Electronic Colloquium on Computational Complexity, Report No. 121 (2009)

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 $\mathbb{R}^{O(\log n)}$, such that the L_2 norm of any vector in $\mathbb{R}^n$ is distorted by at most $1 \pm o(1)$ in at least a fraction of $1 - o(1)$ of the transformations in the set. Albeit the tradeoff between the distortion 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 spherical digons in $\mathbb{S}^{n-1}$, for $\epsilon = o(1)$. This construction leads to an 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). We also construct an ϵ-sample for linear threshold functions on $\mathbb{S}^{n-1}$, for $\epsilon = o(1)$.

*Faculty of Computer Science, Technion, Haifa 32000, Israel. Email: {zkarnin,shpilka}@cs.technion.ac.il. Research supported by the Israel Science Foundation (grant number 439/06).

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

ISSN 1433-8092

1 Introduction

In this paper we construct a small set of explicit dimension reducing linear transformations mapping vectors in ℓ_2^n into vectors in ℓ_2^d, for $d \ll n$. The celebrated Johnson-Lindenstrauss Lemma [JL84] states the following. In any Hilbert space, a random linear mapping into ℓ_2^d preserves the norm of any vector up to a factor of $1 \pm \epsilon$ with probability at least $1 - \exp(-\epsilon^2 d)$. In fact, quite simple sample spaces suffice for points in ℓ_2^n; see [DG99, Ach03, Mat08]. Thus, in order to preserve approximately all pairwise distances among n points in 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., [JN09] 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 [AC06]). 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 [AC06, AL08]. The probabilistic constructions can be derandomized using the method of conditional expectations [EIO02, Siv02]. However, there is no construction that uses a $\mathrm{poly}(n)$ 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 $\Omega((\log n)^{3/2})$ random bits [SZ99]. A construction using $O(\log n)$ random bits would yield a fixed collection of $\mathrm{poly}(n)$ mappings that contains, for every configuration of n points, a JL transform for that configuration.

We construct a set $\mathcal{A}_{n,d}$ of linear mappings $A : \ell_2^n \to \ell_2^d$ of cardinality

$$|\mathcal{A}_{n,d}| = n^{1+o(1)} \cdot 2^{O(\log^{3/2} d)}$$

(note that if $d = \exp(o(\log^{2/3} n))$, then $|\mathcal{A}_{n,d}| = n^{1+o(1)}$). We show that $\mathcal{A}_{n,d}$ satisfies the following.

Theorem 1.1. *There is a constant $c > 0$ such that for every n and for every d, for every vector $x \in \ell_2^n$, a fraction of at least $1 - d^{-c}$ of $A \in \mathcal{A}_{n,d}$ satisfies that*

$$(1 - d^{-c}) \cdot \|x\|_2 \leq \|Ax\|_2 \leq (1 + d^{-c}) \cdot \|x\|_2 \,.$$

Even though our explicit construction falls short of providing a $\mathrm{poly}(n)$-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 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 $F \in \mathcal{F}$,

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

In our case, Ω is the unit sphere $\mathbb{S}^{n-1}$, endowed with the uniform (Haar) 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 $u, v \in \mathbb{S}^{n-1}$. It is easy to show that sampling $O(n/\epsilon)$ points i.i.d. from μ gives an ϵ-sample with high probability. We construct, for every $\epsilon > 0$, a set $P_\epsilon \subset \mathbb{S}^{n-1}$ of cardinality $|P_\epsilon| = n^{1+o(1)} \cdot 2^{O(\log^{3/2}(1/\epsilon))}$. (Notice that for $\epsilon > 1/\exp(o(\log^{2/3} n))$, P_ϵ has nearly linear size.) We prove the following theorem.

Theorem 1.2. *P_ϵ is an ϵ-sample for spherical digons.*

Our methods also give 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 measure on $\mathbb{S}^{n-1}$, 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 $\theta \in \mathbb{R}$. In this case, too, it is easy to show that sampling $O(n/\epsilon)$ points i.i.d. from μ gives an ϵ-sample with high probability. We construct, for every $\epsilon > 0$, a set $Q_\epsilon \subset \mathbb{R}^n$ of cardinality $|Q_\epsilon| = n^{1+o(1)} \cdot 2^{O(\log^{3/2}(1/\epsilon))}$, and prove the following.

Theorem 1.3. *Q_ϵ is an ϵ-sample for halfspaces.*

Hitting sets (or ϵ-nets) of size $\text{poly}(n/\epsilon)$ for linear threshold functions on the binary cube and on the unit sphere were recently constructed in [RS09]. Also recently, Diakonikolas et al. [DGJ+09] constructed an ϵ-sample of cardinality $n^{O(\epsilon^{-2}\log^2(1/\epsilon))}$ for linear threshold functions on the binary cube. To the best of our knowledge, our paper presents the first construction of a non-trivial ϵ-sample for linear threshold functions on the unit sphere. In addition, the size of our ϵ-sample is significantly smaller than that of [DGJ+09], and in particular is nearly linear for $\epsilon > 1/\exp(o(\log^{2/3} n))$. Recently, [MZ09] gave explicit constructions of ϵ-samples (a.k.a. pseudo-random-generators) for threshold functions of degree d polynomials, over the boolean cube. For $\epsilon = o(1)$ their ϵ-sample has a polynomial size when $d = 1$ (i.e. the underlying function is a linear threshold function) and $\epsilon > 1/\text{poly}(\log n)$. For a constant ϵ, their construction has polynomial size for every d (the size is $n^{1/\epsilon^{O(d)}}$).

Construction of ϵ-samples (often called pseudo-random generators) 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, 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 using the method of conditional expectations in conjunction with Nisan's [Nis92] deterministic simulation of RL [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.[1] Henceforth we refer to such a derandomization

[1] Because 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.

as an *oblivious* derandomization.

In order to understand the connection between Theorem 1.2 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 constructs a cut by choosing a vector $x \in \mathbb{S}^{n-1}$ uniformly at random, then separates the mapped nodes by the hyperplane through the origin 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.

To illustrate our proof techniques, we 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 projections of $\mathbb{S}^{n-1}$ onto a lower dimensional space $\mathbb{R}^d$. These projections, stipulated by Theorem 1.1, are generated as follows. Using a construction of Indyk [Ind07], we first embed $\mathbb{S}^{n-1}$ 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. (In other words, we embed $\mathbb{S}^{n-1}$ into an n-dimensional almost Euclidean linear subspace of ℓ_1^N.) We then produce samples of d coordinates of the image using known sampling techniques [Gol97]. Each sample, properly scaled, gives a projection of $\mathbb{S}^{n-1}$ onto $\mathbb{R}^d$. Such a projection preserves ℓ_2 distances approximately, and therefore it preserves angles approximately, with high probability over the choice of sample. Due to the low dimension, we can produce in $\mathbb{S}^{d-1}$ a poly(n)-size ϵ-sample for spherical digons using a pseudo-random generator for space bounded computation [Nis92, SZ99]. Our construction lifts this ϵ-sample for spherical digons in low dimension 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 $\mathbb{R}^d$. The ϵ-sample for spherical digons in $\mathbb{S}^{n-1}$ is composed essentially of the entire collection of lifted low dimensional sample points.

1.1 Organization

In Section 3.1 we prove Theorem 1.1. We then use it to give an ϵ-sample for digons in Section 4.1, thus proving Theorem 1.2. In Section 5 we show how to derandomize the Goemans-Williamson algorithm using the ϵ-sample for digons. In section 5.2 we use similar techniques to derandomize the graph coloring algorithm of [KMS98]. In Section 6 we give an ϵ-sample for linear threshold functions (Theorem 1.3).

