58th Annual IEEE Symposium on Foundations of Computer Science

Optimal compression of approximate inner products and dimension reduction

Noga Alon
Sackler School of Mathematics
and Blavatnik School of Computer Science
Tel Aviv University
Tel Aviv 69978, Israel
Email: nogaa@tau.ac.il

Bo'az Klartag
Sackler School of Mathematics
Tel Aviv University
Tel Aviv 69978, Israel
and Department of Mathematics
Weizmann Institute of Science
Rehovot 7610001, Israel
Email: klartagb@tau.ac.il

***Abstract*—Let X be a set of n points of norm at most 1 in the Euclidean space R^k, and suppose $\varepsilon > 0$. An ε-distance sketch for X is a data structure that, given any two points of X enables one to recover the square of the (Euclidean) distance between them up to an *additive* error of ε. Let $f(n,k,\varepsilon)$ denote the minimum possible number of bits of such a sketch. Here we determine $f(n,k,\varepsilon)$ up to a constant factor for all $n \geq k \geq 1$ and all $\varepsilon \geq \frac{1}{n^{0.49}}$. Our proof is algorithmic, and provides an efficient algorithm for computing a sketch of size $O(f(n,k,\varepsilon)/n)$ for each point, so that the square of the distance between any two points can be computed from their sketches up to an additive error of ε in time linear in the length of the sketches. We also discuss the case of smaller $\varepsilon > 2/\sqrt{n}$ and obtain some new results about dimension reduction in this range. In particular, we show that for any such ε and any $k \leq t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$ there are configurations of n points in R^k that cannot be embedded in R^ℓ for $\ell < ck$ with c a small absolute positive constant, without distorting some inner products (and distances) by more than ε. On the positive side, we provide a randomized polynomial time algorithm for a bipartite variant of the Johnson-Lindenstrauss lemma in which scalar products are approximated up to an additive error of at most ε. This variant allows a reduction of the dimension down to $O(\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2})$, where n is the number of points.*

***Keywords*-compression scheme; dimension reduction; Gaussian correlation; epsilon-net;**

I. Introduction

A crucial tool in several important algorithms is the ability to generate a compact representation (often called a sketch) of high dimensional data. Examples include streaming algorithms [5], [20], compressed sensing [7] and data structures supporting nearest neighbors search [1], [11]. A natural problem in this area is that of representing a collection of n points in the k-dimensional Euclidean ball in a way that enables one to recover approximately the distances or the inner products between the points. The most basic question about it is the minimum possible number of bits required in such a representation as a function of n, k and the approximation required. Another challenge is to design economic sketches that can be generated efficiently and support efficient procedures for recovering the approximate inner product (or distance) between any two given points.

Consider a sketch that enables one to recover each inner product (or square distance) between any pair of the n points up to an additive error of ε. The Johnson-Lindenstrauss Lemma [14] provides an elegant way to generate such a sketch. The assertion of the lemma is that any set of n points in a Euclidean space can be projected onto a t-dimensional Euclidean space, where $t = \Theta(\frac{\log n}{\varepsilon^2})$, so that all distances and inner products between pairs of points are preserved up to a factor of $1+\varepsilon$. This supplies a sketch obtained by storing the (approximate) coordinates of the projected points. Although the above estimate for t has been recently shown by Larsen and Nelson [18] to be tight up to a constant factor for all $\varepsilon \geq \frac{1}{n^{0.49}}$, improving by a logarithmic factor the estimate in [2], this does not provide a tight estimate for the minimum possible number of bits required for the sketch. The results in Kushilevitz, Ostrovsky and Rabani [17] together with the lower bound in [18], however, determine the minimum possible number of bits required for such a sketch up to a constant factor for all $k \geq \frac{\log n}{\varepsilon^2}$ where $\varepsilon \geq \frac{1}{n^{0.49}}$, leaving a gap in the bounds for smaller dimension k. Our first result here closes this gap.

A. Our contribution

Let X be a set of n points of norm at most 1 in the Euclidean space R^k, and suppose $\varepsilon > 0$. An ε-distance sketch for X is a data structure that, given any two points of X enables one to recover the square of the Euclidean distance between them, and their inner product, up to an *additive* error of ε. Let $f(n,k,\varepsilon)$ denote the minimum possible number of bits of such a sketch. Our first main result is a determination of $f(n,k,\varepsilon)$ up to a constant factor for all $n \geq k \geq 1$ and all $\varepsilon \geq \frac{1}{n^{0.49}}$.

Theorem I.1. *For all n and $\frac{1}{n^{0.49}} \leq \varepsilon \leq 0.1$ the function $f(n,k,\varepsilon)$ satisfies the following*

- *For $\frac{\log n}{\varepsilon^2} \leq k \leq n$,*

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

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.65

IEEE computer society

- *For* $\log n \leq k \leq \frac{\log n}{\varepsilon^2}$,

$$f(n,k,\varepsilon) = \Theta(nk \log(2 + \frac{\log n}{\varepsilon^2 k})).$$

- *For* $1 \leq k \leq \log n$,

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

The proof is algorithmic, and provides an efficient algorithm for computing a sketch of size $O(f(n,k,\varepsilon)/n)$ for each point, so that the square of the distance between any two points can be computed from their sketches up to an additive error of ε in time linear in the length of the sketches. The tight bounds show that if $\varepsilon \geq \frac{1}{n^{0.49}}$ and $\ell \leq c\frac{\log n}{\varepsilon^2}$ for some (small) absolute positive constant c, then $f(n,k,\varepsilon)$ for $k = \frac{\log n}{\varepsilon^2}$ is significantly larger than $f(n,\ell,2\varepsilon)$, supplying an alternative proof of the main result of [18] which shows that the $\frac{\log n}{\varepsilon^2}$ estimate in the Johnson-Lindenstrauss dimension reduction lemma is tight.

An advantage of this alternative proof is that an appropriate adaptation of it works for smaller values of ε, covering all the relevant range. For any $\varepsilon \geq \frac{2}{\sqrt{n}}$, define $t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$. We show that for every $k \leq t$ there is a collection of n points of norm at most 1 in R^k, so that in any embedding of them in dimension ℓ such that no inner product (or distance) between a pair of points is distorted by more than ε, the dimension ℓ must be at least $\Omega(k)$. This extends the main result of [18], where the above is proved only for $\varepsilon \geq \frac{\log^{0.5001} n}{\sqrt{k}}$.

The above result for small values of ε suggests that it may be possible to improve the Johnson-Lindenstrauss Lemma in this range. Indeed, our second main result addresses dimension reduction in this range. Larsen and Nelson [19] asked if for any ε the assertion of the Johnson-Lindenstrauss Lemma can be improved, replacing $\frac{\log n}{\varepsilon^2}$ by $t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$. (Note that this is trivial for $\varepsilon < \frac{1}{\sqrt{n}}$ as in this range $t > n$, and it is true for $\varepsilon > \frac{1}{n^{0.49}}$ as in this case $\log(2+\varepsilon^2 n) = \Theta(\log n)$.) Motivated by this we prove the following bipartite version of this statement.

Theorem I.2. *There exists an absolute positive constant C such that for every vectors $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in R^n$, each of Euclidean norm at most 1, and for every $0 < \varepsilon < 1$ and $t = \lfloor C\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2} \rfloor$ there are vectors $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in R^t$ so that for all i, j*

$$|\langle x_i, y_j \rangle - \langle a_i, b_j \rangle| \leq \varepsilon$$

The proof of the theorem is algorithmic, providing a randomized polynomial time algorithm for computing the vectors x_i, y_j given the vectors a_i, b_j.

B. Related work

As mentioned above, one way to obtain a sketch for the above problem when $k \geq \frac{\log n}{\varepsilon^2}$ and $\varepsilon \geq \frac{1}{n^{0.49}}$ is to apply the Johnson-Lindenstrauss Lemma [14] (see [1] for an efficient implementation) projecting the points into a t-dimensional space, where $t = \Theta(\frac{\log n}{\varepsilon^2})$, and then rounding each point to its closest neighbor in an appropriate ε-net. This provides a sketch of size $O(t \log(1/\varepsilon))$ bits per point, which by the results in [18] is optimal up to a $\log(1/\varepsilon)$ factor for these values of n and k.

A tight upper bound of $O(t)$ bits per point for these values of the parameters, with an efficient recovery procedure, follows from the work of [17]. Their work does not seem to provide tight bounds for smaller values of k.

A very recent paper of Indyk and Wagner [13] addresses the harder problem of approximating the inner products between pairs of points up to a *relative* error of ε, for the special case $k = n$, and determines the minimum number of bits required here up to a factor of $\log(1/\varepsilon)$.

