Lecture 5 — Tuesday, Jan 21, 2020

Lecturer: Mert Pilanci *Scribe: Srivatsan Sridhar, Varun Srivastava*

5.1 A generalized theory of dimension reduction

In this lecture, we study the problem of dimension reduction in general, of which the approximate matrix multiplication, trace estimation and matrix product verification algorithms covered in previous lectures, are special cases. We will define this problem, and state and prove the Johnson-Lindenstrauss (JL) Lemma in this lecture, which is fundamental in the theory of dimension reduction.

Given an arbitrary set of vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, we wish to map them to lower dimensional projections $y_1, \ldots, y_n \in \mathbb{R}^m$ with $m \ll d$. We will consider linear transformations of the form $y_i = Sx_i$. Informally, we want a good dimension reduction method to preserve distances between points, i.e. $\|y_i - y_j\| \approx \|x_i - x_j\|$ for $i, j = 1, \ldots, n$.

For example, when $d = 2$ and $m = 1$, the point x_i in the plane $\mathbb{R}^2$ are projected onto a line in the plane. In this case, S will just be a row vector which gives the direction of the line. If S is a fixed matrix, there will exists a configuration of points $\{x_i\}$ such that the distances will not be preserved. In the 2-dimensional example, suppose points were projected onto the Y-axis. Then all points with the same Y-coordinate get projected to the same point. The distances would not be preserved then (the distance between the projected points is 0). Thus, S will be chosen as a randomized matrix so that distances are preserved within an ϵ-error with high probability. Common choices are i.i.d. Gaussian matrix, i.i.d. Rademacher matrix (± 1 with prob. $\frac{1}{2}$), sampling matrix (exactly one 1 in each row).

5.1.1 Johnson-Lindenstrauss Lemma

Lemma 1 (Johnson-Lindenstrauss Lemma). Given n points $x_1, \ldots, x_n \in \mathbb{R}^d$, and $\epsilon > 0$, there exists a matrix $S \in \mathbb{R}^{m \times d}$ with $m > \frac{9 \log n}{\epsilon^2 - \epsilon^3}$ such that the following holds for all $i, j = 1, \ldots, n$ and $i \neq j$ with probability at least $1/2$:

$$1 - \epsilon < \frac{\|S(x_i - x_j)\|^2}{\|x_i - x_j\|^2} < 1 + \epsilon \tag{5.1}$$

The proof for this lemma will be given ahead. In particular, we will consider the matrix $S = \frac{1}{\sqrt{m}} G$ where $G_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$.

5.1.2 Distance-Preserving Property

If we want the S to be a distance preserving transformation, we expect that

$$\mathbb{E}\|S(x_i - x_j)\|^2 = \|x_i - x_j\|^2 \tag{5.2}$$

In this subsection, we will prove this. Suppose that $S = \frac{1}{\sqrt{m}}G$ where $G_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Define $u_{ij} = \frac{x_i - x_j}{\|x_i - x_j\|}$ so that $\|u_{ij}\| = 1$. Let $G_{(k)}$ denote the k^{th} row of G.

$$
\begin{aligned}
\mathbb{E}\|Su_{ij}\|^2 &= \mathbb{E}\left[\frac{1}{m}]\|Gu_{ij}\|^2\right] \\
&= \frac{1}{m}\mathbb{E}\left[u_{ij}^T G^T G u_{ij}\right] \\
&= \frac{1}{m}u_{ij}^T \mathbb{E}\left[G^T G\right] u_{ij} \\
&= \frac{1}{m}u_{ij}^T \mathbb{E}\left[\sum_{k=1}^{m} G_{(k)} G_{(k)}^T\right] u_{ij} \\
&= \frac{1}{m}u_{ij}^T \left(\sum_{k=1}^{m} \mathbb{E}\left[G_{(k)} G_{(k)}^T\right]\right) u_{ij} \\
&= \frac{1}{m}u_{ij}^T \left(\sum_{k=1}^{m} \mathbb{I}_d\right) u_{ij} \\
&= u_{ij}^T \mathbb{I}_d u_{ij} \\
&= \|u_{ij}\|^2
\end{aligned}
$$

5.1.3 Concentration of Measure for Uniform Distribution

Before we consider the full fledged JL lemma for projection to m dimensions, we consider a simpler concentration result for $m = 1$ which admits an intuitive geometric argument. We have already shown that the projection preserves the norm in expectation i.e.

$$
\mathbb{E}\|Su_{ij}\|^2 = \|u_{ij}\|^2
$$

For one dimensional projection, S is a row vector $S = g^T \in \mathbb{R}^d \sim \mathcal{N}(0, \mathbb{I}_d)$. Then we examine the probability $\mathbb{P}\left(|Su| \geq \epsilon\right) = \mathbb{P}\left(|g^T u| \geq \epsilon\right)$.

