Testing Surface Area

Pravesh Kothari
Department of Computer Science
The University of Texas at Austin
kothari@cs.utexas.edu

Amir Nayyeri*
Department of Computer Science
Carnegie Mellon University
amirn@cs.cmu.edu

Ryan O'Donnell†
Department of Computer Science
Carnegie Mellon University
odonnell@cs.cmu.edu

Chenggang Wu ‡
IIIS
Tsinghua University
wcg06@mails.tsinghua.edu.cn

August 27, 2013

Abstract

We consider the problem of estimating the surface area of an unknown n-dimensional set F given membership oracle access. In contrast to previous work, we do not assume that F is convex, and in fact make no assumptions at all about F. By necessity this means that we work in the property testing model; we seek an algorithm which, given parameters A and ϵ, satisfies:

- if $\mathrm{surf}(F) \leq A$ then the algorithm accepts (whp);
- if F is not ϵ-close to some set G with $\mathrm{surf}(G) \leq \kappa A$, then the algorithm rejects (whp).

We call $\kappa \geq 1$ the "approximation factor" of the testing algorithm.

The $n = 1$ case (in which "$\mathrm{surf}(F) = 2m$" means F is a disjoint union of m intervals) was introduced by Kearns and Ron [KR98], who solved the problem with $\kappa = 1/\epsilon$ and $O(1/\epsilon)$ oracle queries. Later, Balcan et al. [BBBY12] solved it with with $\kappa = 1$ and $O(1/\epsilon^4)$ queries.

We give the first result for higher dimensions n. Perhaps surprisingly, our algorithm completely evades the "curse of dimensionality": for any n and any $\kappa > \frac{4}{\pi} \approx 1.27$ we give a test that uses $O(1/\epsilon)$ queries. For small n we have improved bounds. For $n = 1$ we can achieve $\kappa = 1$ with $O(1/\epsilon^{3.5})$ queries (slightly improving [BBBY12]), or any $\kappa > 1$ with $O(1/\epsilon)$ queries (improving [KR98]). For $n = 2, 3$ we obtain $\kappa \approx 1.08, 1.125$ respectively, with $O(1/\epsilon)$ queries. Getting an arbitrary $\kappa > 1$ for $n > 1$ remains an open problem.

Finally, motivated by the learning results from [KOS08], we extend our techniques to obtain a similar tester for *Gaussian* surface area for any n, with query complexity $O(1/\epsilon)$ and any approximation factor $\kappa > \frac{4}{\pi} \approx 1.27$.

*Supported by NSF grants CCF-1065106.

†Supported by NSF grants CCF-0747250 and CCF-1116594 and a grant from the MSR–CMU Center for Computational Thinking.

‡This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540. Research done while visiting Carnegie Mellon University.

1 Introduction

We consider a fundamental geometric problem — estimating the surface area of a set $F \subseteq \mathbb{R}^n$ given point-query access. This problem arises in several practical scientific areas under the name stereology, as well as in the fields of statistics and computer graphics/vision, e.g., [CFRC07, LYZ$^+$10, YC96, Ras03].

In the special case that F is promised to be *convex*, the problem is closely related to the well-known task in theoretical computer science of estimating volume given membership oracle access (and a point inside F). The task of estimating surface area of convex sets was stated as an open problem in the monumental work of Grötschel et al. [GLS88]. It was apparently first solved by Belkin et al. [BNN06]; they showed that an $\widetilde{O}(n^{8.5})$-query algorithm could be deduced from the work of Dyer et al. [DGH98], and also gave an $\widetilde{O}(n^4)$-query algorithm.

In this work we consider arbitrary sets $F \subseteq \mathbb{R}^n$ and do not make any assumptions of convexity. By necessity, this means that we must relax our goals slightly. For example, given a nice convex set such as a sphere, one can add a very thin tentacle to it with negligible volume but arbitrarily large surface area. No algorithm with only oracle access to the set can hope to find (let alone estimate the surface area of) this tentacle, even given a starting point inside the initial convex set. Relatedly, one can take a nice convex set and make part of its boundary "wiggle violently" at a very small scale, increasing its surface area by an arbitrary amount in a way that is undetectable to oracle algorithms. Thus, without the convexity assumption, we must allow ourselves the flexibility of disregarding or smoothing out tiny-volume pieces of the set. This leads us naturally to work in the model of *property testing* [RS96, BLR93, GGR98].

1.1 Property testing

The natural framing of surface area estimation in the property testing model is as follows. An algorithm is given black-box query access to a set $F \subseteq \mathbb{R}^n$, meaning it can query any point $x \in \mathbb{R}^n$ and learn whether or not $x \in F$. The algorithm is also given as input two parameters: a surface area upper bound A and an error tolerance $\epsilon > 0$. The algorithm — which is allowed to be randomized — makes queries to F and then, must either "accept" or "reject". The goal for it is to accept (with high probability) if F's surface area satisfies $\mathrm{surf}(F) \leq A$ and reject (with high probability) if F is "ϵ-far" from having $\mathrm{surf}(F) \leq A$. Roughly speaking, the second condition means that if F is accepted with non-negligible probability, then it can be made to have surface area at most A after altering at most ϵ volume. (Presumably, this entails deleting some small parts of F, or smoothing out some small crinkles.) One is typically less concerned with the actual running time of the testing algorithm than its query complexity, though, in fact, the testing algorithms in this paper are extremely efficient (running in time linear in the number of queries).

We will also allow one more relaxation to the usual property testing framework: a slight gap between the "completeness" and "soundness". Specifically, we will only require the testing algorithm to reject with high probability if F is ϵ-far from having $\mathrm{surf}(F) \leq \kappa \cdot A$. We call $\kappa \geq 1$ the "approximation factor". This extra flexibility (introduced already in the first work on the topic [KR98]) lets us achieve substantial query savings in $n = 1$ dimension and is necessary for our proof technique in higher dimensions.

To make the property testing definitions precise we need a notion of volume or probability density on $\mathbb{R}^n$. In this paper we consider two different possibilities. For the main part of the paper we will assume that the set F has been translated and scaled so that it lies inside the unit cube $(0, 1)^n$. Lebesgue measure on $(0, 1)^n$ is a probability measure and it gives us a natural notion of two sets $F, G \subseteq (0, 1)^n$ being ϵ-close, that the volume of their symmetric difference be at most ϵ: $\mathrm{vol}(F \triangle G) \leq \epsilon$.

We advise the reader that this setting is probably the most natural in the practical case of low dimensions; e.g., $n = 2$ or 3. In high dimensions, the assumption that $F \subseteq (0, 1)^n$ may cause the error tolerance ϵ to have a rather large effect. For example, any sphere $F \subseteq (0, 1)^n$ has volume $2^{-\Theta(n)}$ and can therefore be considered "negligible" (unless $\epsilon \leq 2^{-\Theta(n)}$). This seems like an unavoidable aspect of working in the property testing model. Indeed, for high dimensions n, we feel that our second setting, *Gaussian space*, is more natural.

Gaussian surface area. Property testing is often viewed as a precursor to Computational Learning (see [Ron08]) and indeed an important motivation for our paper arises from the learning results of Klivans et al. [KOS08]. In this case, the standard n-dimensional Gaussian measure is used to define the volume (and surface area) of sets in $\mathbb{R}^n$. Analogous to the previous setting, the distance between two sets $F, G \subseteq \mathbb{R}^n$ is defined as the *Gaussian* volume of $F \triangle G$. In this setting, Klivans et. al. gave an $n^{O(A^2)}$-time agnostic learning algorithm under the Gaussian distribution for the class of sets with *Gaussian surface area* at most A. Our Gaussian surface area tester fits very nicely in the "testing before learning" framework. The soundness of our tester guarantees that any set F that the tester accepts (with high probability), is ϵ-close to having Gaussian surface area at most $\kappa \cdot A$. The deficiency of being ϵ-close that comes with property testing is taken care of by the agnostic aspect of the [KOS08] learning algorithm, and the approximation factor κ does not make any real difference to the learning algorithm's run-time of $n^{O(A^2)}$.

