Explicit Construction of a Small Epsilon-Net for Linear Threshold Functions

Yuval Rabani
Computer Science Department
Technion
Haifa 32000
Israel
rabani@cs.technion.ac.il

Amir Shpilka
Computer Science Department
Technion
Haifa 32000
Israel
shpilka@cs.technion.ac.il

ABSTRACT

We give explicit constructions of epsilon nets for linear threshold functions on the binary cube and on the unit sphere. The size of the constructed nets is polynomial in the dimension n and in $\frac{1}{\epsilon}$. To the best of our knowledge no such constructions were previously known. Our results match, up to the exponent of the polynomial, the bounds that are achieved by probabilistic arguments.

As a corollary we also construct subsets of the binary cube that have size polynomial in n and covering radius of $\frac{n}{2} - c\sqrt{n \log n}$, for any constant c. This improves upon the well known construction of dual BCH codes that only guarantee covering radius of $\frac{n}{2} - c\sqrt{n}$.

Categories and Subject Descriptors

F.2.2 [**Nonnumerical Algorithms and Problems**]: Geometrical problems and computations; F.2.m [**Analysis of Algorithms and Problem Complexity**]: Miscellaneous

General Terms

Theory

Keywords

Epsilon-Net, Explicit Construction, Linear Threshold Function

1. INTRODUCTION

Influenced by the discovery of unexpected connections linking fundamental questions in geometric functional analysis to problems in theoretical computer science, there has been a recent interest in explicit or algorithmic construction of certain geometric objects that are known to exist via probabilistic arguments. For example, the celebrated dimension reduction lemma of Johnson and Lindenstrauss [18] has been derandomized using the method of conditional expectations [12, 24]. Another example that is still mostly open is the construction of high dimensional nearly-Euclidean linear subspaces of ℓ_1^n [17, 15, 16]. This problem is related to the question of constructing compressed sensing schemes [11]; other probabilistic compressed sensing schemes, using the restricted isometry property [8], also exhibit a geometric flavor. All these geometric objects have numerous applications in areas such as coding theory and data compression, communication complexity, nearest neighbor search, learning theory, and computational linear algebra (see, e.g., the introduction of [15]), hence the desire to discover explicit constructions.

In this paper we study what is perhaps the simplest such question. We construct ϵ-nets for linear threshold functions on the binary cube $\mathcal{B}_n = \{-1, +1\}^n$ as well as on the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. A function $f : \mathbb{R}^n \to \{-1, 1\}$ is called a linear threshold function (LTF) iff for some $v \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ we have that $f(x) = 1$ iff $\langle v, x \rangle \geq \theta$. Notice that when restricted to $\mathbb{S}^{n-1}$, a linear threshold function is simply the indicator function of a closed spherical cap of $\mathbb{S}^{n-1}$. Given a measurable set $\Omega \subset \mathbb{R}^n$ endowed with a measure μ and a family $\mathcal{F}$ of measurable subsets of Ω, an ϵ-net for $\mathcal{F}$ is a set $S \subset \Omega$ such that for every $F \in \mathcal{F}$ with $\mu(F) > \epsilon$, we have that $|S \cap F| > 0$. Constructing ϵ-nets for natural set systems $(\Omega, \mu, \mathcal{F})$ has been studied extensively in some cases. For example, the case where Ω is the convex hull of a d-points set P and $\mathcal{F}$ is the family of all convex hulls of subsets of P received a lot of attention (see, e.g., [10, 5]). The case where $\Omega = [m]^d$ and the set $\mathcal{F}$ is the set of all combinatorial rectangles also received a lot of attention [13, 19]. To the best of our knowledge, the case of linear threshold functions has not been previously considered in this context.

We consider Ω which is either the binary cube or the unit sphere (endowed with the uniform measure), and the family $\mathcal{F}$ includes the subsets $A_f = \{x \in \Omega : f(x) = 1\}$, for all linear threshold functions f. We construct $S \subset \Omega$ of cardinality $\mathrm{poly}(n, 1/\epsilon)$ that includes a point from A_f for every linear threshold function f that satisfies $\mu(A_f) \geq \epsilon$, where μ is the uniform measure on Ω. A random sample of $O(n/\epsilon)$ points is an ϵ-net with high probability, and our goal is to construct such a set explicitly. We prove the following theorem.

THEOREM 1.1. *There exist two universal constants $a, b > 0$ such that for every $\epsilon > 0$ there is an explicit construction of an ϵ-net, $N_\epsilon \subset \mathcal{B}_n$, for linear threshold functions of size $|N_\epsilon| = O(\epsilon^{-b} \cdot n^a)$.*

Note that when $\epsilon = 1/\mathrm{poly}(n)$ the construction above yields a polynomial sized set. As a corollary of our construction, we

STOC'09, May 31–June 2, 2009, Bethesda, Maryland, USA.

get a similar construction for the unit sphere.

THEOREM 1.2. *There exist two universal constants $a, b > 0$ such that for every $\epsilon = \exp(-O(\sqrt{n}))$ there is an explicit construction of an ϵ-net, $S_\epsilon \subset \mathbb{S}^{n-1}$, for spherical caps of size $|S_\epsilon| = O(\epsilon^{-b} \cdot n^a)$.*

As another corollary of our construction we also construct a $\mathrm{poly}(n)$ size subset of $\mathcal{B}_n$ with covering radius of $\frac{n}{2} - \Omega(\sqrt{n \log n})$. The covering radius r of a set of points $S \subset \mathcal{B}_n$ is the smallest ρ such that for every $x \in \mathcal{B}_n$ there is some $s \in S$ with $\mathcal{H}(x, s) \leq \rho$, where $\mathcal{H}$ denotes Hamming distance. We note that this construction improves upon the one guaranteed by dual BCH codes. This result was independently obtained by Alon [2].

COROLLARY 1.3. *There exists $a > 0$ such that for every $c > 0$ there is an explicit construction of a set $C \subset \mathcal{B}_n$ of size $|C| = n^2 \cdot (n^c)^a$ such that for every $z \in \mathcal{B}_n$ there is some $x \in C$ with $\mathcal{H}(z, x) \leq \frac{n}{2} - \sqrt{cn \log n}$.*

We note that linear threshold functions play an important role in both theory and practice. For example, bounded depth TC^0 circuits, composed of a constant number of layers of threshold functions, received considerable attention in complexity theory, and support vector machines use threshold functions as hypothesis in many learning scenarios. Aside from the intrinsic interest in studying linear threshold functions, our work is motivated by the desire to build methodically a theory of pseudorandom generators for geometric functions. In the algebraic setting (over GF[2]), ϵ-biased sample spaces fool linear functions [22]; they were recently composed to construct pseudorandom generators for low-degree polynomials [25]. Analogously, we hope that dealing with linear threshold functions is a good starting point for the gradual construction of more complicated pseudorandom generators for non-linear geometric functions, which are needed to resolve some of the questions mentioned earlier.

Our constructions use several ideas from derandomization theory. The first one is the notion of a k-wise independent distribution. A set of m random variables on a sample space Ω is k-wise independent iff every subset of the random variables of cardinality at most k is independent. There are numerous applications in computer science for k-wise independent distributions with small support. In particular, $\mathrm{poly}(n)$ size k-wise independent distributions on $\mathcal{B}_n$ give a construction of a cover code with covering radius $\frac{n}{2} - \Omega(\sqrt{n})$. We improve the covering radius of a $\mathrm{poly}(n)$ size set to $\frac{n}{2} - \Omega(\sqrt{n \log n})$. The idea is to concatenate $O(\log n)$ samples from a 4-wise independent distribution with $m = n/O(\log n)$ random variables. In order to restrict the size of the constructed set, we need to consider only a subset of all possible concatenations. We use a constant degree expander graph on the sample space Ω (of the 4-wise independent distribution of length m) and consider all the random walks of length $O(\log n)$ on it. Then, for each random walk we concatenate the relevant vectors of length m to get a vector of length n. We then show that this set has the desired covering radius. In order to show the more general goal of an ϵ-net for LTFs, we note that the construction above works if all the coefficients defining a LTF f are roughly of the same magnitude. As this is not always the case, the idea is to partition the set of coordinates into $O(\log n)$ "buckets" such that each of them contains approximately the same weight of coefficients as the other sets. To get a small set of partitions, we use certain explicit constructions of perfect hash functions. We then apply the above construction to each candidate partition to get the desired ϵ-net.

Organization.

In Section 2 we give some formal definitions and the necessary background on k-wise independent distributions, expander graphs and perfect hash functions. We also give some concentration results for threshold functions. In section 3 we give the construction of a cover code. In Section 4 we give our main construction for linear threshold functions and in Section 5 we give the construction for spherical caps.

