UBC CPSC 536N: Sparse Approximations	Winter 2013

Lecture 22 — March 27, 2013

Prof. Nick Harvey — *Scribe: Zachary Drudi*

In this lecture, we prove Tropp's inequality, asuming Lieb's inequality. We first state Lieb's inequality and record some easy consequences. Then we review the proof of the Chernoff bound before proving Tropp's inequality, which can be thought of as a matrix generalization of the Chernoff bound.

1 Preliminaries

Definition 1. If A, B are positive definite, define $A \odot B = \exp(\log(A) + \log(B))$.

This binary operation yields an abelian group on the set of positive definite matrices. In particular, $\odot$ is commutative. Also, if A and B commute then $A \odot B$ is the usual product AB.

Theorem 2. (Lieb) Fix any symmetric H. The map $A \mapsto \operatorname{trace} \exp\big(\log(A) + H\big)$ is concave on positive definite matrices.

This result is difficult, and we will not be doing the proof.

Corollary 3. $\operatorname{trace}(A \odot B)$ is concave in A.

Proof. $\operatorname{trace}(A \odot B) = \operatorname{trace} \exp(\log A + \log B)$. Apply Lieb's theorem with $H = \log B$. ∎

Corollary 4. Let B be fixed, and A a random matrix. Then $\mathbb{E}[\operatorname{trace}(A \odot B)] \leq \operatorname{trace}(\mathbb{E}[A] \odot B)$.

Proof. Apply Jensen's inequality. ∎

Corollary 5. Let $A_1, ..., A_k$ be independent random positive definite matrices. Then

$$\mathbb{E}[\operatorname{trace}(A_1 \odot ... \odot A_k)] \leq \operatorname{trace}(\mathbb{E}[A_1] \odot ... \odot \mathbb{E}[A_k])$$

Proof. Induction, applied to the preceding result. ∎

2 The Chernoff Bound

To highlight the similarities between Tropp's inequality and the Chernoff bound, we first present a complete proof of the Chernoff bound.

Theorem 6. Let $X_1, ..., X_k$ be independent random variables with $0 \leq X_i \leq R$. Let $\mu_{\min} \leq \sum_i \mathrm{E}[\, X_i \,] \leq \mu_{\max}$. Then, for all $\delta \in [0,1]$,

$$\begin{array}{rcccl} \Pr\left[\, \sum_{i=1}^k X_i \geq (1+\delta)\mu_{\max} \,\right] & \overset{(a)}{\leq} & \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}/R} & \overset{(b)}{\leq} & e^{-\delta^2 \mu_{\max}/3R} \\ \Pr\left[\, \sum_{i=1}^k X_i \leq (1-\delta)\mu_{\min} \,\right] & \overset{(c)}{\leq} & \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\max}/R} & \overset{(d)}{\leq} & e^{-\delta^2 \mu_{\min}/2R}. \end{array}$$

Inequality (a) is actually valid for all $\delta \geq 0$.

We now prove inequality (a). Inequalities (b) and (d) are proven in the appendix.

Claim 7.

$$\Pr\left[\sum_{i=1}^k X_i \geq t\right] \;\leq\; \inf_{\theta>0} e^{-\theta t} \cdot \prod_{i=1}^k \mathrm{E}\left[\, e^{\theta X_i} \,\right].$$

Proof. Fix $\theta > 0$.

$$\begin{array}{rcll} \Pr[\, \sum_i X_i \geq t \,] & = & \Pr[\, \sum_i \theta X_i \geq \theta t \,] & \\ & = & \Pr[\, \exp(\sum_i \theta X_i) \geq \exp(\theta t) \,] & \text{(monotonicity of } e^x\text{)} \\ & \leq & e^{-\theta t} \cdot \mathrm{E}[\, \exp(\sum_i \theta X_i) \,] & \text{(Markov's inequality)} \end{array}$$

This expectation can be simplified:

$$\begin{array}{rcll} \mathrm{E}[\, \exp(\sum_i \theta X_i) \,] & = & \mathrm{E}\left[\, \prod_i e^{\theta X_i} \,\right] & \\ & = & \prod_i \mathrm{E}\left[\, e^{\theta X_i} \,\right] & \text{(independence)} \end{array}$$

Combining these proves the claim. ■

Claim 8. Let X be a random variable with $0 \leq X \leq 1$. Then

$$\mathrm{E}\left[\, e^{\theta X} \,\right] \;\leq\; 1 + (e^{\theta} - 1) \cdot \mathrm{E}[\, X \,].$$

