Optimal compression of approximate inner products and dimension reduction

Noga Alon[1] Bo'az Klartag[2]

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.

1 The problem and main results

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. What is the minimum possible number of bits of such a sketch ? Denote this minimum by $f(n, k, \varepsilon)$. Here $1 \leq k \leq n$ and we assume that $\frac{1}{n^{0.49}} \leq \varepsilon \leq 0.1$.

The most basic case is when $k = n$, that is, there is no restriction on the dimension. In this case one can apply the Johnson-Lindenstrauss Lemma [8] to project the points into

[1]Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant 2012/107, by an ISF grant 620/13 and by the Israeli I-Core program.

[2]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. Research supported in part by an ERC grant.

R^m where $m = O(\log n/\varepsilon^2)$ with distortion at most $\varepsilon/2$, and then round each point to the closest one in an $\varepsilon/2$-net in the ball of radius $1+\varepsilon/2$ in R^m. As the size of the net is $[O(1/\varepsilon)]^m$, this enables us to represent each point by $O(m\log(1/\varepsilon))$ bits showing that

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

On the other hand it is not difficult to deduce from the recent construction in [12] that

$$f(n,n,\varepsilon) \geq \Omega(\frac{n\log n}{\varepsilon^2}).$$

A better upper bound follows from the results of Kushilevitz, Ostrovsky and Rabani in [11], where the authors show that all inner products between the pairs of n points on the unit sphere in R^n can be approximated up to a relative error of ε by storing only $O(\frac{\log n}{\varepsilon^2})$ bits per point. This easily implies that $f(n,n,\varepsilon) = \Theta(\frac{n\log n}{\varepsilon^2})$ and in view of the discussion above

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

for all $\frac{\log n}{\varepsilon^2} \leq k \leq n$.

What happens for smaller k ? In this paper we determine $f(n,k,\varepsilon)$ up to a constant factor for all admissible n,k and ε. This is stated in the following Theorem.

Theorem 1.1. *For all n and $\frac{1}{n^{0.49}} \leq \varepsilon \leq 0.1$ the function $f(n,k,\varepsilon)$ satisfies the following*

- *For $\frac{\log n}{\varepsilon^2} \leq k \leq n$,*
$$f(n,k,\varepsilon) = \Theta(\frac{n\log n}{\varepsilon^2}).$$
- *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 $1 \leq k \leq \log n$,*
$$f(n,k,\varepsilon) = \Theta(nk\log(1/\varepsilon)).$$

As mentioned above, the first part of the theorem is known, by the results of [11], [12]. For completeness, and since our proof is different, we include here a proof of this part as well. We present two proofs of the upper bound in the theorem. The first, described in Section 2, is based on a short probabilistic (or volume) argument. Its disadvantage is that it is not constructive and provides neither an efficient algorithm for producing the sketch

for a given set of points X, nor an efficient algorithm for recovering the approximate square distance between two desired points of X, given the sketch. The second proof, presented in Section 3, 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.

The proofs of the upper bound here and in [11] are different. In particular, our proof(s) yield sharp results for all values of k while the argument in [11] is suboptimal for $k = o(\frac{\log n}{\varepsilon^2})$. We describe the lower bound in Section 4.

Theorem 1.1 supplies an alternative proof of the main result of [12] 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

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

Corollary 1.2. *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,ℓ,ε)-Euclidean dimension reduction if and only if $\ell = \Omega(\log n/\varepsilon^2)$.*

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

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

Corollary 1.2 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-Lindesntrauss Lemma), and that if there is an (n,k,ℓ,ε)-Euclidean dimension reduction with $\ell = o(k)$ then necessarily $k \geq \ell \geq c_2\varepsilon^{-2}\log n$, for some absolute constants $c_1, c_2 > 0$. This statement for $k \geq \Omega(\varepsilon^{-2}\log n)$ is proved in [12], and the result for smaller k is an easy consequence.