2. PRELIMINARIES

We will use the following notation. The n-dimensional binary cube is $\mathcal{B}_n = \{-1, 1\}^n$. The $(n-1)$-dimensional unit sphere in $\mathbb{R}^n$ is $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. The Hamming distance on $\mathbb{R}^n$ is denoted by $\mathcal{H}$, so $\mathcal{H}(x, y)$ is the number of coordinates i for which $x_i \neq y_i$. For $x \in \mathbb{R}^n$ and $J = \{i_1, \ldots, i_{|J|}\} \subseteq [n]$ we denote $x_J = (x_{i_1}, \ldots, x_{i_{|J|}})$. We will abuse notation and use (for $A \subset \mathbb{R}^n$) $\mathcal{H}(x, A)$ to denote $\min_{y \in A} \mathcal{H}(x, y)$. For $A \subseteq \mathcal{B}_n$ and $\rho > 0$, we put $A_\rho = \{x \in \mathcal{B}_n : \mathcal{H}(x, A) \leq \rho\}$. The covering radius of a set $C \subset \mathcal{B}_n$ is the minimum ρ such that $C_\rho = \mathcal{B}_n$. Namely, it is the minimal ρ such that for every $x \in \mathcal{B}_n$ there is $y \in A$ with $\mathcal{H}(x, y) \leq \rho$.

In this paper we focus on linear threshold functions. A vector $v \in \mathbb{R}^n$ and a real number $\theta \in \mathbb{R}$ define a linear threshold function $L_{v,\theta} : \mathcal{B}_n \to \{-1, 1\}$ by $L_{v,\theta}(x) = \mathrm{sign}(\langle v, x \rangle - \theta)$. In other words, $L_{v,\theta}(x) = 1$ if $\langle v, x \rangle \geq \theta$ and $L_{v,\theta}(x) = -1$ otherwise. For a linear function $L_{v,\theta}$ we define by $A_{v,\theta}$ it set of accepting inputs. Namely, $A_{v,\theta} = L_{v,\theta}^{-1}(1) = \{x \in \mathcal{B}_n : \langle v, x \rangle \geq \theta\}$. A spherical cap in $\mathbb{R}^n$ is a subset of $\mathbb{S}^{n-1}$ that is contained in a half-space. Namely, for every $v \in \mathbb{R}^n$ and $\theta > 0$ the cap $C_{v,\theta}$ is defined as $C_{v,\theta} = \{x \in \mathbb{S}^{n-1} : \langle v, x \rangle \geq \theta\}$. Stated differently, $C_{v,\theta} = L_{v,\theta}^{-1}(1) \cap \mathbb{S}^{n-1}$ (we now think of $L_{v,\theta}$ as a function from $\mathbb{R}^n$ to $\{-1, 1\}$).

2.1 k-wise independent distributions

A multiset $I \subset \{-1, 1\}^n$ such that for every $j \in \{1, 2, \ldots, k\}$, for every $\{i_1, i_2, \ldots, i_j\} \subset \{1, 2, \ldots, n\}$, and for every $z_1, z_2, \ldots, z_j \in \{-1, 1\}$, satisfy that

$$\left|\left\{x \in I : (x_{i_1}, x_{i_2}, \ldots, x_{i_j}) = (z_1, z_2, \ldots, z_j)\right\}\right| = \frac{|I|}{2^j},$$

is called a k-wise independent sample space. Many explicit constructions of small k-wise independent sample spaces are known. For example, extended binary BCH codes of length $n = 2^m - 1$ and designed distance $2t + 2$ can be used to construct a $(2t+1)$-wise independent sample space of size $2^{mt+1} = 2(n+1)^t$ (see [6, Chapter 16]).

FACT 2.1. *For every integer $k > 0$ there exists an explicit construction of a sample space of size $O(n^{k/2})$ that is k-wise independent.*

Let a multiset $S \subseteq \{-1, 1\}^n$ be a k-wise independent sample spaces. The following is an easy observation.

OBSERVATION 2.2. *For $i \in [n]$ and $\alpha \in \{-1, 1\}$, the multiset $S_{i,\alpha} := \{x \in S : x_i = \alpha\}$ is a $k-1$-wise independent sample space.*

The following result was proved by Berger in [7].

LEMMA 2.3 (LEMMA 3.1 IN [7]). *Let $S \subset \{-1,1\}^n$ be a 4-wise independent sample space. Then for every $x \in \mathbb{S}^{n-1}$ we have that $\mathbb{E}[\langle s,x\rangle] = 0$, $\mathbb{E}[\langle s,x\rangle^2] = 1$ and $\mathbb{E}[\langle s,x\rangle^4] \leq 3$, where all expectations are with respect to a uniform choice of $s \in S$. Moreover, for every $x \in \mathbb{R}^n$ we have that*

$$\Pr_{s\in S}\left[|\langle s,x\rangle| > \frac{\|x\|_2}{\sqrt{3}}\right] \geq \frac{2}{11}.$$

The following lemma is a special case of a lemma of Alon et al [4].

LEMMA 2.4 (LEMMA 3.2 IN [4]). *Let X be a real random variable and suppose that its first, second and forth moments satisfy $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$ and $\mathbb{E}[X^4] \leq 3$. Then $\Pr[X > 1/7] \geq 1/20$. Consequently, if $S \subset \{-1,1\}^n$ is a 4-wise independent sample spaces then for every $x \in \mathbb{S}^{n-1}$ we have that $\Pr_{s\in_U S}[\langle s,x\rangle > 1/7] \geq 1/20$.*

Next is an easy corollary of Observation 2.2 and Lemma 2.4.

LEMMA 2.5. *Let $k > 4$ be an integer, $S \subseteq \{-1,1\}^n$ a k-wise independent sample space and $v = (v_1,\ldots,v_n) \in \mathbb{R}^n$ a unit vector. Let $M \subset [n]$ be such that $|M| = k-4$ and the entries of v corresponding to the coordinates in M are the $k-4$ largest entries of v (namely, for every $j \notin M$ and every $i \in M$ we have that $|v_j| \leq |v_i|$). Then*

$$\Pr_{x\in S}\left[\langle x,v\rangle \geq \|v_M\|_1 + \frac{1}{7}\|v_{[n]\setminus M}\|_2\right] \geq \frac{4}{5}\cdot 2^{-k}.$$

PROOF. Let $S' \subset S$ be the set of all $s \in S$ such that $\operatorname{sign} s_i = \operatorname{sign} v_i$ for every $i \in M$. By definition we have that $|S'| = 2^{-|M|}\cdot|S| = |S|/2^{k-4}$. Moreover, by Observation 2.2 we get that S' is 4-wise independent. Let $v' = (v'_1,\ldots,v'_n)$ be defined as $v'_i = 0$ for $i \in M$ and $v'_i = v_i$ for $i \notin M$. By Lemma 2.4 we have that

$$\Pr_{s\in S'}\left[\langle s,v'\rangle > \frac{1}{7}\|v'\|_2\right] > \frac{1}{20}.$$

By definition of v' we get that $\langle s,v\rangle = \sum_{i\in M} s_i\cdot v_i + \langle s,v'\rangle = \|v_M\|_1 + \langle s,v'\rangle$. Thus,

$$\Pr_{x\in S}\left[\langle x,v\rangle \geq \|v_M\|_1 + \frac{1}{7}\|v_{[n]\setminus M}\|_2\right] \geq \frac{1}{20\cdot 2^{k-4}} = \frac{4}{5}\cdot 2^{-k}.$$

□

2.2 Expander graphs

An undirected graph $G = (V,E)$ is called an (n,d,λ)-expander if $|V| = n$, the degree of each node is d, and the second largest eigenvalue, in absolute value, of the adjacency matrix of G is λ. For every $d = p+1$, where p is a prime congruent to 1 modulo 4, there are explicit constructions for infinitely many n of (n,d,λ)-expanders, where $\lambda \leq 2\sqrt{d-1}$ [20, 1].

A random walk of length ℓ on G is the following random process. First pick a vertex of G uniformly at random. Denote this vertex with v_1. At the i'th step (for $1 < i \leq \ell$ we pick a neighbor of v_{i-1} uniformly at random and label it with v_i. The walk is the ordered list $(v_1, v_2, \ldots, v_\ell)$. We shall need the following theorem of Alon et al. [3]

THEOREM 2.6. *Let G be an $[n,d,\lambda]$-expander. Let $W_1,\ldots,W_\ell \subset V(G)$ be some subsets of G, each of size at least $\mu n \geq 6\lambda n/d$. The probability that a random walk of length ℓ stays inside $W_1, W_2, \ldots, W_\ell$ is at least $\mu(\mu - 2\lambda/d)^{\ell-1}$.*

2.3 Perfect hash functions

A set H of functions $h : \{1,2,\ldots,n\} \to \{1,2,\ldots,m\}$ such that for every $S \subset \{1,2,\ldots,n\}$ with $|S| = s$ there exists $h \in H$ such that $|h(S)| = s$ is called an (n,m,s)-perfect hash family. For all $n,s \in \mathbb{N}$, $s \leq n$, there are explicit constructions of $(n, O(s), s)$-perfect hash families H with $|H| = 2^{O(s+\log\log n)}$ [23]. We shall need the following strengthening which is immediate corollary of the proofs of [14, 23].

