Sketching Algorithms for Big Data	Fall 2017

Lecture 23 — November 21, 2017

Prof. Noga Alon *Scribe: Preetum Nakkiran*

1 Overview

This lecture will be on "Optimal Compression of Approximate Inner Products and Dimension Reduction", following the paper of Alon and Klartag [AK17].

We will study the following object:

Definition 1. *An* ε-inner-product sketch *for points* $X \subseteq \mathbb{R}^k$, $|X| = n$, *such that* $\forall x : ||x||_2 \leq 1$, *is a data structure that enables one to compute the inner-product*

$$\langle x, x' \rangle \ \forall x, x' \in X$$

up to additive error ε.

Let $f(n, k, \varepsilon)$ be the minimum size (in bits) of an ε-inner-product sketch for n points in dimension k. (We can assume $n \geq k$ WLOG, and further $\varepsilon \geq 1/n^{0.5-\delta}$, since this is the interesting case for dimensionality reduction).

There are various applications for this data structure, for example in streaming, compressed sensing, and computational geometry (eg, Nearest-Neighbors).

Main Questions: What is $f(n, k, \varepsilon)$? And, can we compute sketches efficiently?

We will first do a warm-up, and then present the Main Theorem of [AK17] which exactly characterizes $f(n, k, \varepsilon)$.

2 Warm-Up

First, let us consider the basic case of when $k = n$. We claim an easy upper-bound here is

$$f(n, n, \varepsilon) \leq O(n \frac{\log n}{\varepsilon^2} \log(1/\varepsilon))$$

To do this:

1. First use JL to project the n points in $\mathbb{R}^n$ down to dimension $m := O(\frac{\log n}{\varepsilon^2})$, with distortion $O(\varepsilon)$.
2. Then, round the projected points to an ε-net of the unit ball in $\mathbb{R}^k$.

The ε-net in Step 2 requires at most $O(\frac{1}{\varepsilon})^{(1+o(1))m}$ points, so we need $O(m \log(1/\varepsilon))$ bits per point. This yields the claim.

In fact, this bound on $f(n, n, \varepsilon)$ is nearly-tight. The $\log(1/\varepsilon)$ factor can be shaved, as first shown by Kushilevitz, Ostrovsky, and Rabani [KOR98]. The construction is simple:

Construction: Let $v_1, \ldots, v_t$ be random unit vectors in $\mathbb{R}^n$, for $t = O(\frac{\log n}{\varepsilon^2})$. For every $x \in X$, maintain just the t signs: $\{\text{sign}(\langle x, v_i \rangle)\}_{i \in [t]}$.

Proof sketch: The idea is, for two vectors x, x', a random unit vector v_i will distinguish them (ie, $sign(\langle x, v_i \rangle) \neq sign(\langle x', v_i \rangle)$) with probability proportional to the angle between x, x'. Thus, maintaining t signs gives a good enough approximation of this angle, and taking the cosine recovers approximately $\langle x, x' \rangle$. □

Note that this construction does not yield better bounds on $f(n, k, \varepsilon)$ for smaller $k < n$.

3 Main Theorem

Theorem 2 ([AK17]). *Let $f(n, k, \varepsilon)$ be the minimum size (in bits) of an ε-inner-product sketch for n points in dimension k.*

Then, $\forall n \geq k, \varepsilon \geq \frac{1}{n^{0.49}}$:

(A) For $\frac{\log n}{\varepsilon^2} \leq k \leq n$: $f(n, k, \varepsilon) = \Theta(\frac{n \log n}{\varepsilon^2})$

(B) For $\log n \leq k \leq \frac{\log n^2}{\varepsilon^2}$: $f(n, k, \varepsilon) = \Theta(nk \log(2 + \frac{\log n}{\varepsilon^2 k}))$

(C) For $1 \leq k \leq \log n$: $f(n, k, \varepsilon) = \Theta(nk \log(1/\varepsilon))$.

Note that these bounds all agree on their common boundaries.

Remark 1. *The Main Theorem considers* additive *error. Recently Indyk and Wagner [IW17] have shown how to achieve* relative *error guarantees (which is harder). Note, there is still a gap of* $\log(1/\varepsilon)$ *between upper and lower bounds here.*

Remark 2. *The Main Theorem gives an alternative proof of a recent result by Larsen and Nelson [LN17]: There does not exists a dimensionality reduction of n points in $\mathbb{R}^n$ to dimension $< \frac{c \log n}{\varepsilon^2}$ for some small constant c.*

This follows because in the regime $k < \frac{\log n}{\varepsilon^2}$, $f(n, k, \varepsilon)$ decays rapidly with k, in particular $f(n, n, 2\varepsilon) > f(n, k, \varepsilon)$. That is, there are simply too many configurations of points in $\mathbb{R}^n$ to be faithfully compressed in $\mathbb{R}^k$.