Proof. For $x \in [0,1]$ we have $e^{\theta x} \leq 1 + (e^{\theta} - 1) \cdot x$, by convexity of the left-hand side. Since $X \in [0,1]$,

$$\begin{array}{rrcl} & e^{\theta X} & \leq & 1 + (e^{\theta} - 1) \cdot X \\ \Longrightarrow & \mathrm{E}\left[\, e^{\theta X} \,\right] & \leq & 1 + (e^{\theta} - 1) \cdot \mathrm{E}[\, X \,], \end{array}$$

since inequalities are preserved under taking expectation. ■

Proof (of Chernoff Upper Bound). Without loss of generality $R = 1$.

$$\begin{array}{rcll} \prod_{i=1}^k \mathrm{E}\left[\, e^{\theta X_i} \,\right] & \leq & \prod_{i=1}^k \big(1 + (e^{\theta} - 1) \cdot \mathrm{E}[\, X_i \,]\big) & \text{(by Claim 11)} \\ & = & \exp\left(\sum_{i=1}^k \log\big(1 + (e^{\theta} - 1) \cdot \mathrm{E}[\, X_i \,]\big)\right) & \\ & \leq & \exp\left(\sum_{i=1}^k (e^{\theta} - 1) \cdot \mathrm{E}[\, X_i \,]\right) & \text{(using } \log(1+x) \leq x\text{)} \\ & \leq & \exp\big((e^{\theta} - 1)\mu_{\max}\big) & \end{array}$$

Applying Claim 10 with $t = (1+\delta)\mu_{\max}$ and $\theta = \ln(1+\delta)$

$$
\begin{aligned}
\Pr\left[\sum_i X_i \geq (1+\delta)\mu_{\max}\right] &\leq \exp\big(-\ln(1+\delta)\cdot(1+\delta)\mu_{\max}\big)\cdot\exp(\delta\cdot\mu_{\max}) \\
&= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}}
\end{aligned}
$$

■

3 Tropp's Matrix Chernoff Bound

Theorem 9. Let $X_1, ..., X_k$ be independent random $d \times d$ symmetric matrices with $0 \preceq X_i \preceq R \cdot I$. Let $\mu_{\min} \cdot I \preceq \sum_i \mathrm{E}[X_i] \preceq \mu_{\max} \cdot I$. Then, for all $\delta \in [0,1]$,

$$\begin{aligned} \Pr\left[\lambda_{\max}(\textstyle\sum_{i=1}^k X_i) \geq (1+\delta)\mu_{\max}\right] &\overset{(a)}{\leq} d \cdot \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}/R} \overset{(b)}{\leq} d \cdot e^{-\delta^2 \mu_{\max}/3R} \\ \Pr\left[\lambda_{\min}(\textstyle\sum_{i=1}^k X_i) \leq (1-\delta)\mu_{\min}\right] &\overset{(c)}{\leq} d \cdot \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/R} \overset{(d)}{\leq} d \cdot e^{-\delta^2 \mu_{\min}/2R}. \end{aligned}$$

Inequality (a) is actually valid for all $\delta \geq 0$.

We now prove inequality (a). Inequalities (b) and (d) follow from the discussion in the appendix.

Claim 10.

$$\Pr\left[\lambda_{\max}\Big(\sum_{i=1}^k X_i\Big) \geq t\right] \leq \inf_{\theta > 0} e^{-\theta t} \cdot \mathrm{tr}\left(\bigodot_{i=1}^k \mathrm{E}\left[e^{\theta X_i}\right]\right).$$

Proof. Fix $\theta > 0$.

$$\begin{aligned} \Pr[\lambda_{\max}(\textstyle\sum_i X_i) \geq t] &= \Pr[\lambda_{\max}(\textstyle\sum_i \theta X_i) \geq \theta t] && \text{(homogeneity of max eigenvalue)} \\ &= \Pr\left[\exp\left(\lambda_{\max}(\textstyle\sum_i \theta X_i)\right) \geq \exp(\theta t)\right] && \text{(monotonocity of } e^x) \\ &\leq e^{-\theta t} \cdot \mathrm{E}\left[\exp\left(\lambda_{\max}(\textstyle\sum_i \theta X_i)\right)\right] && \text{(Markov's inequality)} \end{aligned}$$