LEMMA 2.7 (PERFECT HASH FUNCTIONS). *For every integer s, there is an explicit family $\mathcal{H}$ of hash functions $h : [n] \to [8s]$ of cardinality $|\mathcal{H}| = 2^{(4+o(1))\cdot s + \log 2s \log\log n}$ [1] such that the following holds for every unit vector $v \in \mathbb{S}^{n-1}$. Let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, and let I_t denote the set $\{i_1, i_2, \ldots, i_t\}$. There exists some $h \in \mathcal{H}$ such that*

1. *The map h is an injection on I_s.*
2. *Let $t \in [s-1]$. If $v_{i_{t+1}}^2 \leq \frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2$, then*

$$\sum_{r\in[8s]} \min\left\{\|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \geq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2. \quad (1)$$

For completeness we give the proof in the appendix. The following is an easy corollary.

COROLLARY 2.8. *Let $24 \leq s \leq n$ be integers and $\mathcal{H}$ the hash family guaranteed by Lemma 2.7. There exists constants c_1 and c_2 such that one of the following conditions holds (using the same notation as in Lemma 2.7):*

1. *either $\sum_{t=\lceil 2s/3\rceil}^{s-1} |v_{i_{t+1}}| \geq \frac{\sqrt{s}}{32}\|v_{[n]\setminus I_s}\|_2$;*
2. *or, there exists some $h \in \mathcal{H}$ such that h is an injection on I_s and for at least $c_1\cdot 8s$ buckets it holds that $\|v_{h^{-1}(r)\setminus I_t}\|_2^2 \geq \frac{c_2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2$.*

PROOF. Let $h \in \mathcal{H}$ be the map guaranteed by Lemma 2.7 (we shall also use the notations of the lemma). We are guaranteed that h is an injection on I_s. We now consider two cases: **Case 1**: there is some $t \in [s-1]$ such that $v_{i_{t+1}}^2 \leq \frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2$. **Case 2**: for every $t \in [s-1]$ we have that $v_{i_{t+1}}^2 > \frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2$.

Consider case 1. We will show that for some constants c_1, c_2 at least $c_1\cdot 8s$ buckets satisfy that $\|v_{h^{-1}(r)\setminus I_t}\|_2^2 \geq \frac{c_2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2$. Assume for a contradiction that less than $c_1\cdot 8s$ buckets have high norm. By the lemma we know that as there is some $t \in [s-1]$ such that $v_{i_{t+1}}^2 \leq \frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2$ then

$$\sum_{r\in[8s]} \min\left\{\|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \geq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

[1] The $\log 2s$ factor can be eliminated at the expense of a slight complication of the construction (adding a preliminary phase that maps $[n]$ to $[s^2]$ and replacing the maps from $[n]$ in the two-phase construction by maps from $[s^2]$). In our application, this does not improve the exponent beyond a $o(1)$ factor, as we use $s = \Theta(\log n)$.

Hence,

$$\frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2 \leq \sum_{r\in[8s]} \min\left\{\|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \leq$$

$$c_1\cdot 8s\cdot\frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2 + 8s\cdot\frac{c_2}{s}\cdot\|v_{[n]\setminus I_t}\|_2 =$$

$$(16c_1+8c_2)\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

Therefore, for $c_1 = \frac{1}{48}$ and $c_2 = \frac{1}{49}$ we get a contradiction, unless $\|v_{[n]\setminus I_t}\|_2^2 = 0$. However, the claim is trivial if this is the case.

Let us now assume that we are in case 2. It follows that

$$\sum_{t=\lceil 2s/3\rceil}^{s-1} |v_{i_{t+1}}| \geq \sum_{t=\lceil 2s/3\rceil}^{s-1} \frac{1}{8\sqrt{s}}\cdot\|v_{[n]\setminus I_t}\|_2 \geq$$

$$\sum_{t=\lceil 2s/3\rceil}^{s-1} \frac{1}{8\sqrt{s}}\cdot\|v_{[n]\setminus I_s}\|_2 \geq \frac{\sqrt{s}}{32}\|v_{[n]\setminus I_s}\|_2,$$

where in the last inequality we used the assumption that $s \geq 24$. □

2.4 Concentration of Threshold functions

In order to construct an ϵ-net for linear threshold functions we need to understand, for every linear threshold function $L_{v,\theta}$, for which values of θ it holds that $\Pr_{x\in\mathcal{B}_n}[L_{v,\theta}(x)=1] > \epsilon$.

THEOREM 2.9. *(Chernoff-Hoeffding) For $v = (v_1,\ldots,v_n)\in\mathbb{R}^n$ and $\theta\in(0,\infty)$ we have that*

$$\Pr_{x\in\mathcal{B}_n}[\langle x,v\rangle > \theta] \leq \exp\left(-\frac{1}{2}\left(\frac{\theta}{\|v\|_2}\right)^2\right).$$

PROOF. Let $x=(x_1,\ldots,x_n)$. We get that for every $t>0$

$$\Pr_{x\in\mathcal{B}_n}[\langle x,v\rangle > \theta] = \Pr_{x\in\mathcal{B}_n}[\exp(t\cdot\langle x,v\rangle) > \exp(t\cdot\theta)] \leq$$

$$\frac{\mathbb{E}[\exp(t\cdot\langle x,v\rangle)]}{\exp(t\cdot\theta)} = \frac{\mathbb{E}\left[\prod_{i=1}^n \exp(t\cdot x_i\cdot v_i)\right]}{\exp(t\cdot\theta)} =$$

$$\frac{\prod_{i=1}^n \mathbb{E}[\exp(t\cdot x_i\cdot v_i)]}{\exp(t\cdot\theta)} = \frac{\prod_{i=1}^n \frac{1}{2}(\exp(t\cdot v_i)+\exp(-t\cdot v_i))}{\exp(t\cdot\theta)}$$

$$\leq \frac{\prod_{i=1}^n \exp((t\cdot v_i)^2/2)}{\exp(t\cdot\theta)} = \exp(t^2\|v\|_2^2/2 - t\cdot\theta).$$

By picking $t=\frac{\theta}{\|v\|_2^2}$ we get that

$$\Pr_{x\in\mathcal{B}_n}[\langle x,v\rangle > \theta] \leq \exp\left(-\frac{1}{2}\left(\frac{\theta}{\|v\|_2}\right)^2\right).$$

□

The following result will be used to determine how large θ can be, for a given $v\in\mathbb{R}^n$ so that $L_{v,\theta}$ accepts an ϵ fraction of the inputs.

COROLLARY 2.10. *Let $v=(v_1,\ldots,v_n)\in\mathbb{R}^n$ and $\delta\in\mathbb{R}^+$. Assume that $|v_1|\geq|v_2|\geq\ldots\geq|v_n|$. Let $1\leq k\leq n$ be an integer. Assume further that $|v_k|>0$. Then*

$$\Pr_{x\in\mathcal{B}_n}\left[\langle x,v\rangle \geq \|v_{[\lceil 2k/3\rceil]}\|_1 + \delta\cdot\|v_{[n]\setminus[k]}\|_2\right] \leq$$

$$\exp(-k/18) + \exp(-\delta^2/2).$$

PROOF. We have that

$$\Pr_{x\in\mathcal{B}_n}\left[\langle x,v\rangle \geq \|v_{[\lceil 2k/3\rceil]}\|_1 + \delta\cdot\|v_{[n]\setminus[k]}\|_2\right] \leq$$

$$\Pr_{x_{[k]}\in\mathbb{Z}_2^k}\left[\langle x_{[k]},v_{[k]}\rangle \geq \|v_{[\lceil 2k/3\rceil]}\|_1\right] +$$

$$\Pr_{x_{[n]\setminus[k]}\in\mathbb{Z}_2^{[n]\setminus[k]}}\left[\langle x_{[n]\setminus[k]},v_{[n]\setminus[k]}\rangle \geq \delta\cdot\|v_{[n]\setminus[k]}\|_2\right].$$

As $|v_1|\geq|v_2|\geq\ldots\geq|v_k|>0$ we see that in order for the inequality

$$\langle x_{[k]},v_{[k]}\rangle \geq \|v_{[\lceil 2k/3\rceil]}\|_1$$

to hold we must have that $\mathrm{sign}(x_i)=\mathrm{sign}(v_i)$ for at least $2k/3$ of the indices. Using the Chernoff-Hoeffding bound we bound this probability with

$$\Pr_{x_{[k]}\in\mathbb{Z}_2^k}\left[\langle x_{[k]},v_{[k]}\rangle \geq \|v_{[\lceil 2k/3\rceil]}\|_1\right] \leq \exp(-k/18).$$

The upper estimate

$$\Pr_{x_{[n]\setminus[k]}\in\mathbb{Z}_2^{[n]\setminus[k]}}\left[\langle x_{[n]\setminus[k]},v_{[n]\setminus[k]}\rangle \geq \delta\cdot\|v_{[n]\setminus[k]}\|_2\right] \leq$$

$$\exp(-\delta^2/2)$$

also follows immediately from the Chernoff-Hoeffding bound. □

When considering caps and not linear threshold functions the results are somewhat easier. Recall that $C_{v,\theta}$ is defined as $C_{v,\theta} = \{x\in\mathbb{S}^{n-1} : \langle v,x\rangle \geq \theta\}$. For a proof of the next lemma see e.g. [21].

