SELECTING THE Kth LARGEST-AREA CONVEX POLYGON

Jeffrey S. Salowe

Computer Science Report No. TR-88-01
January 22, 1988

SELECTING THE Kth LARGEST-AREA CONVEX POLYGON

Jeffrey S. Salowe

Department of Computer Science
University of Virginia
Charlottesville, Virginia 22903

ABSTRACT

Given a set P of n points in the Euclidean plane, consider the convex polygons determined by subsets of P. We show that the problem of selecting the k^{th} largest-area convex polygon is NP-hard by a reduction from the problem of finding the k^{th} largest m-tuple [3] determined by m sets $X_1, X_2, \ldots, X_m$. The problem of finding the convex polygon with k^{th} largest area was introduced by Chazelle [1] as an example of a multidimensional selection problem for which the time complexity of selecting the median seems inherently difficult.

Introduction

The following open problem was submitted by B. Chazelle [1]: given n points in the Euclidean plane, how hard is it to compute the k^{th} largest-area convex polygon formed by any subset of the points? We answer this question by proving that this problem is, in fact, NP-hard.

Geometric selection problems are of relatively recent interest, and there are basically two paradigms for constructing efficient selection algorithms. The first of these was introduced by Chazelle [1], but it has roots in earlier selection literature. This technique consists of mapping the geometric problem onto some type of constrained matrix with a known sublinear-time selection algorithm. The second technique was introduced by [2] and is based on the idea of using parallel algorithms to create efficient sequential algorithms [5]. [6] describes and applies both techniques to construct selection algorithms.

Chazelle noted that for the problems he studied, the time complexity of selecting the largest or smallest element (extremal selection) was approximately the same as the time complexity of selecting the median. At the end of his paper, Chazelle mentioned the largest-area convex polygon problem as one in which the extremal cases are computationally tractable, but for which the selection of the median seems difficult. Note that, disregarding 2-gons, the smallest-area convex polygon determined by a set of n points in the plane is given in $O(n^2)$ time by the smallest-area triangle, and the largest-area convex polygon is given in $O(n \log n)$ time by the convex hull. For general k, we show that Chazelle was correct in suggesting that the problem is inherently difficult.

The Result

We show that the problem of finding the k^{th} largest-area convex polygon determined by a set of n points in the Euclidean plane is NP-hard by reducing the problem of finding the k^{th} largest m-tuple. We begin with a precise statement of both problems; notationally, let $conv\ (P)$ denote the convex hull of point set P.

k^{th} LARGEST m-TUPLE [3,4]

INSTANCE: Sets $X_1, X_2, \ldots, X_m \subseteq Z^+$, a size $s(x) \in Z^+$ for each $x \in X_i$, $1 \le i \le m$, and positive integers k and B.

QUESTION: Are there k or more distinct m-tuples $(x_1, x_2, \ldots, x_m)$ in $X_1 \times X_2 \times \cdots \times X_m$ for which $\sum_{i=1}^{m} s(x_i) \ge B$?

k^{th} LARGEST-AREA CONVEX POLYGON

INSTANCE: Set P of n points in the Euclidean plane, positive real B.

QUESTION: Are there k or more distinct subsets $P' \subseteq P$ for which P' equals the extreme points of $conv\ (P')$ and the area of $conv\ (P')$ is greater than or equal to B?

It is not clear whether the k^{th} largest-area polygon problem is in NP.

The transformation is based on the the following idea. Consider a circle of large radius centered at the origin, and place a representation of each set along the circumference at intervals of $\frac{2\pi}{m}$ radians, starting from polar angle 0. In addition, place "enforcing" points along the radius of the circle at intervals of $\frac{2\pi}{m}$ radians, starting at polar angle $\frac{\pi}{m}$. More precisely, let $\delta = \max\{s(x) : x \in X_1 \cup \cdots \cup X_m\}$, and let the radius of the circle be r. Input element $x_{ji} \in X_i$, $s(x_{ji}) = \alpha$, is mapped to the point with polar coordinates $(r+\frac{\alpha r}{\delta c}, \frac{2\pi i}{m})$, so that the representation of set X_i is placed along the line segment with endpoints $(r, \frac{2\pi i}{m})$ and $(r+\frac{r}{c}, \frac{2\pi i}{m})$. The m enforcers are located at positions $(r, \frac{(2i+1)\pi}{m})$. We refer to the points $(r, \frac{\pi i}{m})$ as basepoints.