2 Preliminaries

For $n \in \mathbb{N}$ denote $[n] \stackrel{\Delta}{=} \{1, \ldots, n\}$. For $x \in \mathbb{R}^n$ and a subset $S = \{i_1, i_2, \ldots, i_{|S|}\} \subseteq [n]$ define x_S as the restriction of x to the indices of S. That is, $x_S \stackrel{\Delta}{=} (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S|}})$ where

$i_1 < i_2 < \ldots < i_{|S|}$. 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 $\alpha, \beta \in \mathbb{R}^n$ define $\sphericalangle(\alpha, \beta)$ to be the angle between them (the angle ranges between $-\pi$ 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 J-L 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 $t^{-\Omega(1)}$, in almost all of the linear transformations in $\mathcal{A}$. Our methods work for any $t = \exp\left(O\left(\sqrt{\log(n)}\right)\right)$. The size of $\mathcal{A}$ grows with t. However, for any $t = \exp\left(o\left(\sqrt{\log(n)}\right)\right)$, we get $|\mathcal{A}| = n^{1+o(1)}$.

Remark 3.1. *We can prove essentially the same results for any $t = \exp\left(O\left(\log^{2/3}(n)\right)\right)$. To achieve this however, we need to reprove Corollary 5.2 of [Ind07], using an improved version of Theorem 6 of [Ind06] that relies on the small space generator of [SZ99] and not on that of [Nis92]. Here, for simplicity, we use the generator of Nisan and to avoid reproving Corollary 5.2 of [Ind07].*

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

Definition 3.2. *A set $\mathcal{A}$ of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^t$ is called (γ, δ)-norm preserving when for any unit vector $\alpha \in \mathbb{S}^{n-1}$ it holds that*

$$\Pr_{A \in \mathcal{A}}\left[\left|||A\alpha||_2^2 - 1\right| > \delta\right] < \gamma.$$

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

We construct a $\left(t^{-\Omega(1)}, t^{-\Omega(1)}\right)$-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

Intuitively, a vector that is well spread should have a large ratio between its ℓ_1 and ℓ_2 norms (i.e. close to the square root of the dimension). We define the notion of *distortion* of an embedding and then see its relation to the spreadness of the vectors in the image.

Definition 3.3. *Let $n \leq N$ be integers and $\epsilon > 0$. The distortion of $F : \ell_2^n \to \ell_1^N$ is defined as*

$$\Delta(F) \triangleq \sqrt{m} \max_{0 \neq x \in l_2^n} \frac{||F(x)||_2}{||F(x)||_1}.$$

The distortion of an n-dim subspace $V \subseteq \ell_1^N$ is the distortion of the embedding from ℓ_2^n to it.

The next lemma proves that vectors in the image of a low distortion embedding are well spread.

Lemma 3.4. *Let $V \subseteq \mathbb{R}^N$ be a subspace and $\rho \triangleq 1 - 1/\Delta(V)$. Let $x \in V$ be a unit vector. Let $S \in [N]$ be of size $|S| \leq \rho N$. Then it holds that $||x_S||_2^2 \leq 8\rho$.*

Proof. First notice that $||x_S||_1 \leq \sqrt{|S|}||x_S||_2 \leq \sqrt{\rho N}||x_S||_2$. Now, for $\bar{S} \triangleq [N] \setminus S$,

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

We get that

$$1 - \rho - \sqrt{\rho} \cdot \sqrt{1 - ||x_{\bar{S}}||_2^2} \leq ||x_{\bar{S}}||_2.$$

Viewing this as a degree 2 polynomial in $||x_S||_2$ leads to the inequality $||x_S||_2^2 \leq 1 - (1 - 4\rho)^2 \leq 8\rho$. □

Using the low distortion embedding of [Ind07] we get that almost all the entries, of any unit vector in this low distortion subspace, are bounded by $O(\frac{1}{\sqrt{N}})$.

Theorem 3.5. *(Corollary 5.2 of [Ind07]) Let $n = 2^{2k}$ where $k > 0$ is some integer[2]. Let $\epsilon > 0$. There exists an explicit embedding $F : \ell_2^n \to \ell_1^N$ with distortion $\Delta(F) \leq 1 + \epsilon$ where $N = N(n, \epsilon) = n^{1+o(1)} \exp\left(O\left(\log^2\left(\epsilon^{-1}\right)\right)\right)$. Furthermore, for any $x \in \mathbb{R}^n$ it holds that $||Fx||_2 = ||x||_2$.*

Corollary 3.6. *Let $n > 0$ be an integer and $\rho > 0$. There exists an explicit embedding $F : \mathbb{R}^n \to \mathbb{R}^N$ with the following properties: $N = n^{1+o(1)} \exp\left(-O\left(\log^2\left(\rho^{-1}\right)\right)\right)$. For any $x \in \mathbb{R}^n$ it holds that $||Fx||_2 = ||x||_2$. Let $i_1, \ldots, i_N$ be a permutation of $[N]$ s.t. $|(Fx)_{i_1}| \geq |(Fx)_{i_2}| \geq \ldots \geq |(Fx)_{i_N}|$. Then $|(Fx)_{i_{\rho N}}| \leq \frac{\sqrt{8}||Fx||_2}{\sqrt{N}}$ and $\left|\left|(Fx)_{[\rho N]}\right|\right|_2 \leq \sqrt{8\rho}||Fx||_2$.*

Proof. Let F be an embedding of distortion $\Delta(F) = \frac{1}{1-\rho}$ as in Theorem 3.5. By Lemma 3.4 we get

$$|(Fx)_{i_{\rho N}}|^2 \leq \frac{||(Fx)_{[\rho N]}||_2^2}{\rho N} \leq \frac{8\rho||Fx||_2^2}{\rho N}.$$

□

[2]Notice 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 in some arbitrary way and achieve an embedding to $N(n', \epsilon)$ with distortion $1 + \epsilon$.

3.2 Samplers

The previous section gave an embedding F that spreads the coordinates of any nonzero vector. We shall now use this map in order to reduce the dimension while still keeping the ℓ_2 norm. This can be done by taking several different projections of F to subsets of the coordinates. In order to pick these subsets we shall use a combinatorial object called a sampler, whose main property (from our perspective) is that it can be thought of as a tool to estimate well the expectation of any bounded function using a small number of queries (that are independent of f). More accurately, samplers for functions from $[N]$ to $[0,1]$ compute a subset of $[N]$. They estimate the average of each $f : [N] \to [0.1]$ by its average on the subset. Clearly, a deterministic sampler would require $\Omega(N)$ 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 samplers see [Gol97].

Lemma 3.7 (Good sampler)**.** *[Gol97] Let N be an integer and $\epsilon, \gamma > 0$. There exists an explicit family $\mathcal{T}$ of subsets of $[N]$, each of size $t = O\left(\epsilon^{-2}\log(\gamma^{-1})\right)$ with the following properties: $\mathcal{T}$ is of cardinality $|\mathcal{T}| = N \cdot \gamma^{-O(1)} \cdot \epsilon^{-O(1)}$. For any $f : [N] \to [0,1]$,*

$$\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T} f(i) - \mathbb{E}_{i\in[N]}[f(i)]\right| > \epsilon\right] < \gamma.$$

Furthermore, for any $i_1, i_2 \in [N]$, $\Pr_{T\in\mathcal{T}}[i_1 \in T] = \Pr_{T\in\mathcal{T}}[i_2 \in T]$.

