Lecture 11: Final Notes on JL and Subspace Embedding

09-30-2025 Lecturer: Ali Vakilian | Scribe: Manoj Saravanan | Editor: Ali Vakilian

Today, we will finalize our discussion on the Johnson-Lindenstrauss (JL) lemma and subspace embedding, focusing on the motivations and computational aspects of dimensionality reduction.

A main motivation. The primary goal is to reduce the dimensionality of the input with the hope of solving the problem faster. For a set of points $X \subset \mathbb{R}^d$, we seek a mapping f that sends x into $\mathbb{R}^m$ for $m \ll d$. A crucial question then arises: how fast can we compute the map f? Let's explore several approaches to constructing this mapping and their computational complexities.

1. For the original construction of the Johnson-Lindenstrauss lemma [Johnson and Lindenstrauss, 1984], the time required is $O(md)$.

2. Achlioptas [2001] introduced a sparser matrix $\Pi \in \mathbb{R}^{m \times d}$ that provides similar guarantees. The entries of this matrix are independently chosen at random and are equal to:

$$\begin{cases} \frac{1}{\sqrt{s}} & \text{with probability } \frac{1}{2s} \\ -\frac{1}{\sqrt{s}} & \text{with probability } \frac{1}{2s} \\ 0 & \text{with probability } 1 - \frac{1}{s} \end{cases}$$

   Where s is a parameter that controls sparsity. In a common setting, $s = m/3$.

3. **Fast JL Transform:** The main idea here is to pick a **sampling matrix** $S \in \mathbb{R}^{m \times d}$. The matrix S has a single 1 in a random location in each row (zeros elsewhere), with rows chosen at random. Computing the transformation $z \to \Pi z$ is fast, taking only $O(m)$ time. The expected squared norm is preserved:

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{m}}\Pi z\right\|_2^2\right] = \|z\|_2^2,$$

   though the variance can be high, especially when the mass of z is concentrated on a few coordinates. To mitigate this, apply a preconditioning operation R (for an orthogonal matrix R) so that $\frac{\|Rz\|_\infty}{\|Rz\|_2}$ is small with high probability; as a result, Rz becomes *well-spread*, with no single coordinate carrying too much mass. Consequently, the vector $\frac{1}{\sqrt{m}}SRz$ has roughly the same norm as z, and it can be computed in time $O(d \log d + m^3)$. A limitation of this approach is that, despite its speed, it does not exploit sparsity in z when present.

4. **Sparse JL Transform:** If the matrix Π has s non-zero entries per column, the product Πx can be computed in $O(s \cdot \|z\|_0)$ time, where $\|z\|_0$ is the number of non-zero entries in z. To see this, note that the product Πz can be viewed as a linear combination of the columns of Π: $\Pi z = \sum_i z_i \Pi^i$, where Π^i is the i-th column of Π.

   The objective is to make both m (the target dimension) and s (the sparsity) as small as possible. CountSketch provides a version of the Distributional JL transform with $m = O(1/(\epsilon^2 \delta))$ and $s = 1$. A result from Kane and Nelson [2014], similar to CountSketch with $s > 1$, improves the target dimension to $m = O(\log(1/\delta)/\epsilon^2)$ with sparsity $s = O(\epsilon m)$.

1 Sparse JL Transform Details

Define the Sparse JL transform matrix $\Pi \in \mathbb{R}^{m\times d}$ more formally by setting its entries to $\Pi_{r,i} = (\eta_{r,i}\sigma_{r,i})/\sqrt{s}$, where the $\sigma_{r,i}$ are independent Rademacher random variables taking values $+1$ or -1 with equal probability, and the $\eta_{r,i}$ are Bernoulli random variables with the following properties. For all r, i, $\mathbb{E}[\eta_{r,i}] = s/m$. For each fixed column i, the constraint $\sum_{r=1}^{m} \eta_{r,i} = s$ holds, so every column of Π has exactly s nonzero entries. Moreover, the variables $\eta_{r,i}$ are negatively correlated: for any finite $S \subset [m] \times [d]$,

$$\mathbb{E}\left[\prod_{(r,i)\in S} \eta_{r,i}\right] \le \prod_{(r,i)\in S} \mathbb{E}[\eta_{r,i}] \le \left(\frac{s}{m}\right)^{|S|}.$$

Theorem 1.1. *If $m = O(\log(\frac{1}{\delta})/\varepsilon^2)$ and $s = \Theta(\varepsilon m)$, then for any unit norm vector z,* $\Pr(|\|\Pi z\| - 1| > \varepsilon) \le \delta$

Note on Fast JL Transform. Ailon and Chazelle [2006] proposed a method where Πx can be computed in $O(d \log d)$ time. The transformation matrix is defined as $\Pi = \frac{1}{\sqrt{m}} SHD$ where,

- $S_{m\times d}$ is a sampling matrix.
- H is a Hadamard matrix.
- D is a diagonal matrix with independent Rademacher random variables on the diagonal.