Call a subset of the transformed point set good if it contains exactly one point from the image of each input set as well as all of the enforcers; otherwise, the subset is called bad. Note that a bad subset may or may not determine a convex polygon--those that are not the extreme points of a convex polygon were carefully excluded in the definition of the problem. If we assume that c is large enough so that the enforcers are not contained in the convex hull of the points $(r+\frac{r}{c}, \frac{2i\pi}{m})$, then the convex polygons determined by good subsets can be put in a one-to-one correspondence with the m-tuples. In addition, the value of c is chosen so that the area of a convex polygon determined by a good set of points is strictly larger than the area of a convex polygon determined by a bad set of points.

Lemma 1: Let A_g be the smallest possible area of the convex hull of a good set of points, and let A_b be the largest possible area of the convex hull of a bad set of points, provided the bad set consists of the extreme points of its convex hull. Then for every r there is a sufficiently large c such that $A_g > A_b$.

Proof: The area of the smallest polygon determined by a good set of points, A_g, is strictly larger than the area of the convex hull of the basepoints, $A = 2mr^2 \sin\frac{\pi}{2m}\cos\frac{\pi}{2m}$. An overestimation of the largest possible convex polygon corresponding to a bad set of points is given by the convex hull of a set of $2m-1$ points at distance $r+\frac{r}{c}$ from the origin. This area is

$$2m(r+\frac{r}{c})^2 \sin\frac{\pi}{2m}\cos\frac{\pi}{2m} - (r+\frac{r}{c})^2 \sin\frac{\pi}{m}(1-\cos\frac{\pi}{m}),$$

which can be written as

$$A + \frac{2mr^2}{c}(2+\frac{1}{c})\sin\frac{\pi}{2m}\cos\frac{\pi}{2m} - r^2(1+\frac{1}{c})^2 \sin\frac{\pi}{m}(1-\cos\frac{\pi}{m}).$$

Algebraic manipulations show that this sum is strictly less than A whenever

$$c > \frac{\sqrt{m(m-1) + m\cos\frac{\pi}{m}} + m - 1 + \cos\frac{\pi}{m}}{1 - \cos\frac{\pi}{m}}. \square$$

As a result, the area of the convex polygon determined by a particular subset of points is greater than or equal to A if and only if the subset is good. We now show that the area of a good set of points accurately reflects the sum of the corresponding m-tuple.

We define a canonical wedge to be the triangle formed by enforcer $(r, \frac{(2i+1)\pi}{m})$, basepoint $(r, \frac{2i\pi}{m})$ and point $(r+\frac{\alpha r}{\delta c}, \frac{2i\pi}{m})$. The intent is to have the area of this wedge represent $\frac{\alpha}{2}$. That is, the convex hull of the basepoints and point $(r+\frac{\alpha r}{\delta c}, \frac{2i\pi}{m})$ should have an area which is a constant times α units larger than the area A.

The area of the canonical wedge is calculated as follows. The base of the canonical wedge is the line connecting the basepoints, and it has length $2r\sin\frac{\pi}{2m}$. The angle at basepoint $(r, \frac{2i\pi}{m})$ is $\frac{\pi}{2}+\frac{\pi}{m}$. As a result, the height of the canonical wedge is $\frac{\alpha r}{\delta c}\cos\frac{\pi}{m}$, so the area of the canonical wedge described above is

$$r\frac{\alpha r}{\delta c}\sin\frac{\pi}{2m}\cos\frac{\pi}{m}.$$

This observation and the lemma are used to derive the stated reduction.

Theorem 1: The problem of finding the polygon with k^{th} largest area is NP-hard.

Proof: For input sets $X_1,\ldots,X_m$, $m \geq 2$, use the mapping above to construct an instance of the k^{th} largest-area convex polygon problem. If c is large enough so that the enforcers are outside the convex hull of the points $(r+\frac{r}{c}, \frac{2i\pi}{m})$, then, by Lemma 1, one can increase c so that all convex polygons determined by bad point sets have area strictly less than A, the area of the convex hull of the basepoints. Furthermore, all convex polygons determined by good point sets have area larger than A, so each m-tuple can be matched with a convex polygon whose area is greater than A.

