Matrix-valued Chernoff Bounds and Applications

China Theory Week

Anastasios Zouzias

University of Toronto

September 2010

Introduction

- Probability theory: backbone in analysis of randomized algorithms
- Random sampling is the most fundamental technique
- Several inequalities for analyzing approximation: Markov, Chebyshev, Chernoff, Azuma, etc.

In this talk: Discuss recent matrix-valued probabilistic inequalities and their applications

Agenda:

1. Review real-valued probabilistic inequalities
2. Present recent matrix-valued variants
3. A low rank matrix-valued inequality
4. Two applications: matrix sparsification, approximate matrix multiplication

Introduction

- Probability theory: backbone in analysis of randomized algorithms
- Random sampling is the most fundamental technique
- Several inequalities for analyzing approximation: Markov, Chebyshev, Chernoff, Azuma, etc.

In this talk: Discuss recent matrix-valued probabilistic inequalities and their applications

Agenda:

1. Review real-valued probabilistic inequalities
2. Present recent matrix-valued variants
3. A low rank matrix-valued inequality
4. Two applications: matrix sparsification, approximate matrix multiplication

Introduction

- Probability theory: backbone in analysis of randomized algorithms
- Random sampling is the most fundamental technique
- Several inequalities for analyzing approximation: Markov, Chebyshev, Chernoff, Azuma, etc.

In this talk: Discuss recent matrix-valued probabilistic inequalities and their applications

Agenda:

1. Review real-valued probabilistic inequalities
2. Present recent matrix-valued variants
3. A low rank matrix-valued inequality
4. Two applications: matrix sparsification, approximate matrix multiplication

Introduction

- Probability theory: backbone in analysis of randomized algorithms
- Random sampling is the most fundamental technique
- Several inequalities for analyzing approximation: Markov, Chebyshev, Chernoff, Azuma, etc.

In this talk: Discuss recent matrix-valued probabilistic inequalities and their applications

Agenda:

1. Review real-valued probabilistic inequalities
2. Present recent matrix-valued variants
3. A low rank matrix-valued inequality
4. Two applications: matrix sparsification, approximate matrix multiplication

Introduction

- Probability theory: backbone in analysis of randomized algorithms
- Random sampling is the most fundamental technique
- Several inequalities for analyzing approximation: Markov, Chebyshev, Chernoff, Azuma, etc.

In this talk: Discuss recent matrix-valued probabilistic inequalities and their applications

Agenda:

1. Review real-valued probabilistic inequalities
2. Present recent matrix-valued variants
3. A low rank matrix-valued inequality
4. Two applications: matrix sparsification, approximate matrix multiplication

Law of Large Numbers

- Fundamental principle of random sampling:
  Law of Large Numbers (LLN)
  It states that the empirical average converges to true average
- Classical form: for reals rather than matrices
  Let $X_1, \ldots, X_t$ be independent copies of a random variable X
  Goal: estimate the mean $\mathbb{E}[X]$ using samples $X_1, \ldots, X_t$
- Approximate by the empirical mean

$$\frac{1}{t} \sum_{i=1}^{t} X_t \approx \mathbb{E}[X]$$

- How good is the approximation (non-asymptotics)?

Law of Large Numbers

- Fundamental principle of random sampling:
  Law of Large Numbers (LLN)
  It states that the empirical average converges to true average
- Classical form: for reals rather than matrices
  Let $X_1, \ldots, X_t$ be independent copies of a random variable X
  Goal: estimate the mean $\mathbb{E}[X]$ using samples $X_1, \ldots, X_t$
- Approximate by the empirical mean

$$\frac{1}{t}\sum_{i=1}^{t} X_t \approx \mathbb{E}[X]$$

- How good is the approximation (non-asymptotics)?

Question:
Is there a matrix-valued LLN?

Matrix-valued Random Variables

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *matrix-valued random variable* is a measurable function

$$M : \Omega \to \mathbb{R}^{d\times d}$$

- Its expectation is a $d \times d$ matrix, denote by $\mathbb{E}[M] \in \mathbb{R}^{d\times d}$
- Self-adjoint matrix-valued random variable: $M : \Omega \to \mathcal{S}^{d\times d}$
- Caveat: Entries may or may not be correlated with each other

Matrix-valued Random Variables

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *matrix-valued random variable* is a measurable function
$$M : \Omega \to \mathbb{R}^{d \times d}$$
- Its expectation is a $d \times d$ matrix, denote by $\mathbb{E}[M] \in \mathbb{R}^{d \times d}$
- Self-adjoint matrix-valued random variable: $M : \Omega \to \mathcal{S}^{d \times d}$
- Caveat: Entries may or may not be correlated with each other

Matrix-valued random variable

is a random matrix with (possibly) correlated entries

Real-valued Probabilistic Inequalities

Lemma (Markov)

Let $X \geq 0$ be a real-valued random variable (r.v.) and $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Real-valued Probabilistic Inequalities

Lemma (Markov)

Let $X \geq 0$ *be a real-valued random variable (r.v.) and* $a > 0$*. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Lemma (Chernoff-Hoeffding)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Real-valued Probabilistic Inequalities

Lemma (Chernoff-Hoeffding)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Lemma (Bernstein)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$ *and* $\mathrm{Var}(X) \leq \rho^2$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\rho^2 + \gamma\varepsilon/3}\right).$$

Real-valued Probabilistic Inequalities

Lemma (Chernoff-Hoeffding)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Lemma (Bernstein)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$ *and* $\mathrm{Var}(X) \leq \rho^2$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\rho^2 + \gamma\varepsilon/3}\right).$$

...and many more...

Real-valued Probabilistic Inequalities

Lemma (Chernoff-Hoeffding)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left| \frac{1}{t} \sum_{i=1}^{t} X_i - \mathbb{E}[X] \right| > \varepsilon \right) \leq 2 \exp\left(-C \frac{\varepsilon^2 t}{\gamma^2} \right).$$

Lemma (Bernstein)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$ *and* $\mathrm{Var}(X) \leq \rho^2$*, then*

$$\mathbb{P}\left(\left| \frac{1}{t} \sum_{i=1}^{t} X_i - \mathbb{E}[X] \right| > \varepsilon \right) \leq 2 \exp\left(-C \frac{\varepsilon^2 t}{\rho^2 + \gamma\varepsilon/3} \right).$$

Question: How would the matrix-valued generalizations look like?

Real-valued to Matrix-valued

- Is there a meaningful way to generalize the real-valued inequalities to *matrix*-valued?
- Would these inequalities be useful to us?