Note, this also implies that once $k < \frac{n}{\varepsilon^2}$, even dimensionality reduction by a factor of (say) 10 is impossible.

Remark 3. *It turns out that the Main Theorem can be combined with the* Khatri-Sidak Lemma *and* Hargié Inequality *(which are special cases of "Gaussian Correlation"), to work for all $\varepsilon \geq \frac{1}{\sqrt{n}}$. See [AK17] for details.*

For the statement, replace $\log n$ by $\log(2 + \varepsilon^2 n)$ everywhere in the Main Theorem.

4 Proof of Main Theorem

We will now see two proofs of the upper bound (non-constructive and constructive), and one proof of the lower bound.

4.1 Upper Bound

Consider first the regime (A), where $\frac{\log n}{\varepsilon^2} \leq k \leq n$. We wish to show $f(n,k,\varepsilon) = \Theta(\frac{n \log n}{\varepsilon^2})$.

Regime (A) By monotonicity, it is sufficient to show (A) for $k = n$. By JL, we may first project points into dimension $m := \frac{\log n}{\varepsilon^2}$, with distortion $O(\varepsilon)$.

For vectors $w_1, \ldots, w_n \in \mathbb{R}^m$, define the *Gram Matrix* $G(w_1, \ldots, w_n)$ as $G_{i,j} := \langle w_i, w_j \rangle$. Say two Gram Matrices G, G' are *ε-separated* if $\exists i \neq j : \ |G_{i,j} - G'_{i,j}| > \varepsilon$. Let $\mathcal{G}$ be a maximal (w.r.t containment) set of ε-separated Gram Matrices. Then, we clearly have $f(n,k,\varepsilon) \leq \log|\mathcal{G}|$ (simply by remembering, for a point set X, the index of a $g \in \mathcal{G}$ that $G(X)$ is not ε-separated from)

We now want to bound $|\mathcal{G}|$. We will use essentially a volume argument. Let $v_1, \ldots, v_n \in \mathbb{R}^m$ be random vectors, each uniform in a ball of radius 2 about the origin. For all $G(w_1, \ldots w_n) \in \mathcal{G}$, define the event

$$A_G := \{\forall i,j : |\langle v_i, v_j\rangle - \langle w_i, w_j \rangle| \leq \varepsilon/2\}.$$

Notice the events $\{A_G\}_{G \in \mathcal{G}}$ are pairwise disjoint, by ε-separatedness of $\mathcal{G}$. Thus, $\sum_{G\in\mathcal{G}} \mathbb{P}[A_G] \leq 1$. So if the individual probabilities $\mathbb{P}[A_G]$ are large, then there cannot be many such events A_G (thus bounding $|\mathcal{G}|$). It is easy to see that $\mathbb{P}[A_G] \geq \Omega(\varepsilon)^{mn}$ (by noticing that the condition $\{\forall i : ||v_i - w_i||_2 \leq \varepsilon/4\}$ is sufficient). We will show that in fact, a much better bound holds:

$$\mathbb{P}[A_G] \geq \Omega(1)^{mn}.$$

First, let us condition on the event

$$E := \{\forall i : ||v_i - w_i||_2 \leq 1\}.$$

We have $\mathbb{P}[E] = (\frac{1}{2})^{nm}$, by comparing volumes (since we pick w_i uniformly from a ball of radius 2). Moreover, conditioned on this event E, the vector $(v_i - w_i)$ is uniform over a unit ball in $\mathbb{R}^m$. Thus, by concentration (eg Azuma-Hoeffding) we have

$$\forall i \neq j : \mathbb{P}[|\langle v_i - w_i, w_j\rangle| \geq \varepsilon/4 \ \ |E] \leq 2e^{-\Omega(\varepsilon^2 m)} < \frac{1}{2n^2}$$

And symmetrically,

$$\forall i \neq j : \mathbb{P}[|\langle v_i, v_j - w_j\rangle| \geq \varepsilon/4 \ \ |E] \leq 2e^{-\Omega(\varepsilon^2 m)} < \frac{1}{2n^2}$$

This allows us to union bound over all pairs $i \neq j$. In particular, we now know that with probability $\geq \frac{1}{2}(\frac{1}{2})^{nm}$, the following simultaneously holds:

$$\begin{cases} \forall i : ||v_i - w_i||_2 \leq 1 \\ \forall i \neq j : |\langle v_i - w_i, w_j\rangle| < \varepsilon/4 \\ \forall i \neq j : |\langle v_i, v_j - w_j\rangle| < \varepsilon/4 \end{cases}$$