There have been several papers dealing with the tightness of the dimension t in the Johnson-Lindenstrauss lemma, culminating with the recent work of Larsen and Nelson that determines it up to a constant factor for $\varepsilon \geq \frac{1}{n^{0.49}}$ (see [18] and the references therein). For smaller values of ε the situation is more complicated. Our results here, extending the one of [18], show that no reduction to dimension smaller than $t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$ is possible, for any $n \geq k \geq t$ and any $\varepsilon > \frac{2}{\sqrt{n}}$. (For any smaller value of ε, or for any $k < t$ no reduction by more than a constant factor is possible). There is no known improvement in the statement of the Johnson-Lindenstrauss Lemma for small values of ε, and our bipartite version and some related results proved here are the first to suggest that such an improvement may indeed hold.

C. Techniques

Our arguments combine probabilistic and geometric tools. The lower bound for the function $f(n,k,\varepsilon)$ is proved by a probabilistic argument. We provide two proofs of the upper bound. The first is based on a short yet intriguing volume argument. Its main disadvantage is that it is not constructive, and its main advantage is that by combining it with results about Gaussian correlation it can be extended to deal with smaller values of ε as well, for all the relevant range. The second proof is algorithmic and is based on randomized rounding.

The results about improved (bipartite) dimension reduction for small ε are proven using several tools from convex geometry including the low-M^* estimate and the finite volume-ratio theorem (see, e.g., [6]), and basic results about the positive correlation between symmetric convex events with the Gaussian measure. We believe that these tools may be useful in the study of related algorithmic questions in high dimensional geometry.

II. Additional results

Theorem I.1 supplies an alternative proof of the main result of [18] about dimension reduction. For $n \geq k \geq \ell$ and $\varepsilon \geq \frac{1}{n^{0.49}}$ we say that there is an (n,k,ℓ,ε)-Euclidean dimension reduction if for any points $x_1, \ldots, x_n \in R^k$

of norm at most one, there exist points $y_1, \ldots, y_n \in R^\ell$ satisfying

$$\|x_i - x_j\|^2 - \varepsilon \leq \|y_i - y_j\|^2 \leq \|x_i - x_j\|^2 + \varepsilon, \qquad (1)$$

for all $i, j = 1, \ldots, n$.

Corollary II.1. *There exists an absolute positive constant $c > 0$ so that for any $n \geq k > ck \geq \ell$ and for $1/n^{0.49} \leq \varepsilon \leq 0.1$, there is an $(n, k, \ell, \varepsilon)$-Euclidean dimension reduction if and only if $\ell = \Omega(\frac{\log n}{\varepsilon^2})$.*

Moreover, the same holds if we replace additive distortion by multiplicative distortion, i.e., if we replace condition (1) by the following condition

$$(1-\varepsilon) \cdot \|x_i - x_j\|^2 \leq \|y_i - y_j\|^2$$
$$\leq (1+\varepsilon) \cdot \|x_i - x_j\|^2, (i, j = 1, \ldots, n). \qquad (2)$$

Corollary II.1 means that if $k \geq c_1 \log n/\varepsilon^2$, then there is an $(n, k, \varepsilon^{-2} \log n, \varepsilon)$-Euclidean dimension reduction (by the Johnson-Lindenstrauss Lemma), and that if there is an $(n, k, \ell, \varepsilon)$-Euclidean dimension reduction with $\ell = o(k)$ then necessarily $k \geq \ell \geq c_2 \varepsilon^{-2} \log n$, for some absolute constants $c_1, c_2 > 0$.

In Theorem I.1 and Corollary II.1 it is assumed that $\varepsilon \geq \frac{1}{n^{0.49}}$. For smaller ε we can combine some of our techniques with Hargé's Inequality about Gaussian correlation and prove the following extension of Theorem I.1.

Theorem II.2. *For all n and $\varepsilon \geq \frac{2}{\sqrt{n}}$ the function $f(n, k, \varepsilon)$ satisfies the following, where $t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$.*

- *For $t \leq k \leq n$,*

$$\Omega(nt) \leq f(n, k, \varepsilon) \leq O(n \frac{\log n}{\varepsilon^2}).$$

- *For $\log(2 + \varepsilon^2 n) \leq k \leq t$,*

$$f(n, k, \varepsilon) = \Theta(nk \log(2 + \frac{t}{k})).$$

- *For $1 \leq k \leq \log(2 + \varepsilon^2 n)$,*

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

This implies the following result about dimension reduction.

Corollary II.3. *There exists an absolute positive constant $c > 0$ so that for any $n \geq k > ck \geq \ell$ and for all $\varepsilon \geq \frac{2}{\sqrt{n}}$, if there is an $(n, k, \ell, \varepsilon)$-Euclidean dimension reduction then $\ell = \Omega(\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2})$.*

Note that for the range of ε in which $\log(2 + \varepsilon^2 n) = o(\log n)$ the statements of Theorem II.2 and of Corollary II.3 are essentially the ones obtained from those in Theorem I.1 and Corollary II.1 by replacing the term $\frac{\log n}{\varepsilon^2}$ by the expression $t = \frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$. In fact, it is possible that as suggested by Larsen and Nelson [19] for such small values of ε the assertion of the Johnson-Lindenstrauss Lemma can also be improved, replacing $\frac{\log n}{\varepsilon^2}$ by $\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$. Motivated by this we prove a bipartite version of the result, stated as Theorem I.2 in the previous section. We conjecture that the assertion of this theorem can be strengthened, as follows.

Conjecture II.4. *Under the assumptions of Theorem I.2, the conclusion holds together with the further requirement that $\|x_i\| \leq O(1)$ and $\|y_i\| \leq O(1)$ for all $1 \leq i \leq n$.*

Note that the assertion of the conjecture is trivial for $\varepsilon < \sqrt{C/(2n)}$, as in that case $t \geq n$. Note also that for, say, $\varepsilon > 1/n^{0.49}$ the assertion holds by the Johnson-Lindenstrauss Lemma.

We can show that this conjecture, if true, together with our methods here, suffices to establish a tight upper bound up to a constant factor for the number of bits required for maintaining all inner products between n vectors of norm at most 1 in R^n, up to an additive error of ε in each product, for all $\varepsilon \geq \frac{2}{\sqrt{n}}$, closing the gap between the upper and lower bound in the first bullet in Theorem II.2. The conjecture, however, remains open, but we can establish two results supporting it. The first is a proof of the conjecture when t is $n/2$ (or more generally $\Omega(n)$, that is, the case $\varepsilon = \Theta(1/\sqrt{n})$). Our result is as follows:

Theorem II.5. *Let $m \geq n \geq 1, \varepsilon > 0$ and assume that $a_1, \ldots, a_m, b_1, \ldots, b_m \in R^{2n}$ are points of norm at most one. Suppose that $X_1, \ldots, X_m, Y_1, \ldots, Y_m \in R^n$ are independent random vectors, distributed according to standard Gaussian law. Set $\bar{X}_i = X_i/\sqrt{n}$ and $\bar{Y}_i = Y_i/\sqrt{n}$ for all i.*

Assume that $n \geq C_1 \frac{\log(2+\varepsilon^2 m)}{\varepsilon^2}$. Then with probability of at least $\exp(-C_2 nm)$,

$$\left|\langle \bar{X}_i, \bar{Y}_j \rangle - \langle a_i, b_j \rangle\right| \leq \varepsilon, \text{for } i, j = 1, \ldots, m,$$

and moreover $\|\bar{X}_i\| + \|\bar{Y}_i\| \leq C_3$ for all i. Here, $C_1, C_2, C_3 > 0$ are universal constants.

The second result (stated as Theorem VI.1 below) is an estimate, up to a constant factor, of the number of bits required to represent, for a given set of n vectors $a_1, a_2, \ldots, a_n \in R^k$, each of norm at most 1, the sequence of all inner products $\langle a_i, y \rangle$ with a vector y of norm at most 1 in R^k up to an additive error of ε in each such product. This estimate is the same, up to a constant factor, for all dimensions k with $t \leq k \leq n$ and t as in Theorem I.2, as should be expected from the assertion of the Conjecture.

The remainder of this paper is structured as follows. In Section III we provide our first proof of the upper bound in Theorem I.1, which is based on a short probabilistic (or volume) argument. The second proof, presented in Section IV, is algorithmic. It provides an efficient randomized algorithm for computing a sketch consisting of $O(f(n, k, \varepsilon)/n)$ bits for each point of X, so that the square of the distance

between any two points can be recovered, up to an additive error of ε, from their sketches, in time linear in the length of the sketches. Section V is concerned with the lower bound in Theorem I.1. The results on smaller ε, with the exception of Theorem II.5 whose proof is postponed to the final version of the paper due to space limitations (see also [3]), are proven in Section VI using several tools from convex geometry. The final section 7 contains some concluding remarks and open problems.