Real-valued to Matrix-valued

- Is there a meaningful way to generalize the real-valued inequalities to *matrix*-valued?
- Would these inequalities be useful to us?

$\alpha, \beta \in \mathbb{R}$	$A, B \in S^{d\times d}$	Comments
$\alpha > \beta$	$A \succeq B$	$A - B$ is p.s.d.
$\|\alpha\|$	$\|\|A\|\|$	Spectral norm
e^{α}	e^{A}	Matrix Exponential

Matrix-valued Probabilistic Inequalities

Lemma (Markov)

Let $X \geq 0$ *be a real-valued r.v. and* $a > 0$*. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Matrix-valued Probabilistic Inequalities

Lemma (Markov)

Let $X \geq 0$ *be a real-valued r.v. and* $a > 0$*. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Lemma (Matrix-valued Markov [AW02])

Let $M \succeq 0$ *be a self adjoint matrix-valued r.v. and* $a > 0$*. Then*

$$\mathbb{P}(M \npreceq a \cdot I) \leq \frac{\boldsymbol{tr}(\mathbb{E}[M])}{a}.$$

Remark: $\mathbb{P}(M \npreceq a \cdot I) = \mathbb{P}(\lambda_{\max}(M) > a)$

Matrix-valued Probabilistic Inequalities

Theorem (Chernoff)

Let $X_1, X_2, \ldots, X_t$ be i.i.d. copies of a real-valued r.v. X and $\varepsilon > 0$. If $|X| \leq \gamma$, then

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Matrix-valued Probabilistic Inequalities

Theorem (Chernoff)

Let $X_1, X_2, \ldots, X_t$ be i.i.d. copies of a real-valued r.v. X and $\varepsilon > 0$. If $|X| \leq \gamma$, then

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Theorem (Matrix-valued Chernoff [AW02, WX08])

Let $M_1, M_2, \ldots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d. If $\|M\| \leq \gamma$ a.s., then

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq d\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Matrix-valued Probabilistic Inequalities

Theorem (Chernoff)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Theorem (Matrix-valued Chernoff [AW02, WX08])

Let $M_1, M_2, \ldots, M_t$ *be i.i.d. copies of a self adjoint matrix-valued r.v.* M *of size* d*. If* $\|M\| \leq \gamma$ *a.s., then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq d\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Remark: Proof similar with real-valued case (use of matrix exponential!)

Matrix-valued Probabilistic Inequalities

Theorem (Chernoff)

Let $X_1, X_2, \ldots, X_t$ *be i.i.d. copies of a real-valued r.v.* X *and* $\varepsilon > 0$*. If* $|X| \leq \gamma$*, then*

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Theorem (Matrix-valued Chernoff [AW02, WX08])

Let $M_1, M_2, \ldots, M_t$ *be i.i.d. copies of a self adjoint matrix-valued r.v.* M *of size* d*. If* $\|M\| \leq \gamma$ *a.s., then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq d\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

Question: Can we remove the dependency on the dimensionality (d)?

In general, no!

$$\text{Set } M = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & g_{d-1} & 0 \\ 0 & \dots & 0 & g_d \end{bmatrix}, \; g_i \sim \mathcal{N}(0,1).$$

- Then $\mathbb{E}[M] = 0_{d\times d}$
- $\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| \approx \frac{1}{\sqrt{t}}\|(g_1, g_2, \dots, g_d)\|_\infty$, i.e., maximum deviation of d independent Gaussian r.v.'s

In general, no!

$$\text{Set } M = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & g_{d-1} & 0 \\ 0 & \dots & 0 & g_d \end{bmatrix}, \; g_i \sim \mathcal{N}(0,1).$$

- Then $\mathbb{E}[M] = 0_{d\times d}$
- $\left\| \frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M] \right\| \approx \frac{1}{\sqrt{t}} \|(g_1, g_2, \dots, g_d)\|_\infty$, i.e., maximum deviation of d independent Gaussian r.v.'s

Question:

- Are there any natural assumptions that avoid the dependency on d?

In general, no!

$$\text{Set } M = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & g_{d-1} & 0 \\ 0 & \dots & 0 & g_d \end{bmatrix}, \ g_i \sim \mathcal{N}(0,1).$$

- Then $\mathbb{E}[M] = 0_{d \times d}$
- $\left\| \frac{1}{t} \sum_{i=1}^{t} M_i - \mathbb{E}[M] \right\| \approx \frac{1}{\sqrt{t}} \|(g_1, g_2, \dots, g_d)\|_\infty$, i.e., maximum deviation of d independent Gaussian r.v.'s

Question:

- Are there any natural assumptions that avoid the dependency on d?
- What if M has *rank-one* [RV07, Rud99]?

In general, no!

$$\text{Set } M = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & g_{d-1} & 0 \\ 0 & \dots & 0 & g_d \end{bmatrix}, \; g_i \sim \mathcal{N}(0,1).$$

- Then $\mathbb{E}[M] = 0_{d\times d}$
- $\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| \approx \frac{1}{\sqrt{t}} \|(g_1, g_2, \dots, g_d)\|_\infty$, i.e., maximum deviation of d independent Gaussian r.v.'s

Question:

- Are there any natural assumptions that avoid the dependency on d?
- What if M has *rank-one* [RV07, Rud99]? Low-rank [MZ10]?

Let $M_1, M_2, \dots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d

Theorem ("Restated" Matrix-valued Chernoff)

If $\|M\| \leq \gamma$ *a.s., and* $t = \Omega(\gamma^2/\varepsilon^2 \log d)$ *then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq \frac{1}{poly(d)}.$$

Low Rank Matrix-valued Chernoff

Let $M_1, M_2, \ldots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d

Theorem ("Restated" Matrix-valued Chernoff)

If $\|M\| \leq \gamma$ *a.s., and* $t = \Omega(\gamma^2/\varepsilon^2 \log d)$ *then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq \frac{1}{poly(d)}.$$

Theorem (Low Rank Matrix-valued Chernoff [MZ10])

If $\|M\| \leq \gamma$, ***rank*** $(M) = O(1)$ *a.s.,* $\|\mathbb{E}[M]\| \leq 1$ *and* $t = \Omega(\gamma/\varepsilon^2 \log(\gamma/\varepsilon^2))$ *then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq \frac{1}{poly(t)}.$$

Warm-up (Real-valued case)

- Let's start by proving the real-valued case

Let $X_1, X_2, \ldots, X_t$ be i.i.d. copies of a real-valued r.v. X and $\varepsilon > 0$. If $|X| \leq \gamma$, then

