COMS E6998: Algorithms for Massive Data (Fall'25) Sep 15, 2025

Lecture Lecture 4: FDR Proof

Instructor: *Alex Andoni* Scribes: *Hao Cui*

1 Introduction

In this class, we will prove the statement for the Fast Dimension Reduction (also termed FJL: Fast Johnson-Lindenstrauss elsewhere).

2 Fast Dimension Reduction

First, we recall the Fast Dimension Reduction theorem

Theorem 1 (FDR)**.** *There exists distirbution over $\varphi : \mathbb{R}^n \to \mathbb{R}^k$ such that the following holds:*

1. *φ satisfies DJL, in other words, for all $x \in \mathbb{R}^n$ we have*

$$\Pr\left[\frac{||\varphi(x)||}{||x||} \in 1 \pm \epsilon\right] \geq 1 - \delta.$$

2. *$\varphi(x)$ can be computed in time $O(n \log n + k)$.*
3. *$k = O\left(\frac{\log \frac{1}{\delta}}{\epsilon^2} \log \frac{n}{\delta}\right)$.*

3 Proof of FDR

The idea is as follows: we will show that

$$\varphi(x) := \sqrt{\frac{n}{k}} \cdot PHDx$$

achieves the properties we are looking for. Here, the $\sqrt{n/k}$ comes from the renormalization of the vector, and

$$P = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_n \end{pmatrix}$$

is an $k \times n$ matrix such that each row consists of a vector p_i that has entry 1 at a uniformly random index, and all other entries are 0. The random entries in the vectors p_i are independent of each other. Ideally, we would want

$$\frac{\frac{n}{k}||Px||_2^2}{||x||_2^2} \in 1 \pm \epsilon \tag{1}$$

to hold, which will immediately prove the FDR (without the need for HD matrices), since Px is computed by inspecting each vector p_i, finding where the random 1 bit j_i is, and locating the corresponding value in x_{j_i}. This is a process that will take only a runtime of k. However, we note that this is way too optimistic because if x is a "sparse" vector (we will proceed to define sparsity later on), we would expect that $\frac{n}{k}||Px||_2^2$ would deviate a lot from the original norm. However, for vectors with lower sparsity, we are actually ok:

Lemma 2. 1 *holds with probability* $1-\delta$ *if*

$$k = O\left(\frac{\log\frac{1}{\delta}}{\epsilon^2}\cdot\frac{n||x||_\infty^2}{||x||_2^2}\right).$$

In addition, we will define the quantity $\Delta_x := \frac{n||x||_\infty^2}{||x||_2^2}$ *as the sparsity of* x *(higher means more sparse).*

To get some familiarity with the sparsity, if $x = (\pm 1, \cdots, \pm 1)$, then $\Delta_x = \frac{n\cdot 1}{\sum_{i=1}^n 1} = 1$; on the other hand, if $x = (1, 0, \cdots, 0)$, then $\Delta_x = \frac{n\cdot 1}{1} = n$.

Proof. We have that $||Px||^2 = x_{i_1}^2 + \cdots + x_{i_k}^2$, where i_j's are uniform in $[n]$. Let $z_j = x_{i_j}^2$, then each $z_j \in [0, ||x||_\infty^2]$. We will let $y_i = \frac{z_i}{||x||_\infty^2} \in [0,1]$ so that we can apply Chernoff. Let $Y = \sum y_i$, we calculate

$$\begin{aligned}
\mu &= \mathbb{E}\left[\sum_{i=1}^k y_i\right] \\
&= \frac{k}{||x||_\infty^2}\mathbb{E}[z_i] \\
&= \frac{k}{||x||_\infty^2}\sum_{j=1}^n \frac{1}{n}x_j^2 \\
&= \frac{k}{n}\cdot\frac{||x||_2^2}{||x||_\infty^2}.
\end{aligned}$$

