

Explicit Dimension Reduction and Its Applications

Zohar S. Karnin* Yuval Rabani† Amir Shpilka‡

Abstract

We construct a small set of explicit linear transformations mapping \mathbb{R}^n to \mathbb{R}^t , where $t = O(\log(\gamma^{-1})\epsilon^{-2})$, such that the L_2 norm of any vector in \mathbb{R}^n is distorted by at most $1 \pm \epsilon$ in at least a fraction of $1 - \gamma$ of the transformations in the set. Albeit the tradeoff between the size of the set and the success probability is sub-optimal compared with probabilistic arguments, we nevertheless are able to apply our construction to a number of problems. In particular, we use it to construct an ϵ -sample (or pseudo-random generator) for linear threshold functions on \mathbb{S}^{n-1} , for $\epsilon = o(1)$. We also use it to construct an ϵ -sample for spherical digons in \mathbb{S}^{n-1} , for $\epsilon = o(1)$. This construction leads to an efficient oblivious derandomization of the Goemans-Williamson MAX CUT algorithm and similar approximation algorithms (i.e., we construct a small set of hyperplanes, such that for any instance we can choose one of them to generate a good solution).

Our technique for constructing ϵ -sample for linear threshold functions on the sphere is considerably different than previous techniques that rely on k -wise independent sample spaces.

*Faculty of Computer Science, Technion, Haifa 32000, Israel. Email: zkarnin@cs.technion.ac.il. Research supported by the Israel Science Foundation (grant number 339/10).

†The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: yraabani@cs.huji.ac.il. Research supported by Israel Science Foundation grant number 1109/07 and by US-Israel Binational Science Foundation grant number 2008059.

‡Faculty of Computer Science, Technion, Haifa 32000, Israel and Microsoft Research, Cambridge MA. Email: shpilka@cs.technion.ac.il. Research partially supported by the Israel Science Foundation (grant number 339/10).

1 Introduction

In this paper we construct a small set of explicit dimension reducing linear transformations mapping vectors in ℓ_2^n to vectors in ℓ_2^t , for $t \ll n$ in a way that preserves their norms and show application of these transformations to several derandomization tasks. We first explain the connection to the Johnson-Lindenstrauss lemma and then discuss some applications of our construction.

Johnson-Lindenstrauss lemma. The celebrated Johnson-Lindenstrauss Lemma [JL84] states the following. In any Hilbert space, a random linear mapping into ℓ_2^t preserves the norm of any vector up to a factor of $1 \pm \epsilon$ with probability at least $1 - \exp(-\epsilon^2 t)$. In fact, quite simple sample spaces suffice for points in ℓ_2^n ; see [DG03, Ach03, Mat08]. Thus, in order to preserve approximately all pairwise distances among n points in a Hilbert space, one can reduce the dimension to $O(\epsilon^{-2} \log n)$.

In addition to its intrinsic impact in functional analysis (see, e.g., [JN10] for a recent discussion), the Johnson-Lindenstrauss Lemma is a cornerstone of high dimensional computational geometry. Its numerous applications include approximate nearest neighbor search, learning mixtures of Gaussians, sketching and streaming algorithms, approximation algorithms for clustering high dimensional data, and speeding up linear algebraic computations (see, e.g., the introduction of [AC09]). Thus, understanding the computational aspects of Johnson-Lindenstrauss style dimension reduction, a so-called JL transform, is fundamentally interesting.

A JL transform can be computed very efficiently by probabilistic algorithms [AC09, AL09]. The probabilistic constructions can be derandomized using the method of conditional expectations [EIO02, Siv02]. However, there is no construction that uses a poly(n) size sample space. Simple and efficient probabilistic constructions typically use $\Omega(n)$ random bits; the derandomization via pseudo-random generators for RL [Siv02] is currently best implemented using $\Omega((\log n)^2)$ random bits [Nis92]. Recently, [CW09] gave a different derandomization using the same amount of bits. They prove that a random sign matrix, where the signs are chosen from a $O(\log(n))$ -wise independent distribution is w.h.p. a JL transform. A construction using $O(\log n)$ random bits would yield a fixed collection of poly(n) mappings that contains, for every configuration of n points, a JL transform for that configuration. Such an explicit construction, aside from its fundamental appeal (a simple probabilistic argument proves its existence), would enable, for example, an efficient deterministic parallel implementation of a JL transform.

We construct a set $\mathcal{A}_{n,\gamma,\epsilon}$ of linear mappings $A : \ell_2^n \rightarrow \ell_2^t$, for $t = O(\log(\gamma^{-1})\epsilon^{-2})$ of cardinality

$$|\mathcal{A}_{n,\gamma,\epsilon}| = n^{1+o(1)} \cdot 2^{O(\log^2(\gamma^{-1}\epsilon^{-1}))}.$$

Note that if $(\gamma\epsilon)^{-1} = \exp(o(\log^{1/2} n))$, then $|\mathcal{A}_{n,\gamma,\epsilon}| = n^{1+o(1)}$. We show that $\mathcal{A}_{n,\gamma,\epsilon}$ satisfies the following.

Theorem 1.1. *For every n and for every γ, ϵ , for every vector $x \in \ell_2^n$, a fraction of at least $1 - \gamma$ of $A \in \mathcal{A}_{n,\gamma,\epsilon}$ satisfy that*

$$(1 - \epsilon) \cdot \|x\|_2 \leq \|Ax\|_2 \leq (1 + \epsilon) \cdot \|x\|_2 .$$

We note that very recently, Meka [Mek10] constructed a dimension reducing set \mathcal{A} of size $(n/\gamma)^{O(\log(\log(n/\gamma)\epsilon^{-1}))}$ (in the notations of Theorem 1.1). Kane and Nelson [KN10] gave, in parallel to Meka, a dimension reducing set of size $n^{O(1)} \cdot \gamma^{-O(\log \log(\gamma^{-1}) + \log(\epsilon^{-1}))}$. For small values of γ (i.e. $\gamma = \exp(-\omega(\sqrt{\log(n)}))$), this gives a smaller set \mathcal{A} than our construction. However, in the case where $\gamma, \epsilon = \exp\left(-o\left(\sqrt{\log(n)}\right)\right)$, our construction gives a set of nearly linear size as opposed to polynomial. When using the set to derandomize algorithms, this difference translates into a significantly faster running time.

Application to ϵ -sample spaces. Even though our explicit construction falls short of providing a $\text{poly}(n)$ -size sample space for the JL transform, we nevertheless use it to derive new and interesting corollaries. In particular, we construct an ϵ -sample for halfspaces and an ϵ -sample for spherical digons in the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. Given a measurable set Ω endowed with a probability measure μ and a family \mathcal{F} of measurable subsets of Ω , a (finite) set $P_\epsilon \subset \Omega$ is called an ϵ -sample for $(\Omega, \mu, \mathcal{F})$ iff for every $F \in \mathcal{F}$,

$$\left| \frac{|P_\epsilon \cap F|}{|P_\epsilon|} - \mu(F) \right| \leq \epsilon.$$

Our first result is an ϵ -sample for halfspaces (or linear threshold functions). More precisely, we consider the case where Ω is \mathbb{R}^n , μ is the uniform (Haar) measure on the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, and \mathcal{F} is all sets of the form $\{x \in \mathbb{R}^n : \langle x, u \rangle \geq \theta\}$, for some $u \in \mathbb{S}^{n-1}$ and $\theta \in \mathbb{R}$. It is easy to show that sampling $O(n/\epsilon)$ points i.i.d. from μ gives an ϵ -sample with high probability.

We prove the following theorem.

Theorem 1.2. *There exists an efficient deterministic algorithm that given input $n \in \mathbb{N}$ and $\epsilon > 0$, constructs a set $Q_\epsilon \subset \mathbb{R}^n$ of cardinality $|Q_\epsilon| = n^{1+o(1)} \cdot 2^{O(\log^2(1/\epsilon))}$ which is an ϵ -sample for halfspaces.*

We remark that the set Q_ϵ is a subset of \mathbb{R}^n and not of \mathbb{S}^{n-1} so it may be viewed as a *weak- ϵ -sample*.

It is instructive to compare our results to other works on similar problems. Hitting sets (or ϵ -nets; a much weaker notion than ϵ -samples) of size $\text{poly}(n/\epsilon)$ for linear threshold functions on the Boolean cube and on the unit sphere were constructed in [RS10]. In [DGJ⁺10], Diakonikolas et al. constructed an ϵ -sample (a.k.a. a pseudo-random-generator) of cardinality $n^{O(\epsilon^{-2} \log^2(1/\epsilon))}$ for linear threshold functions on the Boolean cube. Recently, Meka and Zuckerman [MZ10] gave explicit constructions of an ϵ -sample for linear threshold functions over the Boolean cube. Their ϵ -sample has size $n^{O(1)}$ when $\epsilon > 1/\text{poly} \log(n)$ and size $n^{O(\log(1/\epsilon))}$ when $\epsilon > 1/\text{poly}(n)$. Following our result,¹ the same authors showed (in a modified version of their paper [MZ09]) how to obtain an ϵ -sample for linear threshold functions on the unit sphere (and on the Boolean cube) with size $n^{O(1)} \cdot \epsilon^{-O(\log(\epsilon^{-1}))}$. We note that while we only construct weak- ϵ -sample, [MZ09] construct a set of points on the unit sphere. Thus, our (weak) ϵ -sample is smaller than the adaptation of [MZ10] to the

¹A technical report appeared in '09 [KRS09].

unit sphere (near linear in n as opposed to polynomial) and is also much smaller than the constructions of [DGJ⁺10, MZ10] for the Boolean cube. Concluding, the main difference between our construction and the results mentioned above are: we output a weak- ϵ -sample compared to ϵ -samples; we use a completely different set of techniques (i.e., dimension reduction as opposed to k -wise independence); our construction has an almost linear dependence on n , as opposed to a polynomial dependence in the other constructions. The linear vs. polynomial dependence on n (in ϵ -samples in general) can be a significant factor in derandomization of algorithms as it translates into a linear vs. a polynomial overhead in the running time (as in the case of the MAX-CUT, see section 6). Our methods also give an ϵ -sample for spherical digons. In this case, Ω is the unit sphere \mathbb{S}^{n-1} , endowed with the uniform measure μ . The family \mathcal{F} is the set of spherical digons, i.e., all sets of the form $\{x \in \mathbb{S}^{n-1} : \text{sign}(\langle x, u \rangle) \neq \text{sign}(\langle x, v \rangle)\}$, for some $u, v \in \mathbb{S}^{n-1}$. Here too it is easy to show that sampling $O(n/\epsilon)$ points i.i.d. from μ gives an ϵ -sample with high probability.

Theorem 1.3. *There exists an efficient deterministic algorithm that given input $n \in \mathbb{N}$ and $\epsilon > 0$, constructs a set $P_\epsilon \subset \mathbb{S}^{n-1}$ of cardinality $|P_\epsilon| = n^{1+o(1)} \cdot 2^{O(\log^2(1/\epsilon))}$ which is an ϵ -sample for spherical digons.*

In the aforementioned [MZ10], Meka and Zuckerman gave explicit constructions of ϵ -samples for threshold functions of degree d polynomials, over the Boolean cube. The size of their construction is $n^{1/\epsilon^{O(d)}}$ (i.e., seed length is $\log(n)/\epsilon^{O(d)}$). In [DKN10] similar results are obtained for the case of threshold functions of degree 2 polynomials, over the Boolean cube. We note that while our result only concerns digons and not general degree 2 polynomials, the size of our construction is significantly smaller, namely, $n^{1+o(1)} \cdot 2^{O(\log^2(1/\epsilon))}$ (i.e., the seed length is $(1 + o(1)) \log n + O(\log^2(1/\epsilon))$).

Construction of ϵ -samples (usually referred to as pseudo-random generators when working over the Boolean cube) is a core challenge in the study of randomness and computation. It has applications in computational learning theory, combinatorial geometry, derandomization theory, cryptography, and other areas; see, e.g., [KV94, Cha00, AB09, Gol01]. Our construction (Theorem 1.3), in particular, can be used to derandomize the Goemans-Williamson random hyperplane rounding technique for semidefinite programming relaxations [GW95], and its applications in the design and analysis of approximation algorithms. We note that applications of the Goemans-Williamson rounding technique have been derandomized previously [MR99, EIO02, Siv02]. Our derandomization differs from these previous results in it being *oblivious to the instance solved*. In other words, whereas previous derandomization results used a large sample space of possible hyperplanes (and thus had to adapt the choice of hyperplane to the specific instance being solved), we construct a small sample space of hyperplanes, such that for any instance one of those hyperplanes is guaranteed to produce the correct outcome.² Henceforth we refer to such a derandomization as an *oblivious* derandomization. Our oblivious derandomization results in a faster and parallel derandomization of the Goemans-Williamson rounding technique, compared to previous derandomizations. For a more detailed explanation of the differences between the different approaches see the discussion at the end of Section 6.1.

²Because checking a solution can be done in polynomial time, trying all hyperplanes in the support of the sample space guarantees the correct outcome for every instance.

Proof technique. We begin with the methods used for the derandomized version of the Johnson Lindenstrauss Lemma. Using a variant of the construction of Indyk [Ind07] of an embedding of ℓ_2^n into ℓ_1^N , we first embed \mathbb{R}^n in a higher dimensional space \mathbb{R}^N in such a way that the norm of each vector is almost uniformly spread across many coordinates. We then produce samples of t coordinates of the image using known sampling techniques [Zuc97]. Each sample, properly scaled, gives a projection of \mathbb{R}^n onto \mathbb{R}^t (where $t \ll n$). Such a projection preserves L_2 distances with high probability.

We elaborate further about Indyk’s construction and our requirements of the embedding. We require that in any unit vector in the image of the embedding, at most a small, sub-constant, fraction of the coordinates have absolute value that is much larger than the average (i.e., $1/\sqrt{N}$). We also require that the total weight of these “bad” coordinates is negligible. With these properties, we are guaranteed an accurate estimation of the samples. In the original embedding by [Ind07], a (too large) constant fraction of the coordinates may be “bad”. To overcome this we use the following observation: We view a vector in the image (i.e., in \mathbb{R}^N), not as an N dimensional vector over \mathbb{R} but as an N/r dimensional vector over \mathbb{R}^r for some integer r . That is, as a “block vector”. Define the absolute value of a block (i.e., of a vector in \mathbb{R}^r) to be the L_2 norm of the vector. We show that in any unit vector in the image of the embedding, at most a sufficiently small fraction of the blocks have an absolute value that is much larger than the average (i.e., $1/\sqrt{N/r}$). Also, the total weight of these “bad blocks” is negligible. Now, instead of sampling a small number of indices from N we sample a small number of blocks. As the block size is much smaller than n , the dimension is still substantially reduced. The target dimension will not be as small as in the randomized constructions, however, it will be sufficiently small so that standard methods, given in e.g. [Siv02] or [CW09], can reduce the dimension further with negligible cost to the size of the sample space.

We now briefly discuss the construction of an ϵ -sample for spherical digons. The measure of a spherical digon is proportional to the angle between the two hyperplanes that bound it. The first step of the construction is to apply many *norm preserving* projections of \mathbb{S}^{n-1} onto a lower dimensional space \mathbb{R}^t . These projections also preserve angles approximately, with high probability over the choice of sample. Due to the low target dimension, we can produce in \mathbb{S}^{t-1} a poly(n)-size ϵ -sample for spherical digons using a pseudo-random generator for space bounded computation [Nis92]. We then use the adjoint operators of our projections to lift this ϵ -sample, for low dimensional spherical digons, back to \mathbb{R}^n . Each low dimensional sample point is lifted many times, once for each of the constructed projections of \mathbb{S}^{n-1} into \mathbb{R}^t . The ϵ -sample for spherical digons in \mathbb{S}^{n-1} is composed essentially of the entire collection of lifted low dimensional sample points. The construction of a sample space for spherical caps is very similar. In this case the important observation is that the volume of a spherical cap is determined only by the dimension n and its distance from the origin.

