Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 7**

Lipschitz Functions of Gaussian Variables and the Entropy Method

Lecturer: Yun Wei *Scribe: Mengfan Fu*

1 Lipschitz Functions of Gaussian Variables

We first show that Lipschitz functions of a standard Gaussian vector are sub-Gaussian.

Definition 1.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *L-Lipschitz* with respect to the Euclidean norm $\|\cdot\|_2$ if

$$|f(x) - f(y)| \le L \|x - y\|_2 \qquad \text{for all } x, y \in \mathbb{R}^n. \tag{1}$$

Theorem 1.2 ([1, Theorem 2.26])**.** *Let $X = (X_1, \dots, X_n) \sim N(0, I_n)$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz with respect to $\|\cdot\|_2$. Then $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian with parameter at most L, and hence*

$$\mathbb{P}\left(|f(X) - \mathbb{E}[f(X)]| \ge t\right) \le 2e^{-\frac{t^2}{2L^2}} \qquad \textit{for all } t \ge 0. \tag{2}$$

Remark 1.3. The sub-Gaussian parameter in Theorem 1.2 depends on f only through L and does not depend on the dimension n. Thus an L-Lipschitz function of an n-dimensional standard Gaussian vector obeys the same tail bound as a scalar $N(0, L^2)$ variable, however large n is. The tail bound (2) follows from the sub-Gaussian property by applying the Chernoff bound from Lecture 1 to $f(X)$ and to $-f(X)$.

We do not prove Theorem 1.2 here; a proof of the weaker statement with sub-Gaussian parameter $\pi L/2$ can be found in [1, Section 2.3]. We now consider three applications.

Example 1.4 (Order statistics, [1, Example 2.29])**.** Given a random vector $(X_1, \dots, X_n)$, its order statistics are obtained by reordering its entries in non-decreasing order,

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

In particular, $X_{(n)} = \max_{1 \le i \le n} X_i$ and $X_{(1)} = \min_{1 \le i \le n} X_i$. For $k \in \{1, \dots, n\}$, let $f_k : \mathbb{R}^n \to \mathbb{R}$ denote the map $x \mapsto x_{(k)}$, so that $X_{(k)} = f_k(X_1, \dots, X_n)$. The Lipschitz constant of f_k is controlled by the following lemma.

Lemma 1.5. *For all $x, y \in \mathbb{R}^n$ and $k \in \{1, \dots, n\}$,*

$$\left|x_{(k)} - y_{(k)}\right| \le \|x - y\|_\infty .$$

The proof of Lemma 1.5 is left as Exercise 1. Since $\|x - y\|_\infty = \max_{1 \le i \le n} |x_i - y_i| \le \|x - y\|_2$, the lemma gives

$$|f_k(x) - f_k(y)| \le \|x - y\|_2 \qquad \text{for all } x, y \in \mathbb{R}^n,$$

so each f_k is 1-Lipschitz. Consequently, if $(X_1, \dots, X_n) \sim N(0, I_n)$, Theorem 1.2 with $L = 1$ yields, for every $k = 1, \dots, n$,

$$\mathbb{P}\left(\left|X_{(k)} - \mathbb{E}[X_{(k)}]\right| \ge t\right) \le 2e^{-\frac{t^2}{2}} \qquad \text{for all } t \ge 0. \tag{3}$$

In particular, the fluctuations of the maximum $X_{(n)}$ around its mean are of order one, uniformly in n.

Example 1.6 (Gaussian complexity, [1, Example 2.30])**.** Let $A \subseteq \mathbb{R}^n$ be a bounded set and $W \sim N(0, I_n)$, and define

$$Z(A) := \sup_{a \in A} \langle a, W \rangle .$$

The expectation $\mathbb{E}[Z(A)]$ is called the *Gaussian complexity* of A. It is the Gaussian counterpart of the Rademacher complexity from Lecture 6, with the Rademacher vector replaced by W.