Throughout the proofs we make no serious attempt to optimize the absolute constants involved. We write $c, \tilde{C}, c_1, \ldots$ etc. for various positive universal constants, whose values may change from one line to the next. We usually use upper-case C to denote universal constants that we consider "sufficiently large", and lower-case c to denote universal constants that are sufficiently small. For convenience we sometimes bound $f(n,k,2\varepsilon)$ or $f(n,k,5\varepsilon)$ instead of $f(n,k,\varepsilon)$, the corresponding bounds for $f(n,k,\varepsilon)$ follow, of course, by replacing ε by $\varepsilon/2$ or $\varepsilon/5$ in the expressions we get, changing the estimates only by a constant factor. All logarithms are in the natural basis e unless otherwise specified.

III. The upper bound

It is convenient to split the proof of the upper bound in Theorem I.1 into three lemmas, dealing with the different ranges of k. The proof of the upper bound in Theorem II.2, presented in Section VI, combines a similar reasoning with results of Khatri, Sidak [15], [22] and Hargé [10] about the Gaussian correlation Inequality.

Lemma III.1. *For* $\frac{\log n}{\varepsilon^2} \leq k \leq n$,

$$f(n,k,5\varepsilon) \leq O(\frac{n \log n}{\varepsilon^2}).$$

Proof: Since $f(n,k,5\varepsilon)$ is clearly a monotone non-decreasing function of k, it suffices to prove the upper bound for $k = n$. By the Johnson-Lindenstrauss Lemma we can replace the points of $X \subset B^k$, where B^k is the unit ball in R^k, by points in R^m where $m = C\frac{\log n}{\varepsilon^2}$ so that all distances and norms of the points change by at most ε. Hence we may and will assume that our set of points X lies in R^m. Note that given the squares of the norms of two vectors up to an additive error of ε and given their inner product up to an additive error of ε we get an approximation of the square of their distance up to an additive error of 4ε. It thus suffices to show the existence of a sketch that can provide the approximate norm of each of our vectors and the approximate inner products between pairs. The approximate norms can be stored trivially by $O(\log(1/\varepsilon))$ bits per vector. (Note that here the cost for storing even a much better approximation for the norms is negligible, so if the constants are important we can ensure that the norms are known with almost no error). It remains to prepare a sketch for the inner products.

The Gram matrix $G(w_1, w_2, \ldots, w_n)$ of n vectors $w_1, \ldots, w_n$ is the n by n matrix G given by $G(i,j) = \langle w_i, w_j \rangle$. We say that two Gram matrices G_1, G_2 are ε-separated if there are two indices $i \neq j$ so that $|G_1(i,j) - G_2(i,j)| > \varepsilon$. Let $\mathcal{G}$ be a maximal (with respect to containment) set of Gram matrices of ordered sequences of n vectors $w_1, \ldots, w_n$ in R^m, where the norm of each vector w_i is at most 2, so that every two distinct members of $\mathcal{G}$ are ε-separated. Note that by the maximality of $\mathcal{G}$, for every Gram matrix M of n vectors of norms at most 2 in R^m there is a member of $\mathcal{G}$ in which all inner products of pairs of distinct points are within ε of the corresponding inner products in M, meaning that as a sketch for M it suffices to store (besides the approximate norms of the vectors), the index of an appropriate member of $\mathcal{G}$. This requires $\log |\mathcal{G}|$ bits. It remains to prove an upper bound for the cardinality of $\mathcal{G}$. We proceed with that.

Let $V_1, V_2, \ldots, V_n$ be n vectors, each chosen randomly, independently and uniformly in the ball of radius 3 in R^m centered at 0. Let $T = G(V_1, V_2, \ldots, V_n)$ be the Gram matrix of the vectors V_i. For each $G \in \mathcal{G}$ let A_G denote the event that for every $1 \leq i \neq j \leq n$, $|T(i,j) - G(i,j)| < \varepsilon/2$. Note that since the members of $\mathcal{G}$ are ε-separated, all the events A_G for $G \in \mathcal{G}$ are pairwise disjoint. We claim that the probability of each event A_G is at least $0.5(1/3)^{mn}$. Indeed, fix a Gram matrix $G = G(w_1, \ldots, w_n) \in \mathcal{G}$ for some $w_1, \ldots, w_n \in R^m$ of norm at most 2. For each fixed i the probability that V_i lies in the unit ball centered at w_i is exactly $(1/3)^m$. Therefore the probability that this happens for all i is exactly $(1/3)^{nm}$. The crucial observation is that conditioning on that, each vector V_i is uniformly distributed in the unit ball centered at w_i. Therefore, after the conditioning, for each $i \neq j$ the probability that the inner product $\langle V_i - w_i, w_j \rangle$ has absolute value at least $\varepsilon/4$ is at most $2e^{-\varepsilon^2 m/64} < 1/(2n^2)$. (Here we used the fact that the norm of w_j is at most 2 and that the constant C in the definition of m is sufficiently large). Similarly, since the norm of V_i is at most 3, the probability that the inner product $\langle V_i, V_j - w_j \rangle$ has absolute value at least $\varepsilon/4$ is at most $2e^{-\varepsilon^2 m/96} < 1/2n^2$. It follows that with probability bigger than $0.5(1/3)^{nm}$ all these inner products are smaller than $\varepsilon/4$, implying that

$$|\langle V_i, V_j \rangle - \langle w_i, w_j \rangle| \leq |\langle V_i - w_i, w_j \rangle| + |\langle V_i, V_j - w_j \rangle| < \varepsilon/2.$$

This proves that the probability of each event A_G is at least $0.5(1/3)^{nm}$, and as these are pairwise disjoint their number is at most $2 \cdot 3^{nm}$, completing the proof of the lemma. ■

Lemma III.2. *For* $\log n \leq k \leq \frac{\log n}{\varepsilon^2}$,

$$f(n,k,4\varepsilon) \leq O(nk \log(2 + \frac{\log n}{\varepsilon^2 k})).$$

Proof: The proof is nearly identical to the one of the previous lemma. Note, first, that by monotonicity and the

fact that the expression in the right hand side of the statement of the lemma changes only by a constant factor when ε changes by a constant factor, it suffices to prove the required bound for $k = \frac{\delta^2}{\varepsilon^2} \log n$ where $2\varepsilon \leq \delta \leq 1/2$. Let $\mathcal{G}$ be a maximal set of ε-separated Gram matrices of n vectors of norm at most 1 in R^k. (Here it suffices to deal with norm 1 as we do not need to start with the Johnson-Lindenstrauss Lemma which may slightly increase norms). In order to prove an upper bound for $\mathcal{G}$ consider, as before, a fixed Gram matrix $G = G(w_1, \ldots, w_n)$ of n vectors of norm at most 1 in R^k. Let $V_1, V_2, \ldots, V_n$ be random vectors distributed uniformly and independently in the ball of radius 2 in R^k, let T denote their Gram matrix, and let A_G be, as before, the event that $T(i,j)$ and $G(i,j)$ differ by less than $\varepsilon/2$ in each non-diagonal entry. The probability that each V_i lies in the ball of radius, say, $\delta/20$ centered at w_i is exactly $(\delta/40)^{kn}$. Conditioning on that, the probability that the inner product $\langle V_i - w_i, w_j \rangle$ has absolute value at least $\varepsilon/4$ is at most

$$2e^{-\varepsilon^2 400k/32\delta^2} < 1/(2n^2).$$

Similarly, the probability that the inner product $\langle V_i, V_j - w_j \rangle$ has absolute value at least $\varepsilon/4$ is at most

$$2e^{-\varepsilon^2 400k/64\delta^2} < 1/2n^2.$$

As before, this implies that $|\mathcal{G}| \leq 2(40/\delta)^{kn}$, establishing the assertion of the lemma. ■

Lemma III.3. *For $k \leq \log n$,*

$$f(n,k,\varepsilon) \leq O(nk \log(1/\varepsilon)).$$

Proof: Fix an $\varepsilon/2$-net of size $(1/\varepsilon)^{O(k)}$ in the unit ball in R^k. The sketch here is simply obtained by representing each point by the index of its closest neighbor in the net. ■

IV. An algorithmic proof

In this section we present an algorithmic proof of the upper bound of Theorem I.1. We first reformulate the theorem in its algorithmic version. Note that the first part also follows from the results in [17].

Theorem IV.1. *For all n and $\frac{1}{n^{0.49}} \leq \varepsilon \leq 0.1$ there is a randomized algorithm that given a set of n points in the k-dimensional unit ball B^k computes, for each point, a sketch of $g(n,k,\varepsilon)$ bits. Given two sketches, the square of the distance between the points can be recovered up to an additive error of ε in time $O(\frac{\log n}{\varepsilon^2})$ for $\frac{\log n}{\varepsilon^2} \leq k \leq n$ and in time $O(k)$ for all smaller k. The function $g(n,k,\varepsilon)$ satisfies the following*

- *For $\frac{\log n}{\varepsilon^2} \leq k \leq n$,*

$$g(n,k,\varepsilon) = \Theta(\frac{\log n}{\varepsilon^2})$$

and the sketch for a given point can be computed in time $O(k \log k + \log^3 n/\varepsilon^2)$.