In order to understand the connection between Theorem 1.3 and the Goemans-Williamson approximation algorithm for MAX-CUT [GW95], and other similar algorithms, such as the Karger-Motwani-Sudan approximation algorithm for coloring graphs [KMS98], we briefly review their algorithm. The Goemans-Williamson algorithm first solves a semidefinite programming relaxation of MAX-CUT, mapping the nodes of an n -node input graph to points on the unit sphere \mathbb{S}^{n-1} , then the algorithm constructs a cut by choosing a vector $x \in \mathbb{S}^{n-1}$ uniformly at random and separating the mapped nodes by the hyperplane through the origin

which is perpendicular to x . If $u, v \in \mathbb{S}^{n-1}$ are the images of the endpoints of an edge of the input graph, then the set of vectors x that cause the edge to be cut is the union of two antipodal spherical digons (i.e., if x is in one of them, then $-x$ is in the other), hence the immediate connection to the constructed ϵ -sample.

Discussion. As described above, our construction used the adjoint operators of JL type projections in order to lift constructions from low dimensions to high dimensions. In contrast, previous constructions of pseudo-random generators for linear or degree d polynomial threshold functions, over the Boolean cube, used a k -wise independent sample space for k that is polynomial in $1/\epsilon^d$ [DGJ⁺10, MZ10, DKN10].³ Such constructions have seed length $\log(n)/\text{poly}(1/\epsilon^d)$ which is significantly longer than ours. In particular, such techniques will not be able to provide pseudo-random-generators with a short seed for higher degree polynomial threshold functions. One can hope that by completely derandomizing the JL lemma, such short seed pseudo-random-generators would follow using an approach similar to ours.

Organization. In Section 3 we prove Theorem 1.1. We then use it in Section 4 to give an ϵ -sample for linear threshold functions (Theorem 1.2). In Section 5 we construct an ϵ -sample for digons, thus proving Theorem 1.3. We give some applications of Theorem 1.3 in Section 6. Namely, we show how to derandomize the Goemans-Williamson algorithm and the graph coloring algorithm of [KMS98].

2 Preliminaries

For $n \in \mathbb{N}$ denote $[n] \triangleq \{1, \dots, n\}$. For $x \in \mathbb{R}^n$ and a subset $S = \{i_1, i_2, \dots, i_{|S|}\} \subseteq [n]$ define x_S as the restriction of x to the indices of S . That is, $x_S \triangleq (x_{i_1}, x_{i_2}, \dots, x_{i_{|S|}})$ where $i_1 < i_2 < \dots < i_{|S|}$. A unit vector $x \in \mathbb{R}^n$ is a vector satisfying $\|x\|_2 = 1$ (where $\|\cdot\|_2$ is the Euclidean norm). For non-zero $v, u \in \mathbb{R}^n$ define $\angle(v, u)$ to be the angle between them (the angle ranges between 0 and π). We write $a = b \pm c$ to indicate that $b - c \leq a \leq b + c$. ℓ_2^n denotes the Euclidian space of dimension n (\mathbb{R}^n equipped with the $\|\cdot\|_2$ norm) and ℓ_1^N denotes \mathbb{R}^N equipped with the $\|\cdot\|_1$ norm.

3 Derandomization of the JL Lemma

In this section we prove Theorem 1.1. Namely, we construct a set \mathcal{A} of polynomially many linear transformations from \mathbb{R}^n to \mathbb{R}^t such that any unit vector has its L_2 -norm preserved, up to additive distortion of ϵ , in at least $1 - \gamma$ of the linear transformations in \mathcal{A} . The parameter t is a function of ϵ, γ as in the JL lemma. Specifically, as in the known random constructions, $t = O(\log(\gamma^{-1})\epsilon^{-2})$. Our methods work for any $(\gamma\epsilon)^{-1} = \exp\left(O\left(\sqrt{\log(n)}\right)\right)$. The size of \mathcal{A} grows with ϵ, γ . However, when $(\epsilon\gamma)^{-1} = \exp\left(o\left(\sqrt{\log(n)}\right)\right)$, we get $|\mathcal{A}| = n^{1+o(1)}$.

³More accurately, [MZ10] use k -wise independent distributions only for degree 2 or higher polynomial threshold functions and for linear threshold functions, over the Boolean cube, they gave a construction with seed length similar to ours, using other methods.

Before we begin, we note that there exists a simpler derandomization of the JL Lemma which follows from the techniques of [AMS99]. The derandomization works for similar parameters, however, the size of the set constructed is $\Omega(n^4)$ even for constant ϵ, γ . While the exponent of n is usually not crucial when constructing pseudo-random-generators, it does have a significant effect on the running time of our derandomization of MAX-CUT, where it is beneficial to have a set \mathcal{A} of cardinality $n^{1+o(1)}$ as opposed to $\Omega(n^4)$ (see Section 6.1). We give the details of the simpler construction in Appendix B.

We begin with a formal definition of the norm-preserving property.

Definition 3.1. *A set \mathcal{A} of linear transformations from \mathbb{R}^n to \mathbb{R}^t is called (γ, ϵ) -norm preserving when for every unit vector $v \in \mathbb{S}^{n-1}$ it holds that*

$$\Pr_{A \in \mathcal{A}} [|\|Av\|_2^2 - 1| > \epsilon] < \gamma.$$

I.e., the norm of v remains the same up to a multiplicative factor of $1 \pm \epsilon$ with probability $\geq 1 - \gamma$.

We construct a (γ, ϵ) -norm preserving set \mathcal{A} in the following way: First, we embed \mathbb{R}^n in \mathbb{R}^N for some $N > n$. This embedding has the property that all the vectors in its image are ‘well spread’. Intuitively, this means that all the vectors have most of their entries within a certain factor from their average (i.e. around $1/\sqrt{N}$). The set \mathcal{A} is then composed out of various samples of subsets of the rows of the embedding matrix. We give a construction of the required embedding in Section 3.1 and discuss how to sample subsets of its rows in Section 3.2. Finally, we present the construction and analysis of the set \mathcal{A} in Section 3.3, where we also give the proof of Theorem 1.1.

3.1 Euclidean sections of ℓ_1^n

A part of our construction requires embedding \mathbb{R}^n into \mathbb{R}^N ($N \geq n$) such that any vector in the image will have its norm spread throughout its entries. We begin by formally defining the spreadness property of a vector.

Definition 3.2. *A vector $y \in \mathbb{R}^d$ is (α, η) -spread when for any $S \subseteq [d]$ of size $|S| \leq \alpha d$ it holds that $\|y_S\|_2 \leq \eta \|y\|_2$.*

We use a construction of an embedding of ℓ_2^n into ℓ_1^N by [Ind07] in order to obtain a linear operator whose image consists of well spread vectors. For some integers d, r such that $N = dr$, we consider vectors in \mathbb{R}^N as elements in $(\mathbb{R}^r)^d$. Namely, as vectors of d entries where each entry is a vector in \mathbb{R}^r . Addition of vectors in $(\mathbb{R}^r)^d$ is defined in the natural way: $(a_1, \dots, a_d) + (b_1, \dots, b_d) = (a_1 + b_1, \dots, a_d + b_d)$ (each a_i, b_i is an element of \mathbb{R}^r). In this section we prove the following theorem.

Theorem 3.3. *Let $n > 0$ be an integer and $\rho > 0$. There exists an explicit linear transformation $F : \mathbb{R}^n \rightarrow (\mathbb{R}^r)^d$ with the following properties: $d = n^{1+o(1)} \cdot \rho^{-O(\log \log(n))}$ and $r = O(\log \log(n)/\rho)$. For any $y = (y_1, \dots, y_d) = F(z)$, the vector $x = (\|y_1\|_2, \dots, \|y_d\|_2)$ is $(\rho, \sqrt{8\rho})$ -spread. In addition, $\|x\|_2 = \|z\|_2$.*

The embedding by [Ind07] is not shown to have the required properties. However, a sub-procedure in [Ind07] does achieve them.

Lemma 3.4 (Theorem 1.1 in [Ind07]). *Let $n = 2^{2k}$ where $k > 0$ is some integer⁴ and let $\rho > 0$. There exists an explicit linear transformation $F : \mathbb{R}^n \rightarrow (\mathbb{R}^r)^d$ with the following properties: $d = n^{1+o(1)} \cdot \rho^{-O(\log \log(n))}$ and $r = O(\log \log(n)/\rho)$. For any $y = (y_1, \dots, y_d) = F(z)$, the vector $x = (\|y_1\|_2, \dots, \|y_d\|_2)$ is such that $\|x\|_1 \geq (1-\rho)\sqrt{d}\|x\|_2$ and $\|x\|_2 = \|z\|_2$.*

The following lemma shows that a vector with a large L_1 norm, compared to its L_2 norm, is well spread. Theorem 3.3 is a direct consequence.

Lemma 3.5. *Let $d \in \mathbb{N}$, $\rho > 0$ and let $x \in \mathbb{R}^d$ be such that $\|x\|_1 \geq (1-\rho)\sqrt{d}\|x\|_2$. Then x is $(\rho, \sqrt{8\rho})$ -spread.*

Proof. Assume w.l.o.g. that $\|x\|_2 = 1$. Let $S \subseteq [d]$ be of size $|S| \leq \rho d$. Notice that $\|x_S\|_1 \leq \sqrt{|S|}\|x_S\|_2 \leq \sqrt{\rho d}\|x_S\|_2$. Now, for $\bar{S} \triangleq [d] \setminus S$,

$$\|x_{\bar{S}}\|_2 \geq \frac{\|x_{\bar{S}}\|_1}{\sqrt{d}} = \frac{\|x\|_1 - \|x_S\|_1}{\sqrt{d}} \geq (1-\rho)\|x\|_2 - \sqrt{\rho}\|x_S\|_2 = 1 - \rho - \sqrt{\rho}\|x_S\|_2.$$

Hence,

$$1 - \rho - \sqrt{\rho}\|x_S\|_2 \leq \sqrt{1 - \|x_S\|_2^2}.$$

Viewing this as a degree two polynomial in $\|x_S\|_2$ we get the equation

$$(1+\rho)\|x_S\|_2^2 + (-2(1-\rho)\sqrt{\rho})\|x_S\|_2 + (\rho^2 - 2\rho) \leq 0.$$

with standard analysis, we get the inequality $\|x_S\|_2^2 \leq 8\rho$. It works for 8. \square

3.2 Samplers

The previous section gave an embedding F that spreads the coordinates of any nonzero vector. We use this map in order to reduce the dimension while preserving the L_2 norm, by taking several different projections of F to subsets of the coordinates. In order to pick these subsets we use a combinatorial object called an averaging sampler, whose main property is that it can be thought of as a tool to estimate the expectation of any bounded function f using a small number of queries, that are independent of f . More accurately, averaging samplers for functions from $[d]$ to $[0, 1]$ compute a subset of $[d]$. They estimate the average of a function by its average on the subset. Clearly, a deterministic sampler would require $\Omega(d)$ queries to achieve a small error. However, if we allow the sampler to be randomized then the number of samples significantly drops. For more on averaging samplers see [Zuc97].

In [Zuc97], it is shown that the task of constructing an efficiently computable averaging sampler is essentially the same as constructing an efficiently computable *seeded extractor*. As we do not use these objects in any direct manner, we will not get into the details of their uses or construction. For more on extractors we refer the reader to [Sha02]. We require an extractor, given in [GUV09], which combined with the result of [Zuc97] gives the required sampler. For completeness, we give a more thorough explanation in Appendix C.

⁴Notice that for any integer n that is not a natural power of 4 we may initially embed \mathbb{R}^n in $\mathbb{R}^{n'}$ for an integer $n' < 4n$ that is a power of 4 and obtain essentially the same parameters.

Lemma 3.6 (Proposition 2.7 in [Zuc97] combined with Theorem 4.19 in [GUV09]). *Let d be a power of 2.⁵ Let $\epsilon > 0$ be a parameter. For any $1 > \xi > 2/\log(d)$ there exists an efficiently constructible family \mathcal{T} of subsets of $[d]$ with the following properties: Each set $T \in \mathcal{T}$ is of size $t = (\log(d)/\epsilon)^{O(\log(\xi^{-1}))}$. The number of sets in \mathcal{T} is $|\mathcal{T}| = d^{1+O(\xi)}$. For any function $f : [d] \rightarrow [0, 1]$,*

$$\Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}[f(i)] - \mathbb{E}_{j \in [d]}[f(j)]| > \epsilon] < \gamma$$

where $\gamma = d^{-\Omega(\xi)}$. Furthermore, for any $i, j \in [d]$, $\Pr_{T \in \mathcal{T}}[i \in T] = \Pr_{T \in \mathcal{T}}[j \in T]$.

We would like to use samplers in order to project F (the embedding from \mathbb{R}^n to \mathbb{R}^N where $N = dr$) into a subspace of \mathbb{R}^N (the samplers will choose the coordinates to project on). For this we shall define for every vector $x \in \mathbb{R}^d$ a function $f_x : [d] \rightarrow \mathbb{R}$ by $f_x(i) = d \cdot \|(Fx)_{\{(i-1)r+1, \dots, ir\}}\|_2^2$. By Theorem 3.3, $f_x(i)$ is usually at most 8 (that is, it is almost a function from $[d]$ to $[0, 8]$). However, it may obtain large values on some elements of $[d]$. It is not difficult to see that the expectation of f over $[d]$ is almost equal to its expectation over the points in which $f_x(i) \in [0, 8]$. Furthermore, the number of points in $[d]$ in which $f_x \notin [0, 8]$ is negligible. We show that due to these two properties, it is possible to evaluate $\mathbb{E}[f_x]$ by using an averaging sampler. To formalize the required property, we say that a function $f : [d] \rightarrow \mathbb{R}$ is η -bounded in a segment $I \subseteq \mathbb{R}$ when the following holds: $\Pr_{i \in [d]}[f(i) \in I] \geq 1 - \eta$ and $\mathbb{E}_{i \in [d]}[f(i)] = \mathbb{E}_{i \in [d]}[f(i)|f(i) \in I] \pm \eta$.⁶ Intuitively, in our case one should think of η as being much smaller than $1/t$.

Theorem 3.7. [*Sampler for η -bounded functions*] *Let d be an integer and $\epsilon, \eta > 0$ where $\eta < \epsilon/2$. Let $1 > \xi > 2/\log(d)$ and let $f : [d] \rightarrow \mathbb{R}$ be some η -bounded function in the segment $I = [0, 1]$. Let \mathcal{T} be the family defined in Lemma 3.6 for parameters $d, \epsilon' = \epsilon - \eta$ and ξ . Then,*

$$\Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}[f(i)] - \mathbb{E}_{j \in [d]}[f(j)]| > \epsilon] < d^{-\Omega(\xi)} + t\eta = d^{-\Omega(\xi)} + \eta (\log(d)/\epsilon)^{O(\log(\xi^{-1}))}.$$

Proof. Denote by μ the expectation of f over a uniform distribution on $[d]$. Define $S \subset [d]$ as the set of points in which $f(i) \notin I$. Since f is η -bounded in I we have that $\mu_{\bar{S}} \triangleq \mathbb{E}_{i \notin S}[f(i)] = \mu \pm \eta$. Let $g : [d] \rightarrow I$ be the following function: For all $i \notin S$, set $g(i) = f(i)$. For $i \in S$, set $g(i) = \mu_{\bar{S}}$. By the union bound we get that

$$\Pr_{T \in \mathcal{T}} [\exists i \in T \text{ s.t. } f(i) \neq g(i)] \leq t\eta$$

where $t = (\log(d)/(\epsilon - \eta))^{O(\log(\xi^{-1}))} = (\log(d)/\epsilon)^{O(\log(\xi^{-1}))}$ is the size of the sets T within the family \mathcal{T} (as in Lemma 3.6). Now, Lemma 3.6 and the fact that the expectation of g is $\mu_{\bar{S}}$ imply that

$$\Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}[g(i)] - \mu| > \epsilon] \leq \Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}[g(i)] - \mu_{\bar{S}}| > \epsilon - \eta] < d^{-\Omega(\xi)}.$$

⁵This restriction does not contradict the generality of the proof since the value of d is dictated by Lemma 3.4 that always gives d that is a power of 2.

⁶Unless mentioned otherwise, all expectations are computed with respect to the uniform probability.

