Testing Geometric Convexity

Luis Rademacher** and Santosh Vempala**

Mathematics Department and CSAIL, MIT.
Email: {lrademac, vempala}@math.mit.edu

Abstract. We consider the problem of determining whether a given set S in $\mathbb{R}^n$ is *approximately* convex, i.e., if there is a convex set $K \in \mathbb{R}^n$ such that the volume of their symmetric difference is at most $\epsilon \operatorname{vol}(S)$ for some given ϵ. When the set is presented only by a membership oracle and a random oracle, we show that the problem can be solved with high probability using $\mathsf{poly}(n)(c/\epsilon)^n$ oracle calls and computation time. We complement this result with an exponential lower bound for the natural algorithm that tests convexity along "random" lines. We conjecture that a simple 2-dimensional version of this algorithm has polynomial complexity.

1 Introduction

Geometric convexity has played an important role in algorithmic complexity theory. Fundamental problems (sampling, optimization, etc.) that are intractable in general can be solved efficiently with the assumption of convexity. The algorithms developed for these problems assume that the input is a convex set and are often not well-defined for arbitrary sets. Nevertheless, sampling-based approaches for optimization might be extendable to approximately convex sets, since there is hope that approximately convex sets can be sampled efficiently. This raises a basic question: How can we test if a given compact set in $\mathbb{R}^n$ is convex? Similarly, do short proofs of convexity or non-convexity of a set exist? Can one find these proofs efficiently?

To address these questions, we first need to decide how the set (called S henceforth) is specified. At the least, we need a membership oracle, i.e., a black-box that takes as input a point $x \in \mathbb{R}^n$ and answers YES or NO to the question "Does x belong to S?" This is enough to prove that a set is not convex. We find 3 points $x, y, z \in R^n$ such that $x, z \in S$, $y \in [x, z]$ and $y \notin S$. Since a set is convex iff it is convex along every line, such a triple constitutes a proof of non-convexity.

On the other hand, how can we prove that a set *is* convex? Imagine the perverse situation where a single point is deleted from a convex set. We would have to test an uncountable number of points to detect the non-convexity. So we relax the goal to determining if a set is approximately convex. More precisely,

** Partially supported by NSF grant CCR-0307536

given $0 < \epsilon \leq 1$, either determine that S is not convex or that there is a convex set K such that

$$\mathrm{vol}(S \setminus K) + \mathrm{vol}(K \setminus S) \leq \epsilon \, \mathrm{vol}(S) \ .$$

In words, the condition above says that at most an ϵ fraction of S has to be changed to make it convex. We will call this the problem of testing approximate convexity.

This formulation of the problem fits the *property testing* framework developed in the literature ([1]). In fact there has been some work on testing convexity of discrete 1-dimensional functions ([2]), but the general problem is open.

Testing approximate convexity continues to be intractable if S is specified just by a membership oracle. Consider the situation where a small part of S is very far from the rest. How do we find it? To counter this, we assume that we also have access to uniform random points in S, i.e., a random oracle[1]. (There are other alternatives, but we find this to be the cleanest). In this paper, we address the question of testing approximate convexity of a set given by a membership oracle and a random oracle. The complexity of an algorithm is measured by the number of calls to these oracles and the additional computation time.

We begin with a proof that the problem is well-defined, i.e., there exists a closest convex set. Then we give a simple algorithm with complexity $\mathsf{poly}(n)(c/\epsilon)^n$ for any set S in $\mathbb{R}^n$. The algorithm uses random sampling from a convex polytope as a subroutine. Next, we consider what is perhaps the most natural algorithm for testing approximate convexity: get a pair of random points from the set and test if the intersection of the line through them with S is convex. This is motivated by the following conjecture: If the intersection of S with "most" lines is convex, then S itself is approximately convex. Many property testing algorithms in the literature have this flavor, i.e., get a random subset and test if the subset has the required property. Surprisingly, it turns out that the number of tests needed can be *exponential* in the dimension. We construct an explicit family of sets for which the lines through most (all but an exponentially small fraction) pairs of points have convex intersections with the set (i.e., they intersect S in intervals), yet the set is far from convex. Finally, we conjecture that if "most" 2-dimensional sections of a set S are convex, then S is approximately convex.

