

# Testing convex truncation

Anindya De, Shivam Nadimpalli, and Rocco A. Servedio

**Abstract.** We study the basic statistical problem of testing whether normally distributed  $n$ -dimensional data has been *truncated*, i.e., altered by only retaining points that lie in some unknown truncation set  $S \subseteq \mathbb{R}^n$ . As our main algorithmic results,

- (1) we give an  $O(n)$ -sample algorithm that can distinguish the standard normal distribution  $N(0, I_n)$  from  $N(0, I_n)$  conditioned on an unknown and arbitrary *convex set*  $S$ ;
- (2) we give a different  $O(n)$ -sample algorithm that can distinguish  $N(0, I_n)$  from  $N(0, I_n)$  conditioned on an unknown and arbitrary *mixture of symmetric convex sets*.

Both our algorithms are computationally efficient and run in  $O(n^2)$  time, which is linear in the size of the input. These results stand in sharp contrast with known results for learning or testing convex bodies with respect to the normal distribution or learning convex-truncated normal distributions, where state-of-the-art algorithms require essentially  $n^{O(\sqrt{n})}$  samples. An easy argument shows that no finite number of samples suffices to distinguish  $N(0, I_n)$  from an unknown and arbitrary mixture of general (not necessarily symmetric) convex sets, so no common generalization of results (1) and (2) above is possible. We also prove that any algorithm (computationally efficient or otherwise) that can distinguish  $N(0, I_n)$  from  $N(0, I_n)$  conditioned on an unknown symmetric convex set must use  $\Omega(n)$  samples. This shows that the sample complexity of each of our algorithms is optimal up to a constant factor.

## 1. Introduction

Understanding distributions which have been *truncated*, i.e., subjected to some type of conditioning, is one of the oldest and most intensively studied questions in probability and statistics. Research on truncated distributions goes back the work of Bernoulli [5], Galton [23], Pearson [33], and other pioneers; we refer the reader to the introductions of [13, 29] for historical context, and to [1, 11, 36] for contemporary book-length studies of statistical truncation.

In recent years a nascent line of work [13–15, 22] has considered various different learning and inference problems for truncated distributions from a modern theoretical computer science perspective (see Section 1.3 for a more detailed discussion

of these works and how they relate to the results of this paper). The current paper studies an arguably more basic statistical problem than learning or inference, namely *distinguishing* between a null hypothesis (that there has been no truncation) and an alternative hypothesis (that some unknown truncation has taken place).

In more detail, we consider a high-dimensional version of the fundamental problem of determining whether given input data was drawn from a known underlying probability distribution  $\mathcal{P}$ , versus from  $\mathcal{P}$  conditioned on some unknown *truncation set*  $S$  (we write  $\mathcal{P}|_S$  to denote such a truncated distribution). In our work, the known high-dimensional distribution  $\mathcal{P}$  is the  $n$ -dimensional standard normal distribution  $N(0, I_n)$ , and we consider a very broad and natural class of possible truncations, corresponding to conditioning on an unknown *convex set* (and variations of this class).

As we discuss in detail in Section 1.3, the sample complexity and running time of known algorithms for a number of related problems, such as learning convex-truncated normal distributions [29], learning convex sets under the normal distribution [26], and testing whether an unknown set is convex under the normal distribution [10], all scale exponentially in  $\sqrt{n}$ . In sharp contrast, all of our distinguishing algorithms have sample complexity *linear* in  $n$  and running time at most  $\text{poly}(n)$ . Thus, our results can be seen as an exploration of one of the most fundamental questions in testing—namely, *can we test faster than we can learn?* What makes our work different is that we allow the algorithm only to have access to random samples, which is weaker than the more powerful query access that is standardly studied in the complexity theoretic literature on property testing. However, from the vantage point of statistics and machine learning, having only sample access is arguably more natural than allowing queries. Indeed, motivated by the work of Dicker [21] in statistics, a number of recent results in computer science [9, 27, 28] have explored the distinction between *testing versus learning* from random samples, and our work is another instantiation of this broad theme. To complement our algorithmic upper bounds, we also give a number of information theoretic lower bounds on sample complexity, which in some cases nearly match our algorithmic results. We turn to a detailed discussion of our results below.

## 1.1. Our results

We give algorithms and lower bounds for a range of problems on distinguishing the normal distribution from various types of convex truncations.

**1.1.1. Efficient algorithms.** We first note that an upper bound on the Gaussian measure or *Gaussian volume* of the truncation set  $S$  is a necessary assumption, since the limiting case where the Gaussian volume of  $S$  equals 1 is the same as having no trun-

cation. In this case, it is information-theoretically impossible to detect truncation of the background distribution  $N(0, I_n)$  with any finite number of samples.

Our most basic algorithmic result is an algorithm for symmetric convex sets.<sup>1</sup>

**Theorem 1** (Symmetric convex truncations, informal statement). *There is an algorithm, known as SYMM-CONVEX-DISTINGUISHER, which uses  $O(n/\varepsilon^2)$  samples, runs in  $O(n^2/\varepsilon^2)$  time, and distinguishes between the standard  $N(0, I_n)$  distribution and any distribution  $\mathcal{D} = N(0, I_n)|_S$ , where  $S \subset \mathbb{R}^n$  is any symmetric convex set with Gaussian volume at most  $1 - \varepsilon$ .*

The algorithm SYMM-CONVEX-DISTINGUISHER is quite simple: it estimates the expected squared length of a random draw from the distribution and checks whether this value is significantly smaller than it should be for the  $N(0, I_n)$  distribution. (See Section 1.2 for a more thorough discussion of SYMM-CONVEX-DISTINGUISHER and the techniques underlying its analysis.) By extending its analysis, we are able to show that the same algorithm in fact succeeds for a broader class of truncations, namely truncation by any mixture of symmetric convex distributions.

**Theorem 2** (Mixtures of symmetric convex truncations, informal statement). *The algorithm SYMM-CONVEX-DISTINGUISHER uses  $O(n/\varepsilon^2)$  samples, runs in  $O(n^2/\varepsilon^2)$  time, and distinguishes between the standard  $N(0, I_n)$  distribution and any distribution  $\mathcal{D}$  which is a normal distribution conditioned on a mixture of symmetric convex sets such that  $d_{\text{TV}}(N(0, I_n), \mathcal{D}) \geq \varepsilon$  (where  $d_{\text{TV}}(\cdot, \cdot)$  denotes total variation distance).*

It is not difficult to see that the algorithm SYMM-CONVEX-DISTINGUISHER, which only uses the empirical mean of the squared length of samples from the distribution, cannot succeed in distinguishing  $N(0, I_n)$  from a truncation of  $N(0, I_n)$  by a general (non-symmetric) convex set. To handle truncation by general convex sets, we develop a different algorithm which uses both the estimator of SYMM-CONVEX-DISTINGUISHER and also a second estimator corresponding to the squared length of the empirical mean of its input data points. We show that this algorithm succeeds for general convex sets.

**Theorem 3** (General convex truncations, informal statement). *There is an algorithm, CONVEX-DISTINGUISHER, which uses  $O(n/\varepsilon^2)$  samples, runs in  $O(n^2/\varepsilon^2)$  time, and distinguishes between the standard  $N(0, I_n)$  distribution and any distribution  $\mathcal{D} = N(0, I_n)|_S$ , where  $S \subset \mathbb{R}^n$  is any convex set such that  $d_{\text{TV}}(N(0, I_n), N(0, I_n)|_S) \geq \varepsilon$ .*

Given Theorems 2 and 3, it is natural to wonder about a common generalization to mixtures of general convex sets. However, an easy argument (which we sketch in

---

<sup>1</sup>A set  $S \subseteq \mathbb{R}^n$  is *symmetric* if  $x \in S$  whenever  $-x \in S$ .

Appendix A) shows that no finite sample complexity is sufficient for this distinguishing problem, so no such common generalization is possible.

**1.1.2. An information-theoretic lower bound.** Next we show that the sample complexity of both of our algorithms CONVEX-DISTINGUISHER and SYMM-CONVEX-DISTINGUISHER are essentially the best possible, by giving an  $\Omega(n/\varepsilon)$ -sample lower bound for any algorithm that successfully distinguishes  $N(0, I_n)$  from  $N(0, I_n)|_K$  where  $K$  is an unknown symmetric convex set of volume  $1 - \varepsilon$ .

**Theorem 4** (Lower bound, informal statement). *For  $\varepsilon \in (0, 1/2]$ , if  $A$  is any algorithm which, given access to samples from an unknown distribution  $\mathcal{D}$ , successfully distinguishes the case that  $\mathcal{D} = N(0, I_n)$  from the case that  $\mathcal{D} = N(0, I_n)|_K$ , where  $K$  is an unknown symmetric convex set of volume  $1 - \varepsilon$ , then  $A$  must draw  $\Omega(n/\varepsilon)$  samples from  $\mathcal{D}$ .*

## 1.2. Techniques

In this section, we give a technical overview of our upper and lower bounds, starting with the former.

**Upper bounds.** To build intuition, let us first consider the case of a single symmetric convex body  $K$ . It can be shown, using symmetry and convexity of  $K$ , that draws from  $N(0, I_n)|_K$  will on average lie closer to the origin than draws from  $N(0, I_n)$ , so it is natural to use this as the basis for a distinguisher. The proof of this relies on the background distribution being  $N(0, I_n)$  in a crucial manner. We thus are led to consider our first estimator,

$$\mathbf{M} := \frac{1}{T} \sum_{i=1}^T \|\mathbf{x}^{(i)}\|^2, \quad (1.1)$$

where  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$  are independent draws from the unknown distribution (which is either  $N(0, I_n)$  or  $N(0, I_n)|_K$ ). We analyze this estimator using the notion of *convex influence* from the recent work [18]. In particular, we use a version of a Poincaré inequality for convex influence to relate the mean of  $\mathbf{M}$  to the Gaussian volume  $\text{Vol}(K)$  of the truncation set  $K$ , and combine this with the fact that the statistical distance between  $N(0, I_n)$  and  $N(0, I_n)|_K$  is precisely  $1 - \text{Vol}(K)$ . With some additional technical work in the analysis, this same tester turns out to work even for conditioning on a mixture of symmetric convex sets rather than a single symmetric convex set.

The estimator described above will not succeed for general (non-symmetric) convex sets; e.g., if  $K$  is a convex set that is “far from the origin”, then  $\mathbb{E}_{\mathbf{x} \sim N(0, I_n)|_K}[\|\mathbf{x}\|]$  can be larger than  $\mathbb{E}_{\mathbf{x} \sim N(0, I_n)}[\|\mathbf{x}\|]$ . However, if  $K$  is “far from the origin,” then the center of mass of a sample of draws from  $N(0, I_n)|_K$  should be “far from the origin,”

whereas the center of mass of a sample of draws from the standard normal distribution should be “close to the origin”; this suggests that a distinguisher based on estimating the center of mass should work for convex sets  $K$  that are far from the origin. The intuition behind our distinguisher for general convex sets is to trade off between the two cases that  $K$  is “far from the origin” versus “close to the origin”. This is made precise via a case analysis based on whether or not the set  $K$  contains a “reasonably large” origin-centered ball.<sup>2</sup>

**Lower bound.** Our lower bound is proved by considering a randomly rotated “symmetric slab” (that is, a set of the form  $\{x \in \mathbb{R}^n : |\langle x, v \rangle| \leq c_\varepsilon\}$  for some unit vector  $v \in \mathbb{S}^{n-1}$  and threshold  $c_\varepsilon \in \mathbb{R}$ ) of Gaussian volume  $1 - \varepsilon$ . Using this distribution for our lower bound construction makes it tractable to give a precise analysis of the chi-squared divergence (see Fact 6 and equation (2.2)) between a sample of  $T$  draws from the standard  $n$ -dimensional Gaussian conditioned on such a slab, versus a sample of  $T$  draws from  $N(0, I_n)$ . The analysis then reduces to getting good estimates on the Hermite spectrum (cf. Appendix B) of the one-dimensional symmetric interval functions  $\mathbf{1}\{-c_\varepsilon \leq x \leq c_\varepsilon\}$ .