Finally, by the union bound we obtain

$$\Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}[f(i)] - \mu| > \epsilon] \leq \Pr_{T \in \mathcal{T}} [|\mathbb{E}_{i \in T}g(i) - \mu| > \epsilon] + \Pr_{T \in \mathcal{T}} [\exists i \in T \text{ s.t. } f(i) \neq g(i)] < d^{-\Omega(\xi)} + \eta (\log(d)/\epsilon)^{O(\log(\xi^{-1}))}.$$

□

3.3 The Norm Preserving Set

We now describe the construction of a set of (ϵ, γ) -norm preserving transformations from \mathbb{R}^n to $\mathbb{R}^{O(\epsilon^{-2} \log(\gamma^{-1}))}$. We first reduce the dimension from n to $(\log(n)/(\epsilon\gamma))^{O(\log(\xi^{-1}))}$ for some $1 > \xi > 0$ which is sub-constant, yet sufficiently large so that the target dimension will not be too large. Then, by using standard methods, we further reduce the dimension to $O(\epsilon^{-2} \log(\gamma^{-1}))$ (the same target dimension as in the randomized constructions).

Let $N = dr$ and let $F : \mathbb{R}^n \rightarrow \mathbb{R}^{dr}$ be the linear transformation guaranteed by Theorem 3.3 with parameter $\rho = (\log(n)/(\epsilon\gamma))^{-c_1 \log(\xi^{-1})}$ for some sufficiently large constant c_1 and some $1 > \xi > 0$ that will be determined later. Let \mathcal{T} be the family of subsets of $[d]$ guaranteed by Theorem 3.7 w.r.t. the parameters d, ϵ and ξ (we will choose $1 > \xi > 2/\log(d)$ so applying the theorem will be possible). Denote by t' the cardinality of each $T \in \mathcal{T}$. For every $T \in \mathcal{T}$ define $A_T : \mathbb{R}^N \rightarrow \mathbb{R}^{t'r}$ as the projection to the indices of T , when N is considered as a set of d blocks. Specifically, for $T \subseteq [d]$, let $\hat{T} \subseteq [N]$ be $\hat{T} = \cup_{i \in T} \{(i-1)r + 1, \dots, ir\}$. Then $A_T(x) = x_{\hat{T}}$ (i.e., the projection of x to the indices of \hat{T}). The set \mathcal{A}_1 is defined as

$$\mathcal{A}_1 \triangleq \left\{ \sqrt{d/t'} \cdot A_T \cdot F \mid T \in \mathcal{T} \right\}.$$

The following lemma gives the main dimension reduction.

Lemma 3.8. *Assuming $1 > \xi > \max\{2, c_2 \log(\gamma^{-1})\}/\log(n)$ (and in particular that $\gamma > n^{-1/c_2}$) for sufficiently large constant c_2 , the set \mathcal{A}_1 defined above is (γ, ϵ) -norm preserving. Its cardinality is $|\mathcal{A}_1| = n^{1+O(\xi)+o(1)} \cdot (\log(n)/(\epsilon\gamma))^{O(\log(\xi^{-1}) \log \log(n))}$. The projections in \mathcal{A}_1 map \mathbb{R}^n to $\mathbb{R}^{(\log(n)/(\epsilon\gamma))^{O(\log(\xi^{-1}))}}$.*

Proof. The claim regarding $|\mathcal{A}_1|$ and the dimension of the projections follows directly from Theorem 3.3 and Lemma 3.6 (Lemma 3.6 can be applied since $1 > \xi > 2/\log(n) > 2/\log(d)$). Let $w \in \mathbb{R}^n$ be some fixed vector. Assume w.l.o.g. that $\|w\|_2 = 1$. Let $f : [d] \rightarrow \mathbb{R}$ be the function $f(i) \triangleq d \cdot \|(Fw)_{\{(i-1)r+1, \dots, ir\}}\|_2^2$. Notice that the expectation of f is equal to $\|F(w)\|_2^2 = \|w\|_2^2 = 1$. Theorem 3.3 shows that

$$\Pr_{i \in [d]} [f(i) \notin [0, 8]] < \rho, \quad \mathbb{E}_{i \in [d]} [f(i) | f(i) \in [0, 8]] \geq 1 - 8\rho = \mathbb{E}[f(i)] - 8\rho.$$

In other words, the function $f/8$ is ρ -bounded in $[0, 1]$. We also have that for any $T \subseteq [d]$ of size $|T| = t'$,

$$\left\| \sqrt{d/t'} \cdot A_T F(w) \right\|_2^2 = \frac{1}{|T|} \sum_{i \in T} f(i) = \mathbb{E}_{i \in T}[f(i)].$$

By Theorem 3.7, applied to the function $f/8$,

$$\begin{aligned} \Pr_{T \in \mathcal{T}} \left[\left| \left\| \sqrt{d/t'} \cdot A_T F(w) \right\|_2^2 - 1 \right| > \epsilon \right] &= \Pr_{T \in \mathcal{T}} \left[\left| \mathbb{E}_{i \in T} [f(i)] - \mathbb{E}_{i \in [d]} [f(i)] \right| > \epsilon \right] \\ &< d^{-\Omega(\xi)} + \rho t' = d^{-\Omega(\xi)} + \rho (\log(d)/\epsilon)^{O(\log(\xi^{-1}))} = n^{-\Omega(\xi)} + \rho (\log(n)/\epsilon)^{O(\log(\xi^{-1}))}, \end{aligned}$$

where we used the facts that $n \leq d \leq n^{1+o(1)}$ and $t' = (\log(d)/\epsilon)^{O(\log(\xi^{-1}))} = (\log(n)/\epsilon)^{O(\log(\xi^{-1}))}$. Notice that as $\xi \geq c_2 \log(\gamma^{-1})/\log(n)$ we get that $n^{-\Omega(\xi)} = (\gamma^{c_2})^{\Omega(1)}$. Therefore, if both c_2 and c_1 (the constant in the exponent of ρ) are sufficiently large, we get that

$$\begin{aligned} \Pr_{T \in \mathcal{T}} \left[\left| \left\| \sqrt{d/t} \cdot A_T F(w) \right\|_2^2 - 1 \right| > \epsilon \right] &< n^{-\Omega(\xi)} + \rho (\log(n)/\epsilon)^{O(\log(\xi^{-1}))} = \\ &(\gamma^{c_2})^{\Omega(1)} + (\log(n)/(\epsilon\gamma))^{-c_1 \log(\xi^{-1})} \cdot (\log(n)/\epsilon)^{O(\log(\xi^{-1}))} < \gamma/2 + \gamma/2 = \gamma. \end{aligned}$$

□

Denote by t'' the target dimension of the projections in \mathcal{A}_1 . We further reduce the dimension via the following lemma of Clarkson and Woodruff.

Lemma 3.9 (Theorem 2.2 in [CW09]). *Let $1 > \epsilon, \gamma > 0$ and let n be an integer. There exist universal constants c_3, c_4 for which the following holds: Let \mathcal{A}_2 be the sample space of sign matrices of dimension $t \times n$ whose signs are chosen from an s -wise independent distribution, where $s = c_3 \log(\gamma^{-1})$ and $t = c_4 \log(\gamma^{-1})\epsilon^{-2}$. Then \mathcal{A}_2 is a (γ, ϵ) -norm preserving set.*

In other words, if we think of each matrix as a vector of length tn , then when sampling the matrix from an s -wise independent distribution we get that it is norm preserving with high probability.

We note that for all s, m , efficient deterministic constructions of s -wise independent sample spaces over $\{-1, 1\}^m$, of size m^s exist⁷. Hence, when the above lemma is applied to reduce the dimension from t'' to t , the size of \mathcal{A}_2 is at most

$$(t \cdot t'')^{O(\log(\gamma^{-1}))} = \exp \left(O \left(\log(\xi^{-1}) \log(\gamma^{-1}) \log \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right)$$

The following theorem immediately follows.

Theorem 3.10. *Let n be an integer and let $1 > \epsilon, \gamma > 0$. Let $1 > \xi > \max\{2, c \log(\gamma^{-1})\}/\log(n)$ for some universal constant c (we assume that $\gamma > n^{-1/c}$ so such a value of ξ exists). The set*

$$\mathcal{A} = \{A = A_2 \cdot A_1 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

is $(2\gamma, 3\epsilon)$ -norm preserving. It is of cardinality

$$|\mathcal{A}| = n^{1+O(\xi)+o(1)} \cdot \exp \left(O \left(\log(\xi^{-1}) \log^2 \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right).$$

The projections in \mathcal{A} are from \mathbb{R}^n to $\mathbb{R}^{O(\log(\gamma^{-1})\epsilon^{-2})}$.

⁷We elaborate further on s -wise independent sample spaces in Appendix B as a part of the simple construction of a norm-preserving set (that is based on the L_2 approximation algorithm of [AMS99]).

Proof. We first show that \mathcal{A} is norm preserving. Consider a matrix A chosen uniformly at random from \mathcal{A} , by picking A_1 from \mathcal{A}_1 and A_2 from \mathcal{A}_2 uniformly at random and independently. Let $v \in \mathbb{R}^n$ be some fixed unit vector. The probability that both A_1 preserves the norm of v and A_2 preserves the norm of $A_1 v$ is at least $1 - 2\gamma$ (by the union bound). Hence, with probability at least $1 - 2\gamma$

$$\|A_2 A_1 v\|_2 = (1 \pm \epsilon) \|A_1 v\|_2 = (1 \pm \epsilon)^2 \|v\|_2 = 1 \pm 3\epsilon.$$

To calculate the size of \mathcal{A} , simply notice that $|\mathcal{A}| = |\mathcal{A}_1| |\mathcal{A}_2|$.

$$\begin{aligned} |\mathcal{A}_2| &= \exp \left(O \left(\log(\xi^{-1}) \log(\gamma^{-1}) \log \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right) = \\ &\exp \left(O \left(\log(\xi^{-1}) \cdot \log^2 \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right). \end{aligned}$$

Similarly,

$$\begin{aligned} |\mathcal{A}_1| &= n^{1+O(\xi)+o(1)} \cdot (\log(n)/(\epsilon\gamma))^{O(\log(\xi^{-1}) \log \log n)} = \\ &n^{1+O(\xi)+o(1)} \cdot \exp \left(O \left(\log(\xi^{-1}) \log \log(n) \cdot \log \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right) = \\ &n^{1+O(\xi)+o(1)} \exp \left(O \left(\log(\xi^{-1}) \cdot \log^2 \left(\frac{\log(n)}{\epsilon\gamma} \right) \right) \right). \end{aligned}$$

The claim regarding $|\mathcal{A}|$ easily follows. □

Theorem 1.1 stems directly from Theorem 3.10, by picking

$$\xi = \max \left\{ \frac{3}{\log(n)}, \frac{c \log(\gamma^{-1}) + 1}{\log(n)}, \exp \left(-\sqrt{\log(n) / \log^2 \left(\frac{\log(n)}{\gamma\epsilon} \right)} \right) \right\}.$$

The following corollary will be useful in the next sections.

Corollary 3.11. *Let $n \in \mathbb{N}$, and let $0 < \delta < 1$. There exists an explicit construction of a set \mathcal{A} of transformation from \mathbb{R}^n to \mathbb{R}^t such that $t = O(\delta^{-2} \log(\delta^{-1}))$ and \mathcal{A} is a (δ, δ) -norm preserving set of size $|\mathcal{A}| = n^{1+o(1)} \exp(O(\log^2(\delta^{-1})))$.*

4 Fooling Linear Threshold Functions

A linear threshold function (LTF) is a function $f : \mathbb{S}^{n-1} \rightarrow \{-1, 1\}$ of the following form: $f_{w,\theta}(x) = \text{sign}(\langle w, x \rangle - \theta)$ where $w \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ and $\text{sign}(0)$ is defined as 1. Functions of this form are indicator functions of spherical caps. In this section we construct a sample space for spherical caps. We note that the volume of a spherical cap relies only on the dimension n and the threshold θ . As a result we get that a norm-preserving set can reduce the problem to that of finding a sample space for spherical caps of dimension $t \ll n$. We then use Nisan's pseudo-random generator for log-space machines (see Appendix A) to construct a sample

space for spherical caps of dimension t . Informally, we construct a set $Q \subset \mathbb{R}^n$ for which it holds that

$$\mathbb{E}_{x \in Q} [f_{w,\theta}(x)] \approx \mathbb{E}_{y \in \mathbb{S}^{t-1}} [f_{w',\theta,\sqrt{n/t}}(y)] \approx \mathbb{E}_{x \in \mathbb{S}^{n-1}} [f_{w,\theta}(x)]$$

where w' is some unit vector in \mathbb{S}^{t-1} . Let ϵ be a parameter. We now explain how to construct Q , an ϵ -sample for spherical caps of dimension n .

Construction 1. Let \mathcal{A} be a (δ, δ) -norm preserving set of linear transformations from \mathbb{R}^n to \mathbb{R}^t , where $\delta = c\epsilon$ for some sufficiently small constant c and $t = \epsilon^{-C}$ for some sufficiently large constant C . Let $Q' \subseteq \mathbb{R}^t$ be a δ -sample for LTFs in t dimensions. $Q_{\mathcal{A}}$ is defined as

$$Q_{\mathcal{A}} \triangleq \left\{ \sqrt{t/n} \cdot A^T x' \mid x' \in Q', A \in \mathcal{A} \right\}.$$

By Corollary 3.11, a (δ, δ) -norm preserving set exists as long as the ratio between C and c is sufficiently large. Its size is $|\mathcal{A}| = n^{1+o(1)} \exp(O(\log^2(\epsilon^{-1})))$. In Appendix A we give a construction of a sample space for spherical caps in low dimensions. The main tool in its construction is a Pseudo Random Generator for bounded space machines by Nisan. Specifically, we prove the following lemma (it is a direct corollary of Theorem A.15).

Lemma 4.1. For any $t \in \mathbb{N}$ there exists an explicit construction of a set $Q' \subseteq \mathbb{R}^t$ of size $|Q'| = \exp(O(\log^2(t)))$ that is an (weak) ϵ -sample for spherical caps w.r.t. the uniform distribution over \mathbb{S}^{t-1} with $\epsilon = O(1/t)$.

Using Lemma 4.1 in Construction 1 we get $Q_{\mathcal{A}}$ of size $|Q_{\mathcal{A}}| = n^{1+o(1)} \exp(O(\log^2(\epsilon^{-1})))$. Notice that the elements of $Q_{\mathcal{A}}$ are not necessarily unit vectors. As a result we refer to $Q_{\mathcal{A}}$ as a weak ϵ -sample since, as we shall see, it still has the property that

$$\mathbb{E}_{y \in \mathbb{S}^{n-1}} [f_{w,\theta}(y)] = \mathbb{E}_{x \in Q_{\mathcal{A}}} [f_{w,\theta}(x)] \pm \epsilon.$$

Our main result of this section is given in the next theorem.

Theorem 4.2. Let f be a LTF. It holds that

$$|\mathbb{E}_{x \in \mathbb{S}^{n-1}} [f(x)] - \mathbb{E}_{x \in Q_{\mathcal{A}}} [f(x)]| = O(\delta).$$

Before giving the analysis of $Q_{\mathcal{A}}$ we state some known facts regarding $\mathbb{E}_{x \in \mathbb{S}^{n-1}} [f_{w,\theta}(x)]$, basically showing a connection between a fixed projection of a random unit vector and the standard normal distribution $\mathcal{N}(0, 1)$. Denote by $\Phi(z)$ the probability that a random variable from the normal gaussian distribution takes a value larger than z . That is $\Phi(z) \triangleq \Pr_{Y \sim \mathcal{N}(0,1)} [Y > z]$. The following two lemmas are well known. The first relates $\mathcal{N}(0, 1)$ to the random variable $\langle x, w \rangle$ where w is some constant unit vector and x is chosen uniformly from the unit sphere. The proof of the lemma can be found in e.g. [DF87]. The second is a technical lemma regarding the c.d.f. Φ .

Lemma 4.3. [Special case of Equation 1 in [DF87]] Let d be an integer and $w \in \mathbb{S}^{d-1}$ be some fixed unit vector. For any $z \in \mathbb{R}$ it holds that

$$\Pr_{x \in \mathbb{S}^{d-1}} \left[\langle x, w \rangle > z/\sqrt{d} \right] = \Phi(z) \pm O(1/d).$$

Lemma 4.4. *Let $z > 0$ and $0 < \delta < 1/4$. Then*

$$\Phi(z \cdot (1 \pm \delta)) = \Phi(z) \pm O(\delta).$$

Proof. Let $z' = z(1 \pm \delta)$.

$$\begin{aligned} |\Phi(z') - \Phi(z)| &= \left| \frac{1}{\sqrt{2\pi}} \int_z^{z'} \exp(-\tau^2/2) d\tau \right| < \frac{1}{\sqrt{2\pi}} \cdot z\delta \cdot \exp(-z^2(1 - \delta)^2/2) = \\ &O(\delta \cdot z \cdot \exp(-z^2/4)) = O(\delta). \end{aligned}$$

□

We can now prove Theorem 4.2.