LEMMA 2.11. *Let $v\in\mathbb{S}^{n-1}$ be a unit vector. Then*

$$\Pr_{x\in\mathbb{S}^{n-1}}[x\in C_{v,\theta}] \leq \exp\left(-\frac{1}{2}n\theta^2\right),$$

where we consider the uniform probability measure on $\mathbb{S}^{n-1}$.

3. CONSTRUCTION OF A COVERING CODE

As a warm up for the proof of Theorem 1.1 we give an explicit construction of a cover code of covering radius $\frac{n}{2} - c\sqrt{n\log n}$ for $\mathcal{B}_n$. Later on we will build on the ideas of the proof to get the more general result. For convenience we repeat the claim of Corollary 1.3 here.

Corollary 1.3: There exists $a>0$ such that for every $c>0$ there is an explicit construction of a set $C\subset\mathcal{B}_n$ of size $|C| = n^2\cdot(n^c)^a$ such that for every $z\in\mathcal{B}_n$ there is some $x\in C$ with $\mathcal{H}(z,x)\leq\frac{n}{2}-\sqrt{cn\log n}$.

PROOF. Fix $c>0$, and let $n\in\mathbb{N}$. Put $t=\lceil c_1\log n\rceil$, for a sufficiently large constant c_1 that will be later determined. For simplicity we assume that t divides n. Let $J_1,J_2,\ldots,J_t$ be the partition of $[n]$ defined by $J_i=\{(i-1)\cdot n/t+1,\ldots,i\cdot n/t\}$ (in fact, we can take the J_i-s to be any partition of the coordinates into t disjoint sets, each of size n/t). Let $S\subset\{-1,1\}^{n/t}$ be a 4-wise independent distribution. Let $m=|S|$ and recall that by Fact 2.1 we can assume that $m=O((n/t)^2)$.

Denote $S = \{s_0, \ldots, s_{m-1}\}$. The set C is defined as follows. For every sequence of signs $\alpha = (\alpha_1, \ldots, \alpha_t) \in \{-1,1\}^t$ and every $0 \leq j \leq m-1$, let $x^{\alpha,j} \in \mathcal{B}_n$ be defined as the concatenation $(\alpha_1 \cdot s_j) \circ \ldots \circ (\alpha_t \cdot s_{(j+t-1 \bmod m)})$. That is, $x^{\alpha,j}_{J_i} = \alpha_i \cdot s_{(j+i-1 \bmod m)}$. In other words, we concatenate t consecutive elements of S for each of the 2^t possible sign flips. The set C is the collection of all the $x^{\alpha,j}$-s, i.e. $C = \{x^{\alpha,j} : \alpha \in \{-1,1\}^t,\ 0 \leq j < m\}$. Hence, the size of C is $2^t \cdot m = O((\frac{n}{t})^2 \cdot 2^t) \leq n^2 \cdot n^{c_1}$.

We now proceed with the analysis of this construction. As S is 4-wise independent we get by Lemma 2.3 that for every $y \in \{-1,1\}^{n/t}$

$$\Pr\left[\,|\langle y, s\rangle| > \sqrt{n/3t}\right] \geq \frac{2}{11}.$$

Fix $z \in \mathcal{B}_n$. Denote the event that $|\langle z_{J_i}, s_{j+i-1 \bmod m}\rangle| > \sqrt{n/3t}$ with X_i (where $0 \leq j \leq m-1$ is picked uniformly at random). Recall that $\mathbb{E}[X_i] \geq 2/11$ and so, by linearity of expectation, we get that $\mathbb{E}\left[\sum_{i=1}^t X_i\right] \geq 2t/11$. Therefore, for every $z \in \mathcal{B}_n$ there exists $j_z \in \{0, \ldots, m-1\}$ such that

$$\left|\left\{i : |\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle| \geq \sqrt{n/3t}\right\}\right| \geq \frac{2t}{11}.$$

Set $\alpha \in \{-1,1\}^t$ as $\alpha_i = \mathrm{sign}(\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle)$. It follows that

$$\left\langle z, x^{\alpha,j}\right\rangle = \sum_{i=1}^{t} |\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle| \geq \frac{2t}{11}\sqrt{n/3t} \geq$$

$$\frac{2\sqrt{c_1}}{11\sqrt{3}}\sqrt{n \log n}.$$

To complete the proof, set $c_1 = 400c$ to get $\left\langle z, x^{\alpha,j}\right\rangle > 2\sqrt{cn \log n}$. We thus obtain that,

$$\mathcal{H}(z, x^{\alpha,j}) = \frac{n}{2} - \frac{1}{2}\left\langle z, x^{\alpha,j}\right\rangle \leq \frac{n}{2} - \sqrt{cn \log n}.$$

Moreover, $|C| \leq n^2 \cdot n^{c_1} = n^2 \cdot (n^c)^{400}$, as required. □

4. THE MAIN CONSTRUCTION

We now give an explicit construction of an ϵ-net set $N_\epsilon \subset \mathcal{B}_n$ for linear threshold functions. In particular we will prove Theorem 1.1. For convenience we repeat it here.

Theorem 1.1 There exists two universal constants $a, b > 0$ such that for every $\epsilon > 0$ there is an explicit construction of an ϵ-net, $N_\epsilon \subset \mathcal{B}_n$, for linear threshold functions of size $|N_\epsilon| = O(\epsilon^{-b} \cdot n^a)$.

PROOF. Before giving the construction we explain what changes are needed from the earlier construction of the covering code. Consider a unit vector $v' \in \{-1,1\}^{n/\log n}$ and let v be the unit vector in $\mathbb{R}^n$ having $v'/\|v'\|_2$ in its first $n/\log n$ coordinates and zeros elsewhere. Consider the linear function $L_{v,\sqrt{\log n}} : \mathcal{B}_n \to \{-1,1\}$. It is not hard to see that with probability $1/\mathrm{poly}(n)$ over the choice of $x \in \mathcal{B}_n$ we have that $L_{v,\sqrt{\log n}}(x) = 1$, for every such v. On the other hand, there exists a v' (and actually a random v' will have the required property) such that for every $y \in C$, where C is the cover code constructed in Section 3, we will have that $|\langle y, v\rangle| = O(1)$. Thus, for every $y \in C$ we have that $L_{v,\sqrt{\log n}}(y) = 0$. Therefore C is not a $1/\mathrm{poly}(n)$-net. The reason for the failure of C is that all the large coordinates of v were concentrated on a set of size $n/\log n$ that was one of the sets in the partition of the coordinates with respect to which we constructed C. To overcome this difficulty we construct sets in analogous way to the construction of C but with respect to different partitions of the n coordinates. These partitions will come from the family of perfect hash functions discussed in Section 2.3. Another change that we will have to make is in the way that we concatenate short strings (of length $O(n/\log n)$ in order to get length n strings. Previously we simply concatenated consecutive strings. Now we will have to concatenate them according to an expander walk. The reason being that there will be $O(\log n)$ sets in the partitions from which we will have to make sure that we get the "correct" contribution. We now turn to the actual construction (also replacing $1/\mathrm{poly}(n)$ with ϵ).

Let $\epsilon > 0$ be given. We assume that $\epsilon > 2^{-n/100}$ as otherwise we can pick $N = \mathcal{B}_n$. Let $t = \lceil c \log 2/\epsilon \rceil$, for some absolute constant c that will be later determined. We assume w.l.o.g. that $t \geq 24$. We will later need this assumption (without explicitly referring to it) for applying the result of Corollary 2.8. Set $k = 5$ and $d = 2^{18}$. Similarly to the case of cover codes, let $S \subset \{-1,1\}^n$ be a k-wise independent sample space. Let $m = |S|$. By Fact 2.1 we can assume that $m = |S| = O(n^{k/2})$. Denote $S = \{s_i\}_{i=1}^m$. As mentioned above we will need to consider many different partitions of the coordinates, so let $\mathcal{H}$ be the $(n, 8t, t)$-perfect hash family guaranteed by Lemma 2.7. We think of every $h \in \mathcal{H}$ as partitioning the coordinates to $8t$ sets $\{J_{h,1}, \ldots, J_{h,8t}\}$ with $J_{h,i} = h^{-1}(i)$. Let $J_h = \{J_{h,1}, \ldots, J_{h,8t}\}$ be the collection of the sets in the partition. Note that the sets in J_h are not necessarily of the same size. In order to concatenate elements of S to create a word in $\mathcal{B}_n$ we need to consider random walks on an expander graph. Let G be an $(m, d, d/100)$-expander with node set S. In other words, we identify the i-th node of G with s_i. In particular a random walk $(w_1, \ldots, w_\ell)$ on G is a sequence of ℓ elements from S. We now explain how to mix all these ingredients together to get the final construction.

The set N_ϵ contains all the points $x^{h,w}$ (that will be soon defined), where $h \in H$ and w is a walk of length $8t$ in G. We now explain how to construct $x^{h,w}$. Let $h \in \mathcal{H}$ be an hash function and $w = (w_1, \ldots, w_{8t}) \in S^{8t}$ be a random walk on G. Let $i \in \{1, 2, \ldots, 8t\}$. Let w'_i be the first $|J_{h,i}|$ bits of w_i. The reason for this is that it may be the case (and it is most likely the case) that $|J_{h,i}| < n$ and so we need to cut the last bits of w_i to get a vector of length exactly $|J_{h,i}|$. We now define

$$x^{h,w}|_{J_{h,i}} = w'_i = \text{ first } |J_{h,i}| \text{ bits of } w_i.$$

As the collection $\{J_{h,i}\}_{i=1}^{8t}$ is a partition of $[n]$ we get that indeed $x^{h,w} \in \mathcal{B}_n$.