Proof of Oblivious Subspace Embedding

Next, we revisit subspace embeddings and give a complete proof with no additional assumptions (unlike the simplified proof from Lecture 10).

Theorem 1.2. *Suppose $E \subset \mathbb{R}^n$ is a linear subspace of dimension d. Let $\Pi \in \mathbb{R}^{k\times n}$ be a random projection matrix (e.g., with entries from $\mathcal{N}(0,1)$) with $k = O\left(\frac{d}{\varepsilon^2} \log\left(\frac{1}{\delta}\right)\right)$ rows. Then with probability $(1-\delta)$, for every $x \in E$,*

$$\left\| \frac{1}{\sqrt{k}} \Pi x \right\|_2 = (1 \pm \varepsilon)\|x\|_2$$

In other words, the Johnson-Lindenstrauss Lemma extends smoothly from preserving the geometry of a single vector (a 1-dimensional subspace) to preserving the geometry of an arbitrary d-dimensional subspace.

Proof Challenges. How do we prove that Π works for all $x \in E$, which is an infinite set? A simple application of the union bound, as used for a finite set of points, will not work here. The key idea, as also discussed in Lecture 10, is *net argument*. The proof strategy then is as follows:

- **Step 1:** Choose a large but finite set of vectors $T \subset E$ carefully. This set is called an ε-net.
- **Step 2:** Prove that Π preserves the length of all vectors in the finite set T. This can be done using the standard JL Lemma and the union bound over the points in T.
- **Step 3:** Argue that any vector $x \in E$ is sufficiently close to some vector in T. By leveraging the properties of linear maps and the triangle inequality, we can show that if the lengths of vectors in T are preserved, then the length of x is also preserved.

To define formally, a subset T is an ε-net for a space S if for every point $p \in S$, there is a point $x \in T$ such that the distance between them is small. For example,

- In ℓ_2-space, it requires $\|x - p\|_2 \leq \varepsilon$,
- In ℓ_∞-space, it requires $\|x - p\|_\infty \leq \varepsilon$.

Proof of Theorem 1.2. A $(1 \pm \varepsilon)$ ℓ_2-subspace embedding for the column space of an $n \times d$ matrix A is a matrix S such that for all $x \in \mathbb{R}^d$,

$$(1-\varepsilon)\|Ax\|_2^2 \leq \|SAx\|_2^2 \leq (1+\varepsilon)\|Ax\|_2^2.$$

Let $U \in \mathbb{R}^{n \times d}$ have orthonormal columns spanning $\mathrm{col}(A)$. Then S is an ℓ_2-subspace embedding for $\mathrm{col}(A)$ if and only if, for all $x \in \mathbb{R}^d$,

$$(1-\varepsilon)\|x\|_2^2 \leq \|SUx\|_2^2 \leq (1+\varepsilon)\|x\|_2^2,$$

since $\|Ux\|_2^2 = \|x\|_2^2$ by orthonormality. Equivalently, it suffices to prove

$$\sup_{x \in \mathcal{S}^{d-1}} \left| \|SUx\|_2^2 - 1 \right| \leq \varepsilon,$$

because the condition is homogeneous and therefore determined by its restriction to the unit sphere $\mathcal{S}^{d-1}$. In particular, we will work with the unit vectors in the subspace $\mathrm{col}(U)$, namely the set $\{Ux : x \in \mathcal{S}^{d-1}\}$, and establish the displayed bound. To pass from this infinite family to a finite one, fix an ε-net $N \subset \mathcal{S}^{d-1}$, meaning that for every $x \in \mathcal{S}^{d-1}$ there exists $y \in N$ with $\|x - y\|_2 \leq \varepsilon$. Such a net can be constructed greedily, and standard volume arguments give the bound

$$|N| \cdot \mathrm{vol}\big(B(\varepsilon/2)\big) \leq \mathrm{vol}\big(B(1+\varepsilon/2)\big) \quad \Rightarrow \quad |N| \leq \left(1 + \tfrac{2}{\varepsilon}\right)^d,$$

where $B(r)$ denotes the d-dimensional Euclidean ball of radius r. Project N into the subspace via U and set

$$M := \{Ux : x \in N\}.$$

Claim 1.3. For every unit vector $z \in \mathrm{col}(U)$ there exists $y \in M$ such that $\|z - y\|_2 \leq \varepsilon$.

