COS 598I Lecture 3: Tail bounds for matrix concentration

Lecturer: Pravesh Kothari   Scribe: Ştefan Tudose

Spring 2025

In this lecture we will see how to estimate the variance of the largest eigenvalue of random matrices with bounded entries by using the Efron-Stein theorem (also known as the tensorization trick). We will also obtain a bound on the largest eigenvalue of random symmetric matrices with Gaussian entries, as well as tail bounds for the eigenvalues and the operator norm. These will be consequences of the Gaussian Poincaré inequality and Gaussian concentration. At the end we briefly discuss about Talagrand's inequality, which is an analogue of Gaussian concentration for bounded random variables.

Contents

1 The Efron-Stein inequality

Given independent random variables $X_1, \ldots, X_n$, a function $f(X_1, \ldots, X_n)$ tends to stay close to its mean as long as it is not too sensitive to changes in any entry. The Efron-Stein theorem is an instance of this phenomenon.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, define $\mathrm{Var}_{x_i} f : \mathbb{R}^{n-1} \to \mathbb{R}$ by

$$(\mathrm{Var}_{x_i} f)(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \mathbb{E}_{x_i} \left(f(x_1, \ldots, x_n) - \mathbb{E}_{x_i} f(x_1, \ldots, x_n) \right)^2$$

In other words, $\mathrm{Var}_{x_i} f$ is the variance of f with respect to the ith variable when we freeze all the other variables. When we will write $\mathbb{E}\mathrm{Var}_{x_i} f$ below, we refer to the expectation with respect to all the variables.

Theorem 1 (Efron-Stein)**.** Let $X_1, \ldots, X_n$ be independent random variables and let $f : \mathbb{R}^n \to \mathbb{R}$. Then

$$\mathrm{Var} f \leq \sum_{i=1}^{n} \mathbb{E}\mathrm{Var}_{x_i} f$$

Proof. For $1 \leq k \leq n$, define $\mathbb{E}_{[k]} f = \mathbb{E}_{x_1, \ldots, x_k} f$ the expectation of f with respect to the first k variables. We also define $\mathbb{E}_{[0]} f = f$. We can then write

$$f - \mathbb{E} f = \sum_{k=0}^{n-1} \left(\mathbb{E}_{[k]} f - \mathbb{E}_{[k+1]} f \right)$$

Note that if $k < l$ then

$$\mathbb{E}\left(\left(\mathbb{E}_{[k]}f - \mathbb{E}_{[k+1]}f\right)\left(\mathbb{E}_{[l]}f - \mathbb{E}_{[l+1]}f\right)\right)$$

is zero. Indeed, the only term inside the expectation which depends on x_{k+1} is $\mathbb{E}_{[k]}f$ and

$$\mathbb{E}_{x_{k+1}}\mathbb{E}_{[k]}f = \mathbb{E}_{[k+1]}f$$

Therefore

$$\mathrm{Var} f = \mathbb{E}(f - \mathbb{E}f)^2 = \sum_{k=0}^{n-1}\mathbb{E}\left(\mathbb{E}_{[k]}f - \mathbb{E}_{[k+1]}f\right)^2$$

By Jensen we have

$$\left(\mathbb{E}_{[k]}f - \mathbb{E}_{[k+1]}f\right)^2 = (\mathbb{E}_{[k]}(f - \mathbb{E}_{x_{k+1}}f))^2 \leq \mathbb{E}_{[k]}(f - \mathbb{E}_{x_{k+1}}f)^2$$

and so

$$\mathrm{Var} f \leq \sum_{k=0}^{n-1}\mathbb{E}(f - \mathbb{E}_{x_{k+1}}f)^2 = \sum_{k=1}^{n}\mathbb{E}\mathbb{E}_{x_k}(f - \mathbb{E}_{x_k}f)^2 = \sum_{k=1}^{n}\mathbb{E}\mathrm{Var}_{x_k}f$$

□

`Remark:` Note that the inequality is sharp as shown by linear functions

$$f(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n + b$$

We make the observation that given $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathbb{E}_x\mathrm{Var}_{x_i}f = \frac{1}{2}\mathbb{E}_{x,y}(f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, y_i, \ldots, x_n))^2$$

where $y = (y_1, \ldots, y_n)$ is an independent copy of $x = (x_1, \ldots, x_n)$. The Efron-Stein alternatively says that

$$\mathrm{Var} f \leq \frac{1}{2}\mathbb{E}_{x,y}\sum_{i=1}^{n}(f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, y_i, \ldots, x_n))^2$$

