List-decodable zero-rate codes

Noga Alon * Boris Bukh † Yury Polyanskiy ‡

Abstract

We consider list-decoding in the zero-rate regime for two cases: the binary alphabet and the spherical codes in Euclidean space. Specifically, we study the maximal $\tau \in [0,1]$ for which there exists an arrangement of M balls of relative Hamming radius τ in the binary hypercube (of arbitrary dimension) with the property that no point of the latter is covered by L or more of them. As $M \to \infty$ the maximal τ decreases to a well-known critical value τ_L. In this work, we prove several results on the rate of this convergence.

For the binary case, we show that the rate is $\Theta(M^{-1})$ when L is even, thus extending the classical results of Plotkin and Levenshtein for $L = 2$. For $L = 3$ the rate is shown to be $\Theta(M^{-\frac{2}{3}})$.

For the similar question about spherical codes, we prove the rate is $\Omega(M^{-1})$ and $O(M^{-\frac{2L}{L^2-L+2}})$.

1 Introduction

This work concerns list-decoding under *worst-case* errors in the zero-rate regime. We consider the case of the binary alphabet in Sections 1-7 and the case of the unit sphere in Hilbert space in Section 8.

Specifically, suppose we transmit a sequence of n symbols from $\{0,1\}$ over a channel that can adversarially change less than fraction τ of the symbols. The locations of corrupted symbols are unknown to the receiver. The goal is to find a code $C \subset \{0,1\}^n$ such that the receiver can always produce a list of fewer than L messages containing the transmitted message. In other words, we seek a code C such that for every $w \in \{0,1\}^n$ there are fewer than L codewords within Hamming distance τn from w. We call such code $<L$-*list-decodable* with radius τ. The largest such τ is denoted by $\tilde{\rho}_L(C)$ and is called the *(normalized) L-radius* of the code.

Let

$$\tau_L \stackrel{\text{def}}{=} \frac{1}{2} - \frac{\binom{2k}{k}}{2^{2k+1}} \quad \text{if } L = 2k \text{ or } L = 2k+1. \tag{1}$$

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and CMSA, Harvard University, Cambridge, MA 02138, USA. Email: nogaa@tau.ac.il. Research supported in part by a BSF grant, an ISF grant and a GIF grant.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by Sloan Research Fellowship and by U.S. taxpayers through NSF CAREER grant DMS-1555149 and NSF grant DMS-1301548. Part of the work was done during the visit to Université Paris-Est Marne-la-Vallée supported by LabEx Bézout (ANR-10-LABX-58).

‡Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA. Email: yp@mit.edu. Supported in part by the NSF under Grant No CCF-13-18620 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

It is known [3] that for radius $\tau < \tau_L$ the largest $<L$-list-decodable code is exponentially large in n, and for radius $\tau > \tau_L$ the largest $<L$-list-decodable code is of constant size. The aim of this paper is to understand how this constant varies as τ approaches τ_L from above. We define

$$\mathrm{maxcode}_L(\varepsilon) \stackrel{\text{def}}{=} \max\bigl\{|C| : C \subset \{0,1\}^n \text{ is } <L\text{-list-decodable of radius } \tau_L + \varepsilon\bigr\}.$$

Note that in this definition we do not restrict the block length n. The maximum is over all choices of $n \in \mathbb{N}$.

We are aware of three results on $\mathrm{maxcode}_L(\varepsilon)$. First, a construction due to Levenshtein [8] shows that the so-called Plotkin bound is sharp in the unique decoding case, namely

$$\mathrm{maxcode}_2(\varepsilon) = \frac{1}{4\varepsilon} + O(1).$$

Levenshtein's construction uses Hadamard matrices, and so requires ε to be of a special form. As a part of Theorem 1 below we present a construction without a condition on ε.

Second, Blinovsky [3] proved that $\mathrm{maxcode}_L(\varepsilon)$ is finite for every L and every $\varepsilon > 0$. His proof iterates Ramsey's theorem, and gives a very large bound on $\mathrm{maxcode}_L(\varepsilon)$ (which is not made explicit in the paper). Finally, in [5, Theorem 1] Blinovsky claims an upper bound on $\mathrm{maxcode}_L(\varepsilon)$ of the form $\mathrm{maxcode}_L(\varepsilon) = O(1/\varepsilon)$. Below we construct a counterexample to this bound for $L = 3$.[1]

We next overview our results for the binary alphabet.

1.1 Our results (binary alphabet)

Our first result is a version of Levenshtein's construction for any fixed L. In comparison to Levenshtein's result, we have no restriction on ε, but our codes are longer (the value of n is larger).

Theorem 1. *Let $L \geq 2$, suppose m is a positive integer, and let $M = \binom{2m}{m}$. Let $c_L = 2^{-L}\lfloor L/2 \rfloor \binom{L-1}{\lfloor L/2 \rfloor}$. Then there exists an $<L$-list-decodable code in $\{0,1\}^M$ of size $2m$ and radius $\tau = \tau_L + c_L/2m + O(m^{-2})$. In particular,*

$$\mathrm{maxcode}_L(\varepsilon) \geq c_L \varepsilon^{-1} + O(1).$$

Theorem 2. *Let $L \geq 2$ be even. Then*

$$\mathrm{maxcode}_L(\varepsilon) = O(\varepsilon^{-1})$$

We believe that in fact $\max_L(\varepsilon) = c_L\varepsilon^{-1} + O(1)$ for even L.

The case of odd L appears to be significantly harder: the principal reason for this is Lemma 8(b) below. By a different method, however, we were able to make progress for $L = 3$:

Theorem 3. *We have* $\mathrm{maxcode}_3(\varepsilon) = \Theta(\varepsilon^{-3/2})$.

[1] The mistake appears to stem from the second paragraph of the proof of [5, Theorem 1], which proposes a certain extension procedure for codes and claims that it does not decrease L-radius. A simple counter-example to the claim is a code $C = \{0,1\}^2$ with 4-radius equal to 1, but its extension results in reduction of the 4-radius to $\frac{1}{2}$.

2 Mean radii

Definitions For $x \in \mathbb{R}^n$, let $\|x\| \stackrel{\text{def}}{=} (1/n)\sum|x_i|$. In particular, for $x, y \in \{0,1\}^n$ the quantity $\|x-y\|$ is the (normalized) Hamming distance between bit strings x and y.

For points $x^{(1)}, \ldots, x^{(L)} \in \{0,1\}^n$ let

$$\operatorname{rad}(x^{(1)}, \ldots, x^{(L)}) = \min_{y \in [0,1]^n} \max_i \|x^{(i)} - y\|. \tag{2}$$

Note that we allow the coordinates of y to be arbitrary real numbers between 0 and 1. This relaxation makes only slight effect, as the next proposition shows.

Proposition 4. *Let $x^{(1)}, \ldots, x^{(L)} \in \{0,1\}^n$. If $\tau = \operatorname{rad}(x^{(1)}, \ldots, x^{(L)})$, then there is a point $y \in \{0,1\}^n$ such that $\|x^{(i)} - y\| \le \tau + \frac{L}{2n}$.*

Proof. For any bit $z \in \{0,1\}$ and real $w \in [0,1]$ define $\ell(z,w) = w$ if $z = 0$ and $\ell(z,w) = 1 - w$ if $z = 1$. Note that with this notation, for every i, $\|x^{(i)} - y\| = \frac{1}{n}\sum_{j=1}^n \ell(x_j^{(i)}, y_j)$ is an affine function of the variables y_j.

The assumption $\operatorname{rad}(x^{(1)}, \ldots, x^{(L)}) \le \tau$ is equivalent to the fact that the polytope in the variables y_j defined by the inequalities $0 \le y_j \le 1$ for all $1 \le j \le n$ and $\frac{1}{n}\sum_{j=1}^n \ell(x_j^{(i)}, y_j) \le \tau$ for all $1 \le i \le L$ is nonempty. Hence it contains a vertex $y' = (y'_1, \ldots, y'_n)$. In this vertex there are at most L variables y'_j which are neither 0 nor 1, and the desired result is obtained by rounding each such y'_j to the closest integer y_j and by taking $y_j = y'_j$ for all other coordinates y'_j. □

Let ω be a probability distribution on $\{1, 2, \ldots, L\}$ and let Ω_L be the set of all probability measures on the set $[L]$. Then for an L-tuple $x = (x^{(1)}, \ldots, x^{(L)}) \in (\{0,1\}^n)^L$ of codewords, we define their *mean-radius* with respect to ω by

$$\operatorname{mrad}_\omega(x) \stackrel{\text{def}}{=} \min_{y \in [0,1]^n} \mathop{\mathbb{E}}_{i \sim \omega} \|x^{(i)} - y\|. \tag{3}$$

Because $\mathbb{E}_{i\sim\omega}\|x^{(i)} - y\|$ can be written as a sum over the individual coordinates, the y attaining minimum in (3) may be taken to have all of its coordinates in $\{0,1\}$. This leads to an alternative formula for $\operatorname{mrad}_\omega(x)$:

$$\operatorname{mrad}_\omega(x) = \mathop{\mathbb{E}}_{j \in [n]} \min\left(\sum_{x_j^{(i)}=0} \omega(i), \sum_{x_j^{(i)}=1} \omega(i) \right) \tag{4}$$

$$= \tfrac{1}{2} - \tfrac{1}{2} \mathop{\mathbb{E}}_{j \in [n]} \left| \sum_{x_j^{(i)}=0} \omega(i) - \sum_{x_j^{(i)}=1} \omega(i) \right| \tag{5}$$

Duality From the comparison of (2) and (3) it is clear that $\operatorname{rad}(x) \ge \operatorname{mrad}_\omega(x)$ for any ω. The key observation is that a suitable converse holds as well.