Suppose that for a given m-tuple, $(x_1, x_{2,...,} x_m)$, $s(x_j) = \alpha_j$, and $\sum_{i=1}^{m} \alpha_i = B$. Then the area of the corresponding polygon is $A + B'$, $B' = \sum_{i=1}^{m} 2r \frac{\alpha_i r}{\delta c} \sin\frac{\pi}{2m} \cos\frac{\pi}{m} = B \sum_{i=1}^{m} 2 \frac{r^2}{\delta c} \sin\frac{\pi}{2m} \cos\frac{\pi}{m}$. Consequently, if there are k m-tuples whose sum is B or larger, then there are exactly k convex polygons with area $A + B'$ or larger. On the other hand, a polygon whose area is equal to $A + B'$ corresponds to a subset whose sum is $\sum_{i=1}^{m} s(x_i) = B$, so if there are k polygons with area $A + B'$ or larger, then there are k m-tuples whose sum is B or larger. In conclusion, the test for whether k or more m-tuples have sum greater than or equal to B is resolved by testing whether or not there are k or more convex polygons in the transformed problem with area greater than or equal to $A + B'$. □

Remarks

The enumeration problem corresponding to the problem of finding the k^{th} largest m-tuple, "How many distinct m-tuples have sum greater than or equal to B?" is #P-complete. Our transformation to the k^{th} largest-area convex polygon problem is parsimonious [3] in the sense that the number of m-tuples with sum greater than or equal to B is precisely the number of convex polygons with area greater than or equal to $A + B'$. As a result, the enumeration problem, "How many convex polygons determined by subsets of P have area greater than or equal to B?" is #P-hard. Since it is easily seen to be in #P, the enumeration problem is #P-complete.

We also remark that it is possible to construct a pseudo-polynomial time algorithm for the k^{th} largest-area convex polygon problem. For a given set of points, algorithm SELECT has time complexity which is a polynomial in k and n, assuming that we can add and multiply real numbers at constant cost (see Figure 1). If we don't make this assumption, an appropriate adjustment to the time complexity can be made.

The general idea of this algorithm is to list k point sets whose convex hulls form k distinct convex polygons of area greater than or equal to B. If fewer than k point sets are listed, then the corresponding point set has fewer than k distinct subsets that determine convex polygons of area greater than or equal to B. The method of enumerating subsets starts with the convex hull of P; it is clear that the convex hull of the point set has the largest area. It is also clear that only candidates for the convex polygon with second largest area are the convex polygons determined by

SELECT (P, B, k)

```
0. k' <- 0

1. Construct conv (P)

2. If area (conv (P)) ≥ B then
        Add P to expand-queue
        Add P to checklist
        k' <- k' + 1

3. While expand-queue ≠ ∅ and k' < k
3a.     Dequeue front element P' from expand-queue
3b.     For each extreme point p in conv (P')
3ba.            Calculate conv (P'-p)
3bb.            If area (conv (P'-p)) ≥ B and
                P'-p is not in checklist then
                        Add P'-p to expand-queue
                        Add P'-p to checklist
                        k' <- k' + 1

4. If k' ≥ k then
        Return ("yes")
   Else
        Return ("no")
```

Figure 1 — Algorithm *SELECT*

removing an extreme point from the convex hull. There are two possibilities for the convex polygon with third largest area. One possibility is that it is one of the convex polygons determined by removing an extreme point from the convex hull of P, and the second possibility is that it is one of the convex polygons determined by removing an extreme point from the polygon with second largest area. Algorithm SELECT systematically implements this idea.

With each iteration, an extreme point is removed. If the convex hull of the remaining point set has area greater than or equal to B, it is added to the queue for later expansion. Otherwise, it is the case that the convex hull has area smaller than B, so no subset determines a convex polygon of area greater than or equal to B. At most k convex polygons are expanded, and the maximum length of the *expand−queue* is k. Inspecting the *checklist* to see whether a set of points was previously expanded takes $O(nk)$ time. One may find the convex hull of a set of points in $O(n \log n)$ time, and then calculate the area by triangulating in time $O(n)$. The overall time complexity of this procedure is therefore $O(kn^2 \log n + k^2 n^2)$, which is bounded by a polynomial in n and k. Good choices for data structures and a more careful enumeration of the points would decrease the time complexity.

References

1. B. Chazelle, New Techniques for Computing Order Statistics in Euclidean Space , *Proc. of the 1st Annual Symposium on Computational Geometry*, 1985, pp. 125-134 .
2. R. Cole, J. S. Salowe, W. L. Steiger and E. Szemeredi, Selecting Slopes , Submitted to Siam J. on Computing, 1987.
3. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, San Francisco, CA, 1979.
4. D. B. Johnson and T. Mizoguchi, Selecting the k^{th} Element in $X+Y$ and $X_1+X_2+\