Gaussian Approximation of Convex Sets by Intersections of Halfspaces*

Anindya De
Department of Computer Science
University of Pennsylvania
Philadelphia, USA
anindyad@cis.upenn.edu

Shivam Nadimpalli
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, USA
shivamn@mit.edu

Rocco A. Servedio
Department of Computer Science
Columbia University
New York, USA
ras2105@columbia.edu

***Abstract*—We study the approximability of general convex sets in $\mathbb{R}^n$ by intersections of halfspaces, where the approximation quality is measured with respect to the *standard Gaussian distribution* and the complexity of an approximation is the number of halfspaces used. While a large body of research has considered the approximation of convex sets by intersections of halfspaces under distance metrics such as the Lebesgue measure and Hausdorff distance, prior to our work there has not been a systematic study of convex approximation under the Gaussian distribution.**

We establish a range of upper and lower bounds, both for general convex sets and for specific natural convex sets that are of particular interest. Our results demonstrate that the landscape of approximation is intriguingly different under the Gaussian distribution versus previously studied distance measures.

Our results are proved using techniques from many different areas. These include classical results on convex polyhedral approximation, Cramér-type bounds on large deviations from probability theory, and—perhaps surprisingly—a range of topics from computational complexity, including computational learning theory, unconditional pseudorandomness, and the study of influences and noise sensitivity in the analysis of Boolean functions.

***Index Terms*—convex geometry, polyhedral approximation, Boolean functions**

I. Introduction

A long line of mathematical research has investigated *convex polyhedral approximation of convex bodies*, i.e. the broad question of how to best approximate general convex bodies using intersections of halfspaces (or equivalently, convex hulls of finite point sets). Research on questions of this sort dates back at least to the first half of the twentieth century, see for example the early works of Sas [65], Fejes Tóth [71] and Macbeath [53], among others. Contemporary motivation for the study of polyhedral approximation of convex bodies arises from many areas including discrete and computational geometry, geometric convexity, the study of finite-dimensional normed spaces, and optimization. Given this breadth of connections, it is not surprising that the existing body of work on the subject is vast, as witnessed by the hundreds of references that appear in multiple surveys (including the 1983 [33] and 1993 [34] surveys of Gruber and the 2008 survey of Bronstein [15]) on approximation of convex bodies by polyhedra.

A.D. is supported by NSF grants CCF-1910534 and CCF-2045128. S.N. is supported by NSF grants IIS-1838154, CCF-2106429, CCF-2211238, CCF1763970, and CCF-2107187. R.A.S. is supported by NSF grants IIS-1838154, CCF-2106429, and CCF-2211238.

In the polyhedral approximation literature to date, a number of different distance notions have been used to measure the accuracy of convex polyhedral approximations of convex sets. These include the Hausdorff distance, the Nikodym metric (volume of the symmetric difference), the Banach-Mazur distance, distances generated by various L_p metrics, and more. (See [15] for descriptions of each of these metrics and for an extensive overview of approximability results under each of those distance measures.)

This paper adopts a new perspective, by considering approximation of convex bodies in $\mathbb{R}^n$ using the *(standard) Gaussian volume of the symmetric difference* as the distance measure. We refer to this distance measure simply as the *Gaussian distance*. This is a natural and well-studied distance measure in theoretical computer science, as witnessed by the many works that have considered learning, testing, derandomization, and other theoretical computer science problems using it (see e.g. [17], [20], [21], [27], [29], [37], [39]–[41], [43], [45]–[47], [55], [61], [72] and references therein). Given this body of work, it is perhaps surprising that Gaussian distance does not appear to have been previously studied in a systematic way for convex polyhedral approximation.

Let us make our notion of approximation precise:

Definition 1. Given (measurable) sets $K, L \subseteq \mathbb{R}^n$, we define the *Gaussian distance* between K and L to be

$$\mathrm{dist}_{\mathrm{G}}(K,L) := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim N(0,I_n)}[\boldsymbol{x} \in K \triangle L]$$

where $N(0, I_n)$ denotes the n-dimensional standard Gaussian distribution and $K \triangle L = (K \setminus L) \cup (L \setminus K)$ is the symmetric difference of the sets K and L.

We consider approximators which are intersections of finitely many n-dimensional halfspaces, and we measure the complexity of such an approximator by the number of halfspaces (facets) that it contains; we refer to this as its *facet complexity*. Thus, we are concerned with the following broad question:

> Given a convex set K in $\mathbb{R}^n$ and a value $0 < \varepsilon < 1/2$, what is the minimum facet complexity of an

intersection of halfspaces L such that $\mathrm{dist}_{\mathrm{G}}(K, L) \leq \varepsilon$?

Motivation

The broad approximation question stated above arises naturally in the context of contemporary theoretical computer science. As alluded to earlier, an increasing number of results across different areas of TCS involve convex sets and the Gaussian distance; taking the next step on various natural problems in these areas would seem to require an understanding of how well intersections of halfspaces can approximate convex sets in $\mathbb{R}^n$. We give two specific examples below:

- **Testing Convex Sets:** A number of researchers have considered the property testing problem of efficiently determining whether an unknown set $K \subseteq \mathbb{R}^n$ is convex versus far from convex [6], [8], [9], [18], [19], [63]. The Gaussian distance provides a clean mathematical framework for this problem, and one in which a natural analogy emerges between the well-studied problem of monotonicity testing over the Boolean hypercube and convexity testing [19]. Understanding how convex sets can be approximated by "simpler objects" under the Gaussian distance seems likely to be helpful in developing algorithms and lower bound arguments for testing convexity. (In support of this thesis, we recall that the study of approximating linear threshold functions by "simpler objects" [66] yielded insights and technical ingredients, namely the notion of the "critical index," that played an essential role in the development of efficient algorithms for testing linear threshold functions [55].) Indeed, a very recent work has used constructions that are inspired by the construction employed in our upper bound in Section V to give the first super-constant lower bound for testing convexity over $\mathbb{R}^n$ [19]. We also recall that several papers have given efficient testing algorithms for single halfspaces [55] and intersections of k halfspaces [20], [21] over $\mathbb{R}^n$ under the Gaussian distance; combining these results with results about approximating general convex sets by intersections of halfspaces offers a potential avenue to developing efficient testers for general convex sets.
- **Properly Learning Convex Sets:** It has long been known [16] that monotone Boolean functions over $\{-1,1\}^n$ can be learned to any constant accuracy under the uniform distribution in time $2^{\tilde{O}(\sqrt{n})}$, via the "low-degree" algorithm based on the Fourier decomposition. Similarly, it has been known for some time [45] that convex sets can be learned to any constant accuracy under the Gaussian distance, via a "Hermite polynomial" variant of the low-degree algorithm. However, neither the algorithms of [16] nor [45] are *proper*: in the monotone case the output hypothesis of the learning algorithm is not a monotone function, and in the convex setting the output hypothesis is not a convex set. Exciting recent work [49] has given a $2^{\tilde{O}(\sqrt{n})}$-time uniform-distribution learning algorithm for monotone Boolean functions over $\{-1,1\}^n$ that *is* proper (even in the demanding *agnostic learning* model, see [50]); given this, it is a natural goal to seek a $2^{\tilde{O}(\sqrt{n})}$-time proper learning algorithm for convex sets under the Gaussian distance. What would the output hypothesis of such a hoped-for proper learning algorithm look like? It seems quite plausible that it would be an intersection of halfspaces; this begs the question of understanding the capabilities and limitations of intersections of halfspaces as approximators for general convex sets in $\mathbb{R}^n$.

We further note that intersections of halfspaces have been intensively studied in many branches of "concrete complexity theory;" indeed, many of the TCS results alluded to earlier that involve Gaussian distance are more specifically about intersections of halfspaces, e.g. [17], [20], [21], [29], [37], [45], [72].

We also remark that the suite of tools which turn out to be relevant in our study link our investigations closely to theoretical computer science. Perhaps unexpectedly, we will see that there are close connections between the study of facet complexity of polyhedral convex approximators and a number of fundamental ingredients and results (such as noise sensitivity, random restrictions, influence of variables/directions, various extremal constructions, etc.) in the *analysis of Boolean functions* over the discrete domain $\{-1,1\}^n$. From this perspective, our study can be seen as a continuation of the broad theme of exploring the emerging analogy between convex sets in $\mathbb{R}^n$ under the Gaussian distribution and monotone Boolean functions over $\{-1,1\}^n$ under the uniform distribution (see e.g. [24], [25], [28], [37]).

As a point aspect of motivation, we mention that results on the (in)approximability of general convex sets by intersections of few halfspaces can be useful for the study of probability theory in its own right. Indeed, the lower bounds that we prove in Section VII of this paper have very recently been used to give a lower bound on *sparsifying Gaussian processes* while approximately preserving their suprema [22].

Before turning to an overview of our results and techniques, let us fix some convenient notation and terminology:

Notation 2. If $L \subseteq \mathbb{R}^n$ is an intersection of N halfspaces, then as stated earlier we say that the *facet complexity* of L is N. Given a convex set $K \subseteq \mathbb{R}^n$ and a value $0 \leq \varepsilon < 1/2$, we write $\mathrm{FC}(K, \varepsilon)$ to denote the minimum value N such that there is a convex set L that is an intersection N halfspaces satisfying $\mathrm{dist}_{\mathrm{G}}(K, L) \leq \varepsilon$.

II. Overview of Results and Techniques

We start with an overview of our positive results.

A. Overview of Positive Results

We start with our positive results on approximability of convex sets under the Gaussian distance by intersections of halfspaces.

a) Universal Approximation via Hausdorff Distance Approximation: As a warmup, we begin in Section IV-A by giving a fairly simple "universal approximation" result that upper bounds $\mathrm{FC}(K, \varepsilon)$ for any convex set $K \subseteq \mathbb{R}^n$. The key observation (implicit in the proof of Theorem 18) is that for any convex set K, if L is an intersection of halfspaces which is a high-accuracy *outer*[1] approximator of K under the Hausdorff distance measure, then L is also a (slightly lower accuracy) $\mathrm{dist}_{\mathrm{G}}$-approximator for K. (Recall that the Hausdorff distance between two sets $S, T \subseteq \mathbb{R}^n$ is

$$\max\left\{\sup_{s\in S}\|s-T\|, \sup_{t\in S}\|t-S\|\right\}$$

where $\|x-Y\| = \inf_{y\in Y}\|x-y\|$ and $\|x-y\|$ is the Euclidean distance between x and y.) This is easy to establish by simply integrating Keith Ball's universal $O(n^{1/4})$ upper bound [3] on the *Gaussian surface area* of any convex set (see Theorem 15), using the fact that for any $t > 0$ and any convex set K, the "t-enlargement" K_t of K is also a convex set (see the proof of Theorem 18).

Combining the above observation with standard Gaussian tail bounds and classical upper bounds on the facet complexity of an outer Hausdorff approximator for any convex body with bounded radius, we obtain the following:

Theorem 3 (Informal version of Theorem 18)**.** For any convex set K and any $\varepsilon > 0$, we have $\mathrm{FC}(K, \varepsilon) \le (n/\varepsilon)^{O(n)}$.

b) A "Relative-Error" Universal Approximation Sharpening of Theorem 3: Suppose that $K \subseteq \mathbb{R}^n$ is a convex set whose Gaussian volume is very large, i.e. $1-\delta$ where δ is very small. In this situation the trivial approximator which is all of $\mathbb{R}^n$ already has $\mathrm{dist}_{\mathrm{G}}(K, \mathbb{R}^n) = \delta$, so a natural goal for approximation is to achieve small error *relative to* the δ amount of mass which lies outside of K, i.e. to construct an intersection of halfspaces L for which

$$\mathrm{dist}_{\mathrm{G}}(K, L) \le \varepsilon\delta.$$

While Theorem 3 shows that $(n/(\varepsilon\delta))^{O(n)}$ halfspaces suffice for this, in Section IV-B we give a stronger result which has a significantly improved dependence on δ:

Theorem 4 (Informal version of Theorem 19)**.** Let $0 < \varepsilon, \delta < 1$ and let $K \subsetneq \mathbb{R}^n$ be a convex set with $\mathrm{Vol}(K) = 1-\delta$. Then

$$\mathrm{FC}(K, \varepsilon\delta) \le \frac{1}{\delta}\cdot\left(\frac{n}{\varepsilon}\log\left(\frac{1}{\delta}\right)\right)^{O(n)}.$$

(We remark that the strengthening which Theorem 19 achieves over Theorem 18 will be crucial for the construction of our ℓ_p ball approximators, discussed below.)

We briefly explain the main idea underlying the construction of the approximator of Theorem 19. First, we recall a basic property of the $N(0, I_n)$ distribution, which is that along any ray $\{tv : t \ge 0\}$ from the origin in $\mathbb{R}^n$, the distribution of Gaussian mass along that ray follows the chi-distribution (see Section III-B). Since $\mathrm{Vol}(K) = 1-\delta$, this means that the "amount of chi-distribution mass" which is "lopped off" by K along direction v, averaged over all directions v, is δ. The high-level idea of our construction of the approximator L is to place tangent hyperplanes on the boundary of K in such a way as to "lop off" at least a $(1-\varepsilon)$-fraction of the total δ amount of Gaussian mass that lies outside the set K. We do this via an analysis that proceeds direction by direction: We can ignore directions in which K "lops off" only a very small amount (less than $\varepsilon\delta/4$, say) of the chi-distribution's mass. For the other directions, we use a bucketing scheme, a suitable net construction for the points in each bucket, and a mixture of careful geometric arguments and tail bounds to argue that not too many halfspaces are required to adequately handle all of the other directions.

c) Sub-Exponential Approximation for Specific Convex Sets: Going beyond the universal approximation results mentioned above, it is natural to wonder whether improved approximation bounds can be obtained for specific interesting convex sets. Perhaps the two most natural convex sets to consider in this context are the ℓ_1 and ℓ_2 balls of Gaussian volume $1/2$, which we briefly discuss below.[2]

The ℓ_1 ball of Gaussian volume $1/2$, also known as the "spectrahedron" or "cross-polytope," is the convex set

$$\left\{x \in \mathbb{R}^n \ : \|x\|_1 \le cn\right\},$$

where $c = \sqrt{2/\pi} \pm o_n(1)$ (this is an easy consequence of the Berry-Esseen theorem). This set is an intersection of exactly 2^n halfspaces, namely

$$b_1x_1 + \cdots + b_nx_n \le cn \text{ for all } (b_1, \ldots, b_n) \in \{-1, 1\}^n,$$

and it is natural to wonder whether this set can be approximated as an intersection of significantly fewer than 2^n halfspaces. (As we will explain later, it is slightly more convenient for us to work with a slight rescaling of this ball, namely the set

$$B_1 := \left\{x \in \mathbb{R}^n \ : \ \|x\|_1 \le \sqrt{\frac{2}{\pi}}n\right\},$$

which can easily be shown to have $\mathrm{Vol}(B_1) = 1/2 \pm o_n(1)$.)

The ℓ_2 ball of Gaussian volume $1/2$ is simply the Euclidean ball

$$\left\{x \in \mathbb{R}^n \ : \|x\|^2 \le \mathrm{median}(\chi^2(n))\right\}.$$

(Of course, writing this set *exactly* as an intersection of halfspaces requires infinitely many halfspaces.) Using well-known bounds on the chi-squared distribution we have that the Euclidean ball of radius $\sqrt{n}$, which we denote $B(\sqrt{n})$, has Gaussian volume $1/2 + o_n(1)$; it will be slightly more convenient for us to use this as the ℓ_2 ball that we seek to approximate.

[1] Recall that L is an *outer* (respectively *inner*) approximator to K if $K \subseteq L$ (respectively $L \subseteq K$).

