Sample-Based High-Dimensional Convexity Testing[*][†]

Xi Chen[1], Adam Freilich[2], Rocco A. Servedio[3], and Timothy Sun[4]

1 **Columbia University, New York, NY, USA**
   `xichen@cs.columbia.edu`
2 **Columbia University, New York, NY, USA**
   `freilich@cs.columbia.edu`
3 **Columbia University, New York, NY, USA**
   `rocco@cs.columbia.edu`
4 **Columbia University, New York, NY, USA**
   `tim@cs.columbia.edu`

Abstract

In the problem of *high-dimensional convexity testing*, there is an unknown set $S \subseteq \mathbb{R}^n$ which is promised to be either convex or ε-far from every convex body with respect to the standard multivariate normal distribution $\mathcal{N}(0,1)^n$. The job of a testing algorithm is then to distinguish between these two cases while making as few inspections of the set S as possible.

In this work we consider *sample-based* testing algorithms, in which the testing algorithm only has access to labeled samples $(\boldsymbol{x}, S(\boldsymbol{x}))$ where each $\boldsymbol{x}$ is independently drawn from $\mathcal{N}(0,1)^n$. We give nearly matching sample complexity upper and lower bounds for both one-sided and two-sided convexity testing algorithms in this framework. For constant ε, our results show that the sample complexity of one-sided convexity testing is $2^{\tilde{\Theta}(n)}$ samples, while for two-sided convexity testing it is $2^{\tilde{\Theta}(\sqrt{n})}$.

1 Introduction

Over the past few decades the field of property testing has developed into a fertile area with many different branches of active research. Several distinct lines of work have studied the testability of various kinds of *high-dimensional* objects, including probability distributions (see e.g. [9, 37, 4, 38, 2, 13, 1]), Boolean functions (see e.g. [17, 34, 15, 31, 28] and many other works), and various types of codes and algebraic objects (see e.g. [3, 23, 26, 14] and many other works). These efforts have collectively yielded significant insight into the abilities and limitations of efficient testing algorithms for such high-dimensional objects. A distinct line of work has focused on testing (mostly low-dimensional) *geometric properties*. Here too a considerable body of work has led to a good understanding of the testability of various low-dimensional geometric properties, see e.g. [19, 18, 36, 12, 11, 10].

This paper is about a topic which lies at the intersection of the two general strands (high-dimensional property testing and geometric property testing) mentioned above: we study the problem of *high-dimensional convexity testing.* Convexity is a fundamental property

---

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017).
Editors: Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala; Article No. 37; pp. 37:1–37:20

which is intensively studied in high-dimensional geometry (see e.g. [24, 8, 39] and many other references) and has been studied in the property testing of images (the two-dimensional case) [36, 12, 11, 10], but as we discuss in Section 1.2 below, very little is known about high-dimensional convexity testing.

We consider $\mathbb{R}^n$ endowed with the standard normal distribution $\mathcal{N}(0,1)^n$ as our underlying space, so the distance $\text{dist}(S,C)$ between two subsets $S, C \subseteq \mathbb{R}^n$ is

$$\Pr_{\boldsymbol{x} \leftarrow \mathcal{N}(0,1)^n}[\boldsymbol{x} \in S \triangle C],$$

where $S \triangle C$ denotes their symmetric difference. The standard normal distribution $\mathcal{N}(0,1)^n$ is arguably one of the most natural, and certainly one of the most studied, distributions on $\mathbb{R}^n$. Several previous works have studied property testing over $\mathbb{R}^n$ with respect to $\mathcal{N}(0,1)^n$, such as the work on testing halfspaces [31, 6] and the work on testing surface area [30, 33].

1.1 Our results

In this paper we focus on *sample-based* testing algorithms for convexity. Such an algorithm has access to independent draws $(\boldsymbol{x}, S(\boldsymbol{x})) \in \mathbb{R}^n \times \{0,1\}$, where $\boldsymbol{x}$ is drawn from $\mathcal{N}(0,1)^n$ and $S \subseteq \mathbb{R}^n$ is the unknown set being tested for convexity (so in particular the algorithm cannot select points to be queried) with $S(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in S$. We say such an algorithm is an *ε-tester for convexity* if it accepts S with probability at least $2/3$ when S is convex and rejects with probability at least $2/3$ when it is ε-far from convex, i.e., $\text{dist}(S,C) \geq \varepsilon$ for all convex sets $C \subseteq \mathbb{R}^n$. The model of sample-based testing was originally introduced by Goldreich, Goldwasser, and Ron almost two decades ago [21], where it was referred to as "passive testing;" it has received significant attention over the years [27, 20, 6, 22], with an uptick in research activity in this model over just the past year or so [5, 16, 12, 11, 10].

We consider sample-based testers for convexity that are allowed both one-sided (i.e., the algorithm always accepts S when it is convex) and two-sided error. In each case, for constant $\varepsilon > 0$ we give nearly matching upper and lower bounds on sample complexity. Our results are as follows:

▶ **Theorem 1** (One-sided lower bound). *Any one-sided sample-based algorithm that is an ε-tester for convexity over $\mathcal{N}(0,1)^n$ for some $\varepsilon < 1/2$ must use $2^{\Omega(n)}$ samples.*

▶ **Theorem 2** (One-sided upper bound). *For any $\varepsilon > 0$, there is a one-sided sample-based ε-tester for convexity over $\mathcal{N}(0,1)^n$ which uses $(n/\varepsilon)^{O(n)}$ samples.*

▶ **Theorem 3** (Two-sided lower bound). *There exists a positive constant ε_0 such that any two-sided sample-based algorithm that is an ε-tester for convexity over $\mathcal{N}(0,1)^n$ for some $\varepsilon \leq \varepsilon_0$ must use $2^{\Omega(\sqrt{n})}$ samples.*

▶ **Theorem 4** (Two-sided upper bound). *For any $\varepsilon > 0$, there is a two-sided sample-based ε-tester for convexity over $\mathcal{N}(0,1)^n$ which uses $n^{O(\sqrt{n}/\varepsilon^2)}$ samples.*

These results are summarized in Table 1. We discuss the main ideas and techniques behind them in Section 1.3, and prove Theorem 3 in Section 3 and Theorem 1 in Section 4. We leave proofs of Theorems 2 and 4 to the full version due to space limitations.

1.2 Related work

Convexity testing. As discussed above, [36, 10, 11, 12] studied the testing of 2-dimensional convexity under the uniform distribution, either within a compact body such as $[0,1]^2$ [10, 11]

▪ **Table 1** Sample complexity bounds for sample-based convexity testing. In line four, $\varepsilon_0 > 0$ is some absolute constant.

Model	Sample complexity bound	Reference
One-sided	$2^{\Omega(n)}$ samples (for $\varepsilon < 1/2$)	Theorem 1
	$2^{O(n \log(n/\varepsilon))}$ samples	Theorem 2
Two-sided	$2^{\Omega(\sqrt{n})}$ samples (for $\varepsilon < \varepsilon_0$)	Theorem 3
	$2^{O(\sqrt{n} \log(n)/\varepsilon^2)}$ samples	Theorem 4

or over a discrete grid $[n]^2$ [36, 12]. The model of [10, 11] is more closely related to ours: [11] showed that $\Theta(\varepsilon^{-4/3})$ samples are necessary and sufficient for one-sided sample-based testers, while [10] gave a one-sided general tester (which can make adaptive queries to the unknown set) for 2-dimensional convexity with only $O(1/\varepsilon)$ queries.

The only prior work that we are aware of that deals with testing high-dimensional convexity is that of [35]. However, the model considered in [35] is different from ours in the following important aspects. First, the goal of an algorithm in their model is to determine whether an unknown $S \subseteq \mathbb{R}^n$ is not convex or is ε-close to convex in the following sense: the (Euclidean) volume of $S \triangle C$, for some convex C, is at most an ε-fraction of the volume of S. Second, in their model an algorithm both can make membership queries (to determine whether a given point x belongs to S), and can receive samples which are guaranteed to be drawn independently and uniformly at random from S. The main result of [35] is an algorithm which uses $(cn/\varepsilon)^n$ many random samples *drawn from S*, for some constant c, and $\text{poly}(n)/\varepsilon$ membership queries.

Sample-based testing. A wide range of papers have studied sample-based testing from several different perspectives, including the recent works [12, 11, 10] which study sample-based testing of convexity over two-dimensional domains. In earlier work on sample-based testing, [6] showed that the class of linear threshold functions can be tested to constant accuracy under $\mathcal{N}(0,1)^n$ with $\tilde{\Theta}(n^{1/2})$ samples drawn from $\mathcal{N}(0,1)^n$. (Note that a linear threshold function is a convex set of a very simple sort, as every convex set can be expressed as an intersection of (potentially infinitely many) linear threshold functions.) The work [6] in fact gave a characterization of the sample complexity of (two-sided) sample-based testing, in terms of a combinatorial/probabilistic quantity called the "passive testing dimension." This is a distribution-dependent quantity whose definition involves both the class being tested and the distribution from which samples are obtained; it is not *a priori* clear what the value of this quantity is for the class of convex subsets of $\mathbb{R}^n$ and the standard normal distribution $\mathcal{N}(0,1)^n$. Our upper and lower bounds (Theorems 4 and 3) may be interpreted as giving bounds on the passive testing dimension of the class of convex sets in $\mathbb{R}^n$ with respect to the $\mathcal{N}(0,1)^n$ distribution.

1.3 Our techniques

1.3.1 One-sided lower bound

Our one-sided lower bound has a simple proof using only elementary geometric and probabilistic arguments. It follows from the fact (see Lemma 17) that if $q = 2^{\Theta(n)}$ many points are

drawn independently from $\mathcal{N}(0,1)^n$, then with probability $1 - o(1)$ no one of the points lies in the convex hull of the $q - 1$ others. This can easily be shown to imply that more than q samples are required (since given only q samples, with probability $1 - o(1)$ there is a convex set consistent with any labeling and thus a one-sided algorithm cannot reject).

