C&O 750: Randomized Algorithms
Winter 2011
Lecture 11 Notes

Nicholas Harvey
`http://www.math.uwaterloo.ca/~harvey/`

Abstract

The Chernoff bound proves concentration for sums of independent Bernoulli random variables. As we have seen, the same general method works for other sorts of random variables too, e.g., geometric random variables. The topic of this lecture is the Ahlswede-Winter inequality, which uses similar ideas to prove concentration for sums of random matrices.

1 Basic Review

We begin by reviewing some facts about matrices.

Let A be a real, symmetric matrix of size $d \times d$. Let I denote the $d \times d$ identity matrix. The spectral theorem states that there exists a (real) orthogonal matrix U (meaning that $U^\mathsf{T} U = I$) and a real, diagonal matrix D such that

$$A \;=\; UDU^\mathsf{T}.$$

This is called a ***spectral decomposition*** of A. Let u_i be the i^{th} column of U and let λ_i denote the i^{th} diagonal entry of D. Then $\{u_1, ..., u_d\}$ is an orthonormal basis consisting of eigenvectors of A, and λ_i is the eigenvalue corresponding to A.

The matrix A is called ***positive semi-definite*** if all of its eigenvalues are non-negative. This is denoted $A \succeq 0$. Furthermore, for any two symmetric matrices A and B, we write $A \succeq B$ if $A - B \succeq 0$. One can show that this defines a partial order on all symmetric matrices.

Note that, for any t, the eigenvalues of $A - tI$ are $\lambda_1 - t, ..., \lambda_d - t$. The ***spectral norm*** of A, denoted $\|A\|$, is defined to be $\max_i |\lambda_i|$. Thus $-\|A\| \cdot I \preceq A \preceq \|A\| \cdot I$.

The ***trace*** of A, denoted $\operatorname{tr}(A)$, is $\sum_{i=1}^{d} \lambda_d$, and it is also the sum of the diagonal entries of A.

Claim 1. Let A, B and C be symmetric $d \times d$ matrices satisfying $A \succeq 0$ and $B \preceq C$. Then $\operatorname{tr}(A \cdot B) \leq \operatorname{tr}(A \cdot C)$.

Corollary 2. If $A \succeq 0$ then $\operatorname{tr}(A \cdot B) \leq \|B\| \cdot \operatorname{tr}(A)$.

Proof. Apply Claim 1 with $C = \|B\| \cdot I$ and note that $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr}(A)$ for any scalar α. ■

For any function $f : \mathbb{R} \to \mathbb{R}$, we extend f to a function on symmetric matrices as follows. Let $A = UDU^\mathsf{T}$ be a spectral decomposition of A. Then we define

$$f(A) \;=\; U \begin{pmatrix} f(D_{1,1}) & & \\ & \ddots & \\ & & f(D_{d,d}) \end{pmatrix} U^\mathsf{T}.$$

In other words, $f(A)$ is defined simplying by applying the function f to the eigenvalues of A.

Claim 3. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ satisfy $f(x) \le g(x)$ for all $x \in [l, u] \subset \mathbb{R}$. Let A be a symmetric matrix for which all eigenvalues lie in $[l, u]$ (i.e., $lI \preceq A \preceq uI$). Then $f(A) \preceq g(A)$.

We will mainly be interested in the case $f(x) = e^x$. For any symmetric matrix A, note that e^A is positive semi-definite. Whereas in the scalar case $e^{a+b} = e^a e^b$ holds, it is not necessarily true in the matrix case that $e^{A+B} = e^A \cdot e^B$. However, the following useful inequality does hold:

Theorem 4 (Golden-Thompson Inequality). Let A and B be symmetric $d \times d$ matrices. Then $\operatorname{tr}(e^{A+B}) \le \operatorname{tr}(e^A \cdot e^B)$.

2 The Ahlswede-Winter Inequality

Let X be a random $d \times d$ matrix, i.e., a matrix whose entries are all random variables. We define $\mathrm{E}[X]$ to be the matrix whose entries are the expectations of the entries of X. Since expectation and trace are both linear, they commute:

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

Let $X_1, ..., X_n$ be random, 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 eigenvalues of S_n are at most t (i.e., $S_n \preceq tI$). For any $\lambda > 0$, this is equivalent to all eigenvalues of $e^{\lambda S_n}$ being at most $e^{\lambda t}$ (i.e., $e^{\lambda S_n} \preceq e^{\lambda t I}$). If this event fails to hold then then certainly $\operatorname{tr} e^{\lambda S_n} > e^{\lambda t}$, since all eigenvalues of $e^{\lambda S_n}$ are non-negative. Thus we have argued that

$$\Pr[\text{some eigenvalue of } S_n \text{ is greater than } t] \le \Pr\left[\operatorname{tr} e^{\lambda S_n} > e^{\lambda t}\right] \le \mathrm{E}\left[\operatorname{tr} e^{\lambda S_n}\right] / e^{\lambda t}, \tag{1}$$

by Markov's inequality. Now, as in the proof of the Chernoff bound, we want to bound this expectation by a product of expectations, which will lead to an exponentially decreasing tail bound. This is where the Golden-Thompson inequality is needed.

$$\begin{aligned}
\mathrm{E}\left[\operatorname{tr} e^{\lambda S_n}\right] &= \mathrm{E}\left[\operatorname{tr} e^{\lambda X_n + \lambda S_{n-1}}\right] && (\text{since } S_n = X_n + S_{n-1}) \\
&\le \mathrm{E}\left[\operatorname{tr}(e^{\lambda X_n} \cdot e^{\lambda S_{n-1}})\right] && (\text{by Golden-Thompson}) \\
&= \mathrm{E}_{X_1,\ldots,X_{n-1}}\left[\mathrm{E}_{X_n}\left[\operatorname{tr}(e^{\lambda X_n} \cdot e^{\lambda S_{n-1}})\right]\right] && (\text{since the } X_i\text{'s are mutually independent}) \\
&= \mathrm{E}_{X_1,\ldots,X_{n-1}}\left[\operatorname{tr}\left(\mathrm{E}_{X_n}\left[e^{\lambda X_n} \cdot e^{\lambda S_{n-1}}\right]\right)\right] && (\text{since trace and expectation commute}) \\
&= \mathrm{E}_{X_1,\ldots,X_{n-1}}\left[\operatorname{tr}\left(\mathrm{E}_{X_n}\left[e^{\lambda X_n}\right] \cdot e^{\lambda S_{n-1}}\right)\right] && (\text{since } X_n \text{ and } S_{n-1} \text{ are independent}) \\
&\le \mathrm{E}_{X_1,\ldots,X_{n-1}}\left[\left\|\mathrm{E}_{X_n}\left[e^{\lambda X_n}\right]\right\| \cdot \operatorname{tr} e^{\lambda S_{n-1}}\right] && (\text{by Corollary 2}) \\
&= \left\|\mathrm{E}_{X_n}\left[e^{\lambda X_n}\right]\right\| \cdot \mathrm{E}_{X_1,\ldots,X_{n-1}}\left[\operatorname{tr} e^{\lambda S_{n-1}}\right]
\end{aligned}$$

Applying this inequality inductively, we get

$$\mathrm{E}\left[\operatorname{tr} e^{\lambda S_n}\right] \le \prod_{i=1}^{n} \left\|\mathrm{E}\left[e^{\lambda X_i}\right]\right\| \cdot \operatorname{tr} e^{\lambda 0},$$

where 0 is the zero matrix of size $d \times d$. So $e^{\lambda 0} = I$ and

$$\mathrm{E}\left[\operatorname{tr} e^{\lambda S_n}\right] \;\leq\; d \cdot \prod_{i=1}^{n} \left\| \mathrm{E}\left[e^{\lambda X_i} \right] \right\|.$$

Combining this with Eq. (1), we obtain

$$\Pr[\text{some eigenvalue of } S_n \text{ is greater than } t] \;\leq\; de^{-\lambda t} \prod_{i=1}^{n} \left\| \mathrm{E}\left[e^{\lambda X_i} \right] \right\|.$$

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

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

This is the basic inequality. Much like the Chernoff bound, numerous variations and generalizations are possible.

3 Rudelson's Theorem

In this section, we use the Ahlswede-Winter inequality to prove a concentration inequality for random vectors due to Rudelson. His original proof was quite different.

The motivation for Rudelson's inequality comes from the problem of approximately computing the volume of a convex body, which has been a topic of significant interest in theoretical computer science for the last two decades. (In fact, this problem is quite relevant for this class, because any efficient algorithm for this problem must necessarily be randomized.) When solving this problem, a convenient first step is to transform the body into "isotropic position", which is a technical way of saying "roughly like the unit sphere". To perform this first step, one requires a concentration inequality for randomly sampled vectors, which is provided by Rudelson's theorem.