### 1.3. Related work

As noted earlier in the introduction, this paper can be viewed in the context of a recent body of work [13–15, 22, 29] studying a range of statistical problems for truncated distributions from the perspective of theoretical computer science. In particular, [15] gives algorithms for non-parametric density estimation of sufficiently smooth multi-dimensional distributions in low dimension, while [22] gives algorithms for parameter estimation of truncated product distributions over discrete domains, and [14] gives algorithms for truncated linear regression.

The results in this line of research that are closest to our paper are those of [13] and [29], both of which deal with truncated normal distributions (as does our work). In [13], the authors consider the problem of inferring the parameters of an *unknown* high-dimensional normal distribution given access to samples from a *known* truncation set  $S$ , which is provided via access to an oracle for membership in  $S$ . Note that in contrast, in our work the high-dimensional normal distribution is known to be  $N(0, I_n)$  but the truncation set is unknown, and we are interested only in detecting

---

<sup>2</sup>Splitting into these two cases is reminiscent of the case split in the analysis of a weak learning algorithm for convex sets in [19], though the technical details of the analysis are quite different in our work versus [19]. In particular, [19] relies on a “density increment” result for sets with large inradius, whereas we do not use a density increment argument but instead make crucial use of an extension of the Brascamp–Lieb inequality due to Vempala [37].

whether or not truncation has occurred rather than performing any kind of estimation or learning. Like [13], the subsequent work of [29] considered the problem of estimating the parameters of an unknown high-dimensional normal distribution, but allowed for the truncation set  $S$  to also be unknown. They gave an estimation algorithm whose performance depends on the *Gaussian surface area*  $\Gamma(S)$  of the truncation set  $S$ ; when the set  $S$  is an unknown convex set in  $n$  dimensions, the sample complexity and running time of their algorithm is  $n^{O(\sqrt{n})}$ . In contrast, our algorithm for the distinguishing problem requires only  $O(n)$  samples and  $\text{poly}(n)$  running time when  $S$  is an unknown  $n$ -dimensional convex set.

Other prior works which are related to ours are [26] and [10], which dealt with Boolean function learning and property testing, respectively, of convex sets under the normal distribution. [26] gave an  $n^{O(\sqrt{n})}$ -time and sample algorithm for (agnostically) learning an unknown convex set in  $\mathbb{R}^n$  given access to labeled examples drawn from the standard normal distribution, and proved an essentially matching lower bound on sample complexity. [10] studied algorithms for testing whether an unknown set  $S \subset \mathbb{R}^n$  is convex versus far from every convex set with respect to the normal distribution, given access to random labeled samples drawn from the standard normal distribution. [10] gave an  $n^{O(\sqrt{n})}$ -sample algorithm and proved a near-matching  $2^{\Omega(\sqrt{n})}$  lower bound on sample-based testing algorithms.

We note that our techniques are very different from those of [13, 29] and [10, 26]. The paper [26] is based on analyzing the Gaussian surface area and noise sensitivity of convex sets using Hermite analysis, while [10] uses a well-known connection between testing and learning [24] to leverage the [26] learning algorithm result for its testing algorithm, and analyzes a construction due to Nazarov [31] for its lower bound. [13] uses a projected stochastic gradient descent algorithm on the negative log-likelihood function of the samples together with other tools from convex optimization, while (roughly speaking) [29] combines elements from both [26] and [13] together with moment-based methods. In contrast, our approach mainly uses ingredients from the geometry of Gaussian space, such as the Brascamp–Lieb inequality and its extensions due to Vempala [37], and the already-mentioned “convex influence” notion of [18].

Finally, we note that the basic distinguishing problem we consider is similar in spirit to a number of questions that have been studied in the field of property testing of probability distributions [7]. These are questions of the general form: “given access to samples drawn from a distribution that is promised to satisfy thus-and-such property, is it the uniform distribution or far in variation distance from uniform?” Examples of works of this flavor include the work of Batu et al. [4] on testing whether an unknown monotone or unimodal univariate distribution is uniform; the work of Daskalakis et al. [12] on testing whether an unknown  $k$ -modal distribution is uniform; the work of Rubinfeld and Servedio [35] on testing whether an unknown monotone high-dimensional distribution is uniform; and others. The problems we consider are

roughly analogous to these, but where the unknown distribution is now promised to be normal conditioned on (say) a convex set, and the testing problem is whether it is actually the normal distribution (analogous to being actually the uniform distribution, in the works mentioned above) versus far from normal.

## 2. Preliminaries

In Section 2.1, we set up basic notation and background. We recall preliminaries from convex and log-concave geometry in Sections 2.2 and 2.3, and formally describe the classes of distributions we consider in Section 2.4.

### 2.1. Basic notation and background

**Notation.** We use boldfaced letters such as  $\mathbf{x}$ ,  $\mathbf{f}$ ,  $\mathbf{A}$ , 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 “ $\mathbf{x} \sim \mathcal{D}$ ” to indicate that the random variable  $\mathbf{x}$  is distributed according to probability distribution  $\mathcal{D}$ . For  $i \in [n]$ , we will write  $e_i \in \mathbb{R}^n$  to denote the  $i$ th standard basis vector.

**Geometry.** For  $r > 0$ , we write  $\mathbb{S}^{n-1}(r)$  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.,

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

where  $\|x\|$  denotes the  $\ell_2$ -norm  $\|\cdot\|_2$  of  $x \in \mathbb{R}^n$ . We also write  $\mathbb{S}^{n-1}$  for the unit sphere  $\mathbb{S}^{n-1}(1)$ .

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  $\alpha \in [0, 1]$ . Recall that convex sets are Lebesgue measurable.

For sets  $A, B \subseteq \mathbb{R}^n$ , we write  $A + B$  to denote the Minkowski sum

$$\{a + b : a \in A, b \in B\}.$$

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  $a \in \mathbb{R}^n$  and a set  $B \subseteq \mathbb{R}^n$ , we use  $a + B$  and  $B - a$  to denote  $\{a\} + B$  and  $B + \{-a\}$  for convenience.

**Gaussians distributions.** We write  $N(0, I_n)$  to denote the  $n$ -dimensional standard Gaussian distribution, and denote its density function by  $\varphi_n$ , i.e.,

$$\varphi_n(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}.$$

When the dimension is clear from context, we may simply write  $\varphi$  instead of  $\varphi_n$ . We write  $\Phi: \mathbb{R} \rightarrow [0, 1]$  to denote the cumulative density function of the one-dimensional standard Gaussian distribution, i.e.,

$$\Phi(x) := \int_{-\infty}^x \varphi(y) dy.$$

We write  $\text{Vol}(K)$  to denote the Gaussian volume of a (Lebesgue measurable) set  $K \subseteq \mathbb{R}^n$ , that is

$$\text{Vol}(K) := \Pr_{\mathbf{x} \sim N(0, I_n)}[\mathbf{x} \in K].$$

For a Lebesgue measurable set  $K \subseteq \mathbb{R}^n$ , we write  $N(0, I_n)|_K$  to denote the standard Normal distribution conditioned on  $K$ , so the density function of  $N(0, I_n)|_K$  is

$$\frac{1}{\text{Vol}(K)} \cdot \varphi_n(x) \cdot \mathbf{1}_K(x),$$

where we identify  $K$  with its 0/1-valued indicator function. Note that the total variation distance between  $N(0, I_n)$  and  $N(0, I_n)|_K$  is

$$d_{\text{TV}}(N(0, I_n)|_K, N(0, I_n)) = 1 - \text{Vol}(K), \quad (2.1)$$

and so the total variation distance between  $N(0, I_n)$  and  $N(0, I_n)|_K$  is at least  $\varepsilon$  if and only if  $\text{Vol}(K) \leq 1 - \varepsilon$ .

**Gaussian mean testing.** We will require the following result due to Diakonikolas, Kane, and Pensia [20], which builds on prior work by Canonne et al. [8].

**Proposition 5** ([20, Theorem 1.1 and Remark 1.2]). *Let  $\mathcal{D}$  be a log-concave distribution over  $\mathbb{R}^n$  and  $\varepsilon > 0$ . There exists an algorithm, GAUSSIAN-MEAN-TESTING  $(\mathcal{D}, \varepsilon)$ , which, given i.i.d. sample access to  $\mathcal{D}$ , draws  $\Theta(\max\{1, \sqrt{n}/\varepsilon^2\})$  samples from  $\mathcal{D}$ , does an  $O(n^{3/2}/\varepsilon^2)$ -time computation, and has the following performance guarantee:*

- if  $\mathcal{D} = N(0, I_n)$ , then it outputs “accept” with probability  $99/100$ ; and
- if  $\|\mathbf{E}[\mathbf{x}]\| \geq \varepsilon$  for  $\mathbf{x} \sim \mathcal{D}$ , then it will output “reject” with probability  $99/100$ .

**Distinguishing distributions.** We recall the basic fact that variation distance provides a lower bound on the sample complexity needed to distinguish two distributions from each other.

**Fact 6** (Variation distance distinguishing lower bound). Let  $P, Q$  be two distributions over  $\mathbb{R}^n$  and let  $A$  be any algorithm which is given access to independent samples that are either from  $P$  or from  $Q$ . If  $A$  determines correctly (with probability at least  $9/10$ ) whether its samples are from  $P$  or from  $Q$ , then  $A$  must use at least  $\Omega(1/d_{\text{TV}}(P, Q))$  many samples.



In order to prove our lower bound we will instead rely on the  $\chi^2$ -divergence: Given two probability measures  $P$  and  $Q$  on  $\mathbb{R}^n$ , where  $P$  is absolutely continuous with respect to  $Q$  (i.e., for  $S \subseteq \mathbb{R}^n$ ,  $P(S) = 0$  whenever  $Q(S) = 0$ ), the  $\chi^2$ -divergence between  $P$  and  $Q$  is given by

$$d_{\chi^2}(P \parallel Q) = \mathbf{E}_{\mathbf{x} \sim Q} \left[ \left( \frac{dP(\mathbf{x})}{dQ} \right)^2 - 1 \right].$$

The following relationship between the  $\chi^2$ -divergence and the total variation distance is standard:

$$d_{\text{TV}}(P, Q)^2 \leq d_{\chi^2}(P \parallel Q). \quad (2.2)$$

## 2.2. Convex influences

In what follows, we will identify a set  $K \subseteq \mathbb{R}^n$  with its 0/1-valued indicator function. The following notion of *convex influence* was introduced in [17, 18] as an analog of the well-studied notion of *influence of a variable on a Boolean function* (cf. [32, Chapter 2]). In [17, 18], we defined this notion only for symmetric convex sets; we define it below more generally for arbitrary (Lebesgue measurable) subsets of  $\mathbb{R}^n$ .

