Simple Analysis of Sparse, Sign-Consistent JL

Meena Jagadeesan

Harvard University

RANDOM 2019

Linear dimensionality reduction: ℓ_2-to-ℓ_2

Informal goal: Project vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$ (for $m << n$) with a linear map while "preserving geometry" (i.e. $\|f(x) - f(y)\|_2 \approx \|x - y\|_2$).

Linear dimensionality reduction: ℓ_2-to-ℓ_2

Informal goal: Project vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$ (for $m << n$) with a linear map while "preserving geometry" (i.e. $\|f(x) - f(y)\|_2 \approx \|x - y\|_2$).

Many applications:

- Feature hashing (Weinberger et al. '09, Dahlgaard et al. '17, Freksen et al. '18, etc.)
- Numerical linear algebra (Clarkson and Woodruff '12, Nelson and Nguyen '14, etc.)
- Approximate nearest neighbors (Ailon and Chazelle '09, etc.)
- k-means/k-medians (Makarychev, Makarychev, Razenshteyn '18)

Linear dimensionality reduction: ℓ_2-to-ℓ_2

Informal goal: Project vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$ (for $m << n$) with a linear map while "preserving geometry" (i.e. $\|f(x) - f(y)\|_2 \approx \|x - y\|_2$).

Many applications:

- Feature hashing (Weinberger et al. '09, Dahlgaard et al. '17, Freksen et al. '18, etc.)
- Numerical linear algebra (Clarkson and Woodruff '12, Nelson and Nguyen '14, etc.)
- Approximate nearest neighbors (Ailon and Chazelle '09, etc.)
- k-means/k-medians (Makarychev, Makarychev, Razenshteyn '18)
- <u>Compression in the brain</u> (Allen-Zhu, Gelashvili, Micali, Shavit '15)

Mathematical framework

Mathematical framework

Use a probability distribution $\mathcal{M}$ over linear maps $\mathbb{R}^n \to \mathbb{R}^m$ ($m << n$).

Geometry-preserving property: for *each* $x \in \mathbb{R}^n$

$$\mathbb{P}_{M \in \mathcal{M}}[(1-\epsilon)\|x\|_2 \leq \|Mx\|_2 \leq (1+\epsilon)\|x\|_2] > 1-\delta.$$

Mathematical framework

Use a probability distribution $\mathcal{M}$ over linear maps $\mathbb{R}^n \to \mathbb{R}^m$ ($m << n$).

Geometry-preserving property: for *each* $x \in \mathbb{R}^n$

$$\mathbb{P}_{M \in \mathcal{M}}[(1-\epsilon)\|x\|_2 \leq \|Mx\|_2 \leq (1+\epsilon)\|x\|_2] > 1 - \delta.$$

Fundamental result of linear dimensionality reduction:

Lemma (Distributional Johnson-Lindenstrauss Lemma)

Can obtain $m = \Theta(\epsilon^{-2}\log(1/\delta))$ using $\mathcal{M}$ with i.i.d gaussian entries.

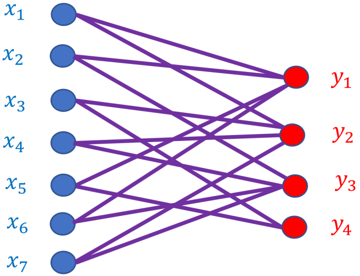
This dimension is actually optimal for any distribution over linear maps (Kane et al. '11, Jayram and Woodruff '11).

Optimality for N-point version (Larsen and Nelson '17)

Application to information compression in the brain

Application to information compression in the brain

Convergent pathways compress information w/o losing the ability to perform computations.