---

[1] A non-trivial example where testing approximate convexity makes sense and the oracles are naturally available is testing approximate convexity of the union of m convex bodies given by membership oracles. In this case, the individual membership oracles give a membership oracle for the union. Also, the membership oracles can simulate random oracles for every convex set (approximately, see [3]), and allow us to approximate the volumes of the convex bodies. Finally, by using a technique similar to the one used to approximate the number of satisfying assignments of a DNF formula (see [4], for example), one can simulate a random oracle for the union (approximately) by means of the individual membership and random oracles and the individual volumes, in time polynomial in m and the other parameters.

2 Preliminaries

The following notation will be used. Let $S \subseteq \mathbb{R}^n$. If S is measurable, $\mathrm{vol}(S)$ denotes the volume of S. The convex hull of S is denoted $\mathrm{conv}(S)$. Let $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$, the usual inner product in $\mathbb{R}^n$.

Let $A, B \subseteq \mathbb{R}^n$ be measurable sets. The *symmetric difference measure distance* (or simply, *distance*) between A and B is

$$d(A, B) = \mathrm{vol}(A \Delta B) .$$

Let $\mathcal{K}$ denote the set of all compact convex sets in $\mathbb{R}^n$ with nonempty interior, and the empty set.

Proposition 1. *Let $S \subseteq \mathbb{R}^n$ compact. Then $\inf_{C \in \mathcal{K}} d(S, C)$ is attained.*

Proof. The set $\mathcal{K}$ with distance d is a metric space. The selection theorem of Blaschke (see the appendix) implies that $\{C \in \mathcal{K}, C \subseteq \mathrm{conv}\, S\}$ is compact. Moreover, $d(S, \cdot) : \mathcal{K} \to \mathbb{R}$ is continuous. Also, it is sufficient to consider convex sets contained in $\mathrm{conv}\, S$, that is,

$$\inf_{C \in \mathcal{K}} d(S, C) = \inf_{C \in \mathcal{K}, C \subseteq \mathrm{conv}\, S} d(S, C) .$$

The last expression is the infimum of a continuous function on a compact set, thus it is attained. $\square$

Definition 2. *Given $S \subseteq \mathbb{R}^n$ compact, a set $C \in \mathrm{argmin}_{C \in \mathcal{K}} d(S, C)$ is called a* closest convex set *of S. S is said to be* ϵ-convex *iff $d(S, C) \leq \epsilon\, \mathrm{vol}(S)$.*

3 Algorithms for Testing Approximate Convexity

We are interested in the following algorithmic problem:

Let $S \subseteq \mathbb{R}^n$ be compact. We are given a *membership oracle* that given $x \in \mathbb{R}^n$ answers "YES" if $x \in S$ and "NO" if $x \notin S$; we also have access to a *random oracle* that when called gives a uniformly sampled random point from S. For any given $\epsilon > 0$, our goal is to determine either that S is ϵ-convex (output "YES") or that S is not convex (output "NO").

In this section, we will give a randomized algorithm for the problem. We will prove that the algorithm works with probability at least 3/4. This can be easily boosted to any desired $1 - \delta$ while incurring an additional factor of $O(\ln(1/\delta))$ in the complexity.

3.1 The One-Dimensional Case

One-dimensional algorithm
INPUT: Access to membership and random oracles of $S \subseteq \mathbb{R}$.

1. Get $12/\epsilon$ points from the random oracle. Let C be their convex hull (the interval containing them).
2. Choose $12/\epsilon$ random points in C. Check if they are all in S using the membership oracle. If so, output "YES", else output "NO".

Theorem 3. *With probability at least* $3/4$*, the one-dimensional algorithm determines that* S *is not convex or that* S *is* ϵ*-convex.*