**Definition 7** (Convex influence). Given a Lebesgue measurable set  $K \subseteq \mathbb{R}^n$  and a unit vector  $v \in \mathbb{S}^{n-1}$ , we define the *convex influence of  $v$  on  $K$* , written  $\mathbf{Inf}_v[K]$ , as

$$\mathbf{Inf}_v[K] := \mathbf{E}_{\mathbf{x} \sim N(0, I_n)} \left[ K(\mathbf{x}) \left( \frac{1 - \langle v, \mathbf{x} \rangle^2}{\sqrt{2}} \right) \right].$$

Furthermore, we define the *total convex influence of  $K$* , written  $\mathbf{I}[K]$ , as

$$\mathbf{I}[K] := \sum_{i=1}^n \mathbf{Inf}_{e_i}[K] = \mathbf{E}_{\mathbf{x} \sim N(0, I_n)} \left[ K(\mathbf{x}) \left( \frac{n - \|\mathbf{x}\|^2}{\sqrt{2}} \right) \right].$$

In [18, Proposition 20], it is shown that the influence of a direction  $v$  captures the rate of change of the Gaussian measure of the set  $K$  under a dilation along  $v$ . Also note that the total convex influence of a set is invariant under rotations. The following is immediate from Definition 7.

**Fact 8.** For Lebesgue measurable  $K \subseteq \mathbb{R}^n$ , we have

$$\mathbf{E}_{\mathbf{x} \sim N(0, I_n) | K} [\mathbf{x}_i^2] = 1 - \frac{\sqrt{2} \cdot \mathbf{Inf}_{e_i}[K]}{\text{Vol}(K)}.$$

We also have that

$$\mathbf{E}_{\mathbf{x} \sim N(0, I_n) | K} [\|\mathbf{x}\|^2] = n - \frac{\sqrt{2} \cdot \mathbf{I}[K]}{\text{Vol}(K)}. \quad (2.3)$$

The following Poincaré-type inequality for convex influences was obtained in the full version of [18, Proposition 23] (see [16]).

**Proposition 9** (Poincaré for convex influences for symmetric convex sets). *For symmetric convex  $K \subseteq \mathbb{R}^n$ , we have*

$$\frac{\mathbf{I}[K]}{\text{Vol}(K)} \geq \Omega(1 - \text{Vol}(K)).$$

The following variant of Proposition 9 for arbitrary convex sets (not necessarily symmetric) is implicit in the proof of [18, Theorem 22] (see [18, equation (16)]). Given a convex set  $K \subseteq \mathbb{R}^n$ , we denote its inradius by  $r_{\text{in}}(K)$ , i.e.,

$$r_{\text{in}}(K) := \max\{r : \text{Ball}(r) \subseteq K\}.$$

When  $K$  is clear from context, we will simply write  $r_{\text{in}}$  instead.

**Proposition 10** (Poincaré for convex influences for general convex sets). *For convex  $K \subseteq \mathbb{R}^n$  with  $r_{\text{in}} > 0$  (and hence  $\text{Vol}(K) > 0$ ), we have*

$$\frac{\mathbf{I}[K]}{\text{Vol}(K)} \geq r_{\text{in}} \cdot \Omega(1 - \text{Vol}(K)).$$

### 2.3. The Brascamp–Lieb inequality

The following result of Brascamp and Lieb [6] generalizes the Gaussian Poincaré inequality to measures which are more log-concave than the Gaussian distribution.

**Proposition 11** (Brascamp–Lieb inequality). *Suppose  $\mathcal{D}$  is a probability distribution on  $\mathbb{R}^n$  with density  $e^{-V(x)} \cdot \varphi_n(x)$  for a convex function  $V: \mathbb{R}^n \rightarrow \mathbb{R}$ . Then for any differentiable function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , we have*

$$\text{Var}_{\mathbf{x} \sim \mathcal{D}}[f(\mathbf{x})] \leq \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\|\nabla f(\mathbf{x})\|^2].$$

Vempala [37] obtained a quantitative version of Proposition 11 in one dimension, which we state next. Note in particular that the following holds for non-centered Gaussians.

**Proposition 12** ([37, Lemma 4.7]). *Fix  $\theta \in \mathbb{R}$  and let  $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  be a log-concave function such that*

$$\mathbf{E}_{\mathbf{x} \sim N(\theta, 1)}[\mathbf{x} f(\mathbf{x})] = 0.$$

*Then  $\mathbf{E}[\mathbf{x}^2 f(\mathbf{x})] \leq \mathbf{E}[f(\mathbf{x})]$  for  $\mathbf{x} \sim N(\theta, 1)$ , with equality if and only if  $f$  is a constant function. Furthermore, if  $\text{supp}(f) \subseteq (-\infty, \varepsilon]$ , then*

$$\mathbf{E}_{\mathbf{x} \sim N(\theta, 1)}[\mathbf{x}^2 f(\mathbf{x})] \leq \left(1 - \frac{1}{2\pi} e^{-\varepsilon^2}\right) \mathbf{E}_{\mathbf{x} \sim N(\theta, 1)}[f(\mathbf{x})].$$

## 2.4. The classes of distributions we consider

We say that a distribution over  $\mathbb{R}^n$  with density  $\varphi$  is *symmetric* if  $\varphi(x) = \varphi(-x)$  for all  $x$ , and that a set  $K \subseteq \mathbb{R}^n$  is symmetric if  $-x \in K$  whenever  $x \in K$ .

We let  $\mathcal{P}_{\text{symm}}$  denote the class of all distributions  $N(0, I_n)|_K$ , where  $K \subseteq \mathbb{R}^n$  may be any symmetric convex set,  $\mathcal{P}_{\text{conv}}$  denote the class of all such distributions, where  $K$  may be any convex set (not necessarily symmetric), and  $\mathcal{P}_{\text{LTF}}$  denote the class of all such distributions, where  $K$  may be any linear threshold function  $\text{sign}(v \cdot x \geq \theta)$ . We let  $\text{Mix}(\mathcal{P}_{\text{symm}})$  denote the class of all convex combinations (mixtures) of distributions from  $\mathcal{P}_{\text{symm}}$ , and we remark that a distribution in  $\text{Mix}(\mathcal{P}_{\text{symm}})$  can be viewed as  $N(0, I_n)$  conditioned on a *mixture* of symmetric convex sets.

The following alternative characterization of  $\text{Mix}(\mathcal{P}_{\text{symm}})$  may be of some interest. Let  $\mathcal{P}_{\text{slcg}}$  denote the class of all symmetric distributions that are log-concave relative to the standard normal distribution, i.e., all distributions that have a density of the form  $e^{-\tau(x)}\varphi_n(x)$ , where  $\tau(\cdot)$  is a symmetric convex function. Let  $\text{Mix}(\mathcal{P}_{\text{slcg}})$  denote the class of all mixtures of distributions in  $\mathcal{P}_{\text{slcg}}$ .

**Claim 13.**  $\text{Mix}(\mathcal{P}_{\text{slcg}}) = \text{Mix}(\mathcal{P}_{\text{symm}})$ .

*Proof.* We will argue below that  $\mathcal{P}_{\text{slcg}} \subseteq \text{Mix}(\mathcal{P}_{\text{symm}})$ . Given this, it follows that any mixture of distributions in  $\mathcal{P}_{\text{slcg}}$  is a mixture of distributions in  $\text{Mix}(\mathcal{P}_{\text{symm}})$ , but since a mixture of distributions in  $\text{Mix}(\mathcal{P}_{\text{symm}})$  is itself a distribution in  $\text{Mix}(\mathcal{P}_{\text{symm}})$ , this means that  $\text{Mix}(\mathcal{P}_{\text{slcg}}) \subseteq \text{Mix}(\mathcal{P}_{\text{symm}})$ . For the other direction, we observe that any distribution in  $\mathcal{P}_{\text{symm}}$  belongs to  $\mathcal{P}_{\text{slcg}}$ ,<sup>3</sup> and hence  $\text{Mix}(\mathcal{P}_{\text{symm}}) \subseteq \text{Mix}(\mathcal{P}_{\text{slcg}})$ .

Fix any distribution  $\mathcal{D}$  in  $\mathcal{P}_{\text{slcg}}$  and let  $e^{-\tau(x)}\varphi_n(x)$  be its density. We have that

$$e^{-\tau(x)}\varphi_n(x) = \mathbf{E}[A_t(x)] \cdot \varphi_n(x), \quad (2.4)$$

where  $A_t(x) = \mathbf{1}[e^{-\tau(x)} \geq t]$  and the expectation in (2.4) is over a uniform  $t \sim [0, 1]$ . Since  $\tau$  is a symmetric convex function, we have that the level set  $\{x \in \mathbb{R}^n : e^{-\tau(x)} \geq t\}$  is a symmetric convex set, so  $\mathcal{D}$  is a mixture of distributions in  $\mathcal{P}_{\text{symm}}$ , as claimed above. ■

## 3. An $O(n/\varepsilon^2)$ -sample algorithm for symmetric convex sets and mixtures of symmetric convex sets

In this section, we give an algorithm (cf. Algorithm 1) to distinguish Gaussians from (mixtures of) Gaussians truncated to a symmetric convex set.

---

<sup>3</sup>Recall that a distribution in  $\mathcal{P}_{\text{symm}}$  has a density which is  $\text{Vol}(K)^{-1} \cdot K(x) \cdot \varphi_n(x)$  for some symmetric convex  $K$ .

### 3.1. Useful structural results

We record a few important lemmas, which are going to be useful for the analysis in this section.

**Lemma 14.** *Let  $K \subseteq \mathbb{R}^n$  be a centrally symmetric convex set. If  $\text{Vol}(K) \leq 1 - \varepsilon$ , then,*

$$\mathbf{E}_{\mathbf{x} \sim N(0, I_n)|_K} [\|\mathbf{x}\|^2] \leq n - c\varepsilon$$

for some absolute constant  $c > 0$ .

*Proof.* We have

$$\mathbf{E}_{\mathbf{x} \sim N(0, I_n)|_K} [\|\mathbf{x}\|^2] = n - \frac{\sqrt{2} \cdot \mathbf{I}[K]}{\text{Vol}(K)} \leq n - \sqrt{2} \cdot c'(1 - \text{Vol}(K)) \leq n - \sqrt{2} \cdot c'\varepsilon,$$

where the equality is equation (2.3), the first inequality is Proposition 9 (Poincaré for convex influences for symmetric convex sets), and the second inequality holds because  $\text{Vol}(K) \leq 1 - \varepsilon$ . ■

**Lemma 15.** *Let  $K \subseteq \mathbb{R}^n$  be a convex set (not necessarily symmetric) and let  $\mathcal{D} = N(0, I_n)|_K$ . Then for any unit vector  $v$ , we have*

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[v \cdot \mathbf{x}] \leq 1.$$

*Proof.* Given  $c > 0$ , we define  $V_c: \mathbb{R}^n \rightarrow \{c, +\infty\}$  to be

$$V_c(x) = \begin{cases} c & \text{if } x \in K, \\ +\infty & \text{if } x \notin K. \end{cases}$$

We note that  $V_c(\cdot)$  is a convex function for any choice of  $c > 0$ , and that for a suitable choice of  $c$ , the density function of  $\mathcal{D}$  is  $e^{-V_c(x)} \cdot \gamma_n(x)$ . Thus, we can apply the Brascamp–Lieb inequality to get, for any differentiable  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , that

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[f(\mathbf{x})] \leq \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\|\nabla f(\mathbf{x})\|^2]. \quad (3.1)$$

Now, we may assume without loss of generality that  $v = e_1$ . Taking  $f(x) = x_1$  in equation (3.1), we get that

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[x_1] \leq 1,$$

which finishes the proof. ■

Now we can bound the variance of  $\|\mathbf{x}\|^2$  when  $\mathbf{x} \sim N(0, I_n)|_K$  for a symmetric convex set  $K$ .