Lemma 5. *For every $x = (x^{(1)}, \dots, x^{(L)}) \in (\{0,1\}^n)^L$ we have*

$$\operatorname{rad}(x) = \max_{\omega \in \Omega_L} \operatorname{mrad}_\omega(x), \tag{6}$$

where the maximum is over all probability measures ω on $\{1, 2, \dots, L\}$.

Proof. Notice that (2) can be rewritten as

$$\operatorname{rad}(x) = \min_{y \in [0,1]^n} \max_{\omega \in \Omega_L} \mathop{\mathbb{E}}_{i \sim \omega} \|x^{(i)} - y\|.$$

Since the function

$$(y, \omega) \mapsto \mathop{\mathbb{E}}_{i \sim \omega} \|x^{(i)} - y\|,$$

is convex in y and affine in ω, von Neumann minimax theorem [12] implies

$$\min_{y \in [0,1]^n} \max_{\omega \in \Omega_L} \mathop{\mathbb{E}}_{i \sim \omega} \|x^{(i)} - y\| = \max_{\omega \in \Omega_L} \min_{y \in [0,1]^n} \mathop{\mathbb{E}}_{i \sim \omega} \|x^{(i)} - y\|.$$

Comparing with (3) completes the proof.

□

Lemma 6. *For every L there exists a finite set of probability measures $\Omega'_L \subset \Omega_L$ such that*

$$\operatorname{rad}(x) = \max_{\omega \in \Omega'_L} \operatorname{mrad}_\omega(x). \tag{7}$$

Furthermore, $|\Omega'_L| \le 4^{L^2}$.

Proof. To each coordinate $j \in [n]$ we can associate the bit string

$$T(j) \stackrel{\text{def}}{=} \left(x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(L)}\right).$$

For a bit string $T \in \{0,1\}^L$, put $P_T = \{j \in [n] : T(j) = T\}$. Let $\alpha_T = |P_T|/n$. We have that

$$\|x^{(i)} - y\| = \sum_{T \in \{0,1\}^L} \alpha_T |y_T - T_i|.$$

To each probability measure $\omega \in \Omega_L$ we can associate a *signature*, which is a function $S_\omega \colon \{0,1\}^L \to \{1, -1\}$ defined by

$$S_\omega(T) \stackrel{\text{def}}{=} \operatorname{sign}\left(\sum_{i:T_i=0} \omega_i - \sum_{i:T_i=1} \omega_i\right) \qquad \text{for } T \in \{0,1\}^L.$$

Since 2^L hyperplanes partition $\mathbb{R}^{L-1}$ into at most $\sum_{j \le L-1} \binom{2^L}{j} \le 2^{L^2}$ regions, the number of possible signatures is at most 2^{L^2}. For each possible signature S, let $\Omega_S \stackrel{\text{def}}{=} \{\omega \in \Omega_L : S_\omega = S\}$. Since Ω_S is an intersection of halfspaces, which, in addition to $S_\omega = S$, includes the additional inequalities $\omega_i \ge 0$ for all i and $\sum_i \omega_i = 1$, it is a convex polytope.

By (4), the maximum of $\mathrm{mrad}_\omega(x)$ over all probability measures $\omega \in \Omega_S$ is the maximum of the following linear function in the variables ω_i:

$$\sum_T \alpha_T f(\omega, T)$$

where

$$f(\omega, T) = \sum_{i:T_i=0} \omega_i \qquad \text{if } S(T) = -1$$

and

$$f(\omega, S, i) = \sum_{i:T_i=1} \omega_i \qquad \text{if } S(T) = +1.$$

This maximum is attained at a vertex of the polytope Ω_S. Thus, in view of the preceding lemma, we may take Ω'_L to be the union of the vertex sets of all polytopes Ω_S, for all signatures S.

Each Ω_S is defined by $m \stackrel{\text{def}}{=} L + 2^L$ inequalities, and so by McMullen's upper bound theorem has at most $\binom{m-\lceil L/2 \rceil}{\lfloor L/2 \rfloor} + \binom{m-\lfloor L/2 \rfloor - 1}{\lceil L/2 \rceil - 1} \leq 2^{L^2}$ vertices. Multiplying by the 2^{L^2} possible signatures, we obtain the result. □

The preceding proof gives an algorithm to compute the set Ω'_L. The results of this computation for small L can be found at http://www.borisbukh.org/code/listdecoding17.html. Interestingly, for $L \leq 4$ the result is very nice. For a set $R \subset [L]$ let $\mathrm{mrad}_R(x)$ be the $\mathrm{mrad}_\omega(x)$ for the probability measure ω that is uniform on R, i.e., $\omega_i = 1/|R|$ if $i \in R$. Then

$$\mathrm{rad}(x) = \max_{|R| \text{ is even}} \mathrm{mrad}_R(x) \qquad \text{if } L \leq 4 \tag{8}$$

Our proof of Theorem 3 will use (8) with $L = 3$ and so establish it formally (generalized to arbitrary ℓ_1-vectors).

Proposition 7. *For any set of three vectors x, y, z in $\mathbb{R}^n$ with respect to the ℓ_1-norm,* $\mathrm{rad}(x, y, z) = \frac{1}{2}\,\mathrm{diam}(x, y, z)$.

Proof. Put $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, $z = (z_1, z_2, \ldots, z_n)$. Let d be the diameter of the set $\{x, y, z\}$. For each i let m_i be the median of x_i, y_i, z_i and define $m = (m_1, m_2, \ldots, m_n)$. Put, also, $a = \|x - m\|, b = \|y - m\|, c = \|z - m\|$, where $\|\cdot\|$ is the ℓ_1-norm. Note that, crucially $\|x - y\| = a + b, \|y - z\| = b + c, \|x - z\| = a + c$. Thus each of these three sums is at most d. If each of the quantities a, b, c is at most $d/2$ then m is a center of an ℓ_1-ball of radius $d/2$ containing x, y, z, showing that in this case the radius is indeed $d/2$, as needed. Otherwise one of the above, say a, is larger than $d/2$. In this case define $w = (1 - \frac{d}{2a})x + \frac{d}{2a}m$. It is easy to check that $x - w = \frac{d}{2a}(x - m)$ and hence $\|x - w\| = \frac{d}{2a}a = d/2$. In addition $m - w = (1 - \frac{d}{2a})(m - x)$ and hence $\|m - w\| = a - \frac{d}{2}$.

Thus, by the triangle inequality, $\|y - w\| \leq \|y - m\| + \|m - w\| = b + a - \frac{d}{2} \leq \frac{d}{2}$, and similarly $\|z - w\| \leq \frac{d}{2}$, completing the proof. □

3 Averaging

Averaging arguments play a major role in this paper. We collect them in this section.

Mean radii of random bit strings Averaging arguments will allow us to show that, in a large code C, mean radii of codewords from C rarely exceed the mean radius of random bit strings. Because of that, we start by computing the mean radius of random bit strings for arbitrary probability measure ω. In particular, we will see that the radius threshold τ_L defined in (1) is the radius of a random L-tuple of bit strings.

Call a random string $w \in \{0,1\}^n$ *p-biased* if each bit of w is 1 with probability p and 0 with probability $1-p$, and the bits are independent of each other. For a probability measure ω on $[L]$, we let

$$\tau_{\omega,p} \stackrel{\text{def}}{=} \mathbb{E}\,\mathrm{mrad}_\omega(w^{(1)},\dots,w^{(L)}) \qquad \text{for independent } p\text{-biased } w^{(1)},\dots,w^{(L)} \in \{0,1\}^n.$$

For brevity, write τ_ω in lieu of $\tau_{\omega,1/2}$. Note that $\tau_{\omega,p}$ is independent of n, in view of (4).