Proof. Clearly, if S is convex then the algorithm answers "YES". So assume that S is not ϵ-convex. We say that the first step succeeds if we get at least one point in the leftmost $\epsilon/4$ fraction of S and another point in the rightmost $\epsilon/4$ fraction of S. The first step fails with probability at most $2(1-\epsilon/4)^{12/\epsilon} \le 2/e^3$. Suppose the first step succeeds. Then,

$$\mathrm{vol}(S \setminus C) \le \mathrm{vol}(S)\frac{\epsilon}{2} \; .$$

This implies that

$$\mathrm{vol}(C \setminus S) \ge \mathrm{vol}(S)\frac{\epsilon}{2} \; .$$

From this, we get

$$\begin{aligned} \mathrm{vol}(C \setminus S) &\ge \max\left\{\frac{\epsilon}{2}\,\mathrm{vol}(S), \mathrm{vol}(C) - \mathrm{vol}(S)\right\} \\ &= \mathrm{vol}(C)\max\left\{\frac{\epsilon}{2}\frac{\mathrm{vol}(S)}{\mathrm{vol}(C)}, 1 - \frac{\mathrm{vol}(S)}{\mathrm{vol}(C)}\right\} \; . \end{aligned} \tag{1}$$

Given that $\epsilon > 0$, the expression

$$\max\left\{\frac{\epsilon}{2}\alpha, 1-\alpha\right\}$$

is minimized as a function of α when $\frac{\epsilon}{2}\alpha = 1-\alpha$, i.e., for $\alpha = \frac{2}{\epsilon+2}$. Thus, from Equation (1) we get

$$\mathrm{vol}(C \setminus S) \ge \frac{\epsilon}{2+\epsilon}\,\mathrm{vol}(C) \; .$$

That is, conditioned on the success of the first step, with probability at least $1-(1-\epsilon/3)^{12/\epsilon} \ge 1-1/e^4$ the algorithm answers "NO". Thus, overall the algorithm answers "NO" with probability at least $(1-1/e^4)(1-2/e^3) \ge 3/4$. □

3.2 The General Case

Here we consider the problem in $\mathbb{R}^n$. It is not evident that the time complexity of the problem can be made independent of the given set S (that is, depending only on ϵ and the dimension). The following algorithm shows such independence ($m = m(\epsilon, n)$ will be chosen later).

n-dimensional algorithm
INPUT: Access to membership and random oracles of $S \subseteq \mathbb{R}^n$.

1. Get m random points from S. Let C be their convex hull.
2. Get $4/\epsilon$ random points from S. If any of them is not in C, output "NO".
3. Get $6/\epsilon$ random points from C. If each of them is in S according to the membership oracle, then output "YES", else output "NO".

Checking if a point y belongs to C is the same as answering whether y can be expressed as a convex combination of the m points that define C. This can be done by solving a linear program. The third step requires random points from C, which is a convex polytope. Sampling convex bodies is a well-studied algorithmic problem and can be done using $O^*(n^3)$ calls to a membership oracle (see [3], for example).

To prove the correctness of the algorithm we will use the following lemmas (the first is from [5] and the second is paraphrased from [6]).

Lemma 4. *Let $C = \operatorname{conv}\{X_1, \ldots, X_m\}$, where the X_i's are independent uniform random samples from a convex body K. Then for any integer $t > 0$, $\mathbb{E}\big((\operatorname{vol}(C)/\operatorname{vol}(K))^t\big)$ is minimized iff K is an ellipsoid.*

Lemma 5. *Let $B_n \subseteq \mathbb{R}^n$ be the unit ball. Let $C = \operatorname{conv}\{X_1, \ldots, X_m\}$, where the X_i's are independent uniform random samples from B_n. There exists a constant c such that, for $m = (cn/\epsilon)^n$,*

$$\mathbb{E}\big(\operatorname{vol}(B_n \setminus C)\big) \leq \epsilon \operatorname{vol}(B_n) .$$

Theorem 6. *Using $m = (224cn/\epsilon)^n$ random points and $\mathsf{poly}(n)/\epsilon$ membership calls, the n-dimensional algorithm determines with probability at least $3/4$ that S is not convex or that S is ϵ-convex.*

Proof. First, assume that S is convex. We want to show that the algorithm outputs "YES" with probability at least $3/4$. Let $X = \operatorname{vol}(S \setminus C)/\operatorname{vol}(S)$. Then by Lemma 4, $\mathbb{E}(X)$ is maximized when K is a ball and using Lemma 5 with our choice of m, we get that

$$\mathbb{E}(X) \leq \frac{\epsilon}{224n} .$$

By Markov's inequality, with probability at least $6/7$,

$$\operatorname{vol}(S \setminus C) \leq \frac{\epsilon}{32} \operatorname{vol}(S) .$$

Given this, Markov's inequality implies that the algorithm will not stop at step 2 with probability at least $3/4$: in step 2, if we let Y be the number of points not in C then

$$\mathbb{E}(Y) \leq \frac{\epsilon}{32} \frac{4}{\epsilon} = \frac{1}{8} ,$$

and therefore, by Markov's inequality,

$$\mathbb{P}(\text{algorithm outputs “NO” in step 2}) = \mathbb{P}(Y \geq 1) = \mathbb{P}\big(Y \geq 8\mathbb{E}(Y)\big) \leq \frac{1}{8} .$$

Thus, the algorithm outputs "YES" with probability at least $\frac{6}{7}\frac{7}{8} = \frac{3}{4}$.