**Lemma 16.** *Let  $\mathcal{D} = N(0, I_n)|_K$  for a symmetric convex set  $K$ . Then, we have that  $\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|^2] \leq 4n$ .*

*Proof.* Taking  $f(x) := \|x\|^2$  in equation (3.1), we have that

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|^2] \leq 4 \cdot \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x}_1^2 + \cdots + \mathbf{x}_n^2].$$

Since  $K$  is symmetric, for each  $i \in [n]$  we have  $\mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x}_i] = 0$ , and hence

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x}_i^2] = \mathbf{Var}[e_i \cdot \mathbf{x}],$$

which is at most 1 by Lemma 15. ■

### 3.2. An $O(n/\varepsilon^2)$ -sample algorithm for symmetric convex sets

We recall Theorem 1.

**Theorem 17** (Restatement of Theorem 1). *For a sufficiently large constant  $C > 0$ , the algorithm SYMM-CONVEX-DISTINGUISHER (Algorithm 1) has the following performance guarantee: given any  $\varepsilon > 0$  and access to independent samples from any unknown distribution  $\mathcal{D} \in \mathcal{P}_{\text{symm}}$ , the algorithm uses  $Cn/\varepsilon^2$  samples, and*

- (1) *If  $\mathcal{D} = N(0, I_n)$ , then with probability at least 9/10 the algorithm outputs “un-truncated”;*
- (2) *If  $\mathbf{d}_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ , then with probability at least 9/10 the algorithm outputs “truncated.”*

As alluded to in Section 1.2, SYMM-CONVEX-DISTINGUISHER uses the estimator from equation (1.1).

<p><b>Algorithm 1:</b> Distinguisher for (mixtures of) symmetric convex sets</p> <p><b>Input:</b> <math>\mathcal{D} \in \mathcal{P}_{\text{conv}}</math>, <math>\varepsilon &gt; 0</math></p> <p><b>Output:</b> “Un-truncated” or “truncated”</p> <p>SYMM-CONVEX-DISTINGUISHER (<math>\mathcal{D}, \varepsilon</math>):</p> <ol style="list-style-type: none"> <li>(1) For <math>T = C \cdot n/\varepsilon^2</math>, sample points <math>\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)} \sim \mathcal{D}</math>.</li> <li>(2) Let <math>\mathbf{M} := \frac{1}{T} \sum_{i=1}^T \ \mathbf{x}^{(i)}\ ^2</math>.</li> <li>(3) If <math>\mathbf{M} \geq n - c\varepsilon/2</math>, output “un-truncated”, else output “truncated”.</li> </ol>
--

*Proof of Theorem 17.* Let  $\mathcal{D}_G := N(0, I_n)$  and  $\mathcal{D}_T := N(0, I_n)|_K$ . Then, for  $\mathbf{x} \sim \mathcal{D}_G$ , the random variable  $\|\mathbf{x}\|^2$  follows the  $\chi^2$  distribution with  $n$  degrees of freedom, and thus we have

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}_G}[\|\mathbf{x}\|^2] = n, \quad \mathbf{Var}_{\mathbf{x} \sim \mathcal{D}_G}[\|\mathbf{x}\|^2] = 2n.$$

On the other hand, if  $d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$  (equivalently,  $\text{Vol}(K) \leq 1 - \varepsilon$ ), then using Lemma 14 and Lemma 16, it follows that

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}_T}[\|\mathbf{x}\|^2] \leq n - c\varepsilon, \quad \mathbf{Var}_{\mathbf{x} \sim \mathcal{D}_T}[\|\mathbf{x}\|^2] \leq 4n.$$

Since in Algorithm 1 the samples  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$  are independent, we have the following:

$$\begin{aligned} \mathbf{E}[\mathbf{M}] &= n, & \mathbf{Var}[\mathbf{M}] &= \frac{3n}{T} & \text{when } \mathcal{D} = \mathcal{D}_G, \\ \mathbf{E}[\mathbf{M}] &= n - c\varepsilon, & \mathbf{Var}[\mathbf{M}] &\leq \frac{4n}{T} & \text{when } \mathcal{D} = \mathcal{D}_T. \end{aligned}$$

By choosing  $T = Cn/\varepsilon^2$  (for a sufficiently large constant  $C$ ), it follows that when  $\mathcal{D} = \mathcal{D}_G$  (resp.  $\mathcal{D} = \mathcal{D}_T$ ), with probability at least 9/10 we have  $\mathbf{M} \geq n - c\varepsilon/2$  (resp.  $\mathbf{M} < n - c\varepsilon/2$ ). This finishes the proof. ■

### 3.3. An $O(n/\varepsilon^2)$ -sample algorithm for mixtures of symmetric convex sets

By extending the above analysis, we can show that Algorithm 1 succeeds for mixtures of (an arbitrary number of) symmetric convex sets as well. In particular, we have the following result.

**Theorem 18.** *For a sufficiently large constant  $C > 0$ , SYMM-CONVEX-DISTINGUISHER (Algorithm 1) has the following performance guarantee: given any  $\varepsilon > 0$  and access to independent samples from any unknown distribution  $\mathcal{D} \in \text{Mix}(\mathcal{P}_{\text{symm}})$ , the algorithm uses  $Cn/\varepsilon^2$  samples, and*

- (1) *if  $\mathcal{D} = N(0, I_n)$ , then with probability at least 9/10 the algorithm outputs “un-truncated”;*
- (2) *if  $d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ , then with probability at least 9/10 the algorithm outputs “truncated”.*

The following lemma, which characterizes the mean and variance of a distribution in  $\text{Mix}(\mathcal{P}_{\text{symm}})$  in terms of the components of the mixture, will be crucial to the proof of Theorem 18.

**Lemma 19.** *Let  $\mathcal{X}$  denote a distribution over Gaussians truncated by symmetric convex sets. Suppose  $\mathcal{D}_{\mathcal{X}} \in \text{Mix}(\mathcal{P}_{\text{symm}})$  is the mixture of  $N(0, I_n)|_{\mathbf{K}}$  for  $\mathbf{K} \sim \mathcal{X}$ . Let  $\mathbf{a}_{\mathbf{K}}$  denote the random variable*

$$\mathbf{a}_{\mathbf{K}} = \mathbf{E}_{\mathbf{x} \sim N(0, I_n)|_{\mathbf{K}}}[\|\mathbf{x}\|^2], \quad \text{where } \mathbf{K} \sim \mathcal{X}.$$

Then

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}_{\mathcal{X}}}[\|\mathbf{x}\|^2] = \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[\mathbf{a}_{\mathbf{K}}], \quad (3.2)$$

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}_{\mathcal{X}}}[\|\mathbf{x}\|^2] \leq 4n + \mathbf{Var}_{\mathbf{K} \sim \mathcal{X}}[\mathbf{a}_{\mathbf{K}}]. \quad (3.3)$$

*Proof.* Note that equation (3.2) follows from linearity of expectation and the definition of  $\mathbf{a}_{\mathbf{K}}$ . For equation (3.3), note that for any symmetric convex set  $K$ , by definition of variance we have

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim N(0, I_n)|_K}[\|\mathbf{x}\|^4] &= (\mathbf{E}_{\mathbf{x} \sim N(0, I_n)|_K}[\|\mathbf{x}\|^2])^2 + \mathbf{Var}_{\mathbf{x} \sim N(0, I_n)|_K}[\|\mathbf{x}\|^2] \\ &\leq \mathbf{a}_K^2 + 4n, \end{aligned}$$

where the inequality is by Lemma 16. By linearity of expectation, it now follows that

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}_{\mathcal{X}}}[\|\mathbf{x}\|^4] \leq 4n + \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[\mathbf{a}_{\mathbf{K}}^2].$$

Combining with equation (3.2), we get equation (3.3).  $\blacksquare$

We are now ready to prove Theorem 18.

*Proof of Theorem 18.* Let  $\mathcal{X}$  denote a distribution over symmetric convex sets. Define  $\mathcal{D}_{\mathcal{X}} \in \text{Mix}(\mathcal{P}_{\text{sym}})$  to be the mixture of  $N(0, I_n)|_{\mathbf{K}}$  for  $\mathbf{K} \sim \mathcal{X}$  and define  $\mathcal{D}_G := N(0, I_n)$ . Using the fact that the samples  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$  are independent, as in the proof of Theorem 17, we have that

$$\mathbf{E}[\mathbf{M}] = n, \quad \mathbf{Var}[\mathbf{M}] = \frac{3n}{T} \quad \text{when } \mathcal{D} = \mathcal{D}_G.$$

As  $T = Cn/\varepsilon^2$  (for a sufficiently large constant  $C$ ), it follows that when  $\mathcal{D} = \mathcal{D}_G$ , with probability at least 9/10 we have that  $\mathbf{M} \geq n - \varepsilon/2$ .

Now we analyze the case that  $\mathcal{D} = \mathcal{D}_{\mathcal{X}}$  has  $d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ . From Lemma 19, it follows that in this case

$$\mathbf{E}[\mathbf{M}] = \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[\mathbf{a}_{\mathbf{K}}], \quad (3.4)$$

$$\mathbf{Var}[\mathbf{M}] = \frac{\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}_{\mathcal{X}}}[\|\mathbf{x}\|^2]}{T} \leq \frac{4n}{T} + \frac{\mathbf{Var}_{\mathbf{K} \sim \mathcal{X}}[\mathbf{a}_{\mathbf{K}}]}{T}. \quad (3.5)$$

Next, observe that

$$\mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[(n - \mathbf{a}_{\mathbf{K}})] \geq c \cdot \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[1 - \text{Vol}(\mathbf{K})] \geq c \cdot d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq c\varepsilon, \quad (3.6)$$

where the first inequality uses Lemma 14 and the second inequality follows from the definition of TV distance.

Now, observing that variance of a random variable is invariant under negation and translation and that  $T = Cn/\varepsilon^2$ , it follows from equation (3.5) that

$$\begin{aligned}\text{Var}[\mathbf{M}] &\leq \frac{4n}{T} + \frac{\text{Var}_{\mathbf{K} \sim \mathcal{X}}[a_{\mathbf{K}}]}{T} \\ &\leq \frac{4\varepsilon^2}{C} + \frac{\varepsilon^2 \cdot \text{Var}_{\mathbf{K} \sim \mathcal{X}}[n - a_{\mathbf{K}}]}{Cn} \\ &\leq \frac{4\varepsilon^2}{C} + \frac{\varepsilon^2 \cdot \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[(n - a_{\mathbf{K}})^2]}{Cn}.\end{aligned}$$

By equation (2.3) and Proposition 9, we have that  $0 \leq a_K \leq n$  for any symmetric convex  $K$ . Thus, we can further upper bound the right-hand side to obtain

$$\text{Var}[\mathbf{M}] \leq \frac{4\varepsilon^2}{C} + \frac{\varepsilon^2 \cdot \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[n - a_{\mathbf{K}}]}{C}.$$

Recalling from equation (3.6) that  $\mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[n - a_{\mathbf{K}}] \geq c\varepsilon$ , a routine computation shows that for a sufficiently large constant  $C$ , we have

$$\text{Var}[\mathbf{M}] \leq \frac{4\varepsilon^2}{C} + \frac{\varepsilon^2 \cdot \mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[n - a_{\mathbf{K}}]}{C} \leq \frac{\mathbf{E}_{\mathbf{K} \sim \mathcal{X}}[n - c\varepsilon/2 - a_{\mathbf{K}}]^2}{100}.$$

Equation (3.4) and Chebyshev's inequality now give that when  $\mathcal{D} = \mathcal{D}_{\mathcal{X}}$ , with probability at least 9/10 we have  $\mathbf{M} \leq n - c\varepsilon/2$ , completing the proof. ■

#### 4. An $O(n/\varepsilon^2)$ -sample algorithm for general convex sets

In this section we present a  $O(n/\varepsilon^2)$ -sample algorithm for distinguishing the standard normal distribution from the standard normal distribution restricted to an arbitrary convex set. More precisely, we prove the following theorem.