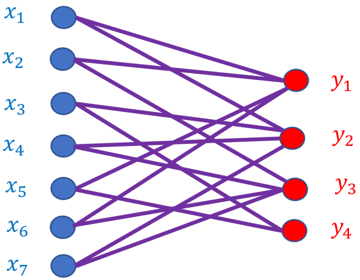


A model (Ganguli, Sompolinsky '12)

- Source information: $x \in \mathbb{R}^n$
- Target information: $y \in \mathbb{R}^m$
- Synaptic connections: a random matrix $M \in \mathbb{R}^{m \times n}$

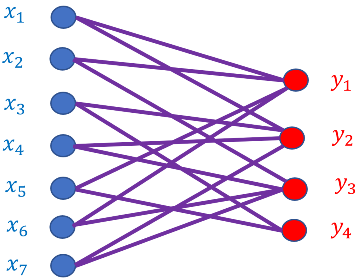
Biological Constraints on Supp($\mathcal{M}$)

(As per Ganguli and Sompolinsky's model)



Biological Constraints on Supp($\mathcal{M}$)

(As per Ganguli and Sompolinsky's model)



Sparsity: every column has $\leq s$ nonzero entries.

- Neurons connected to few post-synaptic neurons

Sign-consistency: in each column, nonzero entries are *all positive* or *all negative*

- Neurons are excitatory or inhibitory

JL for sparse matrices

Sparsity is also more generally useful for reducing projection time.

Informal Construction (Sparse JL)

Uniformly choose s nonzero entries per column; i.i.d signs for nonzero entries

Can set $m = \Theta(\epsilon^{-2} \log(1/\delta))$ and $s = \Theta(\epsilon^{-1} \log(1/\delta)$ (Kane and Nelson, J. ACM '12)

Can set $m = \min(2\epsilon^{-2}/\delta, \Theta(\epsilon^{-2} \log(1/\delta) B))$ and $s = \Theta(\epsilon^{-1} \log(1/\delta)/\log B)$ (Cohen, SODA '16)

Sparse, Sign-Consistent JL (Allen-Zhu et al.)

Sparse, Sign-Consistent JL (Allen-Zhu et al.)

Uniformly choose s nonzero entries per column; i.i.d signs for each column:

Sparse, Sign-Consistent JL (Allen-Zhu et al.)

Uniformly choose s nonzero entries per column; i.i.d signs for each column:

Informal Construction (Sparse, Sign-Consistent JL)

$\mathcal{M}$ is defined so the (r, i)th entry is $\sigma_i \eta_{r,i}/\sqrt{s}$ where:

- *σ_i are i.i.d. Rademachers (random signs)*
- *$\eta_{r,i}$ are $\{0, 1\}$ rvs s.t. $\sum_{r=1}^{m} \eta_{r,i} = s$ and w/ mild assumptions*

Sparse, Sign-Consistent JL (Allen-Zhu et al.)

Uniformly choose s nonzero entries per column; i.i.d signs for each column:

Informal Construction (Sparse, Sign-Consistent JL)

$\mathcal{M}$ is defined so the (r, i)th entry is $\sigma_i \eta_{r,i}/\sqrt{s}$ where:

- *σ_i are i.i.d. Rademachers (random signs)*
- *$\eta_{r,i}$ are $\{0, 1\}$ rvs s.t. $\sum_{r=1}^{m} \eta_{r,i} = s$ and w/ mild assumptions*

Can set $m = \Theta(\epsilon^{-2} \log^2(1/\delta))$, $s = \Theta(\epsilon^{-1} \log(1/\delta))$ (Allen-Zhu, Gelashvili, Micali, and Shavit, PNAS '15)

This work

Simplify and generalize the analysis of sparse, sign-consistent JL.

Theorem (Informal)

For any $e \leq B \leq 1/\delta$*, can set* $m = \Theta(\epsilon^{-2}\log^2(1/\delta)B/\log^2(B))$ *and* $s = \Theta(\epsilon^{-1}\log(1/\delta)/\log B)$ *for sparse, sign-consistent JL.*

This work

Simplify and generalize the analysis of sparse, sign-consistent JL.

Theorem (Informal)

For any $e \leq B \leq 1/\delta$, can set $m = \Theta(\epsilon^{-2}\log^2(1/\delta)B/\log^2(B))$ and $s = \Theta(\epsilon^{-1}\log(1/\delta)/\log B)$ for sparse, sign-consistent JL.