Proof. Let $w \in \mathbb{S}^{n-1}$ and $z \in \mathbb{R}$ be such that $f(x) = \text{sign}(\langle x, w \rangle - z/\sqrt{n})$. For simplicity we assume w.l.o.g. that $z \geq 0$.⁸ By Lemma 4.3,

$$\Pr_{x \in \mathbb{S}^{n-1}} [\langle x, w \rangle > z/\sqrt{n}] = \Phi(z) \pm O(1/n) = \Pr_{x' \in \mathbb{S}^{t-1}} [\langle x', w \rangle > z/\sqrt{t}] \pm O(1/t).$$

Let $\hat{\mathcal{A}} \subset \mathcal{A}$ be the set of all $A \in \mathcal{A}$ such that $\|Aw\|_2 = 1 \pm \delta$. For any $A \in \hat{\mathcal{A}}$ we have

$$\Pr_{x' \in \mathbb{S}^{t-1}} [\langle x', Aw \rangle > z/\sqrt{t}] = \Pr_{x' \in \mathbb{S}^{t-1}} \left[\langle x', w' \rangle > \frac{z}{(1 \pm \delta)\sqrt{t}} \right] \quad (1)$$

where w' is some t -dimensional unit vector. Observe that

$$\begin{aligned} \Pr_{x' \in \mathbb{S}^{t-1}} \left[\langle x', w' \rangle > \frac{z}{(1 \pm \delta)\sqrt{t}} \right] &\stackrel{(1)}{=} \Phi(z/(1 \pm \delta)) \pm O(1/t) \stackrel{(2)}{=} \\ &\Phi(z) \pm O(\delta) \stackrel{(3)}{=} \Pr_{x \in \mathbb{S}^{n-1}} [\langle x, w \rangle > z/\sqrt{n}] \pm O(\delta). \end{aligned} \quad (2)$$

Equalities (1) and (3) stem from Lemma 4.3 and the fact that $\delta > 1/t$. Equality (2) is implied by Lemma 4.4. Calculating we get

$$\begin{aligned} \Pr_{x \in Q_{\mathcal{A}}} [\langle x, w \rangle > z/\sqrt{n}] &\stackrel{(1)}{=} \Pr_{x' \in Q', A \in \mathcal{A}} [\langle x', Aw \rangle > z/\sqrt{t}] \stackrel{(2)}{=} \\ \Pr_{x' \in Q', A \in \hat{\mathcal{A}}} [\langle x', Aw \rangle > z/\sqrt{t}] \pm O(\delta) &\stackrel{(3)}{=} \Pr_{x' \in \mathbb{S}^{t-1}, A \in \hat{\mathcal{A}}} [\langle x', Aw \rangle > z/\sqrt{t}] \pm O(\delta) \stackrel{(4)}{=} \\ \Pr_{x \in \mathbb{S}^{n-1}} [\langle x, w \rangle > z/\sqrt{n}] \pm O(\delta), \end{aligned}$$

where equality (1) follows from the definition of $Q_{\mathcal{A}}$, equality (2) holds since $|\hat{\mathcal{A}}| \geq (1 - O(\delta))|\mathcal{A}|$, equality (3) stems from Q' being a δ -sample for spherical caps in \mathbb{S}^{t-1} and equality (4) is implied by Equations (1) and (2). This proves the claim. □

Corollary 4.5. *Let $\epsilon > 0$ and let n be an integer. There exists an explicit weak ϵ -sample $Q_{\mathcal{A}}$ for spherical caps in \mathbb{S}^{n-1} of cardinality $|Q_{\mathcal{A}}| = \exp(O(\log^2(1/\epsilon))) n^{1+o(1)}$.*

Proof. By using the result of Lemma 4.1 in the construction of $Q_{\mathcal{A}}$ (according to Construction 1), we get by Theorem 4.2 that $Q_{\mathcal{A}}$ is an (weak) ϵ -sample for spherical caps. Its size is $|Q_{\mathcal{A}}| = |\mathcal{A}| \cdot |Q'| = n^{1+o(1)} \exp(O(\log^2(\epsilon^{-1})))$. □

⁸If $z < 0$, we work with $-f$ since $-f(x) = \text{sign}(\langle x, -w \rangle - \frac{-z}{\sqrt{n}})$.

5 An ϵ -Sample for Digons

In this section we prove Theorem 1.3. Namely, we present a method of constructing an ϵ -sample for digons. Recall that digons are characterized by functions of the following form:

$$f_{v,u}(x) \stackrel{\Delta}{=} \text{sign}(\langle v, x \rangle) \cdot \text{sign}(\langle u, x \rangle) \quad (3)$$

where $v, u, x \in \mathbb{S}^{n-1}$. In [GW95], it was shown that the expression $\mathbb{E}_{x \in \mathbb{S}^{n-1}} [f_{v,u}(x)]$ relies only on the angle between v and u . Using this observation we construct the ϵ -sample via the following process: We prove that a norm-preserving set also preserves the angle between any two given vectors (w.h.p.). This leads to a reduction from the problem of constructing an ϵ -sample for digons in \mathbb{S}^{n-1} to the problem in \mathbb{S}^{t-1} for some $t \ll n$. To that end, as in the previous section, we use Nisan's PRG (see Appendix A) for bounded space machines to obtain a sample space for digons.

5.1 Norm preserving implies angle preserving

In this section we prove that a set of linear transformations that is norm preserving is also angle preserving.

Definition 5.1. *A set \mathcal{A} of linear transformations from \mathbb{R}^n to \mathbb{R}^t is called (γ, δ) -angle preserving when for any fixed pair of unit vectors $v, u \in \mathbb{S}^{n-1}$ it holds that*

$$\Pr_{A \in \mathcal{A}} [|\angle(v, u) - \angle(Av, Au)| > \delta] < \gamma.$$

For simplicity, we discuss only the case where \mathcal{A} is (δ, δ) -norm preserving. We start by proving that for two unit vectors v, u it holds that $\cos(\angle(v, u)) = |\langle v, u \rangle|$ is roughly equal to $|\langle Av, Au \rangle|$ which is roughly equal to $\cos(\angle(Av, Au))$.

Lemma 5.2. *Let $v, u \in \mathbb{R}^n$ be unit vectors. Let \mathcal{A} be a (δ, δ) -norm preserving set of linear transformations. Then for a random $A \in \mathcal{A}$, we have that with probability at least $1 - 3\delta$*

$$|\langle Av, Au \rangle - \langle v, u \rangle| \leq 3\delta.$$

Proof. Due to the norm-preserving property of \mathcal{A} , we have, by the union bound, that with probability at least $1 - 3\delta$

$$|\|Av\|_2^2 - 1| < \delta, \quad |\|Au\|_2^2 - 1| < \delta, \quad |\|Av - Au\|_2^2 - \|v - u\|_2^2| < \delta \|v - u\|_2^2 \leq 4\delta.$$

It follows that $\|v - u\|_2^2 = \langle v - u, v - u \rangle = \langle v, v \rangle - 2\langle v, u \rangle + \langle u, u \rangle = 2(1 - \langle v, u \rangle)$ and

$$|\|Av - Au\|_2^2 - 2(1 - \langle Av, Au \rangle)| = |\langle Av - Au, Av - Au \rangle - 2(1 - \langle Av, Au \rangle)| =$$

$$|\langle Av, Av \rangle - 2\langle Av, Au \rangle + \langle Au, Au \rangle - 2 + 2\langle Av, Au \rangle| \leq |\|Av\|_2^2 - 1| + |\|Au\|_2^2 - 1| \leq 2\delta.$$

Hence,

$$|\langle Av, Au \rangle - \langle v, u \rangle| \leq \delta + \frac{|\|Av - Au\|_2^2 - \|v - u\|_2^2|}{2} \leq 3\delta.$$

□

Lemma 5.3. *Let $v, u \in \mathbb{R}^n$ be unit vectors. There exists some universal constant δ_0 such that for $\delta < \delta_0$ the following holds: Let \mathcal{A} be a (δ, δ) -norm preserving set. Then with probability of at least $1 - 3\delta$ we have*

$$|\angle(v, u) - \angle(Av, Au)| \leq 7\sqrt{\delta}.$$

Proof. It suffices to show that if the norms of $v, u, v - u$ were all preserved (up to a $1 \pm \delta$ multiplicative factor) then the claim holds. Define for brevity θ as the angle between v , and u and with θ' the angle between Av and Au . Then

$$\cos(\theta') = \frac{|\langle Av, Au \rangle|}{\|Av\|_2 \|Au\|_2}, \quad \cos(\theta) = |\langle v, u \rangle|$$

and by the previous lemma,

$$\begin{aligned} |\cos(\theta') - \cos(\theta)| &= \left| \frac{|\langle Av, Au \rangle|}{\|Av\|_2 \|Au\|_2} - |\langle v, u \rangle| \right| \leq \\ &\left| \frac{|\langle Av, Au \rangle|}{\|Av\|_2 \|Au\|_2} - |\langle Av, Au \rangle| \right| + |\langle Av, Au \rangle - \langle v, u \rangle| \leq \\ \|Av\|_2 \cdot \|Au\|_2 \left| \left(\frac{1}{\|Av\|_2 \|Au\|_2} - 1 \right) \right| + 3\delta &\leq (1 + \delta)^2 ((1 - \delta)^{-2} - 1) + 3\delta < 6\delta. \end{aligned} \quad (4)$$

The last inequality holds for sufficiently small δ .

We have two cases: In the first case we assume that $|\theta| \leq 2\sqrt{\delta}$ or $|\theta| - \pi \leq 2\sqrt{\delta}$. By symmetry, assume w.l.o.g. that $|\theta| \leq 2\sqrt{\delta}$. Then $\cos(\theta) \geq 1 - \frac{\theta^2}{2} \geq 1 - 2\delta$ (by Taylor expansion) and thus $\cos(\theta') \geq 1 - 8\delta$. As $1 - 8\delta \leq \cos(\theta') \leq 1 - \frac{\theta'^2}{2} + \frac{\theta'^4}{24}$ we get that (for small enough δ_0) $|\theta'| < 5\sqrt{\delta}$ and so $|\theta - \theta'| < 7\sqrt{\delta}$.

In the second case, $2\sqrt{\delta} < \theta < \pi - 2\sqrt{\delta}$. In particular, $|\cos(\theta)| \leq 1 - 2\delta + 2\delta^2/3$. We evaluate θ' as $\arccos(\cos(\theta'))$ via the Taylor expansion of $\arccos(\cdot)$ around the point $\cos(\theta)$:

$$|\theta - \theta'| < \frac{|\cos(\theta) - \cos(\theta')|}{\sqrt{1 - \cos^2(\theta)}} + O\left(|\cos(\theta) - \cos(\theta')|^2 \cdot \frac{\sin(\theta) \cos(\theta)}{(1 - \cos^2(\theta))^{1.5}}\right) \stackrel{(1)}{<}$$

$$6\sqrt{\delta} + O(\delta^2 \cdot \sin(\theta)^{-2} \cos(\theta)) \stackrel{(2)}{=} 6\sqrt{\delta} + O(\delta) \stackrel{(3)}{<} 7\sqrt{\delta}.$$

Inequality (1) follows from Equation(4) and the upper bound on $\cos(\theta)$ (for small enough δ_0).

Equality (2) holds since for $2\sqrt{\delta} < \theta < \pi - 2\sqrt{\delta}$ we have that $\sin(\theta) \geq \sin(2\sqrt{\delta}) = \Omega(\sqrt{\delta})$.

Inequality (3) holds for sufficiently small δ . \square

Corollary 5.4. *There exist universal constants $\delta_0 > 0, c > 0$ such that for any $\delta \leq \delta_0$, if \mathcal{A} is $(c\delta^2, c\delta^2)$ -norm preserving, then it is also (δ, δ) -angle preserving.*

5.2 The Construction

Let $\epsilon > 0$ be a parameter. We shall construct an ϵ -sample for digons over the unit sphere in \mathbb{S}^{n-1} .

Construction 2. Let $\delta = c\epsilon$ and $t = \epsilon^{-C}$ for some sufficiently small constant c and sufficiently large constant C . Let \mathcal{A} be some $(c'\delta^2, c'\delta^2)$ -norm preserving set of transformations from \mathbb{R}^n to \mathbb{R}^t (where c' is the same constant defined in Corollary 5.4). Let $P' \subseteq \mathbb{R}^t$ be a δ -sample for digons over \mathbb{S}^{t-1} . P is defined as⁹

$$P = \left\{ \frac{A^T x'}{\|A^T x'\|_2} \mid x' \in P', A \in \mathcal{A} \right\}.$$

Note that P may be a multi-set.

Assuming the ratio between c and C is sufficiently large, Corollary 3.11 indicates that there exists an explicit $(c'\delta^2, c'\delta^2)$ -norm preserving set of size $|\mathcal{A}| = n^{1+o(1)} \exp(O(\log^2(\epsilon^{-1})))$. Due to our choice of c' , by Corollary 5.4 we have that \mathcal{A} is also (δ, δ) -angle preserving. Similarly to the proof in Section 4, the sample P' for digons in low dimension relies on the Pseudo Random Generator for bounded space machines of Nisan. Specifically, we need the following lemma that we prove in Appendix A (specifically, it will be an immediate corollary of Theorem A.2).

Lemma 5.5. For any $t \in \mathbb{N}$ there exists an explicit construction of a set $P' \subseteq \mathbb{R}^t$ of size $|P'| = \exp(O(\log^2(t)))$ that is an ϵ -sample for digons w.r.t. the uniform distribution over \mathbb{S}^{t-1} with $\epsilon = O(1/t)$.

Theorem 1.3 is implied by the next theorem.

Theorem 5.6. The set $P \subseteq \mathbb{S}^{n-1}$ has size $|P| = n^{1+o(1)} \exp(O(\log^2(\epsilon^{-1})))$ and is an ϵ -sample for digons.

Proof. The claim regarding $|P|$ holds since $|P| = |\mathcal{A}| |P'|$. Notice that for any vector x , any non-zero scalar λ and any digon D , $x \in D$ iff $\lambda x \in D$. Hence, we may analyze P as if it were the set $\{A^T x' \mid x' \in P', A \in \mathcal{A}\}$.

We begin the proof with a lemma showing that the value of the expression $\mathbb{E}_{x \in \mathbb{S}^{n-1}} [f_{v,u}(x)]$ depends only on the angle between v and u .

Lemma 5.7. [Lemma 2.2 of [GW95]] $\mathbb{E}_{x \in \mathbb{S}^{n-1}} [f_{v,u}(x)] = 1 - 2\angle(v, u)/\pi$.

Let $v, u \in \mathbb{S}^{n-1}$ be two fixed vectors. First, We define $\hat{\mathcal{A}}$ as the set of all linear transformation in \mathcal{A} that preserve the angle between the vectors up to a $\pm\delta$ additive factor and additionally satisfy that $Av, Au \neq 0$. Since \mathcal{A} is a $(c'\delta^2, c'\delta^2)$ -norm preserving set, Corollary 5.4 implies that it is also (δ, δ) -angle preserving. Hence, it follows by union bound that $|\hat{\mathcal{A}}| \geq (1 - O(\delta))|\mathcal{A}|$. Therefore,

$$\mathbb{E}_{x \in P} [\text{sign}(\langle x, v \rangle) \cdot \text{sign}(\langle x, u \rangle)] = \mathbb{E}_{x' \in P', A \in \hat{\mathcal{A}}} [\text{sign}(\langle A^T x', v \rangle) \cdot \text{sign}(\langle A^T x', u \rangle)] \pm O(\delta).$$

⁹Note that there may exist $x' \in P'$ and $A \in \mathcal{A}$ such that $A^T x' = 0$. This technical matter can be dealt with by omitting those pairs. We elaborate further on this issue at the end of the section.

