Matrix Concentration

Nick Harvey
University of British Columbia

The Problem

Given any random n×n, symmetric matrices $Y_1,\ldots,Y_k$.
Show that $\sum_i Y_i$ is probably "close" to $E[\sum_i Y_i]$.

Why?

- A matrix generalization of the Chernoff bound.
- Much research on eigenvalues of a random matrix with independent entries. This is more general.

Chernoff/Hoeffding Bound

- **Theorem:**
  Let $Y_1,\ldots,Y_k$ be independent random scalars in $[0,R]$.
  Let $Y = \sum_i Y_i$. Suppose that $\mu_L \leq E[Y] \leq \mu_U$. Then

$$\Pr\left[\sum_i Y_i \leq (1-\epsilon)\mu_L\right] \leq \exp(-\epsilon^2 \mu_L / 2R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \geq (1+\epsilon)\mu_U\right] \leq \exp(-\epsilon^2 \mu_U / 3R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \geq (1+\epsilon)\mu_U\right] \leq \exp\big(-\epsilon \log(\epsilon) \mu_U / 2R\big) \quad \forall \epsilon \geq 1.$$

Rudelson's Sampling Lemma

- **Theorem:** [Rudelson '99]
  Let $Y_1,\ldots,Y_k$ be i.i.d. rank-1, PSD matrices of size n×n s.t. $E[Y_i]=I$, $\|Y_i\| \leq R$. Let $Y = \sum_i Y_i$, so $E[Y]=k \cdot I$. Then

$$\mathrm{E}\left\| \frac{1}{k} \sum_i Y_i - I \right\| \leq O(\sqrt{R \log(n)/k}).$$

- **Example:** Balls and bins
  - Throw k balls uniformly into n bins
  - Y_i = Uniform over $\begin{pmatrix} n & 0 & 0\cdots \\ 0 & 0 & 0\cdots \\ 0 & 0 & 0\cdots \end{pmatrix} \begin{pmatrix} 0 & 0 & 0\cdots \\ 0 & n & 0\cdots \\ 0 & 0 & 0\cdots \end{pmatrix} \begin{pmatrix} 0 & 0 & 0\cdots \\ 0 & 0 & 0\cdots \\ 0 & 0 & n\cdots \end{pmatrix}$
  - If $k = O(n \log n / \epsilon^2)$, all bins same up to factor $1 \pm \epsilon$

Rudelson's Sampling Lemma

- **Theorem:** [Rudelson '99]
  Let $Y_1,\ldots,Y_k$ be i.i.d. rank-1, PSD matrices of size n×n s.t. $E[Y_i]=I$, $\|Y_i\| \leq R$. Let $Y = \sum_i Y_i$, so $E[Y]=k\cdot I$. Then

$$\mathrm{E}\left\|\frac{1}{k}\sum_i Y_i - I\right\| \leq O(\sqrt{R\log(n)/k}).$$

- **Pros:** We've generalized to PSD matrices
- **Mild issue:** We assume $E[Y_i] = I$.
- **Cons:**
  - Y_i's must be identically distributed
  - rank-1 matrices only

Rudelson's Sampling Lemma

- **Theorem:** [Rudelson-Vershynin '07]
  Let $Y_1,\ldots,Y_k$ be i.i.d. rank-1, PSD matrices s.t.
  $E[Y_i]=I$, $\|Y_i\| \leq R$. Let $Y = \sum_i Y_i$, so $E[Y]=k\cdot I$. Then

$$\Pr\left[\lambda_{\min}\left(\sum_i Y_i\right) \leq (1-\epsilon)k\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall\epsilon\in[0,1].$$

$$\Pr\left[\lambda_{\max}\left(\sum_i Y_i\right) \geq (1+\epsilon)k\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall\epsilon\in[0,1].$$

- **Pros:** We've generalized to PSD matrices
- **Mild issue:** We assume $E[Y_i] = I$.
- **Cons:**
  - Y_i's must be identically distributed
  - rank-1 matrices only

Rudelson's Sampling Lemma

- **Theorem:** [Rudelson-Vershynin '07]
  Let $Y_1,\ldots,Y_k$ be i.i.d. rank-1, PSD matrices s.t.
  $E[Y_i]=I$. Let $Y=\sum_i Y_i$, so $E[Y]=k\cdot I$. Assume $Y_i \preceq R\cdot I$. Then

$$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)kI\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall \epsilon\in[0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)kI\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall \epsilon\in[0,1].$$

- **Notation:**
  - $A \preceq B \iff$ B-A is PSD
  - $\alpha I \preceq A \preceq \beta I \iff$ all eigenvalue of A lie in $[\alpha,\beta]$
- **Mild issue:** We assume $E[Y_i] = I$.

