Testing Convexity of Figures Under the Uniform Distribution[*]

Piotr Berman[†] Meiram Murzabulatov[‡] Sofya Raskhodnikova[§]

Abstract

We consider the following basic geometric problem: *Given $\epsilon \in (0, 1/2)$, a 2-dimensional black-and-white figure is ϵ-far from convex if it differs in at least an ϵ fraction of the area from every figure where the black object is convex. How many uniform and independent samples from a figure that is ϵ-far from convex are needed to detect a violation of convexity with constant probability?* This question arises in the context of designing property testers for convexity.

We show that $\Theta(\epsilon^{-4/3})$ uniform samples (and the same running time) are necessary and sufficient for detecting a violation of convexity in an ϵ-far figure and, equivalently, for testing convexity of figures with 1-sided error. Our algorithm beats the $\Omega(\epsilon^{-3/2})$ lower bound by Schmeltz [32] on the number of samples required for learning convex figures under the uniform distribution. It demonstrates that, with uniform samples, we can check if a set is approximately convex much faster than we can find an approximate representation of a convex set.

1 Introduction

Convexity is a fundamental property of geometric objects that plays an important role in algorithms, learning, optimization, and image processing. Some algorithms require that their input is a convex set. However, it is infeasible to check whether an infinite (or a very large) set is indeed convex. How quickly can we check whether it is approximately convex? Can it be done faster than learning an approximate representation of a convex set?

Property testing [31, 15] is a formal study of fast algorithms that determine whether a given object approximately satisfies the desired property. There is a line of work on property testing and sublinear algorithms for geometric convexity[1] and other visual properties (see [25, 24, 30, 18, 19, 20] and references therein). Previous works on testing geometric convexity [25, 24, 30] assume that the tester can query an arbitrary point in the input and find out whether it belongs to the object.

We study the problem of property testing convexity of 2-dimensional figures with only uniform and independent samples from the input. A figure (U, C) consists of a compact convex universe $U \subseteq \mathbb{R}^2$ and a measurable subset $C \subseteq U$. The set C can be thought of as a black object on a white background $U \setminus C$. A figure (U, C) is convex iff C is convex. The relative distance between two figures (U, C) and (U, C') over the same universe is the probability of the symmetric difference

[*]A preliminary full version of this paper appears in RSA [6], https://doi.org/10.1002/rsa.20797.

[†]Computer Science and Engineering, Pennsylvania State University, USA; berman@cse.psu.edu.

[‡]Computer Science and Engineering, Pennsylvania State University, USA; murzabulatov85@gmail.com. This author was supported in part by NSF award CCF-1422975/1832228 and NSF CAREER award CCF-0845701.

[§]Computer Science, Boston University, USA; sofya@bu.edu. This author was supported in part by NSF award CCF-1422975/1832228 and NSF CAREER award CCF-0845701.

[1]Property testing of convexity (and submodularity) of functions has also been investigated [22, 33, 27, 8, 7].

between them under the uniform distribution on U. A figure (U, C) is ϵ-far from convex if the relative distance from (U, C) to every convex figure (U, C') over the same universe is at least ϵ.

Definition 1.1. *Given a proximity parameter $\epsilon \in (0, 1/2)$ and error probability $\delta \in (0, 1)$, a (1-sided error) ϵ-tester for convexity accepts if the figure is convex and rejects with probability at least $1 - \delta$ if the figure is ϵ-far from convex*[2]*. A tester is* uniform *if it accesses its input via uniform and independent samples from U, each labeled with a bit indicating whether it belongs to C.*

Our goal is to determine the smallest number of samples necessary and sufficient for ϵ-testing convexity with a uniform tester.

An easy upper bound for this problem can be obtained from a connection between (proper) PAC-learning and property testing, established by Goldreich et al. [15], and the work of Schmeltz [32] who gives a PAC-learner of convex d-dimensional sets under the uniform distribution. Specifically, for two dimensions, he shows that $\Theta(\epsilon^{-3/2})$ samples are necessary and sufficient[3]. In other words, Schmeltz [32] shows that it suffices to take $O(\epsilon^{-3/2})$ uniform and independent samples from a convex shape of unit area, so that the convex hull of these samples has area at least $1 - \epsilon$ with probability at least 2/3; and moreover, for a disk, $\Omega(\epsilon^{-3/2})$ samples are necessary to satisfy this requirement.

We prove that $\Theta(\epsilon^{-4/3})$ uniform samples are necessary and sufficient for ϵ-testing 2-dimensional convexity with 1-sided error. The running time[4] of our algorithm is asymptotically the same as its sample complexity, specifically, $O(\epsilon^{-4/3})$. Our algorithm beats the $\Omega(\epsilon^{-3/2})$ lower bound (on the sample complexity and thus also on the running time) for learning convex figures under the uniform distribution. It shows that, with uniform samples, we can check if a set is approximately convex much faster than we can find an approximate representation of a convex set.

We also consider convexity testing in the pixel model of [25]. The input representation in that model differs from ours only in that the images are discretized, whereas we consider continuous figures. In the pixel model, an image is specified by an $n \times n$ matrix of Boolean pixel values, representing a discretization of a black-and-white image in $[0, 1]^2$. An algorithm is *adaptive* if it is allowed to query arbitrary entries in the matrix and its queries depend on answers to previous queries. In [25], an adaptive tester for convexity[5] that makes $O(\epsilon^{-2})$ queries is presented. Our upper and lower bounds hold for the pixel model, provided that n is sufficiently large to ensure that every convex area we consider in our analysis has some pixels (i.e., non-zero probability mass when we sample uniformly from the $n \times n$ matrix). The theorem statements for the pixel model (Theorems 5.3 and 5.6) and their proofs are given in Section 5.

Our techniques. Our uniform tester for convexity of 2-dimensional figures is the natural one: it computes the convex hull of sampled black points and rejects iff it contains a sampled white

[2]If δ is not specified, it is assumed to be 1/3. By standard arguments, the error probability can be reduced from 1/3 to an arbitrarily small δ by running the tester $O(\log 1/\delta)$ times.

[3]The PAC learner in [32] is not distribution-free—it works only with respect to the uniform distribution. The VC dimension of convexity, even in two dimensions, is infinite, so convexity is not PAC-learnable under arbitrary distributions.

[4]We analyze the running time in the model where a sample from the input can be taken in unit time. Observe that the running time of an algorithm is bounded from below by its sample complexity.

[5]The VC dimension of convexity of $n \times n$ images is $\Theta(n^{2/3})$, since this is the maximum number of vertices of a convex lattice polygon in an $n \times n$ lattice [2]. Therefore, proper PAC-learners for convex images in the pixel model (that work with respect to all distributions) cannot have complexity independent of n.

point. In other words, it rejects only if it finds a convexity violation. How many points are needed to witness such a violation? The smallest number of points is three: a white point between two black points on the same line. However, a uniform tester is unlikely to sample three points on the same line. If the points are in general position, the smallest number is four: three black and one white in the triangle formed by the three black points. A natural way to exploit this in the analysis is to divide the figure into different parts (which we call patterns) with four regions each, such that we are likely to sample a 4-point witness of non-convexity from the corresponding regions of some pattern. However, the higher the number of regions in each pattern from which we require the tester to sample at least one point, the more samples it needs.

To reduce the number of regions in the patterns, we use a *central point* defined in terms of the Ham Sandwich cut of black points[6]. Such cuts have been studied extensively (see, e.g., [12, p. 356] and [21]), for example, in the context of range queries. Specifically, a *central point* is the intersection of two lines that partition the figure into four regions, each with black area at least $\epsilon/4$. A central point is overwhelmingly likely to end up in the convex hull of sampled black points. So, even though the central point itself is not likely to be sampled, it becomes a de facto part of a witness that comes nearly for free. Conditioned on the central point indeed being in the convex hull of sampled black points, our witness only needs 3 additional points: two black and one white, such that the white point is in the triangle formed by the two black points and the central point. This will ensure that the white point is in the convex hull of sampled black points, that is, a violation of convexity is detected.

The main technical part of the analysis is finding 3-region patterns in every figure that is ϵ-far from convex, such that the algorithm is likely to sample a 3-point witness from at least one of the patterns. Intuitively, a *pattern* consists of a triangle W with vertices v_1, u, v_2, where u is a central point, and disjoint sets B_1 and B_2 of black points, where B_1 (respectively, B_2) is contained in the region formed by the rays uv_1 and v_2v_1 (respectively, uv_2 and v_1v_2). (Patterns are formally defined in Definition 3.3 and depicted in Figure 3.2.) Observe that black points $b_1 \in B_1$ and $b_2 \in B_2$ and a white point $w \in W$, where regions B_1, B_2 and W come from the same pattern, form a witness triple.

We find patterns in the figure such that all black regions of all patterns are *disjoint*. The analysis proceeds in two phases: in the first phase, called the *recoloring phase*, we consider the case when the figure contains many patterns (with mutually disjoint black sets) whose W-regions have large white area; in the second phase, called the *sweeping phase*, we deal with the remaining case. In the former case, we show that the tester is likely to sample points $b_1 \in B_1$ and $b_2 \in B_2$ for some pattern with black regions B_1 and B_2 and, conditioned on that event happening, also sample a point w from the W-region of the same pattern.

The heart of the analysis is the sweeping phase. To partition the black area of the image into black regions of the patterns, we move "sweeping lines" from the boundary of the figure, chopping off small black area of carefully chosen size. At the end, we get a polygon and construct, roughly speaking, a pattern for each of its vertices and each of its sides. Crucially, the central point u remains inside the polygon. To prove this, we use the essential feature of the central point u: specifically, that each line passing through u has black area at least $\epsilon/4$ on each side. We show that W-regions of the patterns jointly cover a fraction of the area proportional to ϵ. We conclude the proof by showing that the tester is likely to sample points $b_1 \in B_1$, $b_2 \in B_2$ and $w \in W$ for some

[6]Our central points are related to the well studied centerpoints [12] and Tukey medians [35]. The guarantee for a centerpoint is that every line that passes through it creates a relatively balanced cut.

pattern with regions B_1, B_2 and W in our construction.

We remind the reader that the coloring and sweeping phases are only used in the analysis. Our algorithm is extremely simple and natural.

To prove our lower bound on the sample complexity of uniform convexity testing with 1-sided error, we construct hard instances, for which a uniform tester for convexity needs to get a 3-point witness, with points coming from different specified regions, in order to detect a violation of convexity. Intuitively, the fact that the number of points in a witness is also 3, as in the analysis of the algorithm, allows us to get a matching lower bound.

Related Work in Property Testing. We already mentioned work on testing geometric convexity [25, 24, 30] in a model similar to ours, but where the tester can query arbitrary points in the input. There is another line of work on testing geometric properties, initiated by Czumaj, Sohler, and Ziegler [11] and Czumaj and Sohler [10], where the input is a set of points represented by their coordinates. The allowed queries and the distance measures on the input space considered in those works are different from ours. The most related problem to ours is that of testing whether points, represented by their coordinates, are in convex position or far from having that property (for example, in the sense that at least an ϵ fraction of points has to be changed to ensure that they are in a convex position). In [10], several sophisticated distance measures and powerful queries to the input are considered. For example, a *range query*, given a range and a natural number i, returns the i-th point in the range. Chazelle and Seshadhri [9], in another related work, give a property tester for convexity of polygons represented by doubly-linked lists of their edges. In contrast to these works, we consider only extremely simple access to the input, measure the distance between figures by the area on which they differ, and can deal with continuous figures.

Uniform testers were first considered by Goldreich, Goldwasser, and Ron [15] and systematically studied by Goldreich and Ron [16] and Fischer et al. [14]. (In these papers, uniform testers are called "sample-based".) In particular, [16, 14] show that certain types of query-based testers yield uniform testers with sublinear (but dependent on the size of the input) sample complexity.

Subsequently to this work, [4] gave an approximation algorithm for the distance to convexity. This algorithm achieves $\pm\epsilon$ approximation with $O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$ uniform samples in time $O(\frac{1}{\epsilon^8})$. Another subsequent article [5] obtained an adaptive, $O(1/\epsilon)$ time algorithm for testing convexity of 2-dimensional figures. It also showed that $\Omega(\epsilon^{-5/4})$ uniform samples are required to test convexity, even with 2-sided error.

Related Work in Computational Geometry. The random process of sampling uniform and independent points from a convex body has been studied extensively. (We stress that, in our problem, the input figure is not guaranteed to be convex. Instead, we are trying to distinguish convex figures from those that are far from convex.) The expected number of vertices of a convex hull of n such samples is well understood. For example, in 2 dimensions, it is $O(n^{1/3})$ when the object is a disk [28] and $O(k \log n)$ when the object is a convex k-gon [29]. (See also [17] for simple proofs of these statements.) Bárány and Füredi [3] analyze the probability that n points chosen from the d-dimensional unit ball are in the convex position. Eldan [13] shows that no algorithm can approximate the volume of a convex body in $\mathbb{R}^d$, with high probability and up to a constant factor, when provided only with a polynomial in d number of random points.

2 Preliminaries on Poissonization

The analysis of our algorithm uses a technique called *Poissonization* [34], in which one modifies a probabilistic experiment to replace a fixed quantity (e.g., the number of samples) with a variable one that follows a Poisson distribution. This breaks up dependencies between different events, and makes the analysis tractable. The Poisson distribution with parameter $\lambda \geq 0$, denoted $\mathrm{Po}(\lambda)$, generates each value $x \in \mathbb{N}$ with probability $e^{-\lambda}\lambda^x/x!$. The expectation and variance of a random variable distributed according to $\mathrm{Po}(\lambda)$ are both λ.