1.2 Prior and related work

The framework of property testing has seen a tremendous amount of work, see, e.g., [Ron08, RS11]; especially in the areas of algebraic properties, e.g., [BLR93, DLM+08, AKK+05], graph properties, e.g., [GGR98, AFNS09], and discrete function properties, e.g., [Bla09, BO10]. For testing in the context of geometric properties see, e.g., the work of Czumaj et al. [CSZ00, CS01].

As far as we are aware, the specific problem studied in this paper — property testing surface area — has been studied previously only in $n = 1$ dimension. In this case, the "surface area" of an open set $F \subseteq [0, 1]$ is the number of points on its boundary; hence the problem is equivalent to testing whether F is a union of at most $A/2$ intervals. In this form, the problem was first proposed by Kearns and Ron [KR98]. They gave a tester solving the problem with excellent query complexity, $O(1/\epsilon)$, but a rather large approximation factor, $\kappa = 1/\epsilon$. The problem was taken up again by Balcan et al. in 2012 [BBBY12]. They gave a tester with query complexity $O(1/\epsilon^4)$ and no approximation factor — i.e., $\kappa = 1$. Actually, their tester naturally achieves $\kappa = 1 + \epsilon$, but they proved the following simple fact (specific to 1 dimension) which allowed them to reduce κ to exactly 1:

Fact 1.1 ([BBBY12]). *Any $F \subset (0, 1)$ which is the union of at most $m(1 + \epsilon)$ intervals is ϵ-close to a union of at most m intervals.*

1.3 Our results

We give the first efficient property testing algorithm for surface area in any dimension $n > 1$. Our testing algorithm is highly efficient — needs just $O(1/\epsilon)$ queries (and running time) — and achieves a small constant approximation factor κ for every dimension n, both in the standard Euclidean and Gaussian settings. We find it somewhat surprising that the "curse of dimensionality" can be completely avoided, especially in the standard Euclidean setting.

To state our results, let's first define the approximation factor κ_n we achieve in dimension n:

Definition 1.2. For $n \in \mathbb{N}^+$ and Γ, the standard Gamma function, define

$$\kappa_n = \frac{4}{\pi} \frac{n}{n+1} \frac{\Gamma(n/2+1)\Gamma(n/2)}{\Gamma(n/2+1/2)^2} = 2(n+1)\left[\frac{\binom{n}{n/2-1/2}}{2^n}\right]^2,$$

so $\kappa_1 = 1$, $\kappa_2 = \frac{32}{3\pi^2} \approx 1.08$, $\kappa_3 = \frac{9}{8} = 1.125$, etc.

Fact 1.3. *κ_n increases to a limit of $\frac{4}{\pi} \approx 1.27$ as $n \to \infty$.*

Proof. The fact that $\kappa_n \to \frac{4}{\pi}$ is a straightforward consequence of Stirling's formula. Thus it remains to show that $r_n = \kappa_n/\kappa_{n+1} \leq 1$ for all n. Since $r_n \to 1$ as $n \to \infty$ it suffices to show that $r_{n+2}/r_n \geq 1$ for all n. But by direct computation, $r_{n+2}/r_n = \frac{(n+4)(n+2)^3}{(n+1)(n+3)^3}$, which is easily shown to decrease to 1. □

We can now state our main theorem for n-dimensional Euclidean surface area. For technical simplicity (to avoid edge effects) we work in the unit torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ rather than the unit cube $(0, 1)^n$.

Theorem 1.4. *There is a randomized testing algorithm with the following properties. Given black-box query access to a set $F \subseteq \mathbb{T}^n$ with piecewise-$\mathcal{C}^1$ boundary, as well as parameters $A > 0$, $0 < \epsilon < 1/2$, and $0 < \eta < 1/2$, the algorithm makes $O(\frac{1/\eta^{2.5}}{\epsilon})$ queries and then either "accepts" or "rejects". The tester satisfies:*

- (Completeness.) *If* $\mathrm{surf}(F) \le A$*, then the tester accepts with probability at least* $9/10$.
- (Soundness.) *If the tester accepts with probability at least* $1/10$*, then there is a set* $G \subseteq \mathbb{T}^n$ *with* $\mathrm{vol}(F \triangle G) \le \epsilon$ *such that* $\mathrm{surf}(G) \le (\kappa_n + \eta) \cdot A$.

Here κ_n is as given by Definition 1.2.

(We remark also that in the soundness statement, we can ensure that G is a finite union of polytopes with $\mathrm{dist}(x, F) \le O(\sqrt{n\eta}\epsilon/A)$ for all $x \in G$.)

Let's consider Theorem 1.4 in combination with Fact 1.3. For $n = 1$, we can achieve any approximation factor $\kappa > 1$ using $O(1/\epsilon)$-queries. This significantly improves on the result in [KR98]. By taking $\eta = \epsilon$ and applying Fact 1.1, we can also achieve no approximation factor (i.e., $\kappa = 1$) with an $O(1/\epsilon^{3.5})$-query tester, slightly improving the result from [BBBY12]. For $n = 2$ we can achieve approximation factor 1.081 using $O(1/\epsilon)$ queries; for $n = 3$ we can achieve approximation factor 1.126 using $O(1/\epsilon)$ queries. Finally, for general dimensions n we can achieve any approximation factor $\kappa > \frac{4}{\pi} \approx 1.27$ using $O(1/\epsilon)$ queries.

Our result for testing Gaussian surface area is similar and completely dimension-independent:

Theorem 1.5. *There is a randomized testing algorithm with the following properties. Given black-box query access to a set $F \subseteq \mathbb{R}^n$ with piecewise-$\mathcal{C}^1$ boundary, as well as parameters $A > 0$, $0 < \epsilon < 1/2$, and $0 < \eta < 1/2$, the algorithm makes $O(\frac{1/\eta^{2.5}}{\epsilon})$ queries and then either "accepts" or "rejects". The tester satisfies:*

- (Completeness.) *If* $\mathrm{surf}_\gamma(F) \le A$*, then the tester accepts with probability at least* $9/10$.
- (Soundness.) *If the tester accepts with probability at least* $1/10$*, then there is a set* $G \subseteq \mathbb{R}^n$ *with* $\mathrm{vol}_\gamma(F \triangle G) \le \epsilon$ *such that* $\mathrm{surf}_\gamma(G) \le (\frac{4}{\pi} + \eta) \cdot A$.

Here $\mathrm{surf}_\gamma(\cdot)$ denotes Gaussian surface area and $\mathrm{vol}_\gamma(\cdot)$ denotes n-dimensional Gaussian probability measure.

We remark that we can't expect the approximation factor in the Gaussian setting to be better than the approximation factor in the Euclidean setting. This is because for sets $F \subseteq \mathbb{R}^n$ very close to the origin, Gaussian and Euclidean surface area are essentially the same, up to a scaling factor (namely, the value of the Gaussian pdf at 0).

1.4 Our methods

In this section we give a rough description of the ideas behind our surface area tester in $\mathbb{T}^n$. In fact, all of the complexity occurs already in the case of $n = 2$ dimensions. We therefore encourage the reader to keep the special case of $n = 2$ in mind (and read "perimeter" for "surface area", and "area" for "volume").