(Bears resemblance to dim/sparsity tradeoffs for sparse JL by Cohen '16.)

This work

Simplify and generalize the analysis of sparse, sign-consistent JL.

Theorem (Informal)

For any $e \leq B \leq 1/\delta$*, can set* $m = \Theta(\epsilon^{-2} \log^2(1/\delta) B / \log^2(B))$ *and* $s = \Theta(\epsilon^{-1} \log(1/\delta) / \log B)$ *for sparse, sign-consistent JL.*

(Bears resemblance to dim/sparsity tradeoffs for sparse JL by Cohen '16.)

Proof method: Different techniques for analyzing moments of JL error terms which turn to be "needed" for this setting

This work

Simplify and generalize the analysis of sparse, sign-consistent JL.

Theorem (Informal)

For any $e \leq B \leq 1/\delta$, *can set* $m = \Theta(\epsilon^{-2}\log^2(1/\delta)B/\log^2(B))$ *and* $s = \Theta(\epsilon^{-1}\log(1/\delta)/\log B)$ *for sparse, sign-consistent JL.*

(Bears resemblance to dim/sparsity tradeoffs for sparse JL by Cohen '16.)

Proof method: Different techniques for analyzing moments of JL error terms which turn to be "needed" for this setting

- Intuition critical in my follow-up work on analyzing sparse JL for feature hashing (NeurIPS '19, to appear).

This work

Simplify and generalize the analysis of sparse, sign-consistent JL.

Theorem (Informal)

For any $e \leq B \leq 1/\delta$, *can set* $m = \Theta(\epsilon^{-2} \log^2(1/\delta) B / \log^2(B))$ *and* $s = \Theta(\epsilon^{-1} \log(1/\delta) / \log B)$ *for sparse, sign-consistent JL.*

(Bears resemblance to dim/sparsity tradeoffs for sparse JL by Cohen '16.)

Proof method: Different techniques for analyzing moments of JL error terms which turn to be "needed" for this setting

- Intuition critical in my follow-up work on analyzing sparse JL for feature hashing (NeurIPS '19, to appear).

Remainder of the talk will focus on the proof method.

High-level approach

Need to show $\mathbb{P}_{M \in \mathcal{M}}[(1-\epsilon)\|x\|_2 \leq \|Mx\|_2 \leq (1+\epsilon)\|x\|_2] > 1 - \delta$.

High-level approach

Need to show $\mathbb{P}_{M \in \mathcal{M}}[(1-\epsilon)\left\|x\right\|_2 \leq \left\|Mx\right\|_2 \leq (1+\epsilon)\left\|x\right\|_2] > 1-\delta$.

Analyze high moment of $Z := \left\|Mx\right\|_2^2 - 1$ for all $x \in B_{\ell_2}$ (like in previous work).

High-level approach

Need to show $\mathbb{P}_{M \in \mathcal{M}}[(1-\epsilon)\|x\|_2 \leq \|Mx\|_2 \leq (1+\epsilon)\|x\|_2] > 1-\delta$.

Analyze high moment of $Z := \|Mx\|_2^2 - 1$ for all $x \in B_{\ell_2}$ (like in previous work).

For sparse, sign-consistent JL:

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s}\sum_{i \neq j}\sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i}\eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

1. Combinatorics (Allen-Zhu et al. '15, Kane and Nelson '12, Freksen et al. '18, etc.)

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

1. Combinatorics (Allen-Zhu et al. '15, Kane and Nelson '12, Freksen et al. '18, etc.)
   - Intricate; unclear how to get dim-sparsity tradeoffs w/ Allen-Zhu et al.

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

1. Combinatorics (Allen-Zhu et al. '15, Kane and Nelson '12, Freksen et al. '18, etc.)
    - Intricate; unclear how to get dim-sparsity tradeoffs w/ Allen-Zhu et al.
2. (Sub)-gaussian quadratic form bounds (Cohen et al. '18, for sparse JL)

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

1. Combinatorics (Allen-Zhu et al. '15, Kane and Nelson '12, Freksen et al. '18, etc.)
    - Intricate; unclear how to get dim-sparsity tradeoffs w/ Allen-Zhu et al.
2. (Sub)-gaussian quadratic form bounds (Cohen et al. '18, for sparse JL)
    - Using the gaussian bound is too weak for this setting!