Let $\mathcal{U}[L]$ denote the uniform probability measure on $[L]$.

Lemma 8.

a) For every probability measure ω on $[L]$ and every p we have

$$\tau_{\omega,p} \le \tau_{\mathcal{U}[L],p}.$$

b) If L is even and $0<p<1$, then the equality holds if and only if $\omega = \mathcal{U}[L]$.

c) $\tau_{\mathcal{U}[L],1/2} = \tau_L$, where τ_L is defined in (1).

d) We have $\tau_{\omega,p} < \tau_L$ whenever $p \ne \frac{1}{2}$.

Proof. Given an ω and a vector of signs $\varepsilon = (\varepsilon_1,\dots,\varepsilon_L) \in \{1,-1\}^L$, define $f_\omega(\varepsilon) \stackrel{\text{def}}{=} \sum_i \varepsilon_i\omega(i)$. By (5), maximization of τ_ω is equivalent to minimization of

$$\mathop{\mathbb{E}}_{\varepsilon}|f_\omega(\varepsilon)| \qquad \text{for } p\text{-biased } \varepsilon \in \{1,-1\}^L.$$

Let Ω be the set of all probability measures on $[L]$ that maximize τ_ω. (The maximum is achieved because $\omega \mapsto \tau_\omega$ is a continuous function on a compact set.) Let $\omega \in \Omega$ be arbitrary. Suppose ω is not uniform. Without loss generality, $\omega(L-1) \ne \omega(L)$. Let ω' be obtained from ω by replacing the values of $L-1$ and L by their averages. If $\varepsilon_{L-1} = \varepsilon_L$, then $f_\omega(\varepsilon) = f_{\omega'}(\varepsilon)$. Also,

$$|f_\omega(\varepsilon_1,\dots,\varepsilon_{L-2},1,-1)|+|f_\omega(\varepsilon_1,\dots,\varepsilon_{L-2},-1,1)| \ge |f_{\omega'}(\varepsilon_1,\dots,\varepsilon_{L-2},1,-1)|+|f_{\omega'}(\varepsilon_1,\dots,\varepsilon_{L-2},-1,1)|.$$

with equality only if $|\sum_{i\le L-2}\varepsilon_i\omega(i)| \ge |\omega(L-1)-\omega(L)|$. Since $\omega \in \Omega$, it follows that the equality does hold, and that $\omega' \in \Omega$ as well.

Since the equality holds, we deduce that

$$|\omega(i)-\omega(i')| \le \min_{\varepsilon}|\sum_{j\notin\{i,i'\}}\varepsilon_j\omega(j)| \qquad \text{for all } i \ne i' \tag{9}$$

From continuity of $\omega \mapsto \tau_\omega$, it follows by repeated pairwise averaging, that if ω' is obtained from ω by replacing the values of ω on any subset of $[L]$ by their averages, then $\omega' \in \Omega$ as well. In particular, $\mathcal{U}[L] \in \Omega$ and so (a) holds.

Suppose that L is even and (b) does not hold. Let $\omega \in \Omega$ be non-uniform. Without loss of generality, $\omega(L-1) \neq \omega(L)$. Let ω' be obtained from ω by replacing values on $[L-2]$ by their averages. Since $\sum_{j \le L-2}(-1)^j \omega'(j) = 0$, the measure ω' fails (9), and so $\omega' \notin \Omega$. This contradicts the assumption that (b) does not hold.

Consider a random walk on $\mathbb{Z}$ starting from 0 that makes a step to the right with probability p and to the left with probability $1-p$. Let $\Delta_{s,p}$ be the probability distribution of the position of this walk after s steps. Relation (5) implies that

$$\tau_{\mathcal{U}[L],p} = \tfrac{1}{2} - \tfrac{1}{2L}\,\mathbb{E}|\Delta_{L,p}|. \tag{10}$$

From [13, Eq. (23) and (32)] we obtain the formula (1) for τ_L.

Finally we turn to (d). In view of (a) we may restrict ourselves to the case $\omega = \mathcal{U}[L]$. Because of (10) and the symmetry under $p \to (1-p)$, it suffices to prove that $\Pr[|\Delta_{s,p}| \ge k] > \Pr[|\Delta_{s,1/2}| \ge k]$ for every $p > 1/2$ and every $s \ge 2$. This follows by induction from

$$\begin{aligned}\Pr[|\Delta_{s+1,p}| \ge k] = {} & \tfrac{1}{2}\Pr[|\Delta_{s,p}| \ge k-1] + \tfrac{1}{2}\Pr[|\Delta_{s,p}| \ge k+1] \\ & + (p - \tfrac{1}{2})\big(\Pr[\Delta_{s,p} \in \{k, k-1\}] - \Pr[\Delta_{s,p} \in \{-k, -k+1\}]\big)\end{aligned}$$ □

Mean radii in large codes Here we show that the average $\mathrm{mrad}_\omega(\cdot)$ over L-tuples in a large $C \subset \{0,1\}^n$ can be only slightly larger than τ_ω. In fact, we will show a generalization of this to codes of possibly small radius.

Lemma 9. *Let ω be a probability measure on $[L]$. Suppose $C \subset \{0,1\}^n$ satisfies $\mathrm{rad}(C) \le p \le \frac{1}{2}$. Then*

$$\mathop{\mathbb{E}}_{w^{(1)},\dots,w^{(L)} \in C} \mathrm{mrad}_\omega(w^{(1)}, \dots, w^{(L)}) \le \tau_{\mathcal{U}[L],p}.$$

Proof. Let $p_j = \Pr_{w \in C}[w_j = 1]$. We have $\mathrm{mrad}_{\mathcal{U}[L]}(C) \le p$ from (6). Without loss of generality (otherwise invert some coordinates), we may assume that y attaining $\mathrm{mrad}_{\mathcal{U}[L]}(C)$ in (4) is $y = 0$. Then we have $\frac{1}{n}\sum_{j \in [n]} p_j \le p$. Denote by $B(q)$ the distribution on $\{1,-1\}^L$ where each coordinate is independently 1 with probability q and -1 with probability $1-q$. Given a vector of signs $\varepsilon = (\varepsilon_1, \dots, \varepsilon_L) \in \{1,-1\}^L$ define $f_\omega(\varepsilon) \stackrel{\mathrm{def}}{=} \sum_i \varepsilon_i \omega(i)$.

From (5) and the proof of part (a) of Lemma 8 we then have

$$1 - 2 \mathop{\mathbb{E}}_{w \in C^L} \mathrm{mrad}_\omega(w) = \mathop{\mathbb{E}}_{j \in [n]} \mathop{\mathbb{E}}_{\varepsilon \sim B(p_j)} |f_\omega(\varepsilon)| \ge \mathop{\mathbb{E}}_{j \in [n]} \mathop{\mathbb{E}}_{\varepsilon \sim B(p_j)} \left|f_{\mathcal{U}[L]}(\varepsilon)\right|$$

By [9, Lemma 8], the function $p \mapsto \mathbb{E}_{\varepsilon \sim B(p)} \left|f_{\mathcal{U}[L]}(\varepsilon)\right|$ is convex. Jensen's inequality and the fact that $\tau_{\mathcal{U}[L],p}$ is an increasing function of p on $[0, \frac{1}{2}]$ then complete the proof. □

Corollary 10. *Let ω be a probability measure on $[L]$. Suppose $C \subset \{0,1\}^n$ is of size $|C| \ge L^2 M$ and satisfies $\mathrm{rad}(C) \le p$. Then there is an L-tuple $w \in C^L$ with distinct codewords such that $\mathrm{mrad}_\omega(w) \le \tau_{\mathcal{U}[L],p} + 1/M$.*

Proof. Let $X \subset C^L$ be the set of all L-tuples with distinct codewords. The corollary follows from $\Pr[w \notin X] \le \binom{L}{2}/|C|$ and $\mathbb{E}_{w\in C} \operatorname{mrad}_\omega(w) \ge \Pr[w \in X]\, \mathbb{E}_{w\in X} \operatorname{mrad}_\omega(w)$. □

4 Abundance of random-like L-tuples

Lemma 11. *Let $\pi\colon \mathbb{R}^n \to \mathbb{R}^m$ be an orthogonal projection on a set of m coordinates. Suppose that $C \subset \{0,1\}^n$ satisfies $\operatorname{rad}(\pi(C)) \le \frac{1}{2} - \varepsilon$. Then there is a $C' \subset C$ of size $|C'| \ge |C|/2$ satisfying $\operatorname{rad}(C') \le \frac{1}{2} - \frac{m}{n}\varepsilon$.*