By Chernoff, we have

$$\Pr[Y \in (1\pm\epsilon)\mu] \geq 1 - 2e^{-\mu\epsilon^2/9}.$$

For this to imply that $\frac{\frac{n}{k}||Px||_2^2}{||x||_2^2} \in 1\pm\epsilon$ with probability at least $1-\delta$, we want this probability to be lower bounded by at least $1-\delta$, so $\mu = \Omega(\frac{\log\frac{1}{\delta}}{\epsilon^2})$. From the calculation above, this implies that $k = \Omega\left(\frac{\log\frac{1}{\delta}}{\epsilon^2}\cdot\frac{n||x||_\infty^2}{||x||_2^2}\right)$ is enough. □

Our conclusion here is that if Δ_x is small, then P_x is good enough. Therefore, our next idea is to introduce the Hadamard matrix, for which the goal is to reduce Δ_x.

3.1 Hadamard

The Hadamard matrix H is a $n\times n$ matrix that has the following property:

1. H is a rotation: $||Hx|| = ||x||$.

2. $H_{ij} \in \pm\{\frac{1}{\sqrt{n}}\}$.

3. It takes $O(n \log n)$ time to compute Hx.

4. If x is sparse, Hx is dense.

The last property is usually referred to as the Uncertainty Principle. However, it could still be the the output Hx is sparse for some *dense* x's. The way to fix this is to define a random diagonal matrix

$$D = \begin{pmatrix} \pm 1 & 0 & \dots & 0 \\ 0 & \pm 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pm 1 \end{pmatrix}$$

that randomly reflects each coordinate. We will show that this is good enough to randomize x away from "bad cases" that result in sparse Hx.

Lemma 3. *Fix x such that $||x||_2 = 1$, let $y = HDx$, then*

$$\Pr_D \left[||y||_\infty^2 > \frac{\log \frac{n}{\delta}}{n} \right] \leq \delta.$$

This is the final piece of puzzle to the proof of the FDR, we can plug in the value of $||y||_\infty$ back into the statement of Lemma 2 (note that since H is a rotation and D is a reflection we have $||y||_2 = 1$) to get the desired value of k.

Proof. We have that

$$y_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} r_j x_j$$

where each r_j is ± 1 with equal probability. We will simplify the problem by considering

$$D = \begin{pmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_n \end{pmatrix}$$

where each $g_i \in N(0, 1)$. Then, it follows that if $\alpha = \Theta(\sqrt{\log \frac{n}{\delta}})$, then

$$\Pr[|y_i| > \alpha/\sqrt{n}] = \Pr[|g| > \alpha] \simeq e^{-\Omega(\alpha^2)} < \delta/n.$$

We will note that the original case is very similar:

Fact 4. $\forall x_1, \cdots, x_n \in \mathbb{R}$, $r_i = \pm 1$, *we have*

$$\Pr_r \left[\left| \sum_{i=1}^{n} r_i x_i \right| > \frac{1}{\sqrt{n}} \alpha ||x||_2 \right] \leq e^{-\alpha^2/2}$$

If we let $\alpha = \sqrt{2 \ln \frac{n}{\delta}}$, then the above probability will be upper bounded by δ/n, hence by a union bound over all n different y_i's we have finished the proof of the lemma, and hence the proof of FDR. $\square$

4 Extra Notes

1. For LSR, we can generally get runtime as good as

$$O(nnz(A) + (d/\epsilon)^{O(1)})$$

   or

$$O((\log \epsilon^{-1})(nnz(A) + d^{O(1)}))$$

   where A is the targeted matrix, d is the source dimension, ϵ is the error, and $nnz(A)$ denotes the number of non-zero entries in A.

2. DJL and OSE can be extended to ℓ_1 norms and beyond using sketching (such as Cauchy projections).

3. a few rounds of HD (for random D) is often used as a peoxy for random rotations, where each D_i is an i.i.d matrix with ± 1 on the diagonal. Normally we would need to use n^2 random numbers for random rotation, but this gives us way to only use $O(n \log n)$ random numbers.