Rudelson's Sampling Lemma

- **Theorem:** [Rudelson-Vershynin '07]
  Let $Y_1,\dots,Y_k$ be i.i.d. rank-1, PSD matrices.
  Let $Z=E[Y_i]$, $Y=\sum_i Y_i$, so $E[Y]=k\cdot Z$. Assume $Y_i \preceq R\cdot Z$. Then

$$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)kZ\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall \epsilon\in[0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)kZ\right] \leq \exp\left(-\Omega(\epsilon^2 k/R\log k)\right) \quad \forall \epsilon\in[0,1].$$

- Apply previous theorem to $\{\, Z^{-1/2}\, Y_i\, Z^{-1/2} : i=1,\dots,k \,\}$.
- Use the fact that $A \preceq B$

$$\Leftrightarrow \quad Z^{-1/2} A Z^{-1/2} \preceq Z^{-1/2} B Z^{-1/2}$$

- So $(1-\epsilon)\, k\, Z \preceq \sum_i Y_i \preceq (1+\epsilon)\, k\, Z$

$$\Leftrightarrow \quad (1-\epsilon)\, k\, I \preceq \sum_i Z^{-1/2} Y_i Z^{-1/2} \preceq (1+\epsilon)\, k\, I$$

Ahlswede-Winter Inequality

- **Theorem:** [Ahlswede-Winter '02]
  Let $Y_1,\ldots,Y_k$ be i.i.d. PSD matrices of size n×n.
  Let $Z=E[Y_i]$, $Y=\sum_i Y_i$, so $E[Y]=k\cdot Z$. Assume $Y_i \preceq R\cdot Z$. Then

$$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)kZ\right] \leq n\exp(-\epsilon^2 k/4R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)kZ\right] \leq n\exp(-\epsilon^2 k/4R) \quad \forall \epsilon \in [0,1].$$

- **Pros:**
  – We've removed the rank-1 assumption.
  – Proof is much easier than Rudelson's proof.
- **Cons:**
  – Still need Y_i's to be identically distributed.
  (More precisely, poor results unless $E[Y_a] = E[Y_b]$.)

Tropp's User-Friendly Tail Bound

- **Theorem:** [Tropp '12]
  Let $Y_1,\ldots,Y_k$ be independent, PSD matrices of size nxn.
  s.t. $\|Y_i\| \leq R$. Let $Y=\sum_i Y_i$. Suppose $\mu_L \cdot I \preceq E[Y] \preceq \mu_U \cdot I$. Then

$$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)\mu_L I\right] \leq n\exp(-\epsilon^2\mu_L/2R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)\mu_U I\right] \leq n\exp(-\epsilon^2\mu_U/3R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)\mu_U I\right] \leq n\exp\left(-\epsilon\log(\epsilon)\mu_U/2R\right) \quad \forall \epsilon \geq 1.$$

- **Pros:**
  - Y_i's do not need to be identically distributed
  - Poisson-like bound for the right-tail
  - Proof not difficult (but uses Lieb's inequality)
- **Mild issue:** Poor results unless $\lambda_{min}(E[Y]) \approx \lambda_{max}(E[Y])$.

Tropp's User-Friendly Tail Bound

- **Theorem:** [Tropp '12]
  Let $Y_1,\ldots,Y_k$ be independent, PSD matrices of size n$\times$n.
  Let $Y=\sum_i Y_i$. Let $Z=E[Y]$. Suppose $Y_i \preceq R \cdot Z$. Then

$$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)Z\right] \leq n \exp(-\epsilon^2/2R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)Z\right] \leq n \exp(-\epsilon^2/3R) \quad \forall \epsilon \in [0,1].$$

$$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)Z\right] \leq n \exp\big(-\epsilon \log(\epsilon)/2R\big) \quad \forall \epsilon \geq 1.$$