We would like to use samplers in order to project F into a subspace of $\mathbb{R}^N$ (the samplers will choose the coordinates to project on). For this we will define for every vector $x \in \mathbb{R}^n$ a function $f_x : [N] \to \mathbb{R}$ by $f_x(i) = N \cdot F(x)_i^2$. By Lemma 3.4, f_x is usually at most 8 (that is, it is almost a function from $[N]$ to $[0,8]$). However, it may obtain large values on some elements of $[N]$. Nevertheless, it is not difficult to see that the expectation of f over $[N]$ is almost equal to its expectation over the points in which $f_x \in [0,8]$. As this happens most of the time we can still estimate $\mathbb{E}[f_x]$ by using the sampler. To formalize this property, we say that a function $f : [N] \to \mathbb{R}$ is η-bounded in a segment $I \subseteq \mathbb{R}$ when the following holds: $\Pr_{i\in[N]}[f(i) \in I] \geq 1 - \eta$ and $\mathbb{E}_{i\in[N]}[f(i)] = \mathbb{E}_{i\in[N]}[f(i)|f(i) \in I] \pm \eta$.

Theorem 3.8. *Let N be an integer and $\delta, \gamma, \eta > 0$ where $\eta < \delta^2$. Let $f : [N] \to \mathbb{R}$ be some η-bounded function in some segment I. Let $\mathcal{T}$ be the family defined in Lemma 3.7 for N, $\epsilon = \delta - \eta$ and γ. Then,*

$$\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T} f(i) - \mu\right| > \delta\right] < \gamma + O\left(\eta\delta^{-2}\log(\gamma^{-1})\right).$$

Proof. Denote by μ the expectation of f over a uniform distribution on $[N]$. Define $S \subset [N]$ as the points in which $f(i) \notin I$. Since f is η-bounded in I we have that $\mu_{\bar{S}} \stackrel{\Delta}{=} E_{i\notin S}[f(i)] = \mu \pm \eta$. Let $g : [N] \to I$ 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

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

Now, Lemma 3.7 and the fact that the expectation of g is $\mu_{\bar{S}}$ imply that

$$\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T}g(i)-\mu_{\bar{S}}\right|>\delta\right]\leq\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T}g(i)-\mu\right|>\delta-\eta\right]<\gamma.$$

Finally, by the union bound we obtain

$$\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T}f(i)-\mu\right|>\delta\right]\leq\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T}g(i)-\mu\right|>\delta\right]+\Pr_{T\in\mathcal{T}}\left[\exists i\in T\text{ s.t. }f(i)\neq g(i)\right]<$$

$$\gamma+\eta t=\gamma+O\left(\eta(\delta-\eta)^{-2}\log(\gamma^{-1})\right)=\gamma+O\left(\eta\delta^{-2}\log(\gamma^{-1})\right).$$

□

For simplicity, we restate the theorem for the parameters that we shall later use.

Corollary 3.9. *Let N,t be integers where $t<N$. There exists a universal constant c such that if $t>c$, then there exists an explicit family $\mathcal{T}$ of subsets of $[N]$ with the following properties: $\mathcal{T}$ is of size $|\mathcal{T}|=N\cdot t^{O(1)}$. Each member of $\mathcal{T}$ is a set of size t. There exists some $\delta=O\left(\log^{1/3}t/\sqrt{t}\right)$ s.t. for any function $f:[N]\to\mathbb{R}$ which is $t^{-1.5}$-bounded in $[0,1]$ it holds that*

$$\Pr_{T\in\mathcal{T}}\left[\left|\frac{1}{t}\sum_{i\in T}f(i)-\mathbb{E}_{i\in[n]}[f(i)]\right|>\delta\right]<\delta.$$

3.3 The Norm Preserving Set

Let $F:\mathbb{R}^n\to\mathbb{R}^N$ be the linear transformation guaranteed in Theorem 3.5 satisfying $\Delta(F)\leq 1/(1-t^{-1.5})$. Let $\mathcal{T}$ be the family of subsets of $[N]$ guaranteed by corollary 3.9. For every $T\in\mathcal{T}$ define $A_T:\mathbb{R}^N\to\mathbb{R}^t$ as the projection to the indices of T. I.e., $A_T(x)=x_T$. The set $\mathcal{A}$ is defined as

$$\mathcal{A}\triangleq\left\{\sqrt{N/t}\cdot A_T\cdot F\;\middle|\;T\in\mathcal{T}\right\}.$$

Theorem 1.1 is an immediate consequence of the next theorem.

Theorem 3.10. *Let $t>0$. The set $\mathcal{A}$ (defined w.r.t. t) is $\left(t^{-\Omega(1)},t^{-\Omega(1)}\right)$-norm preserving. Its cardinality is $|\mathcal{A}|=n^{1+o(1)}\cdot\exp\left(O\left(\log^2(t)\right)\right)$.*

Proof. The claim regarding $|\mathcal{A}|$ stems directly from Theorem 3.5 and corollary 3.9. Let $\alpha\in\mathbb{R}^n$ be some fixed vector. Assume w.l.o.g. that $||\alpha||_2=1$. Let $f:[N]\to\mathbb{R}$ be the function $f(i)\triangleq N\cdot((F(\alpha))_i)^2$. Notice that the expectation of f is equal to $||F(\alpha)||_2^2=||\alpha||_2^2=1$. Corollary 3.6 shows that

$$\Pr_{i\in[N]}\left[f(i)\notin[0,8]\right]<1-\frac{1}{\Delta(F)}\leq t^{-1.5}.$$

We also have that for any $T \subseteq [N]$,

$$\left|\left|\sqrt{N/t} \cdot A_T F(\alpha)\right|\right|_2^2 = \frac{1}{|T|} \sum_{j \in T} f(j).$$

Hence, by the properties of $\mathcal{T}$ and Corollary 3.9 applied to the function $f/8$, we get that for some $\delta = O\left(\log^{1/3} t/\sqrt{t}\right) = t^{-\Omega(1)}$, it holds that $\Pr_{A \in \mathcal{A}}\left[\,|\,||A\alpha||_2^2 - 1\,|\, > \delta\right] < \delta$. □

4 A sample set for digons

In this section we prove Theorem 1.2. Namely, we present a method of constructing an ϵ-sample for digons. Recall that digons are characterized by functions of the following form:

$$f_{\alpha,\beta}(x) \triangleq \text{sign}(\langle \alpha, x \rangle) \cdot \text{sign}(\langle \beta, x \rangle)$$

where $\alpha, \beta, x \in \mathbb{S}^{n-1}$. In [GW95], it was shown that the expression $\mathbb{E}_{x \in \mathbb{S}^{n-1}}\left[f_{\alpha,\beta}(x)\right]$ relies only on the angle between α and β. 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 $t \ll n$. We then construct an ϵ-sample in $\mathbb{R}^t$ of quasi-polynomial (in t) size using a pseuso-random generator for log-space machines (see section 4.1). The parameter t will be set to a sufficiently small value so that the size of the low-dimensional ϵ-sample is polynomial in n.