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right| > \varepsilon\right) \leq 2\exp\left(-C\frac{\varepsilon^2 t}{\gamma^2}\right).$$

- p-th moments, $E_p := \left(\mathbb{E}\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right|^p\right)^{1/p}$
- Approach: Give tight bounds for E_p
- Mimic the real-valued case for matrix-valued

Fact

If $g \sim \mathcal{N}(0, \sigma^2)$, then $(\mathbb{E}\,|g|^p)^{1/p} = O(\sigma\sqrt{p})$

Proof (Warm-up)

Reduce general r.v. X_i to Bernoulli $\epsilon_i \sim \pm 1$ (Symmetrisation Argument)

$$\begin{aligned} E_p &:= \left(\mathbb{E}_{X_i}\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right|^p\right)^{1/p} \\ &\leq \frac{2}{t}\left(\mathbb{E}_{X_i}\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p} \end{aligned}$$

Bound $\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p$. By Khintchine's ineq. we get

$$\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p \leq \left(C \cdot p \sum_{i=1}^{t} X_i^2\right)^{p/2}$$

Proof (Warm-up)

Reduce general r.v. X_i to Bernoulli $\epsilon_i \sim \pm 1$ (Symmetrisation Argument)

$$E_p := \left(\mathbb{E}_{X_i}\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right|^p\right)^{1/p}$$
$$\leq \frac{2}{t}\left(\mathbb{E}_{X_i}\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}$$

Bound $\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p$. By Khintchine's ineq. we get

$$\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p \leq \left(C \cdot p \sum_{i=1}^{t} X_i^2\right)^{p/2}$$

Proof (Warm-up)

Reduce general r.v. X_i to Bernoulli $\epsilon_i \sim \pm 1$ (Symmetrisation Argument)

$$
\begin{aligned}
E_p &:= \left(\mathbb{E}_{X_i}\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right|^p\right)^{1/p} \\
&\leq \frac{2}{t}\left(\mathbb{E}_{X_i}\,\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}
\end{aligned}
$$

Bound $\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p$. By Khintchine's ineq. we get

$$
\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p \leq \left(C \cdot p \sum_{i=1}^{t} X_i^2\right)^{p/2}
$$

Proof (Warm-up) - Continued

$$
\begin{aligned}
E_p &\leq \frac{2}{t}\left(\mathbb{E}_{X_i}\,\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}, && \text{Symmetrisation} \\
&\leq \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t} X_i^2\right)^{p/2}\right)^{1/p}, && \text{Khintchine} \\
&\leq \frac{2C\sqrt{p}}{t}\left(t\gamma^2\right)^{1/2}, && \sum_{i=1}^{t} X_i^2 \leq t\gamma^2 \\
&\leq \frac{2C\gamma\sqrt{p}}{\sqrt{t}}
\end{aligned}
$$

Proof (Warm-up) - Continued

$$
\begin{aligned}
E_p &\leq \frac{2}{t}\left(\mathbb{E}_{X_i}\,\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}, && \text{Symmetrisation}\\
&\leq \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t}X_i^2\right)^{p/2}\right)^{1/p}, && \text{Khintchine}\\
&\leq \frac{2C\sqrt{p}}{t}\left(t\gamma^2\right)^{1/2}, && \sum_{i=1}^{t}X_i^2 \leq t\gamma^2\\
&\leq \frac{2C\gamma\sqrt{p}}{\sqrt{t}}
\end{aligned}
$$

Proof (Warm-up) - Continued

$$\begin{aligned}
E_p &\le \frac{2}{t}\left(\mathbb{E}_{X_i}\,\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}, && \text{Symmetrisation}\\
&\le \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t}X_i^2\right)^{p/2}\right)^{1/p}, && \text{Khintchine}\\
&\le \frac{2C\sqrt{p}}{t}\left(t\gamma^2\right)^{1/2}, && \sum_{i=1}^{t}X_i^2 \le t\gamma^2\\
&\le \frac{2C\gamma\sqrt{p}}{\sqrt{t}}
\end{aligned}$$

Proof (Warm-up) - Continued

$$
\begin{aligned}
E_p &\le \frac{2}{t}\left(\mathbb{E}_{X_i}\,\mathbb{E}_{\epsilon_i}\left|\sum_{i=1}^{t}\epsilon_i X_i\right|^p\right)^{1/p}, && \text{Symmetrisation}\\
&\le \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t}X_i^2\right)^{p/2}\right)^{1/p}, && \text{Khintchine}\\
&\le \frac{2C\sqrt{p}}{t}\left(t\gamma^2\right)^{1/2}, && \sum_{i=1}^{t}X_i^2 \le t\gamma^2\\
&\le \frac{2C\gamma\sqrt{p}}{\sqrt{t}}
\end{aligned}
$$

Theorem (Low Rank Matrix-valued Chernoff [MZ10])

Let $M_1, M_2, \dots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d. If $\|M\| \leq \gamma$, $\textbf{rank}(M) = O(1)$ a.s., $\|\mathbb{E}[M]\| \leq 1$ and $t = \Omega(\gamma/\varepsilon^2 \log(\gamma/\varepsilon^2))$ then

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq \frac{1}{poly(t)}.$$

Let $Z = \left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|$

Goal: Prove a similar bound for $(\mathbb{E} Z^p)^{1/p}$ like before (real case)

Theorem (Low Rank Matrix-valued Chernoff [MZ10])

Let $M_1, M_2, \dots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d. If $\|M\| \le \gamma$, $\textbf{rank}(M) = O(1)$ a.s., $\|\mathbb{E}[M]\| \le 1$ and $t = \Omega(\gamma/\varepsilon^2 \log(\gamma/\varepsilon^2))$ then

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \le \frac{1}{poly(t)}.$$

Let $Z = \left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|$

Goal: Prove a similar bound for $(\mathbb{E} Z^p)^{1/p}$ like before (real case)

Main Problem

There is no Khintchine ineq. for $\|\cdot\|$ space as for reals

Theorem (Low Rank Matrix-valued Chernoff [MZ10])

*Let $M_1, M_2, \ldots, M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M of size d. If $\|M\| \leq \gamma$, **rank**$(M) = O(1)$ a.s., $\|\mathbb{E}[M]\| \leq 1$ and $t = \Omega(\gamma/\varepsilon^2 \log(\gamma/\varepsilon^2))$ then*

$$\mathbb{P}\left(\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\| > \varepsilon\right) \leq \frac{1}{poly(t)}.$$

Let $Z = \left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|$

Goal: Prove a similar bound for $(\mathbb{E} Z^p)^{1/p}$ like before (real case)

Main Problem

There is no Khintchine ineq. for $\|\cdot\|$ space as for reals

...however there is Khintchine ineq. for the Schatten space...