Proof. Let π' be the projection on the remaining $n-m$ coordinates. Classify codewords $c \in C$ based on whether $\|\pi'(c)\| \le \frac{1}{2}$ or $> \frac{1}{2}$. Without loss of generality, at least half of them (call it C') satisfy $\|\pi'(c)\| \le \frac{1}{2}$. Let $y_1 \in \mathbb{R}^m$ be the center minimizing $\operatorname{rad}(\pi(C))$ and define $y \in \mathbb{R}^n$ to be the solution to $\pi(y) = y_1$, $\pi'(y) = 0$. We have for any $c \in C'$

$$\|y - c\| = \tfrac{m}{n}\|\pi(y) - \pi(c)\| + \tfrac{n-m}{n}\|\pi'(y) - \pi'(c)\| \le \tfrac{m}{n}\left(\tfrac{1}{2} - \varepsilon\right) + \tfrac{n-m}{2n} = \tfrac{1}{2} - \tfrac{m}{n}\varepsilon\,. \qquad \square$$

For an L-tuple $x = (x^{(1)}, \dots, x^{(L)}) \in (\{0,1\}^n)^L$, we define $\operatorname{type}(x) \stackrel{\text{def}}{=} \left(\operatorname{type}(x)_u\right)_{u\in\{0,1\}^L}$ to be the probability distribution on $\{0,1\}^L$ with $\operatorname{type}(x)_u \stackrel{\text{def}}{=} \frac{1}{n}\#\{j : x_j^{(i)} = u_i, \forall i \in [L]\}$. The next result shows that the only obstruction to finding a large number of L-tuples with approximately uniform $\operatorname{type}(x)$ is the existence of a large biased subcode.

Lemma 12. *Let s be a natural number. There exist constants $M_0 = M_0(s)$ and $c = c(s)$ such that for any $\varepsilon > 0$ there is a $\delta > 0$ with the following property. For any code $C \subset \{0,1\}^n$ with $M \stackrel{def}{=} |C| \ge M_0$ one of the following two alternatives must hold:*

a) $\exists C' \subset C$ such that $|C'| \ge s$ and $\operatorname{rad}(C') \le \frac{1}{2} - \delta$, or

b) there exists at least $M^L - cM^{L-1}$ many L-tuples of distinct codewords from C such that

$$|\operatorname{type}(x)_u - 2^{-L}| \le \varepsilon \qquad \forall u \in \{0,1\}^L \tag{11}$$

and, in particular, for any ω

$$|\operatorname{mrad}_\omega(x) - \tau_{\omega,1/2}| \le 2^L\varepsilon\,. \tag{12}$$

Consequently, if C does not satisfy a), then the the number of L-tuples of distinct codewords of C violating (11) is of size at most cM^{L-1}.

Proof. Set $2\delta_0 \stackrel{\text{def}}{=} (1 + 2^L\varepsilon)^{1/L} - 1$ and note that with this choice we have $|(\frac{1}{2} \pm \delta_0)^L - 2^{-L}| \le \varepsilon$. Set also $\mu \stackrel{\text{def}}{=} (\frac{1}{2} - \delta_0)^L$ and $\delta \stackrel{\text{def}}{=} \delta_0\mu$. Finally, set $c(s) \stackrel{\text{def}}{=} s2^{L+3}$ and $M_0(s) \stackrel{\text{def}}{=} s2^{L+3}$. Note that (11) implies (12) via (4), and so we only consider (11) below.

Let us assume that a) does not hold. Then in any C'' with $|C''| \ge 4s$ and for any orthogonal projection π_A on a subset of coordinates $A \subset [n]$ there must exist a $c \in C''$ such that

$$\|\pi_A(c)\| \in (1/2 - \delta_0, 1/2 + \delta_0), \tag{13}$$

provided that $\delta_0\frac{|A|}{n} \ge \delta$. Indeed, if all $c \in C''$ violate (13), then at least half of $c \in C$ should satisfy $\|\pi_A(c)\| \le 1/2 - \delta_0$, say. Denote this collection by C_0 and observe that $\operatorname{rad}(\pi_A(C_0)) \le 1/2 - \delta_0$ and

$|C_0| \geq 2s$. By Lemma 11 there must exist $C' \subset C_0$ of size $\geq s$ such that $\mathrm{rad}(C') \leq \frac{1}{2} - \delta$, contradicting assumption.

It follows that for any collection of subsets $A_1, \ldots, A_r$ with $|A_j| \geq \mu n$ for all $j \in [r]$, and any C'' with $|C''| \geq 4sr$ there must exist $c \in C''$ such that (13) holds simultaneously for all $A = A_j$, $j \in [r]$.

We next show that there are more than

$$N_1 = \prod_{j=0}^{L-1} (M - j - 4s \cdot 2^j)$$

L-tuples x of distinct codewords from C that satisfy (11). Indeed, at least $M - 4s$ codewords $x^{(1)}$ have $|\|x^{(1)}\| - \frac{1}{2}| \leq \delta_0$. Once one such codeword $x^{(1)}$ is selected, let $A_0 = \{j \in [n] : x_j^{(1)} = 0\}$ and $A_1 = A_0^c$. Each of these two subsets has cardinality $\geq n(\frac{1}{2} - \delta_0) \geq \mu n$. By the argument above, there are more than $M - 1 - 4s \cdot 2$ codewords $x^{(2)}$ not equal to $x^{(1)}$ such that projections of $x^{(2)}$ on A_0 and A_1 both have weights $\in [\frac{1}{2} - \delta_0, \frac{1}{2} + \delta_0]$. Selecting one such $x^{(2)}$, we define partitions $A_0 = A_{00} \cup A_{01}$ and $A_1 = A_{10} \cup A_{11}$ according to the values of coordinates of $x^{(1)}$ and $x^{(2)}$. Proceeding similarly, we construct $x^{(3)}, \ldots, x^{(L)}$. The resulting L-tuple has distinct elements and satisfies

$$(\tfrac{1}{2} - \delta_0)^L \leq \mathrm{type}(x)_u \leq (\tfrac{1}{2} + \delta_0)^L \qquad \forall u \in \{0,1\}^L \,,$$

which by the choice of δ_0 implies that it satisfies (11) as well.

Note that for $M \geq \max_j k_j$ we have

$$\prod_{j=0}^{L-1} (M - k_j) = M^L \prod_{j=0}^{L-1} (1 - k_j/M) \geq M^L - M^{L-1} \sum_{j=0}^{L-1} k_j \,.$$

Setting $k_j = s2^{j+3} \geq j + 4s \cdot 2^j$ we obtain

$$N_1 \geq M^L - cM^{L-1} \,,$$

provided $M \geq M_0$, completing the proof of the first part.

The final statement of the lemma follows from the fact that there are at most $M^L - cM^{L-1}$ many L-tuples violating (11). □

5 Proof of Theorem 2

Let L be even, and suppose $C \subset \{0,1\}^n$ is an $<L$-list-decodable code of radius $\tau_L + \varepsilon$. We wish to prove that $|C| = O(\varepsilon^{-1})$. Let $\rho_L(C) = \min_{x \in C^L} \mathrm{rad}(x)$ with the minimum taken over all L-tuples x with distinc elements. Unlike $\rho_L(C)$, the L-radius of a code (denoted $\tilde{\rho}_L(C)$) is not a well-behaved quantity. Sadly, our assumptions on C do not imply that $\rho_L(C) \geq \tau_L$. For example, if $L = 2$ then the radius of $\{000, 100\}$ is $1/3 > 1/4 = \tau_2$ whereas $\mathrm{rad}(000, 100) = 1/6$. To get around this, we use the pigeonhole principle to find a subcode C' of size $|C'| \geq 2^{-8L}|C|$ consisting of codewords with the same prefix of length $8L$. Removing the common prefix yields a code of block length $n - 8L$ whose L-radius is at least

$$\frac{n}{n - 8L}(\tau_L + \varepsilon) \geq (1 + 8L/n)\tau_L + \varepsilon \geq \tau_L + \varepsilon + 2L/n.$$

By Proposition 4 we have $\mathrm{rad}(x) \geq \tau_L + \varepsilon$ for every L-tuple x of distinct codewords from this new code. With slight abuse of notation, we rename this new code C (and adjust the value of n accordingly).

