CSE 525: Randomized Algorithms | **Spring 2026**

Lecture 19: Chaining for Norms

Lecturer: Shayan Oveis Gharan | *06/04/25*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

The content of these notes are based on https://homes.cs.washington.edu/~jrl/cse599wi23/notes/lec4.html.

Entropy-number convention. For a metric space (S, ρ), write $e_h(S, \rho)$ for the smallest radius r such that S is coverable by at most 2^{2^h} balls of radius r in metric ρ.

19.1 Norms and the main estimate

Definition 19.1 (Norms and seminorms)**.** *A map $N : \mathbb{R}^n \to \mathbb{R}_+$ is a norm when, for all $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,*

1. $N(\lambda x) = |\lambda|\, N(x)$,
2. $N(x + y) \leq N(x) + N(y)$,
3. $N(x) = 0$ *if and only if* $x = 0$.

When only the first two properties are used, N is a seminorm. The arguments below use the word "norm" in this broad sense.

Let $N_1, \ldots, N_m$ be norms on $\mathbb{R}^n$, and let

$$T \subseteq B_2^n, \qquad B_2^n := \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}.$$

For a standard Gaussian $g \sim N(0, I_n)$, define

$$\kappa := \mathbb{E} \max_{k=1,\ldots,m} N_k(g).$$

Theorem 19.2. *If $\epsilon_1, \ldots, \epsilon_m$ are independent random signs, then*

$$\mathbb{E} \max_{x \in T} \sum_{k=1}^{m} \epsilon_k N_k(x)^2 \lesssim \kappa \log(n) \max_{x \in T} \sqrt{\sum_{k=1}^{m} N_k(x)^2}. \tag{19.1}$$

19.1.1 Example: sums of random matrices

Let

$$A = \sum_{k=1}^{m} \epsilon_k A_k^T A_k,$$

with each $A_k^T A_k$ positive semidefinite. Then

$$\begin{aligned}\|A\|_{\mathrm{op}} &= \max_{\|x\|_2 \le 1} \langle x, Ax\rangle \\ &= \max_{\|x\|_2 \le 1} \sum_{k=1}^m \epsilon_k \langle x, A_k^T A_k x\rangle \\ &= \max_{\|x\|_2 \le 1} \sum_{k=1}^m \epsilon_k \|A_k x\|_2^2 .\end{aligned}$$

This is the preceding setting with $T = B_2^n$ and $N_k(x) = \|A_k x\|_2$.

Translator note. The source text appears to phrase the final identification in squared form. The normalization above is the one for which $\sum_k \epsilon_k N_k(x)^2 = \sum_k \epsilon_k \|A_k x\|_2^2$.

19.2 Dudley's inequality and metric reduction

The process

$$\left\{ \sum_{k=1}^m \epsilon_k N_k(x)^2 : x \in T \right\}$$

is subgaussian with respect to

$$d(x,y) := \left(\sum_{k=1}^m \big(N_k(x)^2 - N_k(y)^2\big)^2 \right)^{1/2} .$$

Dudley's entropy inequality therefore gives

$$\mathbb{E} \max_{x \in T} \sum_{k=1}^m \epsilon_k N_k(x)^2 \lesssim \sum_{h \ge 0} 2^{h/2} e_h(T, d). \tag{19.2}$$

Both sides of (19.1) are homogeneous of degree two in the family $(N_k)_{k=1}^m$. Thus one may rescale and assume

$$\max_{x \in T} \sqrt{\sum_{k=1}^m N_k(x)^2} = 1. \tag{19.3}$$

Define

$$\|x\|_N := \max_{k=1,\dots,m} N_k(x).$$

For $x, y \in T$, use $a^2 - b^2 = (a-b)(a+b)$ to obtain

$$\begin{aligned}
d(x,y) &= \left(\sum_{k=1}^{m} \big(N_k(x) - N_k(y)\big)^2 \big(N_k(x) + N_k(y)\big)^2\right)^{1/2} \\
&\le \left(\sum_{k=1}^{m} (N_k(x-y))^2 (N_k(x) + N_k(y))^2\right)^{1/2} \\
&\le \left(\sum_{k=1}^{m} \max_k N_k(x-y)^2 \cdot (N_k(x) + N_k(y))^2\right)^{1/2} \\
&= \|x-y\|_N \left(\sum_{k=1}^{m} \big(N_k(x) + N_k(y)\big)^2\right)^{1/2} \\
&\le \|x-y\|_N \cdot \left(\sum_{k=1}^{m} 2N_k(x)^2 + 2N_k(y)^2\right)^{1/2} \\
&= 2\,\|x-y\|_N .
\end{aligned}$$

