Lecture 10: Subspace Embedding

09-25-2025 Lecturer: Ali Vakilian | Scribe: Nabin Adhikari | Editor: Ali Vakilian

1 Recap

In our last lecture, we saw how the *Johnson-Lindenstrauss (JL) lemma* allows us to perform dimensionality reduction for a finite set of vectors in a high-dimensional space. The key result is that a random projection can drastically reduce the dimension while approximately preserving all pairwise distances.

However, in many applications, such as in numerical linear algebra, we need to preserve the geometry of an entire *subspace*, not just a finite collection of points. Unfortunately, preserving the norm of *every* vector in a high-dimensional space is impossible with any projection to a lower dimension.

Impossibility for all of $\mathbb{R}^n$. No linear map $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ with $k < n$ can preserve the norms of all vectors. This is because any such map must have a non-trivial kernel, meaning there exists a non-zero vector x such that $\Pi x = 0$, completely destroying its norm.

Given this limitation, we shift our focus from preserving all of $\mathbb{R}^n$ to the more tractable goal of preserving a d-dimensional subspace ($d \leq k$), which is the core idea behind subspace embeddings.

1.1 Proof of Distributional JL and the Gaussian View

The proofs of these JL-type results rely on fundamental properties of Gaussian and other well-behaved random variables.

- **Norm of a Gaussian Vector:** If $Z = (Z_1, \ldots, Z_k)$ is a vector of k i.i.d. standard normal random variables, $Z_i \sim \mathcal{N}(0,1)$, then its squared Euclidean norm, $\|Z\|_2^2$, follows a chi-squared distribution with k degrees of freedom, denoted $\chi^2(k)$.

- χ^2 **Concentration:** A key property of the $\chi^2(k)$ distribution is that it is sharply concentrated around its mean, which is k. For $\epsilon \in (0, 1/2)$, we have:

  $$\Pr\left[(1-\varepsilon)k \leq \|Z\|_2^2 \leq (1+\varepsilon)k\right] \geq 1 - 2e^{-c\varepsilon^2 k}.$$

  This shows that the norm of a random Gaussian vector is very unlikely to be far from its expected value.

- **Rotational Invariance:** A crucial property of the Gaussian distribution is that it is rotationally invariant. This means that for a vector a, the projection of a standard Gaussian vector $Z \sim \mathcal{N}(0, I_d)$ onto a is a one-dimensional Gaussian with variance equal to the squared norm of a: $a^\top Z \sim \mathcal{N}(0, \|a\|_2^2)$.

- **Implication:** These facts together imply that random projections using Gaussian (or Rademacher, i.e., ± 1) matrices approximately preserve ℓ_2-norms, which is the core mathematical insight behind the DJL.

Proof of Theorem 2.1 in Lecture 9. Without loss of generality, we can assume that x is a unit vector, i.e., $\|x\|_2 = 1$. The result for any arbitrary vector x then follows by linearity, as we can scale the projection by $\|x\|_2$.

Let the rows of Π be $z_1, \ldots, z_k \in \mathbb{R}^d$, where each z_i is a vector with i.i.d. standard normal entries. For a fixed x with $\|x\|_2 = 1$, the dot product $z_i \cdot x$ is a linear combination of standard Gaussians, which results in a standard normal random variable itself: $z_i \cdot x \sim \mathcal{N}(0, \|x\|_2^2) = \mathcal{N}(0,1)$.

Define a random variable Y as the sum of the squares of these projections:

$$Y = \sum_{i=1}^{k} (z_i \cdot x)^2.$$

Since each term $(z_i \cdot x)^2$ is the square of a standard normal variable, Y follows a chi-squared distribution with k degrees of freedom, i.e., $Y \sim \chi^2(k)$.

Note that $\|\Pi x\|_2^2 = \sum_{i=1}^{k} (z_i \cdot x)^2 = Y$. Therefore, the expression in the theorem can be rewritten in terms of Y:

$$\frac{1}{k}\|\Pi x\|_2^2 = \frac{Y}{k}.$$

We now use a standard concentration inequality for the chi-squared distribution. For any $0 < \varepsilon < 1/2$, there exists a constant $c > 0$ such that:

$$\Pr\left[|Y - k| \geq \varepsilon k\right] \leq 2\exp(-c\varepsilon^2 k).$$

This is equivalent to saying that the average, Y/k, is close to 1:

$$\Pr\left[\frac{Y}{k} \in (1 \pm \varepsilon)\right] \geq 1 - 2e^{-c\varepsilon^2 k}.$$

