Testing Geometric Objects*

Kathleen Romanik
Department of Computer Science
University of Maryland
College Park, MD 20742

Carl Smith†
Department of Computer Science
University of Maryland
College Park, MD 20742

1 Introduction

Testing plays a part in many areas of computer science, including hardware and software design, learning and robotics. In each case some type of object is tested, such as a circuit, a program or a concept, to determine whether it is correct. For some applications it is sufficient to demonstrate that a tested object is close to the target. For example, a vision system may need to align the position of its camera to within a certain accuracy according to the location of an object in its field of vision. By testing the location where the object would appear when the system is aligned correctly, it can determine if it is within ϵ of the correct alignment. In this situation, the object being used to align the system can be represented as a geometric shape, and the test being used for alignment can determine if a given point in the field of vision is located on the object. This is a probabilistic approach in that the system only determines that it is within ϵ of its correct alignment. Our research develops a general theory of testing with probabilistic error bounds and applies this theory to the testing of geometric shapes.

We define what it means for a class of objects to be testable and we examine various geometric classes to determine their testability. The ideas we use are derived from the probabilistic model of learning developed by Valiant [Val84] and extended by Blumer, Ehrenfeucht, Haussler and Warmuth [BEHW86]. The ideas from this model were used by Haussler and Welzl [HW87] in a computational geometry setting to attack the problem of processing geometric range queries.

*A complete version of this paper appears as a University of Maryland Institute for Advanced Computer Studies technical report, UMIACS-TR-90-69, May 1990. This research was supported in part by NSF Grant CCR-8701104.

†Also supported by University of Maryland Institute for Advanced Computer Studies.

2 Testability

We give the definitions of "object class" and "testable", which are similar to those of "concept class" and "learnable" found in [BEHW86]. We also define four levels of testability - testable, k-testable, two-sided testable and untestable.

An *object* q is a measurable subset $q \subset E^d$ of d-dimensional Euclidean space; in particular, it is a Borel set. An *object class* is a set Q of objects. It is assumed that there is some probability distribution P defined over E^d. For the purposes of this abstract we take P to be the uniform distribution defined over a bounded subspace of E^d.

Given an object $r \in Q$ to be tested and a target object $q \in Q$, r is *consistent* with q on some finite set $t = \{t_1, t_2, \ldots, t_m\} \subset E^d$ if it contains the same subset of t as q. The *error* of r, with respect to q and the probability measure P, is given by $P(q \triangle r)$, where $q \triangle r$ denotes the symmetric difference of the sets. Thus, the error of the object to be tested is measured as the probability of the region that forms the symmetric difference between it and the target object.

Let S denote the set of all finite sets of points in E^d. Let I denote the open interval of rationals $(0,1)$, and let $m: I \rightarrow Z^+$ be a positive integer valued function defined on I. Furthermore, let $\mathbf{T}$ be the set of all computable functions $T: Q \times I \rightarrow S$ and P be a probability measure on E^d. Let $e: Q \times I \rightarrow \Re^+$ be a positive real valued function.

Definition. $T \in \mathbf{T}$ is a *testing algorithm for Q (w.r.t. P) with test set size $m(\epsilon)$* if for all $\epsilon \in I$ and for all $q \in Q$, $T(q,\epsilon) = t \in S, |t| \leq m(\epsilon)$ and for all $r \in Q$, if r is consistent with q on t, then $P(q \triangle r) \leq \epsilon$. $T(q,\epsilon)$ is called a *test set for q with respect to the class Q*. For each $t_i \in T(q,\epsilon)$, if $t_i \in q$ then t_i is a *positive test point*; otherwise, t_i is a *negative test point.*

Definition. $T \in \mathbf{T}$ is a *two-sided testing algorithm for Q (w.r.t. P) with error margin $e(q,\epsilon)$* if T is a testing algorithm for Q (w.r.t. P) and for all $\epsilon \in I$ and for all $q, r \in Q$, if r is inconsistent with q with respect to $T(q,\epsilon)$, then $P(q \triangle r) \geq e(q,\epsilon)$.