Note that since the d dimensional Gausssian distribution $\mathcal{N}(0, \mathbb{I}_d)$ is spherically symmetric, the probability $\mathbb{P}\left(|g^T u| \geq \epsilon\right) = \mathbb{P}\left(|g^T e_1| \geq \epsilon\right) = \mathbb{P}\left(|g_1| \geq \epsilon\right)$. Then we have,

Lemma 2. $\mathbb{P}\left(|g_1| \geq \frac{t\|g\|_2}{\sqrt{d}}\right) \leq 2e^{-\frac{t^2}{2}}$

Proof. Note that the vector $r = \frac{g}{\|g\|_2}$ is distributed *uniformly* on the unit sphere.[1]

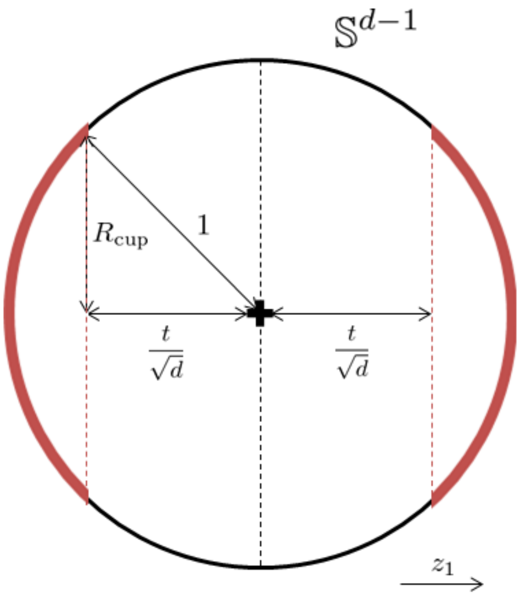
Then, the given lemma is equivalent to

$$
\mathbb{P}\left(|r_1| \geq \frac{t}{\sqrt{d}}\right) \leq 2e^{-\frac{t^2}{2}}
$$

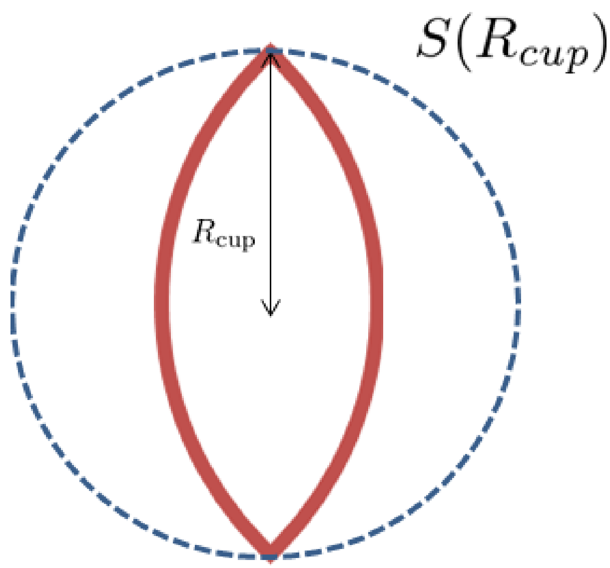
[1] A concise proof can be found at https://math.stackexchange.com/questions/444700/uniform-distribution-on-the-surface-of-unit-sphere

A simple geometric argument is utilized to show this. For a fixed $t > 0$, it can be seen from Figure 5.1.3(a) that the probability of choosing an $r \sim$ Uniform on the unit sphere in $\mathbb{R}^d$ with $|r_1| \geq \frac{t}{\sqrt{d}}$ is the ratio of the area of the d- dimensional spherical cups of radius given by $R_{cup} = \sqrt{1 - \frac{t^2}{d}}$ and the total area of the d dimension unit sphere. Hence,

$$\mathbb{P}\left(|r_1| \geq \frac{t}{\sqrt{d}}\right) = \frac{\text{area}(R_{cup})}{\text{area}(R_{sphere})} \tag{5.3}$$



(a) The spherical cups corresponding to $|r_1| \geq \frac{t}{\sqrt{d}}$ are highlighted in red



(b) Illustration showing that the area of the spherical cups is upper bounded by the area of a sphere with the same radius.

