Stat 991: Multivariate Analysis, Dimensionality Reduction, and Spectral Methods **Lecture: 23A**

Matrix Concentration Derivations

Instructor: Sham Kakade

1 Introduction

Let $X \in \mathbb{R}^{d_1 \times d_2}$ be a random matrix. In many settings, we are interested in the behavior of either:

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{\mathrm{F}} \leq ?, \quad \left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{2} \leq ?$$

where each X_i is sampled i.i.d. from some distribution. Here, $\|\cdot\|_2$ denotes the spectral norm (the largest eigenvalue) and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm.

The following theorem provides a high probability bound on these quantities.

Theorem 1.1. *Assume that $X_i \in \mathbb{R}^{m \times n}$ are sampled i.i.d. Let $d = \min\{d_1, d_2\}$.*

- *(Spectral Norm) Suppose $||X||_2 \leq M$ almost surely. Then with probability greater than $1 - \delta$,*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{2} \leq 6M \sqrt{\frac{1}{n}} \left(\sqrt{\log d} + \sqrt{\log \frac{1}{\delta}} \right) .$$

- *(Frobenius Norm) Suppose $||X||_{\mathrm{F}} \leq M$ almost surely. Then with probability greater than $1 - \delta$,*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right\|_{\mathrm{F}} \leq 6M \sqrt{\frac{1}{n}} \left(1 + \sqrt{\log \frac{1}{\delta}} \right) .$$

1.1 Concentration and Strong-smoothness

Throughout we let $\mathcal{X}$ be a Euclidean vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. We also work with a norm $\|\cdot\|$ over $\mathcal{X}$ (and this norm need not the one induced by $\langle \cdot, \cdot \rangle$).

Definition 1.2. *A function $f : \mathcal{X} \to \mathbb{R}$ is β-strongly smooth w.r.t. a norm $\|\cdot\|$ if f is everywhere differentiable and if for all x, y we have*

$$f(x + y) \leq f(x) + \langle \nabla f(x), y \rangle + \frac{1}{2}\beta \|y\|^2$$

We now point out the role of strong smoothness in proving certain concentration results. In particular, we are interested in the behavior of a function $f(\sum_{i=1}^{n} Z_i)$ where Z_i is a martingale difference sequence. The following simple lemma bounds the expectation of this quantity.

Lemma 1.3. *[Juditsky and Nemirovski; 08] Suppose that Z_i is a martingale difference sequence (where $Z_i \in \mathcal{X}$) and that $||Z_i|| \leq M_i$ almost surely. Also, suppose that f^2 is β-strongly smooth w.r.t. a norm $\|\cdot\|$ on $\mathcal{X}$ and that $f(\mathbf{0}) = 0$.*

$$\mathbb{E} f \left(\sum_{i=1}^{n} Z_i \right) \leq \sqrt{\frac{1}{2}\beta \sum_{i=1}^{n} M_i^2}$$

Proof. By smoothness we have:

$$\begin{aligned}
\mathbb{E}f^2\left(\sum_{i=1}^{n} Z_i\right) &\leq \mathbb{E}f^2\left(\sum_{i=1}^{n-1} Z_i\right) + \mathbb{E}\left\langle \nabla f^2\left(\sum_{i=1}^{n-1} Z_i\right), Z_n\right\rangle + \frac{1}{2}\beta\mathbb{E}\|Z_n\|^2 \\
&= \mathbb{E}f^2\left(\sum_{i=1}^{n-1} Z_i\right) + \mathbb{E}\left[\left\langle \nabla f^2\left(\sum_{i=1}^{n-1} Z_i\right), \mathbb{E}[Z_n|Z_1,\dots Z_{n-1}]\right\rangle\right] + \frac{1}{2}\beta\mathbb{E}\|Z_n\|^2 \\
&\leq \mathbb{E}f^2\left(\sum_{i=1}^{n-1} Z_i\right) + 0 + \frac{1}{2}\beta X_n^2
\end{aligned}$$

where have used that Z_n is a martingale difference sequence. Proceeding recursively and using that $f(\mathbf{0}) = 0$, we have that:

$$\mathbb{E}f^2\left(\sum_{i=1}^{n} Z_i\right) \leq \frac{1}{2}\beta\sum_{i=1}^{n} M_i^2$$

and proof is completed by Jensen's inequality. □

To obtain concentration, we can directly appeal to Hoeffding-Azuma if f is a norm. However, note that in the following lemma we do not require f^2 to be strongly smooth (which is useful for the case of the spectral norm, which is not strongly smooth).

