Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

Today we will prove the Hoeffding-Azuma Inequality, which can be used to prove the Johnson-Lindenstrauss Lemma (though we won’t exactly use it today) and comes up in other important topics. We’ll then re-prove the Johnson-Lindenstrauss lemma.

### 10.1 Hoeffding-Azuma Inequality

**Definition 10.1** Let $X$ be a real-valued random variable. Then we define

$$\|X\|_\infty = \inf\{ c \mid \Pr(|X| \leq c) = 1 \}$$

$X$ is “bounded” means $\|X\|_\infty < \infty$

**Theorem 10.2 Hoeffding-Azuma Inequality** If $\{X_1, \ldots, X_n\}$ are bounded random variables and

$$\mathbb{E}[X_{i_1} \cdots X_{i_k}] = 0 \quad \forall k, \ 1 \leq i_1 < \ldots < i_k \leq n$$

then

$$\Pr\left(\sum_{i=1}^{n} X_i \geq L\right) \leq \exp\left(\frac{-L^2}{2 \sum_{i=1}^{n} \|X\|_\infty^2}\right)$$

where $\exp(x) = e^x$.

Note that in some sense, the above theorem is a generalization of the Chernoff bound.

**Proof:** Recall that

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

Since $e^{ax}$ is a convex function, we have for $x \in [-1, 1]$

$$e^{ax} = \exp\left(a \left(\frac{1+x}{2}\right) - a \left(\frac{1-x}{2}\right)\right) \leq \frac{1+x}{2}e^a + \frac{1-x}{2}e^{-a} \leq \cosh a + x \sinh a$$

(10.1)
Also note that if \( a_i, b_i, 1 \leq i \leq n \) are constants, then
\[
E \left[ \prod_{i=1}^{n} (b_i X_i + a_i) \right] = \prod_{i=1}^{n} a_i
\] (10.4)

Because each term of the expansion with at least one \( X_i \) term has expectation 0, by assumption. Let \( X = X_i/\|X_i\|_\infty \) and \( a = t \cdot \|X_i\|_\infty \). (We will choose \( t \) later to be small and non-negative.) From Equations 10.1 and 10.4 we get:
\[
e^{t X_i} = e^{a X} \\
\leq \cosh a + x \sinh a \\
= \cosh(t \|X_i\|_\infty) + \frac{X_i}{\|X_i\|_\infty} \sinh(t \|X_i\|_\infty)
\]

Hence,
\[
E \left[ \exp \left( t \sum_{i=1}^{n} X_i \right) \right] = E \left[ \prod_{i=1}^{n} e^{t X_i} \right] \\
\leq E \prod_{i=1}^{n} \left( \cosh(t \|X_i\|_\infty) + \frac{X_i}{\|X_i\|_\infty} \sinh(t \|X_i\|_\infty) \right) \\
= \prod_{i=1}^{n} \cosh(t \|X_i\|_\infty) \\
\leq \exp \left( \frac{t^2}{2} \sum_{i=1}^{n} \|X_i\|_\infty^2 \right)
\] (10.8)

where we used the following bound on \( \cosh(x) \)
\[
\cosh(x) = \sum_{k=0}^{\infty} \frac{X^{2k}}{(2k)!} \\
\leq \sum_{k=0}^{\infty} \frac{X^{2k}}{2^k k!} \\
= e^{x^2/2}
\]

(It would be trivial to show that \( \prod \cosh(t \|X_i\|_\infty) \leq e^{t \sum_{i=1}^{n} \|X_i\|_\infty} \), but we need a quadratic term in the exponent to make our bound work, i.e. so that \( t^2 \ll t \).

