A bound on l_1 codes

Alex Samorodnitsky

Abstract

We give reasonably tight bounds on the maximal number of points in the unit ball of ℓ_1^n with pairwise distances at least $2 - \epsilon$, for a small parameter ϵ.

The same result was obtained independently and somewhat earlier by Lee and Moharrami (see [2], where it is stated in terms of metric embeddings). Our approach seems to be different, in that we essentially reduce the problem to the question of existence of certain constant weight binary codes.

1 Introduction

This note deals with the following question of Ilya Razenshteyn [1]:

Given a parameter ϵ, how many points $x_1, \ldots, x_N$ can there be in $\mathbb{R}^n$, such that the ℓ_1 norm of each point is at most 1, and the pairwise ℓ_1 distances between x_i are at least $2 - \epsilon$. Specifically, taking $N(\epsilon) = N(n, \epsilon)$ to be the maximal number of such points, how fast does $\log N(\epsilon)$ go to zero with ϵ?

We prove the following claim:

$$2^{O(\epsilon^2) \cdot n} \le N(\epsilon) \le 2^{O\left(\epsilon^2 \log 1/\epsilon\right) \cdot n} \tag{1}$$

The same result was obtained independently (and earlier) by Lee and Moharrami [2], in terms of metric embeddings. They prove that to embed the metric of a star on $N + 1$ vertices (i.e., a graph with one central vertex connected to N peripheral vertices) in ℓ_1^n with $(1 + \epsilon)$-distortion requires

$$n \ge \Omega\left(\frac{N}{\epsilon^2 \log 1/\epsilon}\right)$$

Our argument seems to be different from that of [2]. In fact, we essentially reduce the question to the question of existence of certain constant weight binary codes. Let $A(n, w, d)$ be the cardinality of the best constant weight binary code of length n, weight w and distance d. Then, for any constant b (independent of n and ϵ) and assuming all relevant values to be integers from now on

$$N(\epsilon) \ge A\left(n, b\epsilon n, (2 - \epsilon) \cdot b\epsilon n\right) \tag{2}$$

On the other hand

$$N(\epsilon) \leq 2^{O\left(\epsilon^4 \log 1/\epsilon\right) \cdot n} \cdot \sum_{r=0}^{4\epsilon n} A\left(2n, r, 2r - 4\epsilon^2 n\right) \tag{3}$$

It should be noted that for any constant $b < 1/2$ holds

$$A\left(n, b\epsilon n, (2-\epsilon) \cdot b\epsilon n\right) \geq 2^{\Omega\left(\epsilon^2\right) \cdot n}$$

and for any w and any constant a (independent of n and ϵ) holds

$$A\left(n, w, 2w - a\epsilon^2 n\right) \leq 2^{O\left(\epsilon^2 \log 1/\epsilon\right) \cdot n}$$

The lower bound in the constant weight analog of the Gilbert-Varshamov bound, and the upper bound follows from a Bassalygo-Elias type of argument, which we provide for completeness (see Lemma 1.4 below).

Lower bound: The lower bound is trivial. Choose $b < 1/2$, take $C = \{y_1, ..., y_N\}$ to be a constant weight binary code of weight $w = b\epsilon n$ and distance $d = b\epsilon n - b\epsilon^2 n$, and then take $x_i = \frac{1}{w} \cdot y_i$. The lower bound in (2) follows.

1.1 The upper bound

We assume, w.l.o.g., that all x_i are of unit length. Fix a parameter $\delta = 1/(4\epsilon n)$. Decompose each x_i as $x_i = u_i + v_i$, where for each coordinate $1 \leq k \leq n$

$$\begin{cases} u_i(k) = x_i(k) & \text{if } |x_i| \leq \delta \\ u_i = 0 & \text{otherwise} \end{cases}$$

and v_i is defined accordingly. Note that u_i and v_i have disjoint supports and therefore $\|x_i\| = \|u_i\| + \|v_i\|$.

Let $S = S(\delta) := \sum_{i=1}^N \|u_i\|$. (Then $\sum_{i=1}^N \|v_i\| = N - S$.) We claim that S can't be too large.