Definition 2.1. *A* Poisson-s tester *is a uniform tester that takes a random number of samples distributed as* $\mathrm{Po}(s)$.

The following lemma is paraphrased from [26, Lemma 5.3], except that the terminology is adjusted to fit in with our application. The proofs from [26] work nearly verbatim. Even though part (a) is not stated in [26], the proof for this part is similar to the proof of part (b). We use part (a) to analyze our algorithm and part (b) to prove lower bounds on the sample complexity (so, we do not need a statement about the running time in part (b)).

We use $[k]$ to denote the integer set $\{1, \ldots, k\}$.

Lemma 2.1 (Poissonization Lemma)**.**

(a) *Uniform algorithms can simulate Poisson algorithms. Specifically, for every Poisson-s tester $\mathcal{A}$ for property $\mathcal{P}$ with error probability δ, there is a uniform tester $\mathcal{A}'$ for $\mathcal{P}$ that uses at most $2s$ samples and has error probability at most $\delta + 4/s$. Moreover,*

- *if $\mathcal{A}$ has 1-sided error, so does $\mathcal{A}'$;*
- *if $\mathcal{A}$ runs in time $t(x)$ when it takes x samples, then $\mathcal{A}'$ has running time $O(t(2s))$.*

(b) *Poisson algorithms can simulate uniform algorithms. Specifically, for every uniform tester $\mathcal{A}$ for property $\mathcal{P}$ that uses at most s samples and has error probability δ, there is a Poisson-$2s$ tester $\mathcal{A}'$ for $\mathcal{P}$ with error probability at most $\delta + 4/s$. Moreover,*

- *if $\mathcal{A}$ has 1-sided error, so does $\mathcal{A}'$.*

(c) *Let Ω be a sample space from which a Poisson-s algorithm makes uniform draws. Suppose we partition Ω into sets $\Omega_1, \ldots, \Omega_k$ (e.g., these sets can correspond to disjoint areas of the figure from which points are sampled), where each outcome is in set Ω_i with probability p_i for $i \in [k]$. Let X_i be the total number of samples in Ω_i seen by the algorithm. Then X_i is distributed as $\mathrm{Po}(p_i \cdot s)$.* ***Moreover, random variables X_i are mutually independent for all $i \in [k]$.***

3 Uniform Tester for Convexity

In this section, we give our optimal uniform convexity tester for figures.

Theorem 3.1 (Main Theorem)**.** *There is a uniform (1-sided error) ϵ-tester for convexity with sample and time complexity $O(\epsilon^{-4/3})$.*

Proof. We start by reducing the problem to the special case when the universe U is an axis-aligned rectangle of unit area. Consider a convex 2-dimensional set U'. By Lemma A.1 (see Appendix A), U' is contained in a rectangle U whose area is at most twice the area of U'. If we consider figures (U, C) instead of (U', C), relative distances between figures decrease by at most a factor of 2. An ϵ-tester that has access to (U', C) can simply simulate and $\epsilon/2$-tester that works on (U, C). As a result, the new tester will have the same asymptotic complexity as the simulated tester. Therefore, we can assume w.l.o.g. that U is a rectangle. Finally, note that if U does not have unit area, the figure can be rescaled, and if U is not axis-aligned, the figure can be rotated.

By the Poissonization Lemma (Lemma 2.1), it is sufficient to prove that there is a 1-sided error Poisson-s convexity tester with $s = O(\epsilon^{-4/3})$, error probability $\delta \leq 0.333$, and linear running time in the number of samples[7]. By standard arguments, such a tester can be obtained from a 1-sided error Poisson-s convexity tester with $s = O(\epsilon^{-4/3})$, but with error probability $\delta \leq 0.33$ and *expected* linear running time in the number of samples. Our Poisson convexity tester satisfying the latter requirements is Algorithm 1. To make the algorithm and its analysis easier to visualize, we color points in C black and points in $U \setminus C$ white. (In the analysis, we recolor some of the black points to make them violet.)

Algorithm 1: Uniform ϵ-*tester* for convexity (when U is an axis-aligned rectangle).

---

input : parameter $\epsilon \in (0, 1/2)$;
access to uniform and independent samples from (U, C).

Set $s = 50\epsilon^{-4/3}$. Sample $\mathrm{Po}(s)$ points from U uniformly and independently at random.

Run bucket sort with s bins to sort the sampled black points by the x-coordinate; call the resulting sorted list S_B. Similarly, compute the sorted list S_W of the sampled white points.

```
// Check if the convex hull of S_B contains a pixel from S_W.
```

Use Andrew's monotone chain convex hull algorithm [1] to compute $\mathrm{UH}(S_B)$ and $\mathrm{LH}(S_B)$, the upper and the lower hulls of S_B, respectively, sorted by the x-coordinate.

Merge sorted lists $S_W, \mathrm{UH}(S_B)$ and $\mathrm{LH}(S_B)$ to determine for each point w in S_W its left and right neighbors in $\mathrm{UH}(S_B)$ and $\mathrm{LH}(S_B)$. If w lies between the corresponding line segments of the upper and lower hulls, **reject**.

Accept.

Sample and Time Complexity. Algorithm 1 samples $q = \mathrm{Po}(s)$ points, where $s = O(\epsilon^{-4/3})$. Since the x-coordinates of the sampled q points are distributed uniformly in the interval corresponding to the length of the rectangle U, they can be sorted in expected time $O(q)$ by subdividing this interval into s subintervals of equal length, and using them as buckets in the bucket sort algorithm. Andrew's monotone chain algorithm finds the convex hull of a set of q sorted points in time $O(q)$. Since the expectation of q is $O(\epsilon^{-4/3})$, Algorithm 1 runs in expected time $O(\epsilon^{-4/3})$. By the discussion preceding the algorithm, we can get a uniform algorithm with the worst case running time $O(\epsilon^{-4/3})$ and with a slightly larger error δ than in Algorithm 1.

[7]Our proof works for sufficiently small ϵ. Suppose an algorithm works for all $\epsilon \leq \epsilon_0$. For $\epsilon > \epsilon_0$, we can run it with parameter ϵ_0 in constant time and obtain the required guarantees, since an ϵ_0-tester is also an ϵ-tester for $\epsilon_0 < \epsilon$.

Correctness. If figure (U, C) is convex, $\mathrm{Hull}(S_B)$ contains only black points, and Algorithm 1 always accepts. From now on, we consider a figure (U, C) that is ϵ-far from convex. We show that Algorithm 1 rejects it with probability at least 0.67.

For a set (region) R, let $\mathrm{Hull}(R)$ denote its convex hull and $A(R)$ denote its area or, equivalently, its probability mass under the uniform distribution of points in U. (It is equivalent because we assumed w.l.o.g. that U has unit area.) For a region R, its area of a certain color (e.g., its black area) is the probability mass of points of that color in R. For example, initially, the black area of R is $A(R \cap C)$.

We start by defining a special point, which belongs, with high probability, to $\mathrm{Hull}(S_B)$ constructed by Algorithm 1.

Definition 3.1 (Central point)**.** *A point is* central *if it is the intersection of two lines such that each of the closed quadrants formed by these lines has black area at least $\epsilon/4$, i.e., the intersection of C and each quadrant has area at least $\epsilon/4$. We say that the two lines* define *this central point.*

For the proof of the next claim, we use the Ham Sandwich Theorem. We only need it for the special case of sets in $\mathbb{R}^2$.

Lemma 3.2 (The Ham Sandwich Theorem in $\mathbb{R}^2$)**.** *Any two measurable sets in a plane can be simultaneously bisected (with respect to their measure) by a line.*

Claim 3.3. If $A(C) \geq \epsilon$, then U contains a central point.

Proof. By a continuity argument, there exists a line that bisects C into two sets of area $A(C)/2$ each. By the Ham Sandwich Theorem (see Lemma 3.2), applied to the two resulting sets, for every such line, there exists another line that bisects both of the resulting sets into sets of area $A(C)/4$ each. By Definition 3.1, the intersection point of the two lines is a central point. □

Since the empty set is convex and (U, C) is ϵ-far from convex, $A(C) \geq \epsilon$. Thus, by Claim 3.3, there is a central point in U. Denote one such point by u. The central point u is fixed throughout the analysis of Algorithm 1. Next, we bound the probability that u is in $\mathrm{Hull}(S_B)$. Note that u is just a point in U, not necessarily a sample point.

Lemma 3.4. *The probability that the central point u is not in Hull(S_B), where S_B is the set of black points sampled by Algorithm 1, is at most 0.01.*

Proof. Let ℓ_1^u and ℓ_2^u be the two lines that define the central point u (see Definition 3.1). If the algorithm samples a black point from each open quadrant formed by ℓ_1^u and ℓ_2^u, then the central point u is in the convex hull of the four points sampled from each quadrant, i.e., it is in the convex hull of all sampled black points. By the Poissonization Lemma 2.1, the number of samples from each quadrant has distribution $\mathrm{Po}(p \cdot s)$, where $p \geq \epsilon/4$. Thus, the probability that the algorithm fails to sample a black point from one particular quadrant is at most $e^{-\epsilon \cdot s/4}$. For $s = 50 \cdot \epsilon^{-4/3}$, we have that $e^{-\epsilon \cdot s/4} \leq e^{-6}$. By a union bound, the probability that the algorithm will not sample a black point from at least one of the four open quadrants is at most $4 \cdot e^{-6} < 0.01$. Thus, the probability that $u \notin \mathrm{Hull}(S_B)$ is at most 0.01. □

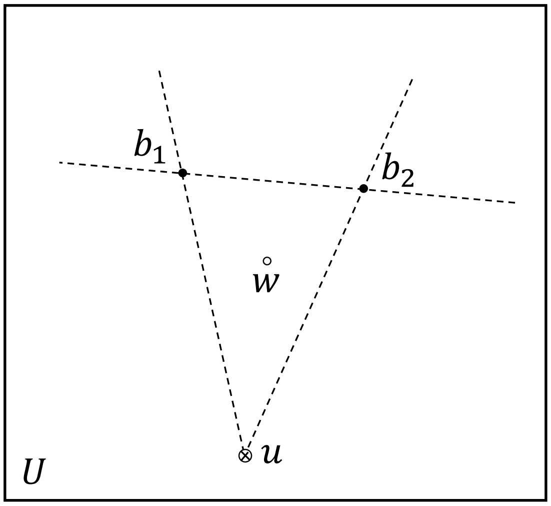

Figure 3.1: A witness triple (b_1, b_2, w).

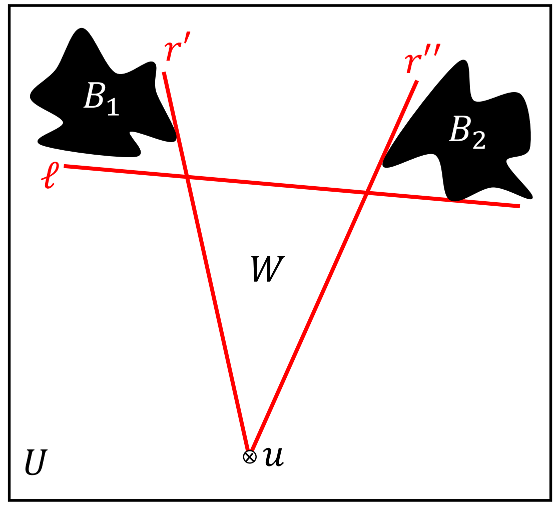

Figure 3.2: A pattern.

Definition 3.2 (A witness triple). *Recall that u is a fixed central point. A triple of points (b_1, b_2, w) is a* witness triple *if b_1 and b_2 are black, and w is a white point contained in the triangle $\triangle ub_1b_2$. (See Figure 3.1. Note that $\triangle ub_1b_2$ could be degenerate, i.e., all three vertices could lie on the same line.)*

If the central point u is indeed contained in the convex hull of all black points sampled by Algorithm 1 and if, in addition, the algorithm samples a witness triple, then the algorithm rejects because it found a white sample w in the convex hull of black samples. By Lemma 3.4, the first event fails to occur with probability at most 0.01. If we get a guarantee that the algorithm fails to sample a witness triple with probability at most 0.32 then, by a union bound, the algorithm fails to reject with probability at most $0.01 + 0.32 = 0.33$, as required.

The required guarantee follows from Propositions 3.6 and 3.8 that we will prove in Sections 3.1 and 3.2, respectively. Both propositions together prove that if the input figure is ϵ-far from convex then the tester samples a witness triple with probability at least 0.68. Proposition 3.6 does it for the case when a certain condition holds (specifically, the number of iterations in the recoloring phase is large) whereas Proposition 3.8 proves it when this condition does not hold. This completes the proof of Theorem 3.1. □

Propositions 3.6 and 3.8 break the analysis into two cases, depending on the number of certain patterns in the input. Patterns are parts of the input from which, intuitively, we are likely to sample a witness triple.

Definition 3.3 (A pattern). *A* pattern *consists of two rays r' and r'', emanating from the central point u, a line ℓ that crosses the two rays, and disjoint sets B_1 and B_2 of black points for which the following conditions hold. The set B_1 (respectively, B_2) has area $t = 0.025 \cdot \epsilon^{3/2}$ and is a subset of the infinite region formed by the line ℓ and the ray r' (respectively, by ℓ and r''). (See Figure 3.2.) If, in addition, the white area of the triangular region W, formed by ℓ, r', and r'', is at least $0.025 \cdot \epsilon^{4/3}$, then the pattern is called a* white-heavy *pattern. Otherwise, it is called a* white-light *pattern.*

Observe that, for any pattern, a point from B_1, a point from B_2, and a white point from W form a witness triple.