We will also use that if X, Y are independent and identically distributed, then

$$\mathbb{E}((X-Y)^2) = \mathbb{E}((X-Y)^2 1_{X>Y} + (X-Y)^2 1_{X=Y} + (X-Y)^2 1_{X<Y}) = 2\mathbb{E}((X-Y)^2 1_{X>Y})$$

and therefore

$$\mathrm{Var} f \leq \mathbb{E}_{x,y}\sum_{i=1}^{n}(f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, y_i, \ldots, x_n))^2 \cdot 1_{f(x_1,\ldots,x_i,\ldots,x_n)\geq f(x_1,\ldots,y_i,\ldots,x_n)}$$

As a consequence, we can estimate the variance of the eigenvalues of random symmetric matrices with bounded entries:

Theorem 2. Let W be a $n \times n$ symmetric random matrix such that $|W_{ij}| \leq K$ for all i, j almost surely. Then

$$\mathrm{Var}(\lambda_1(W)) \leq 16K^2$$

where λ_1 is the largest eigenvalue of W.

Proof. By the above observation, it suffices to understand how λ_1 changes when changing entry (i,j) (and implicitly (j,i) since W is symmetric). Consider then the matrix W'_{ij} which differs from W only in entries (i,j) (and (j,i)). Let v be an unit eigenvector of $\lambda_1(W)$.

We have

$$\lambda_1(W) - \lambda_1(W'_{ij}) = v^T W v - \max_{\|u\|=1} u^T W'_{ij} u \leq v^T W v - v^T W'_{ij} v = v^T (W - W'_{ij}) v \leq 4K|v_i||v_j|$$

where the last inequality follows because $W - W'_{ij}$ has at most two entries which are not equal (namely (i,j) and (j,i)), and the values there are at most $2K$. This means

$$\left(\lambda_1(W) - \lambda_1(W'_{ij})\right)^2 \cdot 1_{\lambda(W) > \lambda(W'_{ij})} \leq 16K^2 v_i^2 v_j^2$$

We use theorem 1 (Efron-Stein) in the form of the observation above to get

$$\mathrm{Var}(\lambda_1(W)) \leq \mathbb{E} \sum_{i \geq j} \left(\lambda_1(W) - \lambda_1(W'_{ij})\right)^2 \cdot 1_{\lambda(W) > \lambda(W'_{ij})} \leq \mathbb{E}\left(\sum_{i \geq j} 16K^2 v_i^2 v_j^2\right) \leq 16K^2$$

where we used that v is a unit vector and

$$\sum_{i \geq j} v_i^2 v_j^2 = \sum_i v_i^2 v_i^2 + \sum_{i > j} v_i^2 v_j^2 \leq \sum_i v_i^2 v_i^2 + 2 \sum_{i > j} v_i^2 v_j^2 = \left(\sum_i v_i^2\right)^2 = 1$$

□

2 Gaussian concentration

We will see in this section that C^1 functions $f : \mathbb{R}^n \to \mathbb{R}$ with a well-behaved gradient (e.g. Lipschitz functions), when evaluated on i.i.d. standard Gaussians, tend to be close to their mean.

We will repeatedly use the following fact: given $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ two C^1 functions and $g_1, \ldots, g_n$ i.i.d. standard Gaussians, we have

$$\mathrm{Cov}(f_1(g_1, \ldots, g_n), f_2(g_1, \ldots, g_n)) = \int_0^1 \mathbb{E} \nabla f_1(g_1, \ldots, g_n) \cdot \nabla f_2(g_1(t), \ldots, g_n(t)) dt$$

where $g_i(t) = t g_i + \sqrt{1-t^2} g'_i$ and $g'_1, \ldots, g'_n$ are independent copies of $g_1, \ldots, g_n$. A proof can be found in the appendix. Note that $g_1(t), \ldots, g_n(t)$ are i.i.d. standard Gaussians as well.