**Theorem 20.** *There is an algorithm, CONVEX-DISTINGUISHER (Algorithm 2), with the following performance guarantee: Given any  $\varepsilon > 0$  and access to independent samples from any unknown distribution  $\mathcal{D} \in \mathcal{P}_{\text{conv}}$ , the algorithm uses  $O(n/\varepsilon^2)$  samples, runs in  $O(n^2/\varepsilon^2)$  time, and*

- (1) *if  $\mathcal{D} = N(0, I_n)$ , then with probability at least 9/10 the algorithm outputs “un-truncated”;*
- (2) *if  $d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ , then with probability at least 9/10 the algorithm outputs “truncated”.*

Note that the estimator  $\mathbf{M}$  in Algorithm 2 is identical to the estimator  $\mathbf{M}$  in Algorithm 1 to distinguish Gaussians restricted to (mixtures of) symmetric convex sets.



**Algorithm 2:** Distinguisher for general convex sets**Input:**  $\mathcal{D} \in \mathcal{P}_{\text{conv}}, \varepsilon > 0$ **Output:** “un-truncated” or “truncated”CONVEX-DISTINGUISHER( $\mathcal{D}, \varepsilon$ ):

- (1) If GAUSSIAN-MEAN-TESTING ( $\mathcal{D}, 0.01$ ) returns “reject,” then halt and output “truncated.”
- (2) For  $T = C \cdot n/\varepsilon^2$ , sample points  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)} \sim \mathcal{D}$ .
- (3) Set  $\mathbf{M} := \frac{1}{T} \sum_{j=1}^T \|\mathbf{x}^{(j)}\|^2$ . Output “truncated” if  $\mathbf{M} \leq n - c\varepsilon/2$  and “un-truncated” otherwise.

As we will see, the analysis of Algorithm 1 via the Poincaré inequality for convex influences (cf. Proposition 9) extends to arbitrary convex sets with “large inradius”. For the “small inradius” case, we further consider sub-cases depending on how close the center of mass of  $\mathcal{D}$ , denoted  $\mu$ , is to the origin (see Figure 1):

**Case 1.** When  $\|\mu\| \gg 0$ , we detect truncation via GAUSSIAN-MEAN-TESTING in Step 2 of Algorithm 2.

**Case 2.** When  $\|\mu\| \approx 0$ , we show that we can detect truncation via  $\mathbf{M}$ . This is our most technically-involved case and relies crucially on (small extensions of) Vempala’s quantitative Brascamp–Lieb inequality (Proposition 12).

#### 4.1. Useful preliminaries

We note below two useful consequences of Vempala’s quantitative one-dimensional Brascamp–Lieb inequality (Proposition 12) which will be useful in our analysis of Algorithm 2.

The following proposition says that if the center of mass of a convex body (with respect to the standard normal distribution) along a direction  $v \in \mathbb{S}^{n-1}$  is the origin, then the convex influence of  $v$  on the body is non-negative.

**Proposition 21.** *Given a convex set  $K \subseteq \mathbb{R}^n$  and  $v \in \mathbb{S}^{n-1}$ , if*

$$\mathbf{E}_{\mathbf{x} \sim N(0, I_n)}[K(\mathbf{x})\langle v, \mathbf{x} \rangle] = 0,$$

*then  $\mathbf{Inf}_v[K] \geq 0$ .*

*Proof.* We may assume without loss of generality that  $v = e_1$ . Note that the function  $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  defined by

$$f(x) := \mathbf{E}_{\mathbf{y} \sim N(0, I_{n-1})}[K(x, \mathbf{y})],$$

is a log-concave function (this is immediate from the Prékopa–Leindler inequality; see [30, 34]). Furthermore, note that by Fact 8,

$$\sqrt{2} \cdot \mathbf{Inf}_v[K] = \mathbf{E}_{\mathbf{x} \sim N(0,1)}[f(\mathbf{x})(1 - \mathbf{x}^2)],$$

and so the result follows by Proposition 12.  $\blacksquare$

We also require a version of Proposition 12 for log-concave functions whose center of mass with respect to the standard normal distribution is not at the origin. Looking ahead, Proposition 22 will come in handy when analyzing Algorithm 2 for Gaussians restricted to convex sets with small inradius and with center of mass close to the origin.

**Proposition 22.** *Let  $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  be a one-dimensional log-concave function with*

$$\mathbf{E}_{\mathbf{x} \sim N(0,1)}[\mathbf{x} f(\mathbf{x})] = \mathbf{E}_{\mathbf{x} \sim N(0,1)}[\mu \cdot f(\mathbf{x})]$$

*for some  $\mu \in \mathbb{R}$ . Then*

$$\mathbf{E}_{\mathbf{x} \sim N(0,1)}[\mathbf{x}^2 f(\mathbf{x})] \leq (1 + \mu^2) \cdot \mathbf{E}_{\mathbf{x} \sim N(0,1)}[f(\mathbf{x})].$$

*Furthermore, if  $\text{supp}(f) \subseteq (-\infty, \varepsilon]$ , then*

$$\mathbf{E}_{\mathbf{x} \sim N(0,1)}[\mathbf{x}^2 f(\mathbf{x})] \leq \left(1 + \mu^2 - \frac{1}{2\pi} e^{-(\varepsilon - \mu)^2}\right) \cdot \mathbf{E}_{\mathbf{x} \sim N(0,1)}[f(\mathbf{x})]. \quad (4.1)$$

We prove Proposition 22 by translating the log-concave function  $f$  so that its center of mass (with respect to a shifted Gaussian) is the origin, and then appealing to Proposition 12.

*Proof.* Note that it suffices to prove equation (4.1). Consider the one-dimensional log-concave function  $\tilde{f}: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  given by

$$\tilde{f}(x) := f(x + \mu).$$

It is clear that  $\text{supp}(\tilde{f}) \subseteq (-\infty, \varepsilon - \mu]$  if  $\text{supp}(f) \subseteq (-\infty, \varepsilon]$ . Note that

$$\mathbf{E}_{\mathbf{x} \sim N(-\mu,1)}[\tilde{f}(\mathbf{x})] = \int_{\mathbb{R}} f(x + \mu) \varphi(x + \mu) dx = \mathbf{E}_{\mathbf{x} \sim N(0,1)}[f(\mathbf{x})]. \quad (4.2)$$

We also have that

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim N(-\mu,1)}[\mathbf{x} \tilde{f}(\mathbf{x})] &= \int_{\mathbb{R}} x f(x + \mu) \varphi(x + \mu) dx \\ &= \int_{\mathbb{R}} (y - \mu) f(y) \varphi(y) dy \\ &= \mathbf{E}_{\mathbf{y} \sim N(0,1)}[y f(y)] - \mathbf{E}_{\mathbf{y} \sim N(0,1)}[\mu \cdot f(y)] \\ &= 0, \end{aligned}$$

where we made the substitution  $y = x - \mu$ . Therefore, by Proposition 12, we have that

$$\mathbf{E}_{\mathbf{x} \sim N(-\mu, 1)}[\mathbf{x}^2 \tilde{f}(\mathbf{x})] \leq \left(1 - \frac{1}{2\pi} e^{-(\varepsilon - \mu)^2}\right) \cdot \mathbf{E}_{\mathbf{x} \sim N(-\mu, 1)}[\tilde{f}(\mathbf{x})]. \quad (4.3)$$

However, we have

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim N(-\mu, 1)}[\mathbf{x}^2 \tilde{f}(\mathbf{x})] &= \int_{\mathbb{R}} x^2 f(x + \mu) \varphi(x + \mu) dx \\ &= \int_{\mathbb{R}} (y - \mu)^2 f(y) \varphi(y) dy \\ &= \mathbf{E}_{\mathbf{y} \sim N(0, 1)}[\mathbf{y}^2 f(\mathbf{y})] - \mathbf{E}_{\mathbf{y} \sim N(0, 1)}[\mu^2 \cdot f(\mathbf{y})]. \end{aligned} \quad (4.4)$$

Equation (4.1) now follows from equations (4.2), (4.3) and (4.4).  $\blacksquare$

## 4.2. Proof of Theorem 20

We can now turn to the proof of Theorem 20.

*Proof of Theorem 20.* Suppose first that  $\mathcal{D} = N(0, I_n)$ . In this case,

$$\mathbf{E}[\mathbf{M}] = \frac{1}{T} \sum_{j=1}^T \mathbf{E}[\|\mathbf{x}^{(j)}\|^2] = \frac{1}{T} \sum_{j=1}^T n = n. \quad (4.5)$$

We also have that

$$\begin{aligned} \mathbf{Var}[\mathbf{M}] &= \frac{1}{T^2} \sum_{j=1}^T \mathbf{Var}[\|\mathbf{x}^{(j)}\|^2] = \frac{1}{T} (\mathbf{Var}_{\mathbf{x} \sim N(0, I_n)}[\|\mathbf{x}\|^2]) \\ &= \frac{1}{T} \sum_{i=1}^n \mathbf{Var}_{\mathbf{x}_i \sim N(0, 1)}[\mathbf{x}_i^2] = \frac{2n}{T}, \end{aligned} \quad (4.6)$$

where we used the fact that  $\mathbf{Var}_{\mathbf{x} \sim N(0, 1)}[\mathbf{x}^2] = 2$ . Looking ahead, we also note that in this case, by Proposition 5 we have the algorithm outputs “truncated” in Step 2 with probability at most 0.01.

Next, suppose that  $\mathcal{D} = N(0, I_n)|_K$  for convex  $K \subseteq \mathbb{R}^n$  with  $d_{TV}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ . Let us write  $r_{\text{in}}$  for the in-radius of  $K$ . Suppose first that  $r_{\text{in}} \geq 0.1$ . In this case, we have that

$$\mathbf{E}[\mathbf{M}] = \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|^2] \leq n - \Omega(\varepsilon) \quad (4.7)$$

by equation (2.1), Fact 8, and Proposition 10. By independence of the  $\mathbf{x}^{(j)}$ ’s, we also have that

$$\mathbf{Var}[\mathbf{M}] = \frac{1}{T^2} \sum_{j=1}^T \mathbf{Var}_{\mathbf{x}^{(j)} \sim \mathcal{D}}[\|\mathbf{x}^{(j)}\|^2].$$

Note, however, that by Proposition 11 we have

$$\mathbf{Var}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|^2] \leq 4 \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\|\mathbf{x}\|^2], \quad \text{and so } \mathbf{Var}[\mathbf{M}] \leq \frac{4n}{T}, \quad (4.8)$$

where the second inequality follows from (4.7). From equations (4.5) and (4.7), we have that the means of  $\mathbf{M}$  under  $N(0, I_n)$  versus  $N(0, I_n)|_K$  differ by  $\Omega(\varepsilon)$ , and from equations (4.6) and (4.8) we have that the standard deviations in both settings are on the order of  $O(\sqrt{n/T})$ . This shows that CONVEX-DISTINGUISHER indeed succeeds in distinguishing  $\mathcal{D} = N(0, I_n)$  from  $\mathcal{D} = N(0, I_n)|_K$  with  $O(n/\varepsilon^2)$  samples in the case that  $r_{\text{in}} \geq 0.1$ .