We continue with a series of equations leading to the required result,

$$\begin{aligned} \mathbb{E}_{x' \in P', A \in \hat{\mathcal{A}}} [\text{sign}(\langle A^T x', v \rangle) \cdot \text{sign}(\langle A^T x', u \rangle)] &= \mathbb{E}_{x' \in P', A \in \hat{\mathcal{A}}} [\text{sign}(\langle x', Av \rangle) \cdot \text{sign}(\langle x', Au \rangle)] \stackrel{(1)}{=} \\ &\mathbb{E}_{A \in \hat{\mathcal{A}}, y \in \mathbb{S}^{t-1}} [\text{sign}(\langle y, Av \rangle) \cdot \text{sign}(\langle y, Au \rangle)] \pm \delta \stackrel{(2)}{=} \mathbb{E}_{A \in \hat{\mathcal{A}}} \left[1 - 2 \frac{\angle(Av, Au)}{\pi} \right] \pm \delta \stackrel{(3)}{=} \\ &1 - \frac{2\angle(v, u)}{\pi} \pm 2\delta \stackrel{(4)}{=} \mathbb{E}_{x \in \mathbb{S}^{n-1}} [\text{sign}(\langle x, v \rangle) \cdot \text{sign}(\langle x, u \rangle)] \pm 2\delta. \end{aligned}$$

Equality (1) stems from P' being a δ -sample for t dimensional digons and from $Av, Au \neq 0$. Equalities (2) and (4) follow from Lemma 5.7 and equality (3) holds due to the definition of $\hat{\mathcal{A}}$. By combining with the previous equation and recalling that $\delta = c\epsilon$ where c is some sufficiently small constant, we get the required result.

We end the proof by dealing with the case of pairs (A, x') such that $A^T x' = 0$. Consider a digon determined by v, u where $u = -v$. By the calculations above we get that

$$\begin{aligned} \mathbb{E}_{x \in P} [\text{sign}(\langle x, v \rangle) \cdot \text{sign}(\langle x, -v \rangle)] &= \\ \mathbb{E}_{x \in \mathbb{S}^{n-1}} [\text{sign}(\langle x, v \rangle) \cdot \text{sign}(\langle x, -v \rangle)] \pm O(\delta) &\leq -1 + O(\delta). \end{aligned}$$

Hence, at most an $O(\delta)$ fraction of vectors in P are equal to the zero vector (recall that P is a multi-set so this is not a trivial statement). It follows that these vectors may be omitted without changing the conclusion. \square

6 Applications

In this section we give two applications of the ϵ -sample for digons that was constructed in Section 5. Specifically, we use the ϵ -sample to derandomize rounding procedures of solutions of semi definite programs (SDP for short). For example, in the famous Goemans-Williamson algorithm, the rounding scheme of the SDP solution is done by picking a random hyperplane and mapping the solution vectors to $\{1, -1\}$ according to the side of the hyperplane they belong to. It is not hard to show that the probability that two vectors will map to different values depends only on the angle between them. In fact, a hyperplane will separate the vectors if and only if $\text{sign}(\langle v, x \rangle) \cdot \text{sign}(\langle u, x \rangle) = -1$, where x is perpendicular to the hyperplane. Hence, in order to choose hyperplanes that appear random to such a process we need an ϵ -sample for digons. Another application for derandomizing the coloring algorithm of [KMS98] appears in Section 6.2.

6.1 Deterministic approximation of Max-Cut

Max-Cut is the following problem: Given a graph $G = (V, E)$, we seek a subset $S \subseteq V$ of vertices that maximizes the number of edges from it to $V \setminus S$. Namely, $\text{Max-Cut}(G) = \max_S E(S, V \setminus S)$. Goemans and Williamson [GW95] gave a randomized approximation

algorithm for the max-cut problem using semi-definite programming (SDP for short) that we now describe. First notice that max-cut can be solved by the following integer program:

$$\text{Maximize } \frac{1}{2} \sum_{1 \leq i < j \leq |V|} w_{i,j}(1 - v_i \cdot v_j) \quad \text{subject to } \forall i \ v_i \in \{-1, 1\}. \quad (5)$$

The following is an SDP relaxation of the integer program.

$$\text{Maximize } \frac{1}{2} \sum_{1 \leq i < j \leq |V|} w_{i,j}(1 - \langle v_i, v_j \rangle) \quad \text{subject to } \forall i \ v_i \in \mathbb{S}^{n-1}. \quad (6)$$

An approximation to the integer problem (5) can be obtained from a solution to (6) in the following way: Choose a random unit vector x and construct the following cut: $S = \{i \mid \langle x, v_i \rangle \geq 0\}$. Denote by W the size of the cut produced this way and $\mathbb{E}[W]$ its expectation. [GW95] analyzed the approximation given by the SDP using the observation that

$$\mathbb{E}[W] = \sum_{i < j} w_{i,j} \frac{\arccos(v_i \cdot v_j)}{\pi}$$

and showing that

$$\sum_{i < j} w_{i,j} \frac{\arccos(v_i \cdot v_j)}{\pi} \geq \alpha \frac{1}{2} \sum_{i < j} w_{i,j}(1 - v_i \cdot v_j) \geq \alpha \cdot \text{OPT}$$

for $\alpha > 0.87856$, where OPT denotes the size of the maximal cut. Using the conditional expectation method, a set S can be found whose corresponding W (cut weight) is at least as large as the expectation, and thus is at least α times the size of the maximum cut [MR99].

We derandomize this process by choosing the vector x , with respect to which we define S , from an ϵ -sample for digons P_ϵ for some $\epsilon = o(1)$ (such a sample space is constructed in Section 5). As we shall soon see we will get that for most $x \in P_\epsilon$ the corresponding cut S is “good”. To prove this we simply go over the steps of the proof of Goemans and Williamson. We note that the only part that is sensitive to the fact that x is not completely random is in the analysis of $\mathbb{E}[W]$. Below we show that $\mathbb{E}[W]$ (almost) does not change when picking $x \in P_\epsilon$ uniformly at random instead of $x \in \mathbb{S}^{n-1}$ (since P_ϵ is small we can go over all $x \in P_\epsilon$ and pick the best one).

Lemma 6.1. $\mathbb{E}_{x \in P_\epsilon}[W] \geq \mathbb{E}_{x \in \mathbb{S}^{n-1}}[W] - 2\epsilon \cdot \text{OPT}$.

Proof. By definition of W :

$$\begin{aligned} \mathbb{E}_{x \in \mathbb{S}^{n-1}}[W] &= \sum_{i < j} w_{i,j} \Pr_{x \in \mathbb{S}^{n-1}}[\text{sign}(v_i \cdot x) \neq \text{sign}(v_j \cdot x)] = \\ &= \sum_{i < j} w_{i,j} \left(\Pr_{x \in P_\epsilon}[\text{sign}(v_i \cdot x) \neq \text{sign}(v_j \cdot x)] \pm \epsilon \right) = \mathbb{E}_{x \in P_\epsilon}[W] \pm \epsilon \sum_{i < j} w_{i,j}. \end{aligned}$$

Notice that OPT is bounded from below by $\frac{1}{2} \sum_{i < j} w_{i,j}$. This is since a set S randomly chosen by picking each vertex with probability $1/2$ will have an expected weight of $\frac{1}{2} \sum_{i < j} w_{i,j}$. Hence,

$$|\mathbb{E}_{x \in P_\epsilon}[W] - \mathbb{E}_{x \in \mathbb{S}^{n-1}}[W]| \leq \epsilon \sum_{i < j} w_{i,j} \leq 2\epsilon \cdot \text{OPT}.$$

□

Thus, by choosing $x \in P_\epsilon$ at random instead of $x \in \mathbb{S}^{n-1}$, we get an $(\alpha - 2\epsilon)$ -approximation algorithm. Keeping in mind that $\epsilon = o(1)$, the ratio is practically the same.

Corollary 6.2. *Let $\epsilon > 0$ and $n \in \mathbb{N}$. There exists an oblivious deterministic algorithm that transforms any solution to the SDP relaxation of Max-Cut into an $(\alpha - \epsilon)$ approximation for Max-Cut.*

Comparison to previous results. We now elaborate on the differences between our derandomization of the Goemans-Williamson algorithm and previous derandomizations of it [MR99, EIO02, Siv02]. The starting point for all current derandomization methods is the solution to the SDP corresponding to the Max-Cut problem. In [MR99] derandomization is performed using the method of conditional expectations and pessimistic estimators. In [Siv02], it is shown that the computation of the weight of a cut formed by a given hyperplane can be done with a log-space machine. Using Nisan’s PRG for such machines, the process is derandomized (non-obliviously). In [EIO02] the authors use an *instance specific* derandomization of the Johnson-Lindenstrauss lemma to reduce the dimension of the solution set (namely, they compute a projection matrix for the vectors that form the solution of the SDP). Then, as the dimension is low, they run over all possible projection vectors to find the one that gives the best cut. On the other hand, our derandomization gives a fixed set of hyperplanes, of size $n^{1+o(1)}$ such that any SDP solution has at least one hyperplane in the set yielding a sufficiently good rounding. Thus, our approach is oblivious to the underlying solution of the SDP. One advantage of our proof is that it can be parallelized. That is, in order to round the SDP solution we can check all the possible roundings given by our construction in parallel and output the best one. In contrast, the derandomization procedures in [MR99, EIO02, Siv02] are sequential in nature and cannot be parallelized. Another advantage of our construction is that the rounding of the SDP can be performed in time $n^{(1+o(1))\omega}$, where ω is the exponent of matrix multiplication (which is at most 2.36 [CW90]), while the fastest algorithm so far, by [EIO02], runs in time $n^{3+o(1)}$.

6.2 Coloring 3-Colorable Graphs

A coloring of a graph is an assignment of colors to its vertices such that each pair of neighbors is colored differently. A graph is said to be k -colorable if it has a coloring with k distinct colors. We deal with the following promise problem: Given a graph on n vertices that is 3-colorable, efficiently find a 3-coloring of the graph. This problem is well known to be NP-hard. [KMS98] gave an approximation algorithm that efficiently colors a 3-colorable graph

with $O(\min\{n^{0.387}, \Delta^{\log_3 2} \log n\})$ colors where Δ is the maximum degree of any vertex¹⁰.

We obtain the following result.

Theorem 6.3. *For any $\epsilon > 0$, There exists an oblivious derandomization to the randomized approximation algorithm of [KMS98], achieving a coloring of $O(\min\{n^{0.387+\epsilon}, \Delta^{\log_3(2)+\epsilon} \log n\})$ colors for a 3-colorable graph with n vertices, where Δ is the maximum degree of any vertex. The running time overhead of the derandomization is $n^{O(\log(\epsilon^{-1}))}$.*

We now give a brief description of the approximation algorithm: First, solve a semi-definite program assigning a vector to each vertex such that the angle between any pair of neighbors is large ($\frac{2\pi}{3}$ radians). Notice that the existence of such an assignment is guaranteed by the 3-colorability of the graph. Next, assign colors to the vertices in the following way: Choose r random unit vectors independently x_1, \dots, x_r . Each vertex will receive r bits. The value of the i 'th bit of a vertex with a corresponding vector v will be set according to the sign of $\langle v, x_i \rangle$. The color of the vertex will be described by the r bits. The probability that two neighboring vertices will have the same i 'th bit is at most $1/3$ due to the large angle between their vectors (same analysis as in the Goemans-Williamson algorithm). As a result we get that the color assigned to two neighboring vectors is equal with probability at most 3^{-r} . The probability that a vertex v will have a neighbor having the same color is at most $\Delta 3^{-r}$ where Δ is the maximum vertex degree. By taking $r = \lceil \log_3(\Delta) + 2 \rceil$ we get that the expectation of the percentage of vertices that have neighbors with the same color is $1/4$. By trying several times we get a ‘‘semi-coloring’’ for which at least half the vertices have no neighbors of the same color. We now repeat this process recursively with a new set of colors on the vertices with neighbors of the same color (we later explain how this repetition is made in the derandomized version). This will result with $O(\Delta^{\log_3 2}) \approx O(\Delta^{0.631})$ colors when $\Delta = \Omega(n^c)$ for some constant $c > 0$ or $O(\Delta^{\log_3 2} \log n)$ colors for general Δ . Notice that Δ may be as high as $n - 1$. In such cases, the approximation can be improved by using the following method: For any vertex whose degree is higher than $\delta \approx n^{0.613}$, color its neighboring vertices (they can be 2-colored efficiently) in 2 new colors. This will use at most $2n/\delta$ colors and reduce the maximum degree to δ . Taking the optimal value for δ ($\delta \approx n^{0.613}$) results in a $\min\{n^{0.387}, \Delta^{\log_3(2)} \log(n)\}$ approximation.

Our derandomization differs in that we choose the r ‘random’ vectors from the set P_ϵ described in the previous section (as opposed to random unit vectors) by taking a random expander walk of length r . For the analysis we require a known result concerning expander graphs that is given in Section 6.3. Let $\epsilon > 0$ be some small constant. The set P_ϵ denotes an ϵ -sample for digons as guaranteed by Theorem 1.3. We describe a randomized algorithm that requires a logarithmic number of random bits. This can be derandomized by going over all settings for the random bits. We choose the vectors x_1, \dots, x_r in the following way: First, construct an expander graph of parameters (n', d, λ) where $n' = |P_\epsilon|$, $d = O(\epsilon^{-2})$ and¹¹

¹⁰In fact, they show two methods where the second method obtains a coloring of $\min\{O(\Delta^{1/3} \log^{1/2} \Delta \log n), O(n^{1/4} \log^{1/2} n)\}$ colors. However, our constructions can only derandomize the first method, yielding the slightly worse approximation.

¹¹We set d as the smallest integer for which there exist an efficient construction for an expander graph with $\lambda \leq \epsilon d$. Since there exist graphs with $\lambda \approx \sqrt{d}$, $d = O(\epsilon^{-2})$ is sufficiently large.

$\lambda \leq \epsilon d$. Label each vertex of the graph as a vector in P_ϵ . Choose $x_1, \dots, x_r \in P_\epsilon$ by taking a random walk of length r in the expander.

The amount of random bits required in order to choose x_1, \dots, x_r is

$$\log(n') + r \log(d) = O(\log(\epsilon^{-1}) \log(n)) .$$

Hence, the support size of the sample space is polynomial in n assuming a constant ϵ (or of size $n^{O(\log(\epsilon^{-1}))}$ for general ϵ). We define this sample space of r -tuples as \mathcal{X} . That is, the r -tuple of vectors is chosen uniformly from \mathcal{X} . The following lemma bounds the probability that the chosen r -tuple does not ‘separate’ two neighboring vectors from the graph (the original graph which we are coloring). It actually proves a slightly stronger statement that will be put to use later:

Lemma 6.4. *Let v_1, v_2 be two vectors corresponding to two neighboring vertices of the graph. Let c_1, c_2 be the colors assigned to each vertex according to the choice of $(x_1, \dots, x_{r'})$ for some $r' \leq r$. Then*

$$\Pr_{(x_1, \dots, x_{r'}) \in \mathcal{X}} [c_1 = c_2] < (1/3 + 2\epsilon)^{r'-1} .$$

Proof. Let B be the set of vectors in P_ϵ for which $\text{sign}(\langle v_1, x \rangle) = \text{sign}(\langle v_2, x \rangle)$. Notice that due to the properties of P_ϵ and the fact that the angle between v_1, v_2 is $\frac{2\pi}{3}$ we have that

$$\frac{|B|}{|P_\epsilon|} = \Pr_{x \in P_\epsilon} [\text{sign}(\langle v_1, x \rangle) = \text{sign}(\langle v_2, x \rangle)] < 1/3 + \epsilon .$$

The vertices corresponding to v_1, v_2 are assigned the same color only when the entire random walk lies within the set B . We bound the probability of this event by using Lemma 6.5 which leads to the following bound:

$$\Pr [\forall i \in [r'], x_i \in B] < \left(\frac{|B|}{|P_\epsilon|} + \frac{\lambda}{d} \right)^{r'-1} \leq (1/3 + 2\epsilon)^{r'-1} .$$

□

Proof of Theorem 6.3. By the above lemma, following the original notations of the algorithm, we may now take $r = \left\lceil \log_{\frac{1}{1/3+2\epsilon}}(\Delta) + 3 \right\rceil$ instead of $r = \lceil \log_3(\Delta) + 2 \rceil$ and the analysis remains the same. Specifically, at least one r tuple in \mathcal{X} will provide a coloring in which at least $n/2$ vertices do not have any neighbor of the same color. As in the original algorithm, we proceed recursively on the set of vertices that have neighbors with the same color. It is easy to see that after repeating this process for at most $\log n$ steps we achieve a coloring using $O(\min\{n^{0.387+O(\epsilon)}, \Delta^{\log_3 2+O(\epsilon)} \log n\})$ many colors with a running time of $n^{O(\log(\epsilon^{-1}))}$. □

6.3 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 $\lambda \leq 2\sqrt{d-1}$ [Mar88, LPS88].