Proof. Write $z = Ux$ with $x \in \mathcal{S}^{d-1}$. By the definition of N, choose $x' \in N$ with $\|x - x'\|_2 \leq \varepsilon$ and set $y := Ux' \in M$. Since U has orthonormal columns, it is an isometry on $\mathbb{R}^d$, hence

$$\|z - y\|_2 = \|Ux - Ux'\|_2 = \|x - x'\|_2 \leq \varepsilon.$$

□

We now conclude the argument using the ε-net M. The goal is to approximate any unit vector $y \in \mathrm{col}(U)$ by a rapidly convergent series of vectors drawn from M.

Claim 1.4. For every $x \in \mathcal{S}^{d-1}$ there exists $y' \in M$ with $\|Ux - y'\|_2 \leq \varepsilon$.

Fix a unit vector $y = Ux$ with $x \in \mathcal{S}^{d-1}$. By the claim, choose $y_1 \in M$ with $\|y - y_1\|_2 \leq \varepsilon$. Define the normalized remainder $\hat{r}_1 := (y - y_1)/\|y - y_1\|_2$ and set $\alpha_1 := \|y - y_1\|_2 \leq \varepsilon$. Apply the claim again to obtain $y_2' \in M$ with $\|\hat{r}_1 - y_2'\|_2 \leq \varepsilon$, and set $y_2 := \alpha_1 y_2'$, so that

$$\|y - y_1 - y_2\|_2 \leq \alpha_1 \varepsilon \leq \varepsilon^2.$$

Iterating this construction yields a sequence $(y_i)_{i\geq 1}$ with $y_i \in \operatorname{span}(M)$ and, for every $k \geq 1$,

$$\Big\| y - \sum_{i=1}^{k} y_i \Big\|_2 \leq \varepsilon^k \quad \text{and} \quad \|y_i\|_2 \leq \varepsilon^{i-1} + \varepsilon^i \leq 2\varepsilon^{i-1}.$$

Hence $y = \sum_{i\geq 1} y_i$ with $\sum_{i\geq 1} \|y_i\|_2 \leq \sum_{i\geq 1} 2\varepsilon^{i-1} = \frac{2}{1-\varepsilon}$. By a union bound over the finite set M (and standard JL tail bounds), with high probability the embedding S simultaneously satisfies for all $y', y'' \in M$,

$$(1-\varepsilon)\|y'\|_2^2 \leq \|Sy'\|_2^2 \leq (1+\varepsilon)\|y'\|_2^2, \quad \big| \langle Sy', Sy'' \rangle - \langle y', y'' \rangle \big| \leq c\,\varepsilon\, \|y'\|_2 \|y''\|_2,$$

for an absolute constant $c > 0$ (the inner-product bound follows from the distance bound via polarization).

Write $y = \sum_i y_i$ and expand:

$$\|Sy\|_2^2 = \Big\| S \sum_i y_i \Big\|_2^2 = \sum_i \|Sy_i\|_2^2 + 2 \sum_{i<j} \langle Sy_i, Sy_j \rangle.$$

Using the displayed estimates,

$$\begin{aligned} \big| \|Sy\|_2^2 - \|y\|_2^2 \big| &\leq \varepsilon \sum_i \|y_i\|_2^2 + 2c\varepsilon \sum_{i<j} \|y_i\|_2 \|y_j\|_2 \\ &\leq c'\,\varepsilon \Big(\sum_i \|y_i\|_2 \Big)^2 \leq c'\,\varepsilon \Big(\frac{2}{1-\varepsilon} \Big)^2, \end{aligned}$$

for an absolute constant c'. Since y is a unit vector, this gives

$$\|Sy\|_2^2 = 1 \pm C\,\varepsilon,$$

for an absolute constant C, which is the desired $(1 \pm O(\varepsilon))$ guarantee. By homogeneity, the bound extends to all $y \in \operatorname{col}(U)$, completing the proof. $\square$