Now we are ready to prove the theorem using Equation 10.5 and Markov’s Inequality.
\[
Pr \left( \sum_{i=1}^{n} X_i \geq L \right) = Pr \left( \exp \left( t \sum_{i=1}^{n} X_i \right) \geq e^{tL} \right) \\
\leq E \left[ \exp(t \sum_{i=1}^{n} X_i) \right] / e^{tL}
\]
\[ \leq \exp \left( \left( \frac{t^2}{2} \right) \left( \sum_{i=1}^{n} \|X_i\|_\infty^2 \right) - tL \right) \]
\[ = \exp \left( -\frac{L^2}{2} \left( \sum_{i=1}^{n} \|X_i\|_\infty^2 \right) \right) \text{ for } t = \frac{L}{\sum_{i=1}^{n} \|X_i\|_\infty^2} \]

Note that it is analogous to prove the other side of the bound. So the Hoeffding-Azuma Inequality actually implies that \( \Pr (|\sum_{i=1}^{n} X_i| \geq L) \leq 2 \exp \left( -\frac{L^2}{2} \left( \sum_{i=1}^{n} \|X_i\|_\infty^2 \right) \right) \).

## 10.2 Johnson-Lindenstrauss Lemma revisited

We present another proof of the Johnson-Lindenstrauss dimension reduction lemma from last lecture, using entirely different techniques.

**Theorem 10.3 JL Lemma:** If \( V \subseteq \mathbb{R}^d \) s.t. \( |V| = n \), then there is for every \( 0 < \epsilon < 1/2 \), a linear map \( A : \mathbb{R}^d \to \mathbb{R}^k \) such that \( \forall v_i, v_j \in V \)
\[ (1 - \epsilon) \|v_i - v_j\|_2 \leq \|Av_i - Av_j\|_2 \leq (1 + \epsilon) \|v_i - v_j\|_2 \]
and \( k = \Theta \left( \frac{\log \frac{n}{\epsilon^2}}{\epsilon^2} \right) \).

**Proof:** Define the random matrix \( A \) by \( A = \frac{1}{\sqrt{k}} \left( X^{(j)}_{i} \right)_{j,i} \), where \( X^{(j)}_{i} \sim \mathcal{N}(0, 1) \) are i.i.d. random variables.

Let \( u = \frac{v_i - v_j}{\|v_i - v_j\|_2} \), so that \( \sum_{i=1}^{n} u_i^2 = 1 \). Then \( \|Au\|_2^2 = \frac{1}{k} \sum_{i=1}^{k} \left( \sum_{j=1}^{n} X^{(j)}_{i} u_i \right)^2 \).

Our proof will use two facts about normal random variables, as well as a convenient function definition, \( \varphi(\lambda) \). Eventually, we want to show that with high probability, \( (1 - \epsilon) \leq \|Au\| \leq (1 + \epsilon) \).

**Fact 1 (2-stability property of normal random variables):** If \( X, Y \sim \mathcal{N}(0, 1) \) are i.i.d., then \( aX + bY \sim \mathcal{N}(0, a^2 + b^2) \).

So let us define random variable \( Y_j = \sum_{i=1}^{n} X^{(j)}_{i} u_i \). From Fact 1, \( Y_j \sim \mathcal{N}(0, 1) \).

**Fact 2:** If \( Y \sim \mathcal{N}(0, 1) \),
\[ \mathbb{E} \left[ e^{\lambda Y^2} \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda y^2} e^{-y^2/2} dy = \frac{1}{\sqrt{1 - 2\lambda}} \]
We now have \( \|Au\|_2^2 = \frac{1}{k} \sum_{j} Y_j^2 \).

Define \( \varphi(\lambda) \) to be
\[ \varphi(\lambda) = \left| \log \mathbb{E} \left[ e^{\lambda(Y^2 - 1)} \right] \right| \]

Then,
\[ \varphi(\lambda) = \left| -\frac{1}{2} \log(1 - 2\lambda) - \lambda \right| = \sum_{k=2}^{\infty} \frac{2^{k-1} \lambda^k}{k} \]
\[
\leq 2\lambda^2(1 + 2\lambda + (2\lambda)^2 + \ldots) \\
\leq \frac{2\lambda^2}{1 - 2\lambda} \text{ for } |\lambda| < \frac{1}{2}
\]