Lemma 1.4. *Let f be a norm. Suppose that Z_i are independent (where $Z_i \in \mathcal{X}$) and that $f(Z_i) \leq M_i$ almost surely. Then with probability greater than $1-\delta$,*

$$f\left(\sum_{i=1}^{n} Z_i\right) \leq \mathbb{E}f\left(\sum_{i=1}^{n} Z_i\right) + \sqrt{8\log\frac{1}{\delta}\sum_{i=1}^{n} M_i^2}$$

Proof. Using that f is a norm,

$$\left| f\left(\sum_i Z_i\right) - f\left(\sum_{i\neq j} Z_i + Z_j'\right)\right| \leq f(Z_j) + f(Z_j')$$

for all $Z_1,\dots Z_n$, and Z_j'. Since the distribution over Z_i's are independent, this implies (Doob's) martingale $D_j = \mathbb{E}[f\left(\sum_{i=1}^{n} Z_i\right)|Z_j,\dots Z_1]$ satisfies the bounded difference property:

$$|D_j - D_{j-1}| \leq 2M_j\,.$$

The result now follows from Hoeffding-Azuma. □

1.2 Matrix Concentration Proofs

The Schatten q-norm is defined as:

$$\frac{1}{2}\|X\|_{S(q)}^2 := \frac{1}{2}\|\sigma(X)\|_q^2$$

where $\sigma(X)$ is the singular values of X and $\|\cdot\|_q^2$ is the usual L_q-norm. The function f^2:

$$f^2(X) = \frac{1}{2}\|X\|_{S(q)}^2$$

is $(q-1)$-strongly smooth, as shown in [Juditsky and Nemirovski; 08] (for $q \geq 2$).

Note that the spectral norm $\|\cdot\|_2$ is just the Schatten ∞-norm $\|\cdot\|_{S(\infty)}$ and the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ is just the Schatten 2-norm $\|\cdot\|_{S(2)}$.

Proof. For the spectral norm case, note that our assumption that $\|X\|_{S(\infty)} \le M$ (almost surely) implies $\|X\|_{S(q)} \le d^{1/q} M$ (almost surely). Let $Z_i = X_i - \mathbb{E}[X]$. By convexity of norms and Jensen's inequality, $\|\mathbb{E}[X]\|_{S(q)} \le \mathbb{E}[\|X\|_{S(q)}] \le d^{1/q} M$. So we have that $\|Z_i\|_{S(q)} \le 2d^{1/q} M$ (almost surely). Hence, by Lemma 1.3:

$$\mathbb{E}\left\|\sum_{i=1}^{n} Z_i\right\|_{S(\infty)} \le \mathbb{E}\left\|\sum_{i=1}^{n} Z_i\right\|_{S(q)} \le \sqrt{4(q-1)nd^{2/q}M^2}$$

and choosing $q = \log d$

$$\mathbb{E}\left\|\sum_{i=1}^{n} Z_i\right\|_{S(\infty)} \le 2Me\sqrt{n \log d}$$

Now let us apply Lemma 1.4 with f as the spectral norm $\|\cdot\|_{S(\infty)}$. Here, we have that $\|Z_i\|_{S(\infty)} \le 2M$ (almost surely), and our first claim follows.

For the Frobenius norm case, again let $Z_i = X_i - \mathbb{E}[X]$. Then by convexity of norms and Jensen's inequality, $\|\mathbb{E}[X]\|_{S(2)} \le \mathbb{E}[\|X\|_{S(2)}] \le M$. So we have that $\|Z_i\|_{S(2)} \le 2M$ (almost surely). Hence, by Lemma 1.3:

$$\mathbb{E}\left\|\sum_{i=1}^{n} Z_i\right\|_{S(2)} \le \sqrt{4nM^2}$$

Using Lemma 1.4 with this norm and $\|Z_i\|_{S(2)} \le 2M$, we have our second claim. $\square$

Acknowledgements

These notes are from discussions with Ambuj Tewari and Shai Shalev-Schwartz. Also, they borrow heavily from ideas in [Juditsky and Nemirovski; 08].