4.1 Pseudo Random Generator for Bounded Space Machines

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

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

for every function $f \in \mathcal{F}$. In other words, G expands a seed of d truly random bits into m bits that look random to any function in $\mathcal{F}$. Notice that by taking $\mathcal{F}$ to be the family of all digon indicator functions, we get that the image of an ϵ-PRG for $\mathcal{F}$ is an ϵ-sample for digons. Denote by space(S) the family of functions that can be computed using at most S bits of space.

Theorem 4.1 ([Nis92])**.** *Let $\epsilon = 2^{-O(S)}$. Then there exists an explicit ϵ-PRG for* space(S)*, $G : \{0,1\}^{O(S \log(m))} \to \{0,1\}^m$.*

Applying the same ideas as [Siv02] we use such PRGs in order to fool digon indicator functions in low dimensions. Notice that such functions (i.e., $f_{\alpha,\beta}$) in $\mathbb{S}^{t-1}$ can be well approximated by rounding the entries of α, β to $O(\log(t))$ bits (i.e., storing only the $O(\log(t))$ significant bits of each entry in α and β). The resulting function will have the same expectation up to

a $1/\text{poly}(t)$ error. Now, this approximating function can be computed using $O(\log(t))$ bits of memory[3]. Hence, in order to "fool" digon indicator functions in low dimensional spaces, it suffices to use the pseudo-random generator of Nisan that fools bounded space machines. Note, that the above argument works without any change also for linear threshold functions. That is, any LTF can be approximated, with error of $1/\text{poly}(t)$, by a function requiring $O(\log(t))$ bits of memory. By using the notations of Theorem 4.1 and taking $m \approx t\log(t)$, $S \approx \log(t)$ we get the following corollary.

Corollary 4.2. *Let $\epsilon = t^{-1}$. There exists an explicit ϵ-PRG for indicator functions of digons (and for linear threshold functions) in $\mathbb{S}^{t-1}$, $G : \{0,1\}^{O(\log^2(t))} \to \{0,1\}^{O(t\log(t))}$, where the output of G is interpreted as a vector in $\mathbb{S}^{t-1}$ in which each entry is written using $O(\log t)$ bits.*

Rephrased in our original terms, the ϵ-sample for digons (and spherical caps) in low dimensional subspaces is the image of G. I.e., the set $\{G(y)\}_{y\in\{0,1\}^d}$ where $d = O\left(\log^2(t)\right)$.

Corollary 4.3. *Let t be some integer. There exist a (t^{-1})-sample for digons (and for spherical caps) in $\mathbb{S}^{t-1}$ of cardinality $t^{O(\log(t))}$ that can be constructed in $t^{O(\log(t))}$ time.*

4.2 Norm preserving implies angle preserving

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

Definition 4.4. *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 $\alpha,\beta \in \mathbb{S}^{n-1}$ it holds that*

$$\Pr_{A\in\mathcal{A}}\left[\left|\sphericalangle(\alpha,\beta) - \sphericalangle\left(A\alpha, A\beta\right)\right| > \delta\right] < \gamma.$$

For simplicity, we discuss only the case where $\mathcal{A}$ is (δ,δ)-norm preserving. We start by proving that for two unit vectors α,β it holds that $\cos\left(\sphericalangle\left(\alpha,\beta\right)\right) = \langle\alpha,\beta\rangle$ is roughly equal to $\langle A\alpha, A\beta\rangle$ which is roughly equal to $\cos\left(\sphericalangle\left(A\alpha, A\beta\right)\right)$.

Lemma 4.5. *Let $\alpha,\beta \in \mathbb{R}^n$ 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 $1-3\delta$*

$$\left|\langle A\alpha, A\beta\rangle - \langle\alpha,\beta\rangle\right| \leq 3\delta.$$

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

$$\left|||A\alpha||^2 - 1\right| < \delta, \quad \left|||A\beta||^2 - 1\right| < \delta, \quad \left|||A\alpha - A\beta||^2 - ||\alpha-\beta||^2\right| < \delta||\alpha-\beta||^2 \leq 4\delta\ .$$

[3]Specifically, the inner products $\langle\alpha,x\rangle$ and $\langle\beta,x\rangle$ need to be computed. Each of these expressions can be computed by keeping an index of the indices ($\log(t)$ bits), the sum so far, up to $1/\text{poly}(t)$ precision ($O(\log(t))$ bits) and computing each product $\alpha_i \cdot x_i$, up to $1/\text{poly}(t)$ precision, separately $O(\log(t))$ bits

It follows that $||\alpha-\beta||^2 = \langle \alpha-\beta, \alpha-\beta\rangle = \langle\alpha,\alpha\rangle - 2\langle\alpha,\beta\rangle + \langle\beta,\beta\rangle = 2(1-\langle\alpha,\beta\rangle)$ and

$$\left|||A\alpha - A\beta||^2 - 2(1-\langle A\alpha, A\beta\rangle)\right| = |\langle A\alpha - A\beta, A\alpha - A\beta\rangle - 2\left(1-\langle A\alpha, A\beta\rangle\right)| =$$

$$|\langle A\alpha, A\alpha\rangle - 2\langle A\alpha, A\beta\rangle + \langle A\beta, A\beta\rangle - 2 + 2\langle A\alpha, A\beta\rangle| \leq \left|||A\alpha||^2 - 1\right| + \left|||A\beta||^2 - 1\right| \leq 2\delta.$$

Hence,

$$|\langle A\alpha, A\beta\rangle - \langle\alpha,\beta\rangle| \leq \delta + \frac{|||A\alpha - A\beta||^2 - ||\alpha-\beta||^2|}{2} \leq 3\delta.$$

□

Lemma 4.6. *Let $\alpha, \beta \in \mathbb{R}^n$ 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*

$$|\sphericalangle(\alpha,\beta) - \sphericalangle(A\alpha, A\beta)| \leq 7\sqrt{\delta}.$$

Proof. It suffices to show that if the norms of $\alpha, \beta, \alpha-\beta$ were all preserved (up to a $1\pm\delta$ multiplicative factor) then the claim holds. Define for briefness θ as the angle between α, and β and with θ' the angle between $A\alpha$ and $A\beta$. Then

$$\cos(\theta') = \frac{\langle A\alpha, A\beta\rangle}{||A\alpha||_2||A\beta||_2} \quad , \quad \cos(\theta) = \langle\alpha,\beta\rangle$$

and by the previous lemma,

$$|\cos(\theta') - \cos(\theta)| = \left|\frac{\langle A\alpha, A\beta\rangle}{||A\alpha||_2||A\beta||_2} - \langle\alpha,\beta\rangle\right| \leq$$

$$\left|\frac{\langle A\alpha, A\beta\rangle}{||A\alpha||||A\beta||} - \langle A\alpha, A\beta\rangle\right| + |\langle A\alpha, A\beta\rangle - \langle\alpha,\beta\rangle| \leq$$

$$|||A\alpha|| \cdot ||A\beta||\left|\left(\frac{1}{||A\alpha||||A\beta||} - 1\right)\right| + 3\delta \leq (1+\delta)^2\left((1-\delta)^{-2} - 1\right) + 3\delta < 6\delta\,. \tag{1}$$

The last inequality holds since $\delta < 1$.

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 $|\theta'| < 5\sqrt{\delta}$ 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 $\cos(\theta)$:

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

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

Inequality (1) follows from Equation(1). Equality (2) holds since for $2\sqrt{\delta} < \theta < \pi - 2\sqrt{\delta}$ we have that $\sin(\theta) \geq \sin(2\sqrt{\delta}) > 2\sqrt{\delta} - 4\delta^{3/2}/3 > \sqrt{\delta}$ (for a small enough δ). Inequality (3) follows for sufficiently small δ. □

Corollary 4.7. *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.*

4.3 The sample set

We now construct the ϵ-sample. First, we state a lemma indicating that the value of the expression $\mathbb{E}_{x\in\mathbb{S}^{n-1}}[f_{\alpha,\beta}(x)]$ depends only on the angle between α and β.