We can bound the maximum eigenvalue by a trace:

$$\begin{aligned} \exp\left(\lambda_{\max}(\textstyle\sum_i \theta X_i)\right) &= \lambda_{\max}\left(\exp(\textstyle\sum_i \theta X_i)\right) && \text{(definition of matrix exponentiation)} \\ &\leq \mathrm{tr}\left(\exp(\textstyle\sum_i \theta X_i)\right) && \text{(max eigenvalue} \leq \text{sum of eigenvalues)} \end{aligned}$$

Taking the expectation gives the bound:

$$\Pr[\lambda_{\max}(\textstyle\sum_i X_i) \geq t] \leq e^{-\theta t} \cdot \mathrm{E}\left[\mathrm{tr}\left(\exp(\textstyle\sum_i \theta X_i)\right)\right].$$

This expectation can be bounded:

$$\begin{aligned} \mathrm{E}\left[\mathrm{tr}\left(\exp(\textstyle\sum_i \theta X_i)\right)\right] &= \mathrm{E}\left[\mathrm{tr}\left(\exp(\textstyle\sum_i \log A_i)\right)\right] && (\text{let } A_i = e^{\theta X_i}) \\ &= \mathrm{E}[\mathrm{tr}(A_1 \odot \cdots \odot A_k)] && (\text{definition of } \odot) \\ &\leq \mathrm{tr}\left(\mathrm{E}[A_1] \odot \cdots \odot \mathrm{E}[A_k]\right) && \text{(by Corollary 5)} \end{aligned}$$

Combining these inequalities proves the claim. ■

Claim 11. Let X be a random symmetric $d \times d$ matrix with $0 \preceq X \preceq I$. Then

$$\mathrm{E}\left[e^{\theta X}\right] \preceq I + (e^{\theta} - 1) \cdot \mathrm{E}[X].$$

Proof. For $x \in [0,1]$ we have $e^{\theta x} \leq 1 + (e^{\theta} - 1) \cdot x$, by convexity of the left-hand side. Since X has all eigenvalues in $[0,1]$, Claim 2 from Lecture 21 gives

$$\begin{aligned} & e^{\theta X} \preceq I + (e^{\theta} - 1) \cdot X \\ \Longrightarrow \quad & \mathrm{E}\left[e^{\theta X}\right] \preceq I + (e^{\theta} - 1) \cdot \mathrm{E}[X], \end{aligned}$$

since the Löwner ordering is preserved under taking expectation (Claim 3 from Lecture 21). ■

Proof (of Matrix Chernoff Upper Bound). Without loss of generality $R = 1$. Our first observation is a bound for a sum of logs:

$$\begin{aligned} \textstyle\sum_{i=1}^k \log \mathrm{E}\left[e^{\theta X_i}\right] &= k \cdot \textstyle\sum_{i=1}^k \frac{1}{k} \log \mathrm{E}\left[e^{\theta X_i}\right] \\ &\preceq k \cdot \log\left(\textstyle\sum_{i=1}^k \frac{1}{k} \mathrm{E}\left[e^{\theta X_i}\right]\right) \quad \text{(by operator concavity of log)} \end{aligned} \tag{1}$$

Next:

$$
\begin{aligned}
&\operatorname{tr}\left(\mathrm{E}\left[e^{\theta X_1}\right] \odot \cdots \odot \mathrm{E}\left[e^{\theta X_k}\right]\right) \\
&\quad= \operatorname{tr}\exp\left(\textstyle\sum_{i=1}^{k} \log \mathrm{E}\left[e^{\theta X_i}\right]\right) && \text{(definition of } \odot\text{)} \\
&\quad\leq \operatorname{tr}\exp\left(k \cdot \log\left(\textstyle\sum_{i=1}^{k} \frac{1}{k}\mathrm{E}\left[e^{\theta X_i}\right]\right)\right) && \text{(by (1) and Claim 5 from Lecture 21)} \\
&\quad\leq d \cdot \lambda_{\max}\left(\exp\left(k \cdot \log\left(\textstyle\sum_{i=1}^{k} \frac{1}{k}\mathrm{E}\left[e^{\theta X_i}\right]\right)\right)\right) && \text{(sum of eigenvalues} \leq d \text{ times maximum)} \\
&\quad\leq d \cdot \exp\left(k \cdot \log \lambda_{\max}\left(\textstyle\sum_{i=1}^{k} \frac{1}{k}\mathrm{E}\left[e^{\theta X_i}\right]\right)\right) && \text{(definition of matrix exp and log)} \\
&\quad\leq d \cdot \exp\left(k \cdot \log \lambda_{\max}\left(I + \textstyle\sum_{i=1}^{k} \frac{1}{k}(e^{\theta}-1)\,\mathrm{E}\left[X_i\right]\right)\right) && \text{(by Claim 11)} \\
&\quad= d \cdot \exp\left(k \cdot \log\left(1 + \tfrac{e^{\theta}-1}{k}\lambda_{\max}(\textstyle\sum_{i=1}^{k} \mathrm{E}\left[X_i\right])\right)\right) \\
&\quad\leq d \cdot \exp\left((e^{\theta}-1) \cdot \lambda_{\max}(\textstyle\sum_{i=1}^{k} \mathrm{E}\left[X_i\right])\right) && \text{(using } \log(1+x) \leq x\text{)} \\
&\quad\leq d \cdot \exp\left((e^{\theta}-1) \cdot \mu_{\max}\right)
\end{aligned}
$$