1.4 Two-sided lower bound

At a high-level, the proof of our two-sided lower bound uses the following standard approach. We first define two distributions $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ over sets in $\mathbb{R}^n$ such that (i) $\mathcal{D}_{\mathsf{yes}}$ is a distribution over convex sets only, and (ii) $\mathcal{D}_{\mathsf{no}}$ is a distribution such that $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$ is ε_0-far from convex with probability at least $1 - o(1)$ for some positive constant ε_0. We then show that every sample-based, q-query algorithm A with $q = 2^{0.01n}$ must have

$$\Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}};\, \boldsymbol{x}} \big[A \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] - \Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}};\, \boldsymbol{x}} \big[A \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] \leq o(1), \tag{1}$$

where $\boldsymbol{x}$ denotes a sequence of q points drawn from $\mathcal{N}(0,1)^n$ independently and $(\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x}))$ denotes the q labeled samples from $\boldsymbol{S}$. Theorem 3 follows directly from (1).

To draw a set $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$, we sample a sequence of $N = 2^{\sqrt{n}}$ points $\mathbf{y}_1, \ldots, \mathbf{y}_N$ from the sphere $S^{n-1}(r)$ of radius r for some $r = \Theta(n^{1/4})$. Each $\mathbf{y}_i$ defines a halfspace $\boldsymbol{h}_i = \{x : x \cdot \mathbf{y}_i \leq r^2\}$. $\boldsymbol{S}$ is then the intersection of all $\boldsymbol{h}_i$'s. (This is essentially a construction used by Nazarov [32] to exhibit a convex set that has large Gaussian surface area, and used by [29] to lower bound the sample complexity of learning convex sets under the Gaussian distribution.) The most challenging part of the two-sided lower bound proof is to show that, with q points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q \leftarrow \mathcal{N}(0,1)^n$, the q bits $\boldsymbol{S}(\boldsymbol{x}_1), \ldots, \boldsymbol{S}(\boldsymbol{x}_q)$ with $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ are "almost" independent. More formally, the q bits $\boldsymbol{S}(\boldsymbol{x}_1), \ldots, \boldsymbol{S}(\boldsymbol{x}_q)$ with $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ have $o(1)$-total variation distance from q independent bits with the ith bit drawn from the marginal distribution of $\boldsymbol{S}(\boldsymbol{x}_i)$ as $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$. On the other hand, it is relatively easy to define a distribution $\mathcal{D}_{\mathsf{no}}$ that satisfies (ii) and at the same time, $\boldsymbol{S}(\boldsymbol{x}_1), \ldots, \boldsymbol{S}(\boldsymbol{x}_q)$ when $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$ has $o(1)$-total variation distance from the same product distribution. (1) follows by combining the two parts.

1.4.1 Structural result

Our algorithms rely on a new structural result which we establish for convex sets in $\mathbb{R}^n$. Roughly speaking, this result gives an upper bound on the Gaussian volume of the "thickened surface" of any bounded convex subset of $\mathbb{R}^n$; it is inspired by, and builds on, the classic result of Ball [7] that upperbounds the Gaussian surface area of any convex subset of $\mathbb{R}^n$.

1.4.2 One-sided upper bound

Our one-sided testing algorithm employs a "gridding-based" approach to decompose the relevant portion of $\mathbb{R}^n$ (namely, those points which are not too far from the origin) into a collection of disjoint cubes. It draws samples and identifies a subset of these cubes as a proxy for the "thickened surface" of the target set; by the structural result sketched above, if the Gaussian volume of this thickened surface is too high, then the one-sided algorithm can safely reject (as the target set cannot be convex). Otherwise the algorithm does random sampling to probe for points which are inside the convex hull of positive examples it has received but are labeled negative (there should be no such points if the target set is indeed convex, so if such a point is identified, the one-sided algorithm can safely reject). If no such points are identified, then the algorithm accepts.

1.4.3   Two-sided upper bound

Finally, the main tool we use to obtain our two-sided testing algorithm is a *learning* algorithm for convex sets with respect to the normal distribution over $\mathbb{R}^n$. The main result of [29] is an (improper) algorithm which learns the class of all convex subsets of $\mathbb{R}^n$ to accuracy ε using $n^{O(\sqrt{n}/\varepsilon^2)}$ independent samples from $\mathcal{N}(0,1)^n$. Using the structural result mentioned above, we show that this can be converted into a *proper* algorithm for learning convex sets under $\mathcal{N}(0,1)^n$, with essentially no increase in the sample complexity. Given this proper learning algorithm, a two-sided algorithm for testing convexity follows from the well-known result of [21] which shows that proper learning for a class of functions implies (two-sided) testability.

2   Preliminaries and Notation

Notation.   We use boldfaced letters such as $\boldsymbol{x}, \boldsymbol{f}, \mathbf{A}$, etc. to denote random variables (which may be real-valued, vector-valued, function-valued, set-valued, etc; the intended type will be clear from the context). We write "$\boldsymbol{x} \leftarrow \mathcal{D}$" to indicate that the random variable $\boldsymbol{x}$ is distributed according to probability distribution $\mathcal{D}$. Given $a, b, c \in \mathbb{R}$ we use $a = b \pm c$ to indicate that $b - c \leq a \leq b + c$.

Geometry.   For $r > 0$, we write $S^{n-1}(r)$ to denote the origin-centered sphere of radius r in $\mathbb{R}^n$ and $\text{Ball}(r)$ to denote the origin-centered ball of radius r in $\mathbb{R}^n$, i.e.,

$$S^{n-1}(r) = \left\{ x \in \mathbb{R}^n : \|x\| = r \right\} \quad \text{and} \quad \text{Ball}(r) = \left\{ x \in \mathbb{R}^n : \|x\| \leq r \right\},$$

where $\|x\|$ denotes the ℓ_2-norm $\|\cdot\|_2$ of x. We also write S^{n-1} for the unit sphere $S^{n-1}(1)$.

Recall that a set $C \subseteq \mathbb{R}^n$ is convex if $x, y \in C$ implies $\alpha x + (1 - \alpha)y \in C$ for all $\alpha \in [0, 1]$. We write $\mathcal{C}_{\text{convex}}$ to denote the class of all convex sets in $\mathbb{R}^n$. Recall that convex sets are Lebesgue measurable. Given a set $C \subseteq \mathbb{R}^n$ we use $\text{Conv}(C)$ to denote the convex hull of C.

For sets $A, B \subseteq \mathbb{R}^n$, we write $A+B$ to denote the Minkowski sum $\{a+b : a \in A \text{ and } b \in B\}$. For a set $A \subseteq \mathbb{R}^n$ and $r > 0$ we write rA to denote the set $\{ra : a \in A\}$. Given a point a and $B \subseteq \mathbb{R}^n$, we use $a + B$ and $B - a$ to denote $\{a\} + B$ and $B + \{-a\}$ for convenience.

Probability.   We use $\mathcal{N}(0,1)^n$ to denote the standard n-dimensional Gaussian distribution with zero mean and identity covariance matrix. We also recall that the probability density function for the one-dimensional Gaussian distribution is

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp(-x^2/2).$$

Sometimes we denote $\mathcal{N}(0,1)^n$ by $\mathcal{N}^n$ for convenience. The squared norm $\|\boldsymbol{x}\|^2$ of $\boldsymbol{x} \leftarrow \mathcal{N}(0,1)^n$ is distributed according to the chi-squared distribution χ_n^2 with n degrees of freedom. The following tail bound for χ_n^2 (see [25]) will be useful:

▶ **Lemma 5** (Tail bound for the chi-squared distribution). *Let $\mathbf{X} \leftarrow \chi_n^2$. Then we have*

$$\mathbf{Pr}\left[|\mathbf{X} - n| \geq tn \right] \leq e^{-(3/16)nt^2}, \quad \text{for all } t \in [0, 1/2].$$

All target sets $S \subseteq \mathbb{R}^n$ to be tested for convexity are assumed to be Lebesgue measurable and we write $\text{Vol}(S)$ to denote $\mathbf{Pr}_{\boldsymbol{x} \leftarrow \mathcal{N}^n}[\boldsymbol{x} \in S]$, the *Gaussian volume* of $S \subseteq \mathbb{R}^n$. Given two Lebesgue measurable subsets $S, C \subseteq \mathbb{R}^n$, we view $\text{Vol}(S \triangle C)$ as the *distance* between S and C, where $S \triangle C$ is the symmetric difference of S and C. Given $S \subseteq \mathbb{R}^n$, we abuse the notation and use S to denote the indicator function of the set, so we may write "$S(x) = 1$" or "$x \in S$" to mean the same thing.

Sample-based property testing.   Given a point $x \in \mathbb{R}^n$, we refer to $(x, S(x)) \in \mathbb{R}^n \times \{0, 1\}$ as a *labeled sample* from a set $S \subseteq \mathbb{R}^n$. A *sample-based testing algorithm for convexity* is a randomized algorithm which is given as input an accuracy parameter $\varepsilon > 0$ and access to an oracle that, each time it is invoked, generates a labeled sample $(\boldsymbol{x}, S(\boldsymbol{x}))$ from the unknown (Lebesgue measurable) *target set* $S \subseteq \mathbb{R}^n$ with $\boldsymbol{x}$ drawn independently each time from $\mathcal{N}(0, 1)^n$. When run with any Lebesgue measurable $S \subseteq \mathbb{R}^n$, such an algorithm must output "accept" with probability at least $2/3$ (over the draws it gets from the oracle and its own internal randomness) if $S \in \mathcal{C}_{\text{convex}}$ and must output "reject" with probability at least $2/3$ if S is ε-*far* from being convex, meaning that for every $C \in \mathcal{C}_{\text{convex}}$ it is the case that $\text{Vol}(S \triangle C) \geq \varepsilon$. (We also refer to an algorithm as an ε-tester for convexity if it works for a specific accuracy parameter ε.) Such a testing algorithm is said to be *one-sided* if whenever it is run on a convex set S it always outputs "accept;" equivalently, such an algorithm can only output "reject" if the labeled samples it receives are not consistent with any convex set. A testing algorithm which is not one-sided is said to be *two-sided*.

Throughout the rest of the paper we reserve the symbol S to denote the unknown target set (a measurable subset of $\mathbb{R}^n$) that is being tested for convexity.

