UBC CPSC 536N: Sparse Approximations	Winter 2013

Lecture 21 — March 25, 2013

Prof. Nick Harvey | *Scribe: Zachary Drudi*

Here we collect some definitions and results on the Löwner ordering and random matrices, in preparation for proving Tropp's inequality on sums of random matrices.

1 Random Matrices

We consider sums of random, independent, bounded matrices. We want to bound the concentration of eigenvalues or singular values. We will obtain a perfect generalization of the Chernoff bound.

Let X be a random matrix of size $d \times d$. There are two different ways to think of a random matrix:

1: A matrix sampled according to a distribution on matrices

2: An array of scalar random variables

Our perspective also impacts how we interpret the expectation of a random matrix.

1: If we consider X as sampled according to some distribution on matrices, then $\mathbb{E}[X] = \sum_A A \cdot \Pr[X = A]$

2: If we consider X as an array of random variables, then $\mathbb{E}[X]$ is the array of the expectations of the entries of X

Given independent, random, symmetric, positive semi-definite matrices $X_1, X_2, ..., X_k$, we want to understand the concentration of $\sum_i X_i$. Tropp's recent result solves this problem.

Theorem 1 (Tropp '12). Let $X_1, ..., X_k$ be independent random $d \times d$ symmetric matrices with $0 \preceq X_i \preceq R \cdot I$.
Let $\mu_{\min} \cdot I \preceq \sum_i \mathrm{E}\left[\, X_i \,\right] \preceq \mu_{\max} \cdot I$. Then, for all $\delta \in [0, 1]$,

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

In order to prove this theorem, we need to gather some definitions and results on symmetric matrices.

2 Löwner Ordering, Monotonicity, Convexity and Concavity

For any $f : \mathbb{R} \to \mathbb{R}$, we can define a function on symmetric matrices A by applying f to the eigenvalues of A. Formally, let $A = UDU^T$ be the spectral decomposition of A. That is, U is orthogonal and D is the diagonal matrix whose diagonal entries are the eigenvalues of A.
Define $f(A) = Uf(D)U^T$, where $f(D)$ is a diagonal matrix with $[f(D)]_{ii} = f(D_{ii})$.

We will use this definition with $f = \exp, \ln$.

Claim 2. Fact: Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ satisfy $f(x) \leq g(x) \forall x \in [l, u]$. Suppose A is symmetric and the spectrum of $A \subset [l, u]$. Then $f(A) \preceq g(A)$.

Proof. Let UDU^T be the spectral decomposition of A. Then

$$g(A) - f(A) = Ug(D)U^T - Uf(D)U^T = U(g(D) - f(D))U^T$$

Since the diagonal entries of D are in the interval $[l, u]$, we see $g(D_{ii}) \geq f(D_{ii})$ for all i. Therefore the diagonal matrix $g(D)-f(D)$ has non-negative entries on the diagonals, and thus is positive semi-definite, so $f(A) \preceq g(A)$. ■

How do functions behave with respect to the Löwner ordering? Usually badly. One might hope that if f is monotone on some interval $[l, u]$, then when we extend f to matrices, we obtain a monotone operator on matrices with eigenvalues in the interval $[l, u]$. That is, $A \preceq B$ and the eigenvalues of A, B are in $[l, u] \implies f(A) \preceq f(B)$. However, this is not true in general.

For a counter example, consider

$$f(x) = x^2, \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

So f is monotone on $\mathbb{R}_{\geq 0}$. Now

$$B - A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \succeq 0,$$

so $A \preceq B$. We claim $f(A) \npreceq f(B)$.

Proof. For any matrix C with decomposition UDU^T,

$$f(C) = Uf(D)U^T = UD^2U^T = (UDU^T)(UDU^T) = C^2,$$

so

$$f(A) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \quad f(B) = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \implies f(B) - f(A) = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$$

The charateristic polynomial of $f(B) - f(A)$ is $t^2 - 3t - 1$. Since the constant term is negative, the roots must be of different signs, and thus $f(B) - f(A)$ is not positive semi-definite, so $f(A) \npreceq f(B)$. ■

Claim 3. If X and Y are random matrices and $X \preceq Y$, then $\mathbb{E}[X] \preceq \mathbb{E}[Y]$.

Proof. Use the linearity of expectation. If $X \preceq Y$, then $Y - X$ is positive semi-definite. Therefore $\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X]$ is positive semi-definite, so $\mathbb{E}[X] \preceq \mathbb{E}[Y]$. ■

While monotone funtions on $\mathbb{R}$ do not necessarily yield monotone functions on symmetric matrices as we saw above, it is true that if f is monotone then $\text{trace}f := A \mapsto \text{trace}(f(A))$ is monotone. In order to establish this, we need a preliminary result concerning the spectrum of two matrices A, B with $A \preceq B$.