Schatten Space

- Let $A \in \mathbb{R}^{d\times d}$. Denote by C_p^d the p-th Schatten norm space in $\mathbb{R}^d$ equipped with the norm

$$\|A\|_{C_p^d} := \left(\sum_{i=1}^{d} \sigma_i(A)^p\right)^{1/p},$$

  where $\sigma_i(A)$ are the singular values of A.

- $p = \infty$: Operator norm
  $p = 2$: Frobenius (Hilbert-Schmidt) norm
  $p = 1$: Nuclear norm
- $\|A\| \le \|A\|_{C_p^d} \le (\mathbf{rank}\,(A))^{1/p}\,\|A\|$ for any $p \ge 1$
- C_p^d space has a Khintchine inequality [LPP91, LP86]!

$$\left(\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|_{C_p^d}^{p}\right)^{1/p} \le O(\sqrt{p})\left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|_{C_p^d}$$

  where $\epsilon_i \sim \pm 1$.

Schatten Space

- Let $A \in \mathbb{R}^{d\times d}$. Denote by C_p^d the p-th Schatten norm space in $\mathbb{R}^d$ equipped with the norm

$$\|A\|_{C_p^d} := \left(\sum_{i=1}^{d} \sigma_i(A)^p\right)^{1/p},$$

where $\sigma_i(A)$ are the singular values of A.

- $p = \infty$: Operator norm
  $p = 2$: Frobenius (Hilbert-Schmidt) norm
  $p = 1$: Nuclear norm
- $\|A\| \leq \|A\|_{C_p^d} \leq (\mathbf{rank}(A))^{1/p} \|A\|$ for any $p \geq 1$
- C_p^d space has a Khintchine inequality [LPP91, LP86]!

$$\left(\mathbb{E}_{\epsilon_i} \left\|\sum_{i=1}^{t} \epsilon_i M_i\right\|_{C_p^d}^p\right)^{1/p} \leq O(\sqrt{p}) \left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|_{C_p^d}$$

where $\epsilon_i \sim \pm 1$.

Schatten Space

- Let $A \in \mathbb{R}^{d\times d}$. Denote by C_p^d the p-th Schatten norm space in $\mathbb{R}^d$ equipped with the norm

$$\|A\|_{C_p^d} := \left(\sum_{i=1}^{d} \sigma_i(A)^p\right)^{1/p},$$

where $\sigma_i(A)$ are the singular values of A.

- $p = \infty$: Operator norm
  $p = 2$: Frobenius (Hilbert-Schmidt) norm
  $p = 1$: Nuclear norm
- $\|A\| \le \|A\|_{C_p^d} \le (\mathbf{rank}(A))^{1/p} \|A\|$ for any $p \ge 1$
- C_p^d space has a Khintchine inequality [LPP91, LP86]!

$$\left(\mathbb{E}_{\epsilon_i} \left\|\sum_{i=1}^{t} \epsilon_i M_i\right\|_{C_p^d}^p\right)^{1/p} \le O(\sqrt{p}) \left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|_{C_p^d}$$

where $\epsilon_i \sim \pm 1$.

Schatten Space

- Let $A \in \mathbb{R}^{d\times d}$. Denote by C_p^d the p-th Schatten norm space in $\mathbb{R}^d$ equipped with the norm

$$\|A\|_{C_p^d} := \left(\sum_{i=1}^{d} \sigma_i(A)^p\right)^{1/p},$$

  where $\sigma_i(A)$ are the singular values of A.

- $p = \infty$: Operator norm
  $p = 2$: Frobenius (Hilbert-Schmidt) norm
  $p = 1$: Nuclear norm

- $\|A\| \le \|A\|_{C_p^d} \le (\textbf{rank}(A))^{1/p} \|A\|$ for any $p \ge 1$

- C_p^d space has a Khintchine inequality [LPP91, LP86]!

$$\left(\mathbb{E}_{\epsilon_i} \left\|\sum_{i=1}^{t} \epsilon_i M_i\right\|_{C_p^d}^{p}\right)^{1/p} \le O(\sqrt{p}) \left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|_{C_p^d}$$

  where $\epsilon_i \sim \pm 1$.

What we proved before...

Real-valued:

$$\left(\mathbb{E}_{X_i}\left|\frac{1}{t}\sum_{i=1}^{t}X_i-\mathbb{E}[X]\right|^p\right)^{1/p} \leq \frac{C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t}X_i^2\right)^{p/2}\right)^{1/p}$$

...what we get now

Real-valued:

$$\left(\mathbb{E}_{X_i}\left|\frac{1}{t}\sum_{i=1}^{t} X_i - \mathbb{E}[X]\right|^p\right)^{1/p} \leq \frac{C\sqrt{p}}{t}\left(\mathbb{E}_{X_i}\left(\sum_{i=1}^{t} X_i^2\right)^{p/2}\right)^{1/p}$$

Lemma (Main Lemma [MZ10])

Let $M_1,\ldots,M_t$ be i.i.d. copies of a self adjoint matrix-valued r.v. M with rank at most r almost surely. Then for every $p \geq 2$

$$\left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p} \leq \frac{C(rt)^{1/p}\sqrt{p}}{t}\left(\mathbb{E}_{M_j}\left\|\sum_{j=1}^{t} M_j^2\right\|^{p/2}\right)^{1/p}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \quad \leq \quad \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p}$$ Symmetrisation

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p_{C_p^d}\right)^{1/p} \quad \|A\| \leq \|A\|_{C_p^d}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p_{\mathrm{C}_p^d}\right)^{1/p} \quad \|A\| \leq \|A\|_{\mathrm{C}_p^d}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \leq \frac{2}{t}\left(\mathbb{E}_{M_i}\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i}\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p_{C_p^d}\right)^{1/p} \quad \|A\| \leq \|A\|_{C_p^d}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i} C^p p^{p/2}\left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|^p_{C_p^d}\right)^{1/p} \quad \text{Khintchine}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p_{C_p^d}\right)^{1/p} \quad \|A\| \leq \|A\|_{C_p^d}$$

$$\leq \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{M_i}\left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|^p_{C_p^d}\right)^{1/p} \quad \text{Khintchine}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\leq \frac{2}{t}\left(\mathbb{E}_{M_i}\,\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|_{\mathrm{C}_p^d}^p\right)^{1/p} \quad \|A\| \leq \|A\|_{\mathrm{C}_p^d}$$

$$\leq \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{M_i}\left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|_{\mathrm{C}_p^d}^p\right)^{1/p} \quad \text{Khintchine}$$