Claim 3.5. For a given pattern and $i \in \{1, 2\}$, let E_i be the event that Algorithm 1 samples a point from B_i. Then $\Pr[E_1 \cap E_2] \geq 0.7\epsilon^{1/3}$, for $\epsilon \in (0, 5^{-6})$.

Proof. Recall that $t = 0.025 \cdot \epsilon^{3/2}$. By definition of a pattern, set B_1 has area t. Therefore, by the Poissonization Lemma (Lemma 2.1), $\Pr[E_1] = 1 - e^{-ts} = 1 - e^{-1.25\epsilon^{1/6}}$. Note that $1 - e^{-x} \geq 0.7x$ for $x \in (0, 1/4)$, and that $1.25\epsilon^{1/6} \in (0, 1/4)$ for $\epsilon \in (0, 5^{-6})$. Thus, for $\epsilon \in (0, 5^{-6})$, we get $\Pr[E_1] \geq 0.7 \cdot 1.25\epsilon^{1/6} \geq 0.87 \cdot \epsilon^{1/6}$. The same bound holds for event E_2. Since sets B_1 and B_2 are disjoint, events E_1 and E_2 are independent. Therefore,

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2] \geq (0.87 \cdot \epsilon^{1/6})^2 \geq 0.7\epsilon^{1/3},$$

as required. $\square$

To explain the two cases considered in Propositions 3.6 and 3.8, we describe our analysis in two phases, *recoloring* and *sweeping*.

3.1 The Recoloring Phase

In the recoloring phase of the analysis, we greedily construct a maximal collection of white-heavy patterns whose black sets are disjoint (but white sets can overlap). We show (in Proposition 3.6) that if the number of patterns in this collection is large, then the tester samples a witness triple with the desired probability. During the construction, we change the color of the points in the black sets of the selected patterns to violet. This is done to ensure that no white-heavy patters remain in the sweeping phase of the analysis.

While there is a white-heavy pattern in the figure, we repeat the following mental experiment.

1. Choose a white-heavy pattern. Let B_1 and B_2 be the associate sets of black points.
2. If it is iteration i of the mental experiment, let $V_1^i = B_1$ and $V_2^i = B_2$.
3. Recolor violet all points in V_1^i and V_2^i (so that they are not used in subsequent iterations and the next phase of the analysis).

Proposition 3.6. Suppose the input figure is ϵ-far from convex. If the number of iterations in the recoloring phase is at least $5 \cdot \epsilon^{-1/3}$, then the tester samples a witness triple with probability at least 0.68.

Proof. Recall that a point from B_1, a point from B_2 and a white point from W of the same pattern form a witness triple. We analyze the probability that we sample such a triple from the same white-heavy pattern.

For each iteration i of the mental experiment, let E_B^i denote the event that a point is sampled both from V_1^i and from V_2^i. By Claim 3.5, $\Pr[E_B^i] \geq 0.7 \cdot \epsilon^{1/3}$. In each iteration, we recolor only black points that were not recolored previously. Therefore, sets V_1^i, V_2^i, V_1^j, and V_2^j are disjoint for all i, j $(i \neq j)$, i.e., events E_B^i are mutually independent for all i. There are at least $5 \cdot \epsilon^{-1/3}$ events E_B^i, one for each iteration of the mental experiment. By Claim 3.7,

$$\Pr\left[\bigcup_i E_B^i\right] \geq 1 - e^{-(0.7\epsilon^{1/3} \cdot 5\epsilon^{-1/3})} > 0.96.$$

That is, with probability at least 0.96, there exists an iteration i' for which event $E_B^{i'}$ occurs.

Let E_W be the event that a white point is sampled from the region W of the white-heavy pattern used in iteration i'. The set of white points in W has area at least $0.025 \cdot \epsilon^{4/3}$, and this set is disjoint from $V_1^{i'}$ and $V_2^{i'}$, which are the sets of points that were black in the input. Thus, $E_B^{i'}$ and E_W are independent. The probability of E_W is at least $1 - e^{-0.025 \cdot \epsilon^{4/3} \cdot s} \geq 1 - e^{-1.25} > 0.71$. So, $\Pr[E_B^{i'} \cap E_W] > 0.96 \cdot 0.71 > 0.68$. That is, we sample a witness triple with probability at least 0.68.

Claim 3.7. Let $E_1, E_2, \ldots, E_r$ be independent events such that $\Pr[E_j] \geq q_j$ for all $j \in [r]$. Let $q = \sum_{j \in [r]} q_j$. Then $\Pr[\bigcup_{j \in [r]} E_j] \geq 1 - e^{-q}$.

Proof. From the given information, $\Pr[\overline{E}_i] \leq 1 - q_j \leq e^{-q_j}$. Since all events $\overline{E}_j$ are independent, $\Pr[\bigcap_1^r \overline{E}_j] \leq \prod_1^r e^{-q_j} = e^{-q}$. Therefore, $\Pr[\bigcup_1^r E_j] \geq 1 - e^{-q}$, as claimed. ☐

This completes the proof of Proposition 3.6. ☐

3.2 The Sweeping Phase

In this section, we prove Proposition 3.8, the main technical component in the proof of Theorem 3.1.

Proposition 3.8. Suppose the input figure is ϵ-far from convex. If the number of iterations in the recoloring phase is less than $5 \cdot \epsilon^{-1/3}$, then the tester samples a witness triple with probability at least 0.68.

Proof. If the recoloring phase has less than $5 \cdot \epsilon^{-1/3}$ iterations then, for $\epsilon < 5^{-6}$, the violet area (that was black in the original input) is at most

$$5 \cdot \epsilon^{-1/3}(2 \cdot 0.025 \cdot \epsilon^{3/2}) = 0.25 \cdot \epsilon^{7/6} < 0.05 \cdot \epsilon. \tag{1}$$

In the sweeping phase of the analysis, we iteratively construct a set of *sweeping* lines L. Each line $\ell \in L$ is associated with a set of black points S_ℓ of area at most $4t$. (Recall from Definition 3.3 that $t = 0.025 \cdot \epsilon^{3/2}$ is the area of each black set in a pattern. Also recall that some of the originally black points became violet in the recoloring phase and thus are no longer black.) The set S_ℓ lies in the half-plane defined by ℓ that does not contain the central point u. Sets S_ℓ associated with different lines ℓ are disjoint. Lines ℓ whose sets S_ℓ have area exactly $4t$ are collected in L^*. For each such line ℓ, we define an *anchor* point p_ℓ. Later, we use the sets S_ℓ of lines $\ell \in L^*$ to create white-light patterns whose associated regions B_1, B_2, are all disjoint from each other and whose regions W jointly cover a large white area. Each S_ℓ will be partitioned into four sets B_i for the patterns. The anchor points are used to partition sets S_ℓ and to choose subsequent sweeping lines. We describe the construction of sweeping lines next. We start by constructing a *bounding rectangle* R formed by initial sweeping lines and then add more sweeping lines.

Recall that w.l.o.g. we can assume that U is a rectangle. Now we construct lines $\ell_0, \ell_1, \ell_2, \ell_3$ that form a *bounding rectangle* R inside U. Let ℓ_0 and ℓ_2 be the horizontal lines such that $A(S_{\ell_0}) = A(S_{\ell_2}) = 4t$, where S_{ℓ_0} (respectively, S_{ℓ_2}) denote the set of all black points above ℓ_0 (respectively, below ℓ_2). Initially, $L = L^* = \{\ell_0, \ell_2\}$.

Next, we define anchor points. Their position depends on the location of the central point u. As we prove later (in Lemma 3.11), none of our sweeping lines pass u. In particular, u is inside the bounding rectangle and each line ℓ that we will construct separates u from S_ℓ.

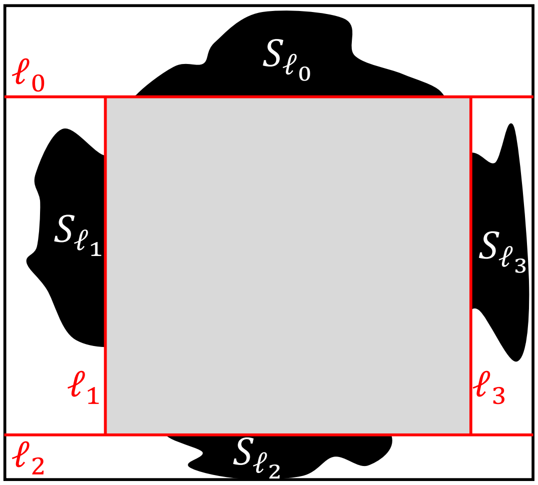

Figure 3.3: Bounding rectangle R.

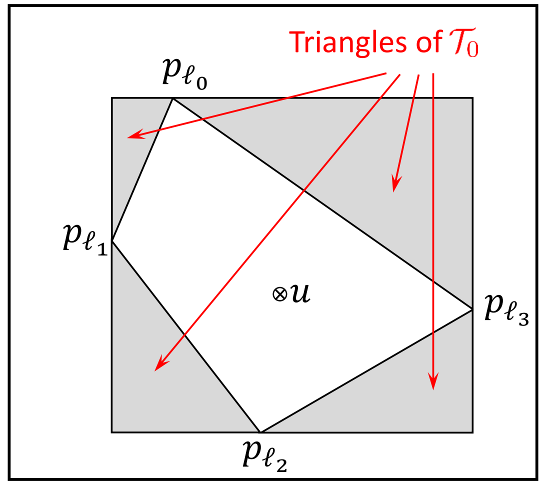

Figure 3.4: Triangles of $\mathcal{T}_0$ when $|P_0| = 4$.

Definition 3.4 (Anchor points)**.** *Consider a line ℓ that does not contain the central point u. Define H_ℓ^u (resp., H_ℓ) to be the closed half-plane formed by ℓ that contains (resp., does not contain) u. For a set S of points in H_ℓ, the* anchor point *of S on ℓ is the intersection of the line ℓ and the ray emanating from u that bisects the set S into two sets of equal area. For a sweeping line $\ell \in L^*$ and the associated set S_ℓ, let p_ℓ denote the anchor point of S_ℓ on ℓ.*

The initial set of anchor points is denoted P_0 and, at first, is equal to $\{p_{\ell_0}, p_{\ell_2}\}$. Now we define the vertical lines ℓ_1 and ℓ_3. The set S_{ℓ_1} (respectively, S_{ℓ_3}) will be the set of all black points to the left of ℓ_1 (respectively, to the right of ℓ_3) between ℓ_0 and ℓ_2. See Figure 3.3. Intuitively, for $i \in \{0, 1, 2, 3\}$, we move the line ℓ_i in parallel starting from the boundary of U and stop moving it immediately when it "sweeps" a set of black points (not "swept" by previous lines) whose area is equal to $4t$. However, the lines ℓ_1 and ℓ_3 will stop before "sweeping" black area $4t$ if they encounter an anchor point. Specifically, for $i = 1, 3$, we require that $A(S_{\ell_i}) \leq 4t$ and that the half-plane $H_{\ell_i}^u$ must contain both anchor points p_{ℓ_0} and p_{ℓ_2}. Let ℓ_1 and ℓ_3 be the vertical lines with the maximum $A(S_{\ell_1})$ and $A(S_{\ell_3})$, respectively, and that satisfy these requirements. For $i = 1, 3$, if $A(S_{\ell_i}) < 4t$, then the line ℓ_i is added only to L. Otherwise, ℓ_i is added to both L and L^* and the anchor point p_{ℓ_i} of S_{ℓ_i} on ℓ_i (given by Definition 3.4) is added to P_0. This completes the construction of P_0.

The bounding rectangle R is formed by the lines $\ell_0, \ell_1, \ell_2, \ell_3$. Note that $2 \leq |P_0| \leq 4$. Let $\mathcal{T}_0$ be the set of (at most four) triangles formed by removing the (possibly degenerate) quadrilateral $\mathrm{Hull}(P_0)$ from the rectangle R. Figures 3.4–3.6 illustrate how $\mathcal{T}_0$ can look like when the size of P_0 is $4, 3$, and 2, respectively.

Definition 3.5 (Line and ray notation.)**.** *For two points x and y, let $r(x, y)$ denote the ray that emanates from x and passes through y, and let $\ell(x, y)$ denote the line through x and y.*

Initially, let P be equal to P_0. The set P will represent the final set of anchor points. We describe a procedure that completes the construction of L, L^*, and P by inductively building sets $\mathcal{T}_i$, starting from the set $\mathcal{T}_0$, defined before. (Recall that this construction is needed only in the analysis, not in the algorithm.)

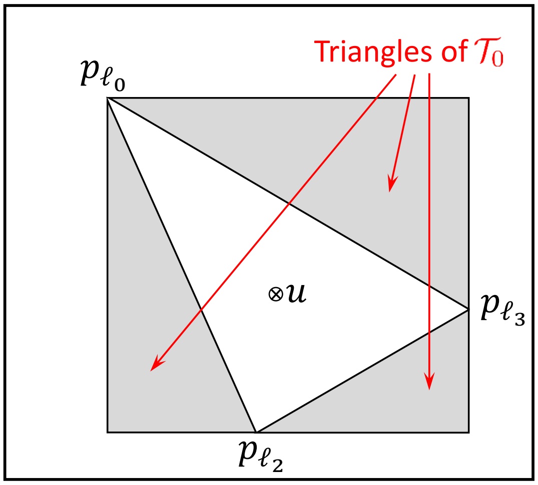

Figure 3.5: Triangles of $\mathcal{T}_0$ when $|P_0| = 3$.