Next, if S is not ϵ-convex, the analysis can be divided into two cases after the first step: either $\operatorname{vol}(S \setminus C) \geq \operatorname{vol}(S)\epsilon/2$ or $\operatorname{vol}(S \setminus C) < \operatorname{vol}(S)\epsilon/2$. In the first case, step 2 outputs "NO" with probability at least $1-\left(1-\frac{\epsilon}{2}\right)^{4/\epsilon} \geq 1-\frac{1}{e^2} \geq \frac{3}{4}$. In the second case we have

$$\mathrm{vol}(C \setminus S) \geq \frac{\epsilon}{2} \mathrm{vol}(S)$$

and by the same analysis as the one-dimensional case, $\mathrm{vol}(C \setminus S) \geq \frac{\epsilon}{3} \mathrm{vol}(C)$. Thus, step 3 outputs "NO" with probability at least $1-(1-\frac{\epsilon}{3})^{6/\epsilon} \geq 3/4$. □

Note that, unlike the one-dimensional case, this algorithm has two-sided error. The complexity of the algorithm is independent of S and depends only on n and ϵ. It makes an exponential number of calls to the random oracle and this dependency is unavoidable for this algorithm. It is known for example that the convex hull of any subset of fewer than c^n points of the ball, contains less than half its volume [7].

The one-dimensional algorithm suggests another algorithm for the general case: let $\ell(x, y)$ be the line through x and y,

Lines-based algorithm
INPUT: Access to membership and random oracles of $S \subseteq \mathbb{R}^n$ compact.

Generate m pairs of random points (x, y) and test if $\ell(x, y) \cap S$ is convex.

How large does m need to be? Somewhat surprisingly, we show in the next section that this algorithm also has an exponential complexity. Testing if $\ell(x, y) \cap S$ is convex is not a trivial task (note that we have a membership oracle for $\ell(x, y) \cap S$ from the oracle for S, but simulating a random oracle is not so simple). However, for the purpose of showing a lower bound in m we will assume that the one-dimensional algorithm checks *exactly* whether $\ell(x, y) \cap S$ is convex (that is, it is an interval).

4 The Lines-Based Algorithm Is Exponential

In this section, we construct an explicit family of compact sets each of which has the following properties: (i) the set is far from convex, and (ii) for all but an exponentially small fraction of pairs of points from the set, the line through the pair of points has a convex intersection with the set. This implies that the lines-based algorithm (described at the end of Section 3.2) has exponential worst-case complexity. Thus, although exact convexity is characterized by "convex along every line," the corresponding reduction of approximate convexity to "convex along most lines" is not efficient.

The proof of the lower bound is in two parts, first we show that the algorithm needs many tests and then that the test family is far from convex (i.e., ϵ is large).

4.1 The Family of Sets: the Cross-Polytope with Peaks

The n-dimensional cross-polytope is an n-dimensional generalization of the octahedron and can be defined as the unit ball with the norm $|x|_1 = \sum_{i=1}^{n} |x_i|$. Let T_n be the "cross-polytope with peaks", that is, the union of the cross-polytope and, for each of its facets $i \in \{1, \ldots, 2^n\}$, the convex hull of the facet and a point $v_i = \lambda d$, where d is the unit outer normal to the facet and $\lambda \geq 1/\sqrt{n}$ is a parameter (that may depend on the dimension). Informally, one adds an n-dimensional simplex on top of each facet of the cross-polytope. The volume of the cross-polytope is a $\frac{1}{\lambda\sqrt{n}}$ fraction of the volume of T_n. We will choose $\lambda = \frac{\sqrt{n}}{n-2}$. In that case, the cross-polytope as a convex set shows that T_n is $O(\frac{1}{n})$-convex. We will prove that T_n is not $\frac{1}{12n^2}$-convex, i.e., for any convex set K, we have $d(K, T_n) > \frac{1}{12n^2} \operatorname{vol} T_n$.