- *For $\log n \leq k \leq \frac{\log n}{\varepsilon^2}$,*

$$g(n,k,\varepsilon) = \Theta(k \log(2 + \frac{\log n}{\varepsilon^2 k})).$$

and the sketch for a given point can be computed in time linear in its length.

- *For $1 \leq k \leq \log n$,*

$$g(n,k,\varepsilon) = \Theta(k \log(1/\varepsilon))$$

and the sketch for a given point can be computed in time linear in its length.

In all cases the length of the sketch is optimal up to a constant factor.

As before, it is convenient to deal with the different possible ranges for k separately. Note first that the proof given in Section III for the range $k \leq \log n$ is essentially constructive, since it is well known (see, for example [4] or the argument below) that there are explicit constructions of ε-nets of size $(1/\varepsilon)^{O(k)}$ in B^k, and it is enough to round each vector to a point of the net which is ε-close to it (and not necessarily to its nearest neighbor).

For completeness we include a short description of a δ-net which will also be used later. For $0 < \delta < 1/4$ and for $k \geq 1$ let $N = N(k,\delta)$ denote the set of all vectors of Euclidean norm at most 1 in which every coordinate is an integral multiple of $\frac{\delta}{\sqrt{k}}$. Note that each member of N can be represented by k signs and k non-negative integers n_i whose sum of squares is at most k/δ^2. Representing each number by its binary representation (or by two bits, say, if it is 0 or 1) requires at most $2k + \sum_i \log_2 n_i$ bits, where the summation is over all positive n_i. Note that $\sum_i \log_2 n_i = 0.5 \log_2(\Pi_i n_i^2)$ which is maximized when all numbers are equal and gives an upper bound of $k \log_2(1/\delta) + 2k$ bits per member of the net. Given a vector in B^k we can round it to a vector of the net that lies within distance $\delta/2$ from it by simply rounding each coordinate to the closest integral multiple of $\delta/\sqrt{k}$. The computation of the distance between two points of the net takes time $O(k)$. The size of the net is $(1/\delta)^k 2^{O(k)}$, as each point is represented by $k \log_2(1/\delta) + 2k$ bits and k signs.

The above description of the net suffices to prove Theorem IV.1 for $k \leq \log n$. We proceed with the proof for larger k.

For $k \geq \frac{40 \log n}{\varepsilon^2}$ we first apply the Johnson-Lindenstrauss Lemma (with the fast version described in [1]) to project the points to R^m for $m = 40 \log n/\varepsilon^2$ without changing any square distance or norm by more than ε. It is convenient to now shrink all vectors by a factor of $1 - \varepsilon$ ensuring they all lie in the unit ball B^m while the square distances, norms and inner products are still within 3ε of their original values. We thus may assume from now on that all vectors lie in B^m.

As done in Section III, we handle norms separately, namely, the sketch of each vector contains some $O(\log(1/\varepsilon))$ bits representing a good approximation for its

norms. The rest of the sketch, which is its main part, will be used for recovering approximate inner products between vectors. This is done by replacing each of our vectors w_i by a randomized rounding of it chosen as follows. Each coordinate of the vector, randomly and independently, is rounded to one of the two closest integral multiples of $1/\sqrt{m}$, where the probabilities are chosen so that its expectation is the original value of the coordinate. Thus, if the value of a coordinate is $(i+p)/\sqrt{m}$ with $0 \leq p \leq 1$ it is rounded to $i/\sqrt{m}$ with probability $(1-p)$ and to $(i+1)/\sqrt{m}$ with probability p. Let V_i be the random vector obtained from w_i in this way. Then the expectation of each coordinate of $V_i - w_i$ is zero. For each $j \neq i$ the random variable $\langle V_i - w_i, w_j \rangle$ is a sum of m independent random variables where the expectation of each of them is 0 and the sum of squares of the difference between the maximum value of each random variable and its minimum value is the square of the norm of w_j divided by m. Therefore this sum is at most $1/m$, and by Hoeffding's Inequality (see [12], Theorem 2) the probability that this inner product is in absolute value at least $\varepsilon/2$ is at most $2e^{-\varepsilon^2 m/8}$ which is smaller than $1/n^5$. Similar reasoning shows that the probability that $\langle V_i, V_j - w_j \rangle$ is of absolute value at least $\varepsilon/2$ is smaller than $1/n^5$. As in the proof in Section III, it follows that with probability at least $1 - 2/n^3$ all inner products of distinct vectors in our rounded set lie within ε of their original values, as needed. The claims about the running time follow from [1] and the description above. This completes the proof of the first part of Theorem IV.1.

The proof of the second part is essentially identical (without the projection step using the Johnson-Lindenstrauss Lemma). The only difference is in the parameters. If $k = \frac{40\delta^2 \log n}{\varepsilon^2}$ with $\varepsilon \leq \delta \leq 1/2$ we round each coordinate randomly to one of the two closest integral multiples of $\delta/\sqrt{k}$, ensuring the expectation will be the original value of the coordinate. The desired result follows as before, from the Hoeffding Inequality. This completes the proof of Theorem IV.1.

V. The lower bound

Lemma V.1. *If*

$$k = \delta^2 \log n/(200\varepsilon^2)$$

where $2\varepsilon \leq \delta \leq 1/2$, *then* $f(n,k,\varepsilon/2) \geq \Omega(kn \log(1/\delta))$

Proof: Fix a maximal set of points N in the unit ball B^k of R^k so that the Euclidean distance between any two of them is at least δ. It is easy and well known that the size of N is $(1/\delta)^{(1+o(1))k}$ (where the $o(1)$-term tends to 0 as δ tends to 0). For the lower bound we construct a large number of ε-separated Gram matrices of n vectors in B^k. Each collection of n vectors consists of a fixed set R of $n/2$ vectors, whose existence is proved below, together with $n/2$ points of the set N. The set R of fixed points will ensure that all the corresponding Gram matrices are ε-separated.

We claim that there is a choice of a set R of $n/2$ points in B^k so that the inner products of any two distinct points from N with some point of R differ by more than ε. Indeed, for any two fixed points of N, the difference between them has norm at least δ, hence the probability that the product of a random point of B^k with this difference is bigger, in absolute value, than ε is at least, say, $e^{-1.5\varepsilon^2 k/\delta^2}$ (with room to spare). It thus suffices to have

$$(1 - e^{-1.5\varepsilon^2 k/\delta^2})^{n/2} < 1/|N|^2$$

hence the following will do:

$$(n/2)e^{-2\varepsilon^2 k/\delta^2} > (2 + o(1))k \log(1/\delta).$$

Thus it suffices to have

$$2\varepsilon^2 k/\delta^2 < \log(n/5k \log(1/\delta))$$

and as the left hand side is equal to $(\log n)/100$ this indeed holds. Thus a set R with the desired properties exists. (Note that here we used the assumption that $\varepsilon \geq \frac{1}{n^{0.49}}$ to conclude that the right-hand-side is at least $(0.02 - o(1)) \log n$.)

Fix a set R as above. Note that every two distinct choices of ordered sets of $n/2$ members of N provide ε-separated Gram matrices. This implies that

$$f(n,k,\varepsilon/2) \geq \log |N|^{n/2}$$

$$= \Omega(n \log |N|) = \Omega(nk \log(1/\delta)),$$

completing the proof of the lemma. ∎

By monotonicity and the case $\delta = 1/2$ in the above Lemma the desired lower bound in Theorem I.1 for all $k \geq \log n$ follows.

It remains to deal with smaller k. Here we fix a set N of size $(1/2\varepsilon)^{(1+o(1))k}$ in B^k so that the distance between any two points is at least 2ε. As before, the inner products with all members of a random set R of $n/2$ points distinguishes, with high probability, between any two members of N by more than ε. Fixing R and adding to it in all possible ways an ordered set of $n/2$ members of N we conclude that in this range

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

completing the proof of the lower bound and hence that of Theorem I.1.

We conclude this section by observing that the proof of the lower bound implies that the size of the sketch per point given by Theorem IV.1 is tight, up to a constant factor, for all admissible values of the parameters. Indeed, in the lower bounds we always have a fixed set R of $n/2$ points and a large net N, so that if our set contains all the points of R then no two distinct points of N can have the same sketch, as for any two distinct $u, v \in N$ there is a member of R whose inner products with u and with v differ by more than ε. The lower bound for the length of the sketch is thus $\log |N|$, by the pigeonhole principle.

VI. Small distortion

In this section we prove several results related the case of smaller ε. In Section VI-A we prove a tight estimate for the number of bits needed to represent ε-approximations of all inner products $\langle a_1, y\rangle, \ldots, \langle a_n, y\rangle$ for a vector $y \in R^k$ of norm at most 1, where $a_1, a_2, \ldots, a_n \in R^k$ are fixed vectors of norm at most 1. In Section VI-B we present the proof of Theorem II.2. The proof of Theorem II.5 is postponed to the final version of the paper, due to space limitations (see also [3]). In Section VI-C we prove Theorem I.2. The techniques here are more sophisticated than those in the previous sections, and rely on several tools from convex geometry.