[2] It is also natural to wonder about the ℓ_∞ ball, but the ℓ_∞ ball $\{x \in \mathbb{R}^n \ : \ \|x\|_\infty \le r\}$ is an intersection of $2n$ halfspaces $-r \le x_i \le r, i \in [n]$.

d) A $2^{O(\sqrt{n})}$-Facet Approximator for the ℓ_2 Ball: For several standard distance measures, including the Hausdorff distance and the symmetric-difference measure, the ℓ_2 ball is the "hardest case" for approximation by intersections of halfspaces, matching or essentially matching known upper bounds for universal approximation. For example, it is known [15] that any intersection of halfspaces which is an ε-approximator to $B(\sqrt{n})$ in Hausdorff distance must have facet complexity $\Omega((\sqrt{n}/\varepsilon)^{(n-1)/2})$, matching the universal approximation upper bound for Hausdorff distance to within a constant factor [2]. For the symmetric-difference measure, Ludwig, Schütt and Werner [52] showed that if L is an intersection of M halfspaces for which

$$\frac{\mathrm{Leb}(L \,\triangle\, B(\sqrt{n}))}{\mathrm{Leb}(B(\sqrt{n}))} \leq c$$

(here Leb denotes volume under the standard Lebesgue measure), where $c > 0$ is some absolute constant, then M must be at least $2^{\Omega(n)}$.

In contrast, for the Gaussian distance we are able to exploit the rotational symmetry of the standard Gaussian distribution and tail bounds on the chi-distribution to get an approximation of the ℓ_2 ball which uses much fewer than 2^n facets, and hence is much better than the universal approximation upper bounds given by Theorem 3 or Theorem 4. For any constant ε, we give a probabilistic construction of a $2^{O(\sqrt{n})}$-facet polytope which is an ε-approximator of the ℓ_2 ball. This construction is analogous to a probabilistic construction of O'Donnell and Wimmer [62], which shows that a variant of a random monotone CNF construction due to Talagrand [69] gives an ε-accurate approximator for the Boolean Majority function. We similarly modify a probabilistic construction of a convex body due to Nazarov [57], and use it to prove the following in Section V:

Theorem 5 (Informal version of Theorem 26)**.** Let B_2 denote the ℓ_2 ball of radius $\sqrt{n}$ in $\mathbb{R}^n$ (so $\mathrm{Vol}(B_2) = 1/2 + o_n(1)$). Then for any constant $\varepsilon > 0$,

$$\mathrm{FC}(B_2, \varepsilon) \leq 2^{O(\sqrt{n})}.$$

e) A $2^{O(n^{3/4})}$-Facet Approximator for the ℓ_1-Ball: Since the ℓ_1 ball does not enjoy the same level of rotational symmetry as the ℓ_2 ball, it is natural to wonder whether it can be similarly approximated with a sub-exponential number of facets. As we now explain, another motivation for this question comes from considering a result of O'Donnell and Wimmer on the inability of small CNF formulas to approximate the "tribes" Boolean DNF formula over $\{0,1\}^n$, in the context of a recently-explored analogy between monotone Boolean functions and symmetric convex sets [24], [25], [28], [37].

In [62] O'Donnell and Wimmer showed that any CNF formula that computes the n-variable DNF tribes function correctly on $(1-\varepsilon)\cdot 2^n$ inputs, for $\varepsilon = 0.1$, must have $2^{\Omega(n/\log n)}$ clauses. As detailed in [25], [28], there is a natural correspondence between s-clause CNF formulas over $\{0,1\}^n$ and symmetric intersections of $2s$ halfspaces over $N(0, I_n)$ (this correspondence is at the heart of the main lower bound of [28] and several of the results of [25], [37]). Thus, in seeking convex sets that may require $2^{\tilde{\Omega}(n)}$ halfspaces to approximate under $N(0,1)^n$, it is natural to look for a Gaussian space convex analogue of of the DNF tribes function over $\{0,1\}^n$. Now,

- The DNF tribes function is the Boolean dual of the CNF Tribes function;
- The Gaussian-space polytope corresponding to the CNF Tribes function is the ℓ_∞ ball [25], [37]; and
- The polytope which is dual to the ℓ_∞ ball is the ℓ_1 ball.

Thus the following question naturally suggests itself:

> Does the ℓ_1 ball B_1 in $\mathbb{R}^n$ require $2^{\tilde{\Omega}(n)}$ halfspaces for constant-accuracy approximation, analogous to how the DNF Tribes function requires $2^{\tilde{\Omega}(n)}$-clause CNFs for constant-accuracy approximation?

While the above conjecture may seem intuitively plausible (and indeed, the authors initially tried to prove it), it turns out to be false. In Section 6 of the full version of this paper [26], we show that the ℓ_1 ball can be approximated to any constant accuracy as an intersection of sub-exponentially many halfspaces:

Theorem 6 (Theorem 29 of [26])**.** Let B_1 denote the origin-centered ℓ_1 ball, i.e.

$$B_1 := \left\{ x \in \mathbb{R}^n \ : \ \|x\|_1 \leq \sqrt{\frac{2}{\pi}} n \right\}.$$

Then for any constant ε, we have

$$\mathrm{FC}(B_1, \varepsilon) \leq 2^{O(n^{3/4})}.$$

We prove Theorem 6 using a probabilistic construction, but one which is very different from the probabilistic construction described above in the sketch of Theorem 5 for the ℓ_2 ball. The high-level intuition is as follows: Determining whether or not a point $x \in \mathbb{R}^n$ lies within B_1 is the same as determining whether or not the sum $|x_1| + |x_2| + \cdots + |x_n|$ exceeds the threshold value $\sqrt{2/\pi}n$. Given this, the main idea is to take a probabilistic approach which exploits both *anti-concentration* and *sampling*:

- Anti-concentration tells us that for $\boldsymbol{x} \sim N(0, I_n)$, only a small fraction of outcomes of $|\boldsymbol{x}_1| + \cdots + |\boldsymbol{x}_n|$ will lie very close to the boundary value of $\sqrt{2/\pi}n$. Thus, at the cost of a small error, we can assume that a typical outcome x of $\boldsymbol{x}$ either has $|x_1| + \cdots + |x_n| \geq \sqrt{2/\pi}n + \tau$ (call this a "heavy" x), or has $|x_1| + \cdots + |x_n| \leq \sqrt{2/\pi}n - \tau$ (call this a "light" x) for a suitable margin parameter τ.
- Given this, we can use a sampling-based approach: if we uniformly sample m coordinates $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m$ from $[n]$ and evaluate $|x_{\boldsymbol{i}_1}| + \cdots + |x_{\boldsymbol{i}_m}|$, then for a suitable threshold θ_1 there will be a noticeable gap between (i) the (extremely large) probability that $|x_{\boldsymbol{i}_1}| + \cdots + |x_{\boldsymbol{i}_m}| \leq \theta_1$ when x is a typical light point, and (ii) the (still very large, but slightly smaller) probability that $|x_{\boldsymbol{i}_1}| + \cdots + |x_{\boldsymbol{i}_m}| \leq \theta_1$ when x is a typical heavy point.

The above gap means that by ANDing together a carefully chosen number M of random sets of the form

$$\{x \in \mathbb{R}^n \ : \ |x_{i_1}| + \cdots + |x_{i_m}| \le \theta_1\}, \tag{1}$$

we get an approximator for B_1 which is accurate on almost all of the points $x \in \mathbb{R}^n$ which are either light or heavy (and, as sketched above, almost all points drawn from $N(0, I_n)$ are either light or heavy, by anti-concentration). The detailed analysis requires sophisticated Cramér-type bounds which give very tight *multiplicative* control on tails of sums of random variables drawn *without* replacement from a finite population of values.

Note that each set of the form (1) is an m-variable "junta" ℓ_1 ball over its m relevant coordinates (and is also an intersection of 2^m halfspaces), so the AND of M such sets is an intersection of $M2^m$ many halfspaces. With careful setting of parameters we get that this is at most $2^{O(n^{3/4})}$ halfspaces, for any constant-factor approximation.

We can in fact extend this approach to obtain a $2^{\tilde{O}(n^{3/4})}$-facet approximator for the ℓ_p ball of volume-$1/2$ for $1 \le p < 2$; we refer the interested reader to the full version of this paper [26] for more details. This upper bound crucially relies on our "relative-error" universal approximation bound, Theorem 4.

B. Overview of Lower Bounds

We now turn to a technical overview of our lower bounds.

a) Non-Explicit Average-Case Lower Bounds: It is well known that standard counting arguments easily yield strong (average-case) lower bounds on the complexity of approximating (non-explicit) *Boolean functions* that map $\{-1, +1\}^n$ to $\{-1, +1\}$. For a range of different computational models, these lower bounds show that even achieving approximation error $1/2 - o_n(1)$ requires any approximator to be exponentially large.

In our context of approximating convex sets in $\mathbb{R}^n$ by intersections of halfspaces, counting arguments are not quite as straightforward because there are infinitely many distinct halfspaces over $\mathbb{R}^n$. Nevertheless, counting arguments can be brought to bear to prove lower bounds. In more detail, [45] used a counting argument to establish the existence of $N = 2^{2^{\Omega(\sqrt{n})}}$ distinct convex sets $K_1, \ldots, K_N$ in $\mathbb{R}^n$, each of which is an intersection of $2^{\Omega(\sqrt{n})}$ many halfspaces, such that $\mathrm{dist}_{\mathrm{G}}(K_i, K_j) \ge 1/44000$ for each $1 \le i \ne j \le N$. Using this result, it is possible to establish the existence of a convex set K (one of the sets $K_1, \ldots, K_N$) such that $\mathrm{FC}(K, c) \ge 2^{c\sqrt{n}}$, where $c > 0$ is a small absolute constant.

In Section VI we give a stronger (but still non-explicit) lower bound, which shows that many halfspaces are required even to achieve error $1/2 - o_n(1)$; in other words, we give an *average-case* lower bound (also known as a *correlation bound*). Our lower bound trades off the accuracy of the approximator against the number of halfspaces:

Theorem 7 (Informal version of Theorem 28)**.** For any $\tau = \omega(n^{-1/2} \cdot \log n)$, there exists some convex set $K \subset \mathbb{R}^n$ that has $\mathrm{FC}(K, \frac{1}{2} - \tau) = 2^{\Omega(\tau \cdot \sqrt{n})}$.

For example, taking $\tau = n^{-1/4}$, Theorem 7 implies the existence of an n-dimensional convex set K such that no intersection of $2^{cn^{1/4}}$ many halfspaces can approximate K to accuracy even as large as $1/2 + n^{-1/4}$; taking τ to be a constant, we recover the non-explicit $2^{\Omega(\sqrt{n})}$ lower bound mentioned in the previous paragraph.

The proof of Theorem 7 combines classical results from statistical learning theory with an information theoretic lower bound on *weak* learning convex sets under $N(0, I_n)$ that was established in recent work [28]. A high-level sketch is as follows: If every convex set could be approximated to high accuracy $(1/2 - \tau)$ by an intersection of "few" halfspaces, then classical results from statistical learning theory would imply the existence of (computationally inefficient but sample-efficient) algorithms to "weakly" learn any convex set K to error $1/2 - \tau/2$. On the other hand, a recent result of De and Servedio [28] gives a strong lower bound on the error that any sample-efficient weak learning algorithm for convex sets must incur. The tension between these two bounds can be shown to imply a lower bound on the number of halfspaces that any $(1/2 - \tau)$-approximator must use; see Section VI for the detailed argument. We remark that the proof of Theorem 7 is the first result we know of in which an information-theoretic sample complexity lower bound for learning is used to establish an *inapproximability* result.

b) Average-Case Lower Bounds via Gaussian Noise Sensitivity: One drawback of Theorem 7 is that it is non-constructive: not only does it not exhibit any particular convex set which is hard to approximate, it does not establish any particular criterion that implies that a convex set is hard to approximate. Section 8 of the full version of this paper [26] gives such a criterion. We show that any set with very high *Gaussian noise sensitivity* (at least $1/2 - 1/\mathrm{poly}(n)$) at very low noise rates (at most $1/\mathrm{poly}(n)$) requires many halfspaces to approximate to accuracy even $1/2 + 1/\mathrm{poly}(n)$:

Theorem 8 (Theorem 49 of [26])**.** If $K \subset \mathbb{R}^n$ is any convex set which satisfies $\mathbf{GNS}_{n^{-c}}(K) \ge \frac{1}{2} - n^{-c}$, then $\mathrm{FC}(K, \frac{1}{2} - n^{-c}) \ge 2^{n^{\Omega(1)}}$, where $c > 0$ is a suitable absolute constant.

Theorem 8 naturally raises the question of whether there in fact exist n-dimensional convex sets that have Gaussian noise sensitivity as high as $1/2 - 1/\mathrm{poly}(n)$ at noise rates that are as low as $1/\mathrm{poly}(n)$. While natural, this question does not appear [60] to have been asked or answered prior to the current work. In Section 8.5 of [26], we combine the information-theoretic lower bounds of [28] on weak learning convex sets with the "low-degree algorithm" for learning bounded functions to show that there do exist n-dimensional convex sets K with $\mathbf{GNS}_{1/\mathrm{poly}(n)}(K) \ge 1/2 - 1/\mathrm{poly}(n)$. We believe that this result may be of independent interest.

c) Lower Bounds via Convex Influences: Neither of the two lower bound techniques discussed thus far yield lower bounds for any explicit convex sets that we know of. For our final lower bound, we use the recently introduced notion of convex influences [24], [25] to prove lower bounds for the ℓ_1 and ℓ_2 balls. In more detail, in Section VII we give a general criterion on a convex set $K \subset \mathbb{R}^n$ which suffices to ensure that $\mathrm{FC}(K,\varepsilon) = 2^{\Omega(\sqrt{n})}$, where ε is a suitable constant. This criterion is that K is symmetric[3] and has maximal *convex influence*, up to a multiplicative constant factor. Since both the ℓ_1 ball of Gaussian volume $1/2$ and the ℓ_2 ball of volume $1/2$ satisfy this condition, we get a $2^{\Omega(\sqrt{n})}$ lower bound for each of these explicit sets.

The notion of convex influence was recently introduced in the paper [25] as an analogue of the classical notion of influence from Boolean function analysis over the discrete cube [38], [59]. In particular, for any unit vector v and convex set K, $\mathbf{I}_v[K]$ is meant to capture how K varies just in the direction v (while averaging out every other direction) — see Definition 34 for the precise formulation. An attractive feature of this notion of influence is that for any orthonormal basis of $\mathbb{R}^n$, say $\{v_1, \ldots, v_n\}$, the sum $\sum_{i=1}^n \mathbf{I}_{v_i}[K]$ is independent of the choice of the basis and depends only on the set K. This leads us to the notion of total influence of a set K, denoted by $\mathbf{I}[K] = \sum_{i=1}^n \mathbf{I}_{v_i}[K]$ (where $\{v_1, \ldots, v_n\}$ is any orthonormal basis of $\mathbb{R}^n$).

[25] proved several structural properties of both influence and total influence. In particular, Proposition 19 of [25] established that the total influence of any convex set K is $O(\sqrt{n})$; furthermore, this bound is tight, as exhibited by $K = B(\sqrt{n})$, which is shown in [25] to be the convex set of maximal total influence.

In this paper, we show that any symmetric set K with maximal total influence (up to a constant factor) requires any approximator to have high facet complexity. In particular, we prove the following theorem:

Theorem 9 (Informal version of Theorem 39)**.** If K is symmetric and $\mathbf{I}[K] = \Omega(\sqrt{n})$, then for a suitable constant $\varepsilon > 0$ we have that $\mathrm{FC}(K,\varepsilon) = 2^{\Omega(\sqrt{n})}$.

The main idea in the proof of the above theorem is a new structural result establishing that if L is a convex set which is an intersection of s halfspaces, then $\mathbf{I}[L] = O(\log s)$ (Proposition 38). Coupled with the fact that $\mathbf{I}[K]$ can be expressed in terms of the degree-2 Hermite coefficients of K, a fairly simple argument using Parseval's identity and Cantelli's inequality leads to Theorem 9.

Once again, we observe that there is an analogy here with the Boolean setting. Namely, Boppana [11] showed that for any s-clause CNF formula, its total influence (as defined over the Boolean cube) is bounded by $O(\log s)$. We note that despite the syntactic analogy with Proposition 38, the underlying techniques are completely different; while Proposition 38 is an inherently geometric argument, the proof of Boppana relies on the method of random restrictions [36].