3   Two-sided lower bound

We recall Theorem 3:

▶ **Theorem 3** (Two-sided lower bound). *There exists a positive constant ε_0 such that any two-sided sample-based algorithm that is an ε-tester for convexity over $\mathcal{N}(0, 1)^n$ for some $\varepsilon \leq \varepsilon_0$ must use $2^{\Omega(\sqrt{n})}$ samples.*

Let $q = 2^{0.01\sqrt{n}}$ and let $\varepsilon_0 > 0$ be a constant to be specified later. To prove Theorem 3, we show that no sample-based, q-query (randomized) algorithm A can achieve the following:

Let $S \subset \mathbb{R}^n$ be a target set that is Lebesgue measurable. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q$ be a sequence of q samples drawn from $\mathcal{N}(0, 1)^n$. Upon receiving $((\boldsymbol{x}_i, S(\boldsymbol{x}_i)) : i \in [q])$, A accepts with probability at least $2/3$ when S is convex and rejects with probability at least $2/3$ when S is ε_0-far from convex.

Recall that a pair (x, b) with $x \in \mathbb{R}^n$ and $b \in \{0, 1\}$ is a labeled sample; a sample-based algorithm A is a randomized map from a sequence of q labeled samples to $\{$"accept","reject"$\}$.

3.1   Proof Plan

Assume for contradiction that there is a q-query (randomized) algorithm A that accomplishes the task above. In Section 3.2 we define two probability distributions $\mathcal{D}_{\text{yes}}$ and $\mathcal{D}_{\text{no}}$ such that (1) $\mathcal{D}_{\text{yes}}$ is a distribution over convex sets in $\mathbb{R}^n$ ($\mathcal{D}_{\text{yes}}$ is a distribution over certain convex polytopes that are the intersection of many randomly drawn halfspaces), and (2) $\mathcal{D}_{\text{no}}$ is a probability distribution over sets in $\mathbb{R}^n$ that are Lebesgue measurable ($\mathcal{D}_{\text{no}}$ is actually supported over a finite number of measurable sets in $\mathbb{R}^n$) such that $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{no}}$ is ε_0-far from convex with probability at least $1 - o(1)$.

Given a sequence $x = (x_1, \ldots, x_q)$ of points, we abuse the notation and write

$$S(x) = (S(x_1), \ldots, S(x_q))$$

and use $(x, S(x))$ to denote the sequence of q labeled samples $(x_1, S(x_1)), \ldots, (x_q, S(x_q))$. It then follows from our assumption on A that

$$\Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}; \, \boldsymbol{x} \leftarrow (\mathcal{N}^n)^q} \big[A \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] \geq 2/3 \qquad \text{and}$$

$$\Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}; \, \boldsymbol{x} \leftarrow (\mathcal{N}^n)^q} \big[A \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] \leq 1/3 + o(1).$$

where we use $\boldsymbol{x} \leftarrow (\mathcal{N}^n)^q$ to denote a sequence of q points sampled independently from $\mathcal{N}^n$ and we usually skip the $\leftarrow (\mathcal{N}^n)^q$ part in the subscript when it is clear from the context. Since A is a mixture of deterministic algorithms, there exists a deterministic sample-based, q-query algorithm A' (equivalently, a deterministic map from sequences of q labeled samples to $\{\text{"Yes"}, \text{"No"}\}$) with

$$\Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}; \, \boldsymbol{x}} \big[A' \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] - \Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}; \, \boldsymbol{x}} \big[A' \text{ accepts } (\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x})) \big] \geq 1/3 - o(1). \tag{2}$$

Let $\mathcal{E}_{\mathsf{yes}}$ (or $\mathcal{E}_{\mathsf{no}}$) be the distribution of $(\boldsymbol{x}, \boldsymbol{S}(\boldsymbol{x}))$, where $\boldsymbol{x} \leftarrow (\mathcal{N}^n)^q$ and $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ (or $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$, respectively). Both of them are distributions over sequences of q labeled samples. Then the LHS of (2), for any deterministic sample-based, q-query algorithm A', is at most the total variation distance between $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}$. We prove the following key lemma, which leads to a contradiction.

▶ **Lemma 6.** *The total variation distance between $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}$ is $o(1)$.*

To prove Lemma 6, it will be convenient for us to introduce a third distribution $\mathcal{E}_{\mathsf{no}}^*$ over sequences of q labeled samples, where $(\boldsymbol{x}, \mathbf{b}) \leftarrow \mathcal{E}_{\mathsf{no}}^*$ is drawn by first sampling a sequence of q points $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)$ from $\mathcal{N}^n$ independently and then for each $\boldsymbol{x}_i$, its label $\mathbf{b}_i$ is set to be 1 independently with a probability that depends only on $\|\boldsymbol{x}_i\|$ (see Section 3.2). Lemma 6 follows from the following two lemmas by the triangle inequality.

▶ **Lemma 7.** *The total variation distance between $\mathcal{E}_{\mathsf{no}}$ and $\mathcal{E}_{\mathsf{no}}^*$ is $o(1)$.*

▶ **Lemma 8.** *The total variation distance between $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}^*$ is $o(1)$.*

The rest of the section is organized as follows. We define $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ (which are used to define $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}$) as well as $\mathcal{E}_{\mathsf{no}}^*$ in Section 3.2 and prove the necessary properties about $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ as well as Lemma 7. We prove Lemma 8 in Sections 3.3 and Appendix A.

3.2 The Distributions

Let $r = \Theta(n^{1/4})$ be a parameter to be specified later, and let $N = 2^{\sqrt{n}}$. We start with the definition of $\mathcal{D}_{\mathsf{yes}}$. A random set $\boldsymbol{S} \subset \mathbb{R}^n$ is drawn from $\mathcal{D}_{\mathsf{yes}}$ using the following procedure:
1. We sample a sequence of N points $\mathbf{y}_1, \ldots, \mathbf{y}_N$ from $S^{n-1}(r)$ independently and uniformly at random. Each point $\mathbf{y}_i$ defines a halfspace

$$\boldsymbol{h}_i = \big\{ x \in \mathbb{R}^n : x \cdot \mathbf{y}_i \leq r^2 \big\}.$$

2. The set $\boldsymbol{S}$ is then the intersection of $\boldsymbol{h}_i$, $i \in [N]$ (this is always nonempty as indeed $\text{Ball}(r)$ is contained in $\boldsymbol{S}$).

It is clear from the definition that $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ is always a convex set.

Next we define $\mathcal{E}_{\mathsf{no}}^*$ (instead of $\mathcal{D}_{\mathsf{no}}$), a distribution over sequences of q labeled samples $(\boldsymbol{x}, \mathbf{b})$. To this end, we use $\mathcal{D}_{\mathsf{yes}}$ to define a function $\rho : \mathbb{R}_{\geq 0} \to [0, 1]$ as follows:

$$\rho(t) = \Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}} \big[(t, 0, \ldots, 0) \in \boldsymbol{S} \big].$$

Due to the symmetry of $\mathcal{D}_{\text{yes}}$ and $\mathcal{N}^n$, the value $\rho(t)$ is indeed the probability that a point $x \in \mathbb{R}^n$ at distance t from the origin lies in $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}$. To draw a sequence of q labeled samples $(\boldsymbol{x}, \mathbf{b}) \leftarrow \mathcal{E}^*_{\text{no}}$, we independently draw q random points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q \leftarrow \mathcal{N}^n$ and then independently set $\mathbf{b}_i = 1$ with probability $\rho(\|\boldsymbol{x}_i\|)$ and $\mathbf{b}_i = 0$ with probability $1 - \rho(\|\boldsymbol{x}_i\|)$.

Given $\mathcal{D}_{\text{yes}}$ and $\mathcal{E}^*_{\text{no}}$, Lemma 8 shows that information-theoretically no sample-based algorithm can distinguish a sequence of q labeled samples $(\boldsymbol{x}, \mathbf{b})$ with $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}$, $\boldsymbol{x} \leftarrow (\mathcal{N}^n)^q$, and $\mathbf{b} = \boldsymbol{S}(\boldsymbol{x})$ from a sequence of q labeled samples drawn from $\mathcal{E}^*_{\text{no}}$. While the marginal distribution of each labeled sample is the same for the two cases, the former is generated in a correlated fashion using the underlying random convex $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}$ while the latter is generated independently.

Finally we define the distribution $\mathcal{D}_{\text{no}}$, prove Lemma 7, and show that a set drawn from $\mathcal{D}_{\text{no}}$ is far from convex with high probability. To define $\mathcal{D}_{\text{no}}$, we let $M \geq 2^{\sqrt{n}}$ be a large enough integer to be specified later. With M fixed, we use

$$0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = 2\sqrt{n}$$

to denote a sequence of numbers such that the origin-centered ball $\text{Ball}(2\sqrt{n})$ is partitioned into M *shells* $\text{Ball}(t_i) \setminus \text{Ball}(t_{i-1})$, $i \in [M]$, and all the M shells have the same probability mass under $\mathcal{N}^n$. By spherical coordinates, it means that the following integral takes the same value for all i:

$$\int_{t_{i-1}}^{t_i} \phi(x, 0, \ldots, 0) x^{n-1} dx, \tag{3}$$

where ϕ denotes the density function of $\mathcal{N}^n$. We show below that when M is large enough,

$$|\rho(x) - \rho(t_i)| \leq 2^{-\sqrt{n}}, \tag{4}$$

for any $i \in [M]$ and any $x \in [t_{i-1}, t_i]$. We will fix such an M and use it to define $\mathcal{D}_{\text{no}}$. (Our results are not affected by the size of M as a function of n; we only need it to be finite.)

To show that (4) holds when M is large enough, we need the continuity of the function ρ, which follows directly from the explicit expression for ρ given later in (6).

▶ **Lemma 9.** *The function* $\rho : \mathbb{R}_{\geq 0} \to [0, 1]$ *is continuous.*

Since ρ is continuous, it is continuous over $[0, 2\sqrt{n}]$. Since $[0, 2\sqrt{n}]$ is compact, ρ is also uniformly continuous over $[0, 2\sqrt{n}]$. Also note that $\max_{i \in [M]}(t_i - t_{i-1})$ goes to 0 as M goes to $+\infty$. It follows that (4) holds when M is large enough.

With $M \geq 2^{\sqrt{n}}$ fixed, a random set $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{no}}$ is drawn as follows. Start with $\boldsymbol{S} = \emptyset$ and for each $i \in [M]$, add the ith shell $\text{Ball}(t_i) \setminus \text{Ball}(t_{i-1})$ to $\boldsymbol{S}$ independently with probability $\rho(t_i)$. Thus an outcome of $\boldsymbol{S}$ is a union of some of the shells and $\mathcal{D}_{\text{no}}$ is supported over 2^M different sets.