A. Inner products with fixed vectors

Theorem VI.1. *Let $a_1, a_2, \ldots, a_n$ be vectors of norm at most 1 in R^k. Suppose $\varepsilon \geq \frac{2}{\sqrt{n}}$ and assume that*

$$\frac{\log(2+\varepsilon^2 n)}{8\varepsilon^2} \leq k \leq n.$$

Then, for a vector y of norm at most 1 the number of bits required to represent all inner products $\langle a_i, y\rangle$ for all $1 \leq i \leq n$ up to an additive error of ε in each such product is

$$\Theta\left(\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}\right).$$

Equivalently, the number of possibilities of the vector

$$\left(\lfloor\frac{\langle a_1, y\rangle}{\varepsilon}\rfloor, \lfloor\frac{\langle a_2, y\rangle}{\varepsilon}\rfloor, \cdots, \lfloor\frac{\langle a_n, y\rangle}{\varepsilon}\rfloor\right)$$

for vectors y of norm at most 1 is

$$2^{\Theta(\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2})}.$$

Proof: As the number of bits required is clearly a monotone non-decreasing function of the dimension it suffices to prove the upper bound for $k = n$ and the lower bound for $k = \frac{\log(2+\varepsilon^2 n)}{8\varepsilon^2}$.

We start with the upper bound. Define $t > 0$ by the equation

$$\varepsilon = \frac{\sqrt{2\log(2+n/t)}}{\sqrt{t}}.$$

(There is a unique solution as the right hand side is a decreasing function of t). Therefore

$$t = \frac{2\log(2+n/t)}{\varepsilon^2}.$$

Since $\varepsilon \geq \frac{2}{\sqrt{n}}$ this implies that $t < n$ since otherwise the right hand side is at most $2\log 3 \cdot n/4 < n$. By the last expression for t, $t \geq \frac{1}{\varepsilon^2}$ and thus $\log(2+n/t) \leq \log(2+\varepsilon^2 n)$ implying that

$$t \leq \frac{2\log(2+\varepsilon^2 n)}{\varepsilon^2}.$$

This implies that

$$\frac{n}{t} \geq \frac{\varepsilon^2 n}{2\log(2+\varepsilon^2 n)}$$

and since $\varepsilon^2 n \geq 4$ it follows that

$$\log(2+n/t) \geq \frac{1}{4}\log(2+\varepsilon^2 n),$$

as can be shown by checking that for $z \geq 4$,

$$2 + \frac{z}{2\log(2+z)} \geq (2+z)^{1/4}.$$

We have thus shown that

$$\frac{\log(2+\varepsilon^2 n)}{2\varepsilon^2} \leq t \leq \frac{2\log(2+\varepsilon^2 n)}{\varepsilon^2}.$$

Define a convex set K in R^n as follows.

$$K = \{x \in R^n : \ |\langle\frac{x}{\sqrt{t}}, a_i\rangle| \leq \varepsilon \ \text{ for all } \ 1 \leq i \leq n\}.$$

By the Khatri-Sidak Lemma ([15], [22], see also [8] for a simple proof), if γ_n denotes the standard Gaussian measure in R^n, then

$$\gamma_n(K) \geq \prod_{i=1}^{n} \gamma_n(\{x \in R^n : \ |\langle\frac{x}{\sqrt{t}}, a_i\rangle| \leq \varepsilon\})$$

$$\geq (1 - 2e^{-\varepsilon^2 t/2})^n$$

$$\geq (1 - 2e^{-\log(2+n/t)})^n = (1 - \frac{2t}{2t+n})^n \geq e^{-3t}.$$

For every measurable centrally symmetric set A in R^n and for any vector $x \in R^n$,

$$\gamma_n(x+A) \geq e^{-\|x\|^2/2}\gamma_n(A).$$

For completeness we repeat the standard argument.

$$\gamma_n(x+A) = \int_A e^{-\|x+y\|^2/2}\frac{1}{(2\pi)^{n/2}}dy$$

$$= e^{-\|x\|^2/2}\gamma_n(A)\int_A e^{-\langle x,y\rangle}e^{-\|y\|^2/2}\frac{1}{\gamma_n(A)(2\pi)^{n/2}}dy.$$

The integral in the right hand side is the expectation, with respect to the Gaussian measure on A, of $e^{-\langle x,y\rangle}$. By Jensen's Inequality this is at least e^z where z is the expectation of $-\langle x, y\rangle$ over A. As $A = -A$ this last expectation is 0 and as $e^0 = 1$ we conclude that $\gamma_n(x+A) \geq e^{-\|x\|^2/2}\gamma_n(A)$, as needed. Taking A as the set K defined above and letting x be any vector b of norm at most 1 in R^n we get

$$\gamma_n(\sqrt{t}b + K) \geq e^{-t/2}\gamma_n(K) > e^{-4t}.$$

Given a vector $b \in R^n$, $\|b\| \leq 1$, let X be a standard random Gaussian in R^n. We bound from below the probability of the event E_b that for every i, $1 \leq i \leq n$,

$$|\langle\frac{X}{\sqrt{t}}, a_i\rangle - \langle b, a_i\rangle| \leq \varepsilon.$$

This, however, is exactly the probability that $X - b\sqrt{t} \in K$, that is, $\gamma_n(\sqrt{t}b + K)$ which as we have seen is at least e^{-4t}.

We can now complete the proof of the upper bound as done in Section III. Let B be a maximum collection of vectors of norm at most 1 in R^n so that for every two distinct $b, b' \in B$ there is some i so that $|\langle b, a_i\rangle - \langle b', a_i\rangle| > 2\varepsilon$. Then the events E_b for $b \in B$ are pairwise disjoint and hence the sum of their probabilities is at most 1. It follows that $|B| \leq e^{4t}$. The upper bound follows as the number of bits needed to represent all inner products $\langle b, a_i\rangle$ for $1 \leq i \leq n$ up to an additive error of 2ε is at most $\lceil \log_2 |B| \rceil$.

We proceed with the proof of the lower bound, following the reasoning in Section V. Put

$$k = \frac{\log(2 + \varepsilon^2 n)}{8\varepsilon^2}.$$

Let B be a collection of, say, $e^{k/8}$ unit vectors in R^k so that the Euclidean distance between any two of them is at least $1/2$. We claim that there are n unit vectors a_i in R^k so that for any two distinct members b, b' of B there is an i so that $|\langle b, a_i\rangle - \langle b', a_i\rangle| > \varepsilon$.

Indeed, taking the vectors a_i randomly, independently and uniformly in the unit ball of R^k the probability that for a fixed pair b, b' the above fails is at most

$$(1 - e^{-4\varepsilon^2 k})^n.$$

Our choice of parameters ensures that

$$\binom{|B|}{2}(1 - e^{-4\varepsilon^2 k})^n < 1.$$

Indeed it suffices to check that

$$e^{-4\varepsilon^2 k} \cdot n > k/4$$

that is $4\varepsilon^2 k < \log(4n/k)$ or

$$k < \frac{\log(4n/k)}{4\varepsilon^2}.$$

It thus suffices to check that

$$\log(2 + \varepsilon^2 n) < 2\log(4n/k) = 2\log(\frac{32\varepsilon^2 n}{\log(2 + \varepsilon^2 n)}).$$

This easily holds since for $\varepsilon \geq 2/\sqrt{n}$,

$$2\log(\frac{32\varepsilon^2 n}{\log(2 + \varepsilon^2 n)}) > \log(2 + \varepsilon^2 n).$$

By the union bound the assertion of the claim follows, implying the desired lower bound as no two members of B can have the same representation. This completes the proof of the theorem. ∎

B. Compression schemes

In this subsection we prove Theorem II.2. The basic approach is similar to the one in the proof of Theorem I.1, the main difference is that in the upper bound proved in Lemma III.1 we replace the simple union bound by a more sophisticated geometric argument based on Hargé's Inequality, which is a special case of the Gaussian correlation conjecture, proved recently by Royen. We start with the following Lemma.

Lemma VI.2. *Let $H_1, \ldots, H_n \subseteq R^k$ be symmetric slabs, where a symmetric slab is a set of the form $\{x \in R^k ; |\langle x, \theta\rangle| \leq 1\}$ for some $\theta \in R^k$. Then,*

$$\frac{Vol_k\left(B^k \cap \bigcap_{i=1}^n H_i\right)}{Vol_k(B^k)} \geq c^k \prod_{i=1}^n \gamma_k(\sqrt{k}H_i),$$

where $c > 0$ is an absolute constant.

