Randomized Algorithms 2023A – Lecture 5
JL Transform and Oblivious Subspace Embedding*

Robert Krauthgamer

1 The JL Transform

JL dimension reduction: We saw the JL lemma which reduces the dimension of n points in $\mathbb{R}^d$. Recall that it uses a random linear map that is drawn obliviously of the data and works with high probability.

Next, we abstract its performance guarantee (ignoring the implementation), because algorithms may have different tradeoffs, e.g., between the target dimension and the runtime. We also change some of the letters (e.g., use $\mathbb{R}^n$ instead of $\mathbb{R}^d$).

Here is a good way to think about the next definition. A matrix $S \in \mathbb{R}^{s\times n}$ is just a linear map $S : \mathbb{R}^n \to \mathbb{R}^s$. It will represent a dimension reduction operation, where b unknown points in $\mathbb{R}^n$ are reduced to points in dimension $s = s(n, b, \varepsilon, \delta)$, and we want this s (the number of rows in S) to be as small as possible. But instead of a single map S, we consider a probability distribution.

Throughout, all vector norms are ℓ_2-norms.

Definition: A random matrix $S \in \mathbb{R}^{s\times n}$ is called an (ε, δ, b)-*Johnson-Lindenstrauss Transform (JLT)* if

$$\forall B \subset \mathbb{R}^n, |B| \le b, \qquad \Pr_S\Big[\forall x \in B,\ \|Sx\| \in (1 \pm \varepsilon)\|x\|\Big] \ge 1 - \delta.$$

We saw in class that a matrix of independent Gaussians (scaled appropriately) attains this guarantee, with a suitable $s = O(\varepsilon^{-2} \log(b/\delta))$. More precisely, we saw it only for $b = 1$, but general b follows easily by applying that result with smaller $\delta' = \delta/b$ and taking a union bound over B.

Notice that the target dimension s does not depend on the ambient dimension n.

We saw also another construction, with bigger target dimension s, but faster matrix-vector multiplication (back then we called it $L = SHD$).

*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.

2 Approximate Matrix Multiplication

Definition: The *Frobenius norm* of a real matrix A is defined as

$$\|A\|_F := (\sum_{i,j} A_{ij}^2)^{1/2}.$$

Problem definition: In *Approximate Matrix Multiplication (AMM)*, the input is $\varepsilon > 0$ and two matrices $A, B \in \mathbb{R}^{n\times m}$, and the goal is to compute a matrix $C \in \mathbb{R}^{m\times m}$ such that

$$\|A^\top B - C\|_F \le \varepsilon\|A\|_F\|B\|_F.$$

Theorem 1: Suppose the matrix $S \in \mathbb{R}^{n\times s}$ is $(\varepsilon', \delta', b')$-JLT, where the parameters satisfy $\varepsilon' = \varepsilon/3$, $\delta' = \delta$, and $b' = O(m^2)$. Then with probability at least $1-\delta$, the matrix $(SA)^\top(SB)$ solves AMM.

Roughly speaking, this theorem reduces the dimension n (of the input matrices) to dimension $s \approx O(\varepsilon^{-2}\log m)$.

Proof: Was seen in class. The main idea is that for fixed $x, y \in \mathbb{R}^n$ with $\|x\| = \|y\| = 1$, we have that

$$2\big[\langle Sx, Sy\rangle - \langle x, y\rangle\big] = \big[\|Sx\|^2 - \|x\|^2\big] + \big[\|Sy\|^2 - \|y\|^2\big] - \big[\|Sx - Sy\|^2 - \|x-y\|^2\big].$$

And now by the JLT guarantee, with high probability $1-\delta'$, each of the three summands is bounded in absolute value.

Remark: The above proof bounds the error in each $C_{i,j}$ (output entry) with high probability, but it clearly suffices to bound the expected squared error, which can be achieved with a smaller matrix S (e.g., no dependence on $b' = O(m^2)$).

3 Oblivious Subspace Embedding

Embedding an entire subspace: In some situations (like regression, as we will see soon), we want a guarantee for a whole subspace, which has infinitely many points.

Observe that a linear subspace $V \subset \mathbb{R}^n$ of dimension d can be described as the column space of $A \in \mathbb{R}^{n\times d}$, i.e., $V = \{Ax : x \in \mathbb{R}^d\}$.

A good way to think about the next definition is that we will solve a problem in $\mathbb{R}^n$ involving an unknown d-dimensional subspace, by reducing the problem to dimension $s = s(n, d, \varepsilon, \delta)$. Thus, we want s (the number of rows in S) to be as small as possible.

Definition: A random matrix $S \in \mathbb{R}^{s\times n}$ is called an (ε, δ, d)-*Oblivious Subspace Embedding (OSE)* if

$$\forall A \in \mathbb{R}^{n\times d}, \qquad \Pr_S\Big[\forall x \in \mathbb{R}^d,\ \|SAx\| \in (1\pm\varepsilon)\|Ax\|\Big] \ge 1-\delta.$$

We next show that it is easy to construct OSE using JLT.

Exer: Show that the OSE property is preserved under right-muliplication by a matrix with orthonormal columns, as follows. If $S \in \mathbb{R}^{s\times n}$ is an (ϵ, δ, d)-OSE matrix, and $U \in \mathbb{R}^{n\times r}$ is a matrix with orthonormal columns, then SU is an $(\epsilon, \delta, \min(r, d))$-OSE matrix (for the space $\mathbb{R}^r$).

Theorem: Let $S \in \mathbb{R}^{s\times n}$ be an (ε, δ, b)-JLT for $\varepsilon < 1/4$. Then S is also an $(O(\varepsilon), \delta, \frac{\ln b}{\ln(1/\varepsilon)})$-OSE.

Remark: To produce OSE for dimension d, we should set in this theorem $d = \frac{\ln b}{\ln(1/\varepsilon)}$, i.e., $b = (1/\varepsilon)^d$, which we can achieve using a Gaussian matrix with $s = O(\varepsilon^{-2}\log(b/\delta)) = O(\varepsilon^{-2}(d\log\frac{1}{\varepsilon} + \log\frac{1}{\delta}))$ rows. A direct construction with sparse columns (and thus fast matrix-vector multiplication) was shown by [Cohen, 2016].

Proof: Was seen in class. The main idea is to use the JLT guarantee on a (3ε)-net N of the unit sphere in $\mathbb{R}^d$, then represent arbitrary unit-length $x \in \mathbb{R}^d$ as an infinite (but converging) sum $x = \sum_{i=0}^{\infty} x_i$, where each x_i is a (scalar) multiple of a net point, and finally use the triangle inequality. We used the next fact, and saw its proof that is based on volume arguments.

Fact: The unit sphere in $\mathbb{R}^d$ has a γ-net N of size $|N| \le (1 + 2/\gamma)^d \le (3/\gamma)^d$.

Remark: It is possible to get a better bound by employing a 1/2-net (instead of ε-net) and expanding $\|SAx\|^2$ including cross terms.