# Advanced Algorithms VI (Fall 2020)

Instructor: Chihao Zhang
Scribed by: Huaijin Wu, Tong Chen

Last modified on October 25, 2020

In this lecture, we prove the Johnson-Lindenstrauss Lemma, which is an important algorithmic application of concentration inequalities. To this end, we introduce the notion of sub-Gaussian random variables and the Bernstein's inequality.

## 1 Johnson-Lindenstrauss Lemma

Given a collection of points in a high-dimensional space, one might try to map the points into a low-dimensional space without distorting the relative distances between points very much. This problem is called *metric embedding* in literature. The Johnson-Lindenstrauss lemma states that such an embedding exists in certain case.

**Lemma 1** (Johnson-Lindenstrauss Lemma). *Let $S$ be a collection of points set that $S \subseteq \mathbb{R}^D$. Then for every $\varepsilon \in (0,1)$, there exists a projection $f : \mathbb{R}^D \to \mathbb{R}^d$ with $d = O\left(\log(|S|/\varepsilon^2)\right)$ such that $\forall x, y \in S, x \neq y$, it holds that*

$$1 - \varepsilon \le \frac{\|f(x) - f(y)\|}{\|x - y\|} \le 1 + \varepsilon \tag{1}$$

Note that in the statement of JL lemma, the dimension $d$ is irrelevant to $D$. So $D$ can be arbitrary large or even infinite. It is surprising that a *random* linear projection from $\mathbb{R}^D$ to $\mathbb{R}^d$ satisfies the requirement of JL with high probability. To see this, suppose we define a matrix $A = (a_{ij})_{\substack{1 \le i \le D \\ 1 \le j \le d}} \in \mathbb{R}^{D \times d}$ where each $a_{ij}$ is chosen from $\{-1, 1\}$ uniformly at random. Then for any vector $u \in \mathbb{R}^D$, $Au \in \mathbb{R}^d$ satisfies

$$\mathbf{E}\left[\|Au\|^2\right] = \mathbf{E}\left[\sum_{i=1}^{d} (Au)_i^2\right] = \sum_{i=1}^{d} \mathbf{E}\left[\left(\sum_{j=1}^{D} a_{ij} \cdot u_j\right)^2\right] = d \cdot \|u\|^2.$$

Therefore, if we choose $f = \frac{A}{\sqrt{d}}$, then $\mathbf{E}\left[\|f(u)\|^2\right] = \|u\|^2$. This implies that for any $x, y \in S$, $\mathbf{E}\left[\|f(x) - f(y)\|^2\right] = \mathbf{E}\left[\|f(x-y)\|^2\right] = \mathbf{E}\left[\|x - y\|^2\right]$. Hence to establish (1), we only need to prove that $\|f(u)\|^2$ is well-concentrated to its expectation, namely

$$\mathbf{Pr}\left[1 - \varepsilon \le \frac{\|f(u)\|^2}{\|u\|^2} \le 1 + \varepsilon\right] \ge 1 - \delta$$

for appropriate $\delta$. Since $f$ is linear, we can assume without loss of generality that $\|u\| = 1$.

For every $i = 1, \ldots, d$, we let $Z_i = \sum_{j=1}^{D} a_{ij} u_j$. Then $\|f(u)\|^2 = \frac{1}{d} \cdot \sum_{i=1}^{d} Z_i^2$. Now we can express our objective as the sum of $d$ independent random variables, so to obtain a concentration result, we might try to apply Chernoff-typed inequalities. However, it seems that we cannot directly apply the Hoeffding inequality here, since $Z_i^2 = \left(\sum_{j=1}^{D} a_{ij}u_j\right)^2$ can be unbounded. Therefore, we need some new tools to tackle random variables of this form.

## 2 Sub-Gaussian Random Variables

Recall the proof of the Chernoff bounds and the Hoeffding inequality. The key to establish these inequalities is an upper bound on the moment generating functions $\mathbf{E}\left[e^{\alpha X}\right]$. We abstract the property and introduce the notion of *sub-Gaussian* random variables.

**Definition 2.** *A random variables $X$ with $\mathbf{E}[X] = 0$ is called* sub-Gaussian *with variance factor $v$, denoted as $X \in \mathcal{G}(v)$, if*

$$\mathbf{E}\left[e^{\alpha X}\right] \le e^{\frac{\alpha^2}{2}v} \quad \textit{for every } \alpha \in \mathbb{R}.$$

The name *sub-Gaussian* comes from the fact that for a Gaussian random variables $X \sim N(0, v)$, it holds that $\mathbf{E}\left[e^{\alpha X}\right] = e^{\frac{\alpha^2}{2}v}$.

The moment generating function of a random variables $X$ is closely related to its $k$-the moment for all $k \in \mathbb{N}$. The following theorem clarifies the relationship, and interested readers can refer to [1, Chapter 2] for more on this.