Proof: Since $T = \sqrt{k}\bigcap_{i=1}^n H_i$ is convex and centrally-symmetric, we may use Hargé's inequality [10], which is a particular case of the Gaussian correlation inequality proven by Royen [21]. This implies that

$$\gamma_k\left(\sqrt{k}\left(B^k \cap \bigcap_{i=1}^n H_i\right)\right)$$
$$\geq \gamma_k(\sqrt{k}B^k) \cdot \gamma_k\left(\sqrt{k}\bigcap_{i=1}^n H_i\right)$$
$$\geq c\prod_{i=1}^n \gamma_k(\sqrt{k}H_i)$$

where the last passage is the Khatri-Sidak lemma. However,

$$\frac{Vol_k\left(B^k \cap \bigcap_{i=1}^n H_i\right)}{Vol_k(B^k)}$$
$$= \frac{Vol_k\left(\sqrt{k}\left(B^k \cap \bigcap_{i=1}^n H_i\right)\right)}{Vol_k(\sqrt{k}B^k)}$$
$$\geq \frac{\gamma_k\left(\sqrt{k}\left(B^k \cap \bigcap_{i=1}^n H_i\right)\right)(2\pi)^{k/2}}{Vol_k(\sqrt{k}B^k)}$$

since the density of γ_k is at most $(2\pi)^{-k/2}$. Since $Vol_k(\sqrt{k}B^k) \leq C^k$, the lemma is proven. ∎

We proceed with the proof of the upper bound in Theorem II.2. For $t \leq k \leq n$ the upper bound (which is probably not tight) is proved by repeating the proof of Lemma III.1 as it is. For $1 \leq k \leq \log(2 + \varepsilon^2 n)$ the upper bound follows by rounding each vector to the closest point in an ε-net in the ball B^k. It remains to deal with the interesting range $\log(2 + \varepsilon^2 n) \leq k \leq t$. By the computation in the beginning of the proof of Theorem VI.1,

$$\varepsilon = \Theta(\frac{\sqrt{2\log(2 + n/t)}}{\sqrt{t}}).$$

Suppose $k = \delta t$, with $\varepsilon^2 \leq \delta \leq b$ for some small absolute positive constant b. Given points $w_1, \ldots, w_n$ in B^k, as in the proof of Lemma III.1 it suffices to prepare a sketch for the inner products between pairs of distinct points. Again, as in that proof, let $\mathcal{G}$ be a maximal (with respect to containment) set of Gram matrices of ordered sequences of n vectors $w_1, \ldots, w_n$ in B^k, so that every two distinct members of $\mathcal{G}$ are ε-separated (that is, have at least one non-diagonal entry in which the two matrices differ by more than ε). By the maximality of $\mathcal{G}$, for every Gram matrix M of n vectors in B^k there is a member of $\mathcal{G}$ in which all inner products of pairs of distinct points are within ε of the corresponding inner products in M, meaning that as a sketch for M it suffices to store (besides the approximate norms of the vectors), the index of an appropriate member of $\mathcal{G}$. This requires $\log |\mathcal{G}|$ bits. It remains to prove an upper bound for the cardinality of $\mathcal{G}$.

Let $V_1, V_2, \ldots, V_n$ be n vectors, each chosen randomly, independently and uniformly in the ball of radius 2 in R^k. Let $T = G(V_1, V_2, \ldots, V_n)$ be the Gram matrix of the vectors V_i. For each $G \in \mathcal{G}$ let A_G denote the event that for every $1 \leq i \neq j \leq n$, $|T(i,j) - G(i,j)| < \varepsilon/2$. Note that since the members of $\mathcal{G}$ are ε-separated, all the events A_G for $G \in \mathcal{G}$ are pairwise disjoint. To complete the proof it thus suffices to show that the probability of each event A_G is at least $e^{-O(nk \log(1/\delta))}$. To see that this is the case, fix a Gram matrix $G = G(w_1, \ldots, w_n) \in \mathcal{G}$ for some $w_1, \ldots, w_n \in B^k$ of norm at most 2. For each fixed i the probability that V_i lies in the ball of radius δ centered at w_i is exactly $(\delta/2)^k$. Therefore the probability that this happens for all i is $(\delta/2)^{nk}$. Conditioning on that, for each i the vector $V_i - w_i$ is uniformly distributed in the ball of radius δ in R^k centered at 0. For each i let, now, A_i be the event that $|\langle V_i - w_i, w_j \rangle| \leq \varepsilon/4$ for all $i < j \leq n$, and that $|\langle V_\ell, V_i - w_i \rangle| \leq \varepsilon/4$ for all $1 \leq \ell < i$. In particular, the event A_1 is that $V_1 - w_1$ lies in the intersection of the $n-1$ slabs $|\langle x, w_j \rangle| \leq \varepsilon/4$ for $j > 1$. More generally, conditioning on the events $A_1, \ldots A_{i-1}$ (as well as on the events that $|V_j - w_j| \leq \delta$ for all j), the event A_i is that $V_i - w_i$ lies in the intersection of the slabs $|\langle x, w_j \rangle| \leq \varepsilon/4$ for $j > i$ and the slabs $|\langle V_\ell, x \rangle| \leq \varepsilon/4$ for $\ell < i$. Note that conditioning on $A_1, \ldots, A_{i-1}$, the vectors $V_1, V_2, \ldots, V_{i-1}$ are of norm at most $1 + \varepsilon/4 < 2$, and once their values are exposed then indeed we have here an intersection of $n-1$ slabs with a ball centered at the origin.

By Lemma VI.2 it follows that the conditional probability of each event A_i given all previous ones $A_1, \ldots, A_{i-1}$ and given that all vectors V_i lie within distance δ of the corresponding vectors w_i is at least

$$C^{-k}(1 - 2e^{-\frac{\varepsilon^2}{64\delta^2}k})^n,$$

where C is an absolute positive constant. Since $k = \delta t$ and $\varepsilon^2 t = \Theta(\log(2 + n/t))$ it follows that

$$e^{-\frac{\varepsilon^2}{64\delta^2}k} \leq e^{-\frac{c \log(2+n/t)}{\delta}} = (\frac{t}{2t+n})^{c/\delta}.$$

As $t \leq n$ the last quantity is at most

$$(\frac{1}{3})^{\frac{c}{2\delta}} \frac{t}{2t+n}$$

provided $\delta < c/2$.

Thus

$$(1 - 2e^{-\frac{\varepsilon^2}{64\delta^2}k})^n \geq [1 - (\frac{1}{3})^{c/2\delta} \frac{t}{2t+n}]^n$$
$$\geq e^{-(1/3)^{c/2\delta}t} \geq e^{-\delta t} = e^{-k}$$

for all $\delta < c'$.

By multiplying all conditional probabilities we conclude that the probability that $V_i - w_i$ is of norm at most δ for all i and that all events A_i hold too is at least $e^{-O(nk \log(1/\delta))}$. However, in this case, for all $i < j$

$$|\langle V_i, V_j \rangle - \langle w_i, w_j \rangle| \leq |\langle V_i - w_i, w_j \rangle| + |\langle V_i, V_j - w_j \rangle| \leq \varepsilon/2$$

and the event A_G occurs. Thus the probability of each event A_G is at least $e^{-O(nk \log(1/\delta))}$, providing the required upper bound for $|\mathcal{G}|$ and hence completing the proof of the upper bound in Theorem II.2.

The proof of the lower bound is similar to the proof of the lower bound in Theorem VI.1. The most interesting case here is again the range

$$\log(2 + \varepsilon^2 n) \leq k \leq t = \frac{\log(2 + \varepsilon^2 n)}{\varepsilon^2}.$$

(Note that the lower bound for $k \geq t$ follows from the case $k = \Theta(t)$.) Here it is convenient to define δ so that $k = \delta^2 \frac{\log(2+\varepsilon^2 n)}{4\varepsilon^2}$ where $\varepsilon \geq \frac{2}{\sqrt{n}}$ and $2\varepsilon \leq \delta < 1$ and to assume we have $2n$ points. Let B be a collection of, say, $(\delta^{-1}/2)^k$ unit vectors in R^k so that the Euclidean distance between any two of them is at least δ. We claim that there are n unit vectors a_i in R^k so that for any two distinct members b, b' of B there is an i so that $|\langle b, a_i \rangle - \langle b', a_i \rangle| > \varepsilon$.