In all the results above $\varepsilon \geq \frac{1}{n^{0.49}}$. Indeed for smaller ε, when $\log(2+\varepsilon^2 n) = o(\log n)$, the arguments break and suggest that it may be possible to replace the $\log n$ term by the expression $\log(2+\epsilon^2 n)$. The lower bound for $n = k$ extends to the entire range $\varepsilon > 2/\sqrt{n}$ if we replace the $\log n$ term by $\log(2+\epsilon^2 n)$. In fact, it is possible that as suggested by Larsen and Nelson [13] for such small values of ε the assertion of the Johnson-Lindenstrauss Lemma can be improved, replacing $\frac{\log n}{\varepsilon^2}$ by $\frac{\log(2+\varepsilon^2 n)}{\varepsilon^2}$. Motivated by this we prove the following slightly weaker result.

Theorem 1.3. *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*

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

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 conjecture that the assertion of the above theorem can be strengthened, as follows.

Conjecture 1.4. *Under the assumptions of Theorem 1.3, the conclusion holds together with the further requirement that $\|x_i\| \leq O(1)$ and $\|y_i\| \leq O(1)$ for all $1 \leq i \leq n$.*

If true, this, 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 $\varepsilon \geq \frac{2}{\sqrt{n}}$. The stronger 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 1.5. *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 $\exp(-C_2 nm)$,

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

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 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 above, as should be expected from the assertion of the Conjecture.

The results on smaller ε are proven in Section 5 using several tools from convex geometry including the low-M^* estimate and the finite volume-ratio theorem (see, e.g., [4])

and basic results about the positive correlation between symmetric convex events with the Gaussian measure. The final section 6 contains some concluding remarks and open problems.

Throughout the proofs we make no serious attempt to optimize the absolute constants involved. 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.

2 The upper bound

It is convenient to split the proof into three lemmas, dealing with the different ranges of k.

Lemma 2.1. *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 increasing function of k, it suffices to prove the upper bound for $k = n$. By [8] 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 $m = 40\frac{\log n}{\varepsilon^2}$ 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 $G(i,j) = w_i^t w_j$. 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. 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 $(V_i - w_i)^t w_j$ has absolute value at least $\varepsilon/2$ is at most $2e^{-\varepsilon^2 m/8} < 1/(2n^2)$. (Here we used the fact that the norm of w_j is at most 2). Similarly, since the norm of V_i is at most 3, the probability that the inner product $V_i^t(V_j - w_j)$ has absolute value at least $\varepsilon/2$ is at most $2e^{-\varepsilon^2 m/12} < 1/2n^2$. It follows that with probability bigger than $0.5(1/3)^{nm}$ all these inner products are smaller than $\varepsilon/2$, implying that

$$|V_i^t V_j - w_i^t w_j| \leq |(V_i - w_i)^t w_j| + |V_i^t(V_j - w_j)| < \varepsilon.$$

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. $\square$

Lemma 2.2.

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 second part of the proof above. Note, first, that by monotonicity and the fact that the expressions above change 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, $\delta/5$ centered at w_i is exactly $(\delta/10)^{kn}$. Conditioning on that, the probability that the inner product $(V_i - w_i)^t w_j$ has absolute value at least $\varepsilon/2$ is at most

$$2e^{-\varepsilon^2 25k/4\delta^2} < 1/(2n^2).$$

Similarly, the probability that the inner product $V_i^t(V_j - w_j)$ has absolute value at least $\varepsilon/2$ is at most

$$2e^{-\varepsilon^2 25k/8\delta^2} < 1/2n^2.$$

As before, this implies that $|\mathcal{G}| \leq 2(10/\delta)^{kn}$, establishing the assertion of the lemma. □

Lemma 2.3.

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. □

3 An algorithmic proof

In this section we present an algorithmic proof of the upper bound of Theorem 1.1. We first reformulate the theorem in its algorithmic version. Note that the first part also follows from the results in [11].