This is sufficient to imply A_G holds, since by the last two conditions and triangle inequality:

$$|\langle v_i, v_j \rangle - \langle w_i, w_j \rangle| = |\langle v_i - w_i, v_j \rangle + \langle w_i, v_j - w_j \rangle| \leq |\langle v_i - w_i, v_j \rangle| + |\langle w_i, v_j - w_j \rangle| \leq \varepsilon/2$$

Thus, $\mathbb{P}[A_G] \geq (\frac{1}{2})^{mn+1}$ and so $|\mathcal{G}| \leq 2^{mn+1}$. And we have $f(n, k, \varepsilon) \leq \log |\mathcal{G}| = O(nm) = O(n\frac{\log n}{\varepsilon^2})$ as desired. □

Remark 4. *What is going on is, once we condition on E, then each w_i is uniform in a unit ball centered around v_i, and so and so by concentration we have $\langle w_i, w_j \rangle \approx \langle v_i, v_j \rangle$.*

Regime (B) This is essentially the same as Regime (A), except without the JL projection.

In particular, for $k = \frac{\delta^2}{\varepsilon^2} \log n$ and $2\varepsilon \leq \delta < 1/2$: Let $\mathcal{G}$ be a maximal set of ε-separated Gram Matrices (of vectors in $\mathbb{R}^k$ with norm ≤ 1). Take $v_1, \ldots, v_n \in \mathbb{R}^k$ random vectors as before, and condition on the event

$$E := \{\forall i : ||v_i - w_i||_2 \leq \delta/40\}$$

Then,

$$\mathbb{P}[E] = (\delta/80)^{kn}$$

and the rest follows as before.

That is, after conditioning on E,

$$\forall i \neq j : \mathbb{P}[|\langle v_i - w_i, w_j \rangle| \geq \varepsilon/4 \ \ |E] \leq 2e^{-\Omega(k\varepsilon^2/\delta^2)} < \frac{1}{2n^2}$$

And we conclude as in part (A). □

Regime (C) Here we can simply round to the closest point in an ε-net for the unit ball. This gives $O(k \log(1/\varepsilon))$ bits per point. □

4.2 Upper Bound: Algorithmic Proof

Here we sketch an alternative, constructive proof of the upper-bound.

Regime (A) Apply JL projection into dimension $m := \frac{\log n}{\varepsilon^2}$, and then do *randomized* rounding to a (1/2)-net. (For example, each coordinate is randomly rounded to an integral multiple of $\frac{1}{\sqrt{2m}}$, preserving its expectation).

By concentration (eg, Chernoff-Hoeffding) the randomized rounding stage does not distort too much w.h.p.

Regime (B) Here we skip the JL projection, and just do randomized rounding of each coordinate to an integral multiple of $\delta/\sqrt{k}$ for $k = \frac{\delta^2}{\varepsilon^2} \log n$.

Regime (C) We simply (deterministically) round each coordinate to the nearest integer multiple of $\varepsilon/\sqrt{k}$.

4.3 Lower Bound

We now describe the main ideas of the lower bound.

The Main Lemma is the following:

Lemma 3. *If* $k = \frac{\delta^2}{200\varepsilon^2} \log n$*, then*

$$f(n, k, \varepsilon) \geq \Omega(nk \log(1/\delta)).$$

Proof Sketch Let N be a maximal δ-packing in $\mathbb{R}^k$ (such that the distance between any two points in N is $\geq \delta$). Note $|N| = (\frac{1}{\delta})^k$.

We can show that w.h.p., a random set R of $(n/2)$ unit vectors in $\mathbb{R}^k$ satisfies:

$$\forall x \neq x' \in N : \exists y \in R : |\langle x, y\rangle - \langle x', y\rangle| > \varepsilon$$

That is, w.h.p. R distinguishes every pair of points in N.

Fix such an R. Then, each (ordered) configuration of R together with $> (n/2)$ points of N requires a *distinct encoding*. Because, any two such configurations must share two points $x, x' \in N$ in common, and then R suffices to distinguish these x, x' – so they cannot be represented by the same encoding. Thus,

$$f(n, k, \varepsilon) \geq \log |N|^{n/2} = \Omega(\frac{n}{2} \log |N|) = \Omega(nk \log(1/\delta))$$

□

References

[AK17] Noga Alon and Bo'az Klartag. Optimal compression of approximate inner products and dimension reduction. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 639–650, 2017.

[IW17] Piotr Indyk and Tal Wagner. Near-optimal (euclidean) metric compression. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 710–723, 2017.

[KOR98] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 614–623, 1998.

[LN17] Kasper Green Larsen and Jelani Nelson. Optimality of the johnson-lindenstrauss lemma. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 633–638, 2017.