Proof Sketch

Let $\widetilde{E}_p := \left(\mathbb{E}\left\|\frac{1}{t}\sum_{i=1}^{t} M_i - \mathbb{E}[M]\right\|^p\right)^{1/p}$

$$\widetilde{E}_p \le \frac{2}{t}\left(\mathbb{E}_{M_i}\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p\right)^{1/p} \quad \text{Symmetrisation}$$

$$\le \frac{2}{t}\left(\mathbb{E}_{M_i}\mathbb{E}_{\epsilon_i}\left\|\sum_{i=1}^{t}\epsilon_i M_i\right\|^p_{C_p^d}\right)^{1/p} \quad \|A\| \le \|A\|_{C_p^d}$$

$$\le \frac{2C\sqrt{p}}{t}\left(\mathbb{E}_{M_i}\left\|\left(\sum_{i=1}^{t} M_i^2\right)^{1/2}\right\|^p_{C_p^d}\right)^{1/p} \quad \text{Khintchine}$$

$$\le \frac{2C(rt)^{1/p}\sqrt{p}}{t}\left(\mathbb{E}_{M_i}\left\|\sum_{i=1}^{t} M_i^2\right\|^{p/2}\right)^{1/p} \quad \|A\|_{C_p^d} \le \textbf{rank}(A)^{1/p}\|A\|$$

SECOND PART

APPLICATIONS

Matrix Sparsification

$$A := \begin{bmatrix}
19 & 3 & 4 & 16 & 7 & 6 & 6 & 7 & 19 & 8 & 13 & 10 & 2 & 4 \\
3 & 9 & 15 & 4 & 12 & 17 & 16 & 4 & 4 & 6 & 8 & 7 & 5 & 14 \\
19 & 19 & 1 & 10 & 5 & 5 & 16 & 16 & 17 & 17 & 11 & 19 & 8 & 4 \\
13 & 16 & 6 & 9 & 16 & 19 & 8 & 7 & 11 & 9 & 9 & 8 & 17 & 8 \\
2 & 20 & 1 & 13 & 6 & 7 & 12 & 11 & 20 & 19 & 2 & 3 & 1 & 13 \\
6 & 14 & 2 & 15 & 11 & 4 & 2 & 4 & 2 & 4 & 5 & 16 & 1 & 16 \\
11 & 1 & 17 & 16 & 14 & 6 & 2 & 13 & 9 & 6 & 3 & 8 & 4 & 2 \\
20 & 17 & 14 & 6 & 18 & 13 & 11 & 6 & 3 & 3 & 4 & 5 & 13 & 19 \\
20 & 19 & 7 & 14 & 20 & 10 & 16 & 14 & 20 & 3 & 5 & 9 & 15 & 16 \\
4 & 14 & 20 & 14 & 11 & 8 & 19 & 14 & 1 & 18 & 9 & 2 & 13 & 10 \\
20 & 16 & 1 & 4 & 3 & 17 & 3 & 15 & 16 & 12 & 1 & 3 & 10 & 9 \\
20 & 15 & 9 & 3 & 3 & 12 & 12 & 10 & 17 & 11 & 19 & 19 & 11 & 9 \\
10 & 8 & 8 & 10 & 6 & 11 & 10 & 2 & 18 & 3 & 19 & 20 & 6 & 7 \\
17 & 14 & 16 & 20 & 17 & 19 & 1 & 5 & 2 & 18 & 10 & 12 & 15 & 11
\end{bmatrix}$$

Matrix Sparsification

$$\widetilde{A} := \begin{bmatrix}
19 & 3 & 0 & 16 & 7 & 6 & 0 & 7 & 0 & 8 & 13 & 10 & 2 & 4 \\
0 & 9 & 15 & 4 & 0 & 17 & 16 & 4 & 4 & 6 & 0 & 7 & 0 & 14 \\
19 & 19 & 1 & 10 & 5 & 5 & 16 & 16 & 17 & 0 & 11 & 0 & 8 & 4 \\
13 & 0 & 6 & 0 & 0 & 19 & 8 & 0 & 11 & 0 & 9 & 8 & 17 & 8 \\
2 & 20 & 1 & 13 & 0 & 7 & 0 & 11 & 20 & 19 & 2 & 3 & 1 & 13 \\
6 & 14 & 2 & 15 & 11 & 4 & 2 & 4 & 2 & 4 & 5 & 16 & 1 & 16 \\
11 & 1 & 0 & 16 & 0 & 6 & 2 & 13 & 0 & 6 & 3 & 8 & 0 & 2 \\
20 & 17 & 14 & 6 & 18 & 13 & 11 & 6 & 3 & 3 & 4 & 5 & 13 & 19 \\
20 & 19 & 7 & 14 & 20 & 10 & 16 & 14 & 20 & 3 & 5 & 0 & 15 & 16 \\
4 & 14 & 20 & 14 & 11 & 8 & 19 & 14 & 1 & 18 & 9 & 2 & 13 & 10 \\
20 & 16 & 1 & 4 & 3 & 0 & 3 & 15 & 0 & 12 & 1 & 3 & 10 & 9 \\
20 & 15 & 9 & 3 & 0 & 12 & 0 & 10 & 0 & 11 & 19 & 19 & 11 & 9 \\
10 & 8 & 8 & 0 & 6 & 11 & 10 & 2 & 18 & 3 & 0 & 20 & 6 & 7 \\
17 & 14 & 16 & 20 & 17 & 19 & 1 & 0 & 2 & 18 & 10 & 12 & 15 & 11
\end{bmatrix}$$