Now to finish the proof:

\[
\Pr \left[ \|Au\|_2^2 \geq 1 + \epsilon \right] = \Pr \left[ \frac{1}{k} \sum_{j=1}^{k} Y_j^2 \geq 1 + \epsilon \right] \\
= \Pr \left[ \frac{1}{k} \sum_{j=1}^{k} (Y_j^2 - 1) \geq \epsilon \right] \\
= \Pr \left[ e^{\lambda \sum_{j=1}^{k} (Y_j^2 - 1)} \geq e^{\lambda \epsilon k} \right] \\
\leq \frac{e^{k\phi(\lambda)}}{e^{\lambda \epsilon k}} \\
\leq e^{(\frac{2\lambda^2}{1 - 2\lambda} - \lambda)k} \\
\leq e^{-2\log n} \text{ for } \lambda = \frac{\epsilon}{4} \text{ and } k = \frac{24\log n}{\epsilon^2}, \epsilon < \frac{1}{2} \\
= \frac{1}{n^2}
\]

The proof above hinges on the fact that we could bound \( \mathbb{E} \left[ e^{\lambda Y^2} \right] \) nicely since the \( X_i^{(j)} \sim N(0,1) \). The proof of the other side of the inequality is also very similar to this. It turns out that we can relax \( X_i^{(j)} \) to be any sub-Gaussian random variable and lose only a constant factor in our choice of \( k \).

**Definition 10.4** \( X \) is a sub-Gaussian random variable means

\[
\mathbb{E} \left[ e^{tX} \right] \leq e^{Ct^2} \text{ for some } C > 0.
\]

For example, if \( X \) is \{±1\} uniformly at random, then \( \mathbb{E} \left[ e^{tX} \right] = \frac{1}{2} e^t + \frac{1}{2} e^{-t} = \cosh(t) \leq e^{t^2/2} \). So \( X \) is sub-Gaussian with \( C = \frac{1}{2} \).

**Claim 10.5** If \( X \) is sub-Gaussian (with constant \( C \)) and \( X_i^{(j)} \sim X \) are i.i.d., then \( \mathbb{E} \left[ e^{\lambda Y^2} \right] \leq \frac{1}{\sqrt{1 - 4C\lambda}} \), where \( Y = \sum_{j=1}^{k} X_i^{(j)} u_j \).

**Proof:** Let \( Z \sim N(0,1) \). So \( \mathbb{E} \left[ e^{\lambda Z} \right] = e^{\lambda^2/2} \). Then,

\[
\mathbb{E} \left[ e^{\lambda Y^2} \right] = \mathbb{E} \left[ e^{(\sqrt{\lambda} X)^2/2} \right] \\
= \mathbb{E}_Y \mathbb{E}_Z \left[ e^{\sqrt{\lambda} YZ} \right] \\
= \mathbb{E}_{Y,Z} \left[ \exp \left( \sum_{i=1}^{k} \sqrt{2\lambda} u_i X_i Z \right) \right]
\]
\[
\begin{align*}
&= \mathbb{E}_Z \left[ \mathbb{E}_Y \left[ e^{\sqrt{2} \sum u_i Z_i} \mid Z \right] \right] \\
&= \mathbb{E}_Z \left[ e^{2\lambda_c Z^2 \sum u_i^2} \right] \\
&= \mathbb{E}_Z \left[ e^{2\lambda_c Z^2} \right] \\
&\leq \frac{1}{\sqrt{1 - 4\lambda_c}}
\end{align*}
\]

Fubini’s Theorem allows us to change the order of expectation in the fourth step. The \(X_i^{(j)}\)’s being sub-Gaussian (with constant \(C\)) gives rise to the fifth step. The sixth step comes from the fact that \(\sum_{i=1}^{k} u_i^2 = 1\). The last step comes from the fact that \(Z \sim \mathcal{N}(0, 1)\) and FACT 2. \(\blacklozenge\)

So we can get a variant of FACT 2 for any matrix \(A\) with independent sub-Gaussian entries. We will lose a factor of \(C\), i.e. \(k = \Theta \left( \frac{C^2 \log n}{\epsilon^2} \right) \).