Lemma 4.8. *[Lemma 2.2 of [GW95]]* $\mathbb{E}_{x\in\mathbb{S}^{n-1}}[f_{\alpha,\beta}(x)] = 1 - 2\sphericalangle(\alpha,\beta)/\pi$.

Recall that the angle preserving set was meant to reduce our problem to a t dimensional space. Let $t > 0$ (we shall eventually set $t = \exp\left(O\left(\sqrt{\log(n)}\right)\right)$) and $P^{(t)} \subseteq \mathbb{S}^{t-1}$ be a $1/t$-sample for digons guaranteed by Corollary 4.3. Notice that $|P^{(t)}| = \exp\left(O\left(\log^2(t)\right)\right)$. The sample (multi) set is defined as

$$P = \left\{ \frac{x'^T \cdot A}{||x'^T \cdot A||} \;\middle|\; x' \in P^{(t)}, A \in \mathcal{A} \right\}$$

where $\mathcal{A}$ is a $(c\delta^2, c\delta^2)$-norm preserving set, guaranteed by Theorem 3.10, for some $\delta = t^{-\Omega(1)}$ and c as in Corollary 4.7. Note, that P may be a multi-set. Also note that there may exist $x' \in P^{(t)}$ and $A \in \mathcal{A}$ s.t. $x'^T A = 0$. Currently, we set their corresponding vector to 0. We shall later see that these 0 vectors can be omitted without affecting the properties of P, so we could indeed assume that P is contained in $\mathbb{S}^{n-1}$. Theorem 1.2 is implied by the next theorem.

Theorem 4.9. *Let $\epsilon \triangleq 5\delta + O(\delta^2)$. The set P is an ϵ-sample for digons and is of cardinality $|P| = n^{1+o(1)} \cdot \exp\left(O\left(\log^2(\epsilon^{-1})\right)\right)$.*

Proof. The claim regarding $|P|$ stems directly from its definition (since $\epsilon = t^{-\Theta(1)}$). Notice that for any non-zero 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 was the set $\left\{x'^T \cdot A \;\middle|\; x' \in P^{(t)}, A \in \mathcal{A}\right\}$.

Let $\alpha, \beta \in \mathbb{S}^{n-1}$ be two given vectors and denote by θ the angle between them. First, We define $\hat{\mathcal{A}}$ as the set of all linear transformation in $\mathcal{A}$ that preserved the angle between the vectors up to a $\pm\delta$ additive factor and additionally[4] satisfy that $A\alpha, A\beta \neq 0$. Notice that since $\mathcal{A}$ is (δ,δ)-angle preserving (by Corollary 4.7) and (δ,δ)-norm preserving, $|\hat{\mathcal{A}}| \geq (1-3\delta)|\mathcal{A}|$. Correspondingly, we define $\hat{P} = \left\{x'^T A \;\middle|\; x' \in P^{(t)}, A \in \hat{\mathcal{A}}\right\}$. Then

$$\mathbb{E}_{x\in P}[\text{sign}(\langle x,\alpha\rangle)\cdot\text{sign}(\langle x,\beta\rangle)] = \mathbb{E}_{x\in\hat{P}}[\text{sign}(\langle x,\alpha\rangle)\cdot\text{sign}(\langle x,\beta\rangle)] \pm 3\delta$$

since at most a fraction of 3δ vectors of P are not in $\hat{P}$. Due to the properties of $\hat{P}, P^{(t)}$ and the fact that $A\alpha, A\beta \neq 0$ we get

$$\mathbb{E}_{x\in\hat{P}}[\text{sign}(\langle x,\alpha\rangle)\cdot\text{sign}(\langle x,\beta\rangle)] = \mathbb{E}_{A\in\hat{\mathcal{A}},x'\in P^{(t)}}[\text{sign}(\langle x',A\alpha\rangle)\cdot\text{sign}(\langle x',A\beta\rangle)] \overset{(1)}{=}$$

[4] The proof of Lemma 4.6 shows that the additional requirement does not reduce the bound on $|\hat{\mathcal{A}}|$.

$$\mathbb{E}_{A\in\hat{\mathcal{A}},y\in\mathbb{S}^{t-1}}\left[\operatorname{sign}\left(\langle y,A\alpha\rangle\right)\cdot\operatorname{sign}\left(\langle y,A\beta\rangle\right)\right]\pm\delta\overset{(2)}{=}\mathbb{E}_{A\in\hat{\mathcal{A}}}\left[1-2\frac{\sphericalangle\left(A\alpha,A\beta\right)}{\pi}\right]\pm\delta\overset{(3)}{=}$$

$$1-\frac{2\sphericalangle\left(\alpha,\beta\right)}{\pi}\pm2\delta\overset{(4)}{=}\mathbb{E}_{x\in\mathbb{S}^{n-1}}\left[\operatorname{sign}\left(\langle x,\alpha\rangle\right)\cdot\operatorname{sign}\left(\langle x,\beta\rangle\right)\right]\pm2\delta$$

where equality (1) stems from $P^{(t)}$ being a δ-sample for t dimensional digons, equalities (2) and (4) follow from Lemma 4.8 and equality (3) holds due to the definition of $\hat{\mathcal{A}}$. By combining with the previous equation we get the required result.

Notice that as defined, P might not be a subset of $\mathbb{S}^{n-1}$ as it might contain zero vectors. However, as $\mathcal{A}$ is $(c\delta^2, c\delta^2)$-norm preserving we see that at most a fraction of $c\delta^2$ of the vectors in P are zeroes and so we can remove them from P and only change ϵ (in Theorem 4.9) by $O(\delta^2)$. □

5 Applications

In this section we give an application of the ϵ-sample for digons constructed in Section 4. Specifically, we use the ϵ-sample to derandomize certain rounding procedures of solutions of semi definite programs. For example, in the famous Goemens-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 $\operatorname{sign}(\langle\alpha,x\rangle)\cdot\operatorname{sign}(\langle\beta,x\rangle)=-1$, where x is perpendicular to the hyperplane. Hence, in order to choose hyperplanes that appear random to such a process we need a sample set for digons. From now on P_ϵ denotes an ϵ-sample for digons. Another application for derandomizing the coloring algorithm of [KMS98] appears in Section 5.2.

5.1 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, 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 programing (SDP for short): First notice that max-cut can be solved by the following integer program:

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

The following is an SDP relaxation of the integer program.

$$\text{Maximize}\quad\frac{1}{2}\sum_{1\le i<j\le|V|}w_{i,j}(1-\langle v_i,v_j\rangle)\qquad\text{subject to}\quad\forall i\ \ v_i\in\mathbb{S}^{n-1}\,.\tag{3}$$

An approximation to the integer problem (2) can be obtained from a solution to (3) in the following way: Choose a random unit vector x and construct the following cut: $S = \{i \mid \langle x, v_i \rangle \geq 0\}$. Let W denote the size of the cut produced this way and $\mathbb{E}[W]$ its expectation. [GW95] proved that the approximation given by the SDP is good by first observing that

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

and then showing that

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

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 will define S, uniformly from the set P_ϵ described in the beginning of this section[5] for some $\epsilon = o(1)$. To prove that this works 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 $x \in P_\epsilon$ instead of $x \in \mathbb{S}^{n-1}$.