Theorem 3.1. *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 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 2 for the range $k \leq \log n$ is essentially constructive, since it is well known (see, for example [3] 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 $n_i \geq 2$. 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 3.1 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 2, 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 $(V_i - w_i)^t w_j$ 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 [6], 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 $V_i^t(V_j - w_j)$ is of absolute value at least $\varepsilon/2$ is smaller than $1/n^5$. As in the proof in Section 2, 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 3.1.

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 3.1. $\square$

4 The lower bound

Lemma 4.1. *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 than ε is at least $e^{-1.5\varepsilon^2 k/\delta^2}$ (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.

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 1.1 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 1.1. □

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 3.1 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.

5 Small distortion

In this section we prove three results related to Conjecture 1.4 regarding the case of smaller ε. In Section 5.1 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 5.2 we prove Theorem 1.5, while in Section 5.3 we prove Theorem 1.3.

5.1 Inner products with fixed vectors

Theorem 5.1. *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 increasing 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 ([9], [14], see also [5] 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 2. 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 4. 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. □

5.2 Halving the dimension

In this subsection we prove Theorem 1.5. Throughout this subsection and the next one 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 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. Theorem 1.5 is equivalent to the following statement:

Theorem 5.2. *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.*

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

$$\left| \left\langle \frac{X_i}{\sqrt{n}}, \frac{Y_j}{\sqrt{n}} \right\rangle - \langle a_i, b_j \rangle \right| \leq \varepsilon \qquad \text{for } i, j = 1, \ldots, m, \tag{3}$$

and moreover $|X_i| + |Y_i| \leq C_3 \sqrt{n}$ for all i.

In the proof of Theorem 5.2 we will use the following theorem, which is the dual version of the finite-volume ratio theorem of Szarek and Tomczak-Jaegermann (see e.g. [4, Section 5.5] and also [10] for an alternative proof). A convex body is a compact, convex set with a non-empty interior, and $B^n = \{x \in R^n \,;\, |x| \leq 1\}$ is the centered unit Euclidean ball in R^n.

Theorem 5.3. *Let $K \subseteq B^{2n}$ be a centrally-symmetric convex body with $Vol_{2n}(K) \geq e^{-10n} Vol_{2n}(B^{2n})$. Then there exists an n-dimensional subspace $E \subseteq R^{2n}$ with*

$$cB^{2n} \cap E \subseteq Proj_E(K),$$

where $Proj_E$ is the orthogonal projection operator onto E in R^{2n}.

For completeness we include a short derivation of this theorem from [4, Theorem 5.5.3].

Proof. The polar body to a centrally-symmetric convex body $K \subseteq R^{2n}$ is

$$K^\circ = \{x \in R^{2n} \,;\, \forall y \in K,\, |\langle x, y \rangle| \leq 1\}.$$

Polarity is an order-reversing involution, i.e., $(K^\circ)^\circ = K$ while $K_1 \subseteq K_2$ implies that $K_1^\circ \supseteq K_2^\circ$. Moreover, $(B^{2n})^\circ = B^{2n}$. Since $K \subseteq B^{2n}$ we know that $B^{2n} \subseteq K^\circ$. By the Santaló inequality (e.g., [4, Theorem 1.5.10]),

$$Vol_n(K^\circ) \leq \frac{Vol_{2n}(B^{2n})^2}{Vol_n(K)} \leq e^{10n} Vol_{2n}(B^{2n}).$$

According to the finite-volume ratio theorem (e.g., [4, Theorem 5.5.3]), there exists an n-dimensional subspace $E \subseteq R^{2n}$ with

$$K^\circ \cap E \subseteq C(B^{2n} \cap E). \tag{4}$$

However, $Proj_E(K)^\circ = K^\circ \cap E$ for any subspace $E \subseteq R^{2n}$. Thus the desired conclusion follows from (4). $\square$

Theorem 5.3 implies the following:

Lemma 5.4. *Let K be as in Theorem 5.3. Then there exists an n-dimensional subspace $E \subseteq R^{2n}$ so that*

$$\forall x \in c_1 B^{2n}, \quad Vol_n(E \cap (x+K)) \geq c_2^n \cdot Vol_n(B^n).$$