4.2 The Non-convexity of the Family Cannot Be Detected by the Lines-Based Algorithm

Proposition 7. *If $\lambda \leq \frac{\sqrt{n}}{n-2}$ then the one-dimensional test has an exponentially low probability of detecting the non-convexity of the cross-polytope with peaks.*

Proof. First, we will prove the following claim:

> Under the hypothesis, every peak is contained in the intersection of the half-spaces determining the n facets of the cross-polytope adjacent to the peak.

It is enough to see that the point $v_i = \lambda d$ (a vertex of the peak) is contained in that intersection. Because of the symmetry, we can concentrate on any particular pair of adjacent facets, say those having normals $d = (1, 1, \ldots, 1)/\sqrt{n}$ and $d' = (-1, 1, \ldots, 1)/\sqrt{n}$. The halfspace determining the facet with normal d is given by $\{x \in \mathbb{R}^n \;:\; \langle x, d' \rangle \leq 1/\sqrt{n}\}$. Then $v_i = \lambda d$ is contained in the halfspace associated to the facet with normal d' (which is sufficient) if

$$\langle \lambda d, d' \rangle \leq \frac{1}{\sqrt{n}} \,.$$

That is,

$$\lambda \leq \frac{\sqrt{n}}{n-2} \,.$$

This proves the claim.

It is sufficient to note that, for the algorithm to answer "NO", we need to choose a line whose intersection with T_n is not convex. Suppose that a line L shows non-convexity. Then it does not intersect the cross-polytope part of T_n a.s. (almost surely), otherwise L intersects exactly 2 facets of the cross-polytope a. s., and intersects only the peaks that are associated to those facets, because of the claim (if one follows the line after it leaves the cross-polytope through one of the facets, it enters a peak, and that peak is the only peak on that side of the facet, because of the claim), and thus $L \cap T_n$ would be convex. Now, while intersecting a peak, L intersects two of its facets at two points that are not at the same distance of the cross-polytope, a.s. The half of L that leaves the peak through the farthest point cannot intersect any other peak because of the claim (the halfspace determined by the respective facet of the cross-polytope containing this peak contains only this peak, and this half of L stays in this halfspace). The half of L that leaves the peak through the closest point will cross the hyperplane determined by one of the adjacent peaks[2] before intersecting any other peak, a.s.; after crossing that hyperplane it can intersect only one peak, namely, the peak associated to that hyperplane, because of the claim. Thus, L has to intersect exactly 2 peaks that have to be adjacent a. s., and L does not intersect the cross-polytope. In other words, the two random points that determine L are in the same peak or in adjacent peaks. The probability of this event is no more than $\frac{n+1}{2^n}$. □

4.3 The Sets in the Family Are Far from Convex

To prove that T_n is far from being convex, we will prove that a close convex set must substantially cover most peaks, and because of this, a significant volume of a close convex set must lie between pairs of adjacent substantially covered peaks, outside of T_n, adding to the symmetric difference. The following lemma will be useful for this part. For $A \subseteq \mathbb{R}^n$ and H a hyperplane and $v \in \mathbb{R}^n$ a unit normal for H, let

$$w_H(A) = \sup_{x \in A} \langle v, x \rangle - \inf_{x \in A} \langle v, x \rangle \ .$$

Lemma 8. *Let $A, B \subseteq \mathbb{R}^n$ compact. Let H be a separating hyperplane[3] for A, B. Let $C = H \cap \operatorname{conv}(A \cup B)$. Then*

$$V_{n-1}(C) \geq \min \left\{ \frac{\operatorname{vol} A}{w_H(A)}, \frac{\operatorname{vol} B}{w_H(B)} \right\} \ .$$