To ensure that the failure probability is at most δ, we set $2e^{-c\varepsilon^2 k} \leq \delta$. Solving for k shows that it is sufficient to choose $k = \Omega(\varepsilon^{-2}\log(1/\delta))$. □

2 Subspace Embeddings

Definition 2.1 (Subspace Embedding). Let $E \subset \mathbb{R}^n$ be a linear subspace of dimension $d < n$. A random linear map $\Pi \in \mathbb{R}^{k \times n}$ is an $(1 \pm \varepsilon)$-subspace embedding for E if for all $x \in E$,

$$\frac{1}{k}\|\Pi x\|^2 = (1 \pm \varepsilon)\|x\|^2.$$

Such a Π is called a *(oblivious) subspace embedding* for E, as it oblivious to the given subspace.

Theorem 2.2 (Subspace Embedding Theorem). *Let $E \subseteq \mathbb{R}^n$ of dimension r, and let $\Pi \in \mathbb{R}^{m \times n}$ have i.i.d. $\mathcal{N}(0,1)$ entries. If $m = O\big(\varepsilon^{-2}(d + \log(1/\delta))\big)$, then with probability at least $1 - \delta$, Π is an $(1 \pm \varepsilon)$-subspace embedding for E.*

The main challenge is that the subspace E contains infinitely many vectors, making our previous union bound strategy for the Metric JL lemma (in Lecture 7) impossible. Instead, we will use a standard discretization argument. The key idea is to first cover the unit sphere within E with a finite set of points, known as an *ε-net*. We then use a union bound to prove that the norm is preserved for every point in this finite net. Finally, we extend this guarantee from the net to all unit vectors in the subspace by an approximation argument.

Proof of Theorem 2.2. Here, we prove a simpler version of the theorem that conveys the high-level ideas of the proof. We make a simplifying assumption that the d-dimensional subspace E is aligned with the first d standard basis vectors of $\mathbb{R}^n$. Thus, any vector in E has non-zero values only in its first d coordinates.

Let $\mathcal{S}^{d-1}$ be the unit sphere within this subspace E. Our strategy is to construct an η-net $T \subset \mathcal{S}^{d-1}$, which is a finite set of points such that for any vector $x \in \mathcal{S}^{d-1}$, there is a nearby point $y \in T$ satisfying

$\|x-y\|_2 \le \eta$. Given our simplifying assumption about E, such a net can be constructed from grid points of width $\frac{\eta}{d}$ in $[-1,1]^d$, yielding a net of size roughly $O((d/\eta)^d)$.

Next, we only need apply JL lemma to each $y \in T$ and the first d standard basis in $\mathbb{R}^n$. With $\varepsilon' = \varepsilon/2$ and $\delta' = \delta/(|T|+d)$, we need $k = O\left(\varepsilon^{-2} \log \frac{|T|}{\delta}\right)$. Thus, applying union bound over all point in the net and the first d basis, with probability at least $1-\delta$, every $y \in T \cup \{e_1, \dots, e_d\}$ satisfies

$$\frac{1}{m}\|\Pi y\|^2 = (1 \pm \varepsilon/2)\|y\|^2.$$

For arbitrary unit $x \in E$, pick $y \in T$ with $\|x-y\| \le \eta$. Then,

$$\frac{1}{m}\|\Pi x\|^2 \le \frac{1}{m}\|\Pi y\|^2 + \frac{1}{m}\|\Pi(x-y)\|^2 \le (1+\frac{\varepsilon}{2}) + (1+\frac{\varepsilon}{2})\sum_{i=1}^{d}|z_i| \le (1+\frac{\varepsilon}{2})(1+\eta)$$

Similarly,

$$\frac{1}{m}\|\Pi x\|^2 \ge \frac{1}{m}\|\Pi y\|^2 - \frac{1}{m}\|\Pi(x-y)\|^2 \ge (1-\frac{\varepsilon}{2}) - (1+\frac{\varepsilon}{2})\sum_{i=1}^{d}|z_i| \le (1+\frac{\varepsilon}{2})(1-\eta)$$

In particular, choosing $\eta = \varepsilon/4$,

$$1-\varepsilon \le \frac{1}{m}\|\Pi x\|^2 \le 1+\varepsilon$$

□

References