C. Discussion

The current paper takes the first steps in studying polytopal approximations of convex sets under Gaussian measure. Several tantalizing questions emerge from our work; we highlight a few of these now.

Perhaps the most natural question is to close the gap between the worst case upper and lower bounds on $\mathrm{FC}(K,\varepsilon)$ in the constant-ε regime. In particular, both Theorem 39 and Theorem 28 guarantee the existence of convex sets K such that $\mathrm{FC}(K,\varepsilon) = 2^{\Omega(\sqrt{n})}$ for some constant $\varepsilon > 0$. On the other hand, Theorems 18 and 19 show that for any constant $\varepsilon > 0$, $\mathrm{FC}(K,\varepsilon) = n^{O(n)}$. Can we close this gap?

In a related vein, it would also be interesting to obtain tighter upper and lower bounds on $\mathrm{FC}(K,\varepsilon)$ when K is the ℓ_p ball of Gaussian volume $1/2$. In particular, for $p > 2$, we do not have an upper bound on the facet complexity (when the error ε is a constant) that is better than $n^{O(n)}$. Is it possible to obtain polytopal approximations for the ℓ_p ball with $2^{n^{1-c}}$ facets for constant $c > 0$ in the constant error regime?

Finally, looking ahead, it would be interesting to study Gaussian-distance polytopal approximation of convex sets via approximators with small *vertex complexity*. Very little seems to be known here and indeed, the vertex complexity analogues of many questions considered in the current paper seem to be wide open. For example, Section V gets a $2^{O(\sqrt{n})}$-facet approximator for the ℓ_2 ball. Can we obtain a similar approximator for the ℓ_2 ball which has subexponential vertex complexity?

III. Preliminaries

Section III-A recalls basic background and sets up notation, Section III-B gives standard facts about the Gaussian and chi-squared distributions, Section III-C recalls known bounds on Gaussian surface area, and finally Section III-D defines the various distance metrics between convex sets that we will use.

A. Basic Notation and Terminology

We use boldfaced letters such as $\boldsymbol{x}, \boldsymbol{f}, \boldsymbol{A}$, etc. to denote random variables (which may be real-valued, vector-valued, function-valued, or set-valued; the intended type will be clear from the context). We write $\boldsymbol{x} \sim \mathcal{D}$ to indicate that the random variable $\boldsymbol{x}$ is distributed according to probability distribution $\mathcal{D}$. We will frequently identify a set $K \subseteq \mathbb{R}^n$ with its $0/1$-valued indicator function.

Notation 10 (Multiplicative approximation)**.** We use the following notation to denote that two nonzero reals a, b are multiplicatively close: For $\nu \geq 0$,

$$a \approx_\nu b \quad \iff \quad e^{-\nu} \leq a/b \leq e^{\nu}$$

(note that this condition is symmetric in a and b).

a) Geometry.: We write $e_i \in \mathbb{R}^n$ to denote the i^{th} standard basis vector. For $r > 0$, we write $\mathbb{S}^{n-1}(r)$ to denote the origin-centered sphere of radius r in $\mathbb{R}^n$ and $B(r)$ to denote the origin-centered ball of radius r in $\mathbb{R}^n$, i.e.,

$$\mathbb{S}^{n-1}(r) = \{x \in \mathbb{R}^n : \|x\| = r\}$$

and

$$B(r) = \{x \in \mathbb{R}^n : \|x\| \leq r\},$$

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

[3]Recall that a set $K \subseteq \mathbb{R}^n$ is *symmetric* if $x \in K$ implies $-x \in K$.

b) Convex Sets and Convex Bodies.: A set $K \subseteq \mathbb{R}^n$ is *convex* if $x, y \in K$ implies $\alpha x + (1-\alpha)y \in K$ for all $\alpha \in [0,1]$. We recall (see e.g. [48]) that all convex sets are Lebesgue measurable.

A convex *body* in $\mathbb{R}^n$ is a compact convex set with non-empty interior. Our results will hold for general convex sets (not only bodies), but since we are working with Gaussian distance, it is easy to see that it suffices to consider only convex bodies. (If a convex set has empty interior then its Gaussian volume is 0; the Gaussian volume of the closure of a convex set is the same as the Gaussian volume of the set; and if a convex set A is unbounded, we can "truncate" it by intersecting it with a sufficiently large ball to obtain a bounded convex set $A' \subset A$ with $\mathrm{dist}_{\mathrm{G}}(A, A') \leq \varepsilon/2$.)

Finally, for sets $K, L \subseteq \mathbb{R}^n$, we write $K + L$ to denote the Minkowski sum $\{x+y : x \in K \text{ and } y \in L\}$. For a set $K \subseteq \mathbb{R}^n$ and $r > 0$ we write rK to denote the set $\{rx : x \in K\}$.

c) Intersections of Halfspaces and Approximation.: For the sake of readability we will mostly state our results in a self-contained way, but the following notation will sometimes be useful:

- We write $\mathrm{Facet}(n, M)$ to denote the class of all convex sets which are intersections of M halfspaces in $\mathbb{R}^n$, i.e. the class of all M-facet convex sets.
- For a convex set $K \subseteq \mathbb{R}^n$ and a value $0 < \varepsilon < 1$, we write $\mathrm{FC}(K, \varepsilon)$ to denote the minimum value M such that there is some intersection of M halfspaces $L \in \mathrm{Facet}(n, M)$ such that $\mathrm{dist}_{\mathrm{G}}(K, L) \leq \varepsilon$, i.e. $\mathrm{FC}(K, \varepsilon)$ is the minimum "facet complexity" of any ε-approximator of K.

B. The Gaussian and Chi-Squared Distributions

We write $N(0, I_n)$ to denote the n-dimensional standard Gaussian distribution, and denote its density function by φ_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 φ instead of φ_n. We write $\mathrm{Vol}(K)$ to denote the Gaussian measure of a (Lebesgue measurable) set $K \subseteq \mathbb{R}^n$, that is

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

We recall the following standard tail bound on Gaussian random variables:

Proposition 11 (Theorem 1.2.6 of [31] or Equation 2.58 of [74])**.** Suppose $\boldsymbol{g} \sim N(0,1)$ is a one-dimensional Gaussian random variable. Then for all $r > 0$,

$$\varphi_1(r)\left(\frac{1}{r} - \frac{1}{r^3}\right) \leq \Pr_{\boldsymbol{g} \sim N(0,1)}[\boldsymbol{g} \geq r] \leq \varphi_1(r)\left(\frac{1}{r} - \frac{1}{r^3} + \frac{3}{r^5}\right)$$

where φ_1 is the one-dimensional Gaussian density.

We will also make use of the Berry-Esseen central limit theorem [7], [32] (alternatively, see Section 11.5 of [59]):

Theorem 12. Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent random variables with $\mathbf{E}[\boldsymbol{X}_i] = 0$ and $\mathbf{Var}[\boldsymbol{X}_i] = \sigma_i^2$, and assume $\sum_{i=1}^n \sigma_i^2 = 1$. Let $\boldsymbol{S} = \sum_{i=1}^n \boldsymbol{X}_i$ and let $\boldsymbol{Z} \sim N(0,1)$ be a standard univariate Gaussian. Then for all $u \in \mathbb{R}$,

$$\big|\mathbf{Pr}[\boldsymbol{S} \leq u] - \mathbf{Pr}[\boldsymbol{Z} \leq u]\big| \leq c\gamma$$

where

$$\gamma = \sum_{i=1}^n \mathbf{E}\left[|\boldsymbol{X}_i|^3\right]$$

and $c \leq 0.56$ is a universal constant.

Recall that the norm of an n-dimensional Gaussian random vector is distributed according to the chi-squared distribution with n degrees of freedom, i.e. if $\boldsymbol{x} \sim N(0, I_n)$ then $\|\boldsymbol{x}\|^2 \sim \chi^2(n)$. It is well known (see e.g. [75]) that the mean of the $\chi^2(n)$ distribution is n, the median is $n\big(1 - \Theta(1/n)\big)$, and for $n \geq 2$ the pdf is everywhere at most 1. We note that an easy consequence of these facts is that the origin-centered ball $B(\sqrt{n})$ of radius $\sqrt{n}$ in $\mathbb{R}^n$ has $\mathrm{Vol}(B(\sqrt{n})) = 1/2 + o_n(1)$.

We will require the following tail bound on $\chi^2(n)$ random variables:

Proposition 13 (Section 4.1 of [51])**.** Suppose $\boldsymbol{y} \sim \chi^2(n)$. Then for any $t > 0$, we have

$$\Pr_{\boldsymbol{y} \sim \chi^2(n)}\left[\boldsymbol{y} \geq n + 2\sqrt{nt} + 2t\right] \leq \exp(-t),$$
$$\Pr_{\boldsymbol{y} \sim \chi^2(n)}\left[\boldsymbol{y} \leq n - 2\sqrt{nt}\right] \leq \exp(-t).$$

C. Bounds on Gaussian Surface Area

Given a measurable set $K \subseteq \mathbb{R}^n$, recall that the *Gaussian surface area of* K—which we will denote as $\mathbf{GSA}(K)$—is given by

$$\mathbf{GSA}(K) := \lim_{\delta \to 0} \frac{\mathrm{Vol}\big(K + B(\delta)\big) - \mathrm{Vol}(K)}{\delta}.$$

For convex sets (more generally, for sets which are sufficiently regular, e.g. have smooth boundary except at a set of measure zero), we have

$$\mathbf{GSA}(K) = \int_{\partial K} \varphi(x)\, d\sigma(x) \tag{2}$$

See [57] or Definition 2 of [45] for further discussion on this point. The following upper bound on the Gaussian surface area of an intersection of halfspaces was obtained by Nazarov; see [42], [45].

Theorem 14 (Nazarov's bound)**.** Suppose $K \subseteq \mathbb{R}^n$ is an intersection of s halfspaces. Then we have

$$\mathbf{GSA}(K) \leq \sqrt{2 \ln s} + 2.$$

We will also require the following upper bound on the Gaussian surface area of an arbitrary convex set in $\mathbb{R}^n$, which was obtained by Ball [3]:

Theorem 15 (Ball's theorem)**.** Suppose $K \subseteq \mathbb{R}^n$ is a convex set. Then we have

$$\mathbf{GSA}(K) \leq O(n^{1/4}).$$

Finally, we recall the Gaussian isoperimetric inequality [12], [67]:

Theorem 16. Suppose $K \subseteq \mathbb{R}^n$. Let $H \subseteq \mathbb{R}^n$ be a halfspace (i.e. $H = \{x \in \mathbb{R}^n : \langle x, v\rangle \le \theta\}$ for $v \in \mathbb{S}^{n-1}$ and $\theta \in \mathbb{R}$) such that $\mathrm{Vol}(H) = \mathrm{Vol}(K)$. Then

$$\mathbf{GSA}(K) \ge \mathbf{GSA}(H) = \varphi_1(\theta)$$

where φ_1 denotes the univariate Gaussian p.d.f.

D. Distance Metrics Between Sets

The primary distance metric we will use throughout this paper is the following: Given two measurable sets $K, L \subseteq \mathbb{R}^n$, we define

$$\mathrm{dist}_{\mathrm{G}}(K, L) := \Pr_{\boldsymbol{x}\sim N(0,I_n)}\left[K(\boldsymbol{x}) \neq L(\boldsymbol{x}).\right]$$

In other words, $\mathrm{dist}_{\mathrm{G}}(K, L) = \mathrm{Vol}(K \triangle L)$, i.e. the Gaussian measure of the symmetric difference of the sets K and L.

We also recall the *Hausdorff distance* between two sets $K, L \subseteq \mathbb{R}^n$, which is defined as

$$\mathrm{dist}_{\mathrm{H}}(K, L) := \max\left\{\sup_{x\in K}\inf_{y\in L} d(x, y), \sup_{y\in L}\inf_{x\in K} d(x, y)\right\}$$

where $d(x, y) := \|x - y\|_2$ denotes the usual Euclidean ℓ_2 distance between the points x and y. We will rely on bounds on polytope approximation under the Hausdorff distance in Section IV-A.

IV. Generic Upper Bounds

We first give upper bounds on the facet complexity of arbitrary convex sets in $\mathbb{R}^n$ under the Gaussian distance metric.

A. Warmup: Universal Approximation via Hausdorff Distance Approximation

The following upper bound on the facet complexity of polytope approximators under the *Hausdorff distance* (cf. Section III-D) was independently obtained by Dudley [30] and by Bronstein and Ivanov [14]; see also Section 4.1 of [15].

Theorem 17. Suppose $K \subseteq \mathbb{R}^n$ is a compact convex set with non-empty interior that is contained in the unit ball in $\mathbb{R}^n$, i.e.

$$K \subseteq B(1).$$

For $0 < \varepsilon < 10^{-3}$, there exists a convex body L which is the intersection of

$$3\sqrt{n}\left(\frac{9}{\varepsilon}\right)^{(n-1)/2} \quad \text{halfspaces}$$

such that $K \subseteq L$ and $\mathrm{dist}_{\mathrm{H}}(K, L) \le \varepsilon$.

We remark that the original results due to [14], [30] consider the *vertex complexity* (i.e. number of vertices) of approximators instead of the facet complexity and furthermore only consider inner approximators to K (i.e. $L \subseteq K$ in the above); their arguments, however, can be easily modified to obtain the above [1]. Finally, we note that the bound in Theorem 17 is known to be close to tight: any outer (or inner) approximation to the unit ball $B(1)$ in $\mathbb{R}^n$ requires $\Omega(1/\varepsilon^{(n-1)/2})$ halfspaces [2], [15].

Theorem 17 and Ball's universal bound on the Gaussian surface area of any convex set (Theorem 15) imply the following upper bound for approximating generic convex sets in $\mathbb{R}^n$ under the Gaussian distance:

Theorem 18 (Universal approximation of convex sets.)**.** Let $0 < \varepsilon < 10^{-3}$. For every convex set $K \subseteq \mathbb{R}^n$, there exists a convex set L which is an intersection of

$$O\left(\frac{n^{5/4} + 2n^{3/4}\sqrt{\ln(2/\varepsilon)}}{\varepsilon}\right)^{(n-1)/2} \quad \text{halfspaces}$$

such that $\mathrm{dist}_{\mathrm{G}}(K, L) \le \varepsilon$, i.e. we have $\mathrm{FC}(K, \varepsilon) \le O\left(\frac{n^{5/4}+2n^{3/4}\sqrt{\ln(2/\varepsilon)}}{\varepsilon}\right)^{(n-1)/2}$.

Proof. For the rest of the argument, set

$$r^* := \sqrt{n} + 2\sqrt{\ln\left(\frac{2}{\varepsilon}\right)} \qquad \text{and} \qquad \varepsilon' := \Theta\left(\frac{\varepsilon}{n^{1/4}}\right).$$

Recalling that the squared norm of a Gaussian vector is distributed according to a $\chi^2(n)$ distribution, we have by Proposition 13 that

$$\begin{aligned}
&\Pr_{\boldsymbol{x}\sim N(0,I_n)}\left[\|\boldsymbol{x}\|^2 > {r^*}^2\right] \\
\le &\Pr_{\boldsymbol{y}\sim\chi^2(n)}\left[\boldsymbol{y}^2 \ge n + 2\sqrt{n\ln\left(\frac{2}{\varepsilon}\right)} + 2\ln\left(\frac{2}{\varepsilon}\right)\right] \\
\le &\frac{\varepsilon}{2}.
\end{aligned} \tag{3}$$

It thus suffices to $(\varepsilon/2)$-approximate $K' := K \cap B(r^*)$ under the Gaussian distance metric.