A good way to understand the construction is the following. We would like to define a point $x = x^{h,w} \in N_\epsilon$. To do so we first map the coordinates of x to $8t$ buckets according to h. Assume that the set $J_{h,i}$ was mapped to the i'th bucket. Now, we would like to assign a value to $x_{J_{h,i}}$ from the k-wise independent set S, and we would like to do so for every $i \in [8t]$. As there are m^{8t} possibilities for such assignments we have to pick a small subset of all possible assignments. We do so by taking an expander walk on an expander with m vertices. Given a walk $w = (w_1, \ldots, w_{8t})$ of length $8t$ we would like to consider the assignment $x_{J_{h,i}} = w_i$. The final thing to notice is that $|J_{h,i}|$ may be smaller than n and so we only consider the

first $|J_{h,i}|$ bits of w_i. Going over all $i \in [8t]$ we get the vector $x^{h,w}$. An easy bound on the size of N_ϵ is

$$|N_\epsilon| = d^{8t-1} \cdot m = O(d^7 \cdot (2/\epsilon)^{8c \log d} \cdot n^{k/2}) =$$

$$O\left(n^a \cdot (1/\epsilon)^b\right),$$

where $a = k/2$ and $b = 8c \log d$ are absolute constants. We now show that N_ϵ is an ϵ-net for linear threshold functions. Let $L_{v,\theta}$ be a linear threshold function, where $\|v\|_2 = 1$, such that

$$\Pr_{x \in \mathcal{B}_n}[L_{v,\theta}(x) = 1] \geq \epsilon.$$

Let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $v_{i_1} \geq v_{i_2} \geq \cdots \geq v_{i_n}$, and let I_r denote the set $\{i_1, i_2, \ldots, i_r\}$. We now show that there exists $x^{h,w}$ in N_ϵ for which $L_{v,\theta}(x^{h,w}) = 1$ which implies that N_ϵ is an ϵ-net for linear threshold functions.

We analyze three different cases. The first is when the support of v is small. The second is when the support is not too small, but most of the mass of v is concentrated on a few coordinates (this case corresponds to the first bullet in Corollary 2.8). The last case is when the mass of v is "nicely" spread. We shall make use of the following notations. Given the k-wise independent set S and an index $i \in [8t]$, consider the first $J_{h,i}$ coordinates of every element in S. Denote this set with $S_{h,i}$. Clearly $S_{h,i}$ is k-wise independent. We also define, for every $i \in [8t]$, $J'_{h,i} = h^{-1}(i) \setminus I_t$.

Case 1: Assume that the size of the support of v is at most t. Clearly, for every $x \in \mathcal{B}_n$ we have that $\langle x, v\rangle \leq \|v\|_1$. We now show that there is some $x^{h,w} \in N_\epsilon$ with $\langle x, v\rangle = \|v\|_1$. This clearly implies that $L_{v,\theta}(x^{h,w}) = 1$. Indeed, Lemma 2.7 guarantees that there is some $h \in \mathcal{H}$ that is injective on I_t. Namely, it maps all the nonzero coordinates of v to different buckets. As a bucket now contains at most one nonzero element, we see that for each $i \in [8t]$ we have that

$$\Pr_{s \in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle = \|v_{I_t \cap J_{h,i}}\|_1\right] \geq \frac{1}{2}, \tag{2}$$

where we used the fact that each bucket contains at most one nonzero element so we only need s to have the correct sign. For every $i \in [8t]$ denote with $A_i \subseteq S_{h,i}$ the set of $s \in S_{h,i}$ that belong to the "good" sets defined in Equation (2). Namely, those elements from $S_{h,i}$ that have a large inner product with $v_{J_{h,i}}$. Clearly, for every i we have that $|A_i|/|S_{h,i}| \geq \frac{1}{2}$. We will now show that there exist a random walk on G such that for every i, $w_i \in A_i$. Indeed, G is an $[n, d, \lambda]$-expander and so Theorem 2.6 guarantees that if $\frac{1}{2} > 2\lambda/d$ then there exists a random walk that hits all the A_i's. As we picked a graph G with $\lambda \leq d/100$ we have the required property. Thus, there exists a walk $w = (w_1, \ldots, w_{8t})$ such that for every i, $w_i \in A_i$. Calculating we get that

$$\left\langle x^{h,w}, v\right\rangle = \sum_{i=1}^{8t} \left\langle w_i, v_{J_{h,i}}\right\rangle =$$

$$\sum_{i \in h(I_t)} \left\langle w_i, v_{J_{h,i}}\right\rangle = \sum_{i \in h(I_t)} |v_i| = \|v\|_1$$

as required. This completes the analysis of the first case.

Case 2: Assume that $\sum_{r=\lceil 2t/3\rceil}^{t-1} |v_{i_{t+1}}| \geq \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2$ (this is the first bullet of Corollary 2.8). Similarly to the first case (or, using Lemma 2.5) we get that there is $x^{h,w} \in N_\epsilon$ such that

$$\left\langle x^{h,w}, v\right\rangle \geq \|v_{I_t}\|_1 \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2. \tag{3}$$

By Corollary 2.10 we get that

$$\Pr_{x \in \mathcal{B}_n}\left[\langle x, v\rangle \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2\right] \leq$$

$$\exp(-t/18) + \exp\left(-\frac{1}{2}\left(\frac{\sqrt{t}}{32}\right)^2\right) = \exp(-\gamma t),$$

for some absolute constant $\gamma > 0$. If we pick c large enough (i.e. $c \geq 1/\gamma$) then for $t = \lceil c \log(2/\epsilon)\rceil$ we get that

$$\Pr_{x \in \mathcal{B}_n}\left[\langle x, v\rangle \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2\right] \leq \exp(-\gamma t) < \epsilon.$$

As we assumed that $\Pr_{x \in \mathcal{B}_n}[L_{v,\theta}(x) = 1] \geq \epsilon$ we have that

$$\theta < \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2. \tag{4}$$

By Equation (3) it now follows that there is $x^{h,w} \in N_\epsilon$ such that

$$\left\langle x^{h,w}, v\right\rangle \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2 > \theta.$$

Hence, for this $x^{h,w}$ we get that $L_{v,\theta}(x^{h,w}) = 1$ as required. This completes the analysis of the second case.

Case 3: We now assume that $\sum_{r=\lceil 2t/3\rceil}^{t-1} |v_{i_{t+1}}| < \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2$. Hence, Corollary 2.8 implies that there exists some $h \in \mathcal{H}$ such that h is an injection on I_t and for at least $c_1 \cdot 8t$ buckets $r \in [8t]$ it holds that $\|v_{h^{-1}(r)\setminus I_t}\|_2^2 \geq \frac{c_2}{t} \cdot \|v_{[n]\setminus I_t}\|_2^2$, for two universal constants c_1 and c_2. Denote the set of $\geq c_1 \cdot 8t$ "good" buckets r with $R \subset [8t]$. It follows that for every $i \in R$

$$\|v_{J'_{h,i}}\|_2 \geq \sqrt{\frac{c_2}{t}} \cdot \|v_{[n]\setminus I_t}\|_2.$$

By Claim 2.5, specialized to $k = 5$, we get that for every $i \in h(I_t)$

$$\Pr_{s \in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle \geq \|v_{I_t \cap J_{h,i}}\|_1 + \frac{1}{7}\|v_{J'_{h,i}}\|_2\right] \geq$$

$$\frac{4}{5} \cdot 2^{-5} = \frac{1}{40}, \tag{5}$$

where we recall that by our assumption on h we have that $|I_t \cap J_{h,i}| = 1$. In addition, Lemma 2.4 implies that for $i \notin h(I_t)$

$$\Pr_{s \in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle \geq \frac{\|v_{J_{h,i}}\|_2}{7}\right] \geq \frac{1}{20}. \tag{6}$$

For every $i \in [8t]$ denote with $A_i \subseteq S_{h,i}$ the set of $s \in S_{h,i}$ that belong to the "good" sets defined in Equations (4), (6). Namely, those elements from $S_{h,i}$ that have large inner product with $v_{J_{h,i}}$. Clearly, for every i we have that $|A_i|/|S_{h,i}| \geq \min(\frac{1}{40}, \frac{1}{20}) = \frac{1}{40}$. We will now show that there exist a random walk on G such that for every i, $w_i \in A_i$. Indeed, G is an