The first inequality uses $|N_k(x) - N_k(y)| \le N_k(x-y)$, and the last equality follows from (19.3). Consequently, $e_h(T,d) \le 2e_h(T, \|\cdot\|_N)$, and (19.2) implies

$$\mathbb{E} \max_{x \in T} \sum_{k=1}^{m} \epsilon_k N_k(x)^2 \lesssim \sum_{h \ge 0} 2^{h/2} e_h(T, \|\cdot\|_N). \tag{19.4}$$

We now split the right-hand side into the ranges $h \le 4\log n$ and $h > 4 \log n$.

19.3 The large-entropy tail

Let

$$B_N := \{x \in \mathbb{R}^n : \|x\|_N \le 1\}.$$

For $x \in T$, the normalization (19.3) gives

$$\|x\|_N \le \sqrt{\sum_{k=1}^{m} N_k(x)^2} \le 1,$$

so $T \subseteq B_N$, hence $e_h(T, \|\cdot\|_N) \le e_h(B_N, \|\cdot\|_N)$.

Claim 19.3. *For any norm on $\mathbb{R}^n$, and any $h \ge 1$,*

$$e_h(B_N, \|\cdot\|_N) \le 4 \cdot 2^{-2^h/n}.$$

Proof. Fix $\delta \in (0,1)$, and choose a maximal collection $x_1, \dots, x_s \in B_N$ with pairwise distances at least 2δ in $\|\cdot\|_N$. Maximality gives the cover

$$B_N \subseteq \bigcup_{j=1}^{s} (x_j + 2\delta B_N).$$

The sets $x_j + \delta B_N$ are pairwise disjoint and contained in $2B_N$, so

$$\mathrm{vol}_n(2B_N) \ge s\ \mathrm{vol}_n(\delta B_N) = s\,(\delta/2)^n\, \mathrm{vol}_n(2B_N).$$

Therefore $s \le (2/\delta)^n$. Taking $\delta = 2 \cdot 2^{-2^h/n}$ yields $s \le 2^{2^h}$ and gives a cover of B_N by at most 2^{2^h} balls of radius $2\delta = 4 \cdot 2^{-2^h/n}$. ☐

Using Claim 19.3, the large-h part of Equation 19.4 obeys

$$\sum_{h>4\log n} 2^{h/2} e_h(T, \|\cdot\|_N) \le 4 \sum_{h>4\log n} 2^{h/2} 2^{-2^h/n} \le O(1).$$

Thus

$$\mathbb{E} \max_{x\in T} \sum_{k=1}^m \epsilon_k N_k(x)^2 \lesssim O(1) + \sum_{0\le h\le 4\log n} 2^{h/2} e_h(T, \|\cdot\|_N).$$

19.4 The relevant entropy range and dual Sudakov

Since $T \subseteq B_2^n$,

$$e_h(T, \|\cdot\|_N) \le e_h(B_2^n, \|\cdot\|_N).$$

The required ingredient is the following dual Sudakov bound.

Lemma 19.4 (Dual Sudakov)**.** *For any norm $\|\cdot\|$ on $\mathbb{R}^n$ and every $h \ge 0$,*

$$e_h(B_2^n, \|\cdot\|) \lesssim 2^{-h/2} \mathbb{E} \|g\|,$$

where $g \sim N(0, I_n)$.

Applying Lemma 19.4 with $\|\cdot\| = \|\cdot\|_N$ gives

$$\mathbb{E} \|g\|_N = \mathbb{E} \max_{k=1,\ldots,m} N_k(g) = \kappa.$$

Therefore

$$\sum_{0\le h\le 4\log n} 2^{h/2} e_h(T, \|\cdot\|_N) \lesssim \kappa \log n.$$

Under the normalization (19.3), this proves (19.1); undoing the rescaling gives the stated form.