Lemma 5.1. $\mathbb{E}_{x \in P_\epsilon}[W] \geq \mathbb{E}_{x \in \mathbb{S}^{n-1}}[W] - 2\epsilon \cdot OPT$.

Proof. By definition of W: $\mathbb{E}_{x \in \mathbb{S}^{n-1}}[W] = \sum_{i<j} w_{i,j} \Pr_{x \in \mathbb{S}^{n-1}}[\mathrm{sign}(v_i \cdot x) \neq \mathrm{sign}(v_j \cdot x)] = \sum_{i<j} w_{i,j} \left(\Pr_{x \in P_\epsilon}[\mathrm{sign}(v_i \cdot x) \neq \mathrm{sign}(v_j \cdot x)] \pm \epsilon\right) = \mathbb{E}_{x \in P_\epsilon}[W] \pm \epsilon \sum_{i<j} w_{i,j}$. 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,

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

□

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

[5] This gives a randomized algorithm using only a logarithmic number of random bits. This algorithm can be derandomized by going over all possible settings for the random bits and choosing the best solution.

5.2 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 coloring of the graph with a minimal number of colors. This problem is well known to be NP-hard. [KMS98] gave an approximation algorithm that efficiently colors a 3-colorable graph with $O\left(\min\left\{n^{0.387}, \Delta^{\log_3 2}\log n\right\}\right)$ colors where Δ is the maximum degree of any vertex. [6]

We now give a brief description of the approximation algorithm: First, solve a semi-definite program assigning each vector to a 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. Then assign colors to the vertices in the following way: Choose r random unit vectors independently $x_1, \ldots, x_r$. Each vertex will receive r bits. The value of the i'th bit of a vertex with 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 $n^{0.387}$ 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 shall require several known results concerning expander graphs that are given in Section 5.2.1. Let $\epsilon > 0$ be some small constant. The set P_ϵ denotes an ϵ-sample for digons as guaranteed by Theorem 1.2. We describe a random algorithm that will use a logarithmic number of bits. This can be derandomized by going over all settings for the random bits. We choose the vectors $x_1, \ldots, x_r$ in the following way.

- Construct an expander graph of parameters (n', d, λ) where $n' = |P_\epsilon|$, $d = O(\epsilon^{-2})$ and[7]

[6] In fact, they show two methods where the second method obtains a coloring of $\min\left\{O\left(\Delta^{1/3}\log^{1/2}\Delta\log n\right), O\left(n^{1/4}\log^{1/2} n\right)\right\}$ colors. However, our methods derandomizes the first method, yielding the slightly worse approximation.

[7] We set d as the smallest integer for which there exist an efficient construction for an expander graph with

$\lambda \leq \epsilon d$.

- Label each vertex of the graph as a vector in P_ϵ.
- Let r be a parameter. Choose $x_1, \ldots, x_r \in P_\epsilon$ by taking a random walk of length r in the expander.

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

$$\log(n') + r\log(d) = \log(|P_\epsilon|) + 2\log_3 2\log(\epsilon^{-1})\log(\Delta) + O(\log(\epsilon^{-1})) = O(\log(\epsilon^{-1})\log n) \ .$$

Hence, the support size of possible r-tuples of vectors is polynomial in n assuming a constant ϵ (or of size $n^{O(\log(\epsilon^{-1}))}$ for general ϵ). We define this set of r-tuples as $\mathcal{X}$.

The following lemma bounds the probability that the chosen r-tuple does not 'separates' 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 5.3. *Let v_1, v_2 be two vectors corresponding to two neighboring vectors of the graph. Let c_1, c_2 be the colors assigned to each vector according to the choice of $(x_1, \ldots, x_{r'})$ for some $r' \leq r$. Then*

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

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 Theorem 5.4 which leads to the following bound:

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

□

Hence, following the original algorithm's notations, we 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. By the outline sketched above, this gives a good coloring of at least $n/2$ vertices, however there may be $n/2$ vertices that were not properly colored. Therefore we need to repeat this process until all of the vertices are isolated (this will require $O(\log n)$ repetitions). Notice that in order to find a coloring of $n/4$ vertices among the remaining $n/2$ vertices we can use the same vectors generated by the random walk and just look for the best one among them. Thus after repeating this process for $\log n$ steps we achieve a coloring using $n^{0.387+\epsilon}$ many colors, for any $\epsilon > 0$, with a running time $n^{O(\log(\epsilon^{-1}))}$.

$\lambda \leq \epsilon d$. Since there exist graphs with $\lambda \approx \sqrt{d}$, $d = O(\epsilon^{-2})$ is sufficiently large.

5.2.1 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, AL88]. In particular, there exist explicit constructions for infinitely many n of $(n, 6, 4)$-expanders.

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 $(v_1, v_2, \ldots, v_t)$. We shall make use of the following theorem regarding such walks

Theorem 5.4. *[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 $(v_1, \ldots, v_\ell)$ 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*

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

6 Fooling Linear Threshold Functions

A linear threshold function (LTF) is a function $f : \mathbb{R}^n \to \{-1, 1\}$ of the following form: $f_{w,\theta}(x) = \text{sign}\left(\langle w, x\rangle - \theta\right)$ where $w \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ and sign(0) is defined as 1. Functions of this form are indicator functions of spherical caps. In this section we construct a sample set for spherical caps. Using similar methods to the case of digons, we show that a norm-preserving set can reduce the problem to that of finding a sample for spherical caps of dimension $t \ll n$. Then we use Nisan's pseudo-random generator for log-space machines (see section 4.1), as in the case of digons. Specifically, we construct a set $Q \subset \mathbb{S}^{n-1}$ for which it holds that

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

where w' is some unit vector in $\mathbb{S}^{t-1}$. We now explain how to construct Q. Let $\mathcal{A}$ be a (δ, δ)-norm preserving set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^t$, where $t = \exp\left(c\sqrt{\log(n)}\right)$ for some constant c and $\delta = t^{-\Omega(1)}$ (see Theorem 3.10). $Q = Q_{\mathcal{A}}$ is defined in the following way: Let $Q' \subseteq \mathbb{R}^t$ be a δ-sample for LTFs in t dimensions. Note that an explicit such set, of size $n^{O(c^2)}$, exists due to corollary 4.3. $Q_{\mathcal{A}}$ is defined as

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

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

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

Our result is given in the next theorem.

Theorem 6.1. *Let f be some LTF. If $\mathcal{A}$ is (δ,δ)-norm preserving, for $\delta > 1/t$ then*

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

Before given 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 the connection to the standard normal distribution $N(0,1)$. Denote by $\Phi(z)$ the probability that a random variable from the normal gaussian distribution takes a value larger than z. That is $\Phi(z) \triangleq \Pr_{Y\sim N(0,1)}\left[Y > z\right]$. The following two lemmas are well known so we postpone their proof to Appendix A.

Lemma 6.2. *Let d be an integer and $w \in \mathbb{S}^{d-1}$ some fixed unit vector. For any $z \in \mathbb{R}$ it holds that*

$$\Pr\left[\langle x, w\rangle > z/\sqrt{d}\right] = \Phi(z) \pm \tilde{O}\left(1/d\right)\ ,$$

where x is a d-dimensional unit vector chosen at random.