Lemma 13. *Any subset $C' \subset C$ satisfying $\mathrm{rad}(C') \leq \frac{1}{2} - \delta$ has size $|C'| < s$ for some s depending on δ.*

Proof. Let $p = \frac{1}{2} - \delta$. Fix $\omega \in \Omega'_L$ and consider any subset $C'' \subset C$ satisfying a) $\mathrm{rad}(C'') \leq \frac{1}{2} - \delta$ and b) $\mathrm{mrad}_\omega(x) \geq \tau_L$ for any L-tuple $x \in (C'')^L$ with distinct elements. From Corollary 10 and the bound $\tau_{\mathcal{U}[L],p} < \tau_L$ from Lemma 8 we know that $|C''| \leq \frac{L^2}{\tau_L - \tau_{\mathcal{U}[L],p}}$. Next, color each L-tuple x of distinct elements of C' according to $\omega \in \Omega'_L$ that solves (7). From finiteness of Ω'_L and Ramsey's theorem, we conclude that had there existed arbitrary large C' then there would have existed an arbitrary large subcode C'' satisfying conditions a) and b) above, violating the established upper bound. □

Let H be the set of all L-tuples $x \in C^L$ such that $\mathrm{mrad}_\omega(x) \geq \tau_L + \varepsilon$ for some $\omega \neq \mathcal{U}[L]$.

Lemma 14. *We have $|H| \leq c_L |C|^{L-1}$ for some constant c_L that depends only on L.*

Proof. Let $\varepsilon = 2^{-L} \min\{\tau_L - \tau_{\omega,1/2} : \omega \in \Omega'_L,\ \omega \neq \mathcal{U}[L]\}$. Since Ω'_L is finite, part (b) of Lemma 8 implies that $\varepsilon > 0$. So, let δ be as in Lemma 12. Let s be the bound from Lemma 13. Note that by the choice of ε, H consists entirely of the L-tuples violating (12) and hence (11). Also by the choice of s, alternative a) in Lemma 12 is impossible. Therefore, by the last statement of the latter Lemma, we have $|H| \leq c(s)|C|^{L-1}$. □

Proof of Theorem 2. Call an L-tuple $x \in C^L$ *good* if all of its codewords are distinct, and $x \notin H$. For a randomly chosen L-tuple $x \in C^L$, the probability that $x^{(i)} = x^{(i')}$ for some $i \neq i'$ is less than $L^2/|C|$. By the preceding lemma and finiteness of the set Ω'_L, the probability that $x \in H$ is also $O(1/|C|)$. So a random x is good with probability $1 - O(1/|C|)$. Lemma 9 then implies that

$$\Pr[x \text{ is good}] \mathop{\mathbb{E}}_{\text{good } x} \mathrm{mrad}_{\mathcal{U}[L]}(x) \leq \tau_{\mathcal{U}[L],1/2} = \tau_L \,.$$

On the other hand, for a good L-tuple we have $\mathrm{rad}(x) = \mathrm{mrad}_{\mathcal{U}[L]}(x)$ and thus the expectation is lower bounded by $\tau_L + \varepsilon$. In all, we conclude that $\frac{\varepsilon}{\tau_L + \varepsilon} = O(1/|C|)$, completing the proof. □

6 Proof of Theorem 1

Proof of Theorem 1. Recall that $M = \binom{2m}{m}$. Consider an $2m$-by-M matrix with $\{0,1\}$ entries whose columns are all possible vectors consisting of exactly m ones. The $2m$ rows of the matrix are the codewords of a code $C \subset \{0,1\}^M$. We claim that $\mathrm{mrad}_{\mathcal{U}[L]}(x) \geq \tau_L + c_L/2m + O(m^{-2})$ for every L-tuple x of distinct codewords from C.

By symmetry, $\mathrm{mrad}_{\mathcal{U}[L]}(x)$ is independent of the actual choice of x. So, fix any x, and pick j at random from $[M]$. Let 0_j be the number of these codewords that have 0 in the j'th column. Similarly, let 1_j be the number of these codewords that have 1 in the j'th column. Let $X_j = \min(0_j, 1_j)/L$. Note that $\mathrm{mrad}_{\mathcal{U}[L]}(x) = \mathbb{E}\, X_j$ by (4).

Suppose $L = 2k+1$ is odd. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_j X_j &= \frac{1}{2k+1}\sum_{l\le k} l\cdot \Pr[\min(0_i,1_i)=l] \\
&= \binom{2m}{m}^{-1}\frac{1}{2k+1}\sum_{l\le k} 2l\binom{2k+1}{l}\binom{2m-2k-1}{m-l} \\
&= \binom{2m}{m}^{-1}\sum_{1\le l\le k} 2\binom{2k}{l-1}\binom{2m-2k-1}{m-l} \\
&= \sum_{1\le l\le k} 2\binom{2k}{l-1}\frac{m(m-1)\cdots(m-l+1)\cdot m(m-1)\cdots(m-2k+l)}{2m(2m-1)\cdots(2m-2k)}
\end{aligned}$$

which, as $m\to\infty$, is

$$\begin{aligned}
&= \sum_{1\le l\le k}\binom{2k}{l-1}2^{-2k}\left[1+\frac{1}{2m}\binom{2k+1}{2}-\frac{1}{m}\binom{l}{2}-\frac{1}{m}\binom{2k-l+1}{2}+O(m^{-2})\right] \\
&= \tau_{2k+1}+2^{-2k-1}k\binom{2k}{k}/2m+O(m^{-2})
\end{aligned}$$

In the last equality here we used the expression for τ_{2k+1}, the formula for the variance of the binomial random variable $B(2k,1/2)$, and the known expression for the expected distance of a balanced random walk of $2k$ steps from the origin.

Similar computations hold if $L=2k$. Denote by $\sum'$ the sum in which the last summand is halved. The expected value of X_j is

$$\begin{aligned}
&\binom{2m}{m}^{-1}\frac{1}{2k}\sum_{l\le k}' 2l\binom{2k}{l}\binom{2m-2k}{m-l} \\
&\qquad= \binom{2m}{m}^{-1}\sum_{l\le k}' 2\binom{2k-1}{l-1}\binom{2m-2k}{m-l} \\
&\qquad= \sum_{l\le k} 2\binom{2k-1}{l-1}\frac{m(m-1)\cdots(m-l+1)\cdot m(m-1)\cdots(m-2k+l+1)}{(2m)(2m-1)\cdots(2m-2k+1)} \\
&\qquad= \sum_{l\le k}\binom{2k-1}{l-1}2^{-2k+1}\left(1+\frac{1}{2m}\binom{2k}{2}-\frac{1}{m}\binom{l}{2}-\binom{2k-l}{2}+O(m^{-2})\right) \\
&\qquad= \tau_{2k}+2^{-2k}k\binom{2k-1}{k}/2m+O(m^{-2})
\end{aligned}$$

□

7 Proof of Theorem 3

We start with the proof of the upper bound, following the approach of Konyagin in [7]. Let C be <3-list-decodable code of vectors in $\{0,1\}^n$ of radius at most $\tau_3+\varepsilon = 1/4+\varepsilon$. By Proposition 7

this implies that among any 3 codewords in C there are two of distance at least $(1/2+2\varepsilon)n$. For each codeword $x=(x_1,x_2,\ldots,x_n)$ define a vector $v=(v_1,v_2,\ldots,v_n)$ in the Euclidean space R^n by $v_i=\frac{(-1)^{x_i}}{\sqrt{n}}$. Note that each such vector is of unit norm, and among any three vectors there are two whose inner product is at most -4ε. Let V be the set of all the vectors obtained from the words in C and put $|V|=m$. Our objective is to show that $m\leq O(1/\varepsilon^{3/2})$. Let $H=(V,E)$ be the graph whose set of vertices is V in which two vertices u,v are connected iff their inner product is larger than -4ε. Fix a vertex $v\in V$ and let $W=N(v)$ be the set of all its neighbors in H. Note that the inner product between any two vertices in W is at most -4ε. Therefore, if $d=|W|$ is the degree of v in H and $\|v\|$ denotes the Euclidean 2-norm of a vector v, then

$$0\leq \left\|\sum_{u\in W} u\right\|^2 \leq d-d(d-1)4\varepsilon \tag{14}$$

implying that $d\leq \frac{1}{4\varepsilon}+1$ and also implying that

$$\left\|\sum_{u\in W} u\right\|^2 \leq d-d(d-1)4\varepsilon = \frac{1}{4\varepsilon}(4\varepsilon d)(1+4\varepsilon-4\varepsilon d)\leq \frac{(1+4\varepsilon)^2}{16\varepsilon}.$$

Therefore, by Cauchy–Schwarz, for every $v\in V$

$$\sum_{u\in N(v)}\langle v,u\rangle \leq \left\|\sum_{u\in N(v)} u\right\| \leq \frac{1+4\varepsilon}{4\sqrt{\varepsilon}}. \tag{15}$$

This gives the following (which can be slightly improved, but as this only changes the error term we prefer to present the simple version below):

$$\begin{aligned}0\leq \left\|\sum_{v\in V} v\right\|^2 &= m+\sum_{v\in V}\sum_{u\in N(v)}\langle v,u\rangle + \sum_{u\neq v\in V, uv\notin E}\langle v,u\rangle \\ &\leq m+m\frac{1+4\varepsilon}{4\sqrt{\varepsilon}}-m\left(m-\frac{1}{4\varepsilon}-2\right)4\varepsilon.\end{aligned}$$

By the last inequality

$$\left(m-\frac{1}{4\varepsilon}-2\right)4\varepsilon \leq 1+\frac{1+4\varepsilon}{4\sqrt{\varepsilon}},$$

implying that

$$m\leq \frac{1}{16\varepsilon^{3/2}}+O\left(\frac{1}{\varepsilon}\right). \tag{16}$$

This completes the proof of the upper bound.