$[n, d, \lambda]$-expander and so Theorem 2.6 guarantees that if $\frac{1}{40} > 2\lambda/d$ then there exists a random walk that hits all the A_i's. As we picked a graph G with $\lambda \leq d/100$ we have the required property. Thus, there exists a walk $w = (w_1, \ldots, w_{8t})$ such that for every i, $w_i \in A_i$. Calculating we get that

$$\left\langle x^{h,w}, v \right\rangle = \sum_{i=1}^{8t} \left\langle w_i, a_{J_{h,i}} \right\rangle$$

$$= \sum_{i \in h(I_t)} \left\langle w_i, v_{J_{h,i}} \right\rangle + \sum_{i \notin h(M)} \left\langle w_i, v_{J_{h,i}} \right\rangle \geq$$

$$\sum_{i \in h(I_t)} (\|v_{I_t \cap J_{h,i}}\|_1 + \frac{1}{7}\|v_{J'_{h,i}}\|_2) + \sum_{i \notin h(I_t)} \frac{\|v_{J_{h,i}}\|_2}{7} =$$

$$\|v_{I_t}\|_1 + \frac{1}{7} \sum_{i \in [8t]} \|v_{J'_{h,i}}\|_2 \geq \|v_{I_t}\|_1 + \frac{1}{7} \sum_{i \in R} \|v_{J'_{h,i}}\|_2 \geq$$

$$\|v_{I_t}\|_1 + \frac{1}{7} \sum_{i \in R} \sqrt{\frac{c_2}{t}} \cdot \|v_{[n] \setminus I_t}\|_2 \geq^{\ddagger}$$

$$\|v_{I_t}\|_1 + \frac{8c_1\sqrt{c_2}}{7} \cdot \sqrt{t} \cdot \|v_{[n] \setminus I_t}\|_2 \geq$$

$$\|v_{I_t}\|_1 + \frac{8c_1\sqrt{c \cdot c_2}}{7} \cdot \sqrt{\log(2/\epsilon)} \cdot \|v_{[n] \setminus I_t}\|_2 \geq^{*}$$

$$\|v_{I_t}\|_1 + \sqrt{2\log(2/\epsilon)} \cdot \|v_{[n] \setminus I_t}\|_2 >^{\dagger} \theta,$$

where inequality $(\ddagger)$ follows from the fact that $|R| \geq c_1 \cdot 8t$, inequality $(*)$ holds for a large enough universal constant c and inequality $(\dagger)$ holds by Equation (4) (the bound on θ from **Case 2** also holds here of course). Thus, $L_{v,\theta}(x^{h,w}) = 1$ as required. This concludes the proof of Theorem 1.1. $\square$

5. CONSTRUCTION OF ϵ-NETS FOR SPHERICAL CAPS

In this section we show how to construct ϵ-nets for spherical caps. In particular we prove Theorem 1.2.

Theorem 1.2 There exists two universal constants $a, b > 0$ such that for every $\epsilon > 0$ there is an explicit construction of an ϵ-net, $S_\epsilon \subset \mathbb{S}^{n-1}$, for spherical caps of size $|S_\epsilon| = O(\epsilon^{-b} \cdot n^a)$.

A first natural attempt is to check whether the ϵ-net for threshold functions is also an ϵ-net for spherical caps. As we are looking for subsets of the sphere $\mathbb{S}^{n-1}$ we consider the natural embedding of $\mathcal{B}_n$ in $\mathbb{S}^{n-1}$ that shrinks every vector by a factor of $\sqrt{n}$. Namely, set $\overline{\mathcal{B}_n} = \{-1/\sqrt{n}, 1/\sqrt{n}\}^n$. In this section whenever we discuss the boolean cube we will refer to the set $\overline{\mathcal{B}_n}$. In particular we will view every subset of $\mathcal{B}_n$ as a subset of $\overline{\mathcal{B}_n}$. To see that the boolean cube (as a subset of $\mathbb{S}^{n-1}$) is not an ϵ-net for a polynomially small ϵ consider the cap defined by $v = (1, 0, \ldots, 0)$ and $\theta = \sqrt{\log(1/\epsilon)/n}$. We see that $L_{v,\theta}(\mathcal{B}_n) = 0$ whereas the cap $C_{v,\theta} = L_{v,\theta}^{-1}(1) \cap \mathbb{S}^{n-1}$ has measure $\mathrm{poly}(\epsilon)$. However, it turns out that if an ϵ-net for LTFs does not hit a large enough cap, then a "rotation" of it does hit the cap. Therefore, the union of an ϵ'-net for linear threshold functions and its rotation yields an ϵ-net for spherical caps. Indeed, the reason that $v = (1, 0, \ldots, 0)$ and $\theta = \sqrt{\log(1/\epsilon)/n}$ show that the boolean cube is not an ϵ-net is that all the mass of v is concentrated on a few coordinates (actually only 1 coordinate). On the other hand, if it was the case that no set of $O(\log(1/\epsilon))$ coordinates contains more than, say, $3/4$ of the total mass of v then the set N_ϵ guaranteed by Theorem 1.1, will hit the cap $C_{v,\sqrt{2\log(1/\epsilon^{1/16})/n}}$ which by Lemma 2.11 is of weight at most $\epsilon^{1/16}$. Indeed, repeating the proof of Theorem 1.1 we see that there is an element $x \in N_\epsilon$ such that if $M \subset [n]$ is the set of $O(\log 1/\epsilon)$ largest coordinates of v then

$$\langle x, v \rangle > \sqrt{2\log(1/\epsilon)/n} \cdot \|v_{[n] \setminus M}\|_2 \geq^{(*)}$$

$$(1/4) \cdot \sqrt{2\log(1/\epsilon)/n} = \sqrt{2\log(1/\epsilon^{1/16})/n},$$

where inequality $(*)$ follows from the fact that at least $1/4$ of the mass of v is supported on the set of coordinates $[n] \setminus M$. Hence, all that we have to do is to find a way of spreading out the coordinates of v so that the mass is "nicely" distributed on many coordinates. Our approach to solving this problem is the following: We show that for the Fourier matrix F, either Fv has the property that its mass is "well spread" or v itself is well spread. Then we simply let $S_\epsilon = N_{\epsilon'} \cup F(N_{\epsilon'})$ for some $\epsilon' = \mathrm{poly}(\epsilon)$ where $N_{\epsilon'}$ is an ϵ'-net for LTF's. We now give the formal proof.

PROOF OF THEOREM 1.2. As before we let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $v_{i_1} \geq v_{i_2} \geq \cdots \geq v_{i_n}$, and I_r denote the set $\{i_1, i_2, \ldots, i_r\}$. Assume that [2] $n = 2^k$ for some integer k. Let F be the $n \times n$ Fourier matrix. In other words, each coordinate of F is in $\{-1/\sqrt{n}, 1/\sqrt{n}\}$ and the rows of F are orthogonal. The following lemma shows that Fv or v are "well spread".

LEMMA 5.1. *For every two subset $M_1, M_2 \subset [n]$ of size $|M_1|, |M_2| \leq \sqrt{n}/20$ and any unit vector $v \in \mathbb{R}^n$ we have that $\|(Fv)_{M_1}\|_2 \leq 3/4$ or $\|v_{M_2}\|_2 \leq 3/4$.*

PROOF. The proof follows the following lemma of [17] (specialized for $L = 2$).

LEMMA 5.2 (LEMMA 4.2 OF [17]). *Let T be a matrix obtained by concatenating rows of two unitary $n \times n$ matrices H_1 and H_2 with coherence[3] δ. Then, for any set of coordinates $M \subset [2n]$ of size $|M| = s$, and any unit vector $v \in \mathbb{R}^n$ we have that $\|(Tv)_M\|_2^2 \leq \frac{1}{2}(1 + \delta s) \cdot \|Tv\|_2^2$.*

Indeed, let T be the matrix whose first n rows are the identity matrix and the last n rows are F. Then, the coherence of T is $\delta = 1/\sqrt{n}$. Given two subsets $M_1, M_2 \subset [n]$ of size $|M_1|, |M_2| \leq \sqrt{n}/20$, let M'_2 be the subset of $\{n+1, \ldots, 2n\}$ obtained by adding n to each element of M_2. Let $M = M_1 \cup M'_2$. Then for any unit vector $v \in \mathbb{R}^n$ it holds that

$$\|(Tv)_M\|_2 \leq \sqrt{\frac{1}{2}(1 + \delta|M|)} \cdot \|Tv\|_2 \leq$$

$$\sqrt{1.1/2} \cdot \|Tv\|_2 < 3/4.$$

This completes the proof of Lemma 5.1. $\square$

[2] If it is not the case then we can work with $n' = 2^k$ such that $n < n' < 2n$.

[3] The coherence of H_1 and H_2 is the largest inner product between a row of H_1 and a row of H_2.