Let's warm up by considering the problem of testing *volume*. This is essentially a trivial problem in the property testing framework; a tester can accurately estimate $\mathrm{vol}(F)$ or $\mathrm{vol}_\gamma(F)$ simply by querying many independent random points and computing the empirical fraction which lie in F. Indeed, here is a simple theorem one can prove for testing Gaussian volume using just a standard Chernoff bound[1]:

Theorem 1.6. *There is a randomized testing algorithm* GaussianVolumeTest *with the following properties. Given black-box query access to a set $F \subseteq \mathbb{R}^n$ as well as parameters $V > 0$, $0 < \tau < 1/2$, the algorithm makes $O(\frac{1/\tau^2}{V})$ queries and then either "accepts" or "rejects". Further:*

- (Completeness.) *If* $\mathrm{vol}_\gamma(F) \le V$ *then the tester accepts with probability at least* $99/100$.

[1] A similar theorem also holds for testing Lebesgue volume.

- (Soundness.) *If* $\mathrm{vol}_\gamma(F) > (1+\tau)\cdot V$ *then the tester rejects with probability at least* $99/100$.

How about surface area? For sets F with nice enough surface we have $\mathrm{surf}(F) = \lim_{\delta\to 0} \frac{\mathrm{vol}(\partial F^{\delta/2})}{\delta}$, where $\partial F^{\delta/2}$ denotes the set of all points within distance $\delta/2$ of F's boundary. This yields a natural idea for estimating $\mathrm{surf}(F)$: choose a small $\delta > 0$, (try to) estimate $\mathrm{vol}(\partial F^{\delta/2})$ by the above sampling approach, and then divide by δ. Since oracle access to F does not directly give us oracle access to $\partial F^{\delta/2}$, we can modify this idea by, say, querying two random points x and y at roughly a distance of δ to see if one is inside F and the other outside.

This modification is backed up by an appropriate version of the Cauchy–Crofton or "Buffon's Needle" Theorem (see Theorem 2.3 below), which roughly says that

$$\mathrm{surf}(F) = c \cdot \frac{\mathbf{E}[\text{\# of intersections between } \partial F \text{ and a random "needle" } \overline{xy} \text{ of length } \delta]}{\delta}, \tag{1}$$

where $c = \Theta(\sqrt{n})$ is a dimension-dependent constant. We can *underestimate* the expectation in the numerator by the probability a random length-δ needle intersects ∂F, which in turn can underestimated by

$$\mathbf{Pr}[\text{two random points } x, y \text{ at distance } \delta \text{ satisfy } x \in F,\, y \notin F \text{ or vice versa}]. \tag{2}$$

This probability — which is something like the "noise sensitivity of F at δ" — can be accurately estimated by a tester. After dividing by δ/c we get an (under)estimate for the true surface area; if this is at most A, the tester can accept. Thus we have a tester whose completeness is ensured by the Buffon's Needle Theorem. We remark that the ideas so far are precisely those used in the $n = 1$ tester of Balcan et al. [BBBY12].

The hope is that this underestimate is close to the truth if δ is "small enough". The difficulty is that a random length-δ needle might pass through F's boundary $k \geq 2$ times; and in this case, even though it should contribute k to the numerator in (1), it will only be counted once (if k is odd) or zero times (if k is even) in the approximation (2). In the "limit" when δ becomes very small, we expect the needle to intersect F's boundary in either zero or one points, but how can the tester know how small does δ need to be? Indeed, the tester does not want to take δ too small, or else it will take too many samples to estimate (2) empirically. What we hope is that by choosing δ roughly proportional to ϵ/A, we get a significant underestimate only when the boundary of F "wiggles at a scale of δ", in which case, we can attempt to find a G satisfying $\mathrm{vol}(F\triangle G) \leq \epsilon$ by "smoothing out" such wiggles.

The difficulty is in defining such a G. As in [BBBY12], we consider a "smoothed-out" version of F's indicator function,

$$g(x) = \mathbf{Pr}[y \in F,\ \text{for } y \text{ a random point at distance roughly } \delta \text{ from x}].$$

(Actually, the "needle length" δ also needs to be randomized to avoid periodicity issues.) Then it appears natural to declare that G contains all x where $g(x)$ is close to 1 and that G doesn't contain any point where $g(x)$ is close to 0. It remains to decide how to define G on points x where $g(x)$ is "in the middle". (These correspond to points that are roughly near the boundary of F.) It can be shown that $\mathrm{vol}(F\triangle G)$ is small no matter what choice is made for such points; the challenge is to fill in the gaps in such a way that the resulting G has small surface area.

In the 1-dimensional case studied in [BBBY12] this task is very easy: the partial definition of G fixes some intervals to be contained in G and some intervals to be excluded from G. At this point, *any* sensible method of filling in the gaps yields appropriately small "surface area" (number of endpoints). In 2 or more dimensions, though, the problem is more complicated; indeed it seems very hard to give a construction for this gap-filling problem. Our solution is to use the Probabilistic Method; we define G to be a random super-level set of g, i.e., $G = \{x : g(x) > t\}$. It turns out that if t is chosen from the *triangular* probability distribution (rather than the naive choice of the uniform distribution) then the *coarea formula* from geometric integration theorem gives a link between surface area and noise sensitivity which is enough to establish the soundness of our tester.

1.5 Organization of the remainder of the paper

In Section 2 we give the technical preliminaries concerning surface area, the Cauchy–Crofton formula, and the coarea formula. In Section 3 we state our testing algorithm for $\mathbb{T}^n$ and prove its completeness. In Section 4 we prove the soundness of the tester. In Section 5 we give all the generalizations to the Gaussian case. Finally, we conclude with some open problems.

2 Technical preliminaries

Notation 2.1. We write $\mathbb{T}^n$ for the torus $\mathbb{R}^n/\mathbb{Z}^n$. We write B^n to denote the unit open ball in $\mathbb{R}^n$ and let S^{n-1} denote its boundary, the unit sphere in $\mathbb{R}^n$.

We define surface area only for "sets of finite perimeter"; for the definition of this term we refer the reader to [AFP00, Section 3.3]. We will use the following:

Notation 2.2. Let $F \subseteq \mathbb{T}^n$ be a "set of finite perimeter". We write $\mathrm{surf}(F)$ for its perimeter (surface area). For sets F with piecewise-$\mathcal{C}^1$ boundary this coincides with $\mathcal{H}^{n-1}(\partial F)$, where $\mathcal{H}^{n-1}$ denotes $(n-1)$-dimensional Hausdorff measure. We further remark that any set of finite perimeter $F \subseteq \mathbb{T}^n$ can be approximated by a sequence of open sets (F_i) each with piecewise-affine boundary[2] such that $\mathrm{vol}(F_i \triangle F) \to 0$, $\mathrm{surf}(F_i) \to \mathrm{surf}(F)$.

Our testing algorithm will rely on an alternate characterization of surface area, namely the following Cauchy–Crofton or "Buffon's Needle"-type fact from integral geometry (see, e.g., [San04, I.8.3, III.15.9]). Note that in this theorem the intuition is that the needle length δ is very short, though in fact it can be of any length.