Recall the definition of $\mathcal{E}_{\text{yes}}$ and $\mathcal{E}_{\text{no}}$ using $\mathcal{D}_{\text{yes}}$ and $\mathcal{D}_{\text{no}}$. We now prove Lemma 7.

Proof of Lemma 7. Let $x = (x_1, \ldots, x_q)$ be a sequence of q points in $\mathbb{R}^n$. We say x is *bad* if either (1) at least one point lies outside of $\text{Ball}(2\sqrt{n})$ or (2) there are two points that lie in the same shell of $\mathcal{D}_{\text{no}}$; we say x is *good* otherwise. We first claim that $\boldsymbol{x} \leftarrow (\mathcal{N}^n)^q$ is bad with probability $o(1)$. To see this, we have from Lemma 5 that event (1) occurs with probability $o(1)$, and from $M \geq 2^{\sqrt{n}}$ and $q = 2^{0.01\sqrt{n}}$ that event (2) occurs with probability $o(1)$. The claim follows from a union bound.

Given that $\boldsymbol{x} \leftarrow (\mathcal{N}^n)^q$ is good with probability $1 - o(1)$, it suffices to show that for any good q-tuple x, the total variation distance between (1) $\boldsymbol{S}(x)$ with $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{no}}$ and (2)

$\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)$ with each bit $\mathbf{b}_i$ being 1 with probability $\rho(\|x_i\|)$ independently, is $o(1)$. Let $\ell_i \in [M]$ be the index of the shell that x_i lies in. Since x is good (and thus, all points lie in different shells), $\boldsymbol{S}(x)$ has the ith bit being 1 independently with probability $\rho(t_{\ell_i})$; for the other distribution, the probability is $\rho(\|x_i\|)$. Using the subadditivity of total variation distance (i.e., the fact that the d_{TV} between two sequences of independent random variables is upper bounded by the sum of the d_{TV} between each pair) as well as (4), we have

$$d_{\mathrm{TV}}(\boldsymbol{S}(x), \mathbf{b}) \leq q \cdot 2^{-\sqrt{n}} = o(1).$$

This finishes the proof. ◀

The next lemma shows that $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$ is ε_0-far from convex with probability $1 - o(1)$, for some positive constant ε_0. In the proof of the lemma we fix both the constant ε_0 and our choice of $r = \Theta(n^{1/4})$. (We remind the reader that ρ and $\mathcal{D}_{\mathsf{no}}$ depend on the value of r.)

▶ **Lemma 10.** *There exist a real value $r = \Theta(n^{1/4})$ with $e^{r^2/2} \geq N/n$ and a positive constant ε_0 such that a set $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$ is ε_0-far from convex with probability at least $1 - o(1)$.*

Proof. We need the following claim but delay its proof to the end of the subsection:

▶ **Claim 11.** *There exist an $r = \Theta(n^{1/4})$ with $e^{r^2/2} \geq N/n$ and a constant $c \in (0, 1/2)$ such that $c < \rho(x) < 1 - c$ for all $x \in [\sqrt{n} - 10, \sqrt{n} + 10]$.*

Let $K \subset [M]$ denote the set of all integers k such that $[t_{k-1}, t_k] \subseteq [\sqrt{n} - 10, \sqrt{n} + 10]$ (note that K is a set of consecutive integers). Observe that (1) the total probability mass of all shells $k \in K$ is at least $\Omega(1)$ (by Lemma 5), and (2) the size $|K|$ is at least $\Omega(M)$ (which follows from (1) and the fact that all shells have the same probability mass).

Consider the following 1-dimensional scenario. We have $|K|$ intervals $[t_{k-1}, t_k]$ and draw $\boldsymbol{T}$ by including each interval independently with probability $\rho(t_k)$. We prove the following claim:

▶ **Claim 12.** *The random set $\boldsymbol{T}$ satisfies the following property with probability at least $1 - o(1)$: For any interval $I \subseteq \mathbb{R}_{\geq 0}$, either I contains $\Omega(M)$ intervals $[t_{k-1}, t_k]$ that are not included in $\boldsymbol{T}$, or $\bar{I}$ contains $\Omega(M)$ intervals $[t_{k-1}, t_k]$ included in $\boldsymbol{T}$.*

Proof. First note that it suffices to consider intervals $I \subseteq \cup_{k \in K}[t_{k-1}, t_k]$ and moreover, we may further assume that both endpoints of I come from endpoints of $[t_{k-1}, t_k]$, $k \in K$. (In other words, for a given outcome T of $\boldsymbol{T}$, if there exists an interval I that violates the condition, i.e., both I and $\bar{I}$ contain fewer than $\Omega(M)$ intervals, then there is such an interval I with both ends from end points of $[t_{k-1}, t_k]$). This assumption allows us to focus on $|K|^2 \leq M^2$ many possibilities for I (as we will see below, our argument applies a union bound over these K^2 possibilities).

Given a candidate such interval I, we consider two cases. If I contains $\Omega(M)$ intervals $[t_{k-1}, t_k]$, $k \in K$, then it follows from Claim 11 and a Chernoff bound that I contains at least $\Omega(M)$ intervals not included in $\boldsymbol{T}$ with probability $1 - 2^{-\Omega(M)}$. On the other hand, if $\bar{I}$ contains $\Omega(M)$ intervals, then the same argument shows that $\bar{I}$ contains $\Omega(M)$ interals included in $\boldsymbol{T}$ with probability $1 - 2^{-\Omega(M)}$. The claim follows from a union bound over all the $|K|^2$ possibilities for I. ◀

We return to the n-dimensional setting and consider the intersection of $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$ with a ray starting from the origin. Note that the intersection of the ray and any convex set is an interval on the ray. As a result, Claim 12 shows that with probability at least $1 - o(1)$ (over the draw of $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{no}}$), the intersection of any convex set with any ray either contains

$\Omega(M)$ intervals $[t_{k-1}, t_k]$ such that shell $k \in K$ is not included in $\boldsymbol{S}$, or misses $\Omega(M)$ intervals $[t_{k-1}, t_k]$ such that shell $k \in K$ is included in $\boldsymbol{S}$. Since by (1) above shells $k \in K$ together have $\Omega(1)$ probability mass under $\mathcal{N}^n$ and each shell contains the same probability mass, we have that with probability $1 - o(1)$, $\boldsymbol{S}$ is ε_0-far from any convex set for some constant $\varepsilon_0 > 0$. (A more formal argument can be given by performing integration using spherical coordinates and applying (3).) ◀

Proof of Claim 11. We start with the choice of r. Let

$$\alpha = \sqrt{n} - 10 \quad \text{and} \quad \beta = \sqrt{n} + 10.$$

Let $\mathrm{cap}(t)$ denote the fractional surface area of the spherical cap $S^{n-1} \cap \{x : x_1 \geq t\}$, i.e.,

$$\mathrm{cap}(t) = \Pr_{\boldsymbol{x} \leftarrow S^{n-1}} \left[\boldsymbol{x}_1 \geq t\right].$$

So cap is a continuous, strictly decreasing function over $[0, 1]$. Since $\mathrm{cap}(0) = 1/2$ and $\mathrm{cap}(1) = 0$, there is a unique $r \in (0, \alpha)$ such that $\mathrm{cap}(r/\alpha) = 1/N = 2^{-\sqrt{n}}$. Below we show that $r = \Theta(n^{1/4})$ and fix it in the rest of the proof. First recall the following explicit expression (see e.g. [29]):

$$\mathrm{cap}(t) = a_n \int_t^1 \left(\sqrt{1 - z^2}\right)^{n-3} dz,$$

where $a_n = \Theta(n^{1/2})$ is a parameter that only depends on n. We also recall the following inequalities from [29] about $\mathrm{cap}(t)$:

$$\mathrm{cap}(t) \leq e^{-nt^2/2}, \quad \text{for all } t \in [0, 1]; \quad \mathrm{cap}(t) \geq \Omega\left(t \cdot e^{-nt^2/2}\right), \quad \text{for } t = O(1/n^{1/4}). \quad (5)$$

By our choice of α and the monotonicity of the cap function, we have $r = \Theta(n^{1/4})$ and

$$1/N = \mathrm{cap}(r/\alpha) \geq \Omega(1/n^{1/4}) \cdot e^{-n(r/\alpha)^2/2}$$
$$\geq \Omega(1/n^{1/4}) \cdot e^{-(r^2/2)(1+O(1/\sqrt{n}))} = \Omega(1/n^{1/4}) \cdot e^{-r^2/2}$$

(using $r = \Theta(n^{1/4})$ for the last inequality), and thus, we have $e^{r^2/2} \geq N/n$.

Next, using the function cap we have the following expression for ρ:

$$\rho(x) = \left(1 - \mathrm{cap}\left(\frac{r}{x}\right)\right)^N. \quad (6)$$

As a side note, ρ is continuous and thus, Lemma 9 follows. Since cap is strictly decreasing, we have that ρ is strictly decreasing as well. To finish the proof it suffices to show that there is a constant $c \in (0, 1/2)$ such that $\rho(\alpha) < 1 - c$ and $\rho(\beta) \geq c$.

$$\rho(\alpha) = (1 - 1/N)^N \approx e^{-1}$$

by our choice of r. In the rest of the proof we show that

$$\mathrm{cap}\left(\frac{r}{\beta}\right) \leq a \cdot \mathrm{cap}\left(\frac{r}{\alpha}\right) = \frac{a}{N}, \quad (7)$$

for some positive constant a. It follows immediately that

$$\rho(\beta) = \left(1 - \mathrm{cap}\left(\frac{r}{\beta}\right)\right)^N \geq \left(1 - \frac{a}{N}\right)^N \geq \left(e^{-2a/N}\right)^N = e^{-2a},$$

using $1 - x \geq e^{-2x}$ for $0 \leq x \ll 1$, and this finishes the proof of the claim.
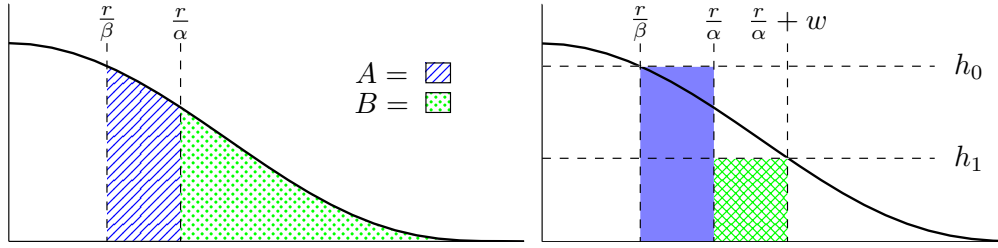
■ **Figure 1** A plot of the integrand $(\sqrt{1 - z^2})^{(n-3)}$. Area A is $\text{cap}(r/\beta) - \text{cap}(r/\alpha)$ and area B is $\text{cap}(r/\alpha)$. The rectangles on the right are an upper bound of A and a lower bound of B.