Claim 4 (Weyl's Monotonicity Theorem). Suppose A and B are symmetric, $n \times n$ matrices. Let $\lambda_i(A)$ be the ith largest eigenvalue of A. If $A \preceq B$, then $\lambda_i(A) \leq \lambda_i(B)$ for all i.

Proof. We use the variational characterization of eigenvalues for symmetric matrices:

$$\begin{aligned} &\lambda_i(A) = \max\{\min\{R_A(x) \mid x \in U \setminus \{0\}\} \mid U \subset V, \dim(U) = i\}, \\ &\text{where } R_A(x) = \frac{x^T A x}{x^T x} \end{aligned}$$

To see this, consider the decomposition of $\mathbb{R}^n$ into the eigenspaces $E_1, ..., E_n$ of A, where $E_j = \text{span}\{v_j\}$, and v_j is a unit eigenvector of A with eigenvalue $\lambda_j(A)$. By taking $U = S_i = \sum_{j=1}^{i} E_j$, we see the RHS above is $\geq R_A(v_i) = \frac{\lambda_i v_i^T v_i}{v_i^T v_i} = \lambda_i$, since v_i minimizes $R_A(x)$ for $x \in S_i$.

On the other hand, let P be the orthogonal projection onto S_i, let U be any subspace with dimension i and consider $P|_U$, the restriction of P to U. If $P|_U$ has trivial kernel, then $rank(P|_U) = dim(U) = rank(P)$, so we conclude $U = \text{im}(P) = S_i$. Otherwise, say $x \in \text{kernel}(P|_U), x \neq 0$. Then x is a linear combination of eigenvectors v_j with $j > i$, so

$$\frac{x^TAx}{x^Tx} = \frac{\sum_{j=i+1}^n \alpha_j^2 \lambda_j}{\sum_{j=i+1}^n \alpha_j^2} \leq \frac{\sum_{j=i+1}^n \alpha_j^2 \lambda_{i+1}}{\sum_{j=i+1}^n \alpha_j^2} = \lambda_{i+1} \leq \lambda_i,$$

thus the minimum of R_A over U is less than or equal to the minimum of R_A over S_i, and the RHS above is $\leq \lambda_i$.

Now we prove the claim. Suppose S_A maximizes the expression $\min\{R_A(x) \mid x \in S_A \setminus \{0\}\}$ among all subspaces with dimension i, and S_B is similiarly the maximizer for B. We have:

$$\begin{aligned}
&\lambda_i(B) - \lambda_i(A) \\
&= \min\{R_B(x) \mid x \in S_B \setminus \{0\}\} - \min\{R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\geq \min\{R_B(x) \mid x \in S_A \setminus \{0\}\} - \min\{R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\overset{(1)}{\geq} \min\{R_B(x) - R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\overset{(2)}{=} \min\{R_{B-A}(x) \mid x \in S_A \setminus \{0\}\} \\
&\geq \min\{R_{B-A}(x) \mid x \in \mathbb{R}^n \setminus \{0\}\} \\
&= \lambda_n(B - A) \\
&\overset{(3)}{\geq} 0
\end{aligned}$$

To obtain (1), say x_A, x_B are the minimizers for A, B in S_A respectively. Then $R_A(x_A) \leq R_A(x_B)$, so

$$R_B(x_B) - R_A(x_A) \geq R_B(x_B) - R_A(x_B) \geq \min\{R_B(x) - R_A(x) \mid x \in S_A \setminus \{0\}\},$$

establishing (1).
For (2), note

$$R_B(x) - R_A(x) = \frac{x^TBx}{x^Tx} - \frac{x^TAx}{x^Tx} = \frac{x^T(B-A)x}{x^Tx} = R_{B-A}(x)$$

(3) follows by the fact that $B - A$ is positive semi-definite. ■

We now establish our result about functions trace$f$, for monotone f.

Claim 5. If f is monotone, then trace$f$ is monotone.

Proof. This follows easily from Claim 4. Say $A \preceq B$. We establish $\mathrm{trace}f(A) \leq \mathrm{trace}f(B)$:

$$\mathrm{trace}f(A) = \sum_{i=1}^n f(\lambda_i(A)) \leq \sum_{i=1}^n f(\lambda_i(B)) = \mathrm{trace}f(B)$$

■

We will use this result for $f = \exp$.

We call f operator concave if $\forall x \in [0,1], \forall A, B$,
$f((1-x)A + xB) \succeq (1-x)f(A) + xf(B)$. Operator convexity is defined similarly, only with a flipped inequality.

As for monotone functions, f convex on $\mathbb{R}$ doesn't imply f is operator convex. For example, exp is not operator convex. However, it is known that log is operator concave on the set of positive definite matrices.