Theorem 2.3. *Assume that $F \subseteq \mathbb{T}^n$ is such that ∂F is a piecewise-$\mathcal{C}^1$ surface. For $x \in \mathbb{T}^n$ and a vector y with $\|y\| = \delta$, let $N(x, y, F)$ denote the "number of times the needle from x to $x + y$ intersects ∂F"; more precisely, $\#\{t \in [0, 1] : x + ty \in \partial F\}$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathbb{T}^n, \boldsymbol{v} \sim S^{n-1}}[N(\boldsymbol{x}, \delta\boldsymbol{v}, F)] = \beta_n \cdot \mathrm{surf}(F) \cdot \delta.$$

Here β_n is the dimension-dependent constant whose definition is given below.

Definition 2.4. The constant β_n in Theorem 2.3 is defined by $\beta_n = \mathbf{E}_{\boldsymbol{v} \sim S^{n-1}}[|\boldsymbol{v}_1|]$. Note that β_n also equals $\mathbf{E}_{v \sim S^{n-1}}[|\langle \boldsymbol{v}, w\rangle|]$ for any fixed unit vector w.

Fact 2.5. *We have $\beta_1 = 1$, $\beta_2 = \frac{2}{\pi}$, $\beta_3 = \frac{1}{2}$, and in general $\beta_n = \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma(n/2+1/2)} \sim \sqrt{\frac{2}{\pi}}\sqrt{\frac{1}{n}}$.*

We will need to introduce a notion of "noise sensitivity" for functions on the torus. This notion was also used by [BBBY12] for their $n = 1$ testing algorithm.

Definition 2.6. Let $f : \mathbb{T}^n \to \mathbb{R}$ be integrable. For $\delta > 0$ we define $\mathrm{S}_\delta f : \mathbb{T}^d \to \mathbb{R}$ by

$$\mathrm{S}_\delta f(x) = \mathop{\mathbf{E}}_{\boldsymbol{z} \sim B^n}[f(x + \delta\boldsymbol{z})].$$

When f has range ± 1 (i.e., f is the ± 1-indicator of a measurable subset of $\mathbb{T}^n$) we also define

$$\mathbf{NS}_\delta[f] = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathbb{T}^n}[\tfrac{1}{2} - \tfrac{1}{2} f(\boldsymbol{x}) \cdot \mathrm{S}_\delta f(\boldsymbol{x})] = \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \mathbb{T}^n, \boldsymbol{z} \sim B^n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x} + \delta\boldsymbol{z})].$$

In the soundness analysis of our algorithm we will consider the superlevel sets of a certain Lipschitz function g; i.e., $g^{>t} = \{x : g(x) > t\}$. The coarea formula connects the expected value of the surface area of superlevel sets and the gradient of the function g. Citations for the following form of the theorem include [AFP00, Theorem 2.93], [EG92, Section 3.4.3].[3]

[2] Or $\mathcal{C}^\infty$ boundary if desired, assuming $n \geq 2$.

[3] These state the result when the domain of g is an open subset of $\mathbb{R}^n$, rather than $\mathbb{T}^n$. However the result continues to hold for $\mathbb{T}^n$.

Theorem 2.7. *Let $g : \mathbb{T}^n \to \mathbb{R}$ be Lipschitz and let $\psi : \mathbb{R}^n \to [0, \infty)$ be bounded. Then*

$$\int_{\mathbb{T}^n} \psi(x)|\nabla g(x)|\, dx = \int_{-\infty}^{\infty} \left(\int_{g^{-1}(t)} \psi(x)\, d\mathcal{H}^{n-1}(x) \right) dt.$$

Corollary 2.8. *Let $g : \mathbb{T}^n \to \mathbb{R}$ be Lipschitz and let ϕ be a bounded probability density function on $\mathbb{R}$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbb{T}^n}[\phi(g(\boldsymbol{x}))|\nabla g(\boldsymbol{x})|] = \mathop{\mathbf{E}}_{\boldsymbol{t}\sim\phi}[\mathrm{surf}(g^{>\boldsymbol{t}})].$$

Here $g^{>\boldsymbol{t}} = \{x \in \mathbb{T}^n : g(x) > \boldsymbol{t}\}$, and this is a "set of finite perimeter" with probability 1.

Proof. Apply Theorem 2.7 with $\psi = \phi \circ g$. On the right-hand side we get

$$\int_{-\infty}^{\infty} \left(\int_{g^{-1}(t)} \phi(t)\, d\mathcal{H}^{n-1}(x) \right) dt = \mathop{\mathbf{E}}_{\boldsymbol{t}\sim\phi}[\mathcal{H}^{n-1}(\{x : g(x) = \boldsymbol{t}\})].$$

But it is known for Lipschitz g that for almost all t the set $g^{>t}$ is of finite perimeter, having surface area equal to $\mathcal{H}^{n-1}(\{x : g(x) = \boldsymbol{t}\})$ (see [AFP00, Theorem 3.40], [Mag12, Theorem 18.1], [EG92, Section 5.5]). □

3 The algorithm and its completeness

We begin with the algorithm that gives us our main theorem (Theorem 1.4).

Given black-box membership access to $F \subseteq \mathbb{T}^n$, as well as parameters $A > 0$ and $0 < \epsilon, \eta < 1/2$:

1. Define $\delta = \frac{\sqrt{\eta}}{\beta_n \cdot A}\epsilon$. (In fact, any $\delta = \Theta(\frac{\sqrt{n}\sqrt{\eta}}{A}\epsilon)$ would be acceptable.)
2. Let $\widetilde{\mathbf{NS}_\delta[f]}$ be an empirical estimate of $\mathbf{NS}_\delta[f]$ computed using $t = \frac{C}{\eta^{2.5}\epsilon}$ samples. (Here C is a large universal constant to be specified later.)
3. Accept if and only if $\widetilde{\mathbf{NS}_\delta[f]} \le \frac{n}{n+1} \cdot \beta_n \cdot A \cdot \delta \cdot (1 + \eta)$ (equivalently, $\frac{n}{n+1} \cdot \sqrt{\eta} \cdot \epsilon \cdot (1 + \eta)$).

We now prove the completeness of this tester:

Theorem 3.1 (Completeness)**.** *Let $F \subseteq \mathbb{T}^n$ be such that ∂F is a piecewise-$\mathcal{C}^1$ surface. Assume that* $\mathrm{surf}(F) \le A$. *Then the tester accepts with probability at least* $9/10$.

Proof. Let $\boldsymbol{r}$ be the random variable distributed as the length of a random vector $\boldsymbol{z} \sim B^n$. It is well known and easy to verify that $\mathbf{E}[\boldsymbol{r}] = \frac{n}{n+1}$. Now

$$\mathbf{NS}_\delta[f] = \mathop{\mathbf{E}}_{\boldsymbol{r}}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\mathbb{T}^d, \boldsymbol{v}\sim S^{d-1}}[f(\boldsymbol{x}) \ne f(\boldsymbol{x} + \delta\boldsymbol{r}\boldsymbol{v})]\right] \le \mathop{\mathbf{E}}_{\boldsymbol{r}}\left[\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{v}}[N(\boldsymbol{x}, \delta\boldsymbol{r}\boldsymbol{v}, F)]\right],$$

where $N(\boldsymbol{x}, \delta\boldsymbol{r}\boldsymbol{v}, F)$ is as in Theorem 2.3. Here we used the fact that if a needle's endpoints are on opposite sides of F then the needle must cross ∂F at least once. Applying Theorem 2.3 we get

$$\mathbf{NS}_\delta[f] \le \mathop{\mathbf{E}}_{\boldsymbol{r}}[\beta_n \cdot \mathrm{surf}(F) \cdot \delta\boldsymbol{r}] = \beta_n \cdot \mathrm{surf}(F) \cdot \delta \cdot \frac{n}{n+1} \le \frac{n}{n+1} \cdot \beta_n \cdot A \cdot \delta =: \mu.$$

It remains to show that the empirical estimate $\widetilde{\mathbf{NS}_\delta[f]}$ will not exceed this by a factor of more than $(1 + \eta)$ except with probability at most $1/10$. Noting that $\mu = \Theta(\sqrt{\eta}\epsilon)$, a standard Chernoff bound implies that

$$\mathbf{Pr}[\widetilde{\mathbf{NS}_\delta[f]} > \mu(1 + \eta)] \le \exp(-\mu t\eta^2/3) = \exp(-\Theta(\eta^{2.5}\epsilon t)) \le 1/10$$

provided the constant C in t's definition is sufficiently large. □

4 Soundness

The goal of this section is to prove the soundness portion of Theorem 1.4. The first step is to analyze the Lipschitz constant of $\mathrm{S}_\delta f$.