Finally we prove (7). Let

$$w = \frac{r}{\alpha} - \frac{r}{\beta} = \Theta\left(\frac{1}{n^{3/4}}\right)$$

since $r = \Theta(n^{1/4})$. Below we show that

$$\int_{r/\beta}^{r/\alpha} \left(\sqrt{1 - z^2}\right)^{n-3} dz \leq a' \cdot \int_{r/\alpha}^{r/\alpha+w} \left(\sqrt{1 - z^2}\right)^{n-3} dz, \tag{8}$$

for some positive constant a'. It follows that

$$\text{cap}\left(\frac{r}{\beta}\right) - \text{cap}\left(\frac{r}{\alpha}\right) \leq a' \cdot \text{cap}\left(\frac{r}{\alpha}\right)$$

and implies (7) by setting $a = a' + 1$. For (8), note that the ratio of the $[r/\beta, r/\alpha]$-integration over the $[r/\alpha, r/\alpha+w]$-integration is at most

$$\left(\frac{\sqrt{1 - (r/\beta)^2}}{\sqrt{1 - (r/\beta + 2w)^2}}\right)^{n-3}$$

as the length of the two intervals are the same and the function $(\sqrt{1 - z^2})^{n-3}$ is strictly decreasing. Figure 1 illustrates this calculation.

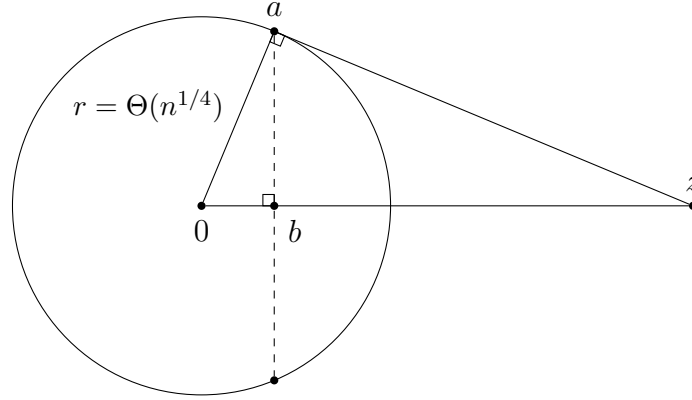Let $\tau = r/\beta = \Theta(1/n^{1/4})$. We can rewrite the above as

$$\left(\frac{1 - \tau^2}{1 - (\tau + 2w)^2}\right)^{(n-3)/2} = \left(1 + \frac{4\tau w + 4w^2}{1 - (\tau + 2w)^2}\right)^{(n-3)/2} = \left(1 + O\left(\frac{1}{n}\right)\right)^{(n-3)/2} = O(1).$$

This finishes the proof of the claim. ◀

3.3 Distributions $\mathcal{E}_{\text{yes}}$ and $\mathcal{E}_{\text{no}}^*$ are close

In the rest of the section we show that the total variation distance between $\mathcal{E}_{\text{yes}}$ and $\mathcal{E}_{\text{no}}^*$ is $o(1)$ and thus prove Lemma 8. Let $z = (z_1, \ldots, z_q)$ be a sequence of q points in $\mathbb{R}^n$. We use $\mathcal{E}_{\text{yes}}(z)$ to denote the distribution of labeled samples from $\mathcal{E}_{\text{yes}}$, conditioning on the samples being z, i.e., $(z, \boldsymbol{S}(z))$ with $\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}$. We let $\mathcal{E}_{\text{no}}^*(z)$ denote the distribution of labeled samples from $\mathcal{E}_{\text{no}}^*$, conditioning on the samples being z, i.e., $(z, \mathbf{b})$ where each $\mathbf{b}_i$ is 1 independently with probability $\rho(\|z_i\|)$. Then

$$d_{\text{TV}}(\mathcal{E}_{\text{yes}}, \mathcal{E}_{\text{no}}^*) = \mathbf{E}_{\mathbf{z} \leftarrow (\mathcal{N}^n)^q}\left[d_{\text{TV}}(\mathcal{E}_{\text{yes}}(\mathbf{z}), \mathcal{E}_{\text{no}}^*(\mathbf{z}))\right]. \tag{9}$$

Figure 2 The fractional surface area of $\mathrm{cover}(z)$, $\mathrm{fsa}(\mathrm{cover}(z))$, is the fraction of $S^{n-1}(r)$ to the right of the dashed line. By similarity of triangles $0az$ and $0ba$, scaling down to the unit sphere, we get (10).

We split the proof of Lemma 8 into two steps. We first introduce the notion of *typical* sequences z of q points and show in this subsection that with probability $1 - o(1)$, $\mathbf{z} \leftarrow (\mathcal{N}^n)^q$ is typical. In the next subsection we show that $d_{\mathrm{TV}}(\mathcal{E}_{\mathsf{yes}}(z), \mathcal{E}_{\mathsf{no}}^*(z))$ is $o(1)$ when z is typical. It follows from (9) that $d_{\mathrm{TV}}(\mathcal{E}_{\mathsf{yes}}, \mathcal{E}_{\mathsf{no}}^*) = o(1)$. We start with the definition of typical sequences.

Given a point $z \in \mathbb{R}^n$, we are interested in the *fraction* of points y (in terms of the area) in $S^{n-1}(r)$ such that $z \cdot y > r^2$. This is because if any such point y is sampled in the construction of $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$, then $z \notin \boldsymbol{S}$. This is illustrated in Figure 2. We refer to the set of such points y as the (*spherical*) *cap covered by* z and we write $\mathrm{cover}(z)$ to denote it. (Note that $\mathrm{cover}(z) = \emptyset$ if $\|z\| \leq r$.)

Given a subset H of $S^{n-1}(r)$ (such as $\mathrm{cover}(z)$), we use $\mathrm{fsa}(H)$ to denote the fractional surface area of H with respect to $S^{n-1}(r)$. Using Figure 2 and elementary geometry, we have the following connection between the fractional surface area of $\mathrm{cover}(z)$ and the cap function (for S^{n-1}):

$$\mathrm{fsa}\big(\mathrm{cover}(z)\big) = \mathrm{cap}\big(r/\|z\|\big). \tag{10}$$

We are now ready to define typical sequences.

▶ **Definition 13.** We say a sequence $z = (z_1, \ldots, z_q)$ of q points in $\mathbb{R}^n$ is *typical* if
1. For every point z_i, we have

$$\mathrm{fsa}\big(\mathrm{cover}(z_i)\big) \in \left[e^{-0.51\,r^2}, e^{-0.49\,r^2} \right]. \tag{11}$$

2. For every $i \neq j$, we have

$$\mathrm{fsa}\big(\mathrm{cover}(z_i) \cap \mathrm{cover}(z_j)\big) \leq e^{-0.96\,r^2}.$$

The first condition of typicality essentially says that every z_i is not too close to and not too far away from the origin (so that we have a relatively tight bound on the fractional surface area of the cap covered by z_i). The second condition says that the caps covered by two points z_i and z_j have very little intersection. We prove the following lemma:

▶ **Lemma 14.** $\mathbf{z} \leftarrow (\mathcal{N}^n)^q$ *is typical with probability at least* $1 - o(1)$.

Proof. We show that $\mathbf{z}$ satisfies each of the two conditions with probability $1 - o(1)$. The lemma then follows from a union bound.

For the first condition, we let $c^* = 0.001$ be a sufficiently small constant. We have from Lemma 5 and a union bound that every $\mathbf{z}_i$ satisfies $(1 - c^*)\sqrt{n} \leq \|\mathbf{z}_i\| \leq (1 + c^*)\sqrt{n}$ with probability $1 - o(1)$. When this happens, we have (11) for every $\mathbf{z}_i$ using (5) and the upper bound of $\mathrm{cap}(t) \leq e^{-nt^2/2}$.

For the second condition, we note that the argument used in the first part implies that

$$\mathbf{E}_{\mathbf{z}_i \leftarrow \mathcal{N}^n} \left[\mathrm{fsa}\big(\mathrm{cover}(\boldsymbol{z}_i)\big) \right] \leq e^{-0.49\,r^2}.$$

Let x_0 be a fixed point in $S^{n-1}(r)$. Viewing the fsa as the following probability:

$$\mathrm{fsa}\big(\mathrm{cover}(z_i)\big) = \Pr_{\boldsymbol{x} \leftarrow S^{n-1}(r)} \big[\boldsymbol{x} \in \mathrm{cover}(z_i) \big],$$

we have

$$\begin{aligned}
e^{-0.49\,r^2} &\geq \mathbf{E}_{\mathbf{z}_i \leftarrow \mathcal{N}^n} \left[\mathrm{fsa}\big(\mathrm{cover}(\boldsymbol{z}_i)\big) \right] && (12) \\
&= \mathbf{E}_{\mathbf{z}_i} \left[\Pr_{\boldsymbol{x} \leftarrow S^{n-1}(r)} \big[\boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_i) \big] \right] \\
&= \Pr_{\boldsymbol{x}, \mathbf{z}_i} \big[\boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_i) \big] = \Pr_{\mathbf{z}_i} \big[x_0 \in \mathrm{cover}(\mathbf{z}_i) \big],
\end{aligned}$$

where the last equation follows by sampling $\boldsymbol{x}$ first and spherical and Gaussian symmetry.