Apply Claim 10 with $t = (1+\delta)\mu_{\max}$ and $\theta = \ln(1+\delta)$:

$$
\begin{aligned}
\Pr\left[\lambda_{\max}(\textstyle\sum_i X_i) \geq (1+\delta)\mu_{\max}\right] &\leq \exp\left(-\ln(1+\delta)\cdot(1+\delta)\mu_{\max}\right) \cdot \left(d \cdot \exp(\delta \cdot \mu_{\max})\right) \\
&= d \cdot \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}}
\end{aligned}
$$

■

4 Appendix

In this appendix, we prove inequalities (b) and (d) from Theorem 6. The same argument also proves inequalities (b) and (d) in Theorem 9.

4.1 Proof of inequality (b)

Claim 12. Suppose $\delta \in [0,1]$. Then $(1+x)\ln(1+x) - x \geq x^2/3$.

Proof. Note that the LHS and RHS both vanish at $x = 0$. So the claim holds if the derivative of the LHS is at least the derivative of the RHS on the interval $[0,1]$. By simple calculus,

$$\frac{d}{dx}\big[(1+x)\ln(1+x) - x\big] = \ln(1+x) \qquad \text{and} \qquad \frac{d}{dx}x^2/3 = 2x/3.$$

At $x = 0$, $\ln(1+x)$ equals $2x/3$. At $x = 1$, we have $\ln(1+x) = \ln(2) > 0.69$ and $2x/3 < 0.67$. Since $\ln(1+x)$ is concave, we have $\ln(1+x) \geq 2x/3$ for all $x \in [0,1]$. ■

Corollary 13. For all $\delta \in [0,1]$,

$$\Pr\left[\sum_{i=1}^{k} X_i \geq (1+\delta)\mu_{\max}\right] \;\leq\; \exp\big(-(\delta^2/3)\mu_{\max}/R\big).$$

Proof. Claim 12 implies that $\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right) \leq e^{-\delta^2/3}$. ■

4.2 Proof of inequality (d)

Claim 14. Suppose $x \in [0,1]$. Then $(1-x)\ln(1-x) + x \geq x^2/2$.

Proof. Note that the LHS and RHS both vanish at $x = 0$. So the claim holds if the derivative of the LHS is at least the derivative of the RHS on the interval $[0,1)$. By simple calculus,

$$\frac{d}{dx}\big[(1-x)\ln(1-x) + x\big] = -\ln(1-x) \qquad \text{and} \qquad \frac{d}{dx}x^2/2 = x.$$

The linear approximation of $-\ln(1-x)$ at $x = 0$ is

$$x \cdot \tfrac{d}{dx}\big(-\ln(1-x)\big)\Big|_{x=0} = x \cdot \big(\tfrac{1}{1-x}\big)\Big|_{x=0} = x.$$

Furthermore, $-\ln(1-x)$ is convex on $[0,1)$ because its second derivative is $1/(1-x)^2 \geq 0$. Thus $-\ln(1-x) \geq x$ on $[0,1)$. ■

Corollary 15. For all $\delta \in [0,1]$,

$$\Pr\left[\sum_{i=1}^{k} X_i \leq (1-\delta)\mu_{\min}\right] \leq \exp\big(-(\delta^2/2)\mu_{\min}/R\big)$$

Proof. Claim 14 implies that $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right) \leq e^{-\delta^2/2}$. ■