Figure 5.1: Illustration of the concentration proof in $\mathbb{R}^2$ [AM]

As seen from Figure 5.1.3(b), the area of the spherical cups with radius $R_{cup} = \sqrt{1 - \frac{t^2}{d}}$ can be upperbounded by the area of a sphere with the same radius.

Hence,

$$\begin{aligned}
\mathbb{P}\left(|r_1| \geq \frac{t}{\sqrt{d}}\right) &= \frac{\text{area}(R_{cup})}{\text{area}(R_{sphere})} \\
&\leq \frac{\text{area(sphere with radius } R_{cup})}{\text{area}(R_{sphere})} \\
&= \sqrt{1 - \frac{t^2}{d}}^{(d-1)} && \text{As the area of unit sphere is 1} \\
&\leq 2e^{-\frac{t^2}{2}} && \text{As } \left(1 - \frac{x}{n}\right)^n \leq e^{-x} \qquad (5.4)
\end{aligned}$$

whenever $0 < t \leq \sqrt{\frac{d}{2}}$

Thus, we have shown for 1-dimensional projections, the probability mass for each component of the random vector r is concentrated in the region $\left[0, \frac{t}{\sqrt{d}}\right]$, and decays exponentially for values greater than $\frac{t}{\sqrt{d}}$ (i.e. each component has a subgaussian distribution)

□

5.1.4 Proof of JL Lemma

In this subsection, we will prove the JL lemma by proving that for any vector $u \in \mathbb{R}^d$ with $\|u\| = 1$ and $\epsilon > 0$, if $S \in \mathbb{R}^{m \times d}$ with $S = \frac{1}{\sqrt{m}} G$ where $G_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $m > \frac{9 \log n}{\epsilon^2 - \epsilon^3}$,

$$\mathbb{P}\left(1 - \epsilon \leq \|Su\|^2 \leq 1 + \epsilon\right) \geq \frac{1}{2} \tag{5.5}$$

Proof.

$$\begin{aligned}
\mathbb{P}\left(\|Su\|^2 > 1 + \epsilon\right) &= \mathbb{P}\left(\frac{1}{m}\|Gu\|^2 > 1 + \epsilon\right) \\
&= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{d} G_{ij}u_j\right)^2 > 1 + \epsilon\right) \\
&= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} z_i^2 > 1 + \epsilon\right)
\end{aligned}$$

Note that $z_i = \sum_j G_{ij}u_j$ is a sum of independent Gaussians and hence is also Gaussian. z_i is independent of z_j for $i \neq j$ since rows of G are independent. $\mathbb{E}[z_i] = \sum_j \mathbb{E}[G_{ij}]u_j = 0$ and $\text{Var}(x_i) = Var\left(\sum_j G_{ij}u_j\right) = \sum_j \text{Var}(G_{ij}u_j) = \sum_j |u_j|^2 = 1$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\|Su\|^2 > 1 + \epsilon\right) &= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} z_i^2 > 1 + \epsilon\right) \\
&= \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{m} z_i^2\right) > \exp\left(\lambda m(1 + \epsilon)\right)\right) \\
&\leq \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{m} z_i^2\right)\right]}{\exp\left(\lambda m(1+\epsilon)\right)} && \text{from Markov's inequality} \\
&= \frac{\mathbb{E}\left[\prod_{i=1}^{m} \exp\left(\lambda z_i^2\right)\right]}{\exp\left(\lambda m(1+\epsilon)\right)} && \text{due to independence of } z_i\text{'s} \\
&= \left(\frac{\exp\left(-2\lambda(1+\epsilon)\right)}{1 - 2\lambda}\right)^{m/2} && \text{Proof in Appendix 5.2.1} \\
&= \left((1+\epsilon)\exp\left(-\epsilon\right)\right)^{m/2} && \text{Putting } \lambda = \frac{\epsilon}{2(1+\epsilon)} \\
&\leq \exp\left(-\left(\epsilon^2 - \epsilon^3\right)\frac{m}{4}\right) && \text{from } 1 + \epsilon \leq \exp\left(\epsilon - (\epsilon^2 - \epsilon^3)/2\right)
\end{aligned}$$

