# Randomized Algorithms 2025A – Lecture 5
# JL Transform, Oblivious Subspace Embedding, and Least Squares Regression*

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, i.e., ignore the implementation, because different 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 matrix $S$, we consider a random matrix, which is the same as a probability distribution over matrices.

Throughout, all vector norms are $\ell_2$-norms.

**Definition:** A random matrix $S \in \mathbb{R}^{s\times n}$ is called an $(\varepsilon, \delta, b)$-*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:** One can solve AMM with probability at least $1-\delta$, by taking a matrix $S \in \mathbb{R}^{n\times s}$ that is $(\varepsilon', \delta', b')$-JLT for parameters $\varepsilon' = \varepsilon/9$, $\delta' = \delta$, $b' = O(m^2)$, and reporting the matrix $(SA)^\top (SB)$.

In essence, this theorem reduces the dimension $n$ (of the columns of the input matrices) to dimension $s = O(\varepsilon^{-2} \log m)$.

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

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

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, instead of a finite set of points.

We shall describe a linear subspace $V \subset \mathbb{R}^n$ of dimension $d$ 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 $(\varepsilon, \delta, d)$-*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 $(\epsilon, \delta, d)$-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 2:** Let $S \in \mathbb{R}^{s\times n}$ be an $(\varepsilon, \delta, b)$-JLT for $\varepsilon < 1/4$. Then $S$ is also an $(O(\varepsilon), \delta, \frac{\ln b}{\ln(1/\varepsilon)})$-OSE.

**How to use this theorem:** To produce OSE for a given dimension $d$, we want $d = \frac{\ln b}{\ln(1/\varepsilon)}$, i.e., we should apply the theorem with $b = (1/\varepsilon)^d$. Recall that We can achieve such JLT using a matrix of Gaussians 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 $(3\varepsilon)$-net $N$ of the unit sphere in $\mathbb{R}^d$, then represent an arbitrary unit length $x \in \mathbb{R}^d$ as an infinite (but converging) series $x = \sum_{i=0}^{\infty} x_i$, where each $x_i$ is a multiple of a net point, and finally use the triangle inequality. The crux is that we apply the JLT guarantee only on the net $N$, then "extend" it to all of $\mathbb{R}^n$. 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 $\gamma$-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 $\varepsilon$-net) and expanding $\|SAx\|^2$ including cross terms.

# 4 Least Squares Regression

**Problem definition:** In *Least Squares Regression*, the input is a matrix $A \in \mathbb{R}^{n\times d}$ and a vector $b \in \mathbb{R}^n$, and the goal is to find $x^* := \text{argmin}\{\|Ax - b\| : x \in \mathbb{R}^d\}$.

Informally, when solving a system $Ax = b$ that is over-constrained ($n \gg d$), we do not expect to find an exact solution, and we want to minimize the sum of squared errors $\sum_i (A_i x - b_i)^2$.

We shall consider $(1 + \varepsilon)$-approximation, i.e., finding $x' \in \mathbb{R}^d$ such that

$$\|Ax' - b\| \le (1 + \varepsilon) \min_{x\in\mathbb{R}^d} \|Ax - b\|. \tag{1}$$

**Theorem 3:** Let $S \in \mathbb{R}^{s\times n}$ be an $(\varepsilon, \delta, d + 1)$-OSE matrix. Then for every regression instance $A \in \mathbb{R}^{n\times d}$ and $b \in \mathbb{R}^n$, with high probability, an optimal solution $x'$ (or even $(1+\varepsilon)$-approximation) to the regression instance $\langle SA, Sb\rangle$ is a $(1 + O(\varepsilon))$-approximation to the instance $\langle A, b\rangle$, i.e., such $x'$ satisfies (1).

This theorem essentially reduces a regression problem with $n$ constraints to regression with $s$ constraints, but we should take into account also the time to compute $SA$.

**Proof:** Was seen in class. It basically follows from applying the OSE guarantee to the subspace spanned by the columns of $A$ and by $b$ (total of $d + 1$ vectors), and then

$$(1 - \varepsilon)\|Ax' - b\| \le \|SAx' - Sb\| = \min_{x\in\mathbb{R}^d} \|SAx - Sb\| \le \|SAx^* - Sb\| \le (1 + \varepsilon)\|Ax^* - b\|.$$

Thus with high probability, $(1-\varepsilon)\mathrm{ALG} \leq (1+\varepsilon)\mathrm{OPT}$.