Proof. We set $c_1 = c/2$ where $c > 0$ is the constant from Theorem 5.3. Thus there exists an n-dimensional subspace E with

$$2c_1 B^{2n} \cap E^\perp \subseteq Proj_{E^\perp}(K) \tag{5}$$

where $E^\perp$ is the orthogonal complement to E in R^{2n}. By Fubini's theorem,

$$e^{-10n} \cdot Vol_{2n}(B^{2n}) \leq Vol_{2n}(K) \leq Vol_n(Proj_{E^\perp}(K)) \cdot \sup_{x \in E^\perp} Vol_n(E \cap (x+K)). \tag{6}$$

The Brunn-Minkowski inequality and the central symmetry of K imply that for any $x \in E^\perp$,

$$Vol_n(E \cap K)^{\frac{1}{n}} \geq \frac{Vol_n(E \cap (x+K))^{\frac{1}{n}} + Vol_n(E \cap (-x+K))^{\frac{1}{n}}}{2} = Vol_n(E \cap (x+K))^{\frac{1}{n}}.$$

Thus the supremum in (6) is attained for $x = 0$. Since $K \subseteq B^{2n}$ we conclude from (6) that

$$Vol_n(K \cap E) \geq e^{-10n} \cdot \frac{Vol_{2n}(B^{2n})}{Vol_n(B^n)} \geq e^{-Cn} \cdot Vol_n(B^n), \tag{7}$$

for some constant $C > 0$. Let $x \in R^{2n}$ satisfy $|x| \leq c_1$. Then $Proj_{E^\perp}(-2x) \in 2c_1 B^{2n} \cap E^\perp$. According to (5) there exists $y \in K$ with $y + 2x \in E$. Thus $(\{y\} + K \cap E)/2 \subseteq K \cap (E - x)$. By (7) and the convexity of K,

$$Vol_n(E \cap (x+K)) = Vol_n(K \cap (E-x)) \geq Vol_n\left(\frac{\{y\} + K \cap E}{2}\right) \geq c_2^n \cdot Vol_n(B^n),$$

for some constant $c_2 > 0$, completing the proof of the lemma. $\square$

As in the previous subsection we write γ_{2n} for the standard Gaussian probability measure in R^{2n}. For a subspace $E \subseteq R^{2n}$, write γ_E for the standard Gaussian measure in the subspace E. For $K \subset R^{2n}$ we denote $\gamma_E(K \cap E)$ by $\gamma_E(K)$.

Corollary 5.5. *Let $K \subseteq R^{2n}$ be a centrally-symmetric convex body with $\gamma_{2n}(K) \geq e^{-n}$. Then there exists an n-dimensional subspace $E \subseteq R^{2n}$ such that for any $v \in R^{2n}$,*

$$|v| \leq \sqrt{n} \qquad \Longrightarrow \qquad \gamma_E(v + CK) \geq c^n.$$

Proof. Write σ_{2n-1} for the uniform probability measure on the unit sphere $S^{2n-1} = \{x \in R^{2n}\,;\, |x| = 1\}$. For $K \subset R^{2n}$ denote $\sigma_{2n-1}(K \cap S^{2n-1})$ by $\sigma_{2n-1}(K)$. Since K is a convex set containing the origin and the Gaussian measure is rotationally-invariant, for any $r > 0$,

$$e^{-n} \leq \gamma_{2n}(K) \leq \gamma_{2n}(rB^{2n}) + \gamma_{2n}(K \setminus rB^{2n}) \leq \gamma_{2n}(rB^{2n}) + \sigma_{2n-1}\left(\frac{K}{r}\right).$$

A standard estimate shows that $\gamma_{2n}(c_1\sqrt{n}B^{2n}) \leq e^{-n}/2$ for some universal constant $c_1 > 0$. It follows that for $K_1 = K \cap c_1\sqrt{n}B^{2n}$,

$$\frac{Vol_{2n}(K_1)}{Vol_{2n}(c_1\sqrt{n}B^{2n})} \geq \sigma_{2n-1}\left(\frac{K_1}{c_1\sqrt{n}}\right) = \sigma_{2n-1}\left(\frac{K}{c_1\sqrt{n}}\right) \geq e^{-n}/2.$$