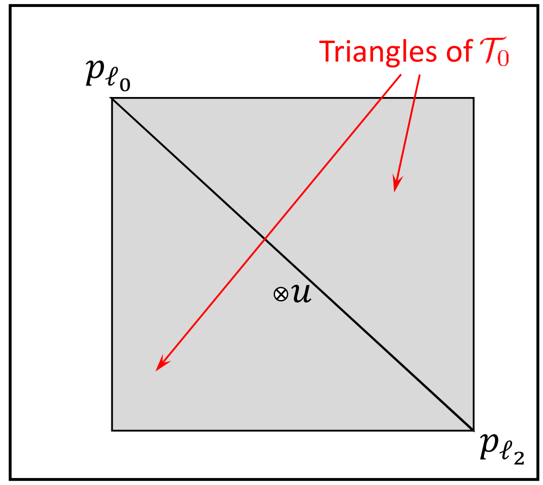

Figure 3.6: Triangles of $\mathcal{T}_0$ when $|P_0| = 2$.

1. Let $m = \log(2/\epsilon)/2$ (w.l.o.g. assume that $\log(2/\epsilon)/2$ is an integer[8]).
2. Initially, $\mathcal{T}_i = \emptyset$, for every $i \in [m]$, and sets L, L^*, P and $\mathcal{T}_0$ are as defined earlier.
3. For every $i = 1, 2, \dots, m$ and every triangle $T \in \mathcal{T}_{i-1}$, do the following:
   (a) Let v be the only vertex of T that is not in P; let $p', p'' \in P$ be its other two vertices.
   (b) If the black area in $T = \triangle vp'p''$ is less than $4t$, then let $\ell = \ell(p', p'')$, define S_ℓ to be the set of black points in $\triangle vp'p''$ and add ℓ to L.
   (c) Otherwise, let ℓ be the line parallel to the base $p'p''$ that intersects the sides vp' and vp'' at v' and v'', respectively, such that the black area of $\triangle vv'v''$ is $4t$ (see Figure 3.7). Let S_ℓ be the set of black points in $\triangle vv'v''$. Let p_ℓ be the anchor point of S_ℓ on ℓ. Add the line ℓ to L and L^*, point p_ℓ to P, and triangles $\triangle p_\ell p'v'$ and $\triangle p_\ell p''v''$ to $\mathcal{T}_i$.
4. This completes the construction of L, L^*, P and $\mathcal{T}_i$, for every $i \in [m]$.

Intuitively, we move a line starting from the vertex v towards the base $p'p''$, keeping it parallel to the base. We stop moving it when it reaches the side $p'p''$ or when it "sweeps" a black area $4t$ in $\triangle vp'p''$.

The goal of sweeping is to eventually construct patterns. Black sets for the patterns will be obtained from the sets S_ℓ for lines in L^*, whereas white regions for the patterns will come from Hull(P). The area between the polygon formed by the sweeping lines and Hull(P) is "uninvestigated" and not useful in the construction of patterns. In order to reduce the uninvestigated area quickly (with a few sweeps), we take sweeping lines *parallel* to the bases of uninvestigated triangles. After sweeping, only triangles in $\mathcal{T}_m$ remain uninvestigated.

Lemma 3.9 (Area of triangles in $\mathcal{T}_m$)**.** *The sum of the areas of all triangles in $\mathcal{T}_m$ is at most $\epsilon/2$.*

[8]If $\epsilon \in (1/2^j, 1/2^{j-2})$ for some odd j, to ϵ-test $\mathcal{P}$ it is enough to ϵ'-test $\mathcal{P}$ with $\epsilon' = 1/2^j$ since $\epsilon' < \epsilon$.

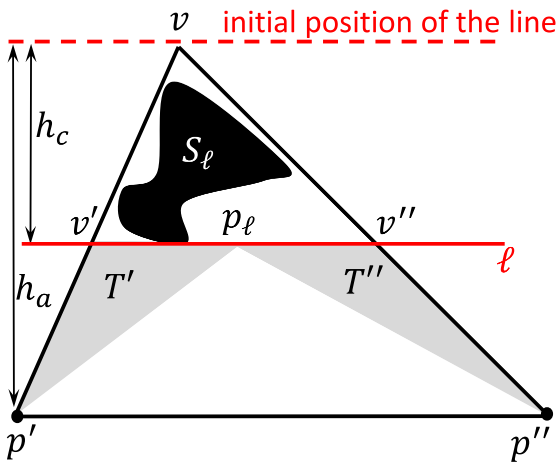

Figure 3.7: Constructing triangles of $\mathcal{T}_i$ in a triangle of $\mathcal{T}_{i-1}$.

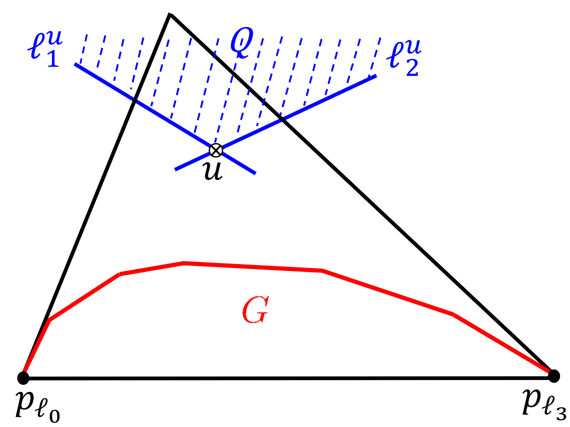

Figure 3.8: An illustration of the central point u outside G.

Proof. Fix $i \in [m]$. Consider a triangle $T \in \mathcal{T}_{i-1}$ and the two triangles $T', T'' \in \mathcal{T}_i$ obtained in the procedure that constructs sets $\mathcal{T}_j$, for $j \in [m]$. Recall that in the procedure, triangle $T = \triangle vp'p''$, where v is the only vertex of T that is not in P and $p', p'' \in P$ are its other two vertices. Moreover, $T' = \triangle p_\ell p' v'$ and $T'' = \triangle p_\ell p'' v''$, where p_ℓ is the anchor point on the line ℓ that is parallel to the base $p'p''$, whereas v' and v'' are the intersection points of ℓ and the sides vp' and vp'', respectively. (See Figure 3.7.)

Claim 3.10. For the triangles $T \in \mathcal{T}_{i-1}$ and $T', T'' \in \mathcal{T}_i$, defined above,

$$A(T') + A(T'') \leq \frac{A(T)}{4}.$$

Proof. Let a and c be the lengths of the sides $p'p''$ and $v'v''$, respectively. Let h_a and h_c be the heights of triangles T and $\triangle vv'v''$, respectively. See Figure 3.7. Then

$$A(T') + A(T'') = A(T) - A(\triangle vv'v'') - A(\triangle p_\ell p' p'') = \frac{ah_a}{2} - \frac{ch_c}{2} - \frac{a(h_a - h_c)}{2} = \frac{(a-c)h_c}{2}.$$

Since triangles T and $\triangle vv'v''$ are similar, $\frac{h_c}{h_a} = \frac{c}{a}$. Thus,

$$\frac{A(T') + A(T'')}{A(T)} = \frac{(a-c)h_c}{2} \cdot \frac{2}{ah_a} = (1 - \frac{c}{a})\frac{h_c}{h_a} = (1 - \frac{c}{a})\frac{c}{a} \leq \frac{1}{4},$$

as claimed. The last inequality holds since $(1-x)x$ is maximized when $x = 1/2$. □

By Claim 3.10, $\sum_{T \in \mathcal{T}_i} A(T) \leq \frac{1}{4} \sum_{T' \in \mathcal{T}_{i-1}} A(T')$ for all $i \in [m]$. The total area of all triangles in T_0 is at most 1. Thus, the total area of all triangles in $\mathcal{T}_m$ is at most $\frac{1}{4^m} = \frac{\epsilon}{2}$, completing the proof of Lemma 3.9. □

Next, we show that the point u cannot be swept by the sweeping lines in L.

Lemma 3.11. *For each line $\ell \in L$, let H_ℓ^* be the closed half-plane that is defined by the line ℓ and that does not contain the set S_ℓ. Let G be the polygon formed by the intersection of the half-planes H_ℓ^*, for all lines $\ell \in L$. Then the central point u is in G.*

Proof. For each line $\ell \in L$, the area of S_ℓ is at most $4 \cdot 0.025\epsilon^{3/2} = 0.1 \cdot \epsilon^{3/2}$. Recall that the central point u is the intersection of lines ℓ_1^u and ℓ_2^u such that, initially, each quadrant defined by these lines has black area at least 0.25ϵ. By (1), the violet area is less than 0.05ϵ and, thus, after the recoloring phase, each of the quadrants has black area at least 0.2ϵ.

The total black area outside of the bounding rectangle R is at most $4 \cdot 0.1\epsilon^{3/2} = 0.4\epsilon^{3/2} < 0.2\epsilon$ for sufficiently small ϵ. Thus, the point u is inside the bounding rectangle R. If u is also in the (possibly degenerate) quadrilateral Hull(P_0) defined by the initial anchor points, as shown in Figures 3.4–3.6, then it is also in G since Hull$(P_0) \subseteq G$. Therefore, the lemma holds in this case.

Consider the remaining case, when the point u is inside one of the initial triangles $T \in \mathcal{T}_0$. Let $L_T = \{\ell \in L \mid S_\ell \cap T \neq \emptyset\}$, i.e., L_T is the set of all sweeping lines that sweep black area inside the triangle T. We need to show that the point u belongs to $\bigcap_{\ell \in L_T} H_\ell^*$, i.e., none of the lines in L_T sweeps the point u.

For the sake of contradiction, assume $u \notin \bigcap_{\ell \in L_T} H_\ell^*$, that is, the point u is outside of G. Then one of the four open quadrants formed by ℓ_1^u and ℓ_2^u does not intersect G. Let Q denote such a quadrant. Note that all black points in Q are swept in the sweeping phase, i.e., each of them belongs to S_ℓ for some $\ell \in L$.

Claim 3.12. The black area swept in Q, that is, the area of $(\bigcup_{\ell \in L_T} S_\ell) \bigcap Q$, is less than 0.2ϵ.

Proof. Quadrant Q does not intersect with the base of the triangle T. Thus, the only triangle of $\mathcal{T}_0$ the quadrant Q intersects with is T. See Figure 3.8. Therefore, Q can be swept only by the sweeping lines in L_T and at most two of the initial sweeping lines $\ell_0, \ell_1, \ell_2, \ell_3$, i.e., by at most $|L_T| + 2$ lines. There are at most 2^{i-1} lines in L_T for each $i \in [m]$, where $m = \log(2/\epsilon)/2$. Thus, for sufficiently small ϵ,

$$|L_T| + 2 \leq \sum_{i=1}^{m} 2^{i-1} = 2^m + 1 = \frac{\sqrt{2}}{\sqrt{\epsilon}} + 1 \leq \frac{2}{\sqrt{\epsilon}}.$$

Each sweeping line sweeps black area at most $0.1\epsilon^{3/2}$. Thus, the total black area swept in Q is at most $0.1\epsilon^{3/2} \cdot 2/\sqrt{\epsilon} = 0.2\epsilon$, as claimed. □

By Claim 3.12 and the fact that the black area of Q is at least 0.2ϵ, some black points in Q are not swept, contradicting our assumption. Thus, u is in $\bigcap_{\ell \in L_T} H_\ell^*$. Recall that u belongs to T. Therefore, the point u is in G, as claimed. □

Lemma 3.13 (White area in Hull(P))**.** *If the number of iterations in the recoloring phase is less than $5 \cdot \epsilon^{-1/3}$, then the white area of Hull(P) is at least 0.1ϵ.*

Proof. We start by giving an upper bound on $|L|$. Recall that $m = \log(2/\epsilon)/2$. The set L consists of the lines that define the sides of the bounding rectangle R and one line for each triangle in $\mathcal{T}_i$ for all $i \in \{0, 1, \ldots, m-1\}$. Therefore,

$$|L| = 4 + \sum_{i=0}^{m-1} |\mathcal{T}_i| \leq 4 + \sum_{i=0}^{m-1} 4 \cdot 2^i = 4 \cdot 2^m \leq \frac{5.7}{\sqrt{\epsilon}}.$$

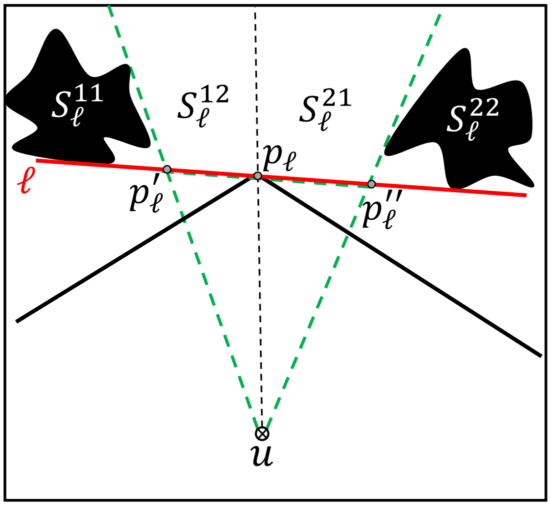

Figure 3.9: A pattern of Type 1.

