Randomized Algorithms 2017A – Lecture 7
Dimension Reduction in ℓ_2*

Robert Krauthgamer

1 The Johnson-Lindenstrauss (JL) Lemma

The Johnson-Lindenstrauss (JL) Lemma: Let $x_1, \ldots, x_n \in \mathbb{R}^d$ and fix $0 < \varepsilon < 1$. Then there exist $y_1, \ldots, y_n \in \mathbb{R}^k$ and $k = O(\varepsilon^{-2} \log n)$, such that

$$\forall i, j \in [n], \qquad \|y_i - y_j\|_2 \in (1 \pm \varepsilon)\|x_i - x_j\|_2.$$

Moreover, there is a randomized linear mapping $L : \mathbb{R}^d \to \mathbb{R}^k$ (oblivious to the given points), such that if we define $y_i = Lx_i$, then with probability at least $1 - 1/n$ all the above inequalities hold.

Throughout, all norms are ℓ_2, unless stated otherwise.

Remark: Note there is no assumption on the input points (e.g., that they lie on a low-dimensonal space).

Idea: The map L is essentially (up to normalization) a matrix of standard Gaussians. In fact, random signs ± 1 would also work!

Since L is linear, $Lx_i - Lx_j = L(x_i - x_j)$, and it suffices to verify that L preserves the norm of arbitrary vector WHP (instead of arbitrary pair of vectors).

Lemma 2 (Main): Let $G \in \mathbb{R}^{d \times k}$ be a random matrix of standard Gaussians, for suitable $k = O(\varepsilon^{-2} \log n)$.

$$\forall v \in \mathbb{R}^d, \qquad \Pr\left[\|Gv\| \notin (1 \pm \varepsilon)\sqrt{k}\|v\|\right] \leq 2/n^3.$$

In fact, the proof shows that the failure probability is at most δ when $k = O(\varepsilon^{-2} \log \frac{1}{\delta})$.

Using main lemma: Let $L = G/\sqrt{k}$, and recall we defined $y_i = Lx_i$. For every $i < j$, apply the lemma to $x_i - x_j$, then with probability at least $1 - 2/n^3$,

$$\|y_i - y_j\| = \|L(x_i - x_j)\| = \|G(x_i - x_j)\|/\sqrt{k} \in (1 \pm \varepsilon)\|x_i - x_j\|.$$

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

Now apply a union bound over $\binom{n}{2}$ pairs.

QED

It remains to prove the main lemma.

Fact 3 (the sum of Gaussians is Gaussian): Let $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ be independent Gaussian random variables. Then $X + Y \sim N(0, \sigma_X^2 + \sigma_Y^2)$.

The proof is by writing the CDF function (integration), recall that PDF is $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Corollary 4 (Gaussians are 2-stable): Let $X_1, \ldots, X_n$ be independent standard Gaussians $N(0,1)$, and let $\sigma_1, \ldots, \sigma_n \in \mathbb{R}$. Then $\sum_i \sigma_i X_i \sim N(0, \sum_i \sigma_i^2)$.

Follows by induction.

Proof of main lemma: Was seen in class, using the next claim.

Claim 5: Let Y have chi-squared distribution with parameter k, i.e., $Y = \sum_{i=1}^k X_i^2$ for independent $X_1, \ldots, X_k \sim N(0,1)$. Then

$$\forall \varepsilon \in (0,1), \qquad \Pr[Y \geq (1+\varepsilon)^2 k] \leq e^{-(3/4)\varepsilon^2 k}.$$

Remark: The claim and its proof are similar to Hoeffding bounds. Indeed, one may compare Claim 5 to the case $Y \sim 2 \cdot B(k, 1/2)$ which has the same expectation.

Proof of Claim 5: Was seen in class.

Exer: Show that the main lemma (and thus the JL Lemma) extends to every matrix G whose entries are iid from a distribution that has mean 0, variance 1, and satisfies a sub-Gaussian tail bound $\mathbb{E}[e^{tX}] \leq e^{Ct^2}$ for some constant $C > 0$. And use it to conclude in particular for a matrix of ± 1.

Hint: Use the following trick. Introduce a standard Gaussian Z independent of X, then $\mathbb{E}[e^{tZ}] = e^{t^2/2}$, and thus

$$\mathbb{E}_X[e^{tX^2}] = \mathbb{E}_X[e^{(\sqrt{2tX^2})^2/2}] = \mathbb{E}_X \mathbb{E}_Z[e^{\sqrt{2tX^2} Z}] = \mathbb{E}_Z \mathbb{E}_X[e^{\sqrt{2tZ^2} X}] \leq \mathbb{E}_Z[e^{2CtZ^2}],$$

and the last term can be evaluated using the previous exercise.

2 Fast JL Transform

Computing the JL map of a vector requires the multiplication of a matrix $L \in \mathbb{R}^{k \times d}$ with a vector $x \in \mathbb{R}^d$, which generally takes $O(kd)$ time, because L is a dense matrix.

Question: Can we compute it faster?