Thus given a target object $q \in Q$ and $\epsilon \in I$, a testing algorithm T produces a test set for q such that any tested object which is consistent with q for this set has error no more than ϵ. If such a T and m exist, then Q is *finitely testable with respect to P with test set size $m(\epsilon)$*. From now on we will refer to finitely testable object classes as simply *testable*, and since we are only examining testability with respect to a uniform distribution in this abstract, we will no longer mention P explicitly. If T is a two-sided testing algorithm, then any object which is inconsistent with q on the test set has error at least $e(q,\epsilon)$. If T produces a constant size test set (i.e. if m is a constant k), then we say that Q is *k-testable.* If no testing algorithm T exists for Q, then Q is *untestable.*

3 Testing vs. Probing

Much work has been done recently in the area of geometric probing [Ber86, CY87, Ski88], which is the algorithmic study of determining an object or some property of the object using a measuring device, or *probe.* One type of probe which is employed is a *finger probe* or *tactile probe* - a device that returns the first point of intersection between a directed line l and the object being investigated. Cole and Yap [CY87] give a finger-probing strategy for determining an n-sided convex polygon which uses $3n$ probes. Another problem in geometric probing is that of verification - counting the number of probes necessary to determine that an object q is indeed the object being probed. Verification problems give a lower bound for determination. Cole and Yap [CY87] give an optimal strategy for verifying an n-sided convex polygon with $2n$ finger probes.

Testing is similar to the problem of probing for verification, because both disciplines must show how to use a finite amount of information to verify an object. In fact, testing can be viewed as a type of geometric probing which uses *point probes* (i.e. test points) as opposed to finger probes. However, point probing is harder because the information returned by a test point is less exact than the information returned by a finger probe. For example, if a test point returns "yes", it does not indicate whether the point is on the boundary or in the interior of the object. For this reason, when we verify an object using test points we can only state that the tested object is within ϵ of the target rather than stating that it is exactly the target. However, a similar result to the one above can be found for verifying a polygon using test points.

Theorem 1. $2n$ test points are necessary and sufficient to verify an n-sided closed convex polygon from the class P of all closed convex polygons.

Proof Sketch. Given an n-sided closed convex polygon p as a target, choose for each vertex v of p a positive test point which is a distance α from v and which is equidistant from each of the two edges adjacent to v. Next, for each edge of p, choose a negative test point midway along the edge and a distance β from it (see Figure 1(a)). By convexity, any consistent polygon q which is tested must at least contain the convex hull of the positive test points and cannot contain any points in the shaded region in Figure 1(b). When α and β are decreased, the area of the region in which p and q can differ decreases; if α and β are chosen to be sufficiently small, this area becomes less than ϵ.