A random walk of length t 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 $1 < i \leq t$) 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, \dots, v_t) . We shall make use of the following lemma regarding such walks

Lemma 6.5. [AKS87, AFWZ95] *Let G be an (n, d, λ) -expander. Let $B \subset V(G)$ be a subset of vertices. Denote by \mathcal{E} the event that a random walk (v_1, \dots, v_ℓ) stays inside B . That is, the event in which $\forall i, v_i \in B$. The probability for the event \mathcal{E} to occur is at most*

$$\left(\frac{|B|}{|V(G)|} + \frac{\lambda}{d} \right)^{\ell-1}.$$

7 Acknowledgments

We would like to thank Jelani Nelson for pointing out a simple construction of a JL-transform sample space using [AMS99]. We thank the anonymous referees for useful comments.

References

- [AB09] S. Arora and B. Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.
- [AC09] N. Ailon and B. Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009.
- [Ach03] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.
- [AFWZ95] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Journal of Computational Complexity*, 5(1):60–75, 1995.
- [AKS87] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in logspace. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC)*, pages 132–140, 1987.
- [AL09] N. Ailon and E. Liberty. Fast dimension reduction using Rademacher series and dual BCH codes. *Discrete and Computational Geometry*, 42:615–630, 2009.
- [AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.
- [AS08] N. Alon and J. Spencer. *The probabilistic method*. J. Wiley, 3 edition, 2008.

- [Cha00] B. Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, New York, NY, USA, 2000.
- [CW90] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progression. *Journal of Symbolic Computation*, 9:251–280, 1990.
- [CW09] K.L. Clarkson and D.P. Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 205–214. ACM, 2009.
- [DF87] P. Diaconis and D. Freedman. A dozen de Finetti-style results in search of a theory. *Annales de l’IHP Probabilités et statistiques*, 23:397–423, 1987.
- [DG03] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.
- [DGJ+10] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. A. Servedio, and E. Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010.
- [DKN10] I. Diakonikolas, D.M. Kane, and J. Nelson. Bounded independence fools degree-2 threshold functions. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.
- [EIO02] L. Engebretsen, P. Indyk, and R. O’Donnell. Derandomized dimensionality reduction with applications. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 705–712, 2002.
- [Gol01] O. Goldreich. *Foundations of cryptography: basic tools*. Cambridge University Press, 2001.
- [GUV09] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM (JACM)*, 56(4):1–34, 2009.
- [GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [Ind07] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1 . In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 615–620, 2007.
- [JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [JN10] W.B. Johnson and A. Naor. The Johnson–Lindenstrauss lemma almost characterizes Hilbert space, but not quite. *Discrete and Computational Geometry*, 43(3):542–553, 2010.

- [KMS98] D. R. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM (JACM)*, 45(2):246–265, 1998.
- [KN10] D.M. Kane and J. Nelson. A derandomized sparse Johnson-Lindenstrauss transform. *Arxiv preprint arXiv:1006.3585*, 2010.
- [KRS09] Z. S. Karnin, Y. Rabani, and A. Shpilka. Explicit dimension reduction and its applications. *Electronic Colloquium on Computational Complexity (ECCC)*, (121), 2009.
- [KV94] M. J. Kearns and U. V. Vazirani. *An introduction to computational learning theory*. MIT Press, Cambridge, MA, USA, 1994.
- [LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [LW05] M. Luby and A. Wigderson. Pairwise independence and derandomization. *Foundations and Trends in Theoretical Computer Science*, 1(4), 2005.
- [Mar88] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, 1988.
- [Mat08] J. Matousek. On variants of the Johnson-Lindenstrauss lemma. *Random Structures and Algorithms*, 33(2):142–156, 2008.
- [Mek10] R. Meka. Almost optimal explicit Johnson-Lindenstrauss transformations. *Electronic Colloquium on Computational Complexity (ECCC)*, (183), 2010.
- [MR99] S. Mahajan and H. Ramesh. Derandomizing approximation algorithms based on semidefinite programming. *SIAM Journal on Computing*, 28(5):1641–1663, 1999.
- [MZ09] R. Meka and D. Zuckerman. Pseudorandom generators for polynomial threshold functions. <http://arxiv.org/abs/0910.4122>, 2009.
- [MZ10] R. Meka and D. Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 427–436, 2010.
- [Nis92] N. Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.
- [PR96] J. K. Patel and C. B. Read. *Handbook of the Normal Distribution*. 1996. CRC Press, Boca Raton, FL, 1996.
- [RS10] Y. Rabani and A. Shpilka. Explicit construction of a small ϵ -net for linear threshold functions. *SIAM Journal on Computing*, 39(8):3501–3520, 2010.

- [Sha02] R. Shaltiel. Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 77:67–95, June 2002.
- [Siv02] D. Sivakumar. Algorithmic derandomization via complexity theory. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 619–626, 2002.
- [Zuc97] D. Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11(4):345–367, 1997.

A Pseudo-Random-Generator for Bounded Space Machines

Let \mathcal{F} be a family of functions from $\{0, 1\}^m$ to $\{-1, 1\}$. Let $r < m$ and $G : \{0, 1\}^r \rightarrow \{0, 1\}^m$ be a function expanding r bits into m bits. We say that G is an ϵ -pseudo-random generator (ϵ -PRG) for \mathcal{F} when

$$|\mathbb{E}_{x \in \{0,1\}^m} [f(x)] - \mathbb{E}_{y \in \{0,1\}^r} [f(G(y))]| < \epsilon$$

for every function $f \in \mathcal{F}$. In other words, G expands a seed of r truly random bits into m bits that seem random to any function in \mathcal{F} . Note that ϵ -samples are equivalent to ϵ -PRGs. For example, by taking \mathcal{F} to be the family of all linear threshold functions, restricted to inputs from $\{-1, 1\}^m$, we get that the image of an ϵ -PRG for \mathcal{F} is an ϵ -sample for half-spaces over the hypercube. The following Theorem of [Nis92] gives a PRG for bounded space machines. Denote by $\text{space}(s)$ the family of functions that can be computed by reading the input bits only once (one pass) using at most s bits of memory.

Theorem A.1 ([Nis92]). *Let $\epsilon = 2^{-O(s)}$. There exists an explicit ϵ -PRG for $\text{space}(s)$, $G : \{0, 1\}^{O(s \log(m))} \rightarrow \{0, 1\}^m$.*

Applying the same ideas as [Siv02] we use such PRGs in order to construct ϵ -samples for digons and for spherical caps when the dimension is low. We will focus on ϵ -sample for digons as the construction and proof for spherical caps is essentially the same. The majority of this section will be devoted to proving the following Theorem.

Theorem A.2. *Let t be an integer and let $0 < \epsilon < 1$. There exists an efficiently constructible set $P \subseteq \mathbb{R}^t$ such that P is an $O(\epsilon + 1/t)$ -sample for digons w.r.t. the uniform measure over the unit sphere and $|P| = \exp(O(\log^2(t/\epsilon)))$.*

Lemma 5.5 is a direct consequence of Theorem A.2. Notice that P is not necessarily a subset of the unit sphere. In the case of digons, this is not a problem since the vectors can be normalized without changing the properties of P . This is since for any digon D , vector x and non-zero scalar λ , $x \in D$ iff $\lambda x \in D$.

Let $\mathcal{F}_{\text{digon}}$ be the family of digon indicator functions. That is, the family of functions from \mathbb{R}^t to $\{-1, 1\}$ of the form

$$f_{v,u}(x) \triangleq \text{sign}(\langle v, x \rangle) \cdot \text{sign}(\langle u, x \rangle).$$

The proof of Theorem A.2 will be in three steps. First we show that we can replace the uniform distribution over the sphere with the Gaussian distribution. Namely, we will show that for any $f \in \mathcal{F}_{\text{digon}}$ it holds that $\mathbb{E}_{x \in \mathbb{S}^{t-1}}[f(x)] \approx \mathbb{E}_{x \sim \mathcal{N}(0, 1/t)^t}[f(x)]$, where $\mathcal{N}(0, 1/t)^t$ is the distribution over \mathbb{R}^t which consists of t independent copies of $\mathcal{N}(0, 1/t)$, the Gaussian distribution with mean 0 and variance $1/t$. In the second step we shall construct a simple finite ϵ -sample S of exponential (in t, ϵ) size for digons w.r.t. the Gaussian measure (this is done in Section A.1). Finally, by using the PRG for bounded space machines, we prove that some smaller subset of S , that we denote by P , is an ϵ -sample w.r.t. to the Gaussian distribution (Section A.2). We begin by showing how to move to the Gaussian distribution.

Lemma A.3. *For any $f \in \mathcal{F}_{\text{digon}}$ it holds that*

$$\left| \mathbb{E}_{x \in \mathbb{S}^{t-1}}[f(x)] - \mathbb{E}_{x \sim \mathcal{N}(0, 1/t)^t}[f(x)] \right| = O(1/t).$$

Proof. We require a result by [DF87] analyzing projections of a uniformly chosen unit vector.

Lemma A.4. *[Special case of Equation 1 in [DF87]] Let t be an integer and let $v, u \in \mathbb{S}^{t-1}$ be a pair of unit vectors. For any event $E(x)$, depending only on $\langle v, x \rangle$ and $\langle u, x \rangle$ it holds that*

$$\left| \Pr_{x \in \mathbb{S}^{t-1}}[E(x)] - \Pr_{x \sim \mathcal{N}(0, 1/t)^t}[E(x)] \right| = O(1/t).$$

Lemma A.3 immediately follows since for any $f_{v,u} \in \mathcal{F}_{\text{digon}}$, the value of $f(x)$ depends only on $\langle v, x \rangle$ and $\langle u, x \rangle$. \square

A.1 A Finite Sample Space

In this section we give a construction of a sample space of exponential size for digons, w.r.t. the measure defined by $\mathcal{N}(0, 1/t)^t$. Let $s \in \mathbb{N}$. To construct the sample space we first identify strings of s bits with elements of \mathbb{R} in a way that a random string would be interpreted, roughly, as a random Gaussian. We begin with some well known facts regarding the Gaussian distribution. Their proof can be found in e.g. [PR96], Chapter 2.

Fact A.5. *Let $z \sim \mathcal{N}(0, 1)$.*

- *(anti-concentration): For any interval I , $\Pr[z \in I] < |I|$ where $|I|$ denotes the length of the interval.*
- *(concentration): For any $\alpha > 0$, $\Pr[z > \alpha] \leq O\left(e^{-\alpha^2/2}\right)$.*
- *For $z' \sim \mathcal{N}(0, 1)$ independent of z and any $\alpha, \beta > 0$ it holds that $\alpha z + \beta z' \sim \mathcal{N}(0, \alpha^2 + \beta^2)$. In particular, for any $\alpha > 0$, $\alpha z \sim \mathcal{N}(0, \alpha^2)$. Also, for a vector $v \in \mathbb{R}^t$ and a vector $a \sim \mathcal{N}(0, 1/t)^t$, it holds that $\langle a, v \rangle \sim \mathcal{N}(0, \|v\|_2^2/t)$.*

Definition A.6. *Let $s \in \mathbb{N}$. Define $\mathcal{I}_s = \{I_i\}_{i \in \{0,1\}^s}$ to be a partition of the interval $(-\infty, +\infty)$ into consecutive intervals such that the measure of each interval under the distribution $\mathcal{N}(0, 1/t)$ is the same. Namely, for $z \sim \mathcal{N}(0, 1/t)$ and any $i \in \{0, 1\}^s$, $\Pr[z \in I_i] = 2^{-s}$.*

Notice that \mathcal{I}_s is uniquely defined, up to reordering of the intervals and deciding whether to move an endpoint of one interval to a ‘touching’ interval. In the rest of this section, we will work with some $s = \Theta(\log(t/\epsilon))$ where the constant in the $\Theta()$ expression is sufficiently large (the exact requirements will appear at a later stage). From Fact A.5 we have that all intervals I in \mathcal{I}_s are of length at least $\sqrt{t}2^{-s} = (t/\epsilon)^{-O(1)}$. Fact A.5 also implies that the absolute value of the endpoints of all *finite* intervals (or their closure) in \mathcal{I}_s , is bounded by $O(\sqrt{s/t}) = O(\sqrt{\log(\epsilon/t)/t})$. In particular, it follows that every interval in \mathcal{I}_s contains integer multiples of $(t/\epsilon)^{-c_1}$ for some constant $c_1 > 0$, where we never need to multiply it by more than $\lfloor (t/\epsilon)^{c_2} \rfloor$ (in absolute value), for some $c_2 > 0$.

Definition A.7. For an interval I denote with \bar{I} the closure of I . Clearly I and \bar{I} can only differ in one or two points. When I is finite we have $\bar{I} = [\alpha, \beta]$, for some α and β , and we denote by $z(I)$ the closest integer multiple of $(t/\epsilon)^{-c_1}$ to $(\alpha + \beta)/2$. I.e. to the mid-point of I . When I is infinite we have that $\bar{I} = [\alpha, +\infty)$ (or $\bar{I} = (-\infty, \alpha]$) for some α and we define $z(I)$ as the closest integer multiple of $(t/\epsilon)^{-c_1}$, inside I , to α . For $i \in \{0, 1\}^s$, let $z(i) \triangleq z(I_i)$. For $z \in \mathbb{R}$, let $i(z)$ be the string identifying the interval in which z resides. Define the Gaussian rounding of z , denoted by \tilde{z} , as $\tilde{z} \triangleq z(I_{i(z)})$. Namely, as $z(I_j)$ where I_j is the interval in which z lies (or, equivalently, as $z(i(z))$).

We note that for any $z \in \mathbb{R}$, the number \tilde{z} can be represented either with s bits by considering $i(z)$ or with $O(s)$ bits when writing it as in integer product of $(t/\epsilon)^{-c_1}$ (recall that $s = \Theta(\log(t/\epsilon))$). The advantage of the second representation is that with it, computing products and additions of two rounded numbers requires $O(s)$ bits of memory. In the rest of the section we will use both methods of representation. The correspondence between real numbers and strings can be extended to vectors (and to longer strings).

Definition A.8. Let $s, t \in \mathbb{N}$. Identify $a \in \{0, 1\}^{st}$ with (a_1, \dots, a_t) where each $a_j \in \{0, 1\}^s$. For $a \in \{0, 1\}^{st}$, its corresponding vector in \mathbb{R}^t is defined as $x(a) = (z(a_1), z(a_2), \dots, z(a_t))$. For a vector $x \in \mathbb{R}^t$, its corresponding string is $a(x) = (i(x_1), i(x_2), \dots, i(x_t))$. Define \tilde{x} , the Gaussian rounding of x , as $(\tilde{x}_1, \dots, \tilde{x}_t)$.

By the definition of \mathcal{I}_s it follows that for a vector $x \sim \mathcal{N}(0, 1/t)^t$, the string $a(x)$ is uniformly distributed in $\{0, 1\}^{st}$. The Gaussian rounding of a vector x should be in some sense, an approximation for it. To that end, for some sufficiently large constant c_3 that will be determined later, we say that a vector $x \in \mathbb{R}^t$ is roundable when $\|x - \tilde{x}\|_2 < (t/\epsilon)^{-c_3}$. Notice that not all vectors are roundable (e.g. when vectors having extremely large coordinates are not roundable). However, in the following lemma we prove that a vector is roundable w.h.p.

Lemma A.9. For any constant $c > 0$ there exists a sufficiently large $s = \Theta(\log(t/\epsilon))$ such that $\|x - \tilde{x}\|_2 < (t/\epsilon)^{-c}$ with probability at least $1 - \epsilon$.

Proof. For a positive $B \in \mathbb{R}$, we say that x is B -bounded when all of the coordinates of x are bounded, in absolute value, by B . Let $B = O(\sqrt{\log(t/\epsilon)/t})$ be such that for $x \sim \mathcal{N}(0, 1/t)^t$, the probability that x is not B -bounded is ϵ . The asymptotic upper bound on B holds due to standard concentration bounds for the normal distribution (see Fact A.5).