Lemma 6.3. *Let $z > 0$ and $0 < \delta < 1$. Then*

$$\Phi\left(z\cdot(1\pm\delta)\right) = \Phi(z) \pm \tilde{O}(\delta)\ .$$

We can now prove Theorem 6.1.

Proof. Let $w \in \mathbb{S}^{n-1}$ and $z \in \mathbb{R}$ be such that $f(x) = \text{sign}\left(\langle x, w\rangle - z/\sqrt{n}\right)$. For simplicity we assume w.l.o.g. that $z \geq 0$.[8] By Lemma 6.2,

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

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

$$\Pr_{x'\in\mathbb{S}^{t-1}}\left[\langle x', Aw\rangle > z/\sqrt{t}\right] = \Pr_{x'\in\mathbb{S}^{t-1}}\left[\langle x', w'\rangle > z/(1\pm\delta)\sqrt{t}\right] \tag{4}$$

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

$$\begin{aligned}\Pr_{x'\in\mathbb{S}^{t-1}}\left[\langle x', w'\rangle > z/(1\pm\delta)\sqrt{t}\right] &\overset{(1)}{=} \Phi\left(z/(1\pm\delta)\right) \pm \tilde{O}\left(1/t\right) \overset{(2)}{=} \\ \Phi(z) \pm \tilde{O}\left(\delta\right) &\overset{(3)}{=} \Pr_{x\in\mathbb{S}^{n-1}}\left[\langle x, w\rangle > z/\sqrt{n}\right] \pm \tilde{O}\left(\delta\right).\end{aligned} \tag{5}$$

Equalities (1) and (3) stem from Lemma 6.2 and the fact that $\delta > 1/t$. Equality (2) is implied by Lemma 6.3. Calculating we get

$$\Pr_{x\in Q_{\mathcal{A}}}\left[\langle x, w\rangle > z/\sqrt{n}\right] \overset{(1)}{=} \Pr_{x'\in Q', A\in\mathcal{A}}\left[\langle x', Aw\rangle > z/\sqrt{t}\right] \overset{(2)}{=}$$

[8] Define $\hat{f} = \text{sign}\left(\langle x, -w\rangle - \frac{-a}{\sqrt{n}}\right)$ and note that $f(x) = -\hat{f}(x)$.

$$\Pr_{x' \in Q', A \in \hat{\mathcal{A}}} \left[\langle x', Aw \rangle > z/\sqrt{t} \right] \pm O(\delta) \overset{(3)}{=} \Pr_{x' \in \mathbb{S}^{t-1}, A \in \hat{\mathcal{A}}} \left[\langle x', Aw \rangle > z/\sqrt{t} \right] \pm O(\delta) \overset{(4)}{=}$$

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

where equality (1) follows by the definition of $Q_{\mathcal{A}}$, equality (2) holds as we replaced A by $\hat{A}$, equality (3) is by the definition of Q' and equality (4) is implied by Equations (4) and (5). This proves the claim. □

Corollary 6.4. *For any $\epsilon > 0$, there exists a weak ϵ-sample $Q_{\mathcal{A}}$ for spherical caps of cardinality $|Q_{\mathcal{A}}| = \exp(\log^2(1/\epsilon))n^{1+o(1)}$. Furthermore, $Q_{\mathcal{A}}$ can be constructed in $\exp(\log^2(1/\epsilon))n^{O(1)}$ time.*

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[AC06] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563, 2006.

[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. *Computational Complexity*, 5(1):60–75, 1995.

[AKS87] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in logspace. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing (STOC)*, pages 132–140, 1987.

[AL88] and P.Sarnak A. Lubotzky, R. Phillips. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[AL08] N. Ailon and E. Liberty. Fast dimension reduction using rademacher series on dual bch codes. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1–9, 2008.

[Cha00] Bernard Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, New York, NY, USA, 2000.

[DG99] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss Lemma. *International Computer Science Institute, Technical Report*, pages 99–006, 1999.

[DGJ+09] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. A. Servedio, and E. Viola. Bounded independence fools halfspaces. *CoRR*, abs/0902.3757, 2009.

[EIO02] L. Engebretsen, P. Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 705–712, 2002.

[Gol97] O. Goldreich. A sample of samplers - a computational perspective on sampling (survey). *Electronic Colloquium on Computational Complexity (ECCC)*, 4(20), 1997.

[Gol01] Oded Goldreich. *Foundations of cryptography: basic tools*. Cambridge University Press, 2001.

[GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[Ind06] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.

[Ind07] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of l2 into l1. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 615–620, 2007.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[JN09] W. B. Johnson and A. Naor. The johnson-lindenstrauss lemma almost characterizes hilbert space, but not quite. In *Proc. of the 19th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 885–891, 2009.

[KMS98] D. R. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.

[KV94] Michael J. Kearns and Umesh V. Vazirani. *An introduction to computational learning theory*. MIT Press, Cambridge, MA, USA, 1994.

[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 Struct. Algorithms*, 33(2):142–156, 2008.

[MR99] S. Mahajan and H. Ramesh. Derandomizing approximation algorithms based on semidefinite programming. *SIAM J. Comput.*, 28(5):1641–1663, 1999.

[MZ09] R. Meka and D. Zuckerman. Pseudorandom generators for polynomial threshold functions. http://arxiv.org/abs/0910.4122, 2009.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[RS09] Y. Rabani and A. Shpilka. Explicit construction of a small epsilon-net for linear threshold functions. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 649–658, 2009.

[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.

[SZ99] M. E. Saks and S. Zhou. Bp $_{\mathrm{h}}$space(s) subseteq dspace(s$^{3/2}$). *J. Comput. Syst. Sci.*, 58(2):376–403, 1999.

A The Distribution of a Projected Random Unit Vector

Before we start proving Lemma 6.2, we present the following well known property of the gaussian distribution:

Lemma A.1. *For any $\alpha > 0$,*

$$\Phi(\alpha) \leq \frac{1}{\alpha\sqrt{2\pi}} \exp\left(-\alpha^2/2\right) .$$

Proof. Note that for every τ and α it holds that $-\tau^2/2 \leq \alpha^2/2 - \alpha\tau$. Hence,

$$\Phi(\alpha) = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\tau^2/2\right) d\tau \leq \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\alpha^2/2 - \alpha\tau) d\tau =$$

$$\frac{\exp(\alpha^2/2)}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp(-\alpha\tau) d\tau = \frac{1}{\alpha\sqrt{2\pi}} \exp\left(-\alpha^2/2\right) .$$

□

We restate Lemma 6.2:

Lemma Let t be an integer and x a t-dimensional unit vector chosen at random. Let $w \in \mathbb{S}^{t-1}$ be some fixed unit vector. For any $\alpha \in \mathbb{R}$, we have

$$Pr\left[\langle x, w\rangle > \frac{\alpha}{\sqrt{t}}\right] = \Phi(\alpha) \pm \tilde{O}\left(1/t\right) .$$

