, \| \cdot \|) \]

\[ \sqrt{kB \log^3 n} \]

\[ n^{kB/u^2} \]

\[ N(\Sigma_k, \| \cdot \|, u) \]

Entropy integral is \( \sqrt{kB \log^3 n} \)
Unrolling everything

\[ \sqrt{\frac{k \log^3 n}{m}} \]

\[ \mathbb{E} \sup_{\|\Phi^T \Phi - I\|} \]

\[ k \log^3 n \]

\[ \gamma_2(\Sigma_k, \|\cdot\|) \]

\[ dB \log^3 n \]

\[ \frac{B}{t} \]

\[ n^{kB/u^2} \]

\[ N(\Sigma_k, \|\cdot\|, u) \]

\[ \sqrt{\frac{B}{t}} \]

\[ \sqrt{k \log^3 n} \]

RIP constant \( \epsilon \leq \sqrt{\frac{k \log^3 n}{m}} \)

Eric Price (MIT) Fast RIP matrices with fewer rows 2013-04-05 51 / 52
Summary and Open Questions

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

- 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 $\log n$ 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?
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?
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
  \[ \gamma_2 \approx \sup_{x} \sum \sqrt{\log |A_{i+1}|} d(x, A_i) \]
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

$$E := \mathbb{E}[\|1/t \sum z_i - x\|] \leq u.$$
Covering Number Bound
Maurey’s empirical method

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

$$E := \mathbb{E}[\| \frac{1}{t} \sum z_i - x \|] \leq u.$$
Covering Number Bound
Maurey’s empirical method

- Answer is $n^t$, where $t$ is such that
  $$E := \mathbb{E}[\| \frac{1}{t} \sum Z_i - X \|] \leq u.$$  

- Symmetrize:
  $$E \lesssim \frac{1}{t} \mathbb{E}[\| \sum g_i z_i \|]$$
  for $g_i \sim N(0, 1)$ i.i.d.

Eric Price (MIT)
Fast RIP matrices with fewer rows
2013-04-05 55 / 52
Covering Number Bound
Maurey’s empirical method

- Answer is \( n^t \), where \( t \) is such that

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

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

- 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 $\ell_2$,
  \[
  \frac{1}{t} \mathbb{E}[\| g \|_2] \leq \frac{1}{t} \mathbb{E}[\| g \|_2^2]^{1/2} = \frac{\sqrt{\text{number nonzero } z_i}}{t} \leq \frac{1}{\sqrt{t}}.
  \]
  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 \( g \).

\[
G(x) = \mathbb{E}_{z,g} \|g\|_A
\]
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 \( g \).

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

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

\[
G(x) \leq G(x_{\text{large}}) + G(x_{\text{small}})
\]

- \( x_{\text{large}} \): 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 \( g \).

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

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

\[
G(x) \leq G(x_{\text{large}}) + G(x_{\text{small}})
\]

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

\[
\|x_{\text{large}}\|_1
\]
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 \( g \).

\[
G(x) = E_{z,g} \|g\|_A
\]

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

\[
G(x) \leq G(x_{\text{large}}) + G(x_{\text{small}})
\]

- \( x_{\text{large}} \): Locations where \( x_i > (\log n)/k \)
  - Bound:
    \[
    \|x_{\text{large}}\|_1
    \]
  - Given \( \|x\|_2^2 \leq 1/k \), maximal \( \|x_{\text{large}}\|_1 \) 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 $g$.

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

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

\[ G(x) \leq G(x_{\text{large}}) + G(x_{\text{small}}) \]

- $x_{\text{large}}$: Locations where $x_i > (\log n)/k$
  - Bound:
    \[ \|x_{\text{large}}\|_1 \]
  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{\text{large}}\|_1$ 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 $g$.

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

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

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

- $x_{\text{large}}$: Locations where $x_i > (\log n)/k$
  - Bound:
    $$\|x_{\text{large}}\|_1 \leq 1 / \log n.$$ 
  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{\text{large}}\|_1$ 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 $g$.

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

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

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

- $x_{\text{large}}$: Locations where $x_i > (\log n)/k$
  - Bound:
    $$\|x_{\text{large}}\|_1 \leq 1/\log n.$$  

  - Given $\|x\|_2^2 \leq 1/k$, maximal $\|x_{\text{large}}\|_1$ 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.
  \[
  \mathcal{G}(x) \leq \mathcal{G}(x_{\text{large}}) + \mathcal{G}(x_{\text{small}})
  \]

- \( x_{\text{large}} \): Locations where \( x_i > (\log n)/k \)
  - Bound:
    \[
    \|x_{\text{large}}\|_1 \leq 1 / \log n.
    \]
    - Given \( \|x\|_2^2 \leq 1/k \), maximal \( \|x_{\text{large}}\|_1 \) 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 $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$. 

---

Eric Price (MIT)   Fast RIP matrices with fewer rows  2013-04-05   57 / 52
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 $\sigma_i^2 = \#z_j$ 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 \( \sigma_i^2 = \frac{\# z_j \text{ at vertex } e_i}{t^2} \approx \frac{x_i}{t} \).
- \( \| A_i g \|_2 \) 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 $\sigma_i^2 = \frac{\#z_j \text{ at vertex } e_i}{t^2} \approx \frac{x_i}{t}$.
- $\|A_i g\|_2$ 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 $\sigma_i^2 = \{\#z_j \text{ at vertex } e_i\}/t^2 \approx x_i/t$.
- $\|A_ig\|_2$ 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
  \[ C = \|A_i\|_{RIP} \cdot \|\sigma\|_\infty \]
- 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 $\sigma_i^2 = \frac{\#z_j \text{ at vertex } e_i}{t^2} \approx x_i/ t$.
- $\|A_i g\|_2$ 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 $\mathbb{E}\|g\|_A = \max \|A_i g\|_2 \lesssim \sqrt{B/t}$. 