Let $N_{\epsilon'} \subset \{-1/\sqrt{n}, 1/\sqrt{n}\}^n$ be an ϵ'-net for linear threshold functions, for some ϵ' that will be later determined. Define $S_\epsilon = N_{\epsilon'} \cup F(N_{\epsilon'})$. In other words, S_ϵ is the union of $N_{\epsilon'}$ with the rotation of $N_{\epsilon'}$ by F. Note that as F is unitary we have that indeed $S_\epsilon \subset \mathbb{S}^{n-1}$. We now show that S_ϵ is indeed an ϵ-net for spherical caps. Let $C_{v,\theta}$ be a spherical cap of weight ϵ. By Lemma 2.11 we see that $\theta \leq \sqrt{2\log(1/\epsilon)/n}$. Let $u = Tv$, where T is the matrix defined in the proof of Lemma 5.1. By Lemma 5.1 we get that no set of $\sqrt{n}/10$ coordinates of u contains more than $3/4$ of the total mass of u. As $u = Tv = (v, Fv)$ (the concatenation of v and Fv) and $\|v\| = \|Fv\|$ we get that either in v or in Fv, no set of $\sqrt{n}/20$ coordinates contains more than $3/4$ of the total mass. Assume w.l.o.g. that in Fv no set of $\sqrt{n}/20$ coordinates contains more than $3/4$ of the total mass (the analysis for v is similar). Let $I_t \subset [n]$ be the set of largest[4] $t = \lceil c\log(1/\epsilon')\rceil \leq \sqrt{n}/20$ coordinates of Fv (note that c, t and I_t are chosen as in the proof of Theorem 1.1). In particular, no coordinate in I_t is the zero coordinate. Following the proof of Theorem 1.1, we note that we are either in **Case 2** or **Case 3** there and hence, for a large enough c, $N_{\epsilon'}$ contains an element $x \in N_{\epsilon'}$ such that[5]

$$\langle x, Fv\rangle \geq^{(\dagger)} \frac{1}{\sqrt{n}} \cdot \sqrt{2\log(1/\epsilon')} \cdot \|(Fv)_{[n]\setminus I_t}\|_2 \geq^{(*)}$$

$$\frac{1}{\sqrt{n}} \cdot \sqrt{2\log(1/\epsilon')} \cdot \frac{1}{4} = \sqrt{2\log(1/\epsilon'^{1/16})/n},$$

where inequality $(\dagger)$ is implied either by Equation (3) (in **Case 2**) or by the conclusion of **Case 3**. Inequality $(*)$ follows from the fact that $\lceil c\log(1/\epsilon')\rceil < \sqrt{n}/20$ and the assumption that every subset of $\sqrt{n}/20$ coordinates of Fv contains at most $3/4$ of the mass of Fv. Hence, $Fx \in F(N_{\epsilon'}) \subset S_\epsilon$ and

$$\langle Fx, v\rangle = \langle x, Fv\rangle \geq \sqrt{2\log(1/\epsilon'^{1/16})/n} =$$

$$\sqrt{2\log(1/\epsilon)/n} \geq \theta,$$

for $\epsilon' = \epsilon^{16}$. This shows that S_ϵ is indeed an ϵ-net for spherical caps. Moreover, we have that

$$|S_\epsilon| \leq 2|N_{\epsilon'}| \leq O(\epsilon'^{-b} \cdot n^a) = O(\epsilon^{-b'} \cdot n^a),$$

for absolute constants a and b'. This completes the proof of Corollary 1.3.

Acknowledgement

We thank Noga Alon and Avi Wigderson for helpful discussions and for bringing [23] to our attention. We also thank Noga for sharing his proof of corollary 1.3 with us.

6. REFERENCES

[1] A. Lubotzky, R. Phillips, and P.Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[2] N. Alon. Private communication, 2008.

[3] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

[4] N. Alon, G. Gutin, and M. Krivelevich. Algorithms with large domination ratio. *J. Algorithms*, 50(1):118–131, 2004.

[5] N. Alon, H. Kaplan, G. Nivasch, M. Sharir, and S. Smorodinsky. Weak ϵ-nets and interval chains. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 1194–1203, 2008.

[6] N. Alon and J. Spencer. *the probabilistic method*. J. Wiley, 3 edition, 2008.

[7] B. Berger. The fourth moment method. *SIAM J. Comput.*, 26(4):1188–1207, 1997.

[8] E. J. Candés and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[9] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18:143–154, 1979.

[10] B. Chazelle. Computational geometry: a retrospective. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing (STOC)*, pages 75–94, 1994.

[11] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[12] L. Engebretsen, P. Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 705–712, 2002.

[13] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Velickovic. Approximations of general independent distributions. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 10–16, 1992.

[14] M. L. Fredman, J. Komlós, and E. Szemerédi. Storing a sparse table with 0(1) worst case access time. *J. ACM*, 31(3):538–544, 1984.

[15] V. Guruswami, J. R. Lee, and A. A. Razborov. Almost euclidean subspaces of l_1^n via expander codes. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 353–362, 2008.

[16] V. Guruswami, J. R. Lee, and A. Wigderson. Euclidean sections of with sublinear randomness and error-correction over the reals. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, 11th International Workshop, APPROX 2008, and 12th International Workshop, RANDOM 2008*, pages 444–454, 2008.

[17] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of l2 into l1. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 615–620, 2007.

[18] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, 26:189ÂŰ206, 1984.

[19] N. Linial, M. Luby, M. E. Saks, and D. Zuckerman. Efficient construction of a small hitting set for combinatorial rectangles in high dimension. *Combinatorica*, 17(2):215–234, 1997.

[20] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problems of*

[4] Recall that we assume that $\epsilon > \exp(-O(\sqrt{n}))$.

[5] The factor of $\frac{1}{\sqrt{n}}$ comes from viewing $\mathcal{B}_n$ as a subset of $\mathbb{S}^{n-1}$. In fact, we can get a much better inner product but we do not try to optimize.

Information Transmission, 24(1):39–46, 1988.

[21] J. Matousek. *Lectures on discrete Geometry*. GTM. springer, 2002.

[22] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[23] J. P. Schmidt and A. Siegel. The analysis of closed hashing under limited randomness (extended abstract). In *Proceedings of the Twenty Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 224–234, 1990.

[24] D. Sivakumar. Algorithmic derandomization via complexity theory. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 619–626, 2002.

[25] E. Viola. The sum of d small-bias generators fools polynomials of degree d. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity (CCC)*, pages 124–127, 2008.

APPENDIX

A. PERFECT HASHING

PROOF OF LEMMA 2.7. Our proof uses the construction of perfect hash families due to Schmidt and Siegel [23]. The construction is a clever oblivious implementation of the Fredman, Komlós, and Szemerédi adaptive hashing scheme [14].

The FKS scheme proceeds in two phases. The first phase of Schmidt and Siegel's construction is identical to the first phase of the FKS scheme. It applies a map $f : [n] \to [s]$, taken from a pairwise independent family of hash functions $\mathcal{F}$. There are known explicit constructions of $\mathcal{F}$ with $|\mathcal{F}| = 2^{\log s + \log\log n + O(1)}$ [9]. A pairwise independent family of hash functions $\mathcal{F}$ has the following property. If f is chosen uniformly at random from $\mathcal{F}$, then for every $x, y \in [n]$, $x \neq y$, it holds that $f(x)$ is distributed uniformly in $[s]$, even when conditioned on $f(y)$. In particular, $\Pr[f(x) = f(y)] = \frac{1}{s}$.

Let $S \subset \{1, 2, \ldots, n\}$ be an arbitrary set of size $|S| \leq s$. Consider the following event.

$$\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2 < 4s. \tag{7}$$

We now show that the probability of this event, when f is chosen uniformly at random from $\mathcal{F}$, is more than $\frac{1}{2}$. Indeed, denoting by χ_p the indicator of an event p, we have that

$$\mathbb{E}\left[\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2\right] = \mathbb{E}\left[\sum_{x,y \in S} \chi_{f(x)=f(y)}\right] =$$

$$\sum_{x,y \in S} \mathbb{E}\left[\chi_{f(x)=f(y)}\right] \leq 2s - 1.$$

By applying Markov's inequality we conclude that

$$\Pr\left[\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2 \geq 4s\right] < \frac{1}{2}. \tag{8}$$

The second phase of the FKS hashing scheme is adaptive, and depends on the hashed set S. The idea is the following. If c_i elements of S landed in bucket $i \in [s]$, then by mapping this bucket to c_i^2 buckets using a pairwise independent family of hash functions, it is likely that no collision between the elements of S occurs. As the first phase guarantees that $\sum_{i \in [s]} c_i^2 = O(s)$, we end up with a hash table of size $O(s)$. The Schmidt and Siegel implementation proceeds as follows. It uses a pairwise independent family of hash functions $\mathcal{G}$. Here it will be convenient to assume that $g \in \mathcal{G}$ maps $[n]$ to bit vectors. So every $g \in \mathcal{G}$ is a function $g : [n] \to \{0,1\}^{2+\log s}$. We can take $|\mathcal{G}| = 2^{\log s + \log\log n + O(1)}$. The second phase uses a selection of s (not necessarily distinct) hash functions from $\mathcal{G}$. The hash functions are selected and used as follows. Take a sequence of $\log s$ hash functions $g_1, g_2, \ldots, g_{\log s} \in \mathcal{G}$. Notice that there are at most $|\mathcal{G}|^{\log s} = 2^{\log^2 s + \log s \log\log n + O(\log s)}$ such sequences. Also take a sequence of s non-negative integers $c_1, c_2, \ldots, c_s$ that satisfy $\sum_{j=1}^{s} c_j = s$ and $\sum_{j=1}^{s} c_j^2 < 4s$. There are at most 2^{2s} such sequences (easily bounded by writing the sequence elements in unary notation, separated by zeros). This sequence is our guess of the bucket loads due to S after the first phase. Also use an assignment $a : [s] \to [\log s]$ such that 1 is assigned to $\frac{s}{2}$ elements of $[s]$, 2 is assigned to $\frac{s}{4}$ elements of $[s]$, and in general i is assigned to $\frac{s}{2^i}$ elements of $[s]$. (Exceptionally $\log s$ is assigned to 2 elements of $[s]$.) The number of such assignments is at most $2^{s \cdot (1 + \sum_{i=1}^{\log s} 2^{-i})} < 2^{2s}$ (write the s assigned values in unary, separated by zeros). The assignment a is our guess as to which of the $\log s$ selected hash functions should be used for each bucket.

