CPSC 536N: Randomized Algorithms	2011-12 Term 2

Lecture 12

Prof. Nick Harvey *University of British Columbia*

1 Concentration for sums of random matrices

Let X be a random real matrix of size $d \times d$. In other words, we have some probability distribution on the space of all $d \times d$ matrices, and we let X be a matrix obtained by sampling from that distribution. Alternatively, we can think of X as a matrix whose entries are real-valued random variables (that are not necessarily independent).

As usual, the expectation of X is simply the weighted average of the possible matrices that X could be, i.e., $\mathrm{E}[X] = \sum_A \Pr[X = A] \cdot A$. Alternatively, we can think of $\mathrm{E}[X]$ as matrix whose entries are the expectations of the entries of X.

Many concentration results are known for matrices whose entries are *independent* random variables from certain real-valued distributions (e.g., Gaussian, subgaussian, etc.) In fact, in Lecture 8 on Compressed Sensing, we proved concentration of the singular values of a matrix whose entries are independent Gaussians. In this lecture, we will look at random matrices whose entries are *not independent*, and we will obtain concentration results by summing multiple *independent copies* of those matrices.

1.1 The Ahlswede-Winter Inequality

The Chernoff bound is a very powerful tool for proving concentration for sums of independent, real-valued random variables. Today we will prove the Ahlswede-Winter inequality, which is a generalization of the Chernoff bound for proving concentration for sums of independent, *matrix*-valued random variables.

Let $X_1, \ldots, X_k$ be random, independent, symmetric matrices of size $d \times d$. Define the partial sums $S_j = \sum_{i=1}^{j} X_i$. We would like to analyze the probability that all eigenvalues of S_k are at most t (i.e., $S_k \preceq tI$). For any $\lambda > 0$, this is equivalent to all eigenvalues of $e^{\lambda S_k}$ being at most $e^{\lambda t}$ (i.e., $e^{\lambda S_k} \preceq e^{\lambda t I}$). If this event fails to hold then then certainly $\operatorname{tr} e^{\lambda S_k} > e^{\lambda t}$, since all eigenvalues of $e^{\lambda S_k}$ are non-negative. Thus we have bounded the probability that some eigenvalue of S_k is greater than t as follows:

$$\Pr[S_k \not\preceq tI] \;\leq\; \Pr[\operatorname{tr} e^{\lambda S_k} > e^{\lambda t}] \;\leq\; \mathrm{E}[\operatorname{tr} e^{\lambda S_k}]/e^{\lambda t}, \tag{1}$$

by Markov's inequality.

Now let us observe a useful property of the trace. Since it is linear, it commutes with expectation:

$$\begin{aligned} \mathrm{E}[\operatorname{tr} X] &= \sum_A \Pr[X = A] \cdot \sum_i A_{i,i} \;=\; \sum_i \sum_A \Pr[X = A] \cdot A_{i,i} \\ &= \sum_i \sum_a \Pr[X_{i,i} = a] \cdot a \;=\; \sum_i \mathrm{E}[X_{i,i}] \;=\; \operatorname{tr}\big(\mathrm{E}[X]\big). \end{aligned}$$

The proof of the Ahlswede-Winter inequality is very similar to the proof of the Chernoff bound; one just has to be a bit careful to do the matrix algebra properly. As in the proof of the Chernoff bound, the main technical step is to bound the expectation in (1) by a product of expectations that each involve a single X_i, because those individual expectations are much easier to analyze. This is where the Golden-Thompson inequality (Theorem 17 in the Notes on Symmetric Matrices) is needed.

$$\begin{aligned}
\mathrm{E}[\operatorname{tr} e^{\lambda S_k}] &= \mathrm{E}[\operatorname{tr} e^{\lambda X_k + \lambda S_{k-1}}] \qquad (\text{since } S_k = X_k + S_{k-1}) \\
&\leq \mathrm{E}[\operatorname{tr}(e^{\lambda X_k} \cdot e^{\lambda S_{k-1}})] \qquad (\text{by Golden} - \text{Thompson}) \\
&= \mathrm{E}_{X_1,\ldots,X_{k-1}}\left[\mathrm{E}_{X_k}[\operatorname{tr}(e^{\lambda X_k} \cdot e^{\lambda S_{k-1}})]\right] \qquad (\text{by independence}) \\
&= \mathrm{E}_{X_1,\ldots,X_{k-1}}\left[\operatorname{tr}\left(\mathrm{E}_{X_k}[e^{\lambda X_k} \cdot e^{\lambda S_{k-1}}]\right)\right] \qquad (\text{trace and expectation commute}) \\
&= \mathrm{E}_{X_1,\ldots,X_{k-1}}\left[\operatorname{tr}\left(\mathrm{E}_{X_k}[e^{\lambda X_k}] \cdot e^{\lambda S_{k-1}}\right)\right] \qquad (X_k \text{ and } S_{k-1} \text{ are independent}) \\
&\leq \mathrm{E}_{X_1,\ldots,X_{k-1}}\left[\|\mathrm{E}_{X_k}[e^{\lambda X_k}]\| \cdot \operatorname{tr} e^{\lambda S_{k-1}}\right] \\
&= \|\mathrm{E}_{X_k}[e^{\lambda X_k}]\| \cdot \mathrm{E}_{X_1,\ldots,X_{k-1}}[\operatorname{tr} e^{\lambda S_{k-1}}],
\end{aligned}$$

where the last inequality follows from Corollary 14 in the Notes on Symmetric Matrices. Applying this inequality inductively, we get

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

since $e^{\lambda 0} = I$ and $\operatorname{tr} I = d$. Combining this with (1), we obtain

$$\Pr[S_k \not\preceq tI] \;\leq\; d e^{-\lambda t} \prod_{i=1}^{k} \|\mathrm{E}[e^{\lambda X_i}]\|.$$

We can also bound the probability that any eigenvalue of S_k is less than $-t$ by applying the same argument to $-S_k$. This shows that the probability that any eigenvalue of S_k lies outside $[-t, t]$ is

$$\Pr[\|S_k\| > t] \;\leq\; d e^{-\lambda t} \left(\prod_{i=1}^{k} \|\mathrm{E}[e^{\lambda X_i}]\| + \prod_{i=1}^{k} \|\mathrm{E}[e^{-\lambda X_i}]\| \right). \tag{2}$$

This is the basic inequality. Much like the Chernoff bound, there are many variations. We will see some next time.