Lemma 1.1:

$$S \leq \left(\frac{1}{2} + o_N(1)\right) \cdot N$$

Proof: Consider the sum of all pairwise distances between x_i:

$$D = \sum_{i \neq j} \|x_i - x_j\|$$

On one hand

$$D \geq (2-\epsilon) \cdot (N-1)N$$

One the other hand

$$D \le \sum_{i \ne j} \|u_i - u_j\| + \sum_{i \ne j} \|v_i - v_j\| =: U + V$$

We bound U and V separately. First

$$V = \sum_{i \ne j} \|v_i - v_j\| \le \sum_{i \ne j} (\|v_i\| + \|v_j\|) = 2 \cdot (N-1)(N-S)$$

(Recall that $\sum \|v_i\| = N - S$.)

We have to work slightly harder to bound U. First we deal with the one-dimensional case.

Lemma 1.2: *Let $y_1, ..., y_N$ be numbers in the interval $[-\delta, \delta]$ with $\sum_{i=1}^N |y_i| = s$. Then, up to negligible error,*

$$\sum_{i \ne j} |y_i - y_j| \le 2Ns - s^2/\delta$$

Proof: (A somewhat sketchy one). By local shifting, the maximum of LHS is attained when all the y_i (maybe except two) are located either at the endpoints $\pm\delta$ or at zero. Simple optimization shows that (up to negligible error) the best case is when the same number $k = s/(2\delta)$ of y_i are at both endpoints, and then the calculation gives the claim of the lemma. ∎

Now, we can bound U. Note that

$$U = \sum_{i \ne j} \|u_i - u_j\| = \sum_{k=1}^{n} \sum_{i \ne j} \|u_i(k) - u_j(k)\|$$

where for each k between 1 and n, $y_i := u_i(k)$ satisfy the conditions of the lemma. Let $s_k = \sum_{i=1}^N |u_i(k)|$. Then $\sum_{k=1}^n s_k = \sum_{i=1}^N \|u_i\| = S$, and, by the lemma, and by simple optimization

$$U = \sum_{k=1}^{n} \sum_{i \ne j} \|u_i(k) - u_j(k)\| \le \sum_{k=1}^{n} \left(2Ns_k - s_k^2/\delta\right) \le 2NS - S^2/(\delta n)$$

where the maximum of RHS is attained when all s_k equal S/n.

Altogether, we have

$$(2-\epsilon)(N-1)N \le D = U + V \le 2(N-1)(N-S) + 2NS - S^2/(\delta n) < 2N^2 - 4\epsilon \cdot S^2,$$

$$S \le (1/2 + o(1)) \cdot N$$

This completes the proof of Lemma 1.1. ∎

By Markov's inequality, omitting the negligible error of $o_N(1)$, this means that at least $\frac{1}{3}N$ of vectors u_i satisfy $\|u_i\| \le 3/4$. Let I be the set of indices of these vectors. Let $i \ne j \in I$. Note that both $\|v_i\|$ and $\|v_j\|$ are at least 1/4 and

$$\|v_i - v_j\| \ge \|x_i - x_j\| - \|u_i - u_j\| \ge (\|x_i\| + \|x_j\|) - (\|u_i\| + \|u_j\|) - \epsilon = \|v_i\| + \|v_j\| - \epsilon$$

Now, we are ready to prove (3). We start with a simple lemma.

Lemma 1.3: *Let C be a set of binary vectors of weight at most $w < n/2$ such that the intersection between (the supports of) any two distinct vectors is of cardinality at most k. Then*

$$|C| \leq \sum_{r=0}^{w} A(n, r, 2r - 2k)$$

Proof: For $0 \leq r \leq w$, let C_r be the set of vectors of weight r in C. Then C_r is a constant weight code of weight r and distance at least $2r - 2k$, and C_r partition C. The claim follows. ∎

Now, for each v_i with $i \in I$, construct a binary vector b_i of length $2n$ as follows

$$\begin{cases} b_i(2k-1), b_i(2k) = 00 & \text{if } v_i(k) = 0 \\ b_i(2k-1), b_i(2k) = 01 & \text{if } v_i(k) < 0 \\ b_i(2k-1), b_i(2k) = 10 & \text{if } v_i(k) > 0 \end{cases}$$

Since all the non-zero coordinates in v_i are of absolute value at least δ, the weight of each b_i is at most $1/\delta$ and, since $\|v_i - v_j\| \geq \|v_i\| + \|v_j\| - \epsilon$, any two distinct vectors intersect in at most $\epsilon/(2\delta)$ positions. Take $w = 1/\delta = 4\epsilon n$, and $k = \epsilon/(2\delta) = 2\epsilon^2 n$. By the preceding lemma, the number of distinct b_i is at most $\sum_{r=0}^{w} A(2n, r, 2r - 2k)$, which accounts for the main term in (3).