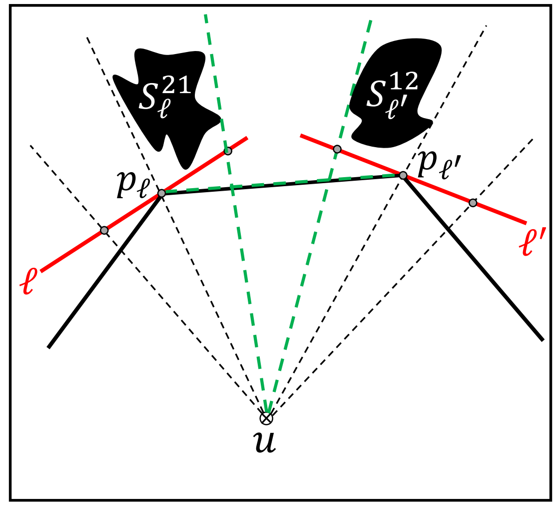

Figure 3.10: A pattern of Type 2.

Each line in L sweeps a black area at most $0.1\epsilon^{3/2}$. Thus, $A(\bigcup_{\ell\in L} S_\ell) \leq 0.1\epsilon^{3/2} \cdot |L| \leq 0.1\epsilon^{3/2} \cdot \frac{5.7}{\sqrt{\epsilon}} \leq 0.57\epsilon$. By Lemma 3.9, the area of all triangles in $\mathcal{T}_m$ is at most 0.5ϵ. By (1), the violet area is less than 0.05ϵ. We obtain a convex figure if we make all black and violet points outside of G white, all white points inside Hull(P) black, and color each triangle in $\mathcal{T}_m$ according to the majority of its area. This recolors area at most $(0.57 + 0.05 + 0.5/2) \cdot \epsilon < 0.9\epsilon$ outside Hull(P). Since the figure is ϵ-far from convex, the white area of Hull(P) is at least 0.1ϵ, as claimed. □

Lemma 3.14. *If the number of iterations in the recoloring phase is less than $5 \cdot \epsilon^{-1/3}$, then there exists a set of white-light patterns whose black regions B_1 and B_2 are all mutually disjoint and whose W regions jointly cover Hull(P).*

Proof. For each line $\ell \in L^*$, recall that p_ℓ denotes the anchor point of the set S_ℓ on ℓ. Denote the two sets, into which the ray $r(u, p_\ell)$ divides S_ℓ, by S_ℓ^1 and S_ℓ^2 (points of S_ℓ^1 come first in the clockwise order). See Figure 3.9. Let p_ℓ' and p_ℓ'' denote the anchor points on ℓ of the sets S_ℓ^1 and S_ℓ^2, respectively. Let the ray $r(u, p_\ell')$ (respectively, $r(u, p_\ell'')$) divide the set S_ℓ^1 (respectively, S_ℓ^2) into sets S_ℓ^{11} and S_ℓ^{12} (respectively, S_ℓ^{21} and S_ℓ^{22}) such that S_ℓ^{11} (respectively, S_ℓ^{21}) is the leftmost subset of S_ℓ^1 (respectively, S_ℓ^2).

Recall what patterns are from Definition 3.3. For each $\ell \in L$, the rays $r(u, p_\ell'), r(u, p_\ell'')$ and the line $\ell(p_\ell', p_\ell'') = \ell$ together with sets $B_1 = S_\ell^{11}$ and $B_2 = S_\ell^{22}$ form a pattern. See Figure 3.9. We say that such a pattern is of *Type 1* and it is *attributed* to ℓ.

Let $\mathcal{T}$ be the set of all triangles $\triangle upp'$, such that p and p' are anchor points and are adjacent vertices of Hull(P). Consider any two lines ℓ and ℓ' from L^*, such that $\triangle up_\ell p_{\ell'}$ is in $\mathcal{T}$ and p_ℓ comes before $p_{\ell'}$ in the clockwise order. We say that the point u is *well-positioned* with respect to the side $p_\ell p_{\ell'}$ if p_ℓ'' and $p_{\ell'}'$ lie in one half-plane defined by the line $\ell(p_\ell, p_{\ell'})$, and u lies in the other. If u is well-positioned w.r.t. $p_\ell p_{\ell'}$ and p_ℓ'' comes before $p_{\ell'}'$ in the clockwise order, then the rays $r(u, p_\ell''), r(u, p_{\ell'}')$ and the line $\ell(p_\ell, p_{\ell'})$ together with the sets $B_1 = S_\ell^{21}$ and $B_2 = S_{\ell'}^{12}$ form a pattern. See Figure 3.10. Such a pattern is of *Type 2* and it is *attributed* to the side $p_\ell p_{\ell'}$.

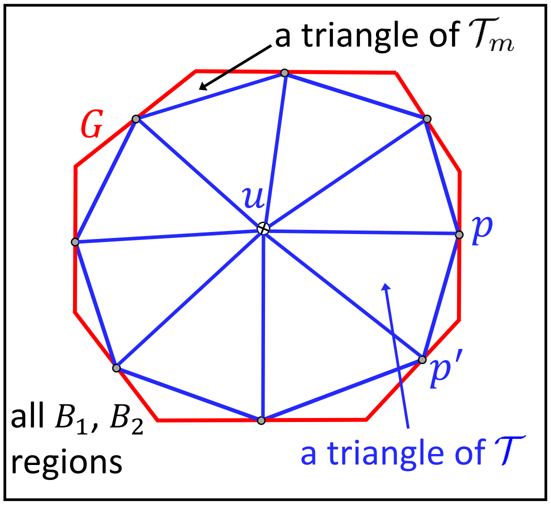

Figure 3.11: An illustration of triangles in $\mathcal{T}_m$ and triangles in $\mathcal{T}$.

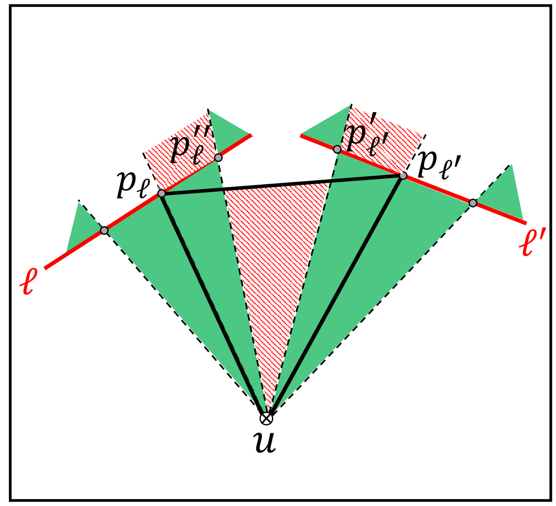

Figure 3.12: When p''_ℓ comes before $p'_{\ell'}$ in the clockwise order.

Recall the polygon G from Lemma 3.11 and recall that the point u is inside G. Note that the region $G \setminus \text{Hull}(P)$ consists of all triangles in $\mathcal{T}_m$ (see Figure 3.11). (If a triangle that constitutes the region $G \setminus \text{Hull}(P)$ is not in $\mathcal{T}_m$, then it belongs to $\mathcal{T}_i$, for some $i < m$. This triangle was not subdivided since it has a small black area. Thus, a sweeping line sweeps the entire black area in that triangle and stops when it reaches the base of the triangle. By Lemma 3.11, every line in L does not sweep the point u. Therefore, every such triangle cannot contain u.) Thus, u can be in Hull(P) or in one of the triangles of $\mathcal{T}_m$. First, assume u is in Hull(P). Then, u is well-positioned w.r.t. every side of Hull(P). Moreover, triangles in $\mathcal{T}$ partition Hull(P). Consider a triangle $\triangle up_\ell p_{\ell'}$ from $\mathcal{T}$. If p''_ℓ comes before $p'_{\ell'}$ in the clockwise order, then the W-regions of the Type 1 patterns attributed to ℓ and ℓ', and the W-region of the Type 2 pattern attributed to $p_\ell p_{\ell'}$, jointly cover $\triangle up_\ell p_{\ell'}$ (see Figure 3.12). Otherwise, the W- regions of the Type 1 patterns attributed to ℓ and ℓ' jointly cover $\triangle up_\ell p_{\ell'}$, as shown in Figure 3.13. (In this case, there is no Type 2 pattern attributed to $p_\ell p_{\ell'}$.) Note that, in both cases, the black regions B_1 and B_2 of the patterns are mutually disjoint. Thus, by considering every triangle in $\mathcal{T}$, we obtain a set of white-light patterns whose regions B_1, B_2 are disjoint and whose W-regions jointly cover all the triangles in $\mathcal{T}$. Therefore, the W-regions of all those white-light patterns jointly cover Hull(P), as shown in Figure 3.14. This figure shows the case when all $p_\ell p_{\ell'}$ have Type 2 patterns attributed to them. Type 1 patterns are green and Type 2 patterns are red. (If printed black-and-white, Type 1 patterns are solid-colored and Type 2 patterns have an ornament.)

Now assume u is in one of the triangles of $\mathcal{T}_m$. Consider the triangle of $\mathcal{T}_m$ in which u is located and let pp' denote its side, where p and p' are adjacent vertices of Hull(P). Let T denote the triangle $\triangle upp' \in \mathcal{T}$. Note that all the triangles in $\mathcal{T} \setminus T$ jointly cover Hull(P). Moreover, for each triangle $\triangle up_\ell p_{\ell'} \in (\mathcal{T} \setminus T)$, the following holds: 1) u is well-positioned w.r.t. $p_\ell p_{\ell'}$, 2) the W-regions of the Type 1 patterns attributed to ℓ and ℓ', and the W-region of the Type 2 pattern attributed to $p_\ell p_{\ell'}$ jointly cover $\triangle up_\ell p_{\ell'}$, as shown in Figure 3.15. (In this figure, Type 1 patterns are green and Type 2 patterns are red. If printed black-and-white, Type 1 patterns are solid-colored and Type 2 patterns have an ornament.) Thus, there are white-light patterns whose W-regions jointly cover

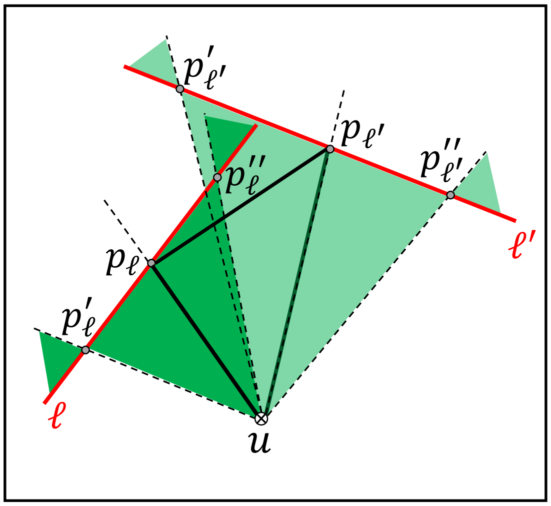

Figure 3.13: When $p'_{\ell'}$ comes before p''_{ℓ} in the clockwise order.

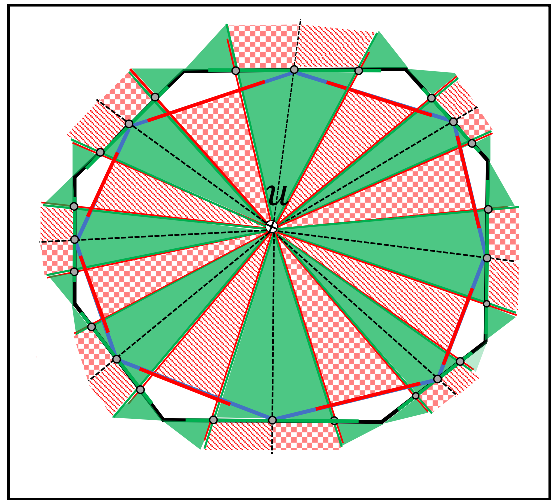

Figure 3.14: An illustration of Type 1 and Type 2 patterns when u belongs to Hull(P).

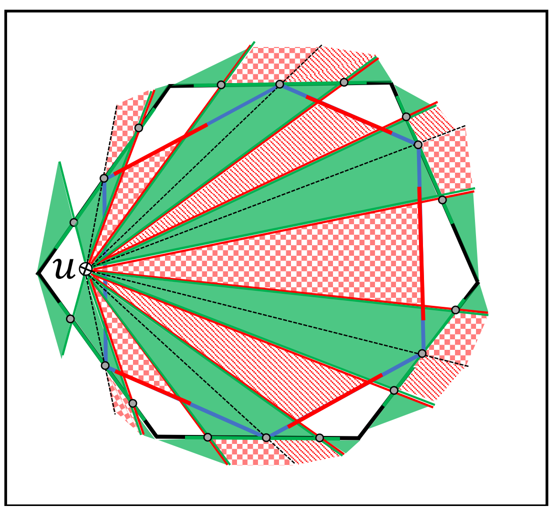

Figure 3.15: Patterns when u is in a triangle from $\mathcal{T}_m$.

Hull(P) and whose regions B_1, B_2 are mutually disjoint, as claimed. $\square$

Now we will show that with probability at least 0.68, the tester samples a witness triple. Intuitively, by Lemmas 3.13 and 3.14, the W-regions of the patterns we constructed cover white area at least 0.1ϵ. So, the tester is likely to sample a point from one of them and, conditioned on that event, it is also likely to sample points from both black sets of the same pattern.