Analyzing the moments of Z

$$Z := \|Mx\|_2^2 - 1 = \frac{1}{s} \sum_{i \neq j} \sum_{r=1}^{m} \sigma_i \sigma_j \eta_{r,i} \eta_{r,j} x_i x_j.$$

Limitations of existing approaches for bounding moments of Z:

1. Combinatorics (Allen-Zhu et al. '15, Kane and Nelson '12, Freksen et al. '18, etc.)
   - Intricate; unclear how to get dim-sparsity tradeoffs w/ Allen-Zhu et al.
2. (Sub)-gaussian quadratic form bounds (Cohen et al. '18, for sparse JL)
   - Using the gaussian bound is too weak for this setting!

My key ingredient: more precise quadratic form bounds

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s}\sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i}\eta_{r,j}\right)$

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s} \sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i} \eta_{r,j} \right) = \sigma^T A_\eta \sigma.$

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s} \sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i} \eta_{r,j} \right) = \sigma^T A_\eta \sigma.$

$\implies \mathbb{E}[Z^p] = \mathbb{E}_\eta \left[\mathbb{E}_\sigma \left[\left(\sigma^T A_\eta \sigma \right)^p \right] \right].$

$\mathbb{E}_\sigma \left[\left(\sigma^T A_\eta \sigma \right)^p \right]$ is a <u>moment of a Rademacher quadratic form</u>.

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s} \sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i} \eta_{r,j}\right) = \sigma^T A_\eta \sigma.$

$\implies \mathbb{E}[Z^p] = \mathbb{E}_\eta \left[\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right]\right].$

$\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right]$ is a <u>moment of a Rademacher quadratic form</u>.

CJN '18 (for sparse JL) uses $\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right] \lesssim \mathbb{E}_g\left[\left(g^T A_\eta g\right)^p\right]$.

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s} \sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i} \eta_{r,j} \right) = \sigma^T A_\eta \sigma.$

$\implies \mathbb{E}[Z^p] = \mathbb{E}_\eta \left[\mathbb{E}_\sigma \left[\left(\sigma^T A_\eta \sigma \right)^p \right] \right].$

$\mathbb{E}_\sigma \left[\left(\sigma^T A_\eta \sigma \right)^p \right]$ is a moment of a Rademacher quadratic form.

CJN '18 (for sparse JL) uses $\mathbb{E}_\sigma \left[\left(\sigma^T A_\eta \sigma \right)^p \right] \lesssim \mathbb{E}_g \left[\left(g^T A_\eta g \right)^p \right]$.

But, we show using $\mathbb{E}_g \left[\left(g^T A_\eta g \right)^p \right]$ is too weak for this setting!

- $x = [1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0]$ already yields sub-optimal dimension m

Expressing Z as a Rademacher quadratic form

$Z = \frac{1}{s} \sum_{i \neq j} \sigma_i \sigma_j x_i x_j \left(\sum_{r=1}^{m} \eta_{r,i} \eta_{r,j}\right) = \sigma^T A_\eta \sigma.$

$\implies \mathbb{E}[Z^p] = \mathbb{E}_\eta \left[\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right]\right].$

$\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right]$ is a moment of a Rademacher quadratic form.

CJN '18 (for sparse JL) uses $\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right] \lesssim \mathbb{E}_g\left[\left(g^T A_\eta g\right)^p\right]$.

But, we show using $\mathbb{E}_g\left[\left(g^T A_\eta g\right)^p\right]$ is too weak for this setting!

- $x = [1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0]$ already yields sub-optimal dimension m

Thus, we need a Rademacher-specific bound for $\mathbb{E}_\sigma\left[\left(\sigma^T A_\eta \sigma\right)^p\right]$.