By Lemma 5.4, there exists an n-dimensional subspace $E \subseteq R^{2n}$ such that

$$\forall x \in c_2 B^{2n}, \quad Vol_n\left(E \cap \left(x + \frac{K_1}{c_1\sqrt{n}}\right)\right) \geq c^n \cdot Vol_n(B^n) \geq \left(\frac{\tilde{c}}{\sqrt{n}}\right)^n.$$

Now that the universal constants c_1 and c_2 are determined, we proceed as follows: For any $v \in R^{2n}$ with $|v| \leq \sqrt{n}$,

$$\begin{aligned}\gamma_E\left(v + \frac{K}{c_1c_2}\right) &\geq \gamma_E\left(v + \frac{K_1}{c_1c_2}\right) \geq e^{-Cn} Vol_n\left(E \cap \left(v + \frac{K_1}{c_1c_2}\right)\right) \\ &\geq \left(\tilde{c}\sqrt{n}\right)^n Vol_n\left(E \cap \left(\frac{c_2 v}{\sqrt{n}} + \frac{K_1}{c_1\sqrt{n}}\right)\right) \geq \bar{c}^n. \end{aligned}$$

□

Before continuing with the proof of Theorem 5.2 recall that as mentioned in the previous subsection, for any $a \in R^n$ and a centrally-symmetric measurable set $T \subseteq R^n$,

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

Proof of Theorem 5.2. We may assume that $n \geq 5 \cdot \varepsilon^{-2} \log(2 + \varepsilon^2 m)$, thus

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

We identify R^n with the subspace of R^{2n} of all vectors whose last n coordinates vanish, thus we may write $R^n \subseteq R^{2n}$. Let $U \in O(2n)$ be an orthogonal matrix to be determined later on. Observe that for all i, j,

$$\left| \left\langle \frac{X_i}{\sqrt{n}}, \frac{Y_j}{\sqrt{n}} \right\rangle - \langle a_i, b_j \rangle \right| \leq \left| \left\langle \frac{UX_i}{\sqrt{n}} - a_i, b_j \right\rangle \right| + \left| \left\langle \frac{X_i}{\sqrt{n}}, \frac{Y_j}{\sqrt{n}} - U^{-1} b_j \right\rangle \right|. \tag{8}$$

We shall bound separately each of the two summands on the right-hand side of (8). Define

$$K = \left\{ x \in R^{2n} ; \left| \left\langle \frac{x}{\sqrt{n}}, b_j \right\rangle \right| \leq \varepsilon \text{ for } j = 1, \ldots, m \right\}.$$

Recall that $\Phi(t) = (2\pi)^{-1/2} \int_t^\infty \exp(-s^2/2) ds$ and $\Phi(t) \leq \exp(-t^2/2)$ for $t \geq 1$. By the Khatri-Sidak lemma

$$\begin{aligned} \gamma_{2n}(K) &\geq \prod_{j=1}^m \gamma_{2n} \left(\left\{ x \in R^{2n} ; \left| \left\langle \frac{x}{\sqrt{n}}, b_j \right\rangle \right| \leq \varepsilon \right\} \right) = \prod_{j=1}^m \left(1 - 2\Phi(\sqrt{n}\varepsilon / |b_j|) \right) \\ &\geq \left(1 - 2\Phi\left(2\sqrt{\log(2 + m/n)} \right) \right)^m \geq \left(1 - \frac{n}{m+n} \right)^m \geq e^{-n}. \end{aligned}$$

From Corollary 5.5, there exists an n-dimensional subspace $E \subseteq R^{2n}$ such that for any $v \in R^{2n}$

$$|v| \leq \sqrt{n} \qquad \Longrightarrow \qquad \gamma_E(v + CK) \geq c^n. \tag{9}$$

Let us now set $U \in O(2n)$ to be any orthogonal transformation with $U(R^n) = E$. We also set C_3 to be a sufficiently large universal constant such that $\mathbb{P}(|X_i| \leq C_3\sqrt{n}) \geq 1 - c^n/2$, where $c > 0$ is the constant from (9). Then

$$\begin{aligned} &\mathbb{P}\left(\forall i,\ UX_i - \sqrt{n} a_i \in CK \quad \text{and} \quad |X_i| \leq C_3\sqrt{n} \right) \\ &= \prod_{i=1}^m \gamma_E\left(\left(\sqrt{n} a_i + CK \right) \cap C_3\sqrt{n} B^{2n} \right) \geq \exp(-\hat{C} n m). \end{aligned} \tag{10}$$

We move on to bounding the second summand on the right-hand side of (8). We condition on the X_i's satisfying the event described in (10). In particular, $|X_i| \leq C_3\sqrt{n}$ for all i. We now define

$$T = \left\{ y \in R^n ; \left| \left\langle \frac{y}{\sqrt{n}}, \frac{X_i}{\sqrt{n}} \right\rangle \right| \leq \varepsilon \text{ for } i = 1, \ldots, m \right\}.$$

Arguing as before, we deduce from the Khatri-Sidak lemma that $\gamma_n(T) \geq e^{-Cn}$. Write $P(x_1, \ldots, x_{2n}) = (x_1, \ldots, x_n)$. Then for any j,

$$\mathbb{P}\left(\forall i, \left|\left\langle \frac{X_i}{\sqrt{n}}, \frac{Y_j}{\sqrt{n}} - U^{-1} b_j \right\rangle\right| \leq \varepsilon\right) = \gamma_n\left(T + \sqrt{n} P(U^{-1} b_j)\right) \geq e^{-n\|b_j\|^2/2} \gamma_n(T) \geq e^{-\tilde{C}n}.$$

Next we set $\tilde{C}_3$ to be a sufficiently large universal constant such that $\mathbb{P}(|Y_i| \leq \tilde{C}_3 \sqrt{n}) \geq 1 - \exp(-\tilde{C}n)/2$.

To summarize, with probability at least $\exp(-\hat{C}nm)$, for all i, j,

$$U X_i - \sqrt{n} a_i \in CK, \quad \left|\left\langle \frac{X_i}{\sqrt{n}}, \frac{Y_j}{\sqrt{n}} - U^{-1} b_j \right\rangle\right| \leq \varepsilon \quad \text{and} \quad |X_i| + |Y_i| \leq \tilde{C}\sqrt{n}.$$

We thus have an upper bound of $\bar{C}\varepsilon$ for the right-hand side of (8) for all i, j, and moreover, $|X_i| + |Y_i| \leq \tilde{C}\sqrt{n}$ for all i. This implies a variant of Theorem 5.2, in which the ε in (3) is replaced by $\bar{C}\varepsilon$. However, by adjusting the constants, this variant is clearly seen to be equivalent to the original formulation, and the proof is complete. □

5.3 Keeping the inner products with small distortion

In this subsection we prove Theorem 1.3. 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., [4, Chapter 7]:

Theorem 5.6. *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 [4]. 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, as in the proof of Corollary 5.5, we see that

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

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 [4, Lemma 5.2.3] it is explained how concentration inequalities upgrade this fact to the desired bound $M(K) \leq C/\sqrt{n}$. $\square$

Our next observation is that the assumption $\gamma_n(K) \geq 1/2$ in Theorem 5.6 is too strong, and may be weakened to the requirement that $\gamma_n(K) \geq \exp(-ct)$.

Theorem 5.7. *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 5.6, is at most $e^{-c_0 t}$.

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

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

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 5.6, with probability at least $1 - C\exp(-ct)$ of selecting E,

$$\tilde{c}\sqrt{t} B_E \subseteq Proj_E(T) = Proj_E(K) + Proj_E\left(\frac{\tilde{c}\sqrt{t}}{2} B^n\right) = Proj_E(K) + \frac{\tilde{c}\sqrt{t}}{2} B_E. \qquad (11)$$

Since B_E and $Proj_E(K)$ are convex, we deduce from (11) that $(\tilde{c}\sqrt{t}/2)B_E \subseteq Proj_E(K)$, completing the proof. $\square$

Corollary 5.8. *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 $v \in R^n$,*

$$|v| \leq \sqrt{t} \qquad \Longrightarrow \qquad E \cap (v + CK) \neq \emptyset. \qquad (12)$$

Proof. Write $F = E^{\perp}$. Condition (12) is equivalent to $\sqrt{t}B_F \subseteq Proj_F(CK)$. The corollary thus follows from Theorem 5.7 with $C = 1/c_1$. $\square$

Proof of Theorem 1.3. 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 5.8. That is, $c_0 t \geq 5 \cdot \varepsilon^{-2} \log(2 + \varepsilon^2 n)$, thus

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

As in the proof of Theorem 5.2 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

$$\left| \left\langle \frac{X_i}{\sqrt{t}}, \frac{Y_j}{\sqrt{t}} \right\rangle - \langle a_i, b_j \rangle \right| \leq \left| \left\langle \frac{UX_i}{\sqrt{t}} - a_i, b_j \right\rangle \right| + \left| \left\langle \frac{X_i}{\sqrt{t}}, \frac{Y_j}{\sqrt{t}} - U^{-1} b_j \right\rangle \right|. \tag{13}$$

We next bound the first summand on the right-hand side of (13). (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 5.8. By the Khatri-Sidak lemma

$$\begin{aligned} \gamma_n(K) &\geq \prod_{j=1}^{n} \gamma_n \left(\left\{ x \in R^n \,;\, \left| \left\langle \frac{x}{\sqrt{t}}, b_j \right\rangle \right| \leq \sqrt{c_0}\varepsilon \right\} \right) = \prod_{j=1}^{n} \left(1 - 2\Phi(\sqrt{c_0 t}\varepsilon / |b_j|) \right) \\ &\geq \left(1 - 2\Phi\left(2\sqrt{\log(2 + n/(c_0 t))} \right) \right)^n \geq \left(1 - \frac{c_0 t}{n + c_0 t} \right)^n \geq e^{-c_0 t}. \end{aligned}$$

By Corollary 5.8 there exists a t-dimensional subspace $E \subseteq R^n$ such that for any $v \in R^n$

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

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 (13) for all i, j, implying a variant of Theorem 1.3 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. $\square$

6 Concluding remarks

- By the first two parts of Theorem 1.1, $f(n, n, 2\varepsilon)$ 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 [12], the $\frac{\log n}{\varepsilon^2}$ bound in the Johnson-Lindenstrauss Lemma [8] is tight. The first part of Corollary 1.2 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 1.1.

- It is worth noting that in the proof of Theorem 3.1 the inner product of each rounded vector with itself is typically not close to the square of its original value 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 1.1 in Section 2 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 that of distinct w_i, w_j.

- The problem of maintaining all square distances between the points up to a relative error of ε is more difficult than the one considered here. Our lower bounds, of course, hold, see [7] for the best known upper bounds. For this problem there is still a logarithmic gap between the upper and lower bounds.

- The assertion of Theorem 1.5 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 1.3 for $\varepsilon = O(1/\sqrt{n})$ is that for any $2n$ by $2n$ matrix A satisfying $\gamma_2(A) \leq 1$ there is a $2n$ by $2n$ matrix B of rank at most n 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 [3], Theorem 1.2) that if A is a $2n$ by $2n$ Hadamard matrix then any B as above has rank at least $2n - O(1)$.

- Conjecture 1.4 remains open, it seems tempting to try to iterate the assertion of Theorem 1.5 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 [11].

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

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

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

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

[7] P. Indyk and T. Wagner, Near-Optimal (Euclidean) Metric Compression, arXiv 1609.06295, 2016.

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

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

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

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

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

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

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