**Theorem 3.** *Let $X$ be a random variable with $\mathbf{E}[X] = 0$*

*(1) If $X \in \mathcal{G}(v)$, then for every integer $j \ge 1$, $\mathbf{E}\left[X^{2j}\right] \le 2^{j+1} j! v^j$*

*(2) If for some positive constant $v$ and for every integer $j \ge 1$, $\mathbf{E}\left[X^{2j}\right] \le j! v^j$, then $X \in \mathcal{G}(4v)$*

*Proof.* We first prove (1). We can assume without loss of generality that $X$ is a continuous random variable. Assume $X \in \mathcal{G}(v)$, then we have

$$\begin{aligned}
\mathbf{E}\left[X^{2j}\right] &= \int_0^\infty \mathbf{Pr}\left[|X|^{2j} \ge x\right] \,\mathrm{d}x \\
&= \int_0^\infty \mathbf{Pr}\left[|X| \ge x^{\frac{1}{2j}}\right] \,\mathrm{d}x \\
&= 2j \int_0^\infty \mathbf{Pr}\left[|X| \ge z\right] z^{2j-1} \,\mathrm{d}z \quad (z = x^{\frac{1}{2j}}).
\end{aligned}$$

For any $\alpha > 0$, we have

$$\mathbf{Pr}[X > z] = \mathbf{Pr}\left[e^{\alpha X} > e^{\alpha z}\right] \le \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha z}} \le e^{\frac{\alpha^2}{2}v - \alpha z}.$$

We choose $\alpha = \frac{z}{v}$ and obtain $\mathbf{Pr}[X > z] \le e^{-\frac{z^2}{2v}}$. Similarly, we can obtain $\mathbf{Pr}[X < -z] \le e^{-\frac{z^2}{2v}}$. Then

$$\begin{aligned}
\mathbf{E}\left[X^{2j}\right] &\le 4j \int_0^\infty z^{2j-1} e^{-\frac{z^2}{2v}} dz \\
&= 4j \int_0^\infty (2vt)^{j-\frac{1}{2}} e^{-t} d(2vt)^{\frac{1}{2}} \quad (t = -\frac{z^2}{2v}) \\
&= 2j(2v)^j \int_0^\infty t^{j-1} e^{-t} dt \\
&= 2^{j+1} j! v^j.
\end{aligned}$$

We proceed to prove (2). Assume $\mathbf{E}\left[X^{2j}\right] \le j! v^j$. To get rid of the odd moments of $X$, we introduce an independent random variable $X'$ who follows the same distribution as $X$. Then by symmetry of $X - X'$ we have

$$\mathbf{E}\left[e^{\alpha X}\right]\mathbf{E}\left[e^{-\alpha X'}\right] = \mathbf{E}\left[e^{\alpha X}e^{-\alpha X'}\right] = \mathbf{E}\left[e^{\alpha(X-X')}\right] = \sum_{j=0}^{\infty}\frac{\alpha^j \cdot \mathbf{E}\left[(X-X')^j\right]}{j!}.$$

For odd $j$, we have

$$\begin{aligned}\mathbf{E}\left[(X-X')^j\right] &= \sum_{k=0}^{j}\binom{j}{k}\mathbf{E}\left[X^k\right]\mathbf{E}\left[(-X')^{j-k}\right]\\ &= \sum_{k=0}^{j}(-1)^{j-k}\cdot\binom{j}{k}\mathbf{E}\left[X^k\right]\mathbf{E}\left[X^{j-k}\right]\\ &= \sum_{k=0}^{\lfloor j/2\rfloor}\binom{j}{k}\left(\left((-1)^k + (-1)^{j-k}\right)\mathbf{E}\left[X^k\right]\mathbf{E}\left[X^{j-k}\right]\right)\\ &= 0.\end{aligned}$$

Therefore,

$$\mathbf{E}\left[e^{\alpha X}\right]\mathbf{E}\left[e^{-\alpha X'}\right] = \sum_{j=0}^{\infty}\frac{(\alpha)^{2j}\mathbf{E}\left[(X-X')^{2j}\right]}{(2j)!}. \tag{2}$$

Since the function $X \to X^{2j}$ is convex, Jensen's inequality yields

$$(X+(-X'))^{2j} = 2^{2j}\left(\frac{1}{2}X + \frac{1}{2}(-X')\right)^{2j} \le 2^{2j}\left(\frac{1}{2}X^{2j} + \frac{1}{2}X'^{2j}\right) = 2^{2j-1}(X^{2j} + X'^{2j}).$$

So

$$\mathbf{E}\left[(X-X')^{2j}\right] \le 2^{2j-1}\mathbf{E}\left[X^{2j} + X'^{2j}\right] = 2^{2j}\mathbf{E}\left[X^{2j}\right].$$