A digression on Rademachers vs. gaussians for linear forms

$\|Y\|_p := (\mathbb{E}[Y^p])^{1/p}$; σ_i i.i.d. Rademachers; g_i i.i.d gaussians;
$|a_1| \geq \ldots \geq |a_n|$ scalar coefficients

A digression on Rademachers vs. gaussians for linear forms

$\|Y\|_p := (\mathbb{E}[Y^p])^{1/p}$; σ_i i.i.d. Rademachers; g_i i.i.d gaussians; $|a_1| \geq \ldots \geq |a_n|$ scalar coefficients

Khintchine bound: $\|\sum_{i=1}^n a_i\sigma_i\|_p \lesssim \|\sum_{i=1}^n a_i g_i\|_p \sim \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2}$

A digression on Rademachers vs. gaussians for linear forms

$\|Y\|_p := (\mathbb{E}[Y^p])^{1/p}$; σ_i i.i.d. Rademachers; g_i i.i.d gaussians; $|a_1| \geq \ldots \geq |a_n|$ scalar coefficients

Khintchine bound: $\left\|\sum_{i=1}^n a_i \sigma_i\right\|_p \lesssim \left\|\sum_{i=1}^n a_i g_i\right\|_p \sim \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2}$

$\lim_{p\to\infty} \left\|\sum_{i=1}^n a_i \sigma_i\right\|_p \leq \sum_{i=1}^n |a_i| < \infty$, but $\lim_{p\to\infty} \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2} = \infty$.

A digression on Rademachers vs. gaussians for linear forms

$\|Y\|_p := (\mathbb{E}[Y^p])^{1/p}$; σ_i i.i.d. Rademachers; g_i i.i.d gaussians; $|a_1| \geq \ldots \geq |a_n|$ scalar coefficients

Khintchine bound: $\|\sum_{i=1}^n a_i \sigma_i\|_p \lesssim \|\sum_{i=1}^n a_i g_i\|_p \sim \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2}$

$\lim_{p\to\infty} \|\sum_{i=1}^n a_i \sigma_i\|_p \leq \sum_{i=1}^n |a_i| < \infty$, but $\lim_{p\to\infty} \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2} = \infty$.

Khintchine is *not tight* for Rademachers for large p.

A digression on Rademachers vs. gaussians for linear forms

$\|Y\|_p := (\mathbb{E}[Y^p])^{1/p}$; σ_i i.i.d. Rademachers; g_i i.i.d gaussians; $|a_1| \geq \ldots \geq |a_n|$ scalar coefficients

Khintchine bound: $\|\sum_{i=1}^n a_i \sigma_i\|_p \lesssim \|\sum_{i=1}^n a_i g_i\|_p \sim \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2}$

$\lim_{p\to\infty} \|\sum_{i=1}^n a_i \sigma_i\|_p \leq \sum_{i=1}^n |a_i| < \infty$, but $\lim_{p\to\infty} \sqrt{p}\sqrt{\sum_{i=1}^n a_i^2} = \infty$.

Khintchine is *not tight* for Rademachers for large p.

Turns out $\|\sum_{i=1}^n a_i \sigma_i\|_p \sim \left|\sum_{i=1}^p a_i\right| + \sqrt{p}\sqrt{\sum_{i>p} a_i^2}$ (Hitzchenko '93).

Generalizing to quadratic forms of Rademachers

Generalizing to quadratic forms of Rademachers

Lemma

If $(A_{i,j})$ is a symmetric $n \times n$ matrix with zero diagonal and p even, then

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \sigma_i \sigma_j \right\|_p \lesssim \left(\sum_{i=1}^{\min(p,n)} \sum_{j=1}^{\min(p,n)} |A_{i,j}| \right) + \sqrt{p} \sqrt{\sum_{i=1}^{n} \left\| \sum_{j>p} A_{i,j} \sigma_j \right\|_p^2}.$$

Generalizing to quadratic forms of Rademachers

Lemma

If $(A_{i,j})$ is a symmetric $n \times n$ matrix with zero diagonal and p even, then

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \sigma_i \sigma_j \right\|_p \lesssim \left(\sum_{i=1}^{\min(p,n)} \sum_{j=1}^{\min(p,n)} |A_{i,j}| \right) + \sqrt{p} \sqrt{\sum_{i=1}^{n} \left\| \sum_{j>p} A_{i,j} \sigma_j \right\|_p^2}.$$

- An interpolation of ℓ_1 and ℓ_2 bounds.