It is easy to see by definition of the Hausdorff distance that

$$\mathrm{dist}_{\mathrm{H}}(K, L) \le \delta \quad \text{is equivalent to} \quad \mathrm{dist}_{\mathrm{H}}(tK, tL) \le t\delta.$$

In particular, consider the convex set $(1/r^*)K' \subseteq B(1)$, and let $L \subseteq \mathbb{R}^n$ be the polytope guaranteed to exist by Theorem 17 such that $K \subseteq L$ and

$$\mathrm{dist}_{\mathrm{H}}\left(\frac{1}{r^*}K', \frac{1}{r^*}L\right) \le \frac{\varepsilon'}{2r^*},$$

which implies that $\mathrm{dist}_{\mathrm{H}}(K', L) \le \frac{\varepsilon'}{2}$. Recalling our choice of ε', note that Theorem 17 guarantees that L is an intersection of

$$O\left(\frac{18\sqrt{n}r^*}{\varepsilon'}\right)^{(n-1)/2} = O\left(\frac{n^{5/4} + 2n^{3/4}\sqrt{\ln(2/\varepsilon)}}{\varepsilon}\right)^{(n-1)/2}$$

halfspaces. We will show that L is in fact an $(\varepsilon/2)$-approximator for K' under the Gaussian distance, which would complete the proof as

$$\mathrm{dist}_{\mathrm{G}}(K, L) \le \mathrm{dist}_{\mathrm{G}}(K, K') + \mathrm{dist}_{\mathrm{G}}(K', L) \le \varepsilon. \tag{4}$$

Indeed, by definition of the Hausdorff distance we have that $K' \subseteq L \subseteq K'_{\varepsilon'}$

$$\text{where } K'_{\varepsilon'} := \left\{ x \in \mathbb{R}^n : \inf_{y \in K} d(x,y) \leq \varepsilon' \right\}. \tag{5}$$

Consequently, the Gaussian distance between the sets K' and L is

$$\begin{aligned}
\mathrm{dist}_{\mathrm{G}}(K', L) &= \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0, I_n)} \left[K'(\boldsymbol{x}) \neq L(\boldsymbol{x}) \right] \\
&= \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0, I_n)} \left[\boldsymbol{x} \in L \setminus K' \right] \\
&\leq \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0, I_n)} \left[\boldsymbol{x} \in K'_{\varepsilon'} \setminus K' \right] \\
&= \int_{t=0}^{\varepsilon'} \int_{x \in \partial K_t} \varphi(x) \, dx \, dt \\
&= \int_{t=0}^{\varepsilon'} \mathbf{GSA}(K_t) \, dt \\
&\leq \varepsilon' \cdot O(n^{1/4}) \qquad (6) \\
&= \frac{\varepsilon}{2}. \qquad (7)
\end{aligned}$$

Here, K_t is defined as in Equation (5); it is easy to see that if K is a convex, then so is K_t for all $t \geq 0$. Note that we used Keith Ball's bound on the Gaussian surface area of convex sets (Theorem 15) to obtain Equation (6). The theorem now follows from Equations (4) and (7). $\square$

B. A Relative-Error Sharpening of Theorem 18

We now establish a sharpening of Theorem 18 that gives a much better bound in the regime where K has volume very close to 1:

Theorem 19. Let $0 < \delta < 1$. Suppose $K \subseteq \mathbb{R}^n$ is a closed convex set with

$$\mathrm{Vol}(K) = 1 - \delta.$$

Then for all $\varepsilon > 0$, there exists a convex set L which is the intersection of

$$\frac{1}{\delta} \cdot \left(\frac{n}{\varepsilon} \log\left(\frac{1}{\delta} \right) \right)^{O(n)}$$

halfspaces such that $\mathrm{dist}_{\mathrm{G}}(K, L) \leq \varepsilon\delta$; i.e., $\mathrm{FC}(K, \varepsilon\delta) \leq \frac{1}{\delta} \cdot \left(\frac{n}{\varepsilon} \log\left(\frac{1}{\delta} \right) \right)^{O(n)}$.

We mention that in the most interesting regime, when δ is smaller than some absolute constant (which can be taken to be 1/10), our proof of Theorem 19 is self-contained and does not rely on the result of Bronstein–Ivanov (Theorem 17).

To contrast Theorem 19 with Theorem 18, note that Theorem 18 would give an $\varepsilon\delta$-approximation to K using

$$\left(\widetilde{O}(n^{5/4}/\varepsilon\delta) \right)^{(n-1)/2} \text{ halfspaces}$$

where the $\widetilde{O}$ hides a logarithmic factor in $\log(1/\varepsilon\delta)$. The crucial difference between Theorem 19 and Theorem 18 is that the dependence of Theorem 19 on δ is only $\frac{1}{\delta} \cdot \log(1/\delta)^{O(n)}$ rather than $(1/\delta)^{O(n)}$. This can make a major difference if δ is very small; for example, if $\delta = 1/2^{\sqrt{n}}$ and $\varepsilon = 0.01$ (observe that with these parameters, the desired approximator of K must correctly "lop off" at least 99% of the $1/2^{\sqrt{n}}$ amount of mass that lies outside of K), then Theorem 18 would only give a bound of $2^{\tilde{O}(n^{3/2})}$ halfspaces whereas Theorem 19 gives a bound of $2^{\tilde{O}(n)}$ halfspaces. Indeed, our $2^{\tilde{O}(n^{3/4})}$-halfspaces approximation of ℓ_p balls for $p \in (1, 2)$, given in the full version [26], will crucially rely on a savings of this sort that comes from the sharper parameters provided by Theorem 19.

1) Proof Overview: We set up some useful notation and give a high-level proof overview of Theorem 19. Throughout this section as well as the next, let $K \subseteq \mathbb{R}^n$ be a fixed convex set with volume $1 - \delta$ as in the statement of Theorem 19. We may assume without loss of generality that K is closed, since $\mathrm{Vol}(K)$ is the same as the volume of the closure of K for any convex set K.

We first remark that if $\delta \geq 1/10$ then the claimed bound is an immediate consequence of Theorem 18, so we henceforth suppose that $\delta < 1/10$.

As in the proof of Theorem 18, we may assume that K is contained in a ball of sufficiently large radius, namely

$$K \subseteq B(R) \text{ where } R := \sqrt{n} + 2\sqrt{\ln\left(\frac{2}{\varepsilon\delta} \right)},$$

so

$$\mathrm{Vol}(R) \geq 1 - \frac{\varepsilon\delta}{2},$$

and that the goal is to obtain an $(\varepsilon\delta/2)$-approximation to K.[4]

Definition 20 ("Length" function)**.** Given a vector $v \in \mathbb{S}^{n-1}$, we define the function $\ell : \mathbb{S}^{n-1} \to \mathbb{R}$ as

$$\ell(v) := \sup_{x \in K} \langle x, v \rangle.$$

In other words, $\ell(v)$ is the "length" of the set K along the direction v from the origin. We remark that since $\mathrm{Vol}(K) > 0.9 > 1/2$, the set K must contain the origin. We further remark that the function $\ell : \mathbb{S}^{n-1} \to \mathbb{R}$ is sometimes called the *Minkowski* or *support functional* of the set K [70].

Notation 21. We write $m(t)$ for

$$m(t) := \mathop{\mathbf{Pr}}_{\boldsymbol{y} \sim \chi^2(n)} \left[\boldsymbol{y} \geq t^2 \right].$$

Note that $m(t)$ is simply the tail mass of the $\chi^2(n)$ distribution beyond the point t^2. We can also view it as the probability that a Gaussian draw $\boldsymbol{x} \sim N(0, I_n)$ has $\|\boldsymbol{x}\|^2 \geq t$, even conditioned on $\boldsymbol{x}$ lying on any particular ray $\{rv : r \in \mathbb{R}_{\geq 0}\}$ for any fixed $v \in \mathbb{S}^{n-1}$. See Figure 1 for a schematic of the setup thus far.

From Definition 20, Notation 21, and the definition of the set K in Theorem 19, we have that

$$\mathop{\mathbf{E}}_{\boldsymbol{v} \sim \mathbb{S}^{n-1}} \left[m\big(\ell(\boldsymbol{v})\big) \right] = \delta$$

[4] In particular, we may intersect K with $B(R)$ to obtain a convex set K' and aim for an $(\varepsilon\delta/2)$-approximation to K' as in Section IV-A; for notational simplicity, however, we do not introduce the new set K' in this section.

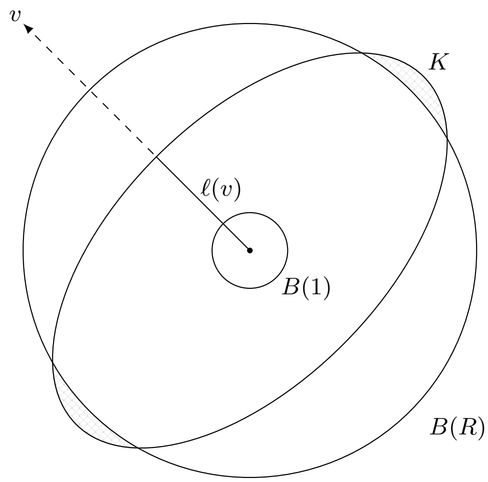

Fig. 1. Setup for the proof of Theorem 19. The length of the solid line is $\ell(v)$, and the conditional Gaussian mass on the dashed ray (conditioned on the Gaussian lying on the ray from the origin in direction v) corresponds to $m(\ell(v))$. The cross-hatched region corresponds to the (at most) $\varepsilon\delta/2$ approximation error incurred by intersecting K with $B(R)$.

since $\mathrm{Vol}(K) = 1 - \delta$. Note that this uses the convexity of the set K and the fact that K contains the origin.

Finally, since $\mathrm{Vol}(K) = 1 - \delta \geq 0.9$, we have that $B(1) \subseteq K$. To see this, note that if there exists $x \in B(1)$ such that $x \notin K$, then the separating hyperplane through x for K "lops off" at least $1 - \Phi(1) > 0.1$ of the Gaussian mass, and hence $\mathrm{Vol}(K) < 0.9$, which is a contradiction. We record this fact for convenience:

Fact 22. We have that $B(1) \subseteq K$.

We now turn to a high-level description of the approximator L. As in the proof of Theorem 18, L will be an outer approximator to the set K (i.e. $K \subseteq L$). We will construct L by placing tangent hyperplanes to ∂K that "lop off" at least a $(1 - \varepsilon)$-fraction of the total δ amount of Gaussian mass that lies outside the set K. In more detail, call a direction $v \in \mathbb{S}^{n-1}$ "good" if $m(\ell(v)) \leq (\varepsilon\delta/4)$, and "bad" otherwise.

- For all "good" directions, we will not worry about lopping off mass. Indeed, even if we failed to lop any probability mass off in all such directions, the total contribution to our overall error would be at most $(\varepsilon\delta/4)$.
- In order to handle "bad" directions, we first bucket them according to the value of $m(\ell(v))$ (which, as stated earlier, is proportional to the amount of Gaussian mass along v outside of K). We will obtain an appropriate covering of the points in each bucket, and we will place tangent hyperplanes to K at the points in our covering. A geometric argument (which will make use of the fact that $B(1) \subseteq K$), together with suitable tail bounds, will then give that the total error incurred from all "bad" directions is at most $(\varepsilon\delta/4)$.

Putting everything together, we will obtain an $(\varepsilon\delta/2)$-approximation to the set K; since we assumed $K \subseteq B(R)$, we already incurred $(\varepsilon\delta/2)$-error, which overall gives

$$\mathrm{dist}_{\mathrm{G}}(K, L) \leq \varepsilon\delta.$$

2) Proof of Theorem 19: As mentioned earlier, we may suppose that $\delta \leq 1/10$. We start by describing the bucketing alluded to in the proof overview.

Definition 23 (Bucketing $\mathbb{S}^{n-1}$)**.** Let $\tau > 0$ be a parameter that we will fix later. For $k \in \mathbb{N}$, we define

$$\mathrm{buck}(k) := \left\{ v \in \mathbb{S}^{n-1} : m(\ell(v)) \in \Big((1+\tau)^{-k}, (1+\tau)^{-k+1} \Big] \right\}.$$

We will fix τ (which will be a small quantity between 0 and 1) later in the course of the proof. Informally, $\mathrm{buck}(k)$ is the set of directions for which the probability that a Gaussian vector in that direction lies outside K is roughly $(1+\tau)^{-k}$.

For large enough k, the amount of mass outside K along the directions in $\mathrm{buck}(K)$ will be small, and so intuitively we should be able to ignore such directions. (We called such directions "good" in the proof overview.) To make this precise, set a parameter k^* to be

$$k^* := \frac{1}{\tau} \ln\left(\frac{4}{\varepsilon\delta}\right).$$

We then have for all $v \in \mathrm{buck}(k)$ with $k > k^*$ that

$$m(\ell(v)) \leq (1+\tau)^{-k+1} \leq \exp(-\tau k^*) \leq \frac{\varepsilon\delta}{4}.$$

As discussed previously, we will only have to worry about the "bad" directions corresponding to

$$\bigsqcup_{k=1}^{k^*} \mathrm{buck}(k) \tag{8}$$

and ignore the "good" directions; this will incur at most $\varepsilon\delta/4$ error. We will use the following lemma to separately handle each bucket $\mathrm{buck}(k)$ for $k \leq k^*$.

Lemma 24. Suppose $A \subseteq \mathbb{S}^{n-1}$ and $\theta \in (0, \pi/2)$. There exists a set $S_A(\theta) \subseteq A$ consisting of at most $(12/\theta)^n$ points such that for all $x \in A$, there exists a point $y \in S_A(\theta)$ such that

$$\angle(x, y) = \arccos\langle x, y\rangle \leq \theta.$$

Proof. Note that $\angle(x, y) \leq \theta$ for $x, y \in \mathbb{S}^{n-1}$ if and only if $\|x - y\| = 2\sin(\theta/2)$. It thus suffices to construct a $2\sin(\theta/2)$-net (cf. Section 4.2 of [73]) of A. From Exercise 4.2.10 of [73], the size of such a net is upper bounded by the size of a $\sin(\theta/2)$-net for $\mathbb{S}^{n-1}$, which in turn is at most

$$\left(\frac{3}{\sin(\theta/2)}\right)^n.$$

This bound is standard; see, for example, Corollary 4.2.13 of [73]. Finally, since $\sin(\cdot)$ is concave on the interval $[0, \pi/2]$, we have that

$$\sin\left(\frac{\theta}{2}\right) \geq \frac{\theta}{4},$$

and so we have that the desired net consists of at most

$$\left(\frac{12}{\theta}\right)^n \text{ points,}$$

which completes the proof. $\square$

We will construct the approximator L by placing tangent hyperplanes to K at the points in S where we define

$$S := \bigsqcup_{k=1}^{k^*} S_{\mathrm{buck}(k)}(\theta^*)$$

where $S_{\mathrm{buck}(k)}(\theta^*)$ is as in Lemma 24 for an appropriate choice of θ^*. It follows that the total number of halfspaces in our approximator L is at most

$$|S| \le k^* \cdot \left(\frac{12}{\theta^*}\right)^n = \frac{1}{\tau}\log\left(\frac{4}{\varepsilon\delta}\right)\cdot\left(\frac{12}{\theta^*}\right)^n, \tag{9}$$

where τ and θ^* are parameters that we will set below.

a) Setting Parameters.: Setting the parameters τ and θ^* fixes our approximator L. We will take

$$\begin{aligned}\tau &= \frac{\varepsilon\delta}{8},\\ \theta^* &= \frac{\varepsilon^2}{64}\left(R\left(2+\frac{16}{\varepsilon}\right)\left(2R^2\left(2+\frac{16}{\varepsilon}\right)-\frac{n}{4}\right)\right)^{-1}.\end{aligned} \tag{10}$$

Our choices above are dictated by the error analysis below; we note that it follows from Equations (9) and (10) that L has the desired number of halfspaces. Before proceeding, we introduce the following notation:

Notation 25. For $k \in [k^*]$, we define

$$\ell(k)_{\max} := \sup_{v\in\mathrm{buck}(k)} \ell(v) \qquad \text{and} \qquad \ell(k)_{\min} := \inf_{v\in\mathrm{buck}(k)} \ell(v).$$

b) Error Analysis.: We will analyze the error in approximating K by L on a direction-by-direction basis; recall that we only need to worry about the "bad" directions (Equation (8)). Fix a "bad" direction $v \in \mathbb{S}^{n-1}$, and suppose $v \in \mathrm{buck}(k)$ (where $k \le k^*$). Consider the two-dimensional setup (corresponding to the two-dimensional plane $\mathrm{span}\{O, U, W\}$) as in Figure 2:

- Let O denote the origin 0^n;
- Let $V \in \partial K$ be the point of intersection of the ray $\{tv : t \ge 0\}$ and ∂K;
- Let $U \in \partial K$, $U \in S_{\mathrm{buck}(k)}(\theta^*)$ be a point in $\mathrm{buck}(k)$ such that $\theta := \angle(\vec{OL}, \vec{ON}) \le \theta^*$; and
- By construction, we put down a halfspace tangent to K at U; let W be the point of intersection of this halfspace with the ray $\{tv : t \ge 0\}$.