Each setting of f, g, c and a defines a hash function $h \in \mathcal{H}$ as follows. For every $x \in [n]$,

$$h(x) = \sum_{i < f(x)} 2^{\lceil 2 \log c_i \rceil} + \bar{g}_{a(f(x))}(x),$$

where $\bar{g}_{a(i)}(x)$ is the first $\lceil 2 \log c_i \rceil$ bits of $g_{a(i)}(x)$. Notice that $|\mathcal{H}| \leq 2^{4s + \log^2 s + \log 2s \log\log n + O(\log s)}$, implying the claim in the lemma.[6] Also notice that each $h \in \mathcal{H}$ maps $[n]$ to

$$\sum_{i=1}^{s} 2^{\lceil 2 \log c_i \rceil} \leq 2 \cdot \sum_{i=1}^{s} c_i^2 < 8s,$$

as required.

Consider a vector $v \in \mathbb{S}^{n-1}$. Let $S = I_s$. For this set S, Equation (7) holds for at least half of the choices of f (by Equation (8)). Fix any such choice f. For $i = 1, 2, \ldots, s$, let $C_i = \{x \in S : f(x) = i\}$. Consider the choice of $c_i = |C_i|$, for $i = 1, 2, \ldots, s$. Fix i. For every $g \in \mathcal{G}$ and $x \in [n]$, let $\bar{g}(x)$ denote the first $\lceil 2 \log c_i \rceil$ bits of $g(x)$. Consider the "bad" event $A_i = A_i(g) = \exists x, y \in C_i,\ x \neq y : \bar{g}(x) = \bar{g}(y)$. As $\mathcal{G}$ is a pairwise independent family of hash functions, if g is chosen uniformly at random in $\mathcal{G}$, then $\Pr[A_i] \leq \binom{c_i}{2} \cdot \frac{1}{c_i^2} < \frac{1}{2}$. Therefore, there exists a choice of g_1 that is good for a set $J_1 \subset [s]$ of buckets of cardinality $|J_1| = \frac{s}{2}$. Similarly, for $j = 2, 3, \ldots, \log s - 1$, there exists a choice of g_j that is good for a set $J_j \subset [s] \setminus \bigcup_{j' < j} J_{j'}$ of cardinality $|J_j| = \frac{s}{2^j}$. Similarly, there exists a choice of $g_{\log s}$ that is good for both elements in $[s] \setminus \bigcup_{j < \log s} J_j$. So, for every f that satisfies Equation (7), there is a choice of g, c, and a such that the resulting hash function h is an injection on I_s.

[6] The factor of $\log 2s$ can be saved by adding, prior to the application of f, a preliminary mapping of $[n]$ to $[s^2]$ using another pairwise independent family of hash functions. The maps f and $g_1, g_2, \ldots, g_{\log s}$ then need to be modified to have domain $[s^2]$ instead of $[n]$. In our application, this does not affect the asymptotic bounds beyond lower order terms.

Finally, we show that if there exists $t \in [s-1]$ such that $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then with high probability f satisfies

$$\sum_{r\in[s]} \min\left\{\|v_{f^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \geq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2. \tag{9}$$

Equation (9) implies Equation (1), as the g_i-s only further split hash buckets.

Let X_j^i be the indicator random variable for the event that $f(j)=i$. As $\Pr[f(j)=i]=\frac{1}{s}$, we have that

$$\mathbb{E}\left[\|v_{f^{-1}(i)\setminus I_t}\|_2^2\right] = \mathbb{E}\left[\sum_{j=t+1}^{n} X_j^i v_j^2\right] =$$

$$\frac{1}{s}\cdot\sum_{j=t+1}^{n} v_j^2 = \frac{1}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

Moreover, as f comes from a pairwise independent family of hash functions, for fixed i the random variables X_j^i are pairwise independent, so

$$\sigma^2\left[\|v_{f^{-1}(i)\setminus I_t}\|_2^2\right] = \sigma^2\left[\sum_{j=t+1}^{n} X_j^i v_j^2\right] =$$

$$\sum_{j=t+1}^{n}\sigma^2[X_j^i]\cdot v_j^4 = \left(1-\frac{1}{s}\right)\cdot\frac{1}{s}\cdot\sum_{j=t+1}^{n} v_j^4.$$

Thus, as $v_j^2 \leq \frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2$ for all $j>t$ (since the $|v_j|$-s are non-increasing), we have that

$$\sigma\left[\|v_{f^{-1}(i)\setminus I_t}\|_2^2\right] \leq \frac{1}{\sqrt{s}}\cdot\sqrt{\sum_{j=t+1}^{n} v_j^4} \leq$$

$$\frac{1}{\sqrt{s}}\cdot\frac{1}{8\sqrt{s}}\cdot\|v_{[n]\setminus I_t}\|_2\cdot\sqrt{\sum_{j=t+1}^{n} v_j^2} = \frac{1}{8s}\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

Using Chebyshev's inequality, we have that

$$\Pr\left[\|v_{f^{-1}(i)\setminus I_t}\|_2^2 \geq \frac{r}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right] \leq \frac{1}{64(r-1)^2}.$$

It follows that

$$\int_{\lambda=2^r}^{2^{r+1}} \lambda\cdot\Pr\left[\|v_{f^{-1}(i)\setminus I_t}\|_2^2 = \frac{\lambda}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right]\mathrm{d}\lambda \leq$$

$$\frac{2^{r+1}}{64(2^r-1)^2}.$$

Thus,

$$\mathbb{E}\left[\max\left\{0, \|v_{f^{-1}(i)\setminus I_t}\|_2^2 - \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\}\right] \leq$$

$$\frac{1}{64s}\cdot\|v_{[n]\setminus I_t}\|_2^2\cdot\sum_{r=1}^{\infty}\frac{2^{r+1}}{(2^r-1)^2} =$$

$$\frac{1}{16s}\cdot\|v_{[n]\setminus I_t}\|_2^2\cdot\sum_{r=1}^{\infty}\frac{2^{r-1}}{(2^r-1)^2} <$$

$$\frac{1}{16s}\cdot\|v_{[n]\setminus I_t}\|_2^2\cdot\sum_{r=1}^{\infty}\frac{1}{2^r-1} < \frac{1}{8s}\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

Let $Y^i = \max\left\{0, \|v_{f^{-1}(i)\setminus I_t}\|_2^2 - \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\}$. We have that $\mathbb{E}\left[\sum_{i\in[s]} Y^i\right] < \frac{1}{8}\cdot\|v_{[n]\setminus I_t}\|_2^2$, so by Markov's Inequality, $\Pr\left[\sum_{i\in[s]} Y^i > \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2\right] < \frac{1}{2}$. We now show that when $\sum_{i\in[s]} Y^i \leq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2$ then Equation (9) holds. Let m be the number of $i\in[s]$ such that $\|v_{f^{-1}(i)\setminus I_t}\|_2^2 > \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2$. We now get that

$$\frac{1}{2}\|v_{[n]\setminus I_t}\|_2^2 \geq \sum_{i\in[s]} Y^i =$$

$$\sum_{i\in[s]}\max\left\{0, \|v_{f^{-1}(i)\setminus I_t}\|_2^2 - \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} =$$

$$\sum_{i:\|v_{f^{-1}(i)\setminus I_t}\|_2^2 > \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2}\left(\|v_{f^{-1}(i)\setminus I_t}\|_2^2 - \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right) =$$

$$\sum_{i:\|v_{f^{-1}(i)\setminus I_t}\|_2^2 > \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2}\|v_{f^{-1}(i)\setminus I_t}\|_2^2 - \frac{2m}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2.$$

Hence,

$$\sum_{i=1}^{s}\min\left\{\|v_{h^{-1}(i)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} =$$

$$\sum_{i:\|v_{f^{-1}(i)\setminus I_t}\|_2^2 \leq \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2}\|v_{h^{-1}(i)\setminus I_t}\|_2^2 + \frac{2m}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2 =$$

$$\|v_{[n]\setminus I_t}\|_2^2 -$$

$$\sum_{i:\|v_{f^{-1}(i)\setminus I_t}\|_2^2 > \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2}\|v_{h^{-1}(i)\setminus I_t}\|_2^2 + \frac{2m}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2 \geq$$

$$\frac{1}{2}\|v_{[n]\setminus I_t}\|_2^2,$$

and Equation (9) holds. Thus, there exists $f\in\mathcal{F}$ that satisfies both Equation (7) and Equation (9). This completes the proof. $\square$