Tropp's User-Friendly Tail Bound

- **Theorem:** [Tropp '12]
  Let $Y_1,\ldots,Y_k$ be independent, PSD matrices of size nxn.
  s.t. $\|Y_i\| \leq R$. Let $Y=\sum_i Y_i$. Suppose $\mu_L \cdot I \preceq E[Y] \preceq \mu_U \cdot I$. Then
  $$\Pr\left[\sum_i Y_i \preceq (1-\epsilon)\mu_L I\right] \leq n \exp(-\epsilon^2 \mu_L / 2R) \quad \forall \epsilon \in [0,1].$$
  $$\Pr\left[\sum_i Y_i \succeq (1+\epsilon)\mu_U I\right] \leq n \exp(-\epsilon^2 \mu_U / 3R) \quad \forall \epsilon \in [0,1].$$

- **Example:** Balls and bins
  - For b=1,…,n
    - For t=1,…,8 log(n)/ϵ^2
      - With prob ½, throw a ball into bin b
      - Let $Y_{b,t} = \begin{pmatrix} 0 & 0 & 0\cdots \\ 0 & 1 & 0\cdots \\ 0 & 0 & 0\cdots \end{pmatrix}$ with prob ½, otherwise 0.

Additive Error

- Previous theorems give **multiplicative error**:

  $$(1-\epsilon)\ \mathrm{E}[\textstyle\sum_i Y_i] \preceq \sum_i Y_i \preceq (1+\epsilon)\ \mathrm{E}[\sum_i Y_i]$$

- Additive error also useful: $\|\sum_i Y_i - \mathrm{E}[\sum_i Y_i]\| \leq \epsilon$

- **Theorem:** [Rudelson & Vershynin '07]
  Let $Y_1,\ldots,Y_k$ be i.i.d. rank-1, PSD matrices.
  Let $Z=\mathrm{E}[Y_i]$. Suppose $\|Z\| \leq 1$, $\|Y_i\| \leq R$. Then

  $$\Pr\left[\| \tfrac{1}{k} \textstyle\sum_i Y_i - Z \| > \epsilon \right] \leq 2 \exp\big(-\Omega(\epsilon^2 k / R \log k) \big) \quad \forall \epsilon \in [0,1]$$

- **Theorem:** [Magen & Zouzias '11]
  If instead rank $Y_i \leq k := \Theta(R \log(R/\epsilon^2)/\epsilon^2)$, then

  $$\Pr\left[\| \tfrac{1}{k} \textstyle\sum_i Y_i - \mathrm{E}[Y_1] \| > \epsilon \right] \leq 1/\mathrm{poly}(k) \quad \forall \epsilon \in [0,1]$$

Proof of Ahlswede-Winter

- **Key idea:** Bound matrix moment generating function
- Let $S_k = \sum_{i=1}^{k} Y_i$

$$\mathrm{E}[\operatorname{tr} e^{S_k}] \quad = \quad \mathrm{E}[\operatorname{tr} e^{Y_k + S_{k-1}}]$$

tr $e^{A+B} \leq$ tr $e^A \cdot e^B$

Golden-Thompson Inequality

Weakness:
This is brutal

By induction, $\mathrm{E}[\operatorname{tr} e^{\lambda S_k}] \;\leq\; \prod_{i=1}^{k} \|\mathrm{E}[e^{\lambda Y_i}]\| \cdot \operatorname{tr} e^{\lambda 0} \;=\; d \cdot \prod_{i=1}^{k} \|\mathrm{E}[e^{\lambda Y_i}]\|$

How to improve Ahlswede-Winter?

- **Golden-Thompson Inequality**
  $\text{tr } e^{A+B} \leq \text{tr } e^{A} \cdot e^{B}$ for all symmetric matrices A, B.

- Does **not** extend to three matrices!
  $\text{tr } e^{A+B+C} \leq \text{tr } e^{A} \cdot e^{B} \cdot e^{C}$ **is FALSE**.

- **Lieb's Inequality**: For any symmetric matrix L,
  the map f : PSD Cone $\rightarrow \mathbb{R}$ defined by
  $$f(A) = \text{tr} \exp(L + \log(A))$$
  is **concave**.

Beyond the basics

- Hoeffding (non-uniform bounds on Y_i's) [Tropp '12]
- Bernstein (use bound on $Var[Y_i]$) [Tropp '12]
- Freedman (martingale version of Bernstein) [Tropp '12]
- Stein's Method (slightly sharper results) [Mackey et al. '12]

- Pessimistic Estimators for Ahlswede-Winter inequality [Wigderson-Xiao '08]

Summary

- We now have beautiful, powerful, flexible extension of Chernoff bound to matrices.
- Ahlswede-Winter has a simple proof; Tropp's inequality is very easy to use.
- Several important uses to date; hopefully more uses in the future.