We proceed with the proof of the lower bound by describing an appropriate construction. Let $G=(V,E)$ be a graph on m vertices, suppose it is a Cayley graph of an elementary abelian 2-group $\mathbb{Z}_2^r$, let A be its adjacency matrix, and let $d=\lambda_1\geq\lambda_2\geq\cdots\lambda_m=-\lambda$ be its eigenvalues, where d is the degree of regularity and $-\lambda$ is the smallest eigenvalue. Assume, further, that G is triangle-free. As G is a Cayley graph of an elementary abelian 2-group, it has an orthonormal basis of eigenvectors $v_1,v_2,...,v_m$ in which each coordinate of each vector is in $\{-1/\sqrt{m},1/\sqrt{m}\}$. Define $B=(A+\lambda I)/\lambda$ where I is the

m-by-m identity matrix. Then B is a positive semidefinite matrix, its diagonal is the all-1 vector, its eigenvalues are $\mu_i = (\lambda_i + \lambda)/\lambda$ and the corresponding eigenvectors are the vectors v_i. Let P be the m-by-m orthogonal matrix whose columns are the vectors v_i, and note that the first v_1 is the constant vector $1/\sqrt{m}$. Let D be the diagonal matrix whose diagonal entries are the eigenvalues μ_i and let $\sqrt{D}$ denote the diagonal matrix whose entries are $\sqrt{\mu_i}$. Then $P^t B P = D$ and thus $B = (P\sqrt{D})(\sqrt{D}P^t)$.

The rows of the matrix $P\sqrt{D}$ are vectors $x_1, x_2, \ldots, x_m$ where $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$. Note that for each j, $x_{ij} \in \{-\sqrt{\mu_j/m}, \sqrt{\mu_j/m}\}$ for all i, and that x_{i1} is positive for all i. In addition $x_i^t x_j = B_{ij}$ for all i, j meaning that the ℓ_2-norm of each vector x_i is 1 and that among any three vectors x_i there is an orthogonal pair. Let y_i be the vector obtained from x_i by removing its first coordinate (the one which is $\sqrt{\mu_1/m} = \sqrt{(d+\lambda)/m\lambda}$). Then each y_i is a vector of ℓ_2-norm $1 - \mu_1/m$ and among any three of them there is a pair with inner product $-\mu_1/m$. We can normalize the vectors by dividing each entry by $\sqrt{1-\mu_1/m}$ to get m unit vectors $z_1, z_2, \ldots, z_m$, where any three of them contain a pair with inner product $-\delta$, where $\delta = \mu_1/(m-\mu_1)$. Moreover, for the vectors $z_i = (z_{ij})$, for each fixed j the absolute value of all z_{ij} is the same for all i, denote this common value by t_j. We can now use the vectors z_i to define functions mapping $[0,1]$ to $\{1,-1\}$ as follows. Split $[0,1]$ into disjoint intervals I_j of length t_j^2 and define f_i to be $\mathrm{sign}(z_{ij})$ on the interval I_j. It is clear that the ℓ_2-norm of each f_i is 1 and the inner product between f_i and f_j is exactly that between z_i and z_j. In particular, each three functions f_i contain a pair whose inner product is at most $-\delta$.

One can replace the functions by vectors of $1, -1$ with essentially the same property, using an obvious rational approximation to the lengths of the intervals.

The graph in [1] is a triangle-free Cayley graph of an elementary abelian 2-group with $d = \big(1/4 + o(1)\big)m^{2/3}$ and $\lambda = \big(9 + o(1)\big)m^{1/3}$. This gives us $\delta = (1/36 + o(1))m^{-2/3}$ and hence the number of vectors is $m = \Theta\big((1/\delta)^{3/2}\big)$. This gives a binary code with $m = \Theta\big((1/\varepsilon)^{3/2}\big)$ codewords of length n so that among any three codewords there are two such that the Hamming distance between them is at least $(1/2 + \varepsilon)n$. According to Proposition 7, this means that the code is <3-list-decodable.

8 Spherical codes in the Hilbert space

Let us now consider a similar question for the case of the real Hilbert space $\mathcal{H}$ (the space of square-summable sequences of real numbers). Similar to the binary alphabet, we may motivate the question by the desire to construct a maximal number M of unit-energy signals, such that when one of them is sent and adversarial noise of bounded energy is applied, it is still possible to reconstruct the original signal, to within a list of size $< L$. We also note that results on adversarial-noise lead to bounds for the average-noise variation, as propounded in [11, Section XII]. We proceed to formal definitions.

We shall employ the same notation as in the rest of the paper, but with the meaning adapted to spherical codes. For example, we denote the norm in $\mathcal{H}$ by $\|\cdot\|$. We redefine $\mathrm{rad}(x)$ similarly: for an arbitary L-tuple $x = (x^{(1)}, \ldots, x^{(L)}) \in \mathcal{H}^L$ we define

$$\mathrm{rad}(x) = \min_{y \in \mathcal{H}} \max_j \|x^{(j)} - y\|\,, \quad \mathrm{diam}(x) = \max_{i,j} \|x^{(i)} - x^{(j)}\|\,.$$

Recall Jung's theorem [6, (2.6)]: For any L-tuple x we have

$$\mathrm{rad}(x) \le \sqrt{\frac{L-1}{2L}}\,\mathrm{diam}(x) \tag{17}$$

with equality if and only if $x^{(1)}, \ldots, x^{(L)}$ are the vertices of an $(L-1)$-simplex, i.e., when x consists of L vectors with pairwise distances all equal.

A *spherical code* C is a finite collection of unit-norm vectors in $\mathcal{H}$ and its L-radius $\rho_L(C)$ is the minimum value of $\mathrm{rad}(x)$ among all L-tuples x of distinct elements of C. We define

$$\mathrm{maxcode}_L(\varepsilon) \stackrel{\text{def}}{=} \sup\left\{|C| : \rho_L(C) \geq \tau_L + \varepsilon\right\},$$

where in this section $\tau_L \stackrel{\text{def}}{=} \sqrt{\frac{L-1}{L}}$. Our formulation corresponds to the problem of packing balls $B(x,r) \stackrel{\text{def}}{=} \{y \in \mathcal{H} : \|x-y\| \leq r\}$ centered on the unit sphere so that no point of $\mathcal{H}$ is covered by more than $(L-1)$ of them. Another equivalent way is to consider the problem of packing spherical caps $C(x,\alpha) = \{y : \|y\| = 1, \langle y, x\rangle \geq \cos\alpha\}$, where $\|x\| = 1$, with the requirement that no point *of the unit sphere* is covered by more than $(L-1)$ of them.

A classical result of Rankin [10] solves the case $L = 2$:

$$\mathrm{maxcode}_2(\varepsilon) = \left\lceil 1 + \frac{1}{2\sqrt{2\varepsilon + 2\varepsilon^2}} \right\rceil = \Theta\left(\frac{1}{\varepsilon}\right). \tag{18}$$

For $L > 2$, Blachman and Few [2] proved that if $\mathcal{H}$ is replaced by $\mathbb{R}^n$ then codes with $\rho_L(C) > \tau_L$ have size polynomial in n, while for $\rho_L(C) < \tau_L$ exponentially large codes exist. This was improved by Blinovsky [4], who demonstrated that codes with $\rho_L(C) > \tau_L$ have a finite upper bound on their size, which does not depend on n. His proof relied on the Ramsey theorem and can be condensed as follows:

Proposition 15 ([4]). *For any $\varepsilon > 0$ $\mathrm{maxcode}_L(\varepsilon)$ is finite.*

Proof. Consider a code C with $\rho_L(C) \geq \tau_L + \varepsilon$. Fix an integer $q \geq 1$, and break $[0,2]$ into q intervals of size $\frac{2}{q}$. Consider a code C and label each pair $(c,c') \in \binom{C}{2}$ according to the interval which contains $\|c - c'\|$. By Ramsey's theorem if C is sufficiently large then there should exist a large subcode C' whose all pairwise distances are in $[a, a+\frac{2}{q})$. From (18) we have $a \leq \sqrt{2} + O(1/|C'|)$ and from (17) we get that $\rho_L(C) \leq \rho_L(C') \leq \tau_L + O(1/|C'|) + O(1/q)$ and hence $|C'| \leq O(1/\varepsilon)$ when $q = O(1/\varepsilon)$. Consequently, C cannot be arbitrary large for a given $\varepsilon > 0$. □

Our main result on spherical codes is the following.