Theorem 3 (Gaussian Poincaré inequality)**.** Given $f : \mathbb{R}^n \to \mathbb{R}$ a C^1 function,

$$\mathrm{Var} f(g_1, \ldots, g_n) \leq \mathbb{E}\left(\|\nabla f\|^2 (g_1, \ldots, g_n)\right)$$

Proof 1. Using the covariance identity above

$$\mathrm{Var} f = \mathrm{Cov}(f, f) = \int_0^1 \mathbb{E} \nabla f(g_1, \ldots, g_n) \cdot \nabla f(g_1(t), \ldots, g_n(t)) dt$$

so by Cauchy-Schwarz

$$\mathrm{Var} f \leq \int_0^1 \left(\mathbb{E}\left(\|\nabla f\|^2 (g_1, \ldots, g_n)\right)\right)^{1/2} \left(\mathbb{E}\left(\|\nabla f\|^2 (g_1(t), \ldots, g_n(t))\right)\right)^{1/2} dt$$

and since $g_1(t), \ldots, g_n(t)$ are i.i.d. standard Gaussians, the conclusion follows. □

Proof 2. Using theorem 1 (Efron-Stein), we can reduce it to the one-dimensional case. Indeed, suppose we have proven the one-dimensional result. Then we can write by Efron-Stein

$$\mathrm{Var} f(g_1,\ldots,g_n) \le \sum_{i=1}^{n} \mathbb{E}\mathrm{Var}_{g_i} f(g_1,\ldots,g_n) \le \sum_{i=1}^{n} \mathbb{E}\mathbb{E}_{g_i} |\partial_{x_i} f(g_1,\ldots,g_n)|^2 = \mathbb{E}\left(\|\nabla f\|^2 (g_1,\ldots,g_n)\right)$$

The advantage of one-dimension is that we have the central limit theorem (CLT) at our disposal. Let $f:\mathbb{R}\to\mathbb{R}$ be a C^1 function and let g be a standard Gaussian. Let $Z_1,\ldots,Z_k$ be independent Rademacher random variables. We then have[1]

$$\begin{aligned}
\mathrm{Var} f(g) &\simeq \mathrm{Var} f\left(\frac{Z_1+\ldots+Z_k}{\sqrt{k}}\right) \text{ by CLT}\\
&\le \sum_{i=1}^{k} \mathbb{E}\mathrm{Var}_i f\left(\frac{Z_1+\ldots+Z_k}{\sqrt{k}}\right) \text{ by Efron-Stein}\\
&= \frac{1}{2}\sum_{i=1}^{k} \mathbb{E}\left(f\left(\frac{1}{\sqrt{k}}\sum_{j\neq i} Z_j + \frac{1}{\sqrt{k}}\right) - f\left(\frac{1}{\sqrt{k}}\sum_{j\neq i} Z_j + \frac{-1}{\sqrt{k}}\right)\right)^2\\
&\simeq \frac{1}{2}\sum_{i=1}^{k} \mathbb{E}\left(\frac{2}{\sqrt{k}} f'\left(\frac{1}{\sqrt{k}}\sum_{j\neq i} Z_j + \frac{-1}{\sqrt{k}}\right)\right)^2 \text{ by Taylor series approximation}\\
&\simeq \frac{1}{2}\sum_{i=1}^{k} \mathbb{E}\left(\frac{2}{\sqrt{k}} f'\left(\frac{1}{\sqrt{k}}\sum_{j\neq i} Z_j\right)\right)^2 \text{ by Taylor series approximation}\\
&\simeq \frac{1}{2}\sum_{i=1}^{k} \frac{4}{k}\mathbb{E}\left(f'(g)^2\right) = 2\mathbb{E}\left(f'(g)\right)^2 \text{ by CLT}
\end{aligned}$$

□

In particular, if the functions is Lipschitz, we obtain the following simple bound:

Corollary 3.1. Let $f:\mathbb{R}^n\to\mathbb{R}$ be a Lipschitz function with

$$|f(x)-f(y)| \le \sigma \|x-y\|$$

If $g_1,\ldots,g_n$ are i.i.d. standard Gaussians, then

$$\mathrm{Var} f(g_1,\ldots,g_n) \le \sigma^2$$

As a consequence, we can bound the variance of the largest eigenvalue of matrices with independent Gaussian entries:

[1] There are some caveats for the one-dimensional proof. We use a Berry-Esséen bound: if $h:\mathbb{R}\to\mathbb{R}$ has bounded derivatives up to third order and $Z_1,\ldots,Z_k$ are i.i.d. Rademacher random variables, then

$$\left|\mathbb{E}h(g) - \mathbb{E}h\left(\frac{Z_1+\ldots+Z_k}{\sqrt{k}}\right)\right| \le C\frac{1}{\sqrt{k}}\sup_{x\in\mathbb{R}}|h'''(x)|$$

We also use that $|h(x_0+\varepsilon)-h(x_0)| \le \varepsilon \sup_{x\in\mathbb{R}}|h'(x)|$, $|h'(x_0+\varepsilon)-h'(x_0)| \le \varepsilon\sup_{x\in\mathbb{R}}|h''(x)|$. In all of these we need more information on the function f from the statement of the theorem than it just being C^1. These issues can be overcome by approximating f by smoother functions which have bounded derivatives.

Theorem 4. Let W be a $n \times n$ symmetric random matrix with the entries on and above the diagonal independent, $W_{ij} \simeq \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$. Then

$$\text{Var } \lambda_1(W) \leq 2 \max_{i,j} \sigma_{ij}^2$$

where $\lambda_1(W)$ is the largest eigenvalue of W. In fact, if $\lambda_1(W) \geq \lambda_2(W) \geq \ldots \geq \lambda_n(W)$ are all the eigenvalues, then

$$\text{Var } \lambda_i(W) \leq 2 \max_{i,j} \sigma_{ij}^2$$

Proof. Let $W = (\mu_{ij} + \sigma_{ij} x_{ij})_{1 \leq i,j \leq n}$ and $W' = (\mu_{ij} + \sigma_{ij} y_{ij})_{1 \leq i,j \leq n}$ with $x_{ij} = x_{ji},\ y_{ij} = y_{ji}$. By Weyl's inequality[2] and Cauchy-Schwarz we have

$$\begin{aligned}
|\lambda_i(W) - \lambda_i(W')| &\leq \left\| W - W' \right\| = \sup_{\|v\|=1} \left| \langle v, (W - W')v \rangle \right| \\
&\leq \sup_{\|v\|=1} \sum_{i,j} |\sigma_{ij}| \cdot |x_{ij} - y_{ij}| \cdot |v_i| \cdot |v_j| \\
&\leq \sup_{\|v\|=1} \left(\sum_{i,j} |\sigma_{ij}|^2 |x_{ij} - y_{ij}|^2 \right)^{1/2} \cdot \left(\sum_{i,j} |v_i|^2 |v_j|^2 \right)^{1/2} \\
&\leq \max |\sigma_{ij}| \cdot \left(\sum_{i,j} |x_{ij} - y_{ij}|^2 \right)^{1/2} \\
&\leq \sqrt{2} \max |\sigma_{ij}| \cdot \left(\sum_{i \geq j} |x_{ij} - y_{ij}|^2 \right)^{1/2}
\end{aligned}$$

so the Lipschitz constant of the function $(x_{ij})_{i \geq j} \mapsto \lambda_i(W)$ is at most $\sqrt{2} \max |\sigma_{ij}|$. By corollary 3.1 we obtain

$$\text{Var } \lambda_i(W) \leq 2 \max_{i,j} \sigma_{ij}^2$$

□

We saw that when evaluated on standard Gaussians, the variance of a Lipschitz function with Lipschitz constant σ is at most σ^2. We can in fact show that the deviation from the mean is subgaussian:

Theorem 5 (Gaussian concentration)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschiz function with

$$|f(x) - f(y)| \leq \sigma \|x - y\|$$

Then

$$\mathbb{P}(|f(g_1, \ldots, g_n) - \mathbb{E} f(g_1, \ldots, g_n)| \geq t) \leq 2e^{-t^2/(2\sigma^2)}$$

[2]For λ_1 we don't actually need Weyl's inequality and we can use the trick we have seen before. If v is a unit eigenvector corresponding to $\lambda_1(W)$, then

$$\lambda_1(W) - \lambda_1(W') = v^T W v - \max_{\|u\|=1} u^T W' u \leq v^T W v - v^T W' v = v^T (W - W') v \leq \|W - W'\|$$

and similarly for the bound on $\lambda_1(W') - \lambda_1(W)$.