We will now show that for sufficiently large s , a B -bounded vector is roundable. To upper bound $\|x - \tilde{x}\|_2$ it suffices to show that all of the coordinates of x fall in intervals whose lengths are at most $t' \triangleq (t/\epsilon)^{-c}/\sqrt{t}$. We now show that for $s = \Theta(\log(t/\epsilon))$ (where the constant in the $\Theta(\cdot)$ depends on c) this is indeed the case. To that end, since the pdf (probability density function) of $\mathcal{N}(0, 1/t)$ is symmetric and decreasing for positive z , it suffices to prove that

$$\Pr_{z \sim \mathcal{N}(0, 1/t)} [z \in [B, B + t']] \geq (t/\epsilon)^{-O(1)}.$$

By assuming w.l.o.g. that $t' < B$, we reach the required conclusion. Denote by $\phi(\cdot)$ the pdf of $\mathcal{N}(0, 1/t)$. It holds that

$$\Pr_{z \sim \mathcal{N}(0, 1/t)} [z \in [B, B + t']] \geq \phi(2B) \cdot t' = (t/\epsilon)^{-O(1)}.$$

□

Lemma A.10. *The set $S \triangleq \{x(a) | a \in \{0, 1\}^{st}\}$ is an $O(\epsilon)$ -sample for digons w.r.t. the measure defined by $\mathcal{N}(0, 1/t)^t$. Specifically, for any digon indicator function $f = f_{v,u}$,*

$$|\mathbb{E}_{x \sim \mathcal{N}(0, 1/t)^t} [f(x)] - \mathbb{E}_{x \in S} [f(x)]| = |\mathbb{E}_{x \sim \mathcal{N}(0, 1/t)^t} [f(x) - f(\tilde{x})]| = O(\epsilon).$$

Proof. The first equality stems from the fact that for a vector $x \sim \mathcal{N}(0, 1/t)^t$, the string $a(x)$ is uniformly distributed in $\{0, 1\}^{st}$. We focus on proving the second equality. Let $E(x)$ be the event where $\|x - \tilde{x}\|_2 \geq \epsilon/\sqrt{t}$ (i.e., x is not roundable), or $|\langle x, v \rangle| < \epsilon/\sqrt{t}$ or $|\langle x, u \rangle| < \epsilon/\sqrt{t}$. When $E(x)$ does not occur we have that $\|x - \tilde{x}\|_2 < \epsilon/\sqrt{t}$. Hence, $\text{sign}(\langle x, v \rangle) = \text{sign}(\langle \tilde{x}, v \rangle)$ and $\text{sign}(\langle x, u \rangle) = \text{sign}(\langle \tilde{x}, u \rangle)$ (as v, u are unit vectors) and so, $f(x) = f(\tilde{x})$. Since $|f(x) - f(\tilde{x})| \leq 2$, it follows that

$$|\mathbb{E}_{x \sim \mathcal{N}(0, 1/t)^t} [f(x) - f(\tilde{x})]| / 2 \leq \Pr_{x \sim \mathcal{N}(0, 1/t)^t} [E(x)] \leq$$

$$\Pr \left[|\langle x, v \rangle| < \epsilon/\sqrt{t} \right] + \Pr \left[|\langle x, u \rangle| < \epsilon/\sqrt{t} \right] + \epsilon = O(\epsilon).$$

The last inequality holds when s is sufficiently large, due to Lemma A.9. The last equality holds due to standard anti-concentration bounds of the normal distribution and since $\langle x, v \rangle$ and $\langle x, u \rangle$ are distributed according to $\mathcal{N}(0, 1/t)$ (see Fact A.5). □

A.2 The Small Sample Space

In this section we construct the sample space P such that for any digon indicator function $f_{v,u}$,

$$\mathbb{E}_{x \in P} [f_{v,u}(x)] \approx \mathbb{E}_{x \in S} [f_{v,u}(x)] \approx \mathbb{E}_{x \in \mathcal{N}(0, 1/t)^t} [f_{v,u}(x)].$$

In the previous section we proved the second equality. We now focus on the first one. Let $m = st$ where $s = O(\log(t/\epsilon))$ is sufficiently large, as defined in the previous section. For a pair of unit vectors v, u , we define $\tilde{f}_{v,u} : \{0, 1\}^m \rightarrow \{-1, 1\}$ as $\tilde{f}_{v,u}(a) = f_{v,u}(x(a)) = \text{sign}(\langle x(a), v \rangle) \cdot \text{sign}(\langle x(a), u \rangle)$. We call $\tilde{f}_{v,u}$ a restricted digon indicator function as it can be viewed as a digon indicator function (over \mathbb{R}^t), restricted to the points in S . Let $\mathcal{F}_{\text{restricted}}$ be the family

of “restricted digon indicator functions”. We will prove that an ϵ -PRG for $\text{space}(s')$, where $s' = O(\log(t/\epsilon))$, is an $O(\epsilon)$ -PRG for $\mathcal{F}_{\text{restricted}}$. Denote by G the PRG and by P the image of the G , interpreted as a set of vectors in \mathbb{R}^t . Namely, $P = \{x(a) | a \in \text{Image}(G)\}$. As G fools digon indicator functions restricted to S , it holds that $\mathbb{E}_{x \in P}[f_{v,u}(x)] \approx \mathbb{E}_{x \in S}[f_{v,u}(x)]$ for any digon indicator function $f_{v,u}$. By Lemma A.10 we get that P is the required sample space.

The outline of the proof is as follows. For any function $\tilde{f} = \tilde{f}_{v,u} \in \mathcal{F}_{\text{restricted}}$ we define a small memory estimate $g_{\tilde{f}}$. This in turn defines a family of small memory estimates $\mathcal{F}_{\text{approx}}$. The functions of the form $g_{\tilde{f}}$ will be in $\text{space}(s')$ for some $s' = O(\log(t/\epsilon))$. This is the setting required for applying Nisan’s PRG, which we use to construct a PRG G for $\mathcal{F}_{\text{approx}}$. We proceed to show that for any $\tilde{f} = \tilde{f}_{u,v} \in \mathcal{F}_{\text{restricted}}$, the expectation of $\tilde{f}(G)$ (i.e., the composition of \tilde{f} with the generator G) is roughly the same as the expectation of $g_{\tilde{f}}(G)$. This is done by considering another family of functions $\mathcal{F}_{\text{error}}$ which are indicator functions to whether a vector x is such that the estimate $g_{\tilde{f}}$ might be wrong due to rounding issues. We shall see that $\mathcal{F}_{\text{error}}$ is also a sub-family of $\text{space}(s')$ and thus G is a PRG for $\mathcal{F}_{\text{error}}$ as well. The required result will follow from (roughly) the triangle inequality.

We begin by formally defining $\mathcal{F}_{\text{approx}}$ and $\mathcal{F}_{\text{error}}$. As a first step we define a standard rounding (as opposed to the Gaussian rounding) for elements in \mathbb{R} and \mathbb{R}^t .

Definition A.11. For $z \in \mathbb{R}$, define the standard rounding of z , denoted by \hat{z} , as the closest integer multiple of $(t/\epsilon)^{-c_1}$ to z , where c_1 is a sufficiently large constant (the same constant appearing in Definition A.7). For a vector $v \in \mathbb{R}^t$ define \hat{v} as $(\hat{v}_1, \dots, \hat{v}_t)$.

For a function $\tilde{f}_{v,u} \in \mathcal{F}_{\text{restricted}}$, the function $g_{\tilde{f}} : \{0, 1\}^m \rightarrow \{-1, 1\}$ is defined as $g_{\tilde{f}}(a) \triangleq \tilde{f}_{\hat{v}, \hat{u}}(a)$. The family $\mathcal{F}_{\text{approx}}$ is defined as $\mathcal{F}_{\text{approx}} \triangleq \{g_{\tilde{f}} | \tilde{f} \in \mathcal{F}_{\text{restricted}}\}$. Since the standard representation of each entry in $\hat{v}, \hat{u}, x(a)$ requires $O(\log(t/\epsilon))$ bits (see the discussion after Definition A.7), the calculation of $\langle x(a), \hat{v} \rangle$ and $\langle x(a), \hat{u} \rangle$ can be done in $\text{space}(O(\log(t/\epsilon)))$. It follows that $\mathcal{F}_{\text{approx}} \subseteq \text{space}(s')$ for some $s' = O(\log(t/\epsilon))$.

Define $h_{\tilde{f}} : \{0, 1\}^m \rightarrow \{-1, 1\}$ such that $h_{\tilde{f}}(a) = -1$ iff $|\langle x(a), \hat{v} \rangle| \geq \epsilon/\sqrt{t}$ and $|\langle x(a), \hat{u} \rangle| \geq \epsilon/\sqrt{t}$. By the same arguments as before it can be shown that $\mathcal{F}_{\text{error}} \triangleq \{h_{\tilde{f}} | \tilde{f} \in \mathcal{F}_{\text{restricted}}\}$ is a sub-family of $\text{space}(s')$ for some $s' = O(\log(t/\epsilon))$.

The following lemma proves that $h_{\tilde{f}}$ is indeed a good measure to whether $g_{\tilde{f}}$ might err.

Lemma A.12. Let $\tilde{f} = \tilde{f}_{v,u} \in \mathcal{F}_{\text{restricted}}$ and let $a \in \{0, 1\}^m$ be such that $h_{\tilde{f}}(a) = -1$. Then $g_{\tilde{f}}(a) = \tilde{f}(a)$

Proof. Let $\alpha \triangleq \max_{a \in \{0, 1\}^m} \{\|x(a)\|_2\}$. It is clear from the definition of Gaussian rounding (see Definition A.7 and the discussion prior to it) that $\alpha = (t/\epsilon)^{O(1)}$. Assuming the standard rounding is sufficiently fine (i.e., c_1 in Definition A.11 is sufficiently large), we have $\|v - \hat{v}\|_2, \|u - \hat{u}\|_2 < \frac{\epsilon}{\alpha\sqrt{t}}$. Hence, when $h(a) = -1$ we have $\text{sign}(\langle x(a), v \rangle) = \text{sign}(\langle x(a), \hat{v} \rangle)$ and $\text{sign}(\langle x(a), u \rangle) = \text{sign}(\langle x(a), \hat{u} \rangle)$. It follows that $g_{\tilde{f}}(a) = \tilde{f}(a)$. \square

As a corollary we get that in order to bound the percentage of points in which $\tilde{f}(a) \neq g_{\tilde{f}}(a)$ (or alternatively, $\mathbb{E} \left[\left| \tilde{f}(a) - g_{\tilde{f}}(a) \right| \right]$), it suffices to bound $\Pr[h_{\tilde{f}}(a) = 1]$. The following lemma proves that $g_{\tilde{f}}$ is indeed a good estimation for \tilde{f} , w.r.t. the uniform distribution over $\{0, 1\}^m$.

Lemma A.13. $\mathbb{E}_{a \in S} [|\tilde{f}(a) - g_{\tilde{f}}(a)|] \leq 2 \Pr[h_{\tilde{f}}(a) = 1] = O(\epsilon)$.

Proof. The first inequality stems immediately from the previous lemma as $|\tilde{f}(a) - g_{\tilde{f}}(a)| \leq 2$. We now analyze the probability that $h_{\tilde{f}}(a) = 1$. It will be convenient to analyze, for $y \sim \mathcal{N}(0, 1/t)^t$, the probability that $h_{\tilde{f}}(a(y)) = 1$. By the definition of \mathcal{I}_s and the function $a(y)$ (Definitions A.6 and A.8) it follows that $a(y)$ is uniformly distributed in $\{0, 1\}^m$, hence $h_{\tilde{f}}(a)$ and $h_{\tilde{f}}(a(y))$ are distributed identically.

Let $y \sim \mathcal{N}(0, 1/t)^t$. Recall that according to Lemma A.9, for sufficiently large $s = O(\log(t/\epsilon))$ (that is the constant in the $O()$ expression should be sufficiently large) it holds that $\|y - \tilde{y}\|_2 < \epsilon/\sqrt{t}$ (i.e., y is roundable) with probability at least $1 - \epsilon$. Assuming the standard rounding is sufficiently fine (i.e., c_1 in Definition A.11 is sufficiently large), we have $1/2 < \|\hat{v}\|_2, \|\hat{u}\|_2 < 2$ since both v and u are unit vectors. Denote by $E(y)$ the event that $|\langle \hat{v}, y \rangle| < 3\epsilon/\sqrt{t}$ or $|\langle \hat{u}, y \rangle| < 3\epsilon/\sqrt{t}$ or y is not roundable. We will now show that

$$\Pr_{y \in \mathcal{N}(0, 1/t)^t} [h_{\tilde{f}}(a(y)) = 1] \leq \Pr_{y \in \mathcal{N}(0, 1/t)^t} [E(y)] = O(\epsilon).$$

To prove the last equality, notice that

$$\Pr[E(y)] \leq \Pr[|\langle \hat{v}, y \rangle| < 3\epsilon/\sqrt{t}] + \Pr[|\langle \hat{u}, y \rangle| < 3\epsilon/\sqrt{t}] + \epsilon.$$

By Fact A.5 we have that $\langle \hat{v}, y \rangle \sim \mathcal{N}(0, \|\hat{v}\|_2^2/t)$. As $\|\hat{v}\|_2 > 1/2$, standard anti concentration bounds for Gaussians (Fact A.5) imply that $\Pr[|\langle \hat{v}, y \rangle| < 3\epsilon/\sqrt{t}] = O(\epsilon)$. Since the same arguments hold for the vector \hat{u} we have that $\Pr[E(y)] = O(\epsilon)$.

Assume that $h_{\tilde{f}}(a(y)) = 1$. If y is not roundable then $E(y)$ has occurred. Else we have $\|y - \tilde{y}\|_2 < \epsilon/\sqrt{t}$ and w.l.o.g. $\|\hat{v}\|_2 < 2$ and $|\langle y, \hat{v} \rangle| < \epsilon/\sqrt{t}$, meaning that $E(y)$ must have occurred. It follows that

$$\Pr_{a \in \{0, 1\}^m} [h_{\tilde{f}}(a) = 1] = \Pr_{y \in \mathcal{N}(0, 1/t)^t} [h_{\tilde{f}}(a(y)) = 1] \leq \Pr_{y \in \mathcal{N}(0, 1/t)^t} [E(y)] = O(\epsilon).$$

□

Lemma A.14. Let $G : \{0, 1\}^r \rightarrow \{0, 1\}^m$ be an ϵ -PRG for both $\mathcal{F}_{\text{approx}}$ and $\mathcal{F}_{\text{error}}$. It holds that G is an $O(\epsilon)$ -PRG for $\mathcal{F}_{\text{restricted}}$.

Proof. Let $\tilde{f} \in \mathcal{F}_{\text{restricted}}$. Since G is an ϵ -PRG for $\mathcal{F}_{\text{approx}}$, it holds that

$$\left| \mathbb{E}_{a \in \{0, 1\}^m} [g_{\tilde{f}}(a)] - \mathbb{E}_{b \in \{0, 1\}^r} [g_{\tilde{f}}(G(b))] \right| \leq \epsilon.$$

Since G is an ϵ -PRG for $\mathcal{F}_{\text{error}}$, it holds that

$$\left| \mathbb{E}_{b \in \{0, 1\}^r} [\tilde{f}(G(b))] - \mathbb{E}_{b \in \{0, 1\}^r} [g_{\tilde{f}}(G(b))] \right| \leq \mathbb{E}_{b \in \{0, 1\}^r} [h_{\tilde{f}}(G(b)) + 1] \leq$$

$$\mathbb{E}_{a \in \{0, 1\}^m} [h_{\tilde{f}}(a) + 1] + \epsilon = 2 \Pr_{a \in \{0, 1\}^m} [h_{\tilde{f}}(a) = 1] + \epsilon = O(\epsilon),$$

where the first inequality is due to Lemma A.12. The last equality is due to Lemma A.13. The same argument also implies that

$$\left| \mathbb{E}_{a \in \{0, 1\}^m} [\tilde{f}(a)] - \mathbb{E}_{a \in \{0, 1\}^m} [g_{\tilde{f}}(a)] \right| = O(\epsilon).$$

The result now follows by the triangle inequality. □

We are now ready to prove Theorem A.2.