Theorem 16. *For any $L \geq 2$ there exist constants $c_1, c_2 > 0$ such that for all $\varepsilon > 0$*

$$c_1 \varepsilon^{-1} \leq \mathrm{maxcode}_L(\varepsilon) \leq c_2 \varepsilon^{-\frac{L^2-L+2}{2L}}. \tag{19}$$

Before proving the theorem we establish two auxiliary lemmas.

Definition. *Call a collection S of unit vectors an (m,ϵ)-*system *if among any m distinct elements $x_1, \ldots, x_m \in S$ there exists a pair with $\langle x_i, x_j\rangle < -\epsilon$.*

Lemma 17. *For each m there exists $C_m > 0$, such that the size of any (m,ϵ)-system S is at most $C_m \epsilon^{-\frac{m}{2}}$ and*

$$\|\sum_{x\in S} x\| \leq C_m \epsilon^{-\frac{m-1}{2}}.$$

Proof. For $m = 2$ this follows from (18) and (14). For $m = 3$ this was shown above in (15) and (16), essentially by reducing to $m = 2$. In general, for arbitrary m we can define a graph with vertices S as in the proof of (16) and notice that the neighborhood $\mathcal{N}(v)$ is an $(m-1,\epsilon)$-system and then apply induction. □

Lemma 18. *For any $L \geq 3$ there exists a non-negative function $\tau(\gamma) = \frac{2\gamma}{L^2-L-2} + O_L(\gamma^2)$, $\gamma \in [0,1]$, with the following property. Consider any L-tuple $x = (x_1, \ldots, x_L)$ of unit-norm vectors with $\operatorname{rad}(x) \geq \tau_L$. If $\langle x_1, x_2\rangle \geq \gamma \geq 0$ then there exist i, j such that $\langle x_i, x_j\rangle \leq -\tau(\gamma)$.*

Proof. Entirely like in (6) we can prove

$$\begin{aligned}\operatorname{rad}(x)^2 &= \max_{\omega\in\Omega_L} \min_{y\in\mathcal{H}} \mathop{\mathbb{E}}_{i\sim\omega} \|x_i - y\|^2 = \max_{\omega\in\Omega_L} \min_{y\in\mathcal{H}} \left(1 - 2\Big\langle \sum_i \omega_i x_i, y\Big\rangle + \|y\|^2\right) \\ &= \max_{\omega} \left(1 - \Big\|\sum_i \omega_i x_i\Big\|^2\right) = 1 - \min_{\omega\in\Omega_L} V(\omega)\,,\end{aligned} \tag{20}$$

where $V(\omega) = \sum_{i,j} v_{i,j}\omega_i\omega_j$ is the quadratic form corresponding to the Gram matrix of x with $v_{i,j} = \langle x_i, x_j\rangle$.

Fix some $0 \leq \tau \leq \frac{1}{L-1}$ and suppose now that x is such that $\langle x_i, x_j\rangle \geq -\tau$ for all i, j. We will show that for some function $\tau(\gamma)$ if $\tau < \tau(\gamma)$ then $\operatorname{rad}(x) < \tau_L$. To that end, we introduce another quadratic form $U(\omega) = \sum_{i,j} u_{i,j}\omega_i\omega_j$ with

$$u_{i,j} = \begin{cases} 1, & i = j, \\ \gamma, & \{i,j\} = \{1,2\}, \\ -\tau, & \text{otherwise.}\end{cases} \tag{21}$$

Note that according to assumptions $v_{i,j} \geq u_{i,j}$ and, therefore, on Ω_L we have $V(\omega) \geq U(\omega)$, and

$$\min_{\Omega_L} V(\omega) \geq \min_{\Omega_L} U(\omega)\,.$$

We next show that U is non-negative definite for all $0 \leq \tau \leq \frac{1}{L-1}$ and all $-\frac{1}{L-1} \leq \gamma \leq 1$. From convexity of the PSD cone, it is sufficient to check the four corners. For $\tau = 0$ the statement is clear. For $\tau = \frac{1}{L-1}$ we consider the two endpoints: $\gamma = -\frac{1}{L-1}$, $\gamma = 1$. For $\gamma = -\frac{1}{L-1}$ the resulting quadratic form equals $U_1(\omega) = \sum_i \omega_i^2 - \frac{1}{L-1}\sum_{i\neq j} \omega_i\omega_j$ and corresponds to the Gram matrix of unit-norm vectors forming an $(L-1)$-simplex centered at the origin. Consequently, U_1 is positive definite. Similarly, for $\gamma = 1$, the quadratic form corresponds to Gram matrix of the following collection: take unit-norm vectors forming an $(L-1)$-simplex, delete one vector and add a copy of another. The resulting quadratic form is non-negative definite.

Since U is convex, we could evaluate $\min_\omega U(\omega)$ by arguing that optimal assignment is symmetric (has equal coordinates $3, \ldots, n$ and $1, 2$). Instead we prefer to proceed indirectly and show another useful property of radii in Hilbert space.

Since $U \succeq 0$, it is a Gram matrix of some other L-tuple x' of unit-norm vectors and we know

$$\operatorname{rad}(x') \geq \operatorname{rad}(x)\,. \tag{22}$$

We temprorarily forget about the special form of U, as defined in (21), and view it as a generic Gram matrix of *some* L-tuple x' of unit-norm vectors with the property that $|\langle x'_i, x'_j\rangle| \le \theta$ for $i \neq j$. We will prove

$$\mathrm{rad}(x')^2 = \tau_L^2 - \frac{1}{L^2}\sum_{i\neq j}\langle x'_i, x'_j\rangle + O(\theta^2)\,, \tag{23}$$

where the $O(\cdot)$ term is uniform in x'. Note that the first two terms of the expression in (23) correspond to $\omega = \mathcal{U}[L]$ in (20). As $\theta \to 0$ the L-tuple x' becomes very close to L orthogonal vectors, and hence in (20) we expect that the optimal $\omega = \mathcal{U}[L] + O(\theta)$, cf. (24). Since we are operating near the minimum of the quadratic form, the $O(\theta)$ deviation of ω translates to $O(\theta^2)$ deviation for the value of U.

Proceeding to a formal proof of (23), first notice that if $\omega_1 = 0$ then as $\theta \to 0$ we must have $1 - U(\omega) \le \tau_{L-1} + o(1)$ (since we are considering only $L-1$ almost orthogonal vectors). But $1 - \min_\omega U(\omega)$ tends to $\tau_L > \tau_{L-1}$, implying that for all sufficiently small θ, the minimizer of $U(\omega)$ is in the interior of Ω_L. At the optimal point ω^* the gradient of U is proportional to a vector of all ones $\mathbf{1}$, from where we find

$$\omega^* = c(I_L + \Delta)^{-1}\,\mathbf{1}\,, \tag{24}$$

where $(I_L + \Delta)$ is the matrix of U, and the normalizing constant c is found from $\langle \omega^*, \mathbf{1}\rangle = 1$ yielding $c = \langle (I_L+\Delta)^{-1}\,\mathbf{1}, \mathbf{1}\rangle^{-1}$. Altogether, we get

$$U(\omega^*) = \langle (I_L+\Delta)\omega^*, \omega^*\rangle = c = \langle I_L\,\mathbf{1}, \mathbf{1}\rangle + \langle \Delta\,\mathbf{1}, \mathbf{1}\rangle + O(\theta^2)\,.$$

Finally, since $\mathrm{rad}(x')^2 = 1 - U(\omega^*)$ we get (23).