Matrix Sparsification

$$\widetilde{A} := \left[\begin{array}{cccccccccccccc}
19 & 3 & 0 & 16 & 7 & 6 & 0 & 7 & 0 & 8 & 13 & 10 & 2 & 4 \\
0 & 9 & 15 & 4 & 0 & 17 & 16 & 4 & 4 & 6 & 0 & 7 & 0 & 14 \\
19 & 19 & 1 & 10 & 5 & 5 & 16 & 16 & 17 & 0 & 11 & 0 & 8 & 4 \\
13 & 0 & 6 & 0 & 0 & 19 & 8 & 0 & 11 & 0 & 9 & 8 & 17 & 8 \\
2 & 20 & 1 & 13 & 0 & 7 & 0 & 11 & 20 & 19 & 2 & 3 & 1 & 13 \\
6 & 14 & 2 & 15 & 11 & 4 & 2 & 4 & 2 & 4 & 5 & 16 & 1 & 16 \\
11 & 1 & 0 & 16 & 0 & 6 & 2 & 13 & 0 & 6 & 3 & 8 & 0 & 2 \\
20 & 17 & 14 & 6 & 18 & 13 & 11 & 6 & 3 & 3 & 4 & 5 & 13 & 19 \\
20 & 19 & 7 & 14 & 20 & 10 & 16 & 14 & 20 & 3 & 5 & 0 & 15 & 16 \\
4 & 14 & 20 & 14 & 11 & 8 & 19 & 14 & 1 & 18 & 9 & 2 & 13 & 10 \\
20 & 16 & 1 & 4 & 3 & 0 & 3 & 15 & 0 & 12 & 1 & 3 & 10 & 9 \\
20 & 15 & 9 & 3 & 0 & 12 & 0 & 10 & 0 & 11 & 19 & 19 & 11 & 9 \\
10 & 8 & 8 & 0 & 6 & 11 & 10 & 2 & 18 & 3 & 0 & 20 & 6 & 7 \\
17 & 14 & 16 & 20 & 17 & 19 & 1 & 0 & 2 & 18 & 10 & 12 & 15 & 11
\end{array}\right]$$

Goal: Given $A \in \mathbb{R}^{n \times n}$ and $\varepsilon > 0$. Find sparse $\widetilde{A}$ s.t. $\left\|\widetilde{A} - A\right\| \leq \varepsilon \|A\|$

Matrix Sparsification

Problem

Given $A \in \mathbb{R}^{n \times n}$ and $\varepsilon > 0$. Find sparse $\widetilde{A}$ s.t. $\left\|\widetilde{A} - A\right\| \leq \varepsilon \|A\|$

- Achlioptas, McSherry [AM07]
  Sparsify each entry (i,j) independently w.p. $\approx |A_{ij}|$
  Analysis: $\widetilde{A} - A$ is a random matrix with independent entries
  Arora et al. [AHK06] simplified their analysis using real-valued Chernoff bounds.
- Drineas, Z. [DZ10]
  Sample each entry (i,j) independently w.p. $\approx A_{ij}^2 / \|A\|_{\mathrm{F}}^2$
  Improve the above results using matrix-valued Chernoff bounds (matrix-valued Bernstein)

Analysis via matrix-valued Chernoff

- Define a matrix-valued r.v. M with $\mathbb{E}[M] = A$
- Each sample of M is a zero $d \times d$ matrix with only one non-zero entry
  Let $p_{ij} = A_{ij}^2 / \|A\|_{\mathrm{F}}^2$ (probability selecting (i, j) entry)

$$\mathbb{P}\left(M = \frac{1}{p_{ij}} A_{ij} e_i e_j^\top\right) = p_{ij}$$

Analysis via matrix-valued Chernoff

$$A := \begin{bmatrix}
19 & 3 & 4 & 16 & 7 & 6 & 6 & 7 & 19 & 8 & 13 & 10 & 2 & 4 \\
3 & 9 & 15 & 4 & 12 & 17 & 16 & 4 & 4 & 6 & 8 & 7 & 5 & 14 \\
19 & 19 & 1 & 10 & 5 & 5 & 16 & 16 & 17 & 17 & 11 & 19 & 8 & 4 \\
13 & 16 & 6 & 9 & 16 & 19 & 8 & 7 & 11 & 9 & 9 & 8 & 17 & 8 \\
2 & 20 & 1 & 13 & 6 & 7 & 12 & 11 & 20 & 19 & 2 & 3 & 1 & 13 \\
6 & 14 & 2 & 15 & 11 & 4 & 2 & 4 & 2 & 4 & 5 & 16 & 1 & 16 \\
11 & 1 & 17 & 16 & 14 & 6 & 2 & 13 & 9 & 6 & 3 & 8 & 4 & 2 \\
20 & 17 & 14 & 6 & 18 & 13 & 11 & 6 & 3 & 3 & 4 & 5 & 13 & 19 \\
20 & 19 & 7 & 14 & 20 & 10 & 16 & 14 & 20 & 3 & 5 & 9 & 15 & 16 \\
4 & 14 & 20 & 14 & 11 & 8 & 19 & 14 & 1 & 18 & 9 & 2 & 13 & 10 \\
20 & 16 & 1 & 4 & 3 & 17 & 3 & 15 & 16 & 12 & 1 & 3 & 10 & 9 \\
20 & 15 & 9 & 3 & 3 & 12 & 12 & 10 & 17 & 11 & 19 & 19 & 11 & 9 \\
10 & 8 & 8 & 10 & 6 & 11 & 10 & 2 & 18 & 3 & 19 & 20 & 6 & 7 \\
17 & 14 & 16 & 20 & 17 & 19 & 1 & 5 & 2 & 18 & 10 & 12 & 15 & 11
\end{bmatrix}$$

Analysis via matrix-valued Chernoff

$$
M_1 := \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 14 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Analysis via matrix-valued Chernoff

$$M_2 := \begin{bmatrix} 0 & 0 & 0 & 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Analysis via matrix-valued Chernoff

$$M_3 := \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&15&0 \end{bmatrix}$$

Analysis via matrix-valued Chernoff

$$M_3 := \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&15&0 \end{bmatrix}$$

$$\cdots\cdots$$

Analysis via matrix-valued Chernoff

$$M_t := \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 13 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Analysis via matrix-valued Chernoff