With this setup in hand, note that the error incurred along direction v is given by

$$\mathrm{error}(v) := m\big(\|V\|\big) - m\big(\|W\|\big).$$

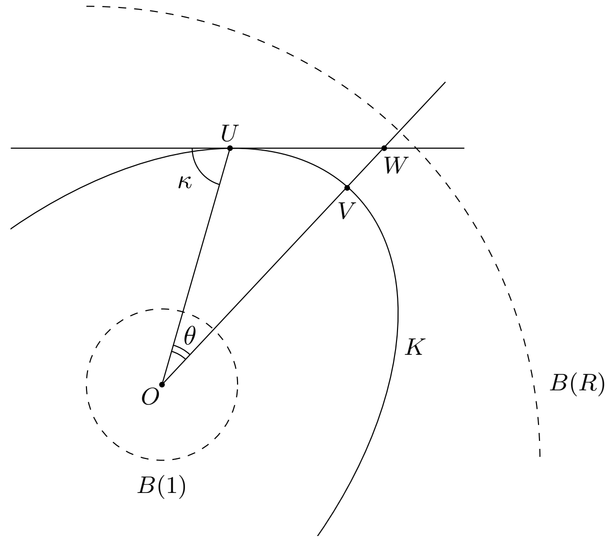

Fig. 2. A two-dimensional setup to analyze the error of our approximator. Here $V \in \mathrm{buck}(k)$ and $U \in S_{\mathrm{buck}(k)}(\theta^*)$ and so $\angle(\vec{OV}, \vec{OU}) = \theta \le \theta^*$

We will next establish that this quantity is at most $\varepsilon\delta/4$, which will imply that the total error incurred from all the "bad" directions is at most $\varepsilon\delta/4$. Note that

$$\begin{aligned}\mathrm{error}(v) &= m\big(\|V\|\big) - m\big(\|W\|\big)\\ &\le m\big(\ell(k)_{\min}\big) - m\big(\|W\|\big)\\ &= \Big(m\big(\ell(k)_{\min}\big) - m\big(\ell(k)_{\max}\big)\Big)\\ &\quad + \Big(m\big(\ell(k)_{\max}\big) - m\big(\|W\|\big)\Big).\end{aligned} \tag{11}$$

It suffices to show that Equation (11) is upper bounded by $\varepsilon\delta/4$. Indeed, recall that we incurred (a) $\varepsilon\delta/2$ error from "capping" the set K by intersecting it with $B(R)$; (b) $\varepsilon\delta/4$ error from the "bad" directions from $\mathrm{buck}(k)$ for $k > k^*$; and finally (c) $\varepsilon\delta/4$ error from all of the "good" directions from Equation (11) (as we will establish in the remainder of the argument); this will complete the proof of Theorem 19.

We will bound each term in Equation (11) separately. By definition of the buckets (Definition 23), we have that

$$m\big(\ell(k)_{\min}\big) - m\big(\ell(k)_{\max}\big) \le \tau(1+\tau)^{-k} \le \tau = \frac{\varepsilon\delta}{8}. \tag{12}$$

In order to bound the second term in Equation (11), recall that by Brahmagupta's formula (cf. Figure 2), we have

$$\frac{\sin(\pi-\kappa)}{\|W\|} = \frac{\sin(\kappa-\theta)}{\|U\|}.$$

This can be rewritten as

$$\|W\| = \frac{\sin\kappa}{\sin(\kappa-\theta)}\cdot\|U\| \le \frac{\sin\kappa}{\sin(\kappa-\theta)}\cdot\ell(k)_{\max}. \tag{13}$$

Recalling Fact 22, since $B(1) \subseteq K \subseteq B(R)$, we have that

$$\tan\kappa = \frac{\mathrm{dist}\Big(O, \overleftrightarrow{UW}\Big)}{\mathrm{dist}\Big(U, O \perp \overleftrightarrow{UW}\Big)} \ge \frac{1}{R}$$

where we write $\mathrm{dist}(O, \overleftrightarrow{UW})$ to mean the distance between the origin and the line $\overleftrightarrow{UW}$, which is guaranteed to be at least 1 thanks to the convexity of K and the fact that $B(1) \subseteq K$; and $\mathrm{dist}(U, O \perp \overleftrightarrow{UW})$ is the distance between U and the point on the line $\overleftrightarrow{UW}$ closest to O, which must be at most $\mathrm{dist}(O, U)$ (since the hypotenuse is the longest side in a right triangle), which in turn is at most R by out setup. (See Figure 2.) Returning to Equation (13), the above lower bound on $\tan\kappa$ implies that

$$\begin{aligned}\frac{\sin\kappa}{\sin(\kappa-\theta)} &= \frac{\sin\kappa}{\sin\kappa\cos\theta - \cos\kappa\sin\theta}\\ &= \frac{1}{\cos\theta - \frac{\sin\theta}{\tan\kappa}}\\ &\leq \frac{1}{\cos\theta - R\sin\theta}.\end{aligned}$$

For $\theta \geq 0$, recalling the standard trigonometric inequalities

$$\sin\theta \leq \theta \qquad \text{and} \qquad \cos\theta \geq 1 - \frac{\theta^2}{2},$$

we thus get that

$$\begin{aligned}\frac{\sin\kappa}{\sin(\kappa-\theta)} &\leq \frac{1}{\cos\theta - R\sin\theta}\\ &\leq \frac{1}{1 - R\theta - \frac{\theta^2}{2}}\\ &\leq 1 + 2R\theta \leq 1 + 2R\theta^*\end{aligned}$$

where the penultimate inequality can be readily verified from the definitions of R and θ, and the final inequality relies on the fact that $\theta \leq \theta^*$ since $U \in S_{\mathrm{buck}(k)}(\theta^*)$ and $V \in \mathrm{buck}(k)$.

Plugging the above bound back into Equation (13), we get that

$$\|W\| \leq \ell(k)_{\max}(1 + 2R\theta^*).$$

As m (cf. Notation 21) is a decreasing function, we have that

$$\begin{aligned}&m\big(\ell(k)_{\max}\big) - m\big(\|W\|\big) && (14)\\ &\leq m\big(\ell(k)_{\max}\big) - m\big(\ell(k)_{\max}(1+2R\theta^*)\big)\\ &= m\big(\ell(k)_{\max}\big)\left(1 - \frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big)}{m\big(\ell(k)_{\max}\big)}\right)\\ &\leq \delta\cdot\left(1 - \frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big)}{m\big(\ell(k)_{\max}\big)}\right) && (15)\end{aligned}$$

where Equation (15) relies on the observation that $m(\ell(v)) \leq \delta$ for any direction $v \in \mathbb{S}^{n-1}$. To see this, note that if $tv \in \partial K$ is such that $m(t) > \delta$, then the supporting hyperplane tangent to K at tv will "lop off" strictly greater than δ mass. Consequently, $\mathrm{Vol}(K) < 1 - \delta$, which is a contradiction.

Now, note that it suffices to show that

$$\frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big)}{m\big(\ell(k)_{\max}\big)} \geq 1 - \frac{\varepsilon}{8} \qquad (16)$$

in order to complete the proof. The remainder of the proof establishes Equation (16). We split into two cases depending on $\ell(k)_{\max}$:

c) Case 1: $\ell(k)_{\max} \geq \sqrt{n-1}$.: . Let $\Delta > 0$ be a parameter that we will set later. We have

$$\begin{aligned}&\frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big)}{m\big(\ell(k)_{\max}\big)}\\ &\geq \frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big) - m\big(\ell(k)_{\max}(1+(2+\Delta)R\theta^*)\big)}{m\big(\ell(k)_{\max}\big) - m\big(\ell(k)_{\max}(1+(2+\Delta)R\theta^*)\big)}.\end{aligned} \qquad (17)$$

Note that the right hand side of the above equation is lower bounded by

$$\begin{aligned}&\frac{\ell(k)_{\max}\Delta R\theta^*\cdot \mathrm{pdf}_{\chi(n)}\big(\ell(k)_{\max}(1+(2+\Delta)R\theta^*)\big)}{\ell(k)_{\max}(2+\Delta)R\theta^*\cdot \mathrm{pdf}_{\chi(n)}(\ell(k)_{\max})} && (18)\\ &= \left(\frac{\Delta}{2+\Delta}\right)\frac{\mathrm{pdf}_{\chi(n)}\big(\ell(k)_{\max}(1+(2+\Delta)R\theta^*)\big)}{\mathrm{pdf}_{\chi(n)}(\ell(k)_{\max})}\\ &\geq \left(\frac{\Delta}{2+\Delta}\right)\cdot\Upsilon(n,R,\theta^*) && (19)\end{aligned}$$

where (18) used the fact that the $\mathrm{pdf}_{\chi(n)}(x)$ is decreasing in x for $x \geq \sqrt{n-1}$, and

$$\Upsilon = \Big((1+(2+\Delta)R\theta^*)^{n-1}\exp(A)\Big), \text{ where}$$

$$A = \frac{-R^2}{2}\Big(2(2+\Delta)R\theta^* + (2+\Delta)^2R^2{\theta^*}^2\Big).$$

In particular, (19) relies on the formula for the density function of the $\chi(n)$-distribution, which is given by

$$\mathrm{pdf}_{\chi(n)}(x) = \begin{cases}\frac{1}{2^{n/2-1}\Gamma(n/2)}x^{n-1}\exp\left(-\frac{x^2}{2}\right) & \text{if } x \geq 0\\ 0 & \text{if } x < 0\end{cases},$$

and the fact that $\ell(k)_{\max} \leq R$. (This is because we truncated K by intersecting it with $B(R)$.)

We will take

$$\Delta := \frac{16}{\varepsilon},$$

as a consequence of which, Equation (16) follows from showing that

$$\Upsilon(n,R,\theta^*) \geq 1 - \frac{\varepsilon^2}{64}.$$

In particular, using the inequality $1 + x \leq e^x$, it suffices to show that

$$\Upsilon(n,R,\theta^*) \geq \exp\left(-\frac{\varepsilon^2}{64}\right).$$

We have that

$$\begin{aligned}\Upsilon(n,R,\theta^*) &\geq \left((1+(2+\Delta)R\theta^*)^{n-1}\exp\left(-2R^3\theta^*(2+\Delta)^2\right)\right)\\ &\geq \exp\left(\theta^*\left(\left(\frac{2+\Delta}{4}\right)nR - 2R^3(2+\Delta)^2\right)\right)\\ &= \exp\left(-\frac{\varepsilon^2}{64}\right),\end{aligned}$$

where the first inequality uses $R\theta^* \ll 1$, the second inequality uses $1 + x \geq e^{x/2}$ for $x \in [0, 1]$, and the third inequality uses the choice of θ^*, which we recall was

$$\theta^* := \frac{\varepsilon^2}{64}\left(R\left(2+\frac{16}{\varepsilon}\right)\left(2R^2\left(2+\frac{16}{\varepsilon}\right)-\frac{n}{4}\right)\right)^{-1}.$$

d) Case 2: $\ell(k)_{\max} < \sqrt{n-1}$.: First, note that $\mathrm{pdf}_{\chi(n)}(x)$ is at most 1 for all x. It follows that

$$\int_{\ell(k)_{\max}}^{\ell(k)_{\max}(1+2R\theta^*)} \mathrm{pdf}_{\chi(n)}(x)\,dx \leq 2\ell(k)_{\max}R\theta^* \leq 2R\theta^*\sqrt{n}. \tag{20}$$

Furthermore, since $\ell(k)_{\max} \leq \sqrt{n-1}$ which is the mode of the $\chi(n)$-distribution it follows that

$$m(\ell(k)_{\max}) \geq \frac{1}{2}.$$

This, together with Equation (20), lets us write

$$\begin{aligned}\frac{m\big(\ell(k)_{\max}(1+2R\theta^*)\big)}{m\big(\ell(k)_{\max}\big)} &\geq \frac{m\big(\ell(k)_{\max}\big)-2R\theta^*\sqrt{n}}{m\big(\ell(k)_{\max}\big)}\\ &\geq 1-4R\theta^*\sqrt{n}\\ &\geq 1-\frac{\varepsilon}{8},\end{aligned}$$

establishing Equation (16) (the final inequality above is straightforward to verify from our choice of θ^*).

Putting both cases together completes the proof of the theorem. □

V. Improved Approximation for the ℓ_2 Ball

For the Hausdorff and Lebesgue distance metrics (cf. Section III-D), the ℓ_2 ball is often an extremal example for known upper bounds on the vertex or facet complexity of polyhedral approximators [5], [13], [30], [52]. In this section, we show that—perhaps surprisingly—the $\sqrt{n}$-radius ℓ_2 ball $B(\sqrt{n})$, which has Gaussian volume $1/2 + o_n(1)$, can be approximated to within any constant error by an intersection of only $2^{O(\sqrt{n})}$ many halfspaces. This is a substantial improvement on the exponential-in-n approximation upper bounds obtained in Section IV for general convex bodies.

Throughout this section, we will write $B_2 := B(\sqrt{n})$ for ease of notation (the "2" subscript is because we are dealing with the ℓ_2 ball).

Theorem 26. Let $0 < \varepsilon < c$ for some sufficiently small absolute constant c. Then there exists a polytope K which is the intersection of

$$s = \exp\left(\Theta\left(\sqrt{n}\cdot\frac{1}{\varepsilon}\log\left(\frac{1}{\varepsilon}\right)\right)\right) \text{ halfspaces}$$

such that $\mathrm{dist}_{\mathrm{G}}(B_2, K) \leq \varepsilon$.

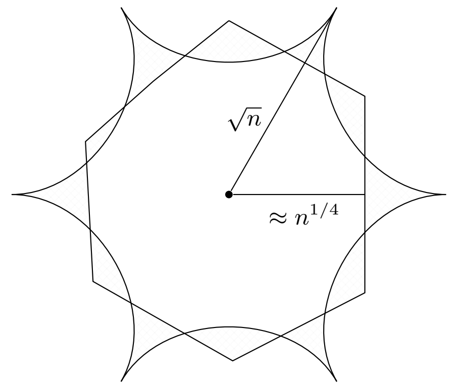

Fig. 3. A cartoon of how a polytope drawn from $\mathrm{Naz}(w, s)$, for suitable $s = 2^{\Theta(\sqrt{n})}$, $w \approx n^{3/4}$, approximates the radius-$\sqrt{n}$ ball in $\mathbb{R}^n$. Our depiction of $B_2 = B(\sqrt{n})$ is inspired by Milman's "hyperbolic" drawings of high-dimensional convex sets [56].

Theorem 26 is inspired by Theorem 2.1 of O'Donnell and Wimmer [62], which shows that the n-bit majority function $\mathrm{Maj}_n : \{0,1\}^n \to \{0,1\}$, defined as

$$\mathrm{Maj}_n(x) := \mathbf{1}\left\{\sum_{i=1}^n x_i \geq n/2\right\},$$

can be approximated by a monotone CNF formula of size $2^{O(\sqrt{n})}$. O'Donnell and Wimmer's construction is probabilistic and bears a close resemblance to Talagrand's random CNF [69]. Our approach for approximating B employs a modification of a probabilistic construction of a convex body due to Nazarov [57]. Looking ahead, in Section VII we will show that Theorem 26 is tight for constant ε; more precisely, we will show that any ε-approximation to B_2 must have $2^{\Omega(\sqrt{n})}$ facets for a suitable small constant $\varepsilon > 0$.

