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 $\varepsilon$-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 $\varepsilon$.

Let $f(n, k, \varepsilon)$ be the minimum size (in bits) of an $\varepsilon$-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 $\varepsilon$-net of the unit ball in $\mathbb{R}^k$. 

The $\varepsilon$-net in Step 2 requires at most $O\left(\frac{1}{\varepsilon^2}(1+o(1))m\right)$ 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 $t = O\left(\frac{\log n}{\varepsilon^2}\right)$. 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, $\text{sign}(\langle x, v_i \rangle) \neq \text{sign}(\langle x', v_i \rangle)$) 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 $\varepsilon$-inner-product sketch for $n$ points in dimension $k$.

Then, $\forall n \geq k, \varepsilon \geq \frac{1}{\sqrt{n^2}}$:

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

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

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

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 Hargie 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 \( f(n, k, \varepsilon) = \Theta\left(\frac{n\log n}{\varepsilon^2}\right) \).

**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 \( \varepsilon \)-separated if \( \exists i \neq j : |G_{i,j} - G'_{i,j}| > \varepsilon \). Let \( G \) be a maximal (w.r.t containment) set of \( \varepsilon \)-separated Gram Matrices. Then, we clearly have \( f(n, k, \varepsilon) \leq \log |G| \) (simply by remembering, for a point set \( X \), the index of a \( g \in G \) that \( G(X) \) is not \( \varepsilon \)-separated from).

We now want to bound \( |G| \). 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 \( G(w_1, \ldots, w_n) \in G \), 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 \( \{ A_G \} \) are pairwise disjoint, by \( \varepsilon \)-separatedness of \( G \). Thus, \( \sum_{G \in G} \mathbb{P}[A_G] \leq 1 \). So if the individual probabilities \( \mathbb{P}[A_G] \) are large, then there cannot be many such events \( A_G \) (thus bounding \( |G| \)). 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:

\[
\begin{align*}
\forall i : ||v_i - w_i||_2 &\leq 1 \\
\forall i \neq j : |\langle v_i - w_i, w_j \rangle| &< \varepsilon/4 \\
\forall i \neq j : |\langle v_i, v_j - w_j \rangle| &< \varepsilon/4
\end{align*}
\]
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, \( P[A_G] \geq (\frac{1}{2})^{mn+1} \) and so \( |G| \leq 2^{mn+1} \). And we have \( f(n, k, \varepsilon) \leq \log |G| = O(nm) = O(n \frac{\log n}{\varepsilon^2}) \) 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 \( G \) be a maximal set of \( \varepsilon \)-separated Gram Matrices (of vectors in \( \mathbb{R}^k \) with norm \( \leq 1 \)). 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,
\[
P[E] = (\delta/80)^{kn}
\]

and the rest follows as before.

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

And we conclude as in part (A). \( \square \)

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

### 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
4.3 Lower Bound

We now describe the main ideas of the lower bound.

The Main Lemma is the following:

**Lemma 3.** If \( k = \frac{\delta^2}{200^2 \varepsilon^2} \log n \), then
\[
f(n, k, \varepsilon) \geq \Omega(n k \log(1/\delta)).
\]

**Proof Sketch**  Let \( N \) be a maximal \( \delta \)-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 \( \frac{n}{2} \) 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 \( > \frac{n}{2} \) 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,
\[
f(n, k, \varepsilon) \geq \log |N|^{n/2} = \Omega(\frac{n}{2} \log |N|) = \Omega(n k \log(1/\delta))
\]

References