Theorem 5. Let $x \in \mathbb{R}^d$ be a random vector such that $\mathrm{E}\left[xx^{\mathsf{T}} \right] = I$. Suppose $\|x\| \leq R$. Let $x_1, ..., x_n$ be independent copies of x. For any $\epsilon \in (0, 1)$, we have

$$\Pr\left[\left\| \frac{1}{n} \sum_{i=1}^{n} x_i x_i^{\mathsf{T}} - I \right\| > \epsilon \right] \;\leq\; 2d \cdot \exp(-\epsilon^2 n / 4R^2).$$

Remark: Note that $R \geq \sqrt{d}$ because

$$d \;=\; \operatorname{tr} I \;=\; \operatorname{tr} \mathrm{E}\left[xx^{\mathsf{T}} \right] \;=\; \mathrm{E}\left[\operatorname{tr}(xx^{\mathsf{T}}) \right] \;=\; \mathrm{E}\left[x^{\mathsf{T}} x \right].$$

Proof. We apply Ahlswede-Winter with

$$X_i \;=\; \frac{1}{2R^2}\left(x_i x_i^{\mathsf{T}} - \mathrm{E}\left[x_i x_i^{\mathsf{T}} \right] \right) \;=\; \frac{1}{2R^2}\left(x_i x_i^{\mathsf{T}} - I \right).$$

Note that $\mathrm{E}[X_i] = 0$, $\|X_i\| \leq 1$, and

$$\begin{aligned}
\mathrm{E}\left[X_i^2\right] &= \frac{1}{4R^4}\,\mathrm{E}\left[(x_i x_i^{\mathsf{T}} - I)^2\right] \\
&= \frac{1}{4R^4}\Big(\mathrm{E}\left[(x_i x_i^{\mathsf{T}})^2\right] - I\Big) \qquad \text{(since } \mathrm{E}\left[x_i x_i^{\mathsf{T}}\right] = I\text{)} \\
&\preceq \frac{1}{4R^4}\,\mathrm{E}\left[(x_i^{\mathsf{T}} x_i) x_i x_i^{\mathsf{T}}\right] \\
&\preceq \frac{R^2}{4R^4}\,\mathrm{E}\left[x_i x_i^{\mathsf{T}}\right] \qquad \text{(since } \|x_i\| \leq R\text{)} \\
&= \frac{I}{4R^2}.
\end{aligned} \tag{3}$$

Now we use Claim 3 together with the inequalities

$$\begin{aligned}
1 + x &\leq e^x \quad \forall x \in \mathbb{R} \\
e^x &\leq 1 + x + x^2 \quad \forall x \in [-1, 1].
\end{aligned}$$

Since $\|X_i\| \leq 1$, for any $\lambda \in [0, 1]$, we have $e^{\lambda X_i} \preceq I + \lambda X_i + \lambda^2 X_i^2$, and so

$$\mathrm{E}\left[e^{\lambda X_i}\right] \;\preceq\; \mathrm{E}\left[I + \lambda X_i + \lambda^2 X_i^2\right] \;\preceq\; I + \lambda^2\,\mathrm{E}\left[X_i^2\right] \;\preceq\; e^{\lambda^2\,\mathrm{E}\left[X_i^2\right]} \;\preceq\; e^{(\lambda^2/4R^2)I},$$

by Eq. (3). Thus $\left\|\mathrm{E}\left[e^{\lambda X_i}\right]\right\| \leq e^{\lambda^2/4R^2}$. The same analysis also shows that $\left\|\mathrm{E}\left[e^{-\lambda X_i}\right]\right\| \leq e^{\lambda^2/4R^2}$. Substituting this into Eq. (2), we obtain

$$\Pr\left[\left\|\sum_{i=1}^{n} \frac{1}{2R^2}\Big(x_i x_i^{\mathsf{T}} - I\Big)\right\| > t\right] \;\leq\; 2d \cdot e^{-\lambda t} \prod_{i=1}^{n} e^{\lambda^2/4R^2} \;=\; 2d \cdot \exp(-\lambda t + n\lambda^2/4R^2)$$

Substituting $t = n\epsilon/2R^2$ and $\lambda = \epsilon$ proves the theorem. ■

Remarks

These notes are adapted from the lecture notes of Roman Vershynin [3]. The Ahlswede-Winter inequality is originally from [1]. Rudelson's inequality is originally from [2].

References

[1] R. Ahlswede and A. Winter. "Strong converse for identification via quantum channels". IEEE Transactions on Information Theory, 2002.

[2] M. Rudelson. "Random vectors in the isotropic position". Journal of Functional Analysis, 1999.

[3] R. Vershynin. "A note on sums of independent random matrices after Ahlswede-Winter". Manuscript, 2009.