Indeed, taking the vectors a_i randomly, independently and uniformly in the unit ball of R^k the probability that for a fixed pair b, b' the above fails is at most

$$(1 - e^{-\frac{\varepsilon^2}{\delta^2}k})^n.$$

Our choice of parameters ensures that

$$\binom{|B|}{2}(1 - e^{-\frac{\varepsilon^2}{\delta^2}k})^n < 1.$$

Indeed it suffices to check that

$$e^{-\frac{\varepsilon^2}{\delta^2}k} \cdot n > 2k \log(1/2\delta)$$

that is

$$\frac{\varepsilon^2}{\delta^2}k < \log(\frac{n}{2k \log(1/2\delta)}),$$

or equivalently

$$k < \frac{\delta^2}{\varepsilon^2}\log(\frac{n}{2k\log(1/2\delta)}).$$

By the definition of

$$k = \delta^2\frac{\log(2+\varepsilon^2 n)}{4\varepsilon^2}$$

it suffices to show that

$$\log(2+\varepsilon^2 n) < 4\log(\frac{n}{2k\log(1/2\delta)})$$

$$= 4\log[\frac{n4\varepsilon^2}{2\delta^2\log(2+\varepsilon^2 n)\log(1/2\delta)}].$$

This easily holds for $\varepsilon \geq 2/\sqrt{n}$.

By the union bound the assertion of the claim follows. The desired result now holds, since every union of the vectors a_i with an ordered set of n members of B must have a different representation, hence the number of bits needed is at least $n\log_2|B| = \Omega(nk\log(1/\delta)$.

The case $k \leq \log(2+\varepsilon^2 n)$ is proved in a similar way by letting B be a 2ε-separated set of points in B^k. We omit the detailed computation. This completes the proof of the theorem.

C. Keeping the inner products with small distortion

In this subsection we prove Theorem I.2. The main result we use is the well-known low M^*-estimate due to Pajor and Tomczack-Jaegermann, which builded upon earlier contributions by Milman and by Gluskin, see e.g., [6, Chapter 7]:

Theorem VI.3. *Let $1 \leq t \leq n$ and let $K \subseteq R^n$ be a centrally-symmetric convex body with $\gamma_n(K) \geq 1/2$. Let $E \subseteq R^n$ be a random subspace of dimension $n-t$. Then with probability at least $1 - C\exp(-ct)$ of selecting E,*

$$\tilde{c}\sqrt{t}B_E \subseteq Proj_E(K).$$

Here, $c, \tilde{c}, C > 0$ are universal constants and $B_E = B^n \cap E$.

Proof: Our formulation is very close to (7.1.1) and Theorem 7.3.1 in [6]. We only need to explain a standard fact, why $\gamma_n(K) \geq 1/2$ implies the bound $M(K) \leq C/\sqrt{n}$ where

$$M(K) := \int_{S^{n-1}} \|x\|_K d\sigma_{n-1}(x)$$

and $\|x\|_K = \inf\{\lambda > 0\,;\, x \in \lambda K\}$. However, it is not difficult to show that

$$\frac{1}{2} \leq \gamma_n(K)$$

$$\leq \gamma_n\left(\frac{\sqrt{n}}{2}B^n\right) + \gamma_n\left(K \setminus \frac{\sqrt{n}}{2}B^n\right)$$

$$\leq e^{-cn} + \sigma_{n-1}\left(\frac{2}{\sqrt{n}}K\right),$$

where σ_{n-1} is the uniform probability measure on the unit sphere S^{n-1}. Hence $\sigma_{n-1}\left(\frac{2}{\sqrt{n}}K\right) \geq 1/2 - \exp(-cn)$. In other words, in a large subset of S^{n-1}, the norm $\|x\|_K$ is at most $2/\sqrt{n}$. In [6, Lemma 5.2.3] it is explained how concentration inequalities upgrade this fact to the desired bound $M(K) \leq C/\sqrt{n}$. ∎

Our next observation is that the assumption $\gamma_n(K) \geq 1/2$ in Theorem VI.3 is too strong, and may be weakened to the requirement that $\gamma_n(K) \geq \exp(-ct)$.

Theorem VI.4. *Let $1 \leq t \leq n$ and let $K \subseteq R^n$ be a centrally-symmetric convex body with $\gamma_n(K) \geq \exp(-c_0 t)$. Let $E \subseteq R^n$ be a random subspace of dimension $n-t$. Then with probability of at least $1 - C\exp(-ct)$,*

$$c_1\sqrt{t}B_E \subseteq Proj_E(K).$$

Proof: We may select the universal constant $c_0 > 0$ so that the probability that a standard normal random variable exceeds $\tilde{c}\sqrt{t}/2$, where $\tilde{c}$ is the constant in the conclusion of Theorem VI.3, is at most $e^{-c_0 t}$.

According to the Gaussian isoperimetric inequality, for a half-space $H \subseteq R^n$,

$$\gamma_n(K)$$

$$= \gamma_n(H) \Longrightarrow \gamma_n(K + (\tilde{c}\sqrt{t}/2)B^n) \geq \gamma_n(H + (\tilde{c}\sqrt{t}/2)B^n).$$

Since $\gamma_n(H) = \gamma_n(K) \geq \exp(-c_0 t)$, the choice of c_0 implies that the distance between the half-space H and the origin is at most $\tilde{c}\sqrt{t}/2$. Consequently, $H + (\tilde{c}\sqrt{t}/2)B^n$ is a half-space containing the origin, thus its Gaussian measure is at least $1/2$. Hence

$$T := K + \frac{\tilde{c}}{2}\sqrt{t}B^n$$

is a centrally-symmetric convex body with $\gamma_n(T) \geq 1/2$. By Theorem VI.3, with probability at least $1 - C\exp(-ct)$ of selecting E,

$$\tilde{c}\sqrt{t}B_E \subseteq Proj_E(T) = Proj_E(K) + Proj_E\left(\frac{\tilde{c}\sqrt{t}}{2}B^n\right)$$

$$= Proj_E(K) + \frac{\tilde{c}\sqrt{t}}{2}B_E. \tag{3}$$

Since B_E and $Proj_E(K)$ are convex, we deduce from (3) that $(\tilde{c}\sqrt{t}/2)B_E \subseteq Proj_E(K)$, completing the proof. ∎

Remark. Consider the case where

$$K = [-r, r]^n$$

is an n-dimensional cube, for $r = c\sqrt{\log(n/\ell)}$. In this case one may easily verify that $\gamma_n(K) \geq \exp(-c_0\ell)$. Thus, according to the last Theorem, with high probability a random $(n-\ell)$-dimensional projection of K contains a

Euclidean ball of radius $\tilde{c}\sqrt{\ell}$. This recovers an inequality by Garnaev and Gluskin [9]. Moreover, the tightness of the Garnaev-Gluskin result shows that the requirement that $\gamma_n(K) \geq \exp(-c_0\ell)$ in the Theorem is optimal.

Corollary VI.5. *Let $K \subseteq R^n$ be a centrally-symmetric convex body with $\gamma_n(K) \geq \exp(-c_0 t)$ with $1 \leq t \leq n$. Then there exists a t-dimensional subspace $E \subseteq R^n$ such that for any $v \in R^n$,*

$$|v| \leq \sqrt{t} \Longrightarrow E \cap (v + CK) \neq \emptyset. \tag{4}$$

Proof: Write $F = E^{\perp}$. Condition (4) is equivalent to $\sqrt{t}B_F \subseteq Proj_F(CK)$. The corollary thus follows from Theorem VI.4 with $C = 1/c_1$. ∎

Proof of Theorem I.2: We may assume that $t \leq n$ as otherwise the conclusion of the theorem is trivial. We may also assume that $C > 5/c_0$ where $c_0 > 0$ is the universal constant from Corollary VI.5. That is, $c_0 t \geq 5 \cdot \varepsilon^{-2} \log(2 + \varepsilon^2 n)$, thus

$$\varepsilon \geq 2\sqrt{\frac{\log(2 + n/(c_0 t))}{c_0 t}}.$$

Identify R^t with the subspace of R^n of all vectors whose last $n - t$ coordinates vanish, thus we may write $R^t \subseteq R^n$. Let $U \in O(n)$ be an orthogonal matrix to be determined later on. For all i, j, and for every vectors X_i, Y_j in R^n

$$\left| \left\langle \frac{X_i}{\sqrt{t}}, \frac{Y_j}{\sqrt{t}} \right\rangle - \langle a_i, b_j \rangle \right| \tag{5}$$

$$\leq \left| \left\langle \frac{UX_i}{\sqrt{t}} - a_i, b_j \right\rangle \right| + \left| \left\langle \frac{X_i}{\sqrt{t}}, \frac{Y_j}{\sqrt{t}} - U^{-1} b_j \right\rangle \right|. \tag{6}$$

We next bound the first summand on the right-hand side of (6). (We will later observe that we can ensure that the second summand vanishes). Define

$$K = \left\{ x \in R^n \,;\, \left| \left\langle \frac{x}{\sqrt{t}}, b_j \right\rangle \right| \leq \sqrt{c_0}\varepsilon \text{ for } j = 1, \ldots, n \right\},$$

where $c_0 > 0$ is still the constant from Corollary VI.5. By the Khatri-Sidak lemma

$$\gamma_n(K) \geq \prod_{j=1}^{n} \gamma_n \left(\left\{ x \in R^n \,;\, \left| \left\langle \frac{x}{\sqrt{t}}, b_j \right\rangle \right| \leq \sqrt{c_0}\varepsilon \right\} \right)$$

$$= \prod_{j=1}^{n} \left(1 - 2\Phi(\sqrt{c_0 t}\varepsilon / |b_j|)\right)$$

$$\geq \left(1 - 2\Phi\left(2\sqrt{\log(2 + n/(c_0 t))}\right)\right)^n$$

$$\geq \left(1 - \frac{c_0 t}{n + c_0 t}\right)^n \geq e^{-c_0 t}.$$

By Corollary VI.5 there exists a t-dimensional subspace $E \subseteq R^n$ such that for any $v \in R^n$

$$|v| \leq \sqrt{t} \Longrightarrow E \cap (v + CK) \neq \emptyset.$$

Let us now set $U \in O(n)$ to be any orthogonal transformation with $U(R^t) = E$, and choose $Ux_i \in E$ so that $Ux_i - \sqrt{t}a_i \in CK$. Finally define $y_j = \sqrt{t}P(U^{-1}b_j)$, where $P(z_1, z_2, \ldots, z_n) = (z_1, z_2, \ldots, z_t)$.