Proof. There exist sections, parallel to H, of A and B that have $(n-1)$-volumes at least $(\operatorname{vol} A)/w_H(A)$ and $(\operatorname{vol} B)/w_H(B)$, respectively. That is, there exist $a, b \in \mathbb{R}^n$ such that $A' = (H + a) \cap A$, $B' = (H + b) \cap B$ satisfy $V_{n-1}(A') \geq (\operatorname{vol} A)/w_H(A)$ and $V_{n-1}(B') \geq (\operatorname{vol} B)/w_H(B)$. Clearly $H \cap \operatorname{conv}(A' \cup B') \subseteq C$ and therefore

$$\begin{aligned} V_{n-1}(C) &\geq V_{n-1}(H \cap \operatorname{conv}(A' \cup B')) \\ &\geq \min\{V_{n-1}(A'), V_{n-1}(B')\} \\ &\geq \min \left\{ \frac{\operatorname{vol} A}{w_H(A)}, \frac{\operatorname{vol} B}{w_H(B)} \right\} \ . \end{aligned}$$

□

[2] "The hyperplane determined by a peak" is the unique hyperplane that contains the facet of the cross-polytope associated to the peak.

[3] That is, a set of the form $H = \{x \in \mathbb{R}^n \ : \ \langle x, y \rangle = \alpha\}$ for some $y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, such that for all $x \in A$ we have $\langle x, y \rangle \leq \alpha$ and for all $x \in B$ we have $\langle x, y \rangle \geq \alpha$.

This bound is sharp: consider a cylinder with a missing slice in the middle, that is, consider in the plane as A a rectangle with axis-parallel sides and non-adjacent vertices $(1,0)$ and $(2,1)$, as B the reflection of A with respect to the y-axis and as the separating line, the y-axis.

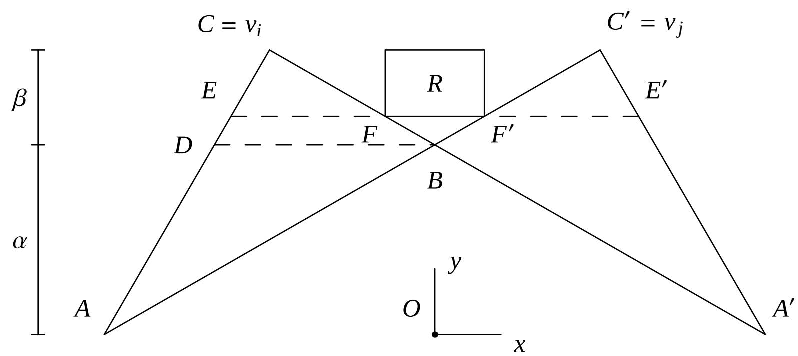

Fig. 1. Projection of the peaks (i,j) of the cross-polytope with peaks onto v_i, v_j for $n=4$.

Lemma 9. *For* $\lambda = \frac{\sqrt{n}}{n-2}$, T_n *is not* $\frac{1}{12n^2}$*-convex.*

Proof. Let C_n be a closest convex set to T_n.

Consider a pair of adjacent peaks (i,j). Figure 1 shows the projection of the pair onto the plane containing the vertices v_i, v_j and the origin. B is the projection of the intersection of the two peaks, an $(n-2)$-dimensional simplex. A and C are the other two vertices of one of the peaks, A' and C' are the respective vertices of the other peak. The plane is orthogonal to the two respective facets of the cross-polytope, the segment AB is the projection of one of them and $A'B$ is the projection of the other facet. D is such that DB is orthogonal to OB, where O is the origin.

First, we will prove that the volume of the preimage (with respect to the projection) of the triangle DBC is a $\frac{1}{n-1}$ fraction of the volume of the peak. To see this, let Q be the preimage of DB in the peak, which is a $(n-1)$-dimensional simplex. Let α be the height of the triangle ABD with respect to A, and let β be the height of the triangle DBC with respect to C. Then the volume of the peak is

$$\frac{1}{n}V_{n-1}(Q)(\alpha+\beta) .$$

Also, the volume of the preimage of DBC is

$$\frac{1}{n}V_{n-1}(Q)\beta .$$

Thus, the volume of the preimage of the triangle DBC is a $\frac{\beta}{\alpha+\beta}$ fraction of the volume of the peak. We can compute α and β. Without loss of generality we can assume that v_i is parallel to $(-1, 1, \ldots, 1)$ and v_j is parallel to $(1, \ldots, 1)$. Then $(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1})$ is a vector in the preimage of B that is in the projection plane, and α is the norm of that vector, that is, $\alpha = 1/\sqrt{n-1}$. An orthonormal basis of the projection plane corresponding to the x, y axes of Figure 1 is

$$\{(1, 0, \ldots, 0), (0, 1/\sqrt{n-1}, \ldots, 1/\sqrt{n-1})\}.$$

Then, $\alpha+\beta$ is the length of the projection of v_j onto $(0, 1/\sqrt{n-1}, \ldots, 1/\sqrt{n-1})$, that is, $\alpha + \beta = \frac{\sqrt{n-1}}{n-2}$ and $\beta = \frac{1}{(n-2)\sqrt{n-1}}$. Thus, $\frac{\beta}{\alpha+\beta} = \frac{1}{n-1}$, as claimed.