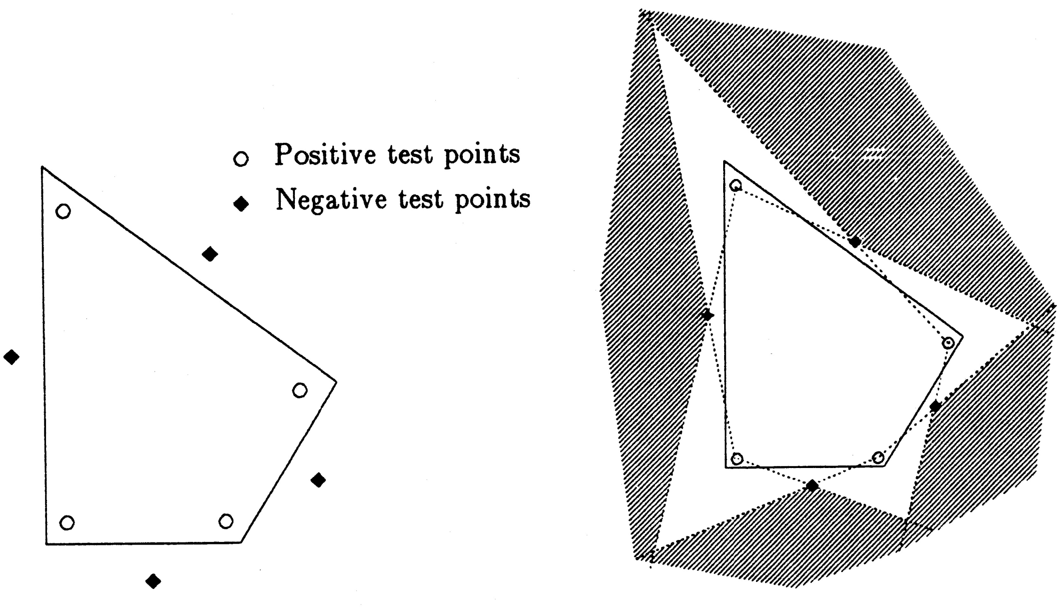
The necessity of $2n$ test points means that for any n-sided convex polygon p, there exists $\epsilon > 0$ such that $2n$ points must be used to insure that any consistent convex polygon is within ϵ. Consider the triangle formed by a line segment drawn between two adjacent edges of p near a vertex. For small enough ϵ, this triangle must contain a positive test point to insure that a consistent polygon does not exclude it. Now consider the region adjacent to an edge e of p which is formed by extending the two edges adjacent to e. This region is always nonempty, so for small enough ϵ it must contain a negative test point to insure that a consistent polygon does not extend into it. □

Theorem 2. The class P of all closed convex polygons is two-sided testable using the same test set defined in the proof of Theorem 1.

Now we show how to test orthogonally closed rectangles and closed balls in the plane.

Proposition 1. The class R of orthogonally closed rectangles in the plane is 6-testable.

Proof. Given $\epsilon > 0$ and $r \in R$ with a longest side length of w, choose positive test points at the lower left and upper right corners of r. Choose four negative test points, each midway along a side of r and α away from it, where $4\alpha(w + \alpha) < \epsilon$. If $q \in R$ is consistent with r on this test



(a) Polygon p with test points (b) Unshaded region contains q

Figure 1: Verifying a convex polygon

set, then it must contain r and cannot be larger than the area of r plus $4\alpha(w+\alpha)$. Therefore, $P(r\Delta q) \leq 4\alpha(w+\alpha) < \epsilon$. □

Proposition 2. The class B of closed balls in the plane is 4-testable.

Proof Sketch. Given $\epsilon > 0$ and $b \in B$ with radius r, draw two orthogonal diameters across b and choose the two endpoints of one diameter as positive test points and two points outside b, each a distance α from the endpoints of the other diameter, as negative test points, where $2\pi\alpha(r+\alpha) < \epsilon$. If $c \in B$ is consistent with b on the test set, then c must at least have diameter $2r$, since this is the distance between the positive test points. Also, c cannot have a radius larger than $\sqrt{\alpha^2+(r+\alpha)^2}$, since the largest ball which includes the positive test points and excludes the negative test points has this radius. Therefore, $P(b\Delta c) \leq 2\pi\alpha(r+\alpha) < \epsilon$. □

4 General Results

We wish to investigate general problems about testing. One such problem is specifying properties of an object class that determine its testability. We use the notions of Vapnik-Chervonenkis dimension (or VC-dimension) and ϵ-net [VC71, BEHW86, HW87], as well as several theorems from [BEHW86], to prove the following result for all well-behaved object classes. Well-behaved is a measure-theoretic condition on object classes given in [BEHW86]. Virtually any object class considered in the context of testing will be well-behaved.

Theorem 3. Let Q be an object class defined over E^d, and let P be a probability measure on E^d. If Q is well-behaved and has finite VC-dimension, then Q is testable with respect to P.