Proof. (of Theorem A.2) By Theorem A.1, there exists an explicit ϵ -PRG $G : \{0, 1\}^r \rightarrow \{0, 1\}^m$ for $\text{space}(s')$, where $s' = O(\log(t)/\epsilon)$, $\mathcal{F}_{\text{approx}}, \mathcal{F}_{\text{error}} \subseteq \text{space}(s')$ and $r = O(\log^2(t/\epsilon))$. As this is an ϵ -PRG for both $\mathcal{F}_{\text{approx}}$ and $\mathcal{F}_{\text{error}}$ it is also an $O(\epsilon)$ -PRG for $\mathcal{F}_{\text{restricted}}$. Let $P \subseteq \mathbb{R}^t$ be the set of vectors corresponding to the strings in the image of G . Namely,

$$P \triangleq \{x(G(b)) \mid b \in \{0, 1\}^r\} .$$

Notice that the size of P is $|P| = \exp(O(\log^2(t/\epsilon)))$. Also notice that for any digon indicator function $f_{v,u}$ (i.e., $f_{v,u} \in \mathcal{F}_{\text{digon}}$),

$$|\mathbb{E}_{x \in S}[f_{v,u}(x)] - \mathbb{E}_{x \in P}[f_{v,u}(x)]| = \left| \mathbb{E}_{a \in \{0,1\}^m}[\tilde{f}_{v,u}(a)] - \mathbb{E}_{b \in \{0,1\}^r}[\tilde{f}_{v,u}(G(b))] \right| = O(\epsilon).$$

From this and from Lemma A.10, it follows that P is an $O(\epsilon)$ -sample for digons w.r.t. the measure defined by $\mathcal{N}(0, 1/t)^t$. Namely, that

$$|\mathbb{E}_{x \sim \mathcal{N}(0,1/t)^t}[f(x)] - \mathbb{E}_{x \in P}[f(x)]| = O(\epsilon)$$

The proof of the theorem follows from this and Lemma A.3. □

A.3 Samples for Spherical Caps

The construction of a sample space for spherical caps is essentially the same as in the case of digons. Lemma 4.1 is an immediate consequence of the following theorem.

Theorem A.15. *Let t be an integer and let $0 < \epsilon < 1$. There exists an efficiently constructible set $Q \subseteq \mathbb{R}^t$ such that Q is an $O(\epsilon + 1/t)$ -sample for digons w.r.t. the uniform measure over the unit sphere and $|Q| = \exp(O(\log^2(t/\epsilon)))$.*

Proof sketch: As in the case of digons, we first establish a finite, yet large sample space for spherical caps w.r.t. the Gaussian distribution. The family $\mathcal{F}_{\text{restricted}}$ of ‘restricted linear threshold functions’ will be defined as functions of the form $\tilde{f}_{v,\theta} : \{0, 1\}^m \rightarrow \{-1, 1\}$ where v is a unit vector, $\theta \in [-1, 1]$ and for a string $a \in \{0, 1\}^m$, $\tilde{f}_{v,\theta}(a) = \text{sign}(\langle x(a), v \rangle - \theta)$. For any $\tilde{f} = \tilde{f}_{v,\theta} \in \mathcal{F}_{\text{restricted}}$ we define an ‘approximation function’ $g_{\tilde{f}} : \{0, 1\}^m \rightarrow \{-1, 1\}$ as $g_{\tilde{f}}(a) = \text{sign}(\langle x(a), \hat{v} \rangle - \hat{\theta})$ and an ‘error function’ $h_{\tilde{f}} : \{0, 1\}^m \rightarrow \{-1, 1\}$ such that $h_{\tilde{f}}(a) = -1$ iff $|\langle x(a), \hat{v} \rangle - \hat{\theta}| \geq \epsilon/\sqrt{t}$. The rest of the analysis is analogous to the case of digons. □

B A Simple Norm Preserving Set

In this section we present a (sketch of a) simpler derandomization of the JL lemma than the one given in Section 3, that was communicated to us by Jelani Nelson. The construction is based on [AMS99] that gave an algorithm for approximating the L_2 norm of a vector in the streaming model. We construct a set \mathcal{A} of linear embeddings from \mathbb{R}^n to \mathbb{R}^t which preserve

the norm of any fixed vector by ϵ with probability $1 - \gamma$. Namely, for any fixed unit vector x ($\|x\|_2 = 1$)

$$\Pr_{A \in \mathcal{A}} [|\|Ax\|_2 - 1| > \epsilon] < \gamma.$$

The output length is $t = O(\epsilon^{-2}\gamma^{-1})$. In order to further reduce the output length to $k = O(\log(\gamma^{-1})\epsilon^{-2})$ as in the randomized constructions, we use the same technique as in Section 3 and apply another norm preserving set of transformations, which reduces the length of the vectors from t to k , of size $\exp(O(\log^2(t)))$ (see Lemma 3.9).

\mathcal{A} will consist of sign matrices whose rows form a pairwise independent sample space over a 4-wise independent sample space over $\{-1, 1\}$. We begin with the definition of k -wise independent sample spaces.

Definition B.1. *Let S be an arbitrary set and let I be a multiset in S^n . I is called a k -wise independent sample space over S when any $j \in [k]$, $1 \leq i_1 < \dots < i_j \leq n$ and $s_1, \dots, s_j \in S$, satisfy that*

$$|\{x \in I | (x_{i_1}, \dots, x_{i_j}) = (s_1, \dots, s_j)\}| = |I|/|S|^j.$$

There are many methods for obtaining a k -wise independent sample space. For the case where $S = \{-1, 1\}$, it can be obtained via standard BCH codes (see e.g. [AS08], Chapter 16). For arbitrary S , it can be obtained via evaluating polynomials over finite fields (see e.g. [LW05]).

Lemma B.2. *There exists a polynomial time algorithm constructing a k -wise independent sample space I in $\{-1, 1\}^n$ of size $O(n^{k/2})$. For a sample space in S^n where $|S| = n$, there exists an explicit construction of a sample space I of size $O(n^k)$.*

We note that both constructions have optimal size, up to constant factors. Let I_1 be a 4-wise independent sample space in $\{-1, 1\}^n$ and let I_2 be a pairwise independent sample space in $(I_1)^t$. The set \mathcal{A} is defined as the set of matrices corresponding to the elements of I_2 , after scaling by a factor of $1/\sqrt{t}$. Indeed, each element of I_2 is actually a t -dimensional vector whose entries are themselves vectors in $\{-1, 1\}^n$. What we do is scale each entry in I_2 by $1/\sqrt{t}$, and thus we can naturally identify elements of I_2 with $t \times n$ matrices whose entries are in $\{-\frac{1}{\sqrt{t}}, \frac{1}{\sqrt{t}}\}$. Clearly, $|\mathcal{A}| = |I_2| = \Theta(|I_1|^2) = \Theta(n^4)$.

Theorem B.3. *\mathcal{A} is a (γ, ϵ) -norm preserving set for $2/(\gamma\epsilon^2) \leq t$. It is of cardinality $|\mathcal{A}| = \Theta(n^4)$.*

Proof. Let $x \in \mathbb{R}^n$ be some fixed unit vector. We now analyze the second and fourth moments of $\langle a, x \rangle$, where a is distributed as follows. First we pick A uniformly at random from \mathcal{A} . Denote by a_1, \dots, a_t the rows of A . Now pick a uniformly at random from the rows of A .

$$\mathbb{E} [\langle a, x \rangle^2] = \sum_{j_1, j_2=1}^n x_{j_1} x_{j_2} \mathbb{E}[a_{j_1} a_{j_2}] = \frac{1}{t} \sum_{j=1}^n x_j^2 = 1/t.$$

As for the fourth moment,

$$\mathbb{E} [\langle a, x \rangle^4] = \sum_{j_1, j_2, j_3, j_4=1}^n x_{j_1} x_{j_2} x_{j_3} x_{j_4} \mathbb{E}[a_{j_1} a_{j_2} a_{j_3} a_{j_4}] =$$

$$\frac{1}{t^2} \left(\sum_{j=1}^n x_j^4 + \binom{4}{2} \sum_{j_1 < j_2} x_{j_1}^2 x_{j_2}^2 \right) \leq \frac{3}{t^2} \left(\sum_{j=1}^n x_j^2 \right)^2 = 3/t^2.$$

It follows that

$$\mathbb{E}_A [\|Ax\|_2^2] = \sum_{i=1}^t \mathbb{E}[\langle a_i, x \rangle^2] = 1$$

and,

$$\mathbb{E} [\|Ax\|_2^4] = \sum_{i_1 \neq i_2 \in [t]} \mathbb{E} [\langle a_{i_1}, x \rangle^2] \mathbb{E} [\langle a_{i_2}, x \rangle^2] + \sum_{i \in [t]} \mathbb{E} [\langle a_i, x \rangle^4] \leq \frac{t(t-1)}{t^2} + \frac{3}{t} = 1 + \frac{2}{t}.$$

Hence,

$$\text{Var} [\|Ax\|_2^2] = \mathbb{E} [\|Ax\|_2^4] - \mathbb{E} [\|Ax\|_2^2]^2 = \mathbb{E} [\|Ax\|_2^4] - 1 \leq 2/t.$$

We now apply Chebyshev's inequality.

$$\Pr [|\|Ax\|_2 - 1| > \epsilon] \leq \Pr [|\|Ax\|_2^2 - 1| > \epsilon] \leq \text{Var} [\|Ax\|_2^2] / \epsilon^2 < 2/(t\epsilon^2) \leq \gamma$$

□

C Averaging Sampler

The goal of this section is to explain how Lemma 3.6 is obtained from [Zuc97, GUV09]. We begin by defining an averaging sampler in its most general form, as defined in [Zuc97].

Definition C.1. An $(n, m, t, \gamma, \epsilon)$ -averaging sampler is a deterministic algorithm which, on input of a uniformly random n bit string, outputs a sequence of t sample points $z_1, \dots, z_t \in \{0, 1\}^m$ such that for any function $f : \{0, 1\}^m \rightarrow [0, 1]$, we have

$$\left| \frac{1}{t} \sum_{i=1}^t f(z_i) - \mathbb{E}[f] \right| \leq \epsilon$$

with probability $\geq 1 - \gamma$.

We now explain the notion of an extractor. An extractor is a function that receives as input two instances of random variables. One is a long string that is “weakly random” and the other is a short string of i.i.d. random bits (independent of the first string). It outputs a long string of bits whose distribution is close to being completely uniform. That is, by using the short truly random string it extracts the randomness out of the long weakly random string. We start by formally defining what a weakly random string is.

Definition C.2. A distribution D on $\{0, 1\}^n$ is called a δ -source if for all $x \in \{0, 1\}^n$, $\Pr_{X \sim D}[X = x] \leq 2^{-\delta n}$.

Definition C.3. $E : \{0, 1\}^n \times \{0, 1\}^s \rightarrow \{0, 1\}^m$ is an $(n, m, s, \delta, \epsilon)$ -extractor if, for x chosen according to any δ -source on $\{0, 1\}^n$ and y chosen uniformly at random from $\{0, 1\}^s$, $E(x, y)$ is within statistical distance ϵ from the uniform distribution on $\{0, 1\}^m$.

The following Theorem by [Zuc97] shows an equivalence between both objects:

Theorem C.4 (Proposition 2.7 in [Zuc97]). *If there is an efficient $(n, m, s, \delta, \epsilon)$ -extractor, then there is an efficiently constructible $(n, m, 2^s, 2^{1-(1-\delta)n}, \epsilon)$ -averaging sampler.*

For our purposes we need an extractor where m is very close to δn . Specifically, for some $\xi > 0$ we require $m = (1 - \xi)\delta n$. We use a result by [GUV09] giving a construction of such an extractor. As it is not formally stated for the parameters we require, we cite the required lemmas to prove the needed result.

Lemma C.5 (Theorem 4.17 in [GUV09]). *For all positive n and all $1 > \delta, \epsilon > 0$, there is an explicit construction of an $(n, m, s, \delta, \epsilon)$ extractor where $m = \lceil \delta n / 2 \rceil$ and $s = O(\log(n) + \log(1/\epsilon))$.*

Lemma C.6 (Lemma 4.18 in [GUV09]). *Suppose E_1 is an $(n, m_1, s_1, \delta_1, \epsilon_1)$ extractor and E_2 is an $(n, m_2, s_2, \delta_2, \epsilon_2)$ extractor. Let r be an integer such that $\delta_2 n \leq \delta_1 n - m - r$. Then $E'(x, (y_1, y_2)) \stackrel{\Delta}{=} E_1(x, y_1) \circ E_2(x, y_2)$ is a $(n, m_1 + m_2, s_1 + s_2, \delta_1, (1/(1 - 2^{-r}))\epsilon_1 + \epsilon_2)$ extractor (where the \circ product is a concatenation of two strings).*

The extractor that we need is given by the following theorem. It was proved in [GUV09] for constant ξ (i.e., $\xi = \Omega(1)$). We require a version for general $\xi > 0$.

Theorem C.7 (Modification of Theorem 4.19 in [GUV09]). *For all integers n and $1 > \epsilon, \delta > 0$ and any $1 \geq \xi > 2/(\delta n)$ there exists an efficiently constructible $(n, m, s, \delta, \epsilon)$ -extractor, with $s = O(\log(\xi^{-1})(\log(n) + \log(1/\epsilon)))$ and $m = (1 - \xi)\delta n$.*

Proof. Similarly to [GUV09], the proof of the theorem follows by applying Lemma C.6 $O(\log(\xi^{-1}))$ times with both extractors being taken from Lemma C.5. Let $E_1^{(1)}$ be the $(n, m_1^{(1)}, s_1^{(1)}, \delta, \epsilon \cdot \xi^C)$ extractor given in Lemma C.5 where C is some sufficiently large constant. For any integer $i > 0$ let $E_2^{(i)}$ be the $(n, m_2^{(i)}, s_2, \delta - (m_1^{(i)} + 1)/n, \epsilon \cdot \xi^C)$ -extractor given in Lemma C.5. For $i > 1$, let $E_1^{(i)}$ be the $(n, m_1^{(i)}, s_1^{(i)}, \delta, \epsilon_1^{(i)})$ extractor obtained by combining $E_1^{(i-1)}$ and $E_2^{(i-1)}$ as in Lemma C.6. Namely, for $x \in \{0, 1\}^n$, $y_1 \in \{0, 1\}^{s_1^{(i-1)}}$, $y_2 \in \{0, 1\}^{s_2}$, $E_1^{(i)}(x, (y_1, y_2)) \stackrel{\Delta}{=} E_1^{(i-1)}(x, y_1) \circ E_2^{(i-1)}(x, y_2)$. Calculating, we get

$$s_1^{(i)} = s_1^{(i-1)} + s_2 = s_1^{(1)} + (i - 1)s_2 = O(i(\log(n) + \log(1/\epsilon) + \log(1/\xi))) = O(i(\log(n) + \log(1/\epsilon))) .$$

The last equality holds since $\xi > 1/n$. As for $\epsilon_1^{(i)}$,

$$\epsilon_1^{(i)} = 2\epsilon_1^{(i-1)} + \epsilon \cdot \xi^C = 2^{O(i)} \cdot \epsilon \cdot \xi^C .$$

Finally,

$$m_1^{(i)} = m_1^{(i-1)} + m_2^{(i-1)} \geq m_1^{(i-1)} + n\delta/2 - (m_1^{(i-1)} + 1)/2 = (\delta n - 1 + m_1^{(i-1)})/2 ,$$

meaning that

$$\delta n - 1 - m_1^{(i)} \leq (\delta n - 1 - m_1^{(1)})/2^{i-1} .$$

It follows that for some $i = O(\log(\xi^{-1}))$, $m_1^{(i)} \geq (1 - \xi)\delta n$. Hence, by taking the first $(1 - \xi)\delta n$ bits of the output of $E_1^{(i)}$ and assigning a sufficiently large value for C , we get an $(n, m, s, \delta, \epsilon)$ -extractor with $m = (1 - \xi)\delta n$ and $s = O(\log(\xi^{-1})(\log(n) + \log(1/\epsilon)))$ as required. \square

By picking $\delta = 1 - \xi$ and combining the results of [GUV09] and [Zuc97] we get

Lemma C.8. *For any integer m and $1 > \xi > 2/m$ there exists an efficiently constructible $(m/(1 + \xi)^2, m, (m/\epsilon)^{O(\log(\xi^{-1}))}, 2^{1 - \frac{\xi m}{(1 - \xi)^2}}, \epsilon)$ -averaging sampler.*

Lemma 3.6 is obtained by setting $d = 2^m$.