Proof. Note that it suffices to show

$$\mathbb{P}(f(g_1, \ldots, g_n) - \mathbb{E}f(g_1, \ldots, g_n) \geq t) \leq e^{-t^2/(2\sigma^2)}$$

since we can then work with $-f$ as well to obtain the desired inequality.

We use the standard exponentiation trick and Markov's inequality:

$$\mathbb{P}(f(g_1, \ldots, g_n) - \mathbb{E}f(g_1, \ldots, g_n) \geq t) = \mathbb{P}(e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))} \geq e^{\lambda t}) \leq e^{-\lambda t}\mathbb{E}e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}$$

If we let $h(\lambda) = \mathbb{E}e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}$, then

$$h'(\lambda) = \mathbb{E}\left(e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}f(g_1, \ldots, g_n)\right) - \mathbb{E}\left(e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}\right)\mathbb{E}f(g_1, \ldots, g_n)$$

so $h'(\lambda) = \mathrm{Cov}(e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}, f(g_1, \ldots, g_n))$. Using the covariance formula in the appendix

$$\begin{aligned}
h'(\lambda) &= \int_0^1 \mathbb{E}\nabla\left(e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}\right) \cdot \nabla f(g_1(t), \ldots, g_n(t))dt \\
&= \int_0^1 \mathbb{E}\lambda e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}\nabla f(g_1, \ldots, g_n) \cdot \nabla f(g_1(t), \ldots, g_n(t))dt \\
&\leq \int_0^1 \mathbb{E}\lambda e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}\left|\nabla f(g_1, \ldots, g_n) \cdot \nabla f(g_1(t), \ldots, g_n(t))\right| dt \\
&\leq \int_0^1 \mathbb{E}\lambda e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))}\left\|\nabla f(g_1, \ldots, g_n)\right\| \cdot \left\|\nabla f(g_1(t), \ldots, g_n(t))\right\| dt \\
&\leq \lambda\sigma^2\mathbb{E}e^{\lambda(f(g_1,\ldots,g_n)-\mathbb{E}f(g_1,\ldots,g_n))} \\
&= \lambda\sigma^2 h(\lambda)
\end{aligned}$$

where we used that $\|\nabla f\| \leq \sigma$ since f is σ-Lipschitz. From $h'(\lambda) \leq \lambda\sigma^2 h(\lambda)$ and $h(0) = 1$ it follows that

$$h(\lambda) \leq e^{\lambda^2\sigma^2/2}$$

Therefore

$$\mathbb{P}(f(g_1, \ldots, g_n) - \mathbb{E}f(g_1, \ldots, g_n) \geq t) \leq e^{-\lambda t}e^{\lambda^2\sigma^2/2}$$

Choosing $\lambda = \dfrac{t}{\sigma^2}$ gives

$$\mathbb{P}(f(g_1, \ldots, g_n) - \mathbb{E}f(g_1, \ldots, g_n) \geq t) \leq e^{-t^2/(2\sigma^2)}$$

□

We can now easily see that for random Gaussian matrices, the deviation of the largest eigenvalue from its mean and the deviation of the operator norm from its mean are sub-gaussian:

Corollary 5.1. Let $A_1, \ldots, A_k$ be $n \times n$ symmetric matrices, $g_1, \ldots, g_k$ are i.i.d. standard Gaussians and let

$$X = \sum_{i=1}^{k} g_i A_i$$

Then $\lambda_1(X) - \mathbb{E}\lambda_1(X)$ and $\|X\| - \mathbb{E}\|X\|$ are sub-gaussian with sub-gaussian norm at most $\sigma_*(X)$, where

$$\sigma_*(X) = \sup_{\|v\|=1}\left(\sum_{i=1}^{n}(v^T A_i v)^2\right)^{1/2}$$

In other words,

$$\mathbb{P}(|\lambda_1(X) - \mathbb{E}\lambda_1(X)| \geq t) \leq 2e^{-t^2/2\sigma_*(X)^2}$$

$$\mathbb{P}(|\, \|X\| - \mathbb{E}\,\|X\|\, | \geq t) \leq 2e^{-t^2/2\sigma_*(X)^2}$$

Proof. We will show that the Lipschitz constants of $\lambda_1(X) - \mathbb{E}\lambda_1(X)$ and $\|X\| - \mathbb{E}\,\|X\|$ are at most $\sigma_*(X)$. The conclusion follows by the Gaussian concentration result in theorem 5. Note that constants do not affect the Lipschitz norm, so it's enough to compute the Lipschitz constants of $\lambda_1(X)$ and $\|X\|$.