$$\widetilde{A} = \begin{bmatrix}
19 & 3 & 0 & 16 & 7 & 6 & 0 & 7 & 0 & 8 & 13 & 10 & 2 & 4 \\
0 & 9 & 15 & 4 & 0 & 17 & 16 & 4 & 4 & 6 & 0 & 7 & 0 & 14 \\
19 & 19 & 1 & 10 & 5 & 5 & 16 & 16 & 17 & 0 & 11 & 0 & 8 & 4 \\
13 & 0 & 6 & 0 & 0 & 19 & 8 & 0 & 11 & 0 & 9 & 8 & 17 & 8 \\
2 & 20 & 1 & 13 & 0 & 7 & 0 & 11 & 20 & 19 & 2 & 3 & 1 & 13 \\
6 & 14 & 2 & 15 & 11 & 4 & 2 & 4 & 2 & 4 & 5 & 16 & 1 & 16 \\
11 & 1 & 0 & 16 & 0 & 6 & 2 & 13 & 0 & 6 & 3 & 8 & 0 & 2 \\
20 & 17 & 14 & 6 & 18 & 13 & 11 & 6 & 3 & 3 & 4 & 5 & 13 & 19 \\
20 & 19 & 7 & 14 & 20 & 10 & 16 & 14 & 20 & 3 & 5 & 0 & 15 & 16 \\
4 & 14 & 20 & 14 & 11 & 8 & 19 & 14 & 1 & 18 & 9 & 2 & 13 & 10 \\
20 & 16 & 1 & 4 & 3 & 0 & 3 & 15 & 0 & 12 & 1 & 3 & 10 & 9 \\
20 & 15 & 9 & 3 & 0 & 12 & 0 & 10 & 0 & 11 & 19 & 19 & 11 & 9 \\
10 & 8 & 8 & 0 & 6 & 11 & 10 & 2 & 18 & 3 & 0 & 20 & 6 & 7 \\
17 & 14 & 16 & 20 & 17 & 19 & 1 & 0 & 2 & 18 & 10 & 12 & 15 & 11
\end{bmatrix}$$

Set $\widetilde{A} := \frac{1}{t}\sum_{i=1}^{t} M_i$

Analysis via matrix-valued Chernoff

- Define a matrix-valued r.v. M with $\mathbb{E}[M] = A$
- Each sample of M is a zero $d \times d$ matrix with only one non-zero entry.
  Let $p_{ij} = A_{ij}^2 / \|A\|_F^2$ (probability selecting (i,j) entry)

$$\mathbb{P}\left(M = \frac{1}{p_{ij}} A_{ij} e_i e_j^\top\right) = p_{ij}$$

- Bounding the number of samples = # of non-zero entries of $\widetilde{A}$
- Matrix-valued Chernoff bounds guarantees $\left\|\widetilde{A} - A\right\| \le \varepsilon \|A\|$

Approximate Matrix Multiplication

Problem

Given $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{n\times p}$ and $\varepsilon > 0$. Approximate matrix product $A^\top B$; compute $\widetilde{A} \in \mathbb{R}^{t\times m}$ and $\widetilde{B} \in \mathbb{R}^{t\times p}$ ($t \ll m, p, n$) such that

$$\left\| \widetilde{A}^\top \widetilde{B} - A^\top B \right\| \leq \varepsilon \|A\| \|B\|.$$

Approaches:

- Randomly project their columns
- Non-uniform row sampling

Related Work:

- Many results w.r.t. Frobenius norm [DKM06, Sar06, CW09]
- "Weak" bounds w.r.t. *spectral* norm [DK01, DKM06, Sar06]
- Similar strong bounds for the special case $A = B$ in [RV07]

Approximate Matrix Multiplication

Problem

Given $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{n\times p}$ and $\varepsilon > 0$. Approximate matrix product $A^\top B$; compute $\widetilde{A} \in \mathbb{R}^{t\times m}$ and $\widetilde{B} \in \mathbb{R}^{t\times p}$ ($t \ll m, p, n$) such that

$$\left\| \widetilde{A}^\top \widetilde{B} - A^\top B \right\| \leq \varepsilon \|A\| \, \|B\| \, .$$

Approaches:

- Randomly project their columns
- Non-uniform row sampling

Related Work:

- Many results w.r.t. Frobenius norm [DKM06, Sar06, CW09]
- "Weak" bounds w.r.t. *spectral* norm [DK01, DKM06, Sar06]
- Similar strong bounds for the special case $A = B$ in [RV07]

Approximate Matrix Multiplication

Problem

Given $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{n\times p}$ and $\varepsilon > 0$. Approximate matrix product $A^\top B$; compute $\widetilde{A} \in \mathbb{R}^{t\times m}$ and $\widetilde{B} \in \mathbb{R}^{t\times p}$ ($t \ll m,p,n$) such that

$$\left\|\widetilde{A}^\top \widetilde{B} - A^\top B\right\| \leq \varepsilon \|A\| \|B\|.$$

Approaches:

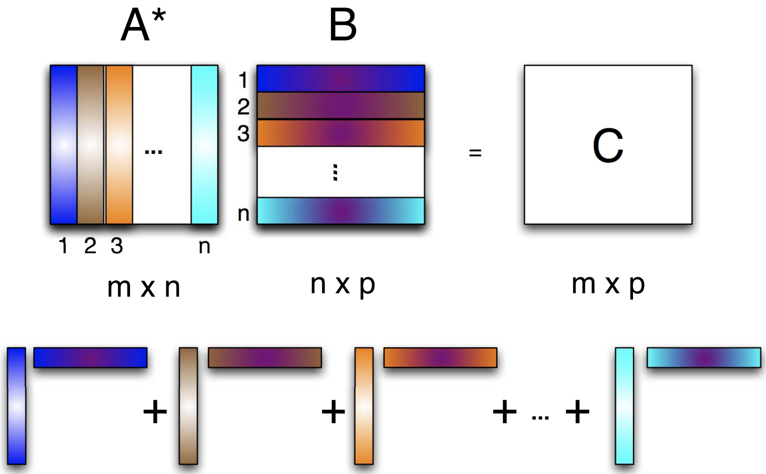
- Randomly project their columns
- Non-uniform row sampling

Related Work:

- Many results w.r.t. Frobenius norm [DKM06, Sar06, CW09]
- "Weak" bounds w.r.t. *spectral* norm [DK01, DKM06, Sar06]
- Similar strong bounds for the special case $A = B$ in [RV07]