To complete the proof of the Lemma, note that from (22), (23) and (21) we have

$$\mathrm{rad}(x)^2 \le \tau_L^2 - \frac{1}{L^2}\big(2\gamma - (L^2 - L - 2)\tau\big) + O(\gamma^2) + O(\tau^2)\,.$$

Consequently, for appropriately defined $\tau(\gamma)$, if $\tau < \tau(\gamma)$ we should have $\mathrm{rad}(x) < \tau_L$. Furthermore, as $\gamma \to 0$ we have that $\tau(\gamma) = \frac{2\gamma}{L^2-L-2} + O_L(\gamma^2)$, as claimed. □

Proof of Theorem 16. Consider a regular $(M-1)$-simplex of unit vectors in $\mathcal{H}$. The pairwise distances are equal $\sqrt{\frac{2M}{M-1}}$ and thus from (17) we have that the radius of any L-tuple is at least $\tau_L\sqrt{\frac{M}{M-1}} = \tau_L + \Omega(1/M)$, proving the lower bound in (19).

We proceed to the upper bound. Fix a code C with $\rho_L(C) \ge \tau_L + \varepsilon$. The main idea is again essentially due to Konyagin: fixing one point $c \in C$ and considering its close neighbors, we notice that the radius constraint (cf. Lemma 18) introduces repulsion between these neighbors (that is they should be widely separated among themselves) and consequently, neighborhoods can not be too large.

We proceed with the argument. First, by (17) any L-tuple with $\mathrm{rad}(x) \ge \tau_L + \varepsilon$ also satisfies $\mathrm{diam}(x) \ge \sqrt{2} + \sqrt{\frac{2}{\tau_L}}\varepsilon$, and thus the code C is also an (L, ϵ')-system, with $\epsilon' = \frac{2}{\sqrt{\tau_L}}\varepsilon$.

Next, let $\epsilon_1 = \varepsilon^{\frac{L-1}{L}}$ and $\epsilon_2 = \tau(\epsilon_1)$, where $\tau(\cdot)$ is from Lemma 18. We consider two types of neighbors c of each point $c_i \in C$, depending on

$$-\epsilon' \le \langle c, c_i\rangle \le \epsilon_1, \text{ or } \langle c, c_i\rangle > \epsilon_1\,. \tag{25}$$

Let $\mathcal{N}'(c_i)$ and $\mathcal{N}''(c_i)$ be the two respective sets of neighbors. The rest of the points are "far away" from c_i and satisfy

$$\langle c, c_i \rangle < -\epsilon' \,. \tag{26}$$

First, notice that since C is an (L, ϵ')-system, we have that $\mathcal{N}'(c_i) \cup \mathcal{N}''(c_i)$ is an (m, ϵ')-system with $m \stackrel{\text{def}}{=} L - 1$. Thus from Lemma 17

$$|\mathcal{N}'(c_i)| \le |\mathcal{N}'(c_i) \cup \mathcal{N}''(c_i)| \le C_m \epsilon'^{-\frac{m}{2}} \,. \tag{27}$$

Second, take any $m = (L-1)$ distinct points in $\mathcal{N}''(c_i)$. Adding c_i to this m-tuple and applying Lemma 18 to the resulting L-tuple, we conclude that $\mathcal{N}''(c_i)$ is an (m, ϵ_2)-system. Therefore, from Lemma 17 we have

$$\Big\| \sum_{c \in \mathcal{N}''(c_i)} c \Big\| \le C_m \epsilon_2^{-\frac{m-1}{2}} \,. \tag{28}$$

Consider

$$\Big\langle c_i, \sum_{c \in C} c \Big\rangle = 1 + \Big\langle c_i, \sum_{c \in \mathcal{N}''(c_i)} c \Big\rangle + \Big\langle c_i, \sum_{c \in \mathcal{N}'(c_i)} c \Big\rangle + \Big\langle c_i, \sum_{c \notin \mathcal{N}' \cup \mathcal{N}'' \cup \{c_i\}} c \Big\rangle \tag{29}$$

$$\le 1 + C_m \epsilon_2^{-\frac{m-1}{2}} + C_m \epsilon_1 \epsilon'^{-\frac{m}{2}} - \epsilon'(|C| - 1 - C_m \epsilon'^{-\frac{m}{2}}) \tag{30}$$

where the second term is estimated by Cauchy–Schwarz and (28), the third term is by the definition of $\mathcal{N}'(c_i)$ and (27), and the fourth term is the combination of (26) and the bound in (27).

Summing (30) over all $c_i \in C$ and using $\sum_{c_i, c \in C} \langle c_i, c \rangle \ge 0$ we get

$$\epsilon' |C| \le 1 + \epsilon' + C_m \epsilon_2^{-\frac{m-1}{2}} + C_m \epsilon_1 \epsilon'^{-\frac{m}{2}} + C_m \epsilon'^{1-\frac{m}{2}} \,,$$

from where, recalling that $\epsilon_1 \asymp \epsilon_2 \asymp \varepsilon^{\frac{L-1}{L}}$ and $\epsilon' \asymp \varepsilon$ we get that the first two terms and the last are negligible compared to the third and fourth, which are comparable and $\asymp \varepsilon^{-\frac{(L-1)(L-2)}{2L}}$. Canceling ϵ' we get an upper bound in (19). □

9 Remarks and open problems

- The $2L$ in Proposition 4 can be improved to $O(\sqrt{L})$ using a combination of the Beck–Fiala floating colors argument with Spencer's six deviations theorem. However, even with this improvement, we do not see a way to prove Theorem 2 with a good value of the implicit constant.

- For odd $L \ge 5$, the best upper bound we have is a tower of exponentials of height L. To that end, one colors L-tuples of codewords according to the measure ω for which $\mathrm{rad}(x) = \mathrm{mrad}_\omega(x)$, uses Ramsey's theorem to get a monochromatic set, and then proceeds similarly to the proof of Theorem 2.

- In the $<L$-list-decodable code in Theorem 1, the code length is exponential in ε^{-1}. One can restrict that code to a random subset of $O(\varepsilon^{-1} \log \varepsilon^{-1})$ coordinates, and obtain a code of asymptotically the same size $c_L \varepsilon^{-1} + O(1)$.

  For $L = 2$ and $L = 4$, the Levenshtein's code has length $O(\varepsilon^{-1})$ and size $c_L \varepsilon^{-1} + O(1)$. However, we do know if one can make the code in Theorem 1 of linear size for general L.

Acknowledgment. We thank Alan Frieze for providing the reference [13].

References

[1] Noga Alon. Explicit Ramsey graphs and orthonormal labelings. *Electron. J. Combin.*, 1:Research Paper 12, approx. 8 pp. 1994.

[2] NM Blachman and L Few. Multiple packing of spherical caps. *Mathematika*, 10(1):84–88, 1963.

[3] V. M. Blinovskiĭ. Bounds for codes in decoding by a list of finite length. *Problemy Peredachi Informatsii*, 22(1):11–25, 1986.

[4] V. Blinovsky. Multiple packing of the Euclidean sphere. *IEEE Trans. Inform. Th.*, 45(4):1334–1337, 1999.

[5] V. Blinovsky. Generalization of Plotkin bound to the case of multiple packing. In *2009 IEEE International Symposium on Information Theory*, pages 2048–2050, June 2009.

[6] L. Danzer, B. Grünbaum, and V Klee. Helly's theorem and its relatives. *Convexity: Proc. of Symposia in Pure Math*, 7:101–180, 1963.

[7] S. V. Konyagin. Systems of vectors in Euclidean space and an extremal problem for polynomials. *Mat. Zametki*, 29(1):63–74, 155, 1981. doi:10.1007/BF01142512 (translation) http://mi.mathnet.ru/mz10048 (Russian original).

[8] V. I. Levenshtein. The application of Hadamard matrices to a problem in coding. *Problemy Kibernetiki*, 5:123–136, 1961. English translation in Problems of Cybernetics 5, 1964 pp. 166–184.

[9] Yury Polyanskiy. Upper bound on list-decoding radius of binary codes. *IEEE Trans. Inf. Theory*, 62:1119–1128, 2016. arXiv:1409.7765.

[10] Robert Alexander Rankin. The closest packing of spherical caps in n dimensions. *Glasgow Mathematical Journal*, 2(3):139–144, 1955.

[11] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell Syst. Tech. J.*, 38:611–656, 1959.

[12] John von Neumann. Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes. *Ergebnisse eines Mathematischen Kolloquiums*, 8:73–83, 1937. Reprinted as "A Model of General Economic Equilibrium, " Review of Economic Studies, vol. 13, 1945.

[13] Eric W. Weisstein. Random walk–1-dimensional. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/RandomWalk1-Dimensional.html (Archived at https://web.archive.org/web/20161121030145/http://mathworld.wolfram.com/RandomWalk1-Dimensional.html).