Proof. Given $q \in Q$ and $0 < \epsilon < 1$, we want to produce a test set t for q such that if $P(p\triangle q) > \epsilon$ for some $p \in Q$, then p is not consistent with q on t. By Lemma 9 of [BEHW86] we know that $Q_{\triangle q} = \{p\triangle q : p \in Q\}$ has the same VC-dimension as Q, namely a finite VC-dimension. Also, if Q is well-behaved, then so is $Q_{\triangle q}$. If we draw at random according to P the number of points specified in Theorem 8 of [BEHW86], then with probability at least $1 - \delta$ we will get an ϵ-net for $Q_{\triangle q}$ with respect to P. If $\delta < 1$ this probability is positive, so an ϵ-net S exists. However, an ϵ-net for $Q_{\triangle q}$ is a test set for q with respect to Q. This is because if $P(p\triangle q) > \epsilon$ for some $p \in Q$, then $(p\triangle q) \cap S \neq \emptyset$ by the definition of an ϵ-net, and therefore p will not be consistent with q on S. □

The reverse implication of this theorem does not hold. For example, in section 3 we showed that the class of all closed convex polygons was testable with test set size $2n$. However, it was shown in [HW87] that this class has infinite VC-dimension.

Closure properties are important to consider when characterizing a family of object classes. When we discuss closure properties for a family of object classes, we assume the space E^d and the probability distribution P are fixed, and operations and testability properties are taken with respect to these. For any given space E^d and probability distribution P, the family of testable object classes with respect to P is closed under intersection and set difference. This is because a test set for an object with respect to one of the original classes will also be a test set with respect to the class obtained from intersecting or finding the set difference of the original classes.

The family of testable concept classes in E^2 with respect to a uniform distribution on a bounded, nonempty subspace of E^2 is not closed under union or complementation. The union of the class R given in Proposition 1 and the class P of finite point sets is untestable. Assume T is a testing algorithm for this class, and let $T(r, \alpha)$ be the finite test set generated by T for $r \in R$ with probability 2α. An object $p \in P$ which is consistent with r on $T(r, \alpha)$ is the set of all positive test points in $T(r, \alpha)$, but $P(p\triangle r) = 2\alpha > \alpha$, so T is not a testing algorithm. Similarly, it can be shown that the complement of the class of closed orthogonal rectangles is not testable.

Results about union and complementation cannot be generalized to any known probability distribution. For example, if a probability distribution P only assigns a positive probability to a finite set of points t, then any object class is k-testable and two-sided testable with respect to P, because t is a test set for any object with respect to any class.

5 Open Problems

Since this is a new area of research, there are many open problems of which we mention a few here. First we would like to characterize exactly when an object class is testable. Also, we are interested in identifying properties which distinguish k-testable object classes and two-sided testable object classes from ones which are not k-testable or two-sided testable. We are also interested in further exploring the relationship of testing to geometric probing.

References

[BEHW86] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred Warmuth. Classifying Learnable Geometric Concepts with the Vapnik-Chervonenkis Dimension. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 273–282, Berkeley, California, May 1986.

[Ber86] Herbert J. Bernstein. Determining the Shape of a Convex n-Sided Polygon By Using $2n + k$ Tactile Probes. *Information Processing Letters*, 22(5):255–260, April 28 1986.

[CY87] Richard Cole and Chee K. Yap. Shape from Probing. *Journal of Algorithms*, 8(1):19–38, March 1987.

[HW87] David Haussler and Emo Welzl. ε-Nets and Simplex Range Queries. *Discrete Computational Geometry*, 2:127–151, 1987.

[Ski88] Steven S. Skiena. *Geometric Probing*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL, April 1988.

[Val84] Leslie G. Valiant. A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[VC71] V.N. Vapnik and A.Ya. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory Probability and its Applications*, 16(2):264–280, 1971.