To prove Theorem 26, we begin by defining a suitable distribution over intersections of randomly chosen halfspaces:

Definition 27. For $w, s > 0$, we write $\mathrm{Naz}(w, s)$ to be the distribution over s-facet polytopes in $\mathbb{R}^n$ where draw from $\mathrm{Naz}(w, s)$ is obtained as follows:

1) For $i \in [s]$, draw i.i.d. $\boldsymbol{g}^{(i)} \sim N(0, I_n)$ and let $\boldsymbol{H}_i$ denote the halfspace
$$\boldsymbol{H}_i := \left\{x \in \mathbb{R}^n : \langle x, \boldsymbol{g}^{(i)}\rangle \leq w\right\}.$$
2) Output the convex set $\mathbf{K} := \bigcap_{i=1}^s \boldsymbol{H}_i$.

With Definition 27 in hand, we turn to the proof of Theorem 26:

Proof of Theorem 26. We will show that there exists an outcome K in the support of $\mathrm{Naz}(w, s)$ for an appropriate choice of parameters w and s that has the desired properties.

To show this, it suffices to show that

$$\mathop{\mathbf{E}}_{\mathbf{K}\sim\mathrm{Naz}(w,s)}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim N(0,I_n)}\big[\mathbf{K}(\boldsymbol{x}) \neq B_2(\boldsymbol{x})\big]\right] \leq \varepsilon,$$

which, by commuting the order of integration, is equivalent to showing that

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim N(0,I_n)}\left[\mathop{\mathbf{Pr}}_{\mathbf{K}\sim\mathrm{Naz}(w,s)}\left[\mathbf{K}(\boldsymbol{x})\neq B_2(\boldsymbol{x})\right]\right]\leq\varepsilon. \tag{21}$$

Equation (21) allows us to control the error on an "x-by-x basis." We set parameters

$$d_{\mathrm{in}}:=\sqrt{n}-\frac{\varepsilon}{4}\qquad\text{and}\qquad d_{\mathrm{out}}:=\sqrt{n}+\frac{\varepsilon}{4}. \tag{22}$$

We will show that when $\|x\|\leq d_{\mathrm{in}}$ or $\|x\|\geq d_{\mathrm{out}}$,

$$\mathop{\mathbf{Pr}}_{\mathbf{K}\sim\mathrm{Naz}(w,s)}\left[\mathbf{K}(x)\neq B_2(x)\right]\leq\frac{\varepsilon}{2}. \tag{23}$$

Since the Gaussian volume of the annulus $\{y\in\mathbb{R}^n : d_{\mathrm{in}}\leq\|y\|\leq d_{\mathrm{out}}\}$ is at most $\frac{\varepsilon}{2}$ (this is an easy consequence of the standard fact that for $n>1$, the pdf of the $\chi^2(n)$-distribution is everywhere at most 1), this establishes Equation (21) which in turn completes the proof.

Note that by construction (Definition 27), for any fixed $x\in\mathbb{R}^n$ we have that

$$\mathop{\mathbf{Pr}}_{\mathbf{K}\sim\mathrm{Naz}(w,s)}\left[x\in\mathbf{K}\right]=\Phi\left(\frac{w}{\|x\|}\right)^s$$

where $\Phi:\mathbb{R}\to[0,1]$ denotes the cumulative density function of the univariate Gaussian distribution $N(0,1)$. To see this, note that

$$\langle x,\boldsymbol{g}^{(j)}\rangle=\sum_{i=1}^n x_i\boldsymbol{g}_i^{(j)}\sim N(0,\|x\|^2)$$

and so

$$\mathop{\mathbf{Pr}}_{\boldsymbol{g}^{(1)},\dots,\boldsymbol{g}^{(s)}}\left[\langle x,\boldsymbol{g}^{(j)}\rangle\leq w\text{ for all }j\in[s]\right]=\Phi\left(\frac{w}{\|x\|}\right)^s$$

due to independence.

This observation informs our setting of the parameters w and s. We take w to satisfy the equation

$$\frac{1-\Phi(w/d_{\mathrm{out}})}{1-\Phi(w/d_{\mathrm{in}})}=\frac{4}{\varepsilon}\ln\left(\frac{4}{\varepsilon}\right) \tag{24}$$

and take s to be

$$s:=\frac{\varepsilon}{4}\left(1-\Phi\left(\frac{w}{d_{\mathrm{in}}}\right)\right)^{-1}. \tag{25}$$

(We will argue that there is a valid solution to Equation (24) later on in the proof.) Given Equations (24) and (25),

- For $x\in\mathbb{R}^n$ with $\|x\|\leq d_{\mathrm{in}}$, we have

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\mathbf{K}\sim\mathrm{Naz}(w,s)}\left[x\in\mathbf{K}\right]&=\Phi\left(\frac{w}{\|x\|}\right)^s\\&\geq\Phi\left(\frac{w}{d_{\mathrm{in}}}\right)^s\\&=\left(1-\frac{\varepsilon}{4s}\right)^s\\&\geq 1-\frac{\varepsilon}{4}\end{aligned}$$

  where the first inequality relies on the fact that $\Phi(\cdot)$ is an increasing function, the following equality relies on our choice of s in Equation (25), and the final inequality makes use of the fact that $(1-a)^b\geq 1-ab$. As $\|x\|\leq d_{\mathrm{in}}$ implies that $x\in B_2$, we pick up at most $\varepsilon/4$ error on such points.
- For $x\in\mathbb{R}^n$ with $\|x\|\geq d_{\mathrm{out}}$ and $\mathbf{K}\sim\mathrm{Naz}(w,s)$,

$$\begin{aligned}\mathbf{Pr}\left[x\in\mathbf{K}\right]&=\Phi\left(\frac{w}{\|x\|}\right)^s\\&\leq\Phi\left(\frac{w}{d_{\mathrm{out}}}\right)^s\\&=\left(1-\frac{4}{\varepsilon}\left(1-\Phi\left(\frac{w}{d_{\mathrm{in}}}\right)\right)\ln\left(\frac{4}{\varepsilon}\right)\right)^s\\&=\left(1-\frac{1}{s}\ln\left(\frac{4}{\varepsilon}\right)\right)^s\\&\leq\frac{\varepsilon}{4}\end{aligned}$$

  where we once again used the fact that $\Phi(\cdot)$ is increasing in the first inequality. The following equalities follow from rearranging Equations (24) and (25), and the final inequality is due to the fact that $(1+x)\leq\exp(x)$. As $x\geq d_{\mathrm{out}}$ implies that $x\notin B_2$, we pick up at most $\varepsilon/4$ error on such points as well.

In particular, the above establishes Equation (23), which in turn establishes Equation (21) and so there exists a set K in the support of $\mathrm{Naz}(w,s)$ that ε-approximates B_2.

It remains to argue that there exists a valid solution to Equation (24), and to bound the number of facets of K (i.e. the parameter s); we start with the former. Using standard Gaussian tail bounds (Proposition 11), provided that $w/d_{\mathrm{in}},w/d_{\mathrm{out}}=\Omega(1)$ (which holds with room to spare; below we will see that these quantities are $\Omega(n^{1/4})$), we can write Equation (24) as

$$\exp\left(\frac{w^2}{2}\left(\frac{1}{d_{\mathrm{in}}^2}-\frac{1}{d_{\mathrm{out}}^2}\right)\right)\cdot\Theta(1)=\frac{4}{\varepsilon}\ln\left(\frac{4}{\varepsilon}\right).$$

Taking logarithms on both sides, for ε at most some sufficiently small constant we get that

$$w^2=\Theta\left(\ln\left(\frac{4}{\varepsilon}\right)+\ln\ln\left(\frac{4}{\varepsilon}\right)\right)\cdot\frac{(d_{\mathrm{in}}d_{\mathrm{out}})^2}{d_{\mathrm{out}}^2-d_{\mathrm{in}}^2}.$$

Recalling our choices of d_{out} and d_{in}, we have

$$\frac{(d_{\mathrm{in}}d_{\mathrm{out}})^2}{d_{\mathrm{out}}^2-d_{\mathrm{in}}^2}=\frac{\left(n-\varepsilon^2/16\right)^2}{\varepsilon\sqrt{n}}=\Theta\left(\frac{n^{3/2}}{\varepsilon}\right).$$

Plugging this back into the previous expression and taking square roots on both sides gives

$$w:=\Theta\left(n^{3/4}\sqrt{\frac{1}{\varepsilon}\left(\ln\left(\frac{4}{\varepsilon}\right)+\ln\ln\left(\frac{4}{\varepsilon}\right)\right)}\right). \tag{26}$$

This lets us bound s, which is the number of facets of K. From Equation (25) and once again using standard tail bounds (Proposition 11), we have

$$s = \frac{\varepsilon}{4}\left(1 - \Phi\left(\frac{w}{d_{\text{in}}}\right)\right)^{-1} \leq \frac{\varepsilon}{4}\exp\left(\frac{w^2}{2d_{\text{in}}^2}\right)\left(\frac{d_{\text{in}}}{w} - \frac{d_{\text{in}}^3}{w^3}\right)$$
$$\leq \exp\left(\Theta\left(\frac{w^2}{d_{\text{in}}^2}\right)\right).$$

From Equations (22) and (26) we get that

$$\frac{w}{d_{\text{in}}} = \Theta\left(n^{1/4}\sqrt{\frac{1}{\varepsilon}\left(\log\left(\frac{4}{\varepsilon}\right) + \log\log\left(\frac{4}{\varepsilon}\right)\right)}\right),$$

from which the claimed bound on s is immediate. ☐

VI. Non-Explicit Average-Case Lower Bounds

The goal of this section is to prove the following theorem:

Theorem 28 (Non-explicit average-case lower bound.)**.** There is an absolute constant $C > 0$ such that for any sufficiently large n and any $\varepsilon = \omega(n^{-1/2} \cdot \log n)$, there exists some convex set $K \subset \mathbb{R}^n$ such that for any L which is an intersection of at most $M := 2^{C\cdot\varepsilon\cdot\sqrt{n}}$ halfspaces,

$$\mathrm{Vol}(K \triangle L) > \frac{1}{2} - \varepsilon.$$

We note that Theorem 28 gives an *average-case* lower bound, i.e. it establishes "strong inapproximability" by intersections of not-too-many halfspaces. For example, taking $\varepsilon = n^{-1/4}$, it shows that there is an n-dimensional convex set K such that no intersection of $2^{cn^{1/4}}$ many halfspaces can approximate K to accuracy even as large as $1/2 + n^{-1/4}$.

The proof of this theorem will go via the notion of VC-dimension, a fundamental measure of complexity in statistical learning theory. Recall that given a set $\mathcal{F}$ of Boolean functions over some domain X, the *VC-dimension of* $\mathcal{F}$ is the largest size of any $S \subseteq X$ which is *shattered* by $\mathcal{F}$, meaning that every possible $0/1$ labeling of the points in S is achieved by some function in $\mathcal{F}$. (See the book [44] for more details.)

To use VC-dimension in our context, we will also need the notion of agnostic learning. We recall the following standard definition:

Definition 29. A class $\mathcal{F}$ of Boolean functions over $\mathbb{R}^n$ is said to be *agnostically PAC learnable* with sample complexity $m(\varepsilon, \delta)$ if for every $\varepsilon, \delta > 0$, the following holds. There is an algorithm $\mathcal{A}$ which for any $f : \mathbb{R}^n \to \{0, 1\}$ and any distribution $\mathcal{D}$ over $\mathbb{R}^n$, given $m(\varepsilon, \delta)$ many i.i.d. labeled samples of the form $(\boldsymbol{x}, f(\boldsymbol{x}))$ (where each $\boldsymbol{x} \sim \mathcal{D}$), outputs a hypothesis $h : \mathbb{R}^n \to \{0, 1\}$ such that with probability $1 - \delta$,

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}}\big[h(\boldsymbol{x}) \neq f(\boldsymbol{x})\big] \leq \varepsilon + \min_{h^*\in\mathcal{F}} \Pr_{\boldsymbol{x}\sim\mathcal{D}}\big[h^*(\boldsymbol{x}) \neq f(\boldsymbol{x})\big].$$

A central result of statistical learning theory is that any concept class $\mathcal{F}$ of Boolean functions is agnostically PAC learnable where the sample complexity is proportional to the VC-dimension of $\mathcal{F}$. We state a sharp form of the bound below, which is due to Talagrand [68]:

Theorem 30. Any concept class $\mathcal{F}$ of Boolean functions over $\mathbb{R}^n$ is agnostically PAC learnable with sample complexity

$$m_{\mathcal{F}}(\varepsilon, \delta) := \Theta\bigg(\frac{\text{VC-dim}(\mathcal{F})}{\varepsilon^2} + \frac{\log(1/\delta)}{\varepsilon^2}\bigg),$$

where $\text{VC-dim}(\mathcal{F})$ is the VC-dimension of $\mathcal{F}$.

Recall that $\mathrm{Facet}(n, M)$ denotes the class of convex sets in $\mathbb{R}^n$ which are intersection of at most M halfspaces. The following is shown in [10] (and is now a standard fact, see e.g. Exercise 3.4 of [44]):

Fact 31. $\text{VC-dim}(\mathrm{Facet}(n, M)) = O(nM \log M)$.

The final ingredient we will require is a recent result of De and Servedio (Theorem 2 in [28]) which establishes a lower bound on the query complexity of any algorithm which "weakly learns" (meaning that the output hypothesis has an error rate only slightly less than $1/2$) convex sets over the Gaussian space. (This lower bound of course also holds for the sample complexity of any algorithm which only receives independent labeled samples $(\boldsymbol{x}, f(\boldsymbol{x}))$ where each $\boldsymbol{x} \sim N(0, I_n)$.)

Theorem 32 (Theorem 2 of [28])**.** For sufficiently large n, for any $s \geq n$, there is a distribution $\mathcal{D}$ over centrally symmetric convex sets $\mathbf{K} \subset \mathbb{R}^n$ with the following property: for a target convex set $\mathbf{K} \sim \mathcal{D}$, for any black box query algorithm A making at most s many queries to $\mathbf{K}$, the expected error of A (the probability over $\mathbf{K} \sim \mathcal{D}$, over any internal randomness of A, and over a random Gaussian $\boldsymbol{x} \sim N(0, 1^n)$, that the output hypothesis h of A predicts incorrectly on $\boldsymbol{x}$) is at least $1/2 - \frac{O(\log s)}{n^{1/2}}$.

We now combine these ingredients to establish Theorem 28:

Proof of Theorem 28. First, let us assume that for every convex set K, there is a convex set L which is an intersection of at most M halfspaces such that

$$\mathrm{Vol}(K \triangle L) \leq \frac{1}{2} - \varepsilon.$$

Now, consider the task of "weak learning" an unknown convex set $K \subseteq \mathbb{R}^n$ given i.i.d. labeled samples of the form $(\boldsymbol{x}, K(\boldsymbol{x}))$ where $\boldsymbol{x} \sim N(0, I_n)$ and $K(\cdot)$ is identified with the indicator function of the convex set K. To do this, we run the agnostic PAC learning algorithm from Theorem 30 for the class $\mathrm{Facet}(n, M)$ with parameters $\varepsilon/2$ and δ on the samples.