Proposition 4.1. *Let $f : \mathbb{T}^n \to [-1, 1]$ be measurable and let $\delta > 0$. Then $g = \mathrm{S}_\delta f$ is (L_n/δ)-Lipschitz, where $L_1 = 1$, $L_2 = \frac{4}{\pi}$, $L_3 = \frac{3}{2}$, and in general*

$$L_n = 2\frac{\mathrm{vol}(B^{n-1})}{\mathrm{vol}(B^n)} = \frac{2}{\sqrt{\pi}}\frac{\Gamma(n/2+1)}{\Gamma(n/2+1/2)} = \sqrt{\tfrac{2}{\pi}}\sqrt{n} + O(\tfrac{1}{\sqrt{n}}).$$

Proof. Let $x, y \in \mathbb{T}^n$ be at distance $\lambda \le \frac{1}{2}\mathrm{diam}(\mathbb{T}^n)$. Let D_x be the probability density on $\mathbb{T}^n$ of $x + \boldsymbol{z}$, where $\boldsymbol{z} \sim B^n$; similarly define D_y. Since $g(x)$ (respectively, $g(y)$) is the average of f under D_x (respectively, D_y) and since f's range is $[-1, 1]$, it follows that $|g(x) - g(y)| \le 2d_{\mathrm{TV}}(D_x, D_y)$. By the data processing inequality this bound can only increase if we interpret D_x as being the uniform distribution on the Euclidean ball $B_x = x + B^n \subset \mathbb{R}^n$ (and similarly for D_y), rather than taking the distribution mod $\mathbb{Z}^n$. Thus

$$|g(x) - g(y)| \le 2\frac{\mathrm{vol}(B_x \setminus B_y)}{\mathrm{vol}(B_x)}.$$

To estimate this, without loss of generality let x be the origin and let $y = (-\lambda, 0, \dots, 0)$. By simple geometric considerations we have

$$\frac{\mathrm{vol}(B_x \setminus B_y)}{\mathrm{vol}(B_x)} = \frac{\mathrm{vol}(B_x \cap W)}{\mathrm{vol}(B_x)},$$

where $W = \{z \in \mathbb{R}^n : |z_1| \le \frac{\lambda}{2}\}$; note that this still holds even if $\delta > 2\lambda$ (in which case $B_x \setminus B_y = B_x$). In turn we can upper-bound $\mathrm{vol}(B_x \cap W)$ by the volume of a cylinder of width λ and base equal to the $(n-1)$-dimensional radius-δ ball $\{z \in B_x : z_1 = 0\}$. Thus we deduce

$$|g(x) - g(y)| \le 2\frac{\lambda \cdot \delta^{n-1}\mathrm{vol}(B^{n-1})}{\delta^n \mathrm{vol}(B^n)} = \lambda \cdot \frac{L_n}{\delta},$$

as needed. □

The key lemma we'll need to establish the soundness is the following:

Lemma 4.2. *Let $f : \mathbb{T}^n \to \{-1, 1\}$ be the indicator of a measurable set $F \subseteq \mathbb{T}^n$ with smooth boundary. Then for any $0 < \eta < \frac{1}{2}$, there exists a set $G \subseteq \mathbb{T}^n$ satisfying*

$$\mathrm{surf}(G) \le 2L_n \cdot \tfrac{1}{\delta} \cdot \mathbf{NS}_\delta[f] \cdot (1 + 2\eta),$$
$$\mathrm{vol}(F \triangle G) \le \tfrac{2}{\sqrt{\eta}} \cdot \mathbf{NS}_\delta[f].$$

Here L_n is as in Proposition 4.1.

Proof. We use the probabilistic method. Let $g = \mathrm{S}_\delta f$, which by Proposition 4.1 is (L_n/δ)-Lipschitz. The essential idea (which doesn't quite work) is to let $\boldsymbol{t}$ be the random variable drawn from the triangular distribution on $[-1, 1]$ with pdf $\phi(t) = 1 - |t|$, and define $\boldsymbol{G}$ to be the set $g^{>\boldsymbol{t}} = \{x \in \mathbb{T}^n : g(x) > \boldsymbol{t}\}$. As the reader will see shortly, by applying the coarea formula Corollary 2.8 we would get $\mathbf{E}[\mathrm{surf}(\boldsymbol{G})] \le 2L_n \cdot \frac{1}{\delta} \cdot \mathbf{NS}_\delta[f]$, and it is also not hard to show that $\mathbf{E}[\mathrm{vol}(F \triangle \boldsymbol{G})] \le O(\mathbf{NS}_\delta[f])$. Thus in expectation $\boldsymbol{G}$ satisfies the desired properties (with the volume difference being even smaller than claimed). The catch is that we need both to hold simultaneously. It is possible to achieve both with a factor-$O(\frac{1}{\eta})$ loss on $\mathrm{vol}(F \triangle G)$ using Markov's inequality. However it is more effective to slightly change the distribution on $\boldsymbol{t}$. To be precise, we truncate $\boldsymbol{t}$ to the interval $[-1 + \sqrt{\eta}, 1 - \sqrt{\eta}]$; i.e., we let it be drawn according to the following probability density:

$$\phi_\eta(t) = \begin{cases} \frac{1-|t|}{1-\eta} & \text{for } -1 + \sqrt{\eta} \le t \le 1 - \sqrt{\eta} \\ 0 & \text{else.} \end{cases}$$

Now defining $\boldsymbol{G} = g^{>\boldsymbol{t}}$ and applying the coarea formula Corollary 2.8 we obtain

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{t}\sim\phi_\eta}[\mathrm{surf}(\boldsymbol{G})] &= \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbb{T}^n}[\phi_\eta(g(\boldsymbol{x}))|\nabla g(\boldsymbol{x})|] \\
&\leq (L_n/\delta)\cdot \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbb{T}^n}[\phi_\eta(g(\boldsymbol{x}))] && (g \text{ is } (L_n/\delta)\text{-Lipschitz}) \\
&\leq (L_n/\delta)\cdot \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbb{T}^n}\left[\frac{1-|g(\boldsymbol{x})|}{1-\eta}\right] && (\text{definition of } \phi_\eta) \\
&< (L_n/\delta)\cdot(1+2\eta)\cdot \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathbb{T}^n}[1-f(\boldsymbol{x})g(\boldsymbol{x})] && (0<\eta<1/2,\ f(\boldsymbol{x})\in\{-1,1\}) \\
&= (L_n/\delta)\cdot(1+2\eta)\cdot 2\mathbf{NS}_\delta[f].
\end{aligned}$$

Thus $\boldsymbol{G}$ satisfies the claimed surface area bound in expectation. We will complete the proof by showing that $\boldsymbol{G}$ satisfies the claimed volume bound with probability 1; this shows there exists a set of finite perimeter G satisfying both bounds.