For the rest of the proof we can therefore assume that  $r_{\text{in}} < 0.1$ . It follows from the hyperplane separation theorem that there exists  $x^* \in \mathbb{S}^{n-1}(0.1)$  such that  $K$  lies entirely on one side of the hyperplane that is tangent to  $\mathbb{S}^{n-1}(0.1)$  at  $x^*$ . Recalling that the standard normal distribution is invariant under rotation, we can suppose without generality that  $x^*$  is the point  $(0.1, 0^{n-1})$ , so we have that either

$$K \subseteq \{x \in \mathbb{R}^n : x_1 < 0.1\} \quad \text{or} \quad K \subseteq \{x \in \mathbb{R}^n : x_1 \geq 0.1\},$$

corresponding to (a) and (b) respectively in Figure 1. Writing  $\mu$  for the center of mass of  $\mathcal{D}$ , i.e.,

$$\mu := \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x}],$$

we can apply another rotation to obtain  $\mu = (\mu_1, \mu_2, 0^{n-2})$  while maintaining that  $x^* = (0.1, 0^{n-1})$ . Now we consider two cases based on the norm of  $\mu$ :

**Case 1.** If  $\|\mu\|^2 \geq 0.06$ , then we claim that Step 1 of Algorithm 2 will correctly output “truncated” with probability at least 99/100. Indeed, it is readily verified that  $\mathcal{D}$  is log-concave, and so this follows immediately from Proposition 5.

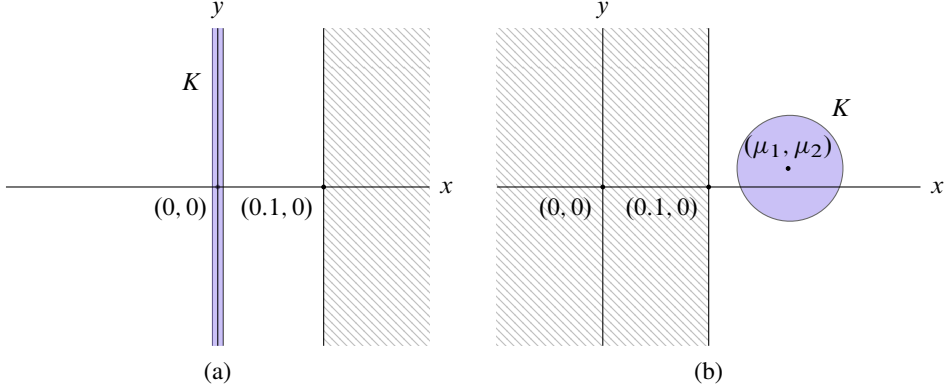
**Case 2.** If  $\|\mu\|^2 < 0.06$ , then we will show that Algorithm 2 will output “truncated” with probability at least 9/10 in Step 3. We will do this by proceeding analogously to the “large inradius” ( $r_{\text{in}} \geq 0.1$ ) setting considered earlier. Recall that

$$\mathbf{E}[\mathbf{M}] = \sum_{i=1}^n \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[x_i^2]. \quad (4.9)$$

For  $i \in \{3, \dots, n\}$ , as  $\mu_i = 0$ , we have by Proposition 21 that  $\mathbf{Inf}_i[K] \geq 0$ , and so

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[x_i^2] \leq 1 \quad \text{for } i \in \{3, \dots, n\} \quad (4.10)$$

by Fact 8.



**Figure 1.** The “small inradius” ( $r_{\text{in}} \leq 0.1$ ) setting in the analysis of Algorithm 2, with  $\mu$  denoting the center of mass of  $K$ . Our estimator for (a) is  $\text{Avg}(\|\mathbf{x}^{(j)}\|^2)$ , whereas for (b) we simply estimate  $\mu$ .

We now consider coordinates 1 and 2. Consider the one-dimensional log-concave functions  $f_1, f_2: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  defined by

$$\begin{aligned} f_1(x) &:= \mathbf{E}_{\mathbf{y} \sim N(0, I_{n-1})}[K(x, \mathbf{y})], \\ f_2(x) &:= \mathbf{E}_{\mathbf{y} \sim N(0, I_{n-1})}[K(\mathbf{y}_1, x, \mathbf{y}_2, \dots, \mathbf{y}_{n-1})]. \end{aligned}$$

Note that  $\mathbf{E}[f_1] = \mathbf{E}[f_2] = \text{Vol}(K)$ . It is also immediate that

$$\mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[x_i^2] = \frac{\mathbf{E}_{\mathbf{x} \sim N(0,1)}[x^2 f_i(\mathbf{x})]}{\text{Vol}(K)}. \quad (4.11)$$

Since we have

$$\mathbf{E}_{\mathbf{x} \sim N(0,1)}[x f_1(\mathbf{x})] = \mu_1 \cdot \text{Vol}(K), \quad \mathbf{E}_{\mathbf{x} \sim N(0,1)}[x f_2(\mathbf{x})] = \mu_2 \cdot \text{Vol}(K),$$

it follows from Proposition 22 that

$$\begin{aligned} \frac{\mathbf{E}_{\mathbf{x} \sim N(0,1)}[x^2 f_1(\mathbf{x})]}{\text{Vol}(K)} &\leq 1 + \mu_1^2 - \frac{e^{-(0.1-\mu_1)^2}}{2\pi}, \\ \frac{\mathbf{E}_{\mathbf{x} \sim N(0,1)}[x^2 f_2(\mathbf{x})]}{\text{Vol}(K)} &\leq 1 + \mu_2^2 \end{aligned} \quad (4.12)$$

(note that we used the fact that  $\text{supp}(f_1) \subseteq (-\infty, 0.1]$  in the first inequality above). Combining equations (4.11) and (4.12) and recalling that  $\|\mu\|^2 < 0.06$ , we get that

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim \mathcal{D}}[x_1^2 + x_2^2] &\leq 2 + \|\mu\|^2 - \frac{1}{2\pi} e^{-(0.1-\mu_1)^2} \\ &< 2.06 - \frac{1}{2\pi} e^{-(0.1+\sqrt{0.06})^2} < 1.95. \end{aligned} \quad (4.13)$$

Combining equations (4.9), (4.10) and (4.13), we get that

$$\mathbf{E}[\mathbf{M}] = \mathbf{E}[\|x\|^2] \leq n - 0.05. \quad (4.14)$$

As in equation (4.8), by the Brascamp–Lieb inequality (Proposition 11) we have that

$$\mathbf{Var}[\mathbf{M}] \leq \frac{4n}{T}, \quad (4.15)$$

and so by equation (4.14), equation (4.15) and Chebyshev’s inequality, for a suitable choice of  $C$  algorithm CONVEX-DISTINGUISHER will output “truncated” in Step 3 (a) with probability at least 0.9. ■

## 5. Lower bound for testing convex truncation

In this section, we present our lower bound for testing convex truncation. Our lower bound is information-theoretic and applies to all algorithms, computationally efficient or otherwise. Formally, we prove the following theorem.

**Theorem 23.** *Let  $\varepsilon \in (0, 1/2]$ . Let  $A$  be any algorithm which is given access to samples from an unknown distribution  $\mathcal{D}$  and has the following performance guarantee:*

- (1) *if  $\mathcal{D} = N(0, I_n)$ , then with probability at least 9/10 the algorithm outputs “un-truncated”;*
- (2) *if  $\mathcal{D} \in \mathcal{P}_{\text{symm}}$  and has  $d_{\text{TV}}(\mathcal{D}, N(0, I_n)) \geq \varepsilon$ , then with probability at least 9/10 the algorithm outputs “truncated”.*

*Then,  $A$  must draw  $T = \Omega(n/\varepsilon)$  samples from  $\mathcal{D}$ .*

As  $\mathcal{P}_{\text{symm}} \subseteq \mathcal{P}_{\text{conv}}$  and  $\mathcal{P}_{\text{symm}} \subseteq \text{Mix}(\mathcal{P}_{\text{symm}})$ , it is immediate from Theorem 23 that CONVEX-DISTINGUISHER and SYMM-CONVEX-DISTINGUISHER are essentially optimal in terms of sample complexity for testing convex truncation and truncation by a mixture of symmetric convex sets respectively.

### 5.1. Useful preliminaries

We refer the reader to Appendix B for background on Hermite analysis.

**Claim 24.** Suppose  $|\delta| \leq 1$ , and let  $(X, Y)$  be  $\delta$ -correlated mean-0 Gaussians, i.e.,

$$(X, Y) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}\right).$$

Let

$$\kappa := \Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right), \quad \text{so } \mathbf{Pr}_{g \sim N(0,1)}[|g| \leq \kappa] = 1 - \varepsilon. \quad (5.1)$$

For all  $\varepsilon \in (0, 1)$ , we have

$$\Pr[|X| \leq \kappa \text{ and } |Y| \leq \kappa] \leq (1 - \varepsilon)^2 + \delta^2 \cdot \varepsilon.$$

*Proof.* Let  $K \subseteq \mathbb{R}$  be the symmetric interval  $K := [-\kappa, \kappa]$ ; we also write  $K: \mathbb{R} \rightarrow \{0, 1\}$  to denote the indicator function of this interval. We have

$$\begin{aligned} \Pr[|X| \leq \kappa \text{ and } |Y| \leq \kappa] &= \mathbf{E}_{\mathbf{g} \sim N(0,1)}[K(\mathbf{g}) \cdot \mathbf{U}_\delta K(\mathbf{g})] \\ &= \sum_{\alpha \in \mathbb{N}} \delta^{|\alpha|} \mathbf{W}^{|\alpha|}[K] \\ &= (1 - \varepsilon)^2 + \sum_{i=1}^{\infty} \delta^i \mathbf{W}^{=i}[K] \\ &= (1 - \varepsilon)^2 + \delta^2 \left( \sum_{i=1}^{\infty} \delta^{2(i-1)} \mathbf{W}^{=2i}[K] \right), \end{aligned} \quad (5.2)$$

where equation (5.2) uses the fact that  $K$  is an even (i.e., symmetric) function, and hence has no Hermite weight on odd levels.

To conclude, note that

$$\begin{aligned} \sum_{i=1}^{\infty} \delta^{2(i-1)} \mathbf{W}^{=2i}[K] &\leq \sum_{i=1}^{\infty} \mathbf{W}^{=2i}[K] \\ &= \sum_{i=1}^{\infty} \mathbf{W}^{=i}[K] \\ &= \mathbf{Var}[K] = \varepsilon \cdot (1 - \varepsilon) < \varepsilon, \end{aligned}$$

where the first equality again uses that  $K$  is an even function. ■

An equivalent rephrasing of Claim 24 is as follows.

**Corollary 25.** *Let  $v, w \in \mathbb{S}^{n-1}$  and let  $\kappa, \varepsilon$  be as above. Then for all  $\varepsilon \in (0, 1)$ , we have*

$$\Pr_{\mathbf{g} \sim N(0,1)^n}[|v \cdot \mathbf{g}| \leq \kappa \text{ and } |w \cdot \mathbf{g}| \leq \kappa] \leq (1 - \varepsilon)^2 + \varepsilon \cdot \langle v, w \rangle^2.$$

We will also require the following standard concentration bound.

**Fact 26** ([2, Lemma 2.2]). Fix  $v \in \mathbb{S}^{n-1}$ . For  $0 \leq \varepsilon < 1$ , we have

$$\Pr_{w \sim \mathbb{S}^{n-1}}[\langle w, v \rangle \geq \varepsilon] \leq \exp\left(\frac{-n\varepsilon^2}{2}\right).$$

## 5.2. Proof of Theorem 23

We start by defining the family of truncations of  $N(0, I_n)$  by symmetric convex sets that we will employ in our lower bound.