Similarly we can express the fractional surface area of $\mathrm{cover}(z_i) \cap \mathrm{cover}(z_j)$ as

$$\mathrm{fsa}\big(\mathrm{cover}(z_i) \cap \mathrm{cover}(z_j)\big) = \Pr_{\boldsymbol{x} \leftarrow S^{n-1}(r)} \big[\boldsymbol{x} \in \mathrm{cover}(z_i) \text{ and } \boldsymbol{x} \in \mathrm{cover}(z_j) \big].$$

We consider the expectation over $\mathbf{z}_i$ and $\mathbf{z}_j$ drawn independently from $\mathcal{N}^n$:

$$\begin{aligned}
&\mathbf{E}_{\mathbf{z}_i, \mathbf{z}_j} \left[\mathrm{fsa}\big(\mathrm{cover}(\mathbf{z}_i) \cap \mathrm{cover}(\mathbf{z}_j)\big) \right] \\
&= \mathbf{E}_{\mathbf{z}_i, \mathbf{z}_j} \left[\Pr_{\boldsymbol{x} \leftarrow S^{n-1}(r)} \big[\boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_i) \text{ and } \boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_j) \big] \right] \\
&= \Pr_{\boldsymbol{x}, \mathbf{z}_i, \mathbf{z}_j} \big[\boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_i) \text{ and } \boldsymbol{x} \in \mathrm{cover}(\mathbf{z}_j) \big] \\
&= \Pr_{\mathbf{z}_i} \big[x_0 \in \mathrm{cover}(\mathbf{z}_i) \big] \cdot \Pr_{\mathbf{z}_j} \big[x_0 \in \mathrm{cover}(\mathbf{z}_j) \big],
\end{aligned}$$

where the last equation follows by sampling $\boldsymbol{x}$ first, independence of $\mathbf{z}_i, \mathbf{z}_j$, and symmetry.

By (12), the expectation of $\mathrm{fsa}(\mathrm{cover}(\mathbf{z}_i) \cap \mathrm{cover}(\mathbf{z}_j))$ is at most $e^{-0.98\,r^2}$, and hence by Markov's inequality, the probability of it being at least $e^{-0.96\,r^2}$ is at most $e^{-0.02\,r^2}$. Using $e^{r^2} \geq (N/n)^2$ and a union bound, the probability of one of the pairs having the fsa at least $e^{-0.96\,r^2}$ is at most

$$q^2 \cdot e^{-0.02r^2} \leq 2^{0.02\,\sqrt{n}} \cdot (n/N)^{0.04} = o(1),$$

since $q = 2^{0.01\sqrt{n}}$ and $N = 2^{\sqrt{n}}$. This finishes the proof of the lemma. ◀

We prove the following lemma in Appendix A to finish the proof of Lemma 8.

▶ **Lemma 15.** *For every typical sequence z of q points, we have*

$$d_{TV}\big(\mathcal{E}_{yes}(z), \mathcal{E}_{no}^*(z)\big) = o(1).$$

4 One-sided lower bound

We recall Theorem 1:

▶ **Theorem 1** (One-sided lower bound). *Any one-sided sample-based algorithm that is an ε-tester for convexity over $\mathcal{N}(0,1)^n$ for some $\varepsilon < 1/2$ must use $2^{\Omega(n)}$ samples.*

We say a finite set $\{x^1, \ldots, x^M\} \subset \mathbb{R}^n$ is *shattered* by $\mathcal{C}_{\mathrm{convex}}$ if for every $(b_1, \ldots, b_M) \in \{0,1\}^M$ there is a convex set $C \in \mathcal{C}_{\mathrm{convex}}$ such that $C(x^i) = b_i$ for all $i \in [M]$. Theorem 1 follows from the following lemma:

▶ **Lemma 16.** *There is an absolute constant $c > 0$ such that for $M = 2^{cn}$, it holds that*

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M\} \text{ is shattered by } \mathcal{C}_{\mathrm{convex}} \right] \geq 1 - o(1).$$

Proof of Theorem 1 using Lemma 16. Suppose that A were a one-sided sample-based algorithm for ε-testing $\mathcal{C}_{\mathrm{convex}}$ using at most M samples. Fix a set S that is ε-far from $\mathcal{C}_{\mathrm{convex}}$ to be the unknown target subset of $\mathbb{R}^n$ that is being tested.[1] Since S is ε-far from convex, it must be the case that

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[A \text{ rejects } (\boldsymbol{x}^1, S(\boldsymbol{x}^1)), \ldots, (\boldsymbol{x}^M, S(\boldsymbol{x}^M)) \right] \geq 2/3. \tag{13}$$

But Lemma 16 together with the one-sidedness of A imply that

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\text{for any } (b^1, \ldots, b^M) \in \{0,1\}^M, A \text{ rejects } (\boldsymbol{x}^1, b^1), \ldots, (\boldsymbol{x}^M, b^M) \right] \leq o(1),$$

since A can only reject if the labeled samples are not consistent with any convex set, which implies that A cannot reject when $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M\}$ is shattered by $\mathcal{C}_{\mathrm{convex}}$. This contradicts with (13) and finishes the proof of the lemma. ◀

In the next subsection we prove Lemma 16 for $c = 1/500$.

4.1 Proof of Lemma 16

Let $M = 2^{cn}$ with $c = 1/500$. We prove the following lemma:

▶ **Lemma 17.** *For $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$ drawn independently from $\mathcal{N}(0,1)^n$, with probability $1 - o(1)$ it is the case that for all $i \in [M]$, no $\boldsymbol{x}^i$ lies in $\mathsf{Conv}(\{\boldsymbol{x}^j : j \in [M] \setminus i\})$.*

If $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$ are such that no $\boldsymbol{x}^i$ lies in $\mathsf{Conv}(\{\boldsymbol{x}^j : j \in [M] \setminus i\})$, then given any tuple $(b^1, \ldots, b^M)$, by taking $C = \mathsf{Conv}(\{\boldsymbol{x}^i : b^i = 1\})$ we see that there is a convex set C such that $C(\boldsymbol{x}^i) = b^i$ for all $i \in [M]$. Thus to establish Lemma 16 it suffices to prove Lemma 17.

To prove Lemma 17, it suffices to show that for each fixed $j \in [M]$ we have

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\boldsymbol{x}^j \in \mathsf{Conv}(\{\boldsymbol{x}^k : k \in [M] \setminus \{j\}\}) \right] \leq M^{-2} \tag{14}$$

---

[1] An example of such a subset S is as follows (we define it as a function $S : \mathbb{R}^n \to \{0,1\}$): Given an odd integer $N > (1/2 - \varepsilon)^{-1} - 1$, let $-\infty = \tau_0 < \tau_1 < \cdots < \tau_N < \tau_{N+1} = +\infty$ be values such that $\mathbf{Pr}_{\boldsymbol{z} \leftarrow \mathcal{N}(0,1)}[\boldsymbol{z} \leq \tau_i] = i/(N+1)$, and let $S : \mathbb{R}^n \to \{0,1\}$ be the function defined by $S(x_1, \ldots, x_n) = \mathbf{1}[i \text{ is even}]$, where $i \in \{0, \ldots, N\}$ is the unique value such that $\tau_i \leq x_1 < \tau_{i+1}$. Fix any $z = (z_2, \ldots, z_n) \in \mathbb{R}^{n-1}$ and we let $S_z : \mathbb{R} \to \{0,1\}$ be the function defined as $S_z(x_1) = S(x_1, z_2, \ldots, z_n)$. An easy argument gives that S_z is $(1/2 - 1/(N+1))$-far (and hence ε-far) from every convex subset of $\mathbb{R}$, and it follows by averaging (using the fact that the restriction of any convex subset of $\mathbb{R}^n$ to a line is a convex subset of $\mathbb{R}$) that S is ε-far from $\mathcal{C}_{\mathrm{convex}}$.

since given this a union bound implies that

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\text{for some } j \in [M], \; \boldsymbol{x}^j \text{ lies in } \mathsf{Conv}(\{\boldsymbol{x}^k : k \in [M] \setminus \{j\}\}) \right] \leq M^{-1} = o(1).$$

By symmetry, to establish (14) it suffices to show that

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\boldsymbol{x}^M \in \mathsf{Conv}(\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{M-1}\}) \right] \leq M^{-2}. \tag{15}$$

In turn (15) follows from the following inequalities (v is a fixed unit vector in the second)

$$\Pr_{\boldsymbol{x} \leftarrow \mathcal{N}(0,1)^n} \left[\|\boldsymbol{x}\| \leq \sqrt{n}/10 \right] < \frac{1}{2}M^{-2} \quad \text{and} \quad \Pr_{\boldsymbol{x} \leftarrow \mathcal{N}(0,1)^n} \left[\boldsymbol{x} \cdot v \geq \sqrt{n}/10 \right] < \frac{1}{2}M^{-3}. \tag{16}$$

The first inequality follows from Lemma 5 using $c = 1/500$. For the second, by the spherical symmetry of $\mathcal{N}(0,1)^n$ we may take $v = (1, 0, \ldots, 0)$. Recall the standard Gaussian tail bound

$$\Pr_{\boldsymbol{z} \leftarrow \mathcal{N}(0,1)} \left[\boldsymbol{z} \geq t \right] \leq e^{-t^2/2}$$

for $t \geq 0$. This gives us that

$$\Pr_{\boldsymbol{x} \leftarrow \mathcal{N}(0,1)^n} \left[\boldsymbol{x} \cdot v \geq \sqrt{n}/10 \right] \leq e^{-n/200} < \frac{1}{2}M^{-3},$$

again using that $M = 2^{cn}$ and $c = 1/500$.

Finally, to see that (15) follows from (16), we observe first that by the first inequality we may assume that $\|\boldsymbol{x}^M\| > \sqrt{n}/10$ (at the cost of failure probability at most $M^{-2}/2$ towards (15)); fix any such outcome x^M of $\boldsymbol{x}^M$. By a union bound over $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{M-1}$ and the second inequality, we have

$$\Pr_{\boldsymbol{x}^i \leftarrow \mathcal{N}(0,1)^n} \left[\text{any } i \in [M-1] \text{ has } \boldsymbol{x}^i \cdot \frac{x^M}{\|x^M\|} \geq \sqrt{n}/10 \right] < \frac{1}{2}M^{-2}.$$

But if every $\boldsymbol{x}^i$ has $\boldsymbol{x}^i \cdot (x^M/\|x^M\|) < \sqrt{n}/10 < \|x^M\|$, then $x^M \notin \mathsf{Conv}(\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{M-1}\})$.

References

1. Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 3591–3599, 2015.

2. Michal Adamaszek, Artur Czumaj, and Christian Sohler. Testing monotone continuous distributions on high-dimensional real cubes. In *SODA*, pages 56–65, 2010.