EF is a segment parallel to DB and at a distance $\frac{\beta}{n+1}$ from it. That way, the volume of the preimage of the triangle EFC is a $(1 - \frac{1}{n+1})^n \geq \frac{1}{e}$ fraction of the volume of the preimage of the triangle DBC, which, as we saw, is a $\frac{1}{n-1}$ fraction of the volume of the peak. That is, the preimage of the triangle EFC is at least a $\frac{1}{e(n-1)}$ fraction of the volume of the peak.

Given a particular peak, we will say that it is *substantially covered* (by C_n) iff the volume of the intersection of C_n and the peak is at least a $1 - \frac{1}{2e(n-1)}$ fraction of the volume of the peak. Because of the choice of EF, if a peak is substantially covered, then at least a $\frac{1}{2e(n-1)}$ fraction of its volume is covered in the preimage of the triangle EFC (that is, above the segment EF).

Now we will prove that every pair of adjacent substantially covered peaks contributes to $C_n \setminus T_n$ at least with a $\frac{1}{6n^2}$ fraction of the volume of a peak, disjoint from the contribution of other pairs. To see this, let U be the subset of C_n intersected with peak i that projects onto EFC and let V be the subset of C_n intersected with peak j that projects onto $F'E'C'$. We will apply Lemma 8 to U, V and every hyperplane which is a preimage of a vertical line intersecting the rectangle R. Moreover, for any such hyperplane H we have that $w_H(U)$ and $w_H(V)$ are no more than the length of DB, which is a $\frac{\beta}{\alpha+\beta} = \frac{1}{n-1}$ fraction of the length of AA' (which is 2), i.e., $\frac{2}{n-1}$. Certainly $W = R \cap \operatorname{conv}(U \cup V)$ is contained in C_n and disjoint from T_n. Because of the choice of EF, the width of the rectangle R is a $\frac{1}{n+1}$ fraction of the distance between C and C', that is, $\frac{2}{(n+1)(n-2)}$. Also, $\operatorname{vol} U$ and $\operatorname{vol} V$ are no less than a $\frac{1}{2e(n-1)}$ fraction of the volume of a peak. Lemma 8 gives that

$$\begin{aligned}
\frac{\operatorname{vol} W}{\operatorname{vol}(\text{one peak})} &\geq (\text{width of } R) \min\left\{\frac{\operatorname{vol} U}{\frac{2}{n-1}}, \frac{\operatorname{vol} V}{\frac{2}{n-1}}\right\} \frac{1}{\operatorname{vol}(\text{one peak})} \\
&\geq \frac{2}{(n+1)(n-2)} \frac{n-1}{2} \frac{1}{2e(n-1)} \\
&\geq \frac{1}{2e(n-2)(n+1)} \\
&\geq \frac{1}{6n^2}\,.
\end{aligned}$$

Let $\epsilon(n) = d(C_n, T_n)$. We claim that the number of peaks that are not substantially covered is a fraction that is at most $en^2\epsilon(n)$ of the total number of peaks. To see this, let $q(n)$ be the fraction of the volume of T_n that the peaks contain. Clearly