**Definition 27.** For  $v \in \mathbb{S}^{n-1}$ , define  $K_v$  as the set

$$K_v := \{x \in \mathbb{R}^n : |x \cdot v| \leq \kappa\},$$

where  $\kappa = \kappa(\varepsilon)$  is as in equation (5.1), and let  $\mathcal{D}_v := N(0, I_n)|_{K_v}$ .

For notational convenience, let  $P$  be the distribution on  $\mathbb{R}^{nT}$  induced by first drawing a Haar-random vector  $v \sim \mathbb{S}^{n-1}$  and then drawing  $T$  i.i.d. samples from  $\mathcal{D}_v$ . Similarly, let  $Q$  be the distribution on  $\mathbb{R}^{nT}$  induced by drawing  $T$  i.i.d. samples from  $N(0, I_n)$ .

It is clear from Definition 27 that  $\text{Vol}(K_v) = 1 - \varepsilon$  for all  $v \in \mathbb{S}^{n-1}$ . Consequently, we get

$$\left(\frac{dP}{dQ}(x^{(1)}, \dots, x^{(T)})\right)^2 = \left(\frac{1}{1 - \varepsilon}\right)^{2T} \mathbf{E}_{v, w \sim \mathbb{S}^{n-1}} \left[ \prod_{i=1}^T K_v(x^{(i)}) \prod_{i=1}^T K_w(x^{(i)}) \right],$$

and so

$$\begin{aligned} d_{\chi^2}(P \parallel Q) &= \mathbf{E}_{x^{(i)} \sim N(0, I_n)} \left[ \left(\frac{1}{1 - \varepsilon}\right)^{2T} \mathbf{E}_{v, w \sim \mathbb{S}^{n-1}} \left[ \prod_{i=1}^T K_v(x^{(i)}) \prod_{i=1}^T K_w(x^{(i)}) \right] \right] - 1 \\ &= \mathbf{E}_{v, w \sim \mathbb{S}^{n-1}} \left[ \left(\frac{1}{1 - \varepsilon}\right)^{2T} \mathbf{E}_{x^{(i)} \sim N(0, I_n)} \left[ \prod_{i=1}^T K_v(x^{(i)}) \prod_{i=1}^T K_w(x^{(i)}) \right] \right] - 1 \\ &= \mathbf{E}_{v, w \sim \mathbb{S}^{n-1}} \left[ \left(\frac{1}{1 - \varepsilon}\right)^{2T} \mathbf{E}_{x \sim N(0, I_n)} [K_v(x) K_w(x)]^T \right] - 1 \\ &= \mathbf{E}_{v, w \sim \mathbb{S}^{n-1}} [\Delta(v, w)^T] - 1, \end{aligned} \tag{5.3}$$

where  $\Delta(v, w)$  is defined as

$$\Delta(v, w) := \left(\frac{1}{1 - \varepsilon}\right)^2 \mathbf{E}_{x \sim N(0, I_n)} [K_v(x) K_w(x)].$$

Note that equation (5.3) relied on the independence of  $x^{(i)} \sim N(0, I_n)$ .

Let  $T$  be as in Theorem 23. It follows from Corollary 25 and the above that

$$\Delta(v, w) \leq 1 + \frac{\varepsilon \cdot \langle v, w \rangle^2}{(1 - \varepsilon)^2},$$



and so

$$\begin{aligned}
d_{\chi^2}(P \parallel Q) &\leq \mathbf{E}_{\mathbf{v}, \mathbf{w} \sim \mathbb{S}^{n-1}} \left[ \left( 1 + \frac{\varepsilon \cdot \langle \mathbf{v}, \mathbf{w} \rangle^2}{(1-\varepsilon)^2} \right)^T - 1 \right] \\
&= \int_{t=0}^{\infty} \mathbf{Pr}_{\mathbf{v}, \mathbf{w} \sim \mathbb{S}^{n-1}} \left[ \left( 1 + \frac{\varepsilon \cdot \langle \mathbf{v}, \mathbf{w} \rangle^2}{(1-\varepsilon)^2} \right)^T - 1 \geq t \right] dt \\
&= 0.1 + \int_{t=0.1}^{\infty} \mathbf{Pr}_{\mathbf{v}, \mathbf{w} \sim \mathbb{S}^{n-1}} \left[ \left( 1 + \frac{\varepsilon \cdot \langle \mathbf{v}, \mathbf{w} \rangle^2}{(1-\varepsilon)^2} \right)^T - 1 \geq t \right] dt \\
&\leq 0.1 + \int_{t=0.1}^{\infty} \mathbf{Pr}_{\mathbf{v}, \mathbf{w} \sim \mathbb{S}^{n-1}} \left[ \exp\left( \frac{T \cdot \varepsilon \cdot \langle \mathbf{v}, \mathbf{w} \rangle^2}{(1-\varepsilon)^2} \right) - 1 \geq t \right] dt \quad (5.4) \\
&= 0.1 + \int_{t=0.1}^{\infty} \mathbf{Pr}_{\mathbf{v}, \mathbf{w} \sim \mathbb{S}^{n-1}} \left[ \langle \mathbf{v}, \mathbf{w} \rangle^2 \geq \frac{(1-\varepsilon)^2 \ln(1+t)}{T \cdot \varepsilon} \right] dt \\
&\leq 0.1 + \int_{t=0.1}^{\infty} 2 \exp\left( \frac{-n(1-\varepsilon)^2 \ln(1+t)}{2T\varepsilon} \right) dt \quad (5.5) \\
&= 0.1 + \int_{t=0.1}^{\infty} 2(1+t)^{-n(1-\varepsilon)^2/2T\varepsilon} dt \\
&\leq 0.11,
\end{aligned}$$

where equation (5.4) relies on the inequality  $1 + x \leq \exp(x)$ , equation (5.5) relies on Fact 26, and the final inequality relies on  $T = \Omega(n/\varepsilon)$ . Theorem 23 now follows from the above with  $T = T_1$ , thanks to equation (2.2) and Fact 6. ■

## A. Hardness for mixtures of general convex sets

Theorem 2 gives an efficient ( $O(n)$ -sample) algorithm that distinguishes  $N(0, I_n)$  from  $N(0, I_n)$  conditioned on a mixture of (any number of) symmetric convex sets, and Theorem 3 gives an efficient ( $O(n)$ -sample) algorithm that distinguishes  $N(0, I_n)$  from  $N(0, I_n)$  conditioned on any single convex set (which may not be symmetric). We observe here that no common generalization of these results, to mixtures of arbitrary convex sets, is possible with any finite sample complexity, no matter how large.

**Theorem 28.** *Let  $\text{Mix}(\mathcal{P}_{\text{conv}})$  denote the class of all convex combinations (mixtures) of distributions from  $\mathcal{P}_{\text{conv}}$ , and let  $N$  be an arbitrarily large integer ( $N$  may depend on  $n$ , e.g., we may have  $N = 2^{2^n}$ ). For any  $0 < \varepsilon < 1$ , no  $N$ -sample algorithm can successfully distinguish between the standard  $N(0, I_n)$  distribution and an unknown distribution  $\mathcal{D} \in \text{Mix}(\mathcal{P}_{\text{conv}})$  which is such that  $d_{\text{TV}}(N(0, I_n), \mathcal{D}) \geq \varepsilon$ .*

*Sketch of the proof.* The argument essentially follows that of the well-known  $\Omega(\sqrt{L})$ -sample lower bound for testing whether an unknown distribution over the discrete set

$\{1, \dots, L\}$  is uniform or  $\Omega(1)$ -far from uniform [3, 25]. Let  $M = \omega(N^2/(1 - \varepsilon))$ , and consider a(n extremely fine) gridding of  $\mathbb{R}^n$  into disjoint hyper-rectangles  $R$  each of which has  $\text{Vol}(R) = 1/M$ . (For convenience we may think of  $M$  as being an  $n$ -th power of some integer, and of  $\varepsilon$  as being of the form  $1/k$  for  $k$  an integer that divides  $M$ .) We note that for any set  $S$  that is a union of such hyper-rectangles, the distribution  $N(0, I_n)|_S$  is an element of  $\text{Mix}(\mathcal{P}_{\text{conv}})$ .

Let  $S$  be the union of a random collection of exactly  $(1 - \varepsilon)M$  many of the hyper-rectangles  $R$ . We have  $\text{Vol}(S) = (1 - \varepsilon)M$ , so  $d_{\text{TV}}(N(0, I_n), N(0, I_n)|_S) = \varepsilon$ , and consequently a successful  $N$ -sample distinguishing algorithm as described in the theorem must be able to distinguish  $N(0, I_n)$  from the distribution  $\mathcal{D} = N(0, I_n)|_S$ . But it is easy to see that any  $o(\sqrt{(1 - \varepsilon)M})$ -sample algorithm will, with  $1 - o(1)$  probability, receive a sample of points that all come from distinct hyper-rectangles; if this occurs, then the sample will be distributed precisely as a sample of the same size drawn from  $N(0, I_n)$ . ■

## B. Hermite analysis over $N(0, I_n)$

Our notation and terminology follow [32, Chapter 11]. We say that an  $n$ -dimensional *multi-index* is a tuple  $\alpha \in \mathbb{N}^n$ , and we define

$$|\alpha| := \sum_{i=1}^n \alpha_i.$$

For  $n \in \mathbb{N}_{>0}$ , we write  $L^2(\mathbb{R}^n)$  to denote the space of functions  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  that have finite second moment under the Gaussian distribution, i.e.,  $f \in L^2(\mathbb{R}^n)$  if

$$\|f\|^2 = \mathbf{E}_{\mathbf{x} \sim N(0, I_n)}[f(\mathbf{x})^2] < \infty.$$

We view  $L^2(\mathbb{R}^n)$  as an inner product space with

$$\langle f, g \rangle := \mathbf{E}_{\mathbf{x} \sim N(0, I_n)}[f(\mathbf{x})g(\mathbf{x})].$$

We recall the Hermite basis for  $L^2(\mathbb{R}, \gamma)$ .

**Definition 29** (Hermite basis). The *Hermite polynomials*  $(h_j)_{j \in \mathbb{N}}$  are the univariate polynomials defined as

$$h_j(x) = \frac{(-1)^j}{\sqrt{j!}} \exp\left(\frac{x^2}{2}\right) \cdot \frac{d^j}{dx^j} \exp\left(-\frac{x^2}{2}\right).$$

The following fact is standard.

**Fact 30** ([32, Proposition 11.33]). The Hermite polynomials  $(h_j)_{j \in \mathbb{N}}$  form a complete, orthonormal basis for  $L^2(\mathbb{R})$ . For  $n > 1$ , the collection of  $n$ -variate polynomials given by  $(h_\alpha)_{\alpha \in \mathbb{N}^n}$  where

$$h_\alpha(x) := \prod_{i=1}^n h_{\alpha_i}(x)$$

forms a complete, orthonormal basis for  $L^2(\mathbb{R}^n)$ .