19.5 Gaussian shift lemma

Lemma 19.5 (Gaussian shift)**.** *Let $K \subseteq \mathbb{R}^n$ be symmetric and convex, and let γ_n denote standard Gaussian measure on $\mathbb{R}^n$. For every $x \in \mathbb{R}^n$,*

$$\gamma_n(K + x) \ge \exp\left(-\frac{\|x\|_2^2}{2}\right) \gamma_n(K).$$

Proof. Using symmetry of K and writing σ for a uniform random sign,

$$\begin{aligned}\gamma_n(K + x) &= (2\pi)^{-n/2} \int_K \exp\left(-\frac{\|x+z\|_2^2}{2}\right) dz \\ &= (2\pi)^{-n/2} \int_K \mathbb{E}_{\sigma\in\{-1,1\}} \exp\left(-\frac{\|\sigma x+z\|_2^2}{2}\right) dz.\end{aligned}$$

Since

$$\mathbb{E}_\sigma \|\sigma x + z\|_2^2 = \|x\|_2^2 + \|z\|_2^2,$$

Jensen's inequality yields

$$\begin{aligned}\gamma_n(K+x) &\geq (2\pi)^{-n/2} \int_K \exp\left(-\frac{\mathbb{E}_\sigma \|\sigma x + z\|_2^2}{2}\right) dz \\ &= (2\pi)^{-n/2} \int_K \exp\left(-\frac{\|x\|_2^2 + \|z\|_2^2}{2}\right) dz \\ &= \exp\left(-\frac{\|x\|_2^2}{2}\right) \gamma_n(K).\end{aligned}$$

□

Translator note. The displayed conclusion above matches the Gaussian-shift bound used in the proof of dual Sudakov; constants are immaterial for the subsequent $\lesssim$ estimate.

19.6 Proof of the dual Sudakov lemma

Let

$$\mathcal{B} := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

be the unit ball of the norm $\|\cdot\|$. Choose $x_1, \ldots, x_s \in B_2^n$ maximally so that the translated sets $x_j + \delta\mathcal{B}$ are pairwise disjoint. Then

$$B_2^n \subseteq \bigcup_{j=1}^{s} (x_j + 2\delta\mathcal{B}), \tag{19.5}$$

so B_2^n is covered by s balls of radius 2δ in the norm $\|\cdot\|$.

For any $\lambda > 0$, the scaled sets $\lambda(x_j + \delta\mathcal{B})$ are pairwise disjoint. Therefore

$$\begin{aligned}1 &\geq \gamma_n\left(\bigcup_{j=1}^{s} \lambda(x_j + \delta\mathcal{B})\right) \\ &= \sum_{j=1}^{s} \gamma_n\big(\lambda x_j + \lambda\delta\mathcal{B}\big) \\ &\geq \sum_{j=1}^{s} \exp\left(-\frac{\lambda^2 \|x_j\|_2^2}{2}\right) \gamma_n(\lambda\delta\mathcal{B}) \\ &\geq s \exp\left(-\frac{\lambda^2}{2}\right) \gamma_n(\lambda\delta\mathcal{B}),\end{aligned}$$

where Lemma 19.5 is used in the third line and $x_j \in B_2^n$ in the final line.

Choose

$$\lambda := \frac{2}{\delta} \mathbb{E} \|g\|.$$

Then, by Markov's inequality,

$$\gamma_n(\lambda\delta\mathcal{B}) = \mathbb{P}\big(\|g\| \leq \lambda\delta\big) = \mathbb{P}\big(\|g\| \leq 2\mathbb{E}\,\|g\|\big) \geq \frac{1}{2}.$$

Combining the previous inequalities gives

$$1 \geq \frac{s}{2}\exp\left(-\frac{1}{2}\left(\frac{2\mathbb{E}\,\|g\|}{\delta}\right)^2\right).$$

Equivalently, up to universal constants,

$$\delta \lesssim \frac{\mathbb{E}\,\|g\|}{\sqrt{\log(s/2)}}.$$

With $s = 2^{2^h}$, the cover in (19.5) has radius $2\delta \lesssim 2^{-h/2}\mathbb{E}\,\|g\|$. Hence

$$e_h(B_2^n, \|\cdot\|) \lesssim 2^{-h/2}\mathbb{E}\,\|g\|,$$

which proves Lemma 19.4.