Index all patterns of Type 1 and Type 2 by natural numbers $1, 2, \ldots$ All of them are white-light patterns, since there are no white-heavy patterns left after the recoloring phase. For a pattern i, let E^i_W denote the event that a white point is sampled from its W region, let a_i be the white area of W, and let E^i_B denote the event that a point is sampled both from its B_1 region and from its B_2 region. By Claim 3.5, $\Pr[E^i_B] \geq 0.7 \cdot \epsilon^{1/3}$. Moreover, $\Pr[E^i_W] = 1 - e^{-a_i s}$. Recall that, by Definition 3.3, $a_i < 0.025\epsilon^{4/3}$ and, thus, $a_i s < 1.5$. We use the fact that $1 - e^{-x} \geq 0.5x$ for all $x \in (0, 1.5)$. We obtain that $\Pr[E^i_W] \geq 0.5 a_i s$. Therefore, the probability that we sample a witness triple from pattern i is

$$\Pr[E^i_B \cap E^i_W] \geq 0.7 \cdot \epsilon^{1/3} \cdot 0.5 a_i s = 17.5 \cdot a_i/\epsilon.$$

By Lemmas 3.13 and 3.14, $\sum a_i \geq 0.1\epsilon$. Therefore, by Claim 3.7, the probability that a witness triple is sampled from at least one pattern is

$$\Pr\left[\bigcup (E^i_B \cap E^i_W)\right] \geq 1 - e^{-\sum_i 17.5 \cdot a_i/\epsilon} \geq 1 - e^{-0.1 \cdot 17.5} > 0.68,$$

as desired. This completes the proof of Proposition 3.8. $\square$

4 Lower Bound for Uniform Testing of Convexity

Theorem 4.1. *Every 1-sided error uniform ϵ-tester for convexity needs $\Omega(\epsilon^{-4/3})$ samples.*

Proof. By the Poissonization Lemma (Lemma 2.1), it is sufficient to prove the lower bound for Poisson algorithms. Observe that a 1-sided error tester can reject only if the samples it obtained are not consistent with any convex figure. For each sufficiently small ϵ, we will construct a set C_ϵ in $U = [0,1]^2$ that is ϵ-far from convex. We will show that there exists a constant c_0 such that every Poisson-s tester with $s = c_0 \cdot \epsilon^{-4/3}$ fails to detect a violation of convexity with probability at least 1/2, for every constructed set C_ϵ.

First, we construct the hard sets C_ϵ. Let $k = \left\lfloor \frac{1}{5} \cdot \epsilon^{-1/2} \right\rfloor$. Let G be a convex regular $2k$-gon inside $[0,1]^2$ with the side length $\frac{1}{2}\sin(\frac{\pi}{2k})$. Number all vertices of G from 1 to $2k$ in the clockwise order (see Figure 4.1). Let G' and G'' be the two regular k-gons obtained by connecting the vertices with odd and even numbers, respectively. Let C_ϵ be the set of points in $G' \cup G''$. That is, in the resulting figure, all points in $G' \cup G''$ are black and all remaining points in $[0,1]^2$ are white.

Lemma 4.2. *The figure (U, C_ϵ) is ϵ-far from convex for all sufficiently small ϵ.*

Proof. The region $G \backslash (G' \cup G'')$ consists of triangles in which all points are white. Call any such triangle *white*. The symmetric difference of G' and G'' consists of triangles in which all points are black. Call any such triangle *black*. Let T be a black triangle and b be its vertex such that it is also a vertex of G (see Figure 4.2). Let d and d' denote the other two vertices of T. Let b_0 be the point on the side dd' such that bb_0 is the height of T. Call triangles $\triangle bb_0d$ and $\triangle bb_0d'$ *teeth*. A *crown* is the convex hull of the vertices of two teeth intersecting in exactly one point (see Figure 4.2).

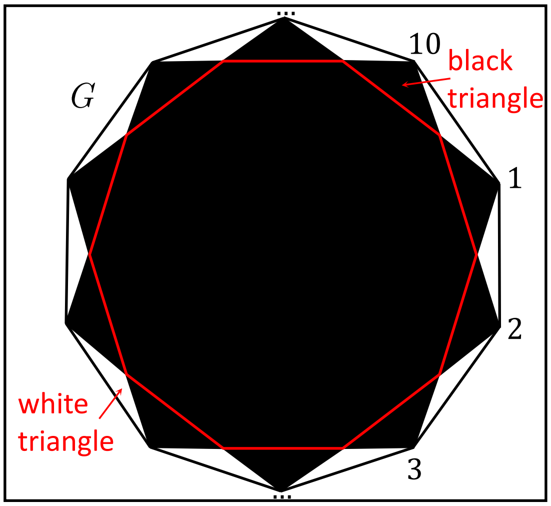

Figure 4.1: An illustration of G for $k = 5$.

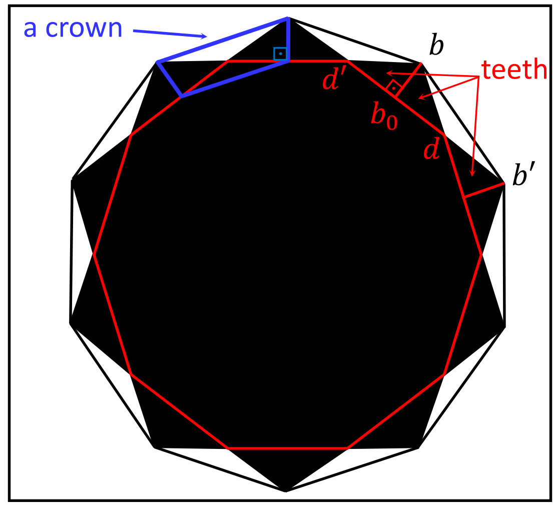

Figure 4.2: Teeth and crowns.

Claim 4.3. Let A_T and A_W be the areas of a tooth and a white triangle, respectively. Then, for sufficiently large k, we have

$$\frac{1}{5k^3} \leq A_T \leq A_W \leq \frac{1}{k^3}.$$

Moreover, area at least $\frac{A_T}{8}$ of each of the $2k$ disjoint crowns must be changed to make C_ϵ convex.

Proof. We start by calculating the areas A_T and A_W. Consider the triangles $\triangle dbd'$ $\triangle bdb'$ shown in Figure 4.2. Let $\theta = \pi/k$. Since the angle $\angle dbd'$ is an interior angle of a regular k-gon, it equals $(k-2)\cdot\pi\cdot\frac{1}{k} = \pi - 2\theta$. By symmetry, $|bd| = |bd'|$. Thus, $\angle bd'b_0 = \angle bdb_0 = \theta$. Consequently, $\angle bdb' = \pi - \theta$ and, by symmetry, $\angle dbb' = \angle db'b = \theta/2$. Let $x = |bd|$ and recall that, by construction, $|bb'| = \frac{1}{2}\sin\frac{\theta}{2}$. The cosine of $\angle dbb'$ is $\cos\frac{\theta}{2} = (|bb'|/2)/x = (\frac{1}{4}\sin\frac{\theta}{2})/x$. Thus,

$$x = \frac{1}{4}\cdot\frac{\sin(\theta/2)}{\cos(\theta/2)} = \frac{1}{4}\cdot\tan\frac{\theta}{2}.$$

By calculating the area for the black and white triangles, we get

$$\begin{aligned}
A_T &= \frac{A(\triangle bdd')}{2} = \frac{|bd|\cdot|bd'|\cdot\sin 2\theta}{4} = \frac{x^2\cdot\sin 2\theta}{4} = \frac{1}{64}\tan^2\frac{\theta}{2}\sin 2\theta;\\
A_W &= A(\triangle bdb') = \frac{|bd|\cdot|b'd|\cdot\sin\theta}{2} = \frac{x^2\cdot\sin\theta}{2} = \frac{1}{32}\tan^2\frac{\theta}{2}\sin\theta.
\end{aligned}$$

Thus, $A_T = \frac{1}{32}\tan^2(\theta/2)\sin\theta\cos\theta = A_W\cdot\cos\theta \leq A_W$.

Since $0.9z \leq \sin z \leq z$ and $z \leq \tan z \leq 2z$ for $z \in [0, 0.78]$, we obtain that, for sufficiently large k,

$$\begin{aligned}
A_T \geq &\quad \tfrac{1}{64}(\tfrac{\pi}{2k})^2\cdot 0.9(\tfrac{2\pi}{k}) &\geq \frac{1}{5}\cdot\frac{1}{k^3};\\
A_W \leq &\quad \tfrac{1}{32}(\tfrac{\pi}{k})^2\cdot(\tfrac{\pi}{k}) &\leq \frac{1}{k^3}.
\end{aligned}$$

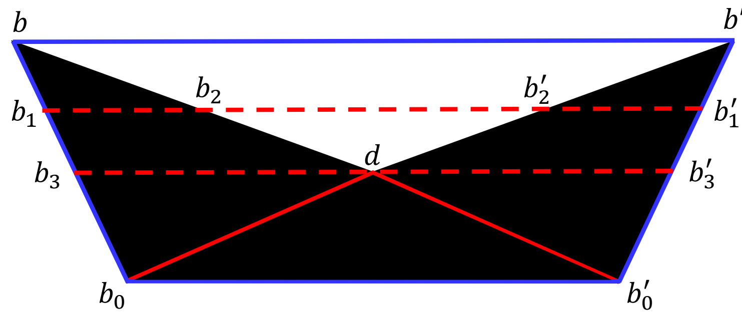

Figure 4.3: A crown.

This completes the proof of the first statement in Claim 4.3.

Now we prove the second statement. Consider a crown with vertices labeled as shown in Figure 4.3. Let b_2 and b_2' be the midpoints of bd and $b'd$, respectively. Let $\ell(b_2, b_2')$ intersect bb_0 and $b'b_0'$ at b_1 and b_1', respectively. Note that $\ell(b_1 b_1')$ is parallel to $\ell(b, b')$. Therefore, triangles $\triangle db_2b_2'$ and $\triangle dbb'$ are similar and

$$A(\triangle db_2b_2') = \frac{A(\triangle dbb')}{4} = \frac{A_W}{4} \geq \frac{A_T}{4}.$$

Consider the line parallel to $\ell(b, b')$ that passes through point d, and let b_3 and b_3' be the intersection of this line with b_0b and $b_0'b'$, respectively. Then $\angle bdb_3 = \angle dbb' = \theta/2$. Recall that $\angle bdb_0 = \theta$ and, consequently, $\angle b_3db_0 = \theta/2$. We get that $A(\triangle b_3db) > A(\triangle b_3db_0)$ because the triangles in question share the side b_3d, both have angle $\theta/2$ adjacent to this side, but $|bd| > |b_0d|$, since bd is the hypothenuse of $\triangle b_0db$. Thus, $A(\triangle b_3db) > A(\triangle b_0db)/2 = A_T/2$. Since $\triangle b_1b_2b$ and $\triangle b_3db$ are similar and since $|bb_2| = |bd|/2$, by the same reasoning as before, we get that $A(\triangle b_1b_2b) = A(\triangle b_3db)/4$. We then obtain $A(\triangle b_1b_2b) \geq A_T/8$. By symmetry, $A(\triangle b_1'b_2'b') \geq A_T/8$ as well.

Consider the triangle $\triangle b_0db_0'$. Note that $|b_0d| = |b_0'd| = x \cdot \cos\theta$. Thus, for sufficiently large k,

$$A(\triangle b_0db_0') = \frac{|b_0d| \cdot |b_0'd| \cdot \sin\theta}{2} = \frac{x^2 \cdot \cos^2\theta \cdot \sin\theta}{2} = A_T \cdot \cos\theta \geq \frac{A_T}{8}.$$

Finally, consider a convex set C' closest to C_ϵ. If the black component of C' contains $\triangle b_2b_2'd$ entirely, then C' and C_ϵ differ on at least $A(\triangle db_2b_2') \geq \frac{A_T}{4}$ area in the crown. If the black component of C' does not contain $\triangle b_2b_2'd$ entirely then, by construction of C_ϵ and convexity of C', one of $\triangle bb_1b_2$, $\triangle b'b_1'b_2'$ or $\triangle b_0db_0'$ is completely outside of the black component of C'. Thus, at least $\frac{A_T}{8}$ area must be modified in a crown in order to obtain C' from C_ϵ. This completes the proof of Claim 4.3. □

There are $2k$ disjoint crowns. Recall that $k = \lfloor \frac{1}{5} \cdot \epsilon^{-1/2} \rfloor$. Thus, by Claim 4.3, to make C_ϵ convex, area at least $\frac{A_T}{8} \cdot 2k \geq 1/(20k^2) \geq \epsilon$ needs to be modified. That is, the figure (U, C_ϵ) is ϵ-far from convex. This completes the proof of Lemma 4.2. □

Now consider how an algorithm can detect a violation of convexity in the hard figures we constructed. First, it is sufficient to change all the points in the white triangles to make such a figure convex. Therefore, a violation can be detected only if a point from a white triangle is in the sample. For any white triangle, it is sufficient to change the points in one of the two black triangles adjacent to it to ensure that the points from the white triangle are not in the convex hull of black points. Therefore, it is necessary to sample a point from both adjacent black triangles. Thus, the probability of detecting a violation of convexity is bounded from above by the probability of detecting a *red-flag triple*, defined next.