Given a function  $f \in L^2(\mathbb{R}^n)$  and  $\alpha \in \mathbb{N}^n$ , we define its *Hermite coefficient on  $\alpha$*  as  $\hat{f}(\alpha) = \langle f, h_\alpha \rangle$ . It follows that  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  can be uniquely expressed as

$$f = \sum_{\alpha \in \mathbb{N}^n} \hat{f}(\alpha) h_\alpha$$

with the equality holding in  $L^2(\mathbb{R}^n, \gamma)$ ; we will refer to this expansion as the *Hermite expansion* of  $f$ . One can check that Parseval's and Plancherel's identities hold in this setting:

$$\langle f, f \rangle = \sum_{\alpha \in \mathbb{N}^n} \hat{f}(\alpha)^2 \quad \text{and} \quad \langle f, g \rangle = \sum_{\alpha \in \mathbb{N}^n} \hat{f}(\alpha) \hat{g}(\alpha).$$

It is also readily verified that  $\mathbf{E}[f(\mathbf{x})] = \hat{f}(0^n)$  and  $\mathbf{Var}[f(\mathbf{x})] = \sum_{\alpha \neq 0^n} \hat{f}(\alpha)^2$ , where  $\mathbf{x} \sim N(0, I_n)$ . We will write  $\mathbf{W}^k[f]$  for the *Hermite weight of  $f$  at level- $k$* , i.e.,

$$\mathbf{W}^k[f] := \sum_{|\alpha|=k} \hat{f}(\alpha)^2,$$

with  $\mathbf{W}^{\leq k}[f]$  defined similarly.

**Definition 31** (Ornstein–Uhlenbeck semigroup). Let  $\rho \in [0, 1]$ . We can define the *Ornstein–Uhlenbeck operator*  $U_\rho$  by its action on  $f \in L^2(\mathbb{R}^n)$  as follows:

$$U_\rho f(x) := \mathbf{E}_{\mathbf{g} \sim N(0, I_n)}[f(\rho x + \sqrt{1 - \rho} \mathbf{g})].$$

The Ornstein–Uhlenbeck semigroup is sometimes referred to as the family of *Gaussian noise operators* or *Mehler transforms*. The Ornstein–Uhlenbeck semigroup acts on the Hermite expansion as follows.

**Fact 32** ([32, Proposition 11.33]). For  $f \in L^2(\mathbb{R}^n, \gamma)$ , the function  $U_\rho f$  has Hermite expansion

$$U_\rho f = \sum_{\alpha \in \mathbb{N}^n} \rho^{|\alpha|} \hat{f}(\alpha) h_\alpha.$$

**Acknowledgments.** This work was partially completed while S. N. was a participant in the program on “Meta-Complexity” at the Simons Institute for the Theory of Computing.

**Funding.** A. De is supported by NSF grants CCF-1910534, CCF-1926872, and CCF-2045128. S. Nadimpalli is supported by NSF grants IIS-1838154, CCF-2106429, CCF-2211238, CCF-1763970, and CCF-2107187. R. A. Servedio is supported by NSF grants IIS-1838154, CCF-2106429, and CCF-2211238. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF).

## References

- [1] N. Balakrishnan and E. Cramer, *The art of progressive censoring*. Stat. Ind. Technol., Birkhäuser/Springer, New York, 2014 Zbl [1365.62001](#) MR [3309531](#)
- [2] K. Ball, *An elementary introduction to modern convex geometry*. In *Flavors of geometry*, pp. 1–58, Math. Sci. Res. Inst. Publ. 31, Cambridge University Press, Cambridge, 1997 Zbl [0901.52002](#) MR [1491097](#)
- [3] T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White, *Testing closeness of discrete distributions*. *J. ACM* **60** (2013), no. 1, article no. 4 Zbl [1281.68227](#) MR [3033221](#)
- [4] T. Batu, R. Kumar, and R. Rubinfeld, *Sublinear algorithms for testing monotone and unimodal distributions*. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pp. 381–390, ACM, New York, 2004 Zbl [1192.68345](#) MR [2121623](#)
- [5] D. Bernoulli, Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir. *Histoire de l’Acad., Roy. Sci. (Paris) avec Mem.* (1760), 1–45
- [6] H. J. Brascamp and E. H. Lieb, *On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation*. *J. Functional Analysis* **22** (1976), no. 4, 366–389 Zbl [0334.26009](#) MR [0450480](#)
- [7] C. L. Canonne, *A survey on distribution testing: Your data is big. But is it blue?* Graduate Surveys 19, Theory of Computing Library, 2020
- [8] C. L. Canonne, X. Chen, G. Kamath, A. Levi, and E. Waingarten, *Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning*. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 321–336, SIAM, Philadelphia, PA, 2021 Zbl [07788359](#) MR [4262454](#)

- [9] X. Chen, A. De, and R. A. Servedio, [Testing noisy linear functions for sparsity](#). In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 610–623, ACM, New York, NY, 2020
- [10] X. Chen, A. Freilich, R. A. Servedio, and T. Sun, Sample-based high-dimensional convexity testing. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, article no. 37, LIPIcs. Leibniz Int. Proc. Inform. 81, Schloss Dagstuhl – Leibniz Zentrum für Informatik, Wadern, 2017 Zbl [1467.68209](#) MR [3695604](#)
- [11] A. C. Cohen, *Truncated and censored samples. Theory and applications*. Stat. Textb. Monogr. 119, Marcel Dekker, New York, 1991 Zbl [0742.62027](#) MR [1153408](#)
- [12] C. Daskalakis, I. Diakonikolas, R. A. Servedio, G. Valiant, and P. Valiant, [Testing  \$k\$ -modal distributions: Optimal algorithms via reductions](#). In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1833–1852, SIAM, Philadelphia, PA, 2012 Zbl [1421.68187](#) MR [3221282](#)
- [13] C. Daskalakis, T. Gouleakis, C. Tzamos, and M. Zampetakis, [Efficient statistics, in high dimensions, from truncated samples](#). In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 639–649, IEEE Computer Society, Los Alamitos, CA, 2018 MR [3899629](#)
- [14] C. Daskalakis, T. Gouleakis, C. Tzamos, and M. Zampetakis, Computationally and statistically efficient truncated regression. In *Proceedings of the Thirty-Second Conference on Learning Theory*, pp. 955–960, Proceedings of Machine Learning Research 99, PMLR, 2019
- [15] C. Daskalakis, V. Kontonis, C. Tzamos, and E. Zampetakis, A statistical Taylor theorem and extrapolation of truncated densities. In *Proceedings of Thirty Fourth Conference on Learning Theory*, pp. 1395–1398, Proceedings of Machine Learning Research 134, PMLR, 2021
- [16] A. De, S. Nadimpalli, and R. A. Servedio, Convex influences. 2021, arXiv:[2109.03107](#)
- [17] A. De, S. Nadimpalli, and R. A. Servedio, Quantitative correlation inequalities via semi-group interpolation. In *12th Innovations in Theoretical Computer Science Conference*, article no. 69, LIPIcs. Leibniz Int. Proc. Inform. 185, Schloss Dagstuhl – Leibniz Zentrum für Informatik, Wadern, 2021 MR [4214313](#)
- [18] A. De, S. Nadimpalli, and R. A. Servedio, Convex influences. In *13th Innovations in Theoretical Computer Science Conference*, article no. 53, LIPIcs. Leibniz Int. Proc. Inform. 215, Schloss Dagstuhl – Leibniz Zentrum für Informatik, Wadern, 2022 Zbl [07829285](#) MR [4459424](#)
- [19] A. De and R. A. Servedio, Weak learning convex sets under normal distributions. In *Proceedings of Thirty-Fourth Conference on Learning Theory*, pp. 1399–1428, Proceedings of Machine Learning Research 134, PMLR, 2021
- [20] I. Diakonikolas, D. M. Kane, and A. Pensia, [Gaussian mean testing made simple](#). In *2023 Symposium on Simplicity in Algorithms (SOSA)*, pp. 348–352, SIAM, Philadelphia, PA, 2023 Zbl [07848516](#) MR [4538689](#)
- [21] L. H. Dicker, [Variance estimation in high-dimensional linear models](#). *Biometrika* **101** (2014), no. 2, 269–284 Zbl [1452.62495](#) MR [3215347](#)

- [22] D. Fotakis, A. Kalavasis, and C. Tzamos, Efficient parameter estimation of truncated Boolean product distributions. In *Proceedings of the Thirty-Third Conference on Learning Theory*, pp. 1586–1600, Proceedings of Machine Learning Research 125, PMLR, 2020
- [23] F. Galton, [An examination into the registered speeds of American trotting horses, with remarks on their value as hereditary data](#). *Lond. R. S. Proc.* **62** (1898), no. 379–387, 310–315 Zbl [29.0195.01](#)
- [24] O. Goldreich, S. Goldwasser, and D. Ron, [Property testing and its connection to learning and approximation](#). *J. ACM* **45** (1998), no. 4, 653–750 Zbl [1065.68575](#) MR [1675099](#)
- [25] O. Goldreich and D. Ron, On testing expansion in bounded-degree graphs. *Electron. Colloquium Comput. Complex.* **TR00** (2000), article no. 20
- [26] A. Klivans, R. O’Donnell, and R. Servedio, [Learning geometric concepts via Gaussian surface area](#). In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 541–550, IEEE Computer Society, Los Alamitos, CA, 2008
- [27] W. Kong, E. Brunskill, and G. Valiant, Sublinear optimal policy value estimation in contextual bandits. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pp. 4377–4387, Proceedings of Machine Learning Research 108, PMLR, 2020
- [28] W. Kong and G. Valiant, Estimating learnability in the sublinear data regime. In *Advances in Neural Information Processing Systems 31*, Curran Associates, 2018
- [29] V. Kontonis, C. Tzamos, and M. Zampetakis, [Efficient truncated statistics with unknown truncation](#). In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1578–1595, IEEE Computer Society, Los Alamitos, CA, 2019 MR [4228243](#)
- [30] L. Leindler, On a certain converse of Hölder’s inequality. II. *Acta Sci. Math. (Szeged)* **33** (1972), no. 3–4, 217–223 Zbl [0245.26011](#) MR [2199372](#)
- [31] 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*, pp. 169–187, Lecture Notes in Math. 1807, Springer, Berlin, 2003 Zbl [1036.52014](#) MR [2083397](#)
- [32] R. O’Donnell, *Analysis of Boolean functions*. Cambridge University Press, New York, 2014 Zbl [1336.94096](#) MR [3443800](#)
- [33] K. Pearson, On the systematic fitting of frequency curves. *Biometrika* **2** (1902), 2–7
- [34] A. Prékopa, On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)* **34** (1973), 335–343 Zbl [0264.90038](#) MR [0404557](#)
- [35] R. Rubinfeld and R. A. Servedio, [Testing monotone high-dimensional distributions](#). *Random Structures Algorithms* **34** (2009), no. 1, 24–44 Zbl [1165.62037](#) MR [2478537](#)
- [36] H. Schneider, *Truncated and censored samples from normal populations*. Stat., Textb. Monogr. 70, Marcel Dekker, New York-Basel, 1986 Zbl [0601.62052](#)
- [37] S. S. Vempala, Learning convex concepts from Gaussian distributions with PCA. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 124–130, IEEE Computer Society, Los Alamitos, CA, 2010 Zbl [08033968](#) MR [3024782](#)

Received 3 June 2023; revised 21 November 2024.

**Anindya De**

Department of Computer and Information Science, University of Pennsylvania,  
3451 Walnut Street, Philadelphia, PA 19104, USA; [anindyad@cis.upenn.edu](mailto:anindyad@cis.upenn.edu)

**Shivam Nadimpalli**

Department of Mathematics, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02139, USA; [shivamn@mit.edu](mailto:shivamn@mit.edu)

**Rocco A. Servedio**

Department of Computer Science, Columbia University, 116th Street and Broadway,  
New York, NY 10027, USA; [ras2105@columbia.edu](mailto:ras2105@columbia.edu)