Let $X = \sum_{i=1}^{k} x_i A_i$ and $Y = \sum_{i=1}^{k} y_i A_i$. Note that

$$|\lambda_1(X) - \lambda_1(Y)| \leq \|X - Y\|$$

$$|\, \|X\| - \|Y\|\, | \leq \|X - Y\|$$

The first inequality follows as in the proof of theorem 4 and the second inequality is just the triangle inequality. We now compute like in the proof of theorem 4

$$\begin{aligned}
\|X - Y\| &= \sup_{\|v\|=1} |\langle v, (X-Y)v\rangle| \\
&= \sup_{\|v\|=1} \left| \sum_{i=1}^{k} (x_i - y_i)\langle v, A_i v\rangle \right| \\
&\leq \sup_{\|v\|=1} \left(\sum_{i=1}^{k} (x_i - y_i)^2 \right)^{1/2} \cdot \left(\sum_{i=1}^{k} \langle v, A_i v\rangle^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^{k} (x_i - y_i)^2 \right)^{1/2} \cdot \sup_{\|v\|=1} \left(\sum_{i=1}^{k} \langle v, A_i v\rangle^2 \right)^{1/2}
\end{aligned}$$

so the Lipschitz norms of the functions

$$(x_1, \ldots, x_k) \mapsto \lambda_1(x_1 A_1 + \ldots + x_k A_k)$$

$$(x_1, \ldots, x_k) \mapsto \|x_1 A_1 + \ldots + x_k A_k\|$$

are at most $\sigma_*(X)$. □

Note that in the second lecture we proved that if $X = \sum_{i=1}^{k} g_i A_i$ is like in Corollary 5.1, then

$$\mathbb{E}\,\|X\| \leq \sigma(X)\sqrt{\log n}$$

where $\sigma(X) = \left\| \sum_{i=1}^{k} A_i^2 \right\|^{1/2} = \left\| \mathbb{E}X^2 \right\|^{1/2}$. One may wonder which of the two parameters, $\sigma(X)$ or $\sigma_*(X)$, is larger. It turns out that it's always the case that $\sigma_*(X) \leq \sigma(X)$:

$$\sigma_*(X)^2 = \sup_{\|v\|=1} \sum_{i=1}^{k} (\langle v, A_i v\rangle)^2 \leq \sup_{\|v\|=1} \sum_{i=1}^{k} (\|v\| \cdot \|A_i v\|)^2 = \sup_{\|v\|=1} \langle v, \sum_{i=1}^{k} A_i^2 v\rangle = \sigma(X)^2$$

3 Talagrand's inequality

We would like to have an analogue of Corollary 3.1 for random variables which are bounded, but it turns out that such a result is not true if we just replace $g_1, \ldots, g_n$ with $X_1, \ldots, X_n$ bounded independent random variables. However, if we impose that the function is also convex, then the following result is true:

Theorem 6 (Talagrand's concentration inequality)**.** Let $X_1, \ldots, X_n$ be independent random variables such that $|X_i| \leq K$ almost surely. If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz

$$|f(x) - f(y)| \leq \sigma \|x - y\|$$

and moreover f *is convex*, then

$$\mathbb{P}(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq t) \leq 2e^{-t^2/(2\sigma^2 K^2)}$$

We will prove this result in the next lecture. To see that convexity is necessary, we consider the following example due to Paata Ivanishvili. Consider the set

$$A = \left\{ (x_1, \ldots, x_n) \in \{0,1\}^n | \sum_{i=1}^{n} x_i \leq \frac{n}{2} \right\}$$

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \inf_{y \in A} \|x - y\|$$

Note that f is 1-Lipschitz (by the triangle inequality), but it is *not* convex.

Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{P}(X_i = 0) = \frac{1}{2}$, $\mathbb{P}(X_i = 1) = \frac{1}{2}$. If Talagrand's concentration inequality was true, then we would have

$$\mathbb{P}(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq n^{1/4}) \leq 2e^{-\sqrt{n}}$$

In particular,

$$\lim_{n \to \infty} \mathbb{P}(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq n^{1/4}) = 0$$

But notice that

$$f(X_1, \ldots, X_n) = \sqrt{\max\left(X_1 + \ldots + X_n - \frac{n}{2}, 0\right)}$$

because $X_i \in \{0, 1\}$ so

$$\begin{aligned}
&\mathbb{P}(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq n^{1/4}) \\
&= \mathbb{P}\left(\left| \sqrt{\max\left(X_1 + \ldots + X_n - \frac{n}{2}, 0\right)} - \mathbb{E}f\left(\sqrt{\max\left(X_1 + \ldots + X_n - \frac{n}{2}, 0\right)} \right) \right| \geq n^{1/4} \right) \\
&= \mathbb{P}\left(\left| \sqrt{\max\left(\frac{\frac{2X_1 - 1}{2} + \ldots + \frac{2X_n - 1}{2}}{\sqrt{n}}, 0 \right)} - \mathbb{E}f\left(\sqrt{\max\left(\frac{\frac{2X_1 - 1}{2} + \ldots + \frac{2X_n - 1}{2}}{\sqrt{n}}, 0 \right)} \right) \right| \geq 1 \right)
\end{aligned}$$

which converges as $n \to \infty$ by the central limit theorem to

$$\mathbb{P}(|\sqrt{\max(g, 0)} - \mathbb{E}\sqrt{\max(g, 0)}| \geq 1)$$

where g is a standard Gaussian.[3] As the above quantity is nonzero, we obtain a contradiction.

[3]Note that

$$\mathbb{E}\sqrt{\max(g, 0)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\max(x, 0)} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}}$$

4 Appendix: The Gaussian covariance formula

Lemma 7. Given $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ two C^1 functions and $g_1, \ldots, g_n$ i.i.d. standard Gaussians, we have

$$\mathrm{Cov}(f_1(g_1, \ldots, g_n), f_2(g_1, \ldots, g_n)) = \int_0^1 \mathbb{E}\nabla f_1(g_1, \ldots, g_n) \cdot \nabla f_2(g_1(t), \ldots, g_n(t))dt$$

where $g_i(t) = tg_i + \sqrt{1-t^2}g_i'$ and $g_1', \ldots, g_n'$ are independent copies of $g_1, \ldots, g_n$.

Proof. Let $E(t) = \mathbb{E}(f_1(g_1, \ldots, g_n)f_2(g_1(t), \ldots, g_n(t)))$. Note that

$$\mathrm{Cov}(f_1(g_1, \ldots, g_n), f_2(g_1, \ldots, g_n)) = E(1) - E(0) = \int_0^1 E'(t)dt$$

We compute using Gaussian integration by parts

$$\begin{aligned}
E'(t) &= \frac{d}{dt}\mathbb{E}f_1(g_1, \ldots, g_n)f_2(g_1(t), \ldots, g_n(t)) \\
&= \sum_{i=1}^n \mathbb{E}f_1(g_1, \ldots, g_n)(\partial_i f_2)(g_1(t), \ldots, g_n(t))g_i'(t) \\
&= \sum_{i=1}^n \mathbb{E}f_1(g_1, \ldots, g_n)(\partial_i f_2)(g_1(t), \ldots, g_n(t))\left(g_i - \frac{t}{\sqrt{1-t^2}}g_i'\right) \\
&= \sum_{i=1}^n \left(\mathbb{E}(\partial_i f_1)(g_1, \ldots, g_n)(\partial_i f_2)(g_1(t), \ldots, g_n(t)) + \sum_{j=1}^n \mathbb{E}f(g_1, \ldots, g_n)(\partial_i\partial_j f_2)(g_1(t), \ldots, g_n(t))t\right) \\
&\quad - \frac{t}{\sqrt{1-t^2}}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}f_1(g_1, \ldots, g_n)(\partial_i\partial_j f_2)(g_1(t), \ldots, g_n(t))\sqrt{1-t^2} \\
&= \sum_{i=1}^n \mathbb{E}(\partial_i f_1)(g_1, \ldots, g_n)(\partial_i f_2)(g_1(t), \ldots, g_n(t)) = \mathbb{E}\nabla f_1(g_1, \ldots, g_n) \cdot \nabla f_2(g_1(t), \ldots, g_n(t))
\end{aligned}$$

where we used that $\frac{\partial}{\partial g_i}g_i(t) = t$, $\frac{\partial}{\partial g_i'}g_i(t) = \sqrt{1-t^2}$. □