Definition 4.1 (A red-flag triple)**.** *A triple of points (w, b_1, b_2) is a* red-flag triple *if w belongs to a white triangles and b_1 and b_2 belong to two different adjacent black triangles.*

Observe that if a sample of points from the figure (U, C_ϵ) does not contain a red-flag triple, then it is consistent with a convex figure. Specifically, an example of such a convex figure is the one that is colored white on all black triangles of (U, C_ϵ) from which no point was sampled and that is identical to (U, C_ϵ) everywhere else. Since there is a convex figure consistent with the sample, every 1-sided error tester must accept in this case.

Lemma 4.4. *Let c_0 be an appropriate constant. For all sufficiently small ϵ, a Poisson-s algorithm with $s = c_0 \cdot \epsilon^{-4/3}$ detects a red-flag triple in the figure (U, C_ϵ) with probability at most 1/2.*

Proof. We define the following random variables for the Poisson-s algorithm: Y counts the total number of sampled red-flag triples, Y_W counts the number of sampled red-flag triples that involve a point w from a white triangle W, variable X_W counts the number of samples in a white triangle W, and X_{B_1} and X_{B_2} count the number of samples in the two black triangles adjacent to W, respectively. To prove the lemma, it is sufficient to show that $\Pr[Y \geq 1] \leq 1/2$.

By the Poissonization Lemma (Lemma 2.1), X_W is a Poisson random variable with expectation $A_W \cdot s$, where A_W is the area of a white triangle. Similarly, $\mathbb{E}[X_{B_1}] = \mathbb{E}[X_{B_2}] = 2A_T \cdot s$, where A_T is the area of a tooth and hence half the area of a black triangle. The random variables X_W, X_{B_1}, X_{B_2} are independent because they are sample counts for disjoint areas. Since $Y_W = X_W \cdot X_{B_1} \cdot X_{B_2}$, we get that

$$\mathbb{E}[Y_W] = \mathbb{E}[X_W] \cdot \mathbb{E}[X_{B_1}] \cdot \mathbb{E}[X_{B_2}] = 4A_W \cdot (A_T)^2 \cdot s^3 \leq 4(A_W)^3 \cdot s^3 \leq \frac{4}{k^9} \cdot s^3.$$

The inequalities above use Claim 4.3 and hold for sufficiently large k (i.e., sufficiently small ϵ).

Note that $k \geq \frac{1}{5\sqrt{\epsilon}} - 1 \geq \frac{1}{6\sqrt{\epsilon}}$, for sufficiently small ϵ. Since there are $2k$ crowns, with identical distributions of samples inside them,

$$\mathbb{E}[Y] = 2k \cdot \mathbb{E}[Y_W] \leq 8\frac{1}{k^8} \cdot s^3 \leq 8 \cdot 6^8 \epsilon^4 \cdot c_0^3 \epsilon^{-4} \leq 1/2,$$

assuming c_0 and ϵ are sufficiently small. By Markov's inequality, the probability of detecting a red-flag is at most $\Pr[Y \geq 1] \leq \mathbb{E}[Y] \leq 1/2$. ☐

Theorem 4.1 follows from Lemmas 4.2 and 4.4. Thus, 1-sided error uniform ϵ-tester for convexity needs $s = \Omega(\epsilon^{-4/3})$ samples. ☐

5 Testing Convexity in the Pixel Model

In this section, we show that Algorithm 1 is an optimal uniform ϵ-tester for convexity in the pixel model. In this model, we focus on black and white images. For simplicity, we only consider square images, but everything in this section can be easily generalized to rectangular images.

Image representation. We represent an image by an $n \times n$ binary matrix M of pixel values, where 0 denotes white and 1 denotes black. The object is a subset of $[n]^2$ corresponding to black pixels; namely, $\{(i,j) \mid M[i,j] = 1\}$.

Definition 5.1 (Convexity of images)**.** *An image is* convex *if the convex hull of all black pixels contains only black pixels. Let the* relative distance *between two images be the fraction of pixels on which they differ. An image is* ϵ-far from being convex *if its relative distance to every convex image is at least* ϵ.

5.1 Uniform Tester for Convexity of Images

The intuition behind why Algorithm 1 works in the pixel model comes from the following theorem which relates the area of a lattice polygon and the number of integer points that the polygon covers. (A lattice polygon is a polygon whose vertices have integer coordinates.)

Theorem 5.1 (Pick's theorem [23])**.** *For a simple lattice polygon G, let α denote the number of lattice points in the interior of G and β denote the number of lattice points on the boundary of G. Then $A(G) = \alpha + \beta/2 - 1$.*

Definition 5.2. *For a polygon G, let $Perim(G)$ denote the perimeter of G and $Pix(G)$ denote the number of pixels covered by G, i.e., pixels in the interior of G and on its boundary.*

By using Theorem 5.1, we prove the following lemma that gives an upper bound on the number of pixels in a polygon in terms of its area and perimeter.

Lemma 5.2. *For a convex polygon G,*

$$Pix(G) \leq A(G) + \frac{Perim(G)}{2} + 1.$$

Proof. If all pixels covered by G are collinear, then $Pix(G) \leq Perim(G)/2 + 1$. This follows from the fact that the length of a line segment in G is at most $Perim(G)/2$ and that the number of pixels on the line segment is at most the length of the line segment plus one. Since

$$Perim(G)/2 + 1 \leq A(G) + Perim(G)/2 + 1,$$

the lemma holds for this case.

Now assume that not all pixels covered by G are collinear. Consider the convex hull of all pixels covered by G. Let α and β denote the number of pixels in the interior and on the boundary of that convex hull, respectively. By Theorem 5.1, we obtain that $\alpha + \beta/2 - 1 \leq A(G)$. Thus,

$$Pix(G) = \alpha + \beta \leq A(G) + \beta/2 + 1 \leq A(G) + Perim(G)/2 + 1.$$

This completes the proof of Lemma 5.2. $\square$

Theorem 5.3 (Main Theorem for the Pixel Model)**.** *There is a uniform (1-sided error) ϵ-tester for convexity of $n \times n$ images with sample and time complexity $O(\epsilon^{-4/3})$, for all $\epsilon \geq 85/n$.*

Proof. Recall from Footnote 7 that it is sufficient to prove the theorem for sufficiently small ϵ also satisfying $\epsilon \geq 85/n$. We show that for all $\epsilon \in [\frac{85}{n}, 5^{-6})$, Algorithm 1 is a uniform ϵ-tester for convexity in the pixel model. The analysis of the algorithm is the same as in Section 3, with a small number of modifications. First, we make adjustments to the definitions to capture the fact that, in the pixel model, the relevant parameter for sampling from some portion of the image is the number of pixels in that portion, rather than its area.

Definition 5.3 (Central point, pattern, anchor point in the pixel model). *The relevant quantities in the pixel model are as defined in Section 3, with the following exceptions:*

1. *A* central point *is the intersection of two lines such that each closed quadrant that they define contains at least* $0.25\epsilon n^2$ *black pixels.*
2. *Each of the sets* B_1 *and* B_2 *in a pattern has exactly* $t = \lfloor 0.025\epsilon^{3/2}n^2 \rfloor$ *black pixels. A pattern is* white-heavy *if the number of white pixels in its* W *region is at least* $0.025\epsilon^{4/3}n^2$*; otherwise, it is* white-light.
3. *Every line* ℓ *that we put in* L *is the furthest line from* u *such that its corresponding set* S_ℓ *of black pixels has size at least* $4t$.
4. *The anchor point* p_ℓ *on a line* ℓ *is defined so that there are at least* $2t$ *pixels of* S_ℓ *in each closed half-plane defined by* $\ell(u, p_\ell)$.

Given the modified definitions, the analysis of the algorithm proceeds as in Section 3, except that the areas of relevant portions of figures are replaced with the number of pixels in the corresponding portion of the image. Recall that $m = \log(2/\epsilon)/2$. For each triangle $T \in \mathcal{T}_m$, we consider the number of pixels in T instead of $A(T)$. Since the number of pixels in a region and the area of the region are not necessarily equal, we prove new versions of Lemma 3.9 (Area of Triangles in $\mathcal{T}_m$) and Lemma 3.13 (White Area in Hull(P)). The new lemmas are Lemma 5.4 (Number of Pixels in Triangles of $\mathcal{T}_m$) and Lemma 5.5 (Number of White Pixels in Hull(P)).

We start by showing that Claim 3.5 holds for the pixel model.

Proof of Claim 3.5 for the pixel model. In a pattern, each of its B_1 and B_2 regions contains $t = \lfloor 0.025\epsilon^{3/2}n^2 \rfloor$ black pixels. Recall the events E_1 and E_2 for a pattern from Claim 3.5 and note that $t \geq 0.024\epsilon^{3/2}n^2 \geq 1$, since $\epsilon \geq 85/n$. Then, $\Pr[E_1] = \Pr[E_2] = 1 - e^{-st/n^2} \geq 1 - e^{-1.2\epsilon^{1/6}}$. Note that $1 - e^{-x} \geq 0.7x$, for $x \in (0, 1/4)$, and that $1.2\epsilon^{1/6} \in (0, 1/4)$. Therefore, $\Pr[E_1] = \Pr[E_2] \geq 1 - e^{-1.2\epsilon^{1/6}} \geq 0.7 \cdot 1.2\epsilon^{1/6} = 0.84\epsilon^{1/6}$ and $\Pr[E_1 \cap E_2] \geq (0.84\epsilon^{1/6})^2 \geq 0.7\epsilon^{1/3}$, as claimed. □

Since Claim 3.5 holds in the pixel model, so does Proposition 3.6 by the same reasoning as in Section 3. Now we show that if $\epsilon \in [\frac{85}{n}, 5^{-6})$, using Lemmas 5.4 and 5.5 instead of Lemmas 3.9 and 3.13, respectively, gives a proof of Proposition 3.8 for the pixel model.

Lemma 5.4 (Number of pixels in triangles of $\mathcal{T}_m$). *If* $\epsilon > \frac{85}{n}$, *then* $\sum_{T \in \mathcal{T}_m} Pix(T) \leq 0.56 \cdot \epsilon n^2$.

Proof. By Lemma 5.2, for every triangle $T \in \mathcal{T}_m$, we have that $Pix(T) \leq A(T) + Perim(T)/2 + 1$. Therefore, $\sum_{T \in \mathcal{T}_m} Pix(T) \leq \sum_{T \in \mathcal{T}_m}(A(T) + Perim(T)/2 + 1)$. Since the area of the whole image is $(n-1)^2$, by Claim 3.10, we get that $\sum_{T \in \mathcal{T}_m} A(T) \leq \frac{1}{4^m}(n-1)^2 = \epsilon(n-1)^2/2$. Note that all triangles in $\mathcal{T}_m$ form a region which is a symmetric difference of two convex polygons such that one is inscribed in the other (see Figure 3.11). Therefore, the sum of the perimeters of all triangles in $\mathcal{T}_m$ is equal to the sum of the perimeters of the polygons. Note that these two convex polygons are inside an $n \times n$ square. Thus, the perimeter of each of them is at most $4n$ and thus, $\sum_{T \in \mathcal{T}_m} Perim(T)/2 \leq (4n + 4n)/2 = 4n$. Recall that $m = \log(2/\epsilon)/2$. Consequently, we obtain that $|\mathcal{T}_m| \leq 2^{m+2} \leq \frac{5.7}{\sqrt{\epsilon}}$. Therefore,

$$\sum_{T \in \mathcal{T}_m} Pix(T) \leq \frac{\epsilon(n-1)^2}{2} + 4n + \frac{5.7}{\sqrt{\epsilon}} \leq \frac{\epsilon n^2}{2} + 4n + 0.7\sqrt{n} \leq \frac{\epsilon n^2}{2} + 4.7n \leq 0.56 \cdot \epsilon n^2,$$

as claimed (recall that $\epsilon \geq 85/n$). □

Lemma 5.5 (Number of white pixels in Hull(P))**.** *If the number of iterations in the recoloring phase is less than* $5 \cdot \epsilon^{-1/3}$*, then the number of white pixels in Hull(P) is at least* $0.1\epsilon n^2$*.*

Proof. The proof of this lemma is very similar to the proof of Lemma 3.9. Note that the number of black pixels "swept" by all lines in L is at most

$$4t \cdot |L| \leq 0.1\epsilon^{3/2} n^2 \cdot |L| \leq 0.1\epsilon^{3/2} n^2 \cdot \frac{5.7}{\sqrt{\epsilon}} \leq 0.57 \cdot \epsilon n^2.$$

By Lemma 5.4, the number of pixels in all triangles of $\mathcal{T}_m$ is at most $0.56 \cdot \epsilon n^2$. Since $\epsilon < 5^{-6}$, the number of all violet pixels after the recoloring phase is at most

$$5 \cdot \epsilon^{-1/3}(2 \cdot 0.025 \cdot \epsilon^{3/2} n^2) = 0.25 \cdot \epsilon^{7/6} n^2 < 0.05 \cdot \epsilon n^2.$$

Recall the polygon G from Lemma 3.11. We obtain a convex image if we make all black and violet pixels outside of G white, all white pixels inside Hull(P) black, and color each triangle in $\mathcal{T}_m$ according to the color of the majority of its pixels. By doing so, we recolor at most $(0.57 + 0.05 + 0.56/2) \cdot \epsilon n^2 = 0.9\epsilon n^2$ pixels outside Hull(P). Since the image is ϵ-far from convex, the number of white pixels in Hull(P) is at least $0.1\epsilon n^2$, as claimed. □

The rest of the proof of Proposition 3.8 for the pixel model is as in Section 3. □

5.2 Lower Bound for Uniform Testing of Convexity of Images

In this section, we prove the lower bound on the number of uniform samples needed for 1-sided error testing of convexity of images.