Now, given that there is a convex set $L \in \mathrm{Facet}(n, M)$ such that $\mathrm{Vol}(K \triangle L) \leq \frac{1}{2} - \varepsilon$, it follows that with probability $1 - \delta$, the algorithm from Theorem 30 outputs a hypothesis h such that

$$\Pr_{\boldsymbol{x}\sim N(0,I_n)}[h(\boldsymbol{x}) \neq K(\boldsymbol{x})] \leq \frac{1}{2} - \varepsilon + \varepsilon/2 = \frac{1}{2} - \varepsilon/2.$$

Further, the sample complexity of this algorithm is given by $S(\varepsilon, \delta)$ defined as

$$S(\varepsilon, \delta) = \Theta\bigg(\frac{\text{VC-dim}(\mathrm{Facet}(n, M))}{\varepsilon^2} + \frac{\log(1/\delta)}{\varepsilon^2}\bigg)$$

Applying Fact 31, we get that

$$S(\varepsilon,\delta) = O\left(\frac{nM\log M}{\varepsilon^2} + \frac{\log(1/\delta)}{\varepsilon^2}\right)$$

In particular, if we set $\delta = \varepsilon/4$, then we get that there is a PAC learning algorithm for convex sets (where the data $\boldsymbol{x} \sim N(0, I_n)$) with sample complexity

$$S(\varepsilon,\varepsilon/4) = O\left(\frac{nM\log M}{\varepsilon^2} + \frac{\log(1/\varepsilon)}{\varepsilon^2}\right)$$

which has expected error at most $1/2 - 3\varepsilon/4$ (where the expectation is over the internal randomness as well as the randomness of the data points $\boldsymbol{x} \sim N^n(0,1)$). On the other hand, by Theorem 32, we get that the expected error must be at least $1/2 - O(n^{-1/2} \cdot \log S(\varepsilon, \varepsilon/4))$. Combining these bounds, we get that

$$\frac{1}{2} - \frac{3\varepsilon}{4} \geq \frac{1}{2} - O(n^{-1/2} \cdot \log S(\varepsilon,\varepsilon/4)).$$

This implies that $\log S(\varepsilon, \varepsilon/4)) = \Omega(\varepsilon \cdot \sqrt{n})$. This implies that

$$\frac{nM\log M}{\varepsilon^2} + \frac{\log(1/\varepsilon)}{\varepsilon^2} = 2^{\Omega(\varepsilon\cdot\sqrt{n})}.$$

Since $\varepsilon = \omega(n^{-1/2} \cdot \log n)$ and $M \geq 1$, the first term on the left hand side is the dominant one, and thus we have

$$M \log M = \frac{\varepsilon^2}{n} \cdot 2^{\Omega(\varepsilon\cdot\sqrt{n})}.$$

Again using $\varepsilon = \omega(n^{-1/2} \cdot \log n)$, we have that

$$M\log M = 2^{\Omega(\varepsilon\cdot\sqrt{n})} \qquad \text{and thus} \qquad M = 2^{\Omega(\varepsilon\cdot\sqrt{n})}.$$

This finishes the proof of Theorem 28. □

VII. Lower Bounds for the ℓ_1 and ℓ_2 Balls (and More) via Convex Influences

Taking ε to be a sufficiently small constant in Theorem 28, we can infer the existence of some convex set in $\mathbb{R}^n$ such that $2^{\Omega(\sqrt{n})}$ halfspaces are required for any ε-approximator, but that result does not let us conclude that any particular convex set is hard to approximate.

In this section we show that the ℓ_1 and ℓ_2 balls B_1 and B_2 (defined in Equation (30) of the full version [26]) are each hard to approximate:

Theorem 33. Any intersection of halfspaces that approximates B_2 to error ε must have at least $2^{\Omega(\sqrt{n})}$ facets, for some absolute positive constant $\varepsilon > 0$. The same is true for B_1.

Theorem 33 implies that the upper bound obtained in Section V is tight in the constant error regime, up to constant factors in the exponent.

Our proof of Theorem 33 will crucially make use of the notion of *convex influence* which was introduced by [24], [25]. More generally, we prove a lower bound on the number of facets required to approximate any symmetric[5] convex set whose convex influence is asymptotically maximal up to constant multiplicative factors (cf. Theorem 39); see Section VII-C for more on this.

[5]Recall that a set $K \subseteq \mathbb{R}^n$ is *symmetric* if $x \in K$ implies $-x \in K$.

A. Convex Influences

The following notion was introduced in [24], [25] as an analogue of the well-studied notion of *influence of a variable on a Boolean function* (cf. Chapter 2 of [59]).

Definition 34. Given a convex set $K \subseteq \mathbb{R}^n$ with $0^n \in K$, the *convex influence of a direction* $v \in \mathbb{S}^{n-1}$ *on* K is defined as

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

where $K(\cdot)$ is the 0/1-valued indicator function of the convex set K. We further define the *total convex influence of* K as

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

We note that the definitions of $\mathbf{Inf}_v[K]$ and $\mathbf{I}[K]$ as defined in [25] include an additional multiplicative factor of $1/\sqrt{2}$ that we omit here. The total convex influence as defined above can be understood as capturing the rate of growth of the Gaussian measure of a convex set under dilations. More formally, we have the following:

Proposition 35 (Dilation formulation of convex influence)**.** Given $K \subseteq \mathbb{R}^n$ with $0^n \in K$, we have

$$\mathbf{I}[K] = \lim_{\delta\to 0} \frac{\mathrm{Vol}\big((1+\delta)K\big) - \mathrm{Vol}(K)}{\delta}.$$

Note that if $K \subseteq \mathbb{R}^n$ is convex with $0^n \in K$, then $\mathbf{I}[K]$ is non-negative. Proposition 35 is analogous to the well-known Margulis–Russo lemma [54], [64] from the analysis of Boolean functions, and a proof of it can be found in Appendix A of [23]. We note that a similar "dilation formulation" holds for the convex influence of a single direction $v \in \mathbb{S}^{n-1}$ on K, although we will not require it here.

We will use the following alternative formulation of the total convex influence of a convex set, which was communicated to us by Joe Neeman [58]:

Lemma 36 (Influence via a surface integral)**.** Given a measurable set $K \subseteq \mathbb{R}^n$ with $0^n \in K$ and a direction $v \in \mathbb{S}^{n-1}$, we have

$$\mathbf{I}[K] = \int_{\partial K} \langle x, \nu_x\rangle \cdot \varphi(x)\, d\sigma(x)$$

where ν_x denotes the unit normal to ∂K at x.

Proof. The proof is a straightforward computation using integration by parts. Recall that via Definition 34, we have

$$\begin{aligned}\mathbf{I}[K] &= \mathop{\mathbf{E}}_{\boldsymbol{x}\sim N(0,I_n)}\left[K(\boldsymbol{x})\Big(n - \|\boldsymbol{x}\|^2\Big)\right] \\ &= \int_K (n - \|x\|^2)\cdot\varphi(x)\,dx \\ &= \int_K \mathcal{L}\left(\frac{\|x\|^2}{2}\right)\cdot\varphi(x)\,dx\end{aligned}$$

where for a function $f : \mathbb{R}^n \to \mathbb{R}$ we define $\mathcal{L}(f) := \Delta f - \langle \nabla f, \nabla f \rangle$. Integrating by parts then gives

$$\mathbf{I}[K] = \int_{\partial K} \langle x, \nu_x \rangle \cdot \varphi(x)\, d\sigma(x),$$

completing the proof. □

Remark 37. We note that the surface integral formulation of total convex influence can be viewed as analogous to the fact that the total influence of a Boolean function is equal to its average sensitivity (cf. Chapter 2 of [59]). Although we will not require it for our purposes, we note that a similar formulation holds for the convex influence of a direction $v \in \mathbb{S}^{n-1}$ on K:

$$\mathbf{Inf}_v[K] = \int_{\partial K} \langle x, v \rangle \cdot \langle \nu_x, v \rangle \cdot \varphi(x)\, d\sigma(x)$$

where as before ν_x denotes the unit normal to ∂K at x.

B. Bounds on the Convex Influence of Polytopes

In this subsection we give an upper bound on the total convex influence of an intersection of halfspaces in terms of the number of halfspaces. An analogous statement for CNF formulas over $\{0,1\}^n$, showing that the influence of any s-clause CNF is at most $O(\log s)$, was first given by Boppana [11] using the technique of random restrictions [35].[6] The proof of Proposition 38 is inspired by the proof of Theorem 14 due to Nazarov; we give a self-contained proof of Proposition 38 below (see [4] for a proof sketch of Nazarov's bound).

Proposition 38 (Convex influence upper bound for intersections of halfspaces.)**.** Let $K \subseteq \mathbb{R}^n$ be an intersection of $s \geq 3$ halfspaces that contains the origin. Then

$$\mathbf{I}[K] < 7 \ln s.$$

We remark that the upper bound of Proposition 38 is best possible up to the hidden constant; the ℓ_∞ ball $K = \{x \in \mathbb{R}^n : \|x\|_\infty \leq r\}$ is an intersection of $2n$ halfspaces, and it is shown in Example 18 of [25] that for a suitable choice of r we have $\mathbf{I}[K] = \Theta(\log n)$.

Proof of Proposition 38. We let $K = \bigcap_{i=1}^s H_i$ for $s > 1$ where each H_i is a halfspace of the form

$$H_i := \left\{ x \in \mathbb{R}^n : \langle x, v_i \rangle \leq \theta_i \right\}$$

where $v_i \in \mathbb{S}^{n-1}$ for $i \in [s]$. Note that each $\theta_i \geq 0$ since K contains the origin. Using Lemma 36, we have

$$\begin{aligned}\mathbf{I}[K] &= \int_{\partial K} \langle x, \nu_x \rangle \cdot \varphi(x)\, d\sigma(x) \\ &= \sum_{i=1}^{s} \left(\int_{\partial H_i \cap \partial K} \langle x, \nu_x \rangle \cdot \varphi(x)\, d\sigma(x) \right). \qquad (27)\end{aligned}$$

Now, we observe that (i) $\mathbf{GSA}(K) = \sum_{i=1}^s \left(\int_{\partial H_i \cap \partial K} \varphi(x)\, d\sigma(x) \right)$; (ii) for $x \in \partial H_i \cap \partial K$ we have $\langle x, \nu_x \rangle = \theta_i$; and (iii) for each $i \in [s]$, we have $\int_{\partial H_i \cap \partial K} \varphi(x)\, d\sigma(x) \leq \int_{\partial H_i} \varphi(x)\, d\sigma(x) = \mathbf{GSA}(H_i)$. Combining these three observations, we get that

$$(27) \leq \max_{i \in [s] : \theta_i \leq \sqrt{2 \ln s}} \theta_i \mathbf{GSA}(K) + \sum_{\substack{j \in [s] \\ \theta_j > \sqrt{2 \ln s}}} \theta_j \mathbf{GSA}(H_j). \qquad (28)$$

We will control each of the two quantities in Equation (28) separately. For the first, we have that

$$\max_{i \in [s] : \theta_i \leq \sqrt{2 \ln s}} \theta(i) \cdot \mathbf{GSA}(K) \leq \sqrt{2 \ln s} \left(\sqrt{2 \ln s} + 2 \right) < 5 \ln s, \qquad (29)$$

where the first inequality is by Nazarov's bound on GSA (Theorem 14). For the second sum, we have that

$$\begin{aligned}\sum_{\substack{j \in [s] \\ \theta_j > \sqrt{2 \ln s}}} \theta_j \cdot \mathbf{GSA}(H_j) &= \sum_{\substack{j \in [s] \\ \theta_j > \sqrt{2 \ln s}}} \theta_j \cdot e^{-\theta_j^2/2} \\ &\leq s \cdot \max_{j : \theta_i > \sqrt{2 \ln s}} \left\{ \theta_j \cdot e^{-\theta_j^2/2} \right\}. \qquad (30)\end{aligned}$$

Since $xe^{-x^2/2}$ is a decreasing function for $x \geq 1$, and since $s > 1$, it follows that $\theta_j \cdot e^{-\theta_j^2/2}$ is maximized for $\theta_j = \sqrt{2 \ln s}$ which lets us conclude that

$$(30) \leq \sqrt{2 \ln s} < 2 \ln s. \qquad (31)$$

The result follows from Equations (28) to (31). □

C. Lower Bounds for Approximating Convex Sets with Maximal Influence

We finally establish the following lower bound on approximating symmetric convex sets with close-to-maximal convex influence. (The specific constants in the theorem below were chosen mostly for concreteness; other constants could have been used instead.)

Theorem 39. Suppose $K \subseteq \mathbb{R}^n$ is a symmetric convex set with $\mathrm{Vol}(K) = 1/2 \pm o_n(1)$ and $\mathbf{I}[K] \geq 0.1\sqrt{n}$. Then any convex polytope L that 6.25×10^{-7}-approximates K must have at least $2^{5.1 \times 10^{-6} \sqrt{n}}$ halfspaces.

Proof. Theorem 39 can be inferred along the lines of the proof of Theorem A.1 from [62], but we give a slightly simpler argument below. Our proof will make use of the Hermite basis; we refer the reader to Appendix A of the full version [26] for a primer on Hermite analysis over the Gaussian measure. Writing $e_i \in \mathbb{R}^n$ for the standard basis vector along the i^{th} coordinate direction, we have that

$$\mathbf{Inf}_{e_i}[K] = \sqrt{2} \cdot \widetilde{K}(2e_i).$$

(This is an immediate consequence of the fact that the degree-2 univariate Hermite polynomial $h_2(x)$ is $\left(1 - x^2\right)/\sqrt{2}$.) Since K is symmetric, it follows from Proposition 9 of [25] that $\mathbf{Inf}_{e_i}[K] \geq 0$. We will use the following simple claim that relies on this fact:

[6] In fact, Boppana [11] obtains an upper bound of $O(\log^{d-1}(s))$ on the total influence of functions computed by depth-d size-s circuits.

Claim 40. Suppose K is as in the statement of Theorem 39. Then at least 0.002-fraction of directions $\{e_1, \ldots, e_n\}$ must have

$$\mathbf{Inf}_{e_i}[K] \geq \frac{0.05}{\sqrt{n}}. \tag{32}$$

Proof. Let $\boldsymbol{i} \sim [n]$ uniformly at random. We have

$$\mathop{\mathbf{E}}_{\boldsymbol{i}\sim[n]}[\mathbf{Inf}_{e_{\boldsymbol{i}}}[K]] \geq \frac{0.1}{\sqrt{n}} \qquad \text{and} \qquad \mathop{\mathbf{Var}}_{\boldsymbol{i}\sim[n]}\left[\mathbf{Inf}_{e_{\boldsymbol{i}}}[K]\right] \leq \frac{1}{n}$$

where the upper bound on the variance follows from Parseval's formula. Recall Cantelli's inequality, which says that for a non-negative random variable $\boldsymbol{X}$ with mean μ and variance σ,

$$\mathbf{Pr}\left[\boldsymbol{X} > \mu - \theta\sigma\right] \geq \frac{\theta^2}{1+\theta^2}$$

for $\theta \in [0, 1]$. (Cantelli's inequality is a straightforward consequence of the Paley–Zygmund inequality.) Since $\mathbf{Inf}_{e_i}[K]$ is non-negative due to the symmetry of K (Proposition 9 of [25]), it follows that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{i}\sim[n]}\left[\mathbf{Inf}_{e_i}[K] > \frac{0.05}{\sqrt{n}}\right] \geq \frac{0.0025}{1.0025} > 0.002,$$

which completes the proof. □

Without loss of generality, let $\{e_1, \ldots, e_t\}$ be the coordinate directions for which $\mathbf{Inf}_{e_i}[K] \geq 0.05/\sqrt{n}$ where $t \geq 0.002n$, as guaranteed by Claim 40.

Now, let L be a convex polytope that 6.25×10^{-7}-approximates K, as in the theorem statement. We observe that L must contain the origin, since if it did not, by the symmetry of K and the fact that $\mathrm{Vol}(K) = \frac{1}{2} \pm o_n(1)$ we would have that $\mathrm{dist}_{\mathrm{G}}(K, L) \geq 0.249$. Hence we can (and will) analyze the influences of various coordinates on L.