3. N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron. Testing Reed-Muller Codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

4. Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 496–505, 2007.

5. Noga Alon, Rani Hod, and Amit Weinstein. On active and passive testing. *Combinatorics, Probability & Computing*, 25(1):1–20, 2016.

6. Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 21–30, 2012.

7. K. Ball. The Reverse Isoperimetric Problem for Gaussian Measure. *Discrete and Computational Geometry*, 10:411–420, 1993.

8   Keith Ball. An elementary introduction to modern convex geometry. In *Flavors of Geometry*, pages 1–58. MSRI Publications, 1997.

9   Tugkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th Symposium on Theory of Computing*, pages 381–390, 2004.

10   Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. The power and limitations of uniform samples in testing properties of figures. In *36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2016, December 13-15, 2016, Chennai, India*, pages 45:1–45:14, 2016.

11   Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Testing convexity of figures under the uniform distribution. In *32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA*, pages 17:1–17:15, 2016.

12   Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Tolerant testers of image properties. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 90:1–90:14, 2016.

13   Arnab Bhattacharyya, Eldar Fischer, Ronitt Rubinfeld, and Paul Valiant. Testing monotonicity of distributions over general partial orders. In *ICS*, pages 239–252, 2011.

14   Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of reed-muller codes. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010*, pages 488–497, 2010.

15   Eric Blais. Testing juntas nearly optimally. In *Proc. 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 151–158, 2009. `doi:10.1145/1536414.1536437`.

16   Eric Blais and Yuichi Yoshida. A characterization of constant-sample testable properties. *CoRR*, abs/1612.06016, 2016. URL: `http://arxiv.org/abs/1612.06016`.

17   M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993. Earlier version in STOC'90.

18   Artur Czumaj and Christian Sohler. Property testing with geometric queries. In *Algorithms – ESA 2001, 9th Annual European Symposium*, pages 266–277, 2001.

19   Artur Czumaj, Christian Sohler, and Martin Ziegler. Property testing in computational geometry. In *Algorithms – ESA 2000, 8th Annual European Symposium*, pages 155–166, 2000.

20   O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samordinsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

21   O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

22   Oded Goldreich and Dana Ron. On sample-based testers. *TOCT*, 8(2):7:1–7:54, 2016.

23   Oded Goldreich and Madhu Sudan. Locally testable codes and pcps of almost-linear length. *J. ACM*, 53(4):558–655, 2006.

24   P.M. Gruber and J.M. Wills, editors. *Handbook of convex geometry, Volume A*. Elsevier, New York, 1993.

25   Iain M. Johnstone. Chi-square oracle inequalities. In *State of the art in probability and statistics*, pages 399–418. Institute of Mathematical Statistics, 2001.

26   Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 403–412, 2008.

27   M. Kearns and D. Ron. Testing problems with sub-learning sample complexity. *Journal of Computer and System Sciences*, 61:428–456, 2000.

28   Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric type theorems. To appear in FOCS, 2015.

29   A. Klivans, R. O'Donnell, and R. Servedio. Agnostically learning convex sets via perimeter. manuscript, 2007.

30   Pravesh Kothari, Amir Nayyeri, Ryan O'Donnell, and Chenggang Wu. Testing surface area. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1204–1214, 2014.

31   K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio. Testing halfspaces. *SIAM J. on Comput.*, 39(5):2004–2047, 2010.

32   F. Nazarov. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric aspects of functional analysis (2001-2002)*, pages 169–187. Lecture Notes in Math., Vol. 1807, Springer, 2003.

33   Joe Neeman. Testing surface area with arbitrary accuracy. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC'14, pages 393–397, 2014.

34   M. Parnas, D. Ron, and A. Samorodnitsky. Testing Basic Boolean Formulae. *SIAM J. Disc. Math.*, 16:20–46, 2002. URL: `citeseer.ifi.unizh.ch/parnas02testing.html`.

35   Luis Rademacher and Santosh Vempala. Testing geometric convexity. In *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science: 24th International Conference, Chennai, India, December 16-18, 2004. Proceedings*, pages 469–480, 2005.

36   Sofya Raskhodnikova. Approximate testing of visual properties. In *Proceedings of RANDOM*, pages 370–381, 2003.

37   R. Rubinfeld and R. Servedio. Testing monotone high-dimensional distributions. In *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 147–156, 2005.

38   Ronitt Rubinfeld and Ning Xie. Testing non-uniform k-wise independent distributions over product spaces. In *Automata, Languages and Programming, 37th International Colloquium, ICALP 2010, Bordeaux, France, July 6-10, 2010, Proceedings, Part I*, pages 565–581, 2010.

39   Stanislaw J. Szarek. Convexity, complexity, and high dimensions. In *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, pages 1599–1621. European Mathematical Society, 2006.

A   Proof of Lemma 15

Fix a typical sequence $z = (z_1, \ldots, z_q)$. Our goal is to show that the total variation distance of $\mathcal{E}_{\mathsf{yes}}(z)$ and $\mathcal{E}_{\mathsf{no}}^*(z)$ is $o(1)$. For this purpose, we define a distribution $\mathcal{F}$ over pairs $(\mathbf{b}, \mathbf{d})$ of strings in $\{0,1\}^q$ (as a coupling of $\mathcal{E}_{\mathsf{yes}}(z)$ and $\mathcal{E}_{\mathsf{no}}^*(z)$), where the marginal distribution of $\mathbf{b}$ as $(\mathbf{b}, \mathbf{d}) \leftarrow \mathcal{F}$ is the same as $\mathcal{E}_{\mathsf{yes}}(z)$ and the marginal distribution of $\mathbf{d}$ is the same as $\mathcal{E}_{\mathsf{no}}^*(z)$. Our goal follows by establishing

$$\Pr_{(\mathbf{b},\mathbf{d}) \leftarrow \mathcal{F}} \big[\mathbf{b} \neq \mathbf{d}\big] = o(1). \tag{17}$$

To define $\mathcal{F}$, we use $\mathbf{M}$ to denote the $q \times N$ $\{0,1\}$-valued random matrix derived from z and $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ (recall that $\boldsymbol{S}$ is the intersection of N random halfspaces $\boldsymbol{h}_j$, $j \in [N]$): the (i,j)th entry $\mathbf{M}_{i,j}$ of $\mathbf{M}$ is 1 if $\boldsymbol{h}_j(z_i) = 1$ (i.e., $z_i \in \boldsymbol{h}_j$) and is 0 otherwise. We use $\mathbf{M}_{i,*}$ to denote the ith row of $\mathbf{M}$, $\mathbf{M}_{*,j}$ to denote the jth column of $\mathbf{M}$, and $\mathbf{M}^{(i)}$ to denote the $i \times N$ sub-matrix of $\mathbf{M}$ that consists of the first i rows of $\mathbf{M}$. (We note that $\mathbf{M}$ is derived from $\boldsymbol{S}$ and they are defined over the same probability space. So we may consider the (conditional) distribution of $\boldsymbol{S} \leftarrow \mathcal{D}_{\mathsf{yes}}$ conditioning on an event involving $\mathbf{M}$, and we may consider the conditional distribution of $\mathbf{M}$ conditioning on an event involving $\boldsymbol{S}$.)

We now define $\mathcal{F}$. A pair $(\mathbf{b}, \mathbf{d}) \leftarrow \mathcal{F}$ is drawn using the following randomized procedure. The procedure has q rounds and generates the ith bits $\mathbf{b}_i$ and $\mathbf{d}_i$ in the ith round:

1. In the first round, we draw a random real number $\mathbf{r}_1$ from $[0, 1]$ uniformly at random. We set $\mathbf{b}_1 = 1$ if $\mathbf{r}_1 \leq \mathbf{Pr}_{\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}}[\boldsymbol{S}(z_1) = 1]$ and set $\mathbf{b}_1 = 0$ otherwise. We then set $\mathbf{d}_1 = 1$ if $\mathbf{r}_1 \leq \rho(\|z_1\|)$ and set $\mathbf{d}_1 = 0$ otherwise. (Note that for the first round, the two thresholds are indeed the same so we always have $\mathbf{b}_1 = \mathbf{d}_1$.) At the end of the first round, we also draw a vector $\mathbf{N}_{1,*}$ according to the distribution of $\mathbf{M}_{1,*}$ conditioning on $\boldsymbol{S}(z_1) = \mathbf{b}_1$.

2. In the ith round, for each i from 2 to q, we draw a random real number $\mathbf{r}_i$ from $[0, 1]$ uniformly at random. We set $\mathbf{b}_i = 1$ if we have

$$\mathbf{r}_i \leq \underset{\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}}{\mathbf{Pr}} \left[\boldsymbol{S}(z_i) = 1 \,\big|\, \mathbf{M}^{(i-1)} = \mathbf{N}^{(i-1)} \right]$$

and set $\mathbf{b}_i = 0$ otherwise. We then set $\mathbf{d}_i = 1$ if $\mathbf{r}_i \leq \rho(\|z_i\|)$ and set $\mathbf{d}_i = 0$ otherwise. At the end of the ith round, we also draw a vector $\mathbf{N}_{i,*}$ according to the distribution of $\mathbf{M}_{i,*}$ conditioning on $\mathbf{M}^{(i-1)} = \mathbf{N}^{(i-1)}$ and $\boldsymbol{S}(z_i) = \mathbf{b}_i$.

It is clear that the marginal distributions of $\mathbf{b}$ and $\mathbf{d}$, as $(\mathbf{b}, \mathbf{d}) \leftarrow \mathcal{F}$, are indeed the same as $\mathcal{E}_{\text{yes}}$ and $\mathcal{E}_{\text{no}}^*$ respectively.

To prove (17), we introduce the following notion of *nice* and *bad* matrices.

▶ **Definition 18.** We say an $i \times N$ $\{0, 1\}$-valued matrix M, for some $i \in [q]$, is *nice* if
1. M has at most $\sqrt{N}$ many 0-entries; and
2. Each column of M has at most one 0-entry.
We say M is *bad* otherwise.

We prove the following two lemmas and use them to prove (17).

▶ **Lemma 19.** $\mathbf{Pr}_{\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}} \big[\mathbf{M} \text{ is bad} \big] = o(1/q)$.

Note that when $\mathbf{M}$ is nice, we have by definition that $\mathbf{M}^{(i)}$ is also nice for every $i \in [q]$.