Sparse JL: Some constructions (see Kane-Nelson, JACM 2014) use a *sparse* matrix L, namely, only an ε-fraction of the entries are nonzero, leading to a speedup by factor ε (and even more if x is sparse).

We will see another approach, where L is dense but its special structure leads to fast multiplication, close to $O(d+k)$ instead of $O(kd)$.

Theorem 6 [Ailon and Chazelle, 2006]: There is a random matrix $L \in \mathbb{R}^{k\times d}$ that satisfies the guarantees of the JL lemma and for which matrix-vector multiplication takes time $O(d\log d + k^3)$.

We will see a simplified version of this theorem (faster but higher dimension).

Theorem 7: For every $d \geq 1$ and $0 < \delta < 1$, there is a random matrix $L \in \mathbb{R}^{k\times d}$ for $k = O(\varepsilon^{-2}\log^2(d/\delta)\log(1/\delta))$, such that

$$\forall v \in \mathbb{R}^d, \qquad \Pr\Big[\|Lv\| \notin (1\pm\varepsilon)\|v\|\Big] \leq 1/\delta,$$

and multiplying L with a vector v takes time $O(d\log d + k)$.

Super-Sparse Sampling: A basic idea is to just sample one entry of v (each time).

Let $S \in \mathbb{R}^{k\times d}$ be a matrix where each row has a single nonzero entry of value $\sqrt{d/k}$ in a uniformly random location. This is sometimes called a sampling matrix (up to appropriate scaling). For every $v \in \mathbb{R}^d$,

$$\mathbb{E}[(Sv)_1^2] = \sum_{j=1}^{d} \tfrac{1}{d}(\sqrt{d/k}\cdot x_j)^2 = \tfrac{1}{k}\|v\|^2.$$

$$\mathbb{E}[\|Sv\|^2] = \sum_{i=1}^{k} \mathbb{E}[(Sv)_i^2] = \|v\|^2.$$

The expectation is correct, however the variance can be huge, e.g., if v has just one nonzero coordinate, then for S to be likely to sample it, we need $k = \Omega(d)$.

We shall first see how to transform v into a vector $y \in \mathbb{R}^d$ with no "heavy" coordinate, meaning that

$$\frac{\|y\|_\infty}{\|y\|_2} \approx \frac{1}{\sqrt{d}}.$$

and later we will prove that super-sparse sampling works for such vectors.

Definition: A *Hadamard matrix* is a matrix $H \in \mathbb{R}^{d\times d}$ that is orthogonal, i.e., $H^T H = I$ and all its entries are in $\{\pm 1/\sqrt{d}\}$.

Observe that by definition $\|Hv\|_2^2 = (Hv)^T(Hv) = v^T v^T = \|v\|_2$.

When d is a power of 2, such a matrix exists, and can be constructed by induction as follows (called a Walsh-Hadamard matrix).

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} /\sqrt{2},$$

$$H_d = \begin{pmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & -H_{d/2} \end{pmatrix} /\sqrt{2}.$$

It is easy to verify it is indeed a Hadamard matrix, i.e., that all entries are $\pm 1/\sqrt{d}$ and $H_d^T H_d = I$.

Lemma 8: Multiplying H_d by a vector can be performed in time $O(d \log d)$.

Exer: Prove this lemma, using divide and conquer.

Randomized Hadamard matrix: Let $D \in \mathbb{R}^{d \times d}$ be a diagonal matrix whose ith diagonal entry is an independent random sign $r_i \in \{\pm 1\}$. Observe that HD is a random Hadamard matrix, because its entries are still $\pm 1/\sqrt{d}$ and $(HD)^T(HD) = D^T H^T H D = D^T D = I$.

Lemma 9: Let HD be a random Hadamard matrix as above. Then

$$\forall 0 \neq v \in \mathbb{R}^d, \qquad \Pr_D \left[\frac{\|HDv\|_\infty}{\|HDv\|_2} \geq \sqrt{\frac{2 \ln(4d/\delta)}{d}} \right] \leq \delta/2.$$

Proof of Lemma 9: Was seen in class, using the following concentration bound.

Hoeffding's (generalized) inequality: Let $X_1, \ldots, X_n$ be independent random variables where $X_i \in [a_i, b_i]$. Then $X = \sum_i X_i$ satisfies

$$\forall t \geq 0, \qquad \Pr\Big[|X - \mathbb{E}[X]| \geq t \Big] \leq 2e^{-2t^2/\sum_i (b_i - a_i)^2}.$$

Lemma 10: Let $S \in \mathbb{R}^{k \times d}$ be a super-sparse sampling matrix (i.e., each row has a single nonzero entry of value $\sqrt{d/k}$ in a uniformly random location). Then

$$\forall y \in \mathbb{R}^d, \|y\|_2 = 1, \|y\|_\infty \leq \lambda, \qquad \Pr_S[\|Sy\|_2^2 \notin (1 \pm \varepsilon)] \leq 2e^{-2\varepsilon^2 k/(d^2 \lambda^4)}.$$

Exer: Prove this lemma using Hoeffding's inequality.

Proof of Theorem 7: Was seen in class, using Lemmas 9 and 10.