Theorem 5.6 (Lower Bound for the Pixel Model)**.** *If* $\epsilon \in \left[\frac{85}{n}, 5^{-6}\right)$, *then every 1-sided uniform* ϵ*-tester for convexity of* $n \times n$ *images in the pixel model needs* $\Omega(\epsilon^{-4/3})$ *samples.*

Proof. To prove the theorem, we construct a hard example using the set U and the polygons G' and G'' from Section 4. We rescale U such that it becomes $[1, n]^2$. We rescale polygons G' and G'' accordingly. Note that the area of every region and the length of every line segment considered in Section 4 are multiplied by $(n-1)^2$ and $n-1$, respectively, after rescaling. Consider the $n \times n$ image M such that $M[i, j] = 1$ iff (i, j) is inside $G' \cup G''$.

Lemma 5.7. *For all* $\epsilon \in \left[\frac{85}{n}, 5^{-6}\right)$*, the image M is ϵ-far from convex.*

Proof. Recall the definition of a crown $bb_0b_0'b'$ and triangles $\triangle b_2b_2'd$, $\triangle bb_1b_2$, $\triangle b'b_1'b_2'$, and $\triangle b_0db_0'$ from the proof of Claim 4.3 in Section 4. Call triangles $\triangle b_2b_2'd$, $\triangle bb_1b_2$, $\triangle b'b_1'b_2'$, and $\triangle b_0db_0'$ *crucial*. Let M' be a closest convex image to M. Note that in every crown, there is a crucial triangle such that M and M' are different on all pixels inside the triangle. Mark exactly one such crucial triangle in every crown. Call every 1×1 square region $[i, i+1] \times [j, j+1]$ *cell*, where $i, j \in [n-1]$. Call every cell that intersects a side of a marked triangle *boundary cell*. Call every cell that lies entirely inside a marked triangle *inner cell*. The number of all boundary cells is at most 2 times the sum of the perimeters of all marked triangles.

Claim 5.8. Let T be a marked triangle in a crown $bb_0b_0'b'$. Then $Perim(T) < |bb_0'| + |b'b_0|$.

Proof. Let $y = |b_0d|$. Recall that $x = |bd| = |b'd|$ and that in $\triangle bb_0d$, the side bd is the hypotenuse. Thus, $x = |bd| > |b_0d| = y$. By the triangle inequality, $Perim(\triangle b_2db_2') < |b_2d| + |b_2'd| + (|b_2d| + |b_2'd|) = x/2 + x/2 + x = 2x < 2x + 2y = |bb_0'| + |b'b_0|$ and $Perim(\triangle b_0db_0') < |b_0d| + |b_0'd| + (|b_0d| + |b_0'd|) = y + y + 2y = 4y < 2x + 2y = |bb_0'| + |b'b_0|$. Now consider $\triangle bb_1b_2$. Recall that $\angle bb_1b_2$ is obtuse. Therefore, $Perim(\triangle b'b_1'b_2') = Perim(\triangle bb_1b_2) < 3 \cdot |bb_2| = 3x/2 < 2x + 2y = |bb_0'| + |b'b_0|$. Since T is one of $\triangle b_2b_2'd$, $\triangle bb_1b_2$, $\triangle b'b_1'b_2'$, or $\triangle b_0db_0'$, the claim holds. □

By Claim 5.8, the sum of the perimeters of all marked triangles is at most $Perim(G') + Perim(G'') \leq 8n$. Therefore, there are at most $16n$ boundary cells and they cover at most $16n$ area. Recall that $k = \lfloor \frac{1}{5} \cdot \epsilon^{-1/2} \rfloor$. By Claim 4.3 and the fact that U is rescaled, the sum of the areas of all marked triangles is at least $(n-1)^2/20k^2 > 1.2\epsilon(n-1)^2$. (Note that Claim 4.3 holds since $\epsilon < 5^{-6}$.)

Thus, the area covered by inner cells is at least $1.2\epsilon(n-1)^2 - 16n \geq \epsilon n^2$ (recall that $\epsilon \geq 85/n$). This implies that the number of inner cells is at least ϵn^2. Since each cell contains 4 pixels and every pixel belongs to at most 4 cells, all inner cells contain at least $\frac{4\epsilon n^2}{4} = \epsilon n^2$ pixels. Thus, at least ϵn^2 pixels must be modified in M to obtain M'. □

Note that to detect nonconvexity of M at least one triple of pixels from a crown must be sampled. The rest of the analysis is done in the same way as in Section 4 and the proof of the theorem is completed. □

6 Conclusion

We showed that, in 2 dimensions, testing convexity of figures with uniform samples can be done faster than learning convex figures under the uniform distribution. It is an interesting open question whether this is also true in higher dimensions. Schmeltz [32] proved that $\Theta(\epsilon^{-(d+1)/2})$ samples are necessary and sufficient for PAC-learning a d-dimensional convex set under the uniform distribution. As mentioned in Section 1, a connection between (proper) PAC-learning and property testing, established by Goldreich et al. [15], implies that the same number of samples are sufficient for ϵ-testing convexity (with uniform algorithms) in d dimensions. However, it is open whether this sample complexity can be improved. No nontrivial lower bounds are known for this problem.

We proved that the running time of the uniform tester presented in this work cannot be improved if 1-sided error is required. The question is open for 2-sided error testers, even though, in subsequent work [5], some progress was made: it was shown that $\Omega(\epsilon^{-5/4})$ uniform samples are required to test convexity, even with 2-sided error.

References

[1] A. M. Andrew. Another efficient algorithm for convex hulls in two dimensions. *Inf. Process. Lett.*, 9(5):216–219, 1979.

[2] Imre Bárány. Extremal problems for convex lattice polytopes: a survey. *Contemporary Mathematics*, 453:87, 2008.

[3] Imre Bárány and Zoltán Füredi. On the shape of the convex hull of random points. *Probability theory and related fields*, 77(2):231–240, 1988.

[4] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Tolerant testers of image properties. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 90:1–90:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[5] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. The power and limitations of uniform samples in testing properties of figures. *Algorithmica*, Jul 2018.

[6] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Testing convexity of figures under the uniform distribution. *Random Structures and Algorithms*, Sep 2018.

[7] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. L_p-testing. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 164–173. ACM, 2014.

[8] Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, 2014*, pages 309–320, 2014.

[9] Bernard Chazelle and C. Seshadhri. Online geometric reconstruction. *J. ACM*, 58(4):14, 2011.

[10] Artur Czumaj and Christian Sohler. Property testing with geometric queries. In *Algorithms - ESA 2001, 9th Annual European Symposium, Aarhus, Denmark, August 28-31, 2001, Proceedings*, pages 266–277, 2001.

[11] Artur Czumaj, Christian Sohler, and Martin Ziegler. Property testing in computational geometry. In *Algorithms - ESA 2000, 8th Annual European Symposium, Saarbrücken, Germany, September 5-8, 2000, Proceedings*, pages 155–166, 2000.

[12] Herbert Edelsbrunner. *Algorithms in Combinatorial Geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*. Springer, 1987.

[13] Ronen Eldan. A polynomial number of random points does not determine the volume of a convex body. *Discrete and Computational Geometry*, 46(1):29–47, 2011.

[14] Eldar Fischer, Oded Lachish, and Yadu Vasudev. Trading query complexity for sample-based testing and multi-testing scalability. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1163–1182. IEEE Computer Society, 2015.

[15] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[16] Oded Goldreich and Dana Ron. On sample-based testers. *ACM Trans. Comput. Theory*, 8(2):7:1–7:54, April 2016.

[17] Sariel Har-Peled. On the expected complexity of random convex hulls. *CoRR*, abs/1111.5340, 2011.

[18] Igor Kleiner, Daniel Keren, Ilan Newman, and Oren Ben-Zwi. Applying property testing to an image partitioning problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(2):256–265, 2011.

[19] Simon Korman, Daniel Reichman, and Gilad Tsur. Tight approximation of image matching. *CoRR*, abs/1111.1713, 2011.

[20] Simon Korman, Daniel Reichman, Gilad Tsur, and Shai Avidan. Fast-match: Fast affine template matching. *International Journal of Computer Vision*, 121(1):111–125, 2017.

[21] Nimrod Megiddo. Partitioning with two lines in the plane. *J. Algorithms*, 6(3):430–433, 1985.

[22] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. On testing convexity and submodularity. *SIAM J. Comput.*, 32(5):1158–1184, 2003.

[23] Georg Pick. Geometrisches zur zahlenlehre. *Sitzenber. Lotos (Prague)*, 19:311–319, 1899.

[24] Luis Rademacher and Santosh Vempala. Testing geometric convexity. In *36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2004, December 16-18, 2004, Chennai, India*, pages 469–480, 2004.

[25] Sofya Raskhodnikova. Approximate testing of visual properties. In *RANDOM-APPROX*, pages 370–381, 2003.

[26] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.*, 39(3):813–842, 2009.

[27] Sofya Raskhodnikova and Grigory Yaroslavtsev. Learning pseudo-Boolean k-DNF and submodular functions. In Sanjeev Khanna, editor, *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1356–1368. SIAM, 2013.

[28] H Raynaud. Sur l'enveloppe convexe des nuages de points aleatoires dans r n. i. *Journal of Applied Probability*, 7(1):35–48, 1970.

[29] Alfréd Rényi and Rolf Sulanke. Über die konvexe hülle von n zufällig gewählten punkten. *Probability Theory and Related Fields*, 2(1):75–84, 1963.

[30] Dana Ron and Gilad Tsur. Testing properties of sparse images. *ACM Trans. Algorithms*, 10(4):17:1–17:52, August 2014.

[31] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[32] Bernd Schmeltz. Learning convex sets under uniform distribution. In *Data Structures and Efficient Algorithms, Final Report on the DFG Special Joint Initiative*, pages 204–213, 1992.

[33] C. Seshadhri and Jan Vondrák. Is submodularity testable? *Algorithmica*, 69(1):1–25, 2014.

[34] Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences.* John Wiley & Sons, Inc., New York, 2001.

[35] John W Tukey. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, pages 523–531, 1975.

A Every Convex Set in $\mathbb{R}^2$ Can Be Approximated by a Rectangle

Recall that for a set (region) R, its area is denoted $A(R)$.

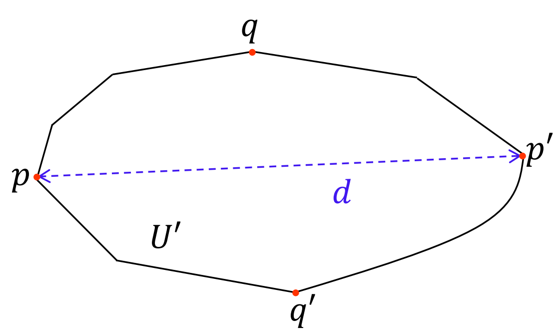

Figure A.1: Points p, p', q, and q'.

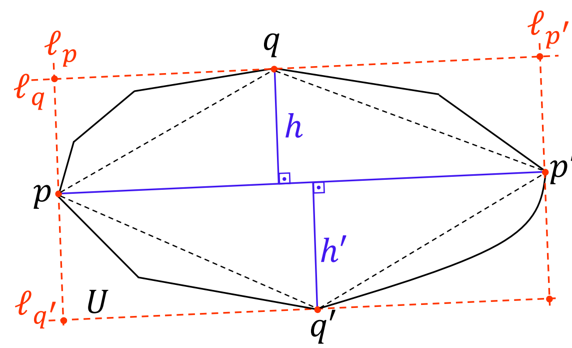

Figure A.2: Lines $\ell_p, \ell_{p'}, \ell_q$, and $\ell_{q'}$.

Lemma A.1. *Let U' be a 2-dimensional convex set. Then U' is contained in a rectangle U such that*

$$A(U) \leq 2 \cdot A(U').$$

Proof. Let d be the diameter of U'. Let p and p' be two points on the boundary of U' at distance d from each other (i.e., $|pp'| = d$). Let H and H' be the two closed half-planes defined by the line $\ell(p, p')$. Let q (respectively, q') be a point in $H \cap U'$ (respectively, a point in $H' \cap U'$) that is furthest from $\ell(p, p')$. See Figure A.1. Let h (respectively, h') be the height of $\triangle pp'q$ from q (respectively, of $\triangle pp'q'$ from q') to its side pp'. Let ℓ_p and $\ell_{p'}$ be the lines through p and p', respectively, that are perpendicular to $\ell(p, p')$. Let ℓ_q and $\ell_{q'}$ be the lines through q and q', respectively, that are parallel to $\ell(p, p')$. See Figure A.2. Note that the lines $\ell_p, \ell_{p'}, \ell_q$, and $\ell_{q'}$ form a rectangle with side lengths d and $h + h'$. Let U denote that rectangle. Since U' is convex, $A(U') \geq A(pqp'q') = A(\triangle pqp') + A(\triangle pq'p') = \frac{1}{2} \cdot (|pp'| \cdot h + |pp'| \cdot h') = \frac{d \cdot (h+h')}{2} = \frac{A(U)}{2}$. Thus, $A(U) \leq 2 \cdot A(U')$. $\square$