This gives an upper bound of $C\sqrt{c_0}\varepsilon$ for the right-hand side of (6) for all i, j, implying a variant of Theorem I.2 in which ε is replaced by $C\sqrt{c_0}\varepsilon$. By adjusting the constants, this variant is equivalent to the original formulation, completing the proof. ∎

Note that the proof of Theorem I.2 leads to a randomized, polynomial-time algorithm for the computation of the x_i, y_j. Indeed, the orthogonal matrix $U \in O(n)$ can be chosen randomly, and according to Theorem VI.4 and Corollary VI.5 such a random matrix works with probability of at least $1 - C\exp(-ct)$. Once the matrix U is known, the computation of x_i such that $Ux_i \in E$ and $Ux_i \in \sqrt{t}a_i + CK$ may be done by linear programming. The computation of the y_j is even quicker, since we set $y_j = \sqrt{t}P(U^{-1}b_j)$. The total running time of the algorithm is clearly polynomial in the input size.

VII. Concluding remarks

- By the first two parts of Theorem I.1, $f(n, n, 2\varepsilon)$ is much bigger than $f(n, k, \varepsilon)$ for any $k < c\frac{\log n}{\varepsilon^2}$ for some absolute constant $c > 0$, implying that, as proved recently by Larsen and Nelson [18], the $\frac{\log n}{\varepsilon^2}$ bound in the Johnson-Lindenstrauss Lemma [14] is tight. The first part of Corollary II.1 follows by a similar reasoning. It can also be derived directly from the result for $k = \log n/\varepsilon^2$. As for the "Moreover" part, it follows by combining the Johnson-Lindenstrauss Lemma with the lower bound of Theorem I.1. Corollary II.3 follows from Theorem II.2 using essentially the same argument.
- It is worth noting that in the proof of Theorem IV.1 the inner product of each rounded vector with itself is typically not close to the square of its original norm and hence it is crucial to keep the approximate norms separately. An alternative, less natural possibility is to store two independent rounded copies of each vector and use their inner product as an approximation for its norm. This, of course, doubles the length of the sketch and there is no reason to do it. For the same reason in the proof of Theorem I.1 in Section III we had to handle norms separately and consider only inner products between distinct vectors. Indeed, in this proof after the conditioning V_i is likely to have much bigger norm than w_i, and yet the inner products of distinct V_i, V_j are typically very close to those of the corresponding distinct w_i, w_j.

- The assertion of Theorem II.5 for $m = 2n$ and $\varepsilon = \frac{C}{\sqrt{n}}$ is tight up to a constant factor even for the case that $a_i = b_i$ for all i and the vectors a_i form an orthonormal basis of R^{2n}. Indeed, it is well known (see, e.g., [2]) that any $2n$ by $2n$ matrix in which every entry differs from the corresponding entry of the identity matrix of dimension $2n$ by less than, say, $\frac{1}{2\sqrt{n}}$ has rank exceeding n.
- For a matrix A, the γ_2-norm of A denoted by $\gamma_2(A)$ is the minimum possible value, over all factorizations $A = XY$, of the product of the maximum ℓ_2-norm of a row of X and the maximum ℓ_2-norm of a column of Y. Therefore, an equivalent formulation of the statement of Theorem I.2 for $\varepsilon = O(1/\sqrt{n})$ is that for any n by n matrix A satisfying $\gamma_2(A) \leq 1$ there is an n by n matrix B of rank at most, say, $n/10$ so that $|A_{ij} - B_{ij}| \leq O(1/\sqrt{n})$ for all i, j. It is worth noting that the assumption that $\gamma_2(A) \leq 1$ here is essential and cannot be replaced by a similar bound on $\max |A_{ij}|$. Indeed, it is known (see [4], Theorem 1.2) that if A is an n by n Hadamard matrix then any B as above has rank at least $n - O(1)$.
- Conjecture II.4 remains open, it seems tempting to try to iterate the assertion of Theorem II.5 in order to prove it. This does not work as the norms of the vectors x_i and y_i obtained in the proof may be much larger than 1 (while bounded), causing the errors in the iteration process to grow too much. An equivalent formulation of this fact is that the γ_2-norm of the matrix $\langle a_i, b_j \rangle$ is 1 whereas that of its approximating lower rank matrix is a larger constant.

ACKNOWLEDGMENT

We thank Jaroslaw Blasiok, Kasper Green Larsen and especially Jelani Nelson for helpful comments, and for noting the relation to the paper [17]. The research of the first author was supported in part by a USA-Israeli BSF grant, by an ISF grant, by a GIF grant and by the Israeli I-Core program. The research of the second author was supported in part by an ERC grant.

REFERENCES

[1] N. Ailon and B. Chazelle, The fast Johnson-Lindenstrauss transform and approximate nearest neighbors, SIAM J. Comput. 39 (2009), 302–322.

[2] N. Alon, Perturbed identity matrices have high rank: proof and applications, Combinatorics, Probability and Computing 18 (2009), 3-15.

[3] N. Alon and B. Klartag, Optimal compression of approximate Euclidean distances, arXiv: 1610.00239, 2016.

[4] N. Alon, T. Lee, A. Shraibman and S. Vempala, The approximate rank of a matrix and its algorithmic applications, Proc. STOC 2013, 675–684.

[5] N. Alon, Y. Matias and M. Szegedy, The space complexity of approximating the frequency moments, Proc. STOC 1996, 20-29. Also: J. Comp. Sys. Sci. 58 (1999), 137-147.

[6] S. Artstein-Avidan, A. Giannopoulos and V. D. Milman, *Asymptotic Geometric Analysis*, American Mathematical Society, 2015.

[7] E. Candés, J. Romberg and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory, 52(2):489–509, 2006.

[8] A. Giannopoulos, On some vector balancing problems, Studia Math. 122 (1997), 225–234.

[9] A. Y. Garnaev and E. D. Gluskin, The widths of an Euclidean ball, Dokl. Akad. Nauk SSSR, Vol. 277, No. 5, (1984), 1048-1052. English translation in Soviet Math. Dokl., Vol. 30, No. 1, (1984), 200–204.

[10] G. Hargé, A particular case of correlation inequality for the Gaussian measure, The Annals of Probability 27 (1999), 1939–1951.

[11] S. Har-Peled, P. Indyk and R. Motwani, Approximate nearest neighbor: Towards removing the curse of dimensionality, Theory of Computing, 8(1):321–350, 2012.

[12] W. Hoeffding, Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association. 58 (301) (1963), 13–30.

[13] P. Indyk and T. Wagner, Near-optimal (Euclidean) metric compression, arXiv: 1609.06295, 2016.

[14] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz maps into a Hilbert space, Contemp Math 26 (1984), 189–206.

[15] C. G. Khatri, On certain inequalities for normal distributions and their applications to simultaneous confidence bounds, Ann. Math. Statist. 38 (1967), 1853–1867.

[16] B. Klartag, A geometric inequality and a low M estimate, Proc. Amer. Math. Soc. 132 (2004), 2619–2628.

[17] E. Kushilevitz, R. Ostrovsky and Y. Rabani, Efficient search for approximate nearest neighbor in high-dimensional spaces, Proc. STOC 1998, 614–623.

[18] K. G. Larsen and J. Nelson, Optimality of the Johnson-Lindenstrauss Lemma, arXiv: 1609.02094, 2016.

[19] K. G. Larsen and J. Nelson, Private Communication.

[20] S. Muthukrishnan, Data streams: Algorithms and applications, Foundations and Trends in Theoretical Computer Science, 1(2), 2005.

[21] T. Royen, A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions, Far East J. Theor. Stat. 48 (2014), 139–145.

[22] Z. Sidak, Rectangular confidence regions for the means of multivariate normal distributions, J. Amer. Statist. Assoc. 62 (1967), 626–633.