Proof. We prove the claim for a positive α as the other case is similar. Denote with X the random variable $X = \langle x, w \rangle$. Since x is randomly chosen from the unit sphere, we may assume w.l.o.g. that w is the vector $e_1 = (1, 0, \ldots, 0)$, meaning that $X = x_1$. Define $\mathcal{S}_t(r)$ as the surface area of the sphere of radius r in $\mathbb{R}^t$. Then

$$\mathcal{S}_t(r) = \pi^{t/2} r^{t-1} / \Gamma(t/2) = \mathcal{S}_t(1) r^{t-1}$$

where Γ is the Γ-Function[9]. Let $\tau, h > 0$. Consider the surface area of the "strip" of the unit sphere in which $x_1 \in (\tau - h, \tau)$. Its surface is bounded from below by $h\mathcal{S}_{t-1}(r)$ where $r = \sqrt{1-\tau^2}$. It is also bounded from above by $h\mathcal{S}_{t-1}(r')$ where $r' = \sqrt{1-(\tau-h)^2}$. By dividing by $\mathcal{S}_t(1)$ we get bounds for the probabilities that $x_1 \in [\tau - h, \tau]$. Therefore, for any $0 \leq \alpha \leq \beta \leq 1$

$$\Pr\left[\frac{\alpha}{\sqrt{t}} < X < \frac{\beta}{\sqrt{t}}\right] = \int_{\frac{\alpha}{\sqrt{t}}}^{\frac{\beta}{\sqrt{t}}} \frac{\mathcal{S}_{t-1}(1) \cdot (1-\tau^2)^{\frac{t-2}{2}}}{\mathcal{S}_t(1)} d\tau \ . \tag{6}$$

To compute the integral we first analyze the ratio $\frac{\mathcal{S}_{t-1}(1)}{\mathcal{S}_t(1)}$. When t is even:

$$\frac{\mathcal{S}_{t-1}(1)}{\mathcal{S}_t(1)} = \frac{\frac{2(2\pi)^{(t-2)/2}}{1\cdot 3\cdot \ldots \cdot (t-3)}}{\frac{(2\pi)^{t/2}}{2\cdot 4\cdot \ldots \cdot (t-2)}} = \frac{2^{t-2} \cdot ((t-2)/2)!^2}{\pi \cdot (t-2)!} \overset{(*)}{=}$$

$$\frac{1}{\sqrt{2\pi}}\sqrt{t-2}\left(1 \pm O\left(\frac{1}{t}\right)\right) = \frac{1}{\sqrt{2\pi}}\sqrt{t}\left(1 \pm O\left(\frac{1}{t}\right)\right) \tag{7}$$

where equality $(*)$ follows from Stirling's approximation. For odd t we get the same result analogically. Let us now analyze the function within the integral in equation 6: For convenience we change τ to be $\frac{\alpha}{\sqrt{t}}$.

$$\left(1 - \frac{\alpha^2}{t}\right)^{(t-2)/2} = \left(1 - \frac{\alpha^2}{t}\right)^{t/2} \cdot \left(1 - \frac{\alpha^2}{t}\right)^{-1} =$$

$$\exp(-\alpha^2/2)\left(1 + O\left(\frac{\alpha^4}{t}\right)\right)\left(1 - \frac{\alpha^2}{t}\right)^{-1} \overset{(1)}{=} \exp(-\alpha^2/2)\left(1 + O\left(\frac{\alpha^4}{t}\right)\right). \tag{8}$$

Equation (1) holds since $e^x = \left(1 + \frac{x}{n}\right)^n \left(1 + O(\frac{x^2}{n})\right)$ when $x = o(\sqrt{n})$. From Equations 6, 7 and 8 we get

$$Pr\left[\frac{\alpha}{\sqrt{t}} < X < \frac{\beta}{\sqrt{t}}\right] = \frac{\sqrt{t}}{\sqrt{2\pi}}\left(1 \pm O\left(\frac{\beta^4}{t}\right)\right)\int_{\frac{\alpha}{\sqrt{t}}}^{\frac{\beta}{\sqrt{t}}} \exp(-\frac{\tau^2 t}{2}) d\tau =$$

$$\frac{1 \pm O\left(\frac{\beta^4}{t}\right)}{\sqrt{2\pi}}\int_{\alpha}^{\beta} \exp(-\frac{\tau^2}{2}) d\tau = \left(1 \pm O\left(\frac{\beta^4}{t}\right)\right)\Phi(\alpha, \beta) \tag{9}$$

[9]For an integer t, the Γ function is given by $\Gamma(t) \triangleq \begin{cases} \left(\frac{t-2}{2}\right)! & t \text{ even} \\ \left(\frac{(t-2)(t-4)\ldots 1}{2^{(t-1)/2}}\right)! & t \text{ odd} \end{cases}$

where $\Phi(\alpha,\beta)$ is defined as the probability of a normal gaussian to take values between α and β. We divide our discussion to two cases. First, assume that $\alpha \geq \sqrt{3\ln(t)}$. We upper bound $\Pr\left[X > \frac{\alpha}{\sqrt{t}}\right]$ by the relative surface area of the strip of the unit sphere in which $x_1 \geq \frac{\alpha}{\sqrt{t}}$:

$$\Pr\left[X > \frac{\alpha}{\sqrt{t}}\right] < \frac{\mathcal{S}_{t-1}(1)\cdot(1-\alpha^2/t)^{\frac{t-2}{2}}}{\mathcal{S}_t(1)} = O\left(\sqrt{t}\cdot\exp\left(-1.5\ln(t)\right)\right) = O(1/t).$$

The second equality stems from equations 7 and 8. By Lemma A.1, $\Phi(\alpha) < \frac{1}{\alpha\sqrt{2\pi}}\exp\left(-\alpha^2/2\right) < 1/t$ and therefore, $\Pr\left[X > \alpha/\sqrt{t}\right] \leq \Phi(\alpha) + O(\frac{1}{t})$. Assume now that $\alpha < \sqrt{3\ln(t)}$. By setting $\beta = \sqrt{3\ln(t)}$ we get from Equation 9 that

$$\Pr\left[X > \frac{\alpha}{\sqrt{t}}\right] = \Pr\left[\frac{\alpha}{\sqrt{t}} < X < \frac{\beta}{\sqrt{t}}\right] + \Pr\left[X > \frac{\beta}{\sqrt{t}}\right] =$$

$$\Phi(\alpha,\beta)\left(1 \pm O\left(\frac{\beta^4}{t}\right)\right) + O\left(\frac{1}{t}\right) = \Phi(\alpha) \pm \tilde{O}\left(\frac{1}{t}\right) .$$

□

We now present the proof of Lemma 6.3.

Lemma Let $\alpha > 0$, $0 < \delta < 1/4$.Then

$$\Phi(\alpha\cdot(1\pm\delta)) = \Phi(\alpha) \pm \tilde{O}(\delta).$$

Proof. Assume first that $\alpha < \sqrt{3\ln(\delta^{-1})}$. Then for any $\gamma = \alpha(1\pm\delta)$,

$$|\Phi(\gamma) - \Phi(\alpha)| = \left|\frac{1}{\sqrt{2\pi}}\int_{\gamma}^{\alpha}\exp(-\tau^2/2)d\tau\right| < \alpha\cdot\delta = \tilde{O}(\delta).$$

Assume now that $\alpha \geq \sqrt{3\ln(\delta^{-1})}$. Let $\gamma \triangleq \alpha\delta$. by Lemma A.1,

$$\Phi(\alpha+\gamma) < \Phi(\alpha) < \Phi(\alpha-\gamma) < \frac{1}{(\alpha-\gamma)\sqrt{2\pi}}\exp\left(-(\alpha-\gamma)^2/2\right) < \exp\left(-(\alpha-\gamma)^2/2\right) \leq$$

$$\exp\left(-\ln\left(\delta^{-1}\right)\cdot 3\left(1-\delta\right)^2/2\right) \leq \delta.$$

The last equality holds since $\delta < 1/4$. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de