Note that every outcome of $\boldsymbol{t} \sim \phi_\eta$ satisfies $|\boldsymbol{t}| \leq 1-\sqrt{\eta}$. Thus whenever $x \in F\triangle\boldsymbol{G}$ — i.e., $f(x) \neq \mathrm{sgn}(g(\boldsymbol{x}) - \boldsymbol{t})$ — we have

$$f(x)g(x) \leq 1-\sqrt{\eta} \qquad \tfrac{1}{2} - \tfrac{1}{2}f(x)g(x) \geq \tfrac{1}{2}\sqrt{\eta}.$$

We deduce that with probability 1 over the choice of $\boldsymbol{G}$,

$$1_{x\in F\triangle\boldsymbol{G}} \leq \tfrac{2}{\sqrt{\eta}}\left(\tfrac{1}{2} - \tfrac{1}{2}f(x)g(x)\right) \qquad \text{for all } x\in\mathbb{T}^n.$$

Taking expectations over $\boldsymbol{x} \in \mathbb{T}^n$ completes the proof. □

We can now complete the proof of soundness of our tester.

Theorem 4.3 (Soundness)**.** *Suppose that the tester accepts with probability at least $1/10$. Then there exists a set $G \subseteq \mathbb{T}^n$ (which is, in fact, a finite union of polytopes) such that:*

$$\begin{aligned}
\mathrm{surf}(G) &\leq (\kappa_n + O(\eta))\cdot A,\\
\mathrm{vol}(F\triangle G) &\leq O(\epsilon),
\end{aligned}$$

where

$$\kappa_n = 2\cdot\frac{n}{n+1}\cdot L_n\cdot\beta_n = 2(n+1)\left[\frac{\binom{n}{n/2-1/2}}{2^n}\right]^2$$

and the $O(\cdot)$'s hide small universal constants.

Proof. By a Chernoff bound very similar to the one in the completeness Theorem 3.1, if $\mathbf{NS}_\delta[f]$ fails to satisfy

$$\mathbf{NS}_\delta[f] \leq \tfrac{n}{n+1}\cdot\beta_n\cdot A\cdot\delta\cdot(1+2\eta) = \tfrac{n}{n+1}\cdot\sqrt{\eta}\cdot\epsilon\cdot(1+2\eta) \tag{3}$$

then the tester rejects with probability at least $9/10$. Thus (3) must hold, whence applying Theorem 4.2 lets us deduce that there exists a set of finite perimeter G satisfying:

$$\begin{aligned}
\mathrm{surf}(G) &\leq 2L_n\cdot\tfrac{n}{n+1}\cdot\beta_n\cdot A\cdot(1+O(\eta)) = (\kappa_n+O(\eta))\cdot A,\\
\mathrm{vol}(F\triangle G) &\leq \tfrac{2}{\sqrt{\eta}}\cdot\tfrac{n}{n+1}\cdot\sqrt{\eta}\cdot\epsilon\cdot(1+2\eta) \leq O(\epsilon).
\end{aligned}$$

□

Remark 4.4. One can approximate the set G in the proof above, arbitrarily well by a set with piecewise-affine boundary, if desired (see remarks in Notation 2.2). Also, for any $x \in G$, as $g(x) = \mathrm{S}_\delta f(x) \geq 1-\sqrt{\eta}$, if $x \notin F$, at least $\sqrt{\eta}/2$ fraction of its δ-neighborhood will lie in F. Thus, in this case, $\mathrm{dist}(x, F) = O(\sqrt{n\eta}\epsilon/A)$.

5 Testing Gaussian surface area

In this section we show how to modify our arguments to obtain similar results for testing *Gaussian* surface area. We need analogous definitions of surface area of sets and the coarea formula in the Gaussian case; here the reader is referred to [CK01] for more details.

Notation 5.1. Let $F \subseteq \mathbb{R}^n$ be a set with finite Gaussian perimeter. We write $\mathrm{vol}_\gamma(F)$ for its Gaussian volume and $\mathrm{surf}_\gamma(F)$ for its Gaussian perimeter (surface area). For sets F with piecewise-$\mathcal{C}^1$ boundary, $\mathrm{surf}_\gamma(F)$ coincides with $\mathcal{H}_\gamma^{n-1}(\partial F)$, where $\mathcal{H}_\gamma^{n-1}$ denotes the $(n-1)$-dimensional Hausdorff measure computed using n-dimensional Gaussian volume as the reference measure.

Theorem 5.2 (Coarea Formula)**.** *Let $g, \psi : \mathbb{R}^n \to \mathbb{R}$ be bounded functions. In addition, assume that g is Lipschitz. Then*

$$\int_{\mathbb{R}^n} \psi(x)|\nabla g(x)|\, d\gamma_n = \int_{-\infty}^{\infty} \left(\int_{g^{-1}(t)} \psi(x)\, d\mathcal{H}_\gamma^{n-1}(x) \right) dt,$$

where γ_n is the n-dimensional Gaussian measure.

Corollary 5.3. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a bounded Lipschitz function and let ϕ be a bounded probability density function on $\mathbb{R}$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\gamma_n}[\phi(g(\boldsymbol{x}))|\nabla g(\boldsymbol{x})|] = \mathop{\mathbf{E}}_{\boldsymbol{t}\sim\phi}[\mathrm{surf}_\gamma(g^{>\boldsymbol{t}})].$$

Here $g^{>\boldsymbol{t}}$ is a "set of finite Gaussian perimeter" with probability 1.

Next we need the appropriate notion of "noise sensitivity", which is defined using the Ornstein–Uhlenbeck operator:

Definition 5.4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be integrable under Gaussian measure. For $0 < \delta < \pi/2$ we define $\mathrm{GS}_\delta f : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathrm{GS}_\delta f(x) = \mathop{\mathbf{E}}_{\boldsymbol{z}\sim\gamma_n}[f((\cos\delta)x + (\sin\delta)\boldsymbol{z})].$$

In case f has range ± 1 we also define

$$\mathbf{GNS}_\delta[f] = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\gamma_n}[\tfrac{1}{2} - \tfrac{1}{2} f(\boldsymbol{x})\mathrm{GS}_\delta f(\boldsymbol{x})] = \mathop{\mathbf{Pr}}_{\boldsymbol{x},\boldsymbol{z}\sim\gamma_n}[f(\boldsymbol{x}) \neq f((\cos\delta)x + (\sin\delta)\boldsymbol{z})].$$

This can be connected to Gaussian surface area, in a manner similar to the Buffon's Needle Theorem 2.3, using the following one-sided inequality due to Ledoux [Led94]:[4]

Theorem 5.5. *Let $F \subseteq \mathbb{R}^n$ have piecewise-$\mathcal{C}^1$ boundary. Then $\mathbf{GNS}_\delta[f] \leq \sqrt{\frac{2}{\pi}} \cdot \delta \cdot \mathrm{surf}(F)$.*

Next we give our testing algorithm. The high level idea of the tester is the same as in the Euclidean case before, but requires a minor adjustment in case the desired value for δ isn't less than $\pi/2$. This minor adjustment uses the GAUSSIANVOLUMETEST algorithm from Theorem 1.6.