Generalizing to quadratic forms of Rademachers

Lemma

If $(A_{i,j})$ is a symmetric $n \times n$ matrix with zero diagonal and p even, then

$$\left\|\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\sigma_i\sigma_j\right\|_p \lesssim \left(\sum_{i=1}^{\min(p,n)}\sum_{j=1}^{\min(p,n)} |A_{i,j}|\right) + \sqrt{p}\sqrt{\sum_{i=1}^{n}\left\|\sum_{j>p} A_{i,j}\sigma_j\right\|_p^2}.$$

- An interpolation of ℓ_1 and ℓ_2 bounds.
- When $p \to \infty$, bounded by $\sum_{i=1}^{\min(p,n)}\sum_{j=1}^{\min(p,n)} |A_{i,j}| < \infty$.

Generalizing to quadratic forms of Rademachers

Lemma

If $(A_{i,j})$ is a symmetric $n \times n$ matrix with zero diagonal and p even, then

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \sigma_i \sigma_j \right\|_p \lesssim \left(\sum_{i=1}^{\min(p,n)} \sum_{j=1}^{\min(p,n)} |A_{i,j}| \right) + \sqrt{p} \sqrt{\sum_{i=1}^{n} \left\| \sum_{j>p} A_{i,j} \sigma_j \right\|_p^2}.$$

- An interpolation of ℓ_1 and ℓ_2 bounds.
- When $p \to \infty$, bounded by $\sum_{i=1}^{\min(p,n)} \sum_{j=1}^{\min(p,n)} |A_{i,j}| < \infty$.

(Tight bound on $\left\| \sigma^T A \sigma \right\|_p$ (Latała '99) messy when A a *random matrix*.)

Follow-up work that uses intuition from these methods

Follow-up work that uses intuition from these methods

"Understanding Sparse JL for Feature Hashing" (NeurIPS 2019)

I study sparse JL on feature vectors.

- Model: limit to vectors x with "small" ℓ_∞-to-ℓ_2 norm ratio
- $s = 1$ understood (Weinberger et al '09, Dahlgaard et al. '17, Freksen et al. '18, etc.)

My main result: Generalization to $s > 1$.

- Tight tradeoff between ℓ_∞-to-ℓ_2 ratio, s, m, ϵ, and δ for sparse JL
- $\implies$ Even (small) $s > 1$ can be much better than $s = 1$.

Follow-up work that uses intuition from these methods

"Understanding Sparse JL for Feature Hashing" (NeurIPS 2019)

I study sparse JL on feature vectors.

- Model: limit to vectors x with "small" ℓ_∞-to-ℓ_2 norm ratio
- $s = 1$ understood (Weinberger et al '09, Dahlgaard et al. '17, Freksen et al. '18, etc.)

My main result: Generalization to $s > 1$.

- Tight tradeoff between ℓ_∞-to-ℓ_2 ratio, s, m, ϵ, and δ for sparse JL
- $\implies$ Even (small) $s > 1$ can be much better than $s = 1$.

Similarly unclear how to adapt combinatorics; gaussian bounds too weak.

Tractable Rademacher-specific bounds are the key technical tool.

Conclusion

- Simplified and generalized the analysis of sparse, sign-consistent JL (Allen-Zhu, Gelashvili, Micali, Shavit '15).
- Specifically obtained dimensionality-sparsity tradeoffs $m = \Theta(\epsilon^{-2} \log^2(1/\delta) B / \log^2(B))$ and $s = \Theta(\epsilon^{-1} \log(1/\delta) / \log B)$.
- Introduced a simple moment bound for Rademacher quadratic forms which enables a simpler analysis of sparse, sign-consistent JL, and could be of broader use.