▶ **Lemma 20.** *For any nice $(i - 1) \times N$ $\{0, 1\}$-valued matrix $M^{(i-1)}$, we have*

$$\underset{\boldsymbol{S} \leftarrow \mathcal{D}_{\text{yes}}}{\mathbf{Pr}} \left[\boldsymbol{S}(z_i) = 1 \,\big|\, \mathbf{M}^{(i-1)} = M^{(i-1)} \right] = \rho(\|z_i\|) \pm o(1/q). \tag{18}$$

Before proving Lemma 19 and 20, we first use them to prove (17). Let $\mathbf{I}_i$ denote the indicator random variable that is 1 if $(\mathbf{b}, \mathbf{d}) \leftarrow \mathcal{E}$ has $\mathbf{b}_i \neq \mathbf{d}_i$ and is 0 otherwise, for each $i \in [q]$. Then (17) is bounded from above by $\sum_{i \in [q]} \mathbf{Pr}[\mathbf{I}_i = 1]$. To bound each $\mathbf{Pr}[\mathbf{I}_i = 1]$ we split the event into

$$\sum_{M^{(i-1)}} \mathbf{Pr} \left[\mathbf{N}^{(i-1)} = M^{(i-1)} \right] \cdot \mathbf{Pr} \left[\mathbf{I}_i = 1 \,|\, \mathbf{N}^{(i-1)} = M^{(i-1)} \right],$$

where the sum is over all $(i - 1) \times N$ $\{0, 1\}$-valued matrices $M^{(i-1)}$, and further split the sum into two sums over nice and bad matrices $M^{(i-1)}$. As $\mathbf{N}^{(i-1)}$ has the same distribution as $\mathbf{M}^{(i-1)}$, it follows from Lemma 19 (and the fact that $\mathbf{M}$ is bad when $\mathbf{M}^{(i-1)}$ is bad) that the sum over bad $M^{(i-1)}$ is at most $o(1/q)$. On the other hand, it follows from Lemma 20 that the sum over nice $M^{(i-1)}$ is $o(1/q)$. As a result, we have $\mathbf{Pr}[\mathbf{I}_i = 1] = o(1/q)$ and thus, $\sum_{i \in [q]} \mathbf{Pr}[\mathbf{I}_i = 1] = o(1)$.

We prove Lemmas 19 and 20 in the rest of the section.

Proof of Lemma 19. We show that the probability of $\mathbf{M}$ violating each of the two conditions in the definition of nice matrices is $o(1/q)$. The lemma then follows by a union bound.

For the first condition, since z is typical the probability of $\mathbf{M}_{i,j} = 0$ is

$$\text{fsa}\big(\text{cover}(z_i)\big) \leq e^{-0.49\, r^2}.$$

By linearity of expectation, the expected number of 0-entries in $\mathbf{M}$ is at most

$$qN \cdot e^{-0.49\,r^2} = o(\sqrt{N}/q),$$

using $e^{r^2/2} \geq N/n$, $N = 2^{\sqrt{n}}$ and $q = 2^{0.01\sqrt{n}}$. It follows directly from Markov's inequality that the probability of $\mathbf{M}$ having more than $\sqrt{N}$ many 0-entries is $o(1/q)$.

For the second condition, again since z is typical, the probability of $\mathbf{M}_{i,j} = \mathbf{M}_{i',j} = 1$ is

$$\mathrm{fsa}\big(\mathrm{cover}(z_i) \cap \mathrm{cover}(z_i')\big) \leq e^{-0.96\,r^2}.$$

By a union bound, the probability of $\mathbf{M}_{i,j} = \mathbf{M}_{i',j} = 1$ for some i, i', j is at most

$$q^2 N \cdot e^{-0.96r^2} = o(1/q).$$

This finishes the proof of the lemma. ◄

Finally we prove Lemma 20. Fix a nice $(i-1) \times N$ matrix M (we henceforth omit the superscript $(i-1)$ since the number of rows of M is fixed to be $i-1$). Recall that $\boldsymbol{S}(z_i) = 1$ if and only if $\boldsymbol{h}_j(z_i) = 1$ for all $j \in [N]$. As a result, we have

$$\Pr_{\boldsymbol{S} \leftarrow \mathcal{D}_{\mathrm{yes}}} \Big[\boldsymbol{S}(z_i) = 1 \,\big|\, \mathbf{M}^{(i-1)} = M \Big] = \prod_{j \in [N]} \Pr_{\boldsymbol{h}_j} \Big[\boldsymbol{h}_j(z_i) = 1 \,\big|\, \mathbf{M}_{*,j}^{(i-1)} = M_{*,j} \Big].$$

On the other hand, letting $\tau = \mathrm{fsa}(\mathrm{cover}(z_i)) = \mathrm{cap}(r/\|z_i\|)$, we have $\rho(\|z_i\|) = (1 - \tau)^N$.

In the next two claims we compare

$$\Pr_{\boldsymbol{h}_j} \Big[\boldsymbol{h}_j(z_i) = 1 \,\big|\, \mathbf{M}_{*,j}^{(i-1)} = M_{*,j} \Big]$$

with $1 - \tau$ for each $j \in [N]$ and show that they are very close. The first claim works on $j \in [N]$ with no 0-entry in $M_{*,j}$ and the second claim works on $j \in [N]$ with one 0-entry in $M_{*,j}$. (These two possibilities cover all $j \in [N]$ since the matrix M is nice.) Below we omit $\mathbf{M}_{*,j}^{(i-1)}$ in writing the conditional probabilities.

▶ **Claim 21.** *For each $j \in [N]$ with no 0-entry in the jth column $M_{*,j}$, we have*

$$\Pr_{\boldsymbol{h}_j} \Big[\boldsymbol{h}_j(z_i) = 1 \,\big|\, M_{*,j} \Big] = (1 - \tau)\left(1 \pm \frac{o(1)}{qN}\right).$$

Proof. Let δ be the probability of $\boldsymbol{h}_j(z_i) = 0$ conditioning on $M_{*,j}$ (which is all-1). Then

$$\delta = \frac{\mathrm{fsa}\left(\mathrm{cover}(z_i) - \bigcup_{j<i} \mathrm{cover}(z_j)\right)}{1 - \mathrm{fsa}\left(\bigcup_{j<i} \mathrm{cover}(z_j)\right)}.$$

Using $e^{-0.51\,r^2} \leq \mathrm{fsa}(\mathrm{cover}(z_j)) \leq e^{-0.49\,r^2}$ and $\mathrm{fsa}(\mathrm{cover}(z_i) \cap \mathrm{cover}(z_j)) \leq e^{-0.96\,r^2}$, we have

$$\delta \leq \frac{\tau}{1 - q \cdot e^{-0.49r^2}} < \tau(1 + 2q \cdot e^{-0.49\,r^2}) = \tau + 2\tau q \cdot e^{-0.49\,r^2}.$$

Using $\tau \leq e^{-0.49\,r^2}$ and $e^{r^2/2} \geq N/n$, we have

$$1 - \delta \geq 1 - \tau - 2\tau q \cdot e^{-0.49\,r^2} \geq 1 - \tau - o\big(1/(qN)\big) \geq (1 - \tau)\big(1 - o(1/(qN))\big).$$

On the other hand, we have $\delta \geq \tau - q \cdot e^{-0.96\,r^2}$ and thus,

$$1 - \delta \leq 1 - \tau + q \cdot e^{-0.96\,r^2} \leq 1 - \tau + o\big(1/(qN)\big) = (1 - \tau)\big(1 + o(1/(qN))\big).$$

This finishes the proof of the claim. ◄

▶ **Claim 22.** *For each $j \in [N]$ with one 0-entry in the jth column $M_{*,j}$, we have*

$$\Pr_{\boldsymbol{h}_j}\Big[\boldsymbol{h}_j(z_i) = 1 \,\big|\, M_{*,j}\Big] \geq 1 - O\big(e^{-0.45\,r^2}\big).$$

Proof. Let i' be the point with $M_{i',j} = 1$ and δ be the conditional probability of $\boldsymbol{h}_j(z_i) = 0$. Then we have

$$\delta \leq \frac{\mathrm{fsa}\big(\mathrm{cover}(z_i) \cap \mathrm{cover}(z_{i'})\big)}{\mathrm{fsa}\Big(\mathrm{cover}(z_i') - \bigcup_{j<i:j\neq i'} \mathrm{cover}(z_j)\Big)} \leq \frac{e^{-0.96\,r^2}}{e^{-0.51\,r^2} - q \cdot e^{-0.96\,r^2}} = O\big(e^{-0.45\,r^2}\big),$$

by our choice of q. This finishes the proof of the claim.   ◀

We combine the two claims to prove Lemma 20.

Proof of Lemma 20. Let h be the number of 0-entries in M. We have $h \leq \sqrt{N}$ since M is nice. By Claims 21, the conditional probability of $\boldsymbol{S}(z_i) = 1$ is at most

$$
\begin{aligned}
\left(\left(1 - \tau\right)\left(1 + o\left(\frac{1}{qN}\right)\right)\right)^{N-h} &= \rho(\|z_i\|) \cdot \frac{1}{(1-\tau)^h} \cdot \left(1 + o\left(\frac{1}{qN}\right)\right)^{N-h} \\
&\leq \rho(\|z_i\|) \cdot (1 + 2\tau)^h \cdot \left(1 + o\left(\frac{1}{qN}\right)\right)^N \\
&\leq \rho(\|z_i\|) \cdot \exp\big(2\tau h + o(1/q)\big) \\
&= \rho(\|z_i\|) \cdot \exp\big(o(1/q)\big) = \rho(\|z_i\|) + o(1/q).
\end{aligned}
$$

Similarly, the conditional probability of $\boldsymbol{S}(z_i) = 1$ is at least

$$
\begin{aligned}
\left(\left(1 - \tau\right)\left(1 - o\left(\frac{1}{qN}\right)\right)\right)^{N-h} &\left(1 - O\big(e^{-0.45\,r^2}\big)\right)^h \\
&\geq \rho(\|z_i\|) \cdot \left(1 - o\left(\frac{1}{qN}\right)\right)^{N-h} \left(1 - O\big(e^{-0.45\,r^2}\big)\right)^h \\
&\geq \rho(\|z_i\|) \cdot \big(1 - o(1/q)\big) \geq \rho(\|z_i\|) - o(1/q).
\end{aligned}
$$

This finishes the proof of the lemma.   ◀