In a very similar manner, we can also show that

$$\begin{aligned}
\mathbb{P}\left(\|Su\|^2 < 1-\epsilon\right) &= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} z_i^2 < 1-\epsilon\right) \\
&= \mathbb{P}\left(\exp\left(-\lambda\sum_{i=1}^{m} z_i^2\right) > \exp\left(-\lambda m(1-\epsilon)\right)\right) \\
&\leq \frac{\mathbb{E}\left[\exp\left(-\lambda\sum_{i=1}^{m} z_i^2\right)\right]}{\exp\left(-\lambda m(1-\epsilon)\right)} \\
&= \left(\frac{\exp\left(2\lambda(1-\epsilon)\right)}{1+2\lambda}\right)^{m/2} \\
&= \left((1-\epsilon)\exp\left(\epsilon\right)\right)^{m/2} \qquad \text{Putting } \lambda = \frac{\epsilon}{2(1-\epsilon)} \\
&\leq \exp\left(-\left(\epsilon^2-\epsilon^3\right)\frac{m}{4}\right)
\end{aligned}$$

Since the concentration results only applies to $u_{ij} = \frac{x_i - x_j}{\|x_i - x_j\|}$ for a particular pair of i, j, we now use the union bound to calculate the concentration around 1 for the entire dataset to conclude the proof

$$\begin{aligned}
\mathbb{P}\left(1-\epsilon \leq \|Su_{ij}\|^2 \leq 1+\epsilon \quad \forall i,j\right) &\geq 1 - \sum_{i,j} 2\exp\left(-\left(\epsilon^2-\epsilon^3\right)\frac{m}{4}\right) \\
&= 1 - 2\binom{n}{2}\exp\left(-\left(\epsilon^2-\epsilon^3\right)\frac{m}{4}\right) \\
&\geq 1 - 2n^2\exp\left(-\left(\epsilon^2-\epsilon^3\right)\frac{m}{4}\right)
\end{aligned}$$

Thus, we will have $\mathbb{P}\left(1-\epsilon \leq \|Su_{ij}\|^2 \leq 1+\epsilon \quad \forall i,j\right) \geq \frac{1}{2}$ if

$$\begin{aligned}
(\epsilon^2-\epsilon^3)\frac{m}{4} &\geq 2\log 2n \\
m &\geq \frac{9\log n}{(\epsilon^2-\epsilon^3)} \qquad \text{provided } n > 256\text{[Mah]}
\end{aligned}$$

□

5.2 Appendix

5.2.1 Proof for computing the exponentiated second moment

As $z_i = \sum_j G_{ij}u_j$ is a sum of independent Gaussians and hence is also Gaussian. z_i is independent of z_j for $i \neq j$ since rows of G are independent.

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{m} \exp\left(\lambda z_i^2\right)\right] &= \prod_{i=1}^{m} \mathbb{E}\left[e^{\lambda z_i^2}\right] \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda z^2} e^{-\frac{z^2}{2}} dz \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{1-2\lambda}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2\lambda}}{\sqrt{2\pi}} e^{-\frac{z^2}{2}(1-2\lambda)} dz \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{1-2\lambda}} \\
&= \left(\frac{1}{1-2\lambda}\right)^{m/2}
\end{aligned}
$$

Bibliography

[AM] Dorina Thanou Aleksander Madry, Alhussein Fawzi, *lecture21.pdf*, https://people.csail.mit.edu/madry/gems/notes/lecture21.pdf, (Accessed on 01/27/2020).

[Mah] Michael W. Mahoney, *1608.04481.pdf*, https://arxiv.org/pdf/1608.04481.pdf, (Accessed on 01/27/2020).