To show that $Z(A)$ concentrates, write $Z(A) = f(W)$ with $f(w) := \sup_{a \in A} \langle a, w \rangle$ for $w \in \mathbb{R}^n$, and let $D(A) := \sup_{a \in A} \|a\|_2 < \infty$. For any $w, w' \in \mathbb{R}^n$,

$$\begin{aligned} |f(w) - f(w')| &= \left| \sup_{a \in A} \langle a, w \rangle - \sup_{a \in A} \langle a, w' \rangle \right| \le \sup_{a \in A} |\langle a, w \rangle - \langle a, w' \rangle| \\ &= \sup_{a \in A} |\langle a, w - w' \rangle| \le \sup_{a \in A} \|a\|_2 \|w - w'\|_2 = D(A) \|w - w'\|_2 . \end{aligned}$$

The second inequality is the Cauchy–Schwarz inequality. Hence f is $D(A)$-Lipschitz, and Theorem 1.2 yields

$$\mathbb{P}\left(|Z(A) - \mathbb{E}[Z(A)]| \ge t \right) \le 2 \exp\left(-\frac{t^2}{2D(A)^2} \right) \qquad \text{for all } t \ge 0. \tag{4}$$

Example 1.7 (Singular values of Gaussian random matrices, [1, Example 2.32])**.** Let $n > d$ and let $X \in \mathbb{R}^{n \times d}$ be a random matrix with i.i.d. $N(0,1)$ entries and ordered singular values $\sigma_1(X) \ge \sigma_2(X) \ge \cdots \ge \sigma_d(X) \ge 0$. Writing $x_1^\top, \ldots, x_n^\top$ for the rows of X, we view the matrix as the vector

$$\operatorname{vec}(X) := \left(x_1^\top, x_2^\top, \ldots, x_n^\top \right)^\top \in \mathbb{R}^{nd},$$

so that $\operatorname{vec}(X) \sim N(0, I_{nd})$ and each $\sigma_k(X)$ is a function of a standard Gaussian vector in $\mathbb{R}^{nd}$. For $X, Y \in \mathbb{R}^{n \times d}$ we have $\|\operatorname{vec}(X) - \operatorname{vec}(Y)\|_2 = \|X - Y\|_F$, so it suffices to bound $|\sigma_k(X) - \sigma_k(Y)|$ by $\|X - Y\|_F$.

Recall that for $M \in \mathbb{R}^{n \times d}$ the spectral norm and the Frobenius norm are

$$\|M\|_2 := \sigma_1(M) = \sup_{a \in \mathbb{R}^d \colon \|a\|_2 = 1} \|Ma\|_2 , \qquad \|M\|_F := \Big(\sum_{i=1}^n \sum_{j=1}^d M_{ij}^2 \Big)^{1/2} = \left(\sigma_1^2(M) + \cdots + \sigma_d^2(M) \right)^{1/2},$$

so that $\|M\|_2 \le \|M\|_F$. By Weyl's inequality (see [1, Exercise 8.3]),

$$\max_{1 \le k \le d} |\sigma_k(X) - \sigma_k(Y)| \le \|X - Y\|_2 \le \|X - Y\|_F . \tag{5}$$

Thus each σ_k is 1-Lipschitz as a function of $\operatorname{vec}(X)$, and Theorem 1.2 gives, for each $k = 1, \ldots, d$,

$$\mathbb{P}\left(|\sigma_k(X) - \mathbb{E}[\sigma_k(X)]| \ge t \right) \le 2e^{-\frac{t^2}{2}} \qquad \text{for all } t \ge 0. \tag{6}$$

We have not computed $\mathbb{E}[\sigma_k(X)]$, but (6) shows that each singular value stays within $O(1)$ of its expectation with high probability, regardless of n and d.

2 Concentration by Entropic Techniques

We now introduce the entropy method for deriving concentration inequalities [1, Section 3.1].

Definition 2.1 (ϕ-entropy)**.** Let $X \sim P$, and let ϕ be a convex function on an interval containing the support of X such that X and $\phi(X)$ are integrable. The *ϕ-entropy* of X is

$$H_\phi(X) := \mathbb{E}[\phi(X)] - \phi(\mathbb{E}[X]). \tag{7}$$

Remark 2.2. By Jensen's inequality, $H_\phi(X) \geq 0$. Moreover, $H_\phi(X) = 0$ if X is almost surely constant or if ϕ is affine. Although the notation suggests a function of X, $H_\phi(X)$ is a number that depends only on the distribution P of X; for instance, $H_\phi(e^{\lambda X})$ below is the ϕ-entropy of the distribution of $e^{\lambda X}$.

Example 2.3. Two familiar quantities are ϕ-entropies.

(a) If $\phi(u) = u^2$, then $H_\phi(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathrm{Var}(X)$.