Suppose that $\mathbf{I}[L] \leq 2.5 \times 10^{-5}\sqrt{n}$. It then follows that at most $t/2$ of $\{e_1, \ldots, e_t\}$ have

$$\mathbf{Inf}_{e_i}[L] \geq \frac{0.05}{2\sqrt{n}}.$$

In other words, for at least $t/2 \geq 0.001n$ of the coordinates in $\{e_1, \ldots, e_t\}$, we have

$$\mathbf{Inf}_{e_i}[L] < \frac{0.05}{2\sqrt{n}}. \tag{33}$$

Call these coordinates "bad" coordinates. Using Parseval's formula and Equations (32) and (33), we get that

$$\begin{aligned}
\mathrm{dist}(K, L) &= \mathop{\mathbf{E}}_{\boldsymbol{x}\sim N(0,I_n)}\left[(K(\boldsymbol{x}) - L(\boldsymbol{x}))^2\right] \\
&= \sum_{\alpha\in\mathbb{N}^n} (\widetilde{K}(\alpha) - \widetilde{L}(\alpha))^2 \\
&\geq \sum_{e_i \text{ is bad}} (\widetilde{K}(2e_i) - \widetilde{L}(2e_i))^2 \\
&> 0.001n \cdot \left(\frac{0.05}{2\sqrt{n}}\right)^2 = 6.25 \times 10^{-7},
\end{aligned}$$

which contradicts the fact that $\mathrm{dist}_{\mathrm{G}}(K, L) \leq 6.25 \times 10^{-7}$. Hence we must have $\mathbf{I}[L] > 2.5 \times 10^{-5}\sqrt{n}$, and by Proposition 38 this means that L must be an intersection of at least

$$2^{\left(\frac{0.25\times10^{-4}}{7\ln 2}\right)\sqrt{n}} > 2^{5.1\times10^{-6}\sqrt{n}}$$

many halfspaces. □

As a consequence of Theorem 39, we immediately obtain lower bounds for approximating the ℓ_2 and ℓ_1 balls of Gaussian measure $\approx 1/2$:

Example 41 (ℓ_2 ball)**.** Example 13 of [25] establishes that

$$\mathbf{I}[B_2] = \Theta(\sqrt{n}).$$

Lemma 36 and Gaussian isoperimetry, however, allow us to obtain a lower bound on the constant hidden by the $\Theta(\cdot)$. Since B_2 is an intersection of infinitely many halfspaces all of which are at distance $\sqrt{n}$ from the origin, we have

$$\mathbf{I}[B_2] = \sqrt{n} \cdot \mathbf{GSA}(B_2) \geq \sqrt{\frac{n}{2\pi}}(1 - o_n(1)) \geq 0.398\sqrt{n}$$

where the first equality is due to Lemma 36 and the second inequality follows from Gaussian isoperimetry. Together with Theorem 39, this immediately implies that any 6.25×10^{-7}-approximation to $B(\sqrt{n})$ requires $2^{\Omega(\sqrt{n})}$ facets. This in turn implies that Theorem 26 is tight up to constant factors when $\varepsilon = 6.25 \times 10^{-7}$.

Examples 42 (Cross-polytope)**.** Let B_1 be the cross-polytope, i.e. the set

$$B_1 := \left\{x \in \mathbb{R}^n : \|x\|_1 \leq \sqrt{\frac{2}{\pi}n}\right\}.$$

Note that B_1 is an intersection of 2^n halfspaces, and from Section 6.1 of the full version [26] that $\mathrm{Vol}(B_1) = 1/2 \pm o_n(1)$. By the Gaussian isoperimetric inequality, we have that $\mathbf{GSA}(B_1) \geq 0.398$, and so using Lemma 36 we have

$$\mathbf{I}[B_1] = \sqrt{\frac{2}{\pi}n} \cdot \mathbf{GSA}(B_1) \geq 0.1\sqrt{n}.$$

Together with Theorem 39, this immediately implies that any 6.25×10^{-7}-approximation to B_1 requires $2^{\Omega(\sqrt{n})}$ facets. (Recall from Theorem 29 of the full version [26] that B_1 can be 0.01-approximated using $2^{\Theta(n^{3/4})}$ halfspaces.)

ACKNOWLEDGMENT

We thank Elisabeth Werner for answering questions about [52]. We thank Ryan O'Donnell and Ashwin Padaki for helpful discussions. This work was partially completed while the authors were visiting the Simons Institute for the Theory of Computing as a part of the program on "Analysis and TCS: New Frontiers."

REFERENCES

[1] A. Abdelkader, S. Arya, G. da Fonseca, and D. M. Mount, "Approximation Algorithms for Geometric Proximity Problems: Part II: Approximating Convex Bodies," 2018, hMI-GCA Workshop 2018. 8

[2] S. Arya, G. D. da Fonseca, and D. M. Mount, "Optimal area-sensitive bounds for polytope approximation," in *Proceedings of the 28th ACM Symposium on Computational Geometry*. ACM, 2012, pp. 363–372. 4, 8

[3] K. Ball, "The Reverse Isoperimetric Problem for Gaussian Measure," *Discrete and Computational Geometry*, vol. 10, pp. 411–420, 1993. 3, 7

[4] ——, "Noise sensitivity and Gaussian surface area," Workshop on Probability & Geometry in High Dimensions, May 2010, available online. 17

[5] K. Ball *et al.*, "An elementary introduction to modern convex geometry," *Flavors of geometry*, vol. 31, no. 1–58, p. 26, 1997. 13

[6] A. Belovs, E. Blais, and A. Bommireddi, "Testing convexity of functions over finite domains," in *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2020, pp. 2030–2045. 2

[7] A. C. Berry, "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941. 7

[8] H. Black, E. Blais, and N. Harms, "Testing and learning convex sets in the ternary hypercube," in *15th Innovations in Theoretical Computer Science Conference, ITCS*, 2024, pp. 15:1–15:21. 2

[9] E. Blais and A. Bommireddi, "On testing and robust characterizations of convexity," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, 2020, pp. 18:1–18:15. 2

[10] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension," *Journal of the ACM*, vol. 36, no. 4, pp. 929–965, 1989. 15

[11] R. B. Boppana, "The average sensitivity of bounded-depth circuits," *Inf. Process. Lett.*, vol. 63, no. 5, pp. 257–261, 1997. 6, 17

[12] C. Borell, "The Brunn-Minkowski inequality in Gauss space," *Inventiones mathematicae*, vol. 30, no. 2, pp. 207–216, 1975. 8

[13] E. M. Bronshteyn and L. D. Ivanov, "The approximation of convex sets by polyhedra," vol. 16, pp. 852–853, 1976. 13

[14] ——, "The approximation of convex sets by polyhedra," *Siberian Mathematical Journal*, vol. 16, no. 5, pp. 852–853, 1975. 8

[15] E. Bronstein, "Approximation of Convex Sets by Polytopes," *Journal of Mathematical Sciences*, vol. 153, no. 6, pp. 727–762, 2008. 1, 4, 8

[16] N. Bshouty and C. Tamon, "On the Fourier spectrum of monotone functions," *Journal of the ACM*, vol. 43, no. 4, pp. 747–770, 1996. 2

[17] E. Chattopadhyay, A. De, and R. A. Servedio, "Simple and Efficient Pseudorandom Generators from Gaussian Processes," in *34th Computational Complexity Conference (CCC)*, ser. LIPIcs, vol. 137, 2019, pp. 4:1–4:33. 1, 2

[18] X. Chen, A. Freilich, R. Servedio, and T. Sun, "Sample-based high-dimensional convexity testing," in *Proceedings of the 17th Int. Workshop on Randomization and Computation (RANDOM)*, 2017, pp. 37:1–37:20. 2

[19] X. Chen, A. De, S. Nadimpalli, R. A. Servedio, and E. Waingarten, "Lower bounds for convexity testing," 2024, manuscript (personal communication). 2

[20] A. De, E. Mossel, and J. Neeman, "Is your function low dimensional?" in *Conference on Learning Theory, COLT*, ser. Proceedings of Machine Learning Research, vol. 99. PMLR, 2019, pp. 979–993. 1, 2

[21] ——, "Robust testing of low dimensional functions," in *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2021, pp. 584–597. 1, 2

[22] A. De, S. Nadimpalli, R. O'Donnell, and R. A. Servedio, "Sparsifying suprema of Gaussian processes," 2024, manuscript (personal communication). 2

[23] A. De, S. Nadimpalli, and R. A. Servedio, "Convex influences," 2021, full version available at arXiv: 2109.03107. [Online]. Available: https://arxiv.org/abs/2109.03107 16

[24] ——, "Quantitative correlation inequalities via semigroup interpolation," in *12th Innovations in Theoretical Computer Science Conference, ITCS 2021*, ser. LIPIcs, vol. 185, 2021, pp. 69:1–69:20. 2, 4, 5, 16

[25] ——, "Convex influences," in *13th Innovations in Theoretical Computer Science Conference, ITCS*, ser. LIPIcs, vol. 215, 2022, pp. 53:1–53:21. 2, 4, 5, 6, 16, 17, 18

[26] ——, "Gaussian Approximation of Convex Sets by Intersections of Halfspaces," 2023. 4, 5, 9, 16, 17, 18

[27] ——, "Testing Convex Truncation," in *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2023, pp. 4050–4082. 1

[28] A. De and R. A. Servedio, "Weak learning convex sets under normal distributions," in *Conference on Learning Theory, COLT 2021*, ser. Proceedings of Machine Learning Research, vol. 134. PMLR, 2021, pp. 1399–1428. 2, 4, 5, 15

[29] I. Diakonikolas, D. M. Kane, T. Pittas, , and N. Zarifis, "The optimality of polynomial regression for agnostic learning under Gaussian marginals in the SQ model," in *ThirtyFourth Annual Conference on Learning Theory (COLT)*, 2021. 1, 2

[30] R. M. Dudley, "Metric entropy of some classes of sets with differentiable boundaries," *Journal of Approximation Theory*, vol. 10, pp. 227–236, 1974. 8, 13

[31] R. Durrett, *Probability: Theory and Examples*, 5th ed., ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. 7

[32] C.-G. Esseen, "On the Liapunoff limit of error in the theory of probability," *Arkiv för matematik, astronomi och fysik*, vol. A, pp. 1–19, 1942. 7

[33] P. Gruber, "Approximation of Convex Bodies," in *Convexity and its Applications*. Birkhäuser, Basel, 1983, pp. 131–162. 1

[34] ——, "Aspects of Approximation of Convex Bodies," in *Handbook of Convex Geometry*, P. Gruber and J. Wills, Eds. Elsevier Science Publishers, 1993, pp. 319–345. 1

[35] J. Håstad, *Computational Limitations for Small Depth Circuits*. Cambridge, MA: MIT Press, 1986. 17

[36] ——, "Almost optimal lower bounds for small depth circuits," in *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, 1986, pp. 6–20. 6

[37] D. J. Hsu, C. H. Sanford, R. A. Servedio, and E. Vlatakis-Gkaragkounis, "Near-Optimal Statistical Query Lower Bounds for Agnostically Learning Intersections of Halfspaces with Gaussian Marginals," in *Conference on Learning Theory (COLT)*, ser. Proceedings of Machine Learning Research, vol. 178. PMLR, 2022, pp. 283–312. 1, 2, 4

[38] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on boolean functions," in *Proc. 29th Annual Symposium on Foundations of Computer Science (FOCS)*, 1988, pp. 68–80. 6

[39] D. Kane, "A Structure Theorem for Poorly Anticoncentrated Gaussian Chaoses and Applications to the Study of Polynomial Threshold Functions," in *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 2012, pp. 91–100. 1

[40] ——, "A Pseudorandom Generator for Polynomial Threshold Functions of Gaussian with Subpolynomial Seed Length," in *IEEE 29th Conference on Computational Complexity, CCC 2014*, 2014, pp. 217–228. 1

[41] D. M. Kane, "The Gaussian Surface Area and Noise Sensitivity of Degree-d Polynomial Threshold Functions," *Computational Complexity*, vol. 20, no. 2, pp. 389–412, 2011. 1

[42] ——, "The average sensitivity of an intersection of half spaces," in *Symposium on Theory of Computing, STOC 2014*, 2014, pp. 437–440. 7

[43] ——, "A Polylogarithmic PRG for Degree 2 Threshold Functions in the Gaussian Setting," in *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, 2015, pp. 567–581. 1

[44] M. Kearns and U. Vazirani, *An Introduction to Computational Learning Theory*. Cambridge, MA: MIT Press, 1994. 15

[45] A. Klivans, R. O'Donnell, and R. Servedio, "Learning geometric concepts via Gaussian surface area," in *Proc. 49th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2008, pp. 541–550. 1, 2, 5, 7

[46] A. R. Klivans and P. Kothari, "Embedding Hard Learning Problems Into Gaussian Space," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014*, 2014, pp. 793–809. 1

[47] P. Kothari, A. Nayyeri, R. O'Donnell, and C. Wu, "Testing surface area," in *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*. SIAM, 2014. 1

[48] R. Lang, "A note on the measurability of convex sets," *Archiv der Mathematik*, vol. 47, pp. 90–92, 1986. 7

[49] J. Lange, R. Rubinfeld, and A. Vasilyan, "Properly learning monotone functions via local correction," in *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022*. IEEE, 2022, pp. 75–86. 2

[50] J. Lange and A. Vasilyan, "Agnostic proper learning of monotone functions: beyond the black-box correction barrier," vol. abs/2304.02700, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.02700 2

[51] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1302–1338, 10 2000. [Online]. Available: https://doi.org/10.1214/aos/1015957395 7

[52] M. Ludwig, C. Schütt, and E. Werner, "Approximation of the euclidean ball by polytopes," *Studia Mathematica*, vol. 173, no. 1, pp. 1–18, 2006. 4, 13, 18

[53] A. Macbeath, "An extremal property of the hypersphere," *Proc. Cambridge Philos. Soc.*, vol. 47, pp. 245–247, 1951. 1

[54] G. Margulis, "Probabilistic characteristics of graphs with large connectivity," *Prob. Peredachi Inform.*, vol. 10, pp. 101–108, 1974. 16

[55] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio, "Testing halfspaces," *SIAM J. on Comput.*, vol. 39, no. 5, pp. 2004–2047, 2010. 1, 2

[56] V. Milman, "Surprising geometric phenomena in high-dimensional convexity theory," in *European Congress of Mathematics: Budapest, July 22–26, 1996 Volume II*. Springer, 1998, pp. 73–91. 13

[57] 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)*. Lecture Notes in Math., Vol. 1807, Springer, 2003, pp. 169–187. 4, 7, 13

[58] J. Neeman, "Personal communication," 2022. 16

[59] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014. 6, 7, 16, 17

[60] ——, "Personal communication," 2023. 5

[61] R. O'Donnell, R. A. Servedio, L. Tan, and D. Kane, "Fooling Gaussian PTFs via local hyperconcentration," 2021, preliminary version in STOC 2020. Revised version includes an appendix by Daniel Kane. 1

[62] R. O'Donnell and K. Wimmer, "Approximation by dnf: examples and counterexamples," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2007, pp. 195–206. 4, 13, 17

[63] L. Rademacher and S. Vempala, "Testing geometric convexity," in *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science: 24th International Conference*, 2005, pp. 469–480. 2

[64] L. Russo, "On the critical percolation probabilities," *Z. Wahrsch. werw. Gebiete*, vol. 43, pp. 39–48, 1978. 16

[65] E. Sas, "Über ein Extremumeigenschaft der Ellipsen," *Compositio Math.*, vol. 6, pp. 468–470, 1939. 1

[66] R. Servedio, "Every linear threshold function has a low-weight approximator," *Comput. Complexity*, vol. 16, no. 2, pp. 180–209, 2007. 2

[67] V. N. Sudakov and B. S. Tsirel'son, "Extremal properties of half-spaces for spherically invariant measures," *Journal of Soviet Mathematics*, vol. 9, no. 1, pp. 9–18, 1978. 8

[68] M. Talagrand, "Sharper bounds for Gaussian and empirical processes," *Annals of Probability*, vol. 22, pp. 28–76, 1994. 15

[69] ——, "How much are increasing sets positively correlated?" *Combinatorica*, vol. 16, no. 2, pp. 243–258, 1996. 4, 13

[70] T. Tkocz, "Asymptotic Convex Geometry Lecture Notes," 2018, available at here. 9

[71] L. F. Tóth, "Approximation by polygons and polyhedra," *Bull. Amer. Math. Soc.*, vol. 54, pp. 431–438, 1948. 1

[72] S. S. Vempala, "Learning convex concepts from gaussian distributions with PCA," in *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS*. IEEE Computer Society, 2010, pp. 124–130. 1, 2

[73] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47. 10

[74] M. Wainwright, "Basic tail and concentration bounds," available online here. 7

[75] Wikipedia contributors, "Chi-squared distribution." Wikipedia, The Free Encyclopedia, accessed September 27, 2023. 7