Given black-box membership access to $F \subseteq \mathbb{R}^n$, as well as parameters $A > 0$ and $0 < \epsilon, \eta < 1/2$:

1. Define $\delta = \frac{\sqrt{\eta}}{A}\epsilon$.
2. If $\delta \geq \pi/2$ then execute GAUSSIANVOLUMETEST(F, $V = \frac{2}{3}\epsilon$, $\tau = \frac{1}{2}$), GAUSSIANVOLUMETEST(F^c, $V = \frac{2}{3}\epsilon$, $\tau = \frac{1}{2}$), and accept if either accepts. Otherwise…
3. Let $\widetilde{\mathbf{GNS}_\delta}[f]$ be an empirical estimate of $\mathbf{GNS}_\delta[f]$ computed using $t = \frac{C}{\eta^{2.5}\epsilon}$ samples. (Here C is a large universal constant to be specified later.)
4. Accept if and only if $\widetilde{\mathbf{GNS}_\delta}[f] \leq \sqrt{\frac{2}{\pi}} \cdot \delta \cdot A \cdot (1+\eta)$ (equivalently, $\sqrt{\frac{2}{\pi}} \cdot \sqrt{\eta} \cdot \epsilon \cdot (1+\eta)$).

[4] Also credited by Ledoux to Pisier [Pis86]. Ledoux only proved the theorem for sets with $\mathcal{C}^\infty$ boundary; however it holds for general Borel sets F when surface area is defined as $\mathcal{H}_\gamma^{n-1}(\partial F)$ [KOS08, Footnote 1], and this coincides with $\mathrm{surf}_\gamma(F)$ when F has piecewise-$\mathcal{C}^1$ boundary [CK01].

We remark that the GaussianVolumeTest algorithm only uses $O(1/\epsilon)$ queries, so the test still needs $O(\frac{1/\eta^{2.5}}{\epsilon})$ queries in total.

We now analyze our tester. Let's first dispense with the case of $\delta \geq \pi/2$; i.e., $A \leq (2/\pi)\sqrt{\eta}\epsilon$.

Lemma 5.6. *Suppose that $\delta \geq \pi/2$. Then the tester accepts with probability at least $9/10$ if $\mathrm{surf}_\gamma(F) \leq A$. If the tester accepts with probability at least $1/10$, then, there exists $G \subseteq \mathbb{R}^n$ with $\mathrm{surf}_\gamma(G) \leq A$ and $\mathrm{vol}_\gamma(F \triangle G) \leq \epsilon$.*

Proof. The second statement ("soundness") is almost trivial. If the tester accepts with probability at least $1/10$ then at least one of GaussianVolumeTest(F), GaussianVolumeTest(F^c), must accept with probability at least $1/20$. By the soundness of GaussianVolumeTest, this means that either $\mathrm{vol}_\gamma(F) \leq \frac{2}{3}\epsilon \cdot (1 + \frac{1}{2}) = \epsilon$ or else $\mathrm{vol}_\gamma(F) \geq 1 - \epsilon$. In the former case we may take $G = \emptyset$; in the latter case we may take $G = \mathbb{R}^n$. Both of these sets have $\mathrm{surf}(G) = 0 \leq A$.

Regarding the first statement of the theorem ("completeness"), suppose $\mathrm{surf}(F) \leq A$, and thus $\mathrm{surf}(F) \leq (2/\pi)\sqrt{\eta}\epsilon \leq \frac{\sqrt{2}}{\pi}\epsilon$. Just from the Gaussian isoperimetric inequality [Tsi76, Bor75] it holds that

$$\mathrm{surf}(F) \geq 2\sqrt{\tfrac{2}{\pi}} \cdot \mathrm{vol}_\gamma(F) \cdot \mathrm{vol}_\gamma(F^c)$$

(see [Led94, (8.25)]). Thus we conclude

$$\mathrm{vol}_\gamma(F) \cdot \mathrm{vol}_\gamma(F^c) \leq \tfrac{1}{2\sqrt{\pi}}\epsilon \leq \tfrac{1}{3}\epsilon$$

and hence either $\mathrm{vol}_\gamma(F) \leq \frac{2}{3}\epsilon$ or $\mathrm{vol}_\gamma(F^c) \leq \frac{2}{3}\epsilon$. In either case, the tester will accept with probability at least $99/100 \geq 9/10$, by the completeness of GaussianVolumeTest. □

In the remainder of this section we analyze the tester under the assumption that $\delta < \pi/2$. We begin by showing that completeness follows easily from Theorem 5.5.

Theorem 5.7 (Completeness)**.** *Let $F \subseteq \mathbb{R}^n$ be such that ∂F is a piecewise-$\mathcal{C}^1$ surface. Assume that $\mathrm{surf}_\gamma(F) \leq A$. Then the tester accepts with probability at least $9/10$.*

Proof. By Theorem 5.5, $\mathbf{GNS}_\delta[f] \leq \sqrt{\frac{2}{\pi}} \cdot \delta \cdot A =: \mu$. Since $\mu = \Theta(\sqrt{\eta}\epsilon)$, the claim now follows exactly as in Theorem 3.1. □

It remains to show the soundness of the tester. As before, we begin with analyzing the Lipschitz constant of $\mathrm{GS}_\delta[f]$.

Proposition 5.8. *Let $f : \mathbb{R}^n \to [-1, 1]$ be measurable and let $0 < \delta < \pi/2$. Then $h = \mathrm{GS}_\delta f$ is $(\sqrt{\frac{2}{\pi}}\cot\delta)$-Lipschitz.*

Proof. Let $x, y \in \mathbb{R}^n$ be at distance λ. Let D_x be the probability density of $(\cos\delta)x + (\sin\delta)\boldsymbol{z}$, where $\boldsymbol{z} \sim \gamma_n$, and similarly define D_y. As in the proof of Proposition 4.1, it suffices to show

$$d_{\mathrm{TV}}(D_x, D_y) \leq \tfrac{1}{\sqrt{2\pi}}(\cot\delta)\lambda.$$

Now D_x and D_y are n-dimensional Gaussian distributions, with standard deviation $\sin\delta$ in each direction and means at distance $(\cos\delta)\lambda$. Without loss of generality we may assume x is the origin and y is $(-(\cos\delta)\lambda, 0, \ldots, 0)$. Then the event achieving the total variation distance between D_x and D_y is the halfspace $H = \{z \in \mathbb{R}^n : z_1 \geq -\frac{1}{2}(\cos\delta)\lambda\}$, and

$$\mathop{\mathbf{Pr}}_{D_x}[H] - \mathop{\mathbf{Pr}}_{D_y}[H] = \mathop{\mathbf{Pr}}_{\boldsymbol{z}\sim\gamma_1}[-\tfrac{1}{2}(\cos\delta)\lambda \leq (\sin\theta)\boldsymbol{z} < \tfrac{1}{2}(\cos\delta)\lambda] \leq \frac{1}{\sqrt{2\pi}\sin\delta} \cdot (\cos\delta)\lambda = \tfrac{1}{\sqrt{2\pi}}(\cot\delta)\lambda,$$

since the pdf of a 1-dimensional Gaussian with standard deviation $\sin\delta$ is bounded above by $\frac{1}{\sqrt{2\pi}\sin\delta}$. This completes the proof. □

Next, we have the following lemma which is crucial for the soundness analysis. Its proof is identical to the proof of Lemma 4.2, using Corollary 5.3 in place of Corollary 2.8 and Proposition 5.8 in place of Proposition 4.1.