Since

$$\frac{(2j)!}{j!} = \prod_{k=1}^{j}(j+k) \ge \prod_{k=1}^{j}2k = 2^j j!,$$

we have

$$(2) = \sum_{j=0}^{\infty}\frac{(\alpha)^{2j}\mathbf{E}\left[(X-X')^{2j}\right]}{(2j)!} \le \sum_{j=0}^{\infty}\frac{(\alpha)^{2j}2^{2j}\mathbf{E}\left[X^{2j}\right]}{(2j)!} \le \sum_{j=0}^{\infty}\frac{(\alpha)^{2j}2^{2j}e^j j!}{(2j)!} \le \sum_{j=0}^{\infty}\frac{\alpha^{2j}v^j 2^j}{j!}.$$

Moreover, again by Jensen's inequality, we know that

$$\mathbf{E}\left[e^{-\alpha X'}\right] \ge e^{-\alpha\mathbf{E}[X']} = 1.$$

So

$$\mathbf{E}\left[e^{\alpha X}\right] \le \sum_{j=0}^{\infty}\frac{\alpha^{2j}v^j 2^j}{j!} = e^{2\alpha^2 v}.$$

That is, $X \in \mathcal{G}(4v)$. □

Given bounds on all moments, we have the following more general concentration inequality, known as *Bernstein's inequality*.

**Theorem 4** (Bernstein's inequality). *Let $X_1, \ldots, X_n$ be independent real-valued random variables. Assume that there exist positive numbers a and b such that*

*(1)* $\sum_{i=1}^{n} \mathbf{E}\left[X_i^2\right] \leq a$

*(2)* $\sum_{i=1}^{n} \mathbf{E}\left[X_i^j\right] \leq \frac{j!}{2} a b^{j-2}$ *for all integers* $q \geq 3$

*Define* $X = \sum_{i=1}^{n}$, $S = X - \mathbf{E}[X]$, *then for all* $t > 0$, *we have*

$$\mathbf{Pr}\left[S \geq \sqrt{2at} + bt\right] \leq e^{-t}.$$

See [1] for a proof of the theorem.

## 3 Proof of Johnson-Lindenstrauss Lemma

We are now ready to prove Lemma 1. Following the discussion in Section 1, for every $i = 1, \ldots, d$, we have

$$\mathbf{E}\left[e^{\alpha Z_i}\right] = \mathbf{E}\left[e^{\alpha \sum_{j=1}^{D} a_{ij} u_j}\right] = \prod_{j=1}^{D} \mathbf{E}\left[e^{\alpha a_{ij} u_j}\right] = \prod_{j=1}^{D} \left(\frac{1}{2}\left(e^{\alpha u_j} + e^{-\alpha u_j}\right)\right).$$

Since

$$\frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \sum_{j=0}^{\infty} \frac{\lambda^{2j}}{(2j)!} \leq \sum_{j=0}^{\infty} \frac{(\lambda)^{2j}}{2^j j!} = e^{\frac{\lambda^2}{2}},$$

we have

$$\mathbf{E}\left[e^{\alpha Z_i}\right] \leq \prod_{j=1}^{D} e^{\frac{\alpha^2 u_j^2}{2}} = e^{\frac{\alpha^2}{2}}.$$

Therefore, $Z_i \in \mathcal{G}(1)$. Let $Y_i = Z_i^2$. According to Theorem 3, we can obtain bounds on mements of $Z_i$ and $Y_i$:

$$\begin{aligned}
\forall j \geq 1 :& \mathbf{E}\left[Z_i^{2j}\right] \leq 2^{j+1} \cdot j!; \\
\forall j \geq 1 :& \mathbf{E}\left[Y_i^{j}\right] \leq 2^{j+1} \cdot j! \leq 4^j \cdot j!; \\
\forall j \geq 1 :& \mathbf{E}\left[Y_i^{2j}\right] \leq 2^{2j+1} \cdot (2j)! \leq 2^{3j+1} \cdot j!.
\end{aligned}$$

Finally we have

$$\mathbf{Pr}\left[\|f(u)\|^2 - 1 > \varepsilon\right] = \mathbf{Pr}\left[\sum_{i=1}^{d} \frac{1}{d} Y_i - 1 > \varepsilon\right] = \mathbf{Pr}\left[\sum_{i=1}^{d} (Y_i - 1) > \varepsilon d\right].$$

We can let $a = 16d$ and $b = 4$ in the Bernstein's inequality (Theorem 4), which gives

$$\mathbf{Pr}\left[\sum_{i=1}^{d} Y_i - 1 \geq 4\sqrt{2dt} + 4t\right] \leq e^{-t}.$$

Applying union bound for every pair of $x, y \in S$, it suffices to let $n^2 \cdot e^{-t} \leq \delta$. So we let $4\sqrt{2dt} + 4t = d\varepsilon$ where $t = \log \frac{n^2}{\delta}$. This requires $d = \Theta\left(\frac{1}{\varepsilon^2} \log \frac{n}{\sqrt{\delta}}\right)$.

## References


[1] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*, 2013. 2, 4