(b) If $\phi(u) = -\log u$ for $u > 0$, then applying ϕ to the positive random variable $e^{\lambda X}$ gives

$$\begin{aligned} H_\phi(e^{\lambda X}) &= \mathbb{E}\left[-\log e^{\lambda X}\right] - \left(-\log \mathbb{E}\left[e^{\lambda X}\right]\right) = -\lambda\mathbb{E}[X] + \log \mathbb{E}\left[e^{\lambda X}\right] \\ &= \log e^{-\lambda \mathbb{E}[X]} + \log \mathbb{E}\left[e^{\lambda X}\right] = \log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right], \end{aligned}$$

which is the cumulant generating function of the centered variable $X - \mathbb{E}[X]$.

From now on we work with the convex function $\phi : [0, \infty) \to \mathbb{R}$ defined by

$$\phi(u) := u \log u \quad \text{for } u > 0, \qquad \phi(0) := 0, \tag{8}$$

where the value at 0 makes ϕ continuous, since $u \log u \to 0$ as $u \to 0^+$. For a non-negative random variable Z we drop the subscript and write

$$H(Z) := H_\phi(Z) = \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z]. \tag{9}$$

Let $\varphi_X(\lambda) := \mathbb{E}\left[e^{\lambda X}\right]$ be the moment generating function of X (not to be confused with the convex function ϕ), and assume that it is finite in a neighborhood of λ, so that $\varphi_X'(\lambda) = \mathbb{E}\left[X e^{\lambda X}\right]$. Taking $Z = e^{\lambda X}$ in (9) and using $\log e^{\lambda X} = \lambda X$, we obtain

$$H(e^{\lambda X}) = \mathbb{E}\left[e^{\lambda X} \cdot \lambda X\right] - \mathbb{E}\left[e^{\lambda X}\right] \log \mathbb{E}\left[e^{\lambda X}\right] = \lambda \varphi_X'(\lambda) - \varphi_X(\lambda) \log \varphi_X(\lambda). \tag{10}$$

Hence $H(e^{\lambda X})$ is determined by the moment generating function of X and its derivative.

Example 2.4 (Entropy of a Gaussian variable, [1, Example 3.1])**.** Let $X \sim N(0, \sigma^2)$. Then $\varphi_X(\lambda) = e^{\lambda^2\sigma^2/2}$ and $\varphi_X'(\lambda) = \lambda\sigma^2 e^{\lambda^2\sigma^2/2}$ for all $\lambda \in \mathbb{R}$, so by (10),

$$H(e^{\lambda X}) = \lambda \cdot \lambda\sigma^2 e^{\frac{\lambda^2\sigma^2}{2}} - e^{\frac{\lambda^2\sigma^2}{2}} \log e^{\frac{\lambda^2\sigma^2}{2}} = \frac{\lambda^2\sigma^2}{2} e^{\frac{\lambda^2\sigma^2}{2}} = \frac{1}{2}\sigma^2\lambda^2\varphi_X(\lambda). \tag{11}$$

Since the moment generating function yields tail bounds through the Chernoff bound, (10) suggests that an upper bound on $H(e^{\lambda X})$ should also lead to a tail bound. The Herbst argument makes this precise for upper bounds of the Gaussian form (11).

Proposition 2.5 (Herbst argument, [1, Proposition 3.2])**.** *Suppose that there is a constant $\sigma > 0$ such that*

$$H(e^{\lambda X}) \leq \frac{1}{2}\sigma^2\lambda^2\varphi_X(\lambda) \qquad \textit{for all } \lambda \in I, \tag{12}$$

where $I = \mathbb{R}$ or $I = [0, \infty)$. Then

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \qquad \textit{for all } \lambda \in I. \tag{13}$$

Remark 2.6. By Example 2.4, every $X \sim N(0, \sigma^2)$ satisfies (12) with equality for all $\lambda \in \mathbb{R}$. When $I = \mathbb{R}$, the conclusion (13) says exactly that X is sub-Gaussian with parameter σ. When $I = [0, \infty)$, it controls the moment generating function only for $\lambda \geq 0$ and so does not by itself give sub-Gaussianity; it does give, through the Chernoff bound from Lecture 1, the one-sided bound

$$\mathbb{P}\left(X - \mathbb{E}[X] \geq t\right) \leq e^{-\frac{t^2}{2\sigma^2}} \qquad \text{for all } t \geq 0.$$

The proof of Proposition 2.5 is deferred; see also [1, Section 3.1.2].

Exercises

1. Prove Lemma 1.5: for all $x, y \in \mathbb{R}^n$ and $k \in \{1, \dots, n\}$, $\left|x_{(k)} - y_{(k)}\right| \leq \|x - y\|_\infty$.

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019.