Lemma 5.9. *Let $f : \mathbb{R}^n \to \{-1, 1\}$ be the indicator of a measurable set $F \subseteq \mathbb{R}^n$. Then for any $0 < \eta < \frac{1}{2}$, there exists a set $G \subseteq \mathbb{R}^n$ satisfying*

$$\begin{aligned} \mathrm{surf}_\gamma(G) &\leq 2\sqrt{\tfrac{2}{\pi}} \cot \delta \cdot \mathbf{GNS}_\delta[f] \cdot (1 + 2\eta), \\ \mathrm{vol}_\gamma(F \triangle G) &\leq \tfrac{2}{\sqrt{\eta}} \cdot \mathbf{GNS}_\delta[f]. \end{aligned}$$

We can now present the analysis of soundness of our Gaussian surface area tester, with approximation factor $\kappa = \frac{4}{\pi} + O(\eta)$:

Theorem 5.10 (Soundness)**.** *Suppose that the tester accepts with probability at least $1/10$. Then there exists a set $G \subseteq \mathbb{R}^n$ such that:*

$$\begin{aligned} \mathrm{surf}_\gamma(G) &\leq (\tfrac{4}{\pi} + O(\eta)) \cdot A, \\ \mathrm{vol}_\gamma(F \triangle G) &\leq O(\epsilon), \end{aligned}$$

where the $O(\cdot)$'s hide small universal constants.

Proof. The proof is identical to that of Theorem 4.3 except for the numerical calculations. By a Chernoff bound we must have

$$\mathbf{NS}_\delta[f] \leq \sqrt{\tfrac{2}{\pi}} \cdot \delta \cdot A \cdot (1 + 2\eta) = \sqrt{\tfrac{2}{\pi}} \cdot \sqrt{\eta} \cdot \epsilon \cdot (1 + 2\eta). \tag{4}$$

Applying Theorem 5.9 now gives us G with $\mathrm{vol}_\gamma(F \triangle G) \leq O(\epsilon)$ and

$$\mathrm{surf}(G) \leq \frac{4}{\pi} \cdot \frac{\delta}{\tan \delta} \cdot A \cdot (1 + 2\eta)^2,$$

which completes the proof because $\frac{\delta}{\tan \delta} \leq 1$ for $0 < \delta < \pi/2$. □

6 Conclusion

In this paper, we have given property testing algorithms which test surface area up to any approximation factor $\kappa > \frac{4}{\pi}$ using $O(1/\epsilon)$ queries. It works for both Euclidean and Gaussian settings, with query complexity and the approximation factor independent of the dimension. A straightforward open question is whether one can achieve approximation factor κ arbitrarily close to 1 in dimensions $n > 1$ (perhaps at the cost of slightly increasing the query complexity). It is possible that our algorithm already achieves this and that it's only our analysis which is deficient. Alternatively, one may seek a lower bound — for example, is it possible that testing surface area in $\mathbb{T}^2$ to approximation factor 1.01 requires $\omega(1/\epsilon)$ or even $(1/\epsilon)^{\omega(1)}$ queries?

Acknowledgments

We are very grateful to Adam Klivans for his significant contributions in the early stages of this work. We would also like to thank Boris Bukh, Brittany Terese Fasy, and Mathoverflow user bezirsk for helpful information.

References

[AFNS09] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009. 1.2

[AKK$^+$05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing reed-muller codes. *Information Theory, IEEE Transactions on*, 51(11):4032–4039, 2005. 1.2

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Clarendon Press Oxford, 2000. 2, 2, 2

[BBBY12] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 21–30. IEEE, 2012. (document), 1.2, 1.1, 1.3, 1.4, 2

[BNN06] Mikhail Belkin, Hariharan Narayanan, and Partha Niyogi. Heat flow and a faster algorithm to compute the surface area of a convex body. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 47–56. IEEE, 2006. 1

[Bla09] Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 151–158. ACM, 2009. 1.2

[BO10] Eric Blais and Ryan O'Donnell. Lower bounds for testing function isomorphism. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 235–246. IEEE, 2010. 1.2

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993. 1, 1.2

[Bor75] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones Mathematicae*, 30(2):207–216, 1975. 5

[CK01] EA Carlen and C Kerce. On the cases of equality in bobkov's inequality and gaussian rearrangement. *Calculus of Variations and Partial Differential Equations*, 13(1):1–18, 2001. 5, 4

[CFRC07] Antonio Cuevas, Ricardo Fraiman, and Alberto Rodríguez-Casal. A nonparametric approach to the estimation of lengths and surface areas. *The Annals of Statistics*, 35(3):1031–1051, 2007. 1

[CS01] Artur Czumaj and Christian Sohler. Property testing with geometric queries. In *Algorithm-słESA 2001*, pages 266–277. Springer, 2001. 1.2

[CSZ00] Artur Czumaj, Christian Sohler, and Martin Ziegler. Property testing in computational geometry. In *Algorithms-ESA 2000*, pages 155–166. Springer, 2000. 1.2

[DLM$^+$08] Ilias Diakonikolas, Homin K Lee, Kevin Matulef, Rocco A Servedio, and Andrew Wan. Efficiently testing sparse gf (2) polynomials. In *Automata, Languages and Programming*, pages 502–514. Springer, 2008. 1.2

[DGH98] Martin Dyer, Peter Gritzmann, and Alexander Hufnagel. On the complexity of computing mixed volumes. *SIAM Journal on Computing*, 27(2):356–400, 1998. 1

[EG92] Lawrence C Evans and Ronald F Gariepy. *Measure theory and fine properties of functions*, volume 5. CRC press, 1992. 2, 2

[GGR98] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998. 1, 1.2

[GLS88] Martin Grötschel, László Lovász, and Alexander Schrijver. Geometric algorithms and combinatorial optimization. 1988. 1

[KR98] Michael Kearns and Dana Ron. Testing problems with sub-learning sample complexity. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 268–279. ACM, 1998. (document), 1.1, 1.2, 1.3

[KOS08] Adam R Klivans, Ryan O'Donnell, and Rocco A Servedio. Learning geometric concepts via gaussian surface area. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 541–550. IEEE, 2008. (document), 1.1, 4

[Led94] Michel Ledoux. Semigroup proofs of the isoperimetric inequality in euclidean and gauss space. *Bulletin des sciences mathématiques*, 118(6):485–510, 1994. 5, 5

[LYZ+10] Yu-Shen Liu, Jing Yi, Hu Zhang, Guo-Qin Zheng, and Jean-Claude Paul. Surface area estimation of digitized 3d objects using quasi-monte carlo methods. *Pattern Recognition*, 43(11):3900–3909, 2010. 1

[Mag12] Francesco Maggi. *Sets of finite perimeter and geometric variational problems: an introduction to Geometric Measure Theory*, volume 135. Cambridge University Press, 2012. 2

[Pis86] Gilles Pisier. *Probabilistic methods in the geometry of Banach spaces*. Springer, 1986. 4

[Ras03] Sofya Raskhodnikova. Approximate testing of visual properties. In Sanjeev Arora, Klaus Jansen, Jos D. P. Rolim, and Amit Sahai, editors, *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques, 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2003 and 7th International Workshop on Randomization and Approx*, volume 2764 of *Lecture Notes in Computer Science*, pages 370–381. Springer, 2003. 1

[Ron08] Dana Ron. Property testing: A learning theory perspective. *Foundations and Trends® in Machine Learning*, 1(3):307–402, 2008. 1.1, 1.2

[RS11] Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM Journal on Discrete Mathematics*, 25(4):1562–1588, 2011. 1.2

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996. 1

[San04] Luis A Santaló. *Integral geometry and geometric probability*. Cambridge University Press, 2004. 2

[Tsi76] VS Tsirel'son. The density of the distribution of the maximum of a gaussian process. *Theory of Probability & Its Applications*, 20(4):847–856, 1976. 5

[YC96] Chee K Yap and Ee-Chien Chang. Issues in the metrology of geometric tolerancing. *Robotics Motion and Manipulation*, pages 393–400, 1996. 1