Randomized Algorithms 2021A – Lecture 6 (second part)
Fast JL Transform*

Robert Krauthgamer

1 The Johnson-Lindenstrauss (JL) Lemma (cont'd)

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

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

Proof of Claim 5: Was seen in class, using the following exercise.

Exer: Prove (by evaluating the integral, and substituting $z = x\sqrt{1-2t}$) that

$$\mathbb{E}[e^{tX_i^2}] = \tfrac{1}{\sqrt{1-2t}}.$$

Exer: Extend the JL Lemma (via the main lemma) to every matrix G whose entries are iid from a distribution that has mean 0, variance 1, and sub-Gaussian tail which means that for some fixed $C > 0$,

$$\forall t > 0, \qquad \mathbb{E}[e^{tX}] \leq e^{Ct^2}.$$

Then 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

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

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

*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 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 $v \in \mathbb{R}^d$,

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

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

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.