$$q(n) = \frac{\lambda - \frac{1}{\sqrt{n}}}{\lambda} = \frac{2}{n} \; .$$

Let X be the number of peaks that are not substantially covered. Then,

$$X \frac{1}{2e(n-1)} q(n) \leq \epsilon(n) \; ,$$

that is,

$$X \leq en(n-1)\epsilon(n) \leq en^2\epsilon(n) \; . \tag{2}$$

We will see now that eventually (as n grows) the number of pairs of adjacent peaks that are substantially covered is a substantial fraction of the total number of adjacent pairs. For a contradiction, assume that, for some subsequence, $\epsilon(n) < \frac{1}{12n^2}$. For n sufficiently large, $en^2\epsilon(n) \leq 1/4$. The number of peaks is 2^n; the number of (unordered) pairs of adjacent peaks is $n2^{n-1}$. A peak that is not substantially covered can participate in at most n pairs of adjacent peaks. Because of (2), there are at most $\frac{1}{4}2^n = 2^{n-2}$ peaks that are not substantially covered (for large n and a subsequence). That way, all the peaks that are not substantially covered can participate in at most $n2^{n-2} = \frac{1}{2}n2^{n-1}$ pairs of adjacent peaks. Thus, at least $1/2$ of the pairs of adjacent peaks involve only substantially covered peaks. For γ equal to the volume of the contribution to $C_n \setminus T_n$ of a pair of substantially covered peaks, this implies that

$$\begin{aligned}
\epsilon(n) &\geq \frac{\operatorname{vol}(C_n \setminus T_n)}{\operatorname{vol} T_n} \\
&\geq \frac{\operatorname{vol}(C_n \setminus T_n)}{\operatorname{vol}(\text{all peaks})} \frac{\operatorname{vol}(\text{all peaks})}{\operatorname{vol} T_n} \\
&\geq \frac{\frac{1}{2}n2^{n-1}\gamma}{2^n \operatorname{vol}(\text{one peak})} q(n) \\
&\geq \frac{n}{4} \frac{1}{6n^2} \frac{2}{n} \\
&\geq \frac{1}{12n^2}
\end{aligned}$$

which is a contradiction. □

5 An Algorithm Based on Planes

In this section, we state a conjecture about approximate convexity. Let S be a compact subset of $\mathbb{R}^n$ whose center of gravity is the origin. For a pair of points $x, y \neq 0$ in $\mathbb{R}^n$ let the subspace spanned by them be $H(x, y)$ and define

$P(x,y) = S \cap H(x,y)$ to be the part of S on this subspace. Our conjecture is the following:

Conjecture. Let μ be the distribution on 2-dimensional sections $P(x,y)$ obtained by picking x and y uniformly at random from S. If

$$\mathbb{P}_\mu\big(P(x,y) \text{ is convex}\big) > 1-\epsilon\ ,$$

then S is $O(n\epsilon)$-convex.

The conjecture motivates the following algorithm (here $p(\cdot)$ and $q(\cdot)$ are fixed polynomials):

Repeat $p(n, 1/\epsilon)$ times

1. Get random points x, y from S.
2. Test if $P(x,y)$ is $q(1/n,\epsilon)$-convex.

Acknowledgements

We would like to thank David Jerison, Dana Ron and Ronitt Rubinfeld for discussions on this topic. We also thank an anonymous member of the program committee for suggesting an interesting application.

References

1. Goldreich, O., Goldwasser, S., Ron, D.: Property testing and its connection to learning and approximation. J. ACM **45** (1998) 653–750
2. Parnas, M., Ron, D., Rubinfeld, R.: On testing convexity and submodularity. SIAM J. Comput. **32** (2003) 1158–1184 (electronic)
3. Lovász, L., Vempala, S.: Hit-and-run from a corner. In: STOC '04. ACM, New York (2004)
4. Motwani, R., Raghavan, P.: Randomized Algorithms. Cambridge University Press (1995)
5. Groemer, H.: On the mean value of the volume of a random polytope in a convex set. Arch. Math. (Basel) **25** (1974) 86–90
6. Bárány, I., Buchta, C.: Random polytopes in a convex polytope, independence of shape, and concentration of vertices. Math. Ann. **297** (1993) 467–497
7. Bárány, I., Füredi, Z.: Computing the volume is difficult. Discrete Comput. Geom. **2** (1987) 319–326
8. Rockafellar, R.T., Wets, R.J.: Variational Analysis. Number 317 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag (1998)

Appendix

For the Hausdorff metric or the symmetric difference volume metric, we have (see [8], Theorem 4.18, for example):

Theorem 10 (Blaschke's selection theorem). *In $\mathbb{R}^n$, any bounded sequence $(C_k)_{k\in\mathbb{N}}$ of nonempty, convex sets has a subsequence converging to some nonempty, compact, convex set C.*