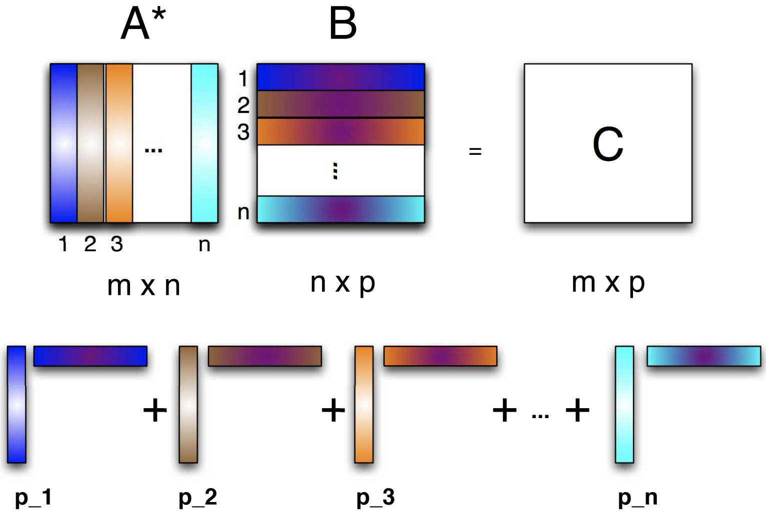
Non-uniform Row Sampling

- Recall that $A^{\top}B = \sum_{i=1}^{n} A_i^{\top} B_i (= \sum_{i=1}^{n} A_i \otimes B_i)$



Non-uniform Row Sampling

- Recall that $A^\top B = \sum_{i=1}^{n} A_i^\top B_i (= \sum_{i=1}^{n} A_i \otimes B_i)$



Non-uniform Row Sampling

Theorem

There exists prob. distribution p_i s.t. if we form an $t \times m$ matrix $\widetilde{A}$ and an $t \times p$ matrix $\widetilde{B}$ by taking t i.i.d. (row indices) samples from p_i with $t = \Omega(\widetilde{r}/\varepsilon^2 \log(\widetilde{r}/\varepsilon^2))$, then

$$\mathbb{P}\left(\left\|\widetilde{A}^\top \widetilde{B} - A^\top B\right\| \leq \varepsilon \|A\| \|B\|\right) \geq 1 - o_{\widetilde{r}}(1),$$

*where $\widetilde{r}$ is **st.rank** $(A) +$ **st.rank** (B).*

- **st.rank** $(A) := \|A\|_{\mathrm{F}}^2 / \|A\| \leq$ **rank** (A)

Proof Sketch

Define a distribution over $\mathbb{R}^{(m+p)\times(m+p)}$ by

$$\mathbb{P}\left(X = \frac{1}{p_i}\begin{bmatrix} 0 & B_i^\top A_i \\ A_i^\top B_i & 0 \end{bmatrix}\right) = p_i.$$

- $\mathbb{E}[X] = \begin{bmatrix} 0 & B^\top A \\ A^\top B & 0 \end{bmatrix}$
- Every (matrix) sample has rank at most *two*.
- $\|X\| \leq \tilde{r}_A + \tilde{r}_B (\leq \tilde{r})$ a.s..

Applying Theorem with $t = \Omega(\tilde{r}/\varepsilon^2 \log(\tilde{r}/\varepsilon^2))$, we get $i_1, i_2, \ldots, i_t$ indices from $[n]$ such that with high probability

$$\left\| \frac{1}{t}\sum_{j=1}^{t} \begin{bmatrix} 0 & \frac{1}{p_{i_j}} B_{i_j}^\top A_{i_j} \\ \frac{1}{p_{i_j}} A_{i_j}^\top B_{i_j} & 0 \end{bmatrix} - \begin{bmatrix} 0 & B^\top A \\ A^\top B & 0 \end{bmatrix} \right\| \leq \varepsilon \|A\| \, \|B\|$$

Conclusion and Open Problems

- Matrix-valued probabilistic inequalities are powerful tools
- Present two application: matrix sparsification and approximate matrix multiplication
- More applications: graph sparsifiers [SS08], matrix completion [Rec09], bounding integrality gaps [Nem07], Cayley graph expansion, etc...

- Many unexplored connections.
- Matrix martingales - Adaptive sampling? See [Tro10]

Thank You

References I

S. Arora, E. Hazan, and S. Kale.
A Fast Random Sampling Algorithm for Sparsifying Matrices.
In *Proceedings of the International Workshop on Randomization and Approximation Techniques (RANDOM)*, pages 272–279, 2006.

D. Achlioptas and F. Mcsherry.
Fast Computation of Low-rank Matrix Approximations.
SIAM J. Comput., 54(2):9, 2007.

R. Ahlswede and A. Winter.
Strong Converse for Identification via Quantum Channels.
IEEE Transactions on Information Theory, 48(3):569–579, 2002.

K. L. Clarkson and D. P. Woodruff.
Numerical Linear Algebra in the Streaming Model.
In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 205–214, 2009.

P. Drineas and R. Kannan.
Fast Monte-Carlo Algorithms for Approximate Matrix Multiplication.
In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 452–459, 2001.

P. Drineas, R. Kannan, and M. W. Mahoney.
Fast Monte Carlo Algorithms for Matrices I: Approximating Matrix Multiplication.
SIAM J. Comput., 36(1):132–157, 2006.

P. Drineas and A. Zouzias.
A Note on Element-wise Matrix Sparsification via Matrix-valued Chernoff Bounds.
Available at arxiv:1006.0407, June 2010.

F. Lust-Piquard.
Inégalités de Khintchine dans C_p $(1 < p < \infty)$.
C. R. Acad. Sci. Paris Sér. I Math., 303(7):289–292, 1986.

References II

F. Lust-Piquard and G. Pisier.
Non Commutative Khintchine and Paley Inequalities.
Arkiv för Matematik, 29(1-2):241–260, December 1991.

A. Magen and A. Zouzias.
Low Rank Matrix-valued Chernoff Bounds and Approximate Matrix Multiplication, 2010.

A. Nemirovski.
Sums of Random Symmetric Matrices and Quadratic Optimization under Orthogonality Constraints.
Mathematical Programming, 109(2):283–317, 2007.

B. Recht.
A Simpler Approach to Matrix Completion.
Available at arxiv:0910.0651, October 2009.

M. Rudelson.
Random Vectors in the Isotropic Position.
J. Funct. Anal., 164(1):60–72, 1999.

M. Rudelson and R. Vershynin.
Sampling from Large Matrices: An Approach through Geometric Functional Analysis.
SIAM J. Comput., 54(4):21, 2007.

T. Sarlos.
Improved Approximation Algorithms for Large Matrices via Random Projections.
In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 143–152, 2006.

D. A. Spielman and N. Srivastava.
Graph Sparsification by Effective Resistances.
In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 563–568, 2008.

References III

J. A. Tropp.
User-Friendly Tail Bounds for Sums of Random Matrices.
Available at arxiv:1004.4389, April 2010.

A. Wigderson and D. Xiao.
Derandomizing the Ahlswede-Winter Matrix-valued Chernoff Bound using Pessimistic Estimators, and Applications.
Theory of Computing, 4(1):53–76, 2008.