We will complete the proof by arguing that at most $2^{O\left(\epsilon^4 \log 1/\epsilon\right) \cdot n}$ distinct vectors v_i are mapped into a given binary vector b. We do this in two steps.

First, note that if at least two vectors are mapped into b then, by preceding reasoning, the weight of b is at most $d = \epsilon/(2\delta) = O\left(\epsilon^2 n\right)$. The set of vectors mapped into b lie in the unit ball of the l_1^d with pairwise distances at least $1/2 - \epsilon$. Therefore, by a standard volume argument, their number is at most $2^{O\left(\epsilon^2\right) \cdot n}$. Therefore, recalling $w = 1/\delta = 4\epsilon n$, and using Lemma 1.4

$$N(n, \epsilon) \leq 2^{O\left(\epsilon^2\right) \cdot n} \cdot \sum_{r=0}^{4\epsilon n} A\left(2n, r, 2r - 4\epsilon^2 n\right) \leq 2^{O\left(\epsilon^2\right) \cdot n} \cdot 2^{O\left(\epsilon^2 \log 1/\epsilon\right) \cdot n} = 2^{O\left(\epsilon^2 \log 1/\epsilon\right) \cdot n}$$

This, in particular, already proves (1).

Next, observe that the vectors $v_1, \ldots, v_M$ mapped into b satisfy $\|v_i\| \geq 1/4 - \epsilon$, and, for distinct i, j, $\|v_i - v_j\| \geq \|v_i\| + \|v_j\| - \epsilon$. It is easy to see that the normalized vectors $y_i = v_i / \|v_i\|$ satisfy $\|y_i - y_j\| \geq 2 - 4\epsilon$. This gives

$$M \leq N(d, 4\epsilon) \leq 2^{O\left(\epsilon^4 \log 1/\epsilon\right) \cdot n}$$

completing the proof of (3).

1.2 An estimate on constant weight codes

Lemma 1.4: *Let a be a constant independent of n and ϵ. Then there exists a constant $c = c(a)$ such that for any $0 \leq w \leq n/2$ and $d = 2w - a\epsilon^2 n$ holds*

$$A(n, w, d) \leq 2^{c \cdot \epsilon^2 \log 1/\epsilon \cdot n}$$

Proof: The claim of the lemma will follow from a combination of two well-known steps: The Johnson bound states that for $d > 2w - 2w^2/n + 1$ holds $A(n, w, d) < n$. In particular, we may assume that $d \leq 2w - 2w^2/n + 1$, since otherwise we are done.

The Bassalygo-Elias argument observes ([3]), that for any two radii $0 < w < w' < n/2$, the Hamming sphere of radius w' is covered by at most $O\left(n^2\right) \cdot \frac{\binom{n}{w'}}{\binom{n}{w}}$ spheres of radius w. Consequently, taking w_0 to be the largest integer such that $d > 2w_0 - 2w_0^2/n + 1$, we have

$$A(n, w, d) \leq O\left(n^3\right) \cdot \frac{\binom{n}{w}}{\binom{n}{w_0}}$$

The claim of the lemma follows by a simple calculation, observing $w_0 \leq w \leq w_0 + a\epsilon^2 n/2$, and using the asymptotic estimate $\binom{n}{k} = 2^{H(k/n)\cdot n \pm o(n)}$.

∎

References

[1] Ilya Razenshteyn, personal communication, 2013.

[2] J. R. Lee and M. Moharrami, *Lower bounds for dimension reduction for trees in l1*, arXiv:1302.6542, 2013.

[3] V. I. Levenshtein, *On the minimum redundancy of binary error-correcting codes*, Problems Inform. Transmission 10, No. 2 (1974), 110-123; Inform. and Control 28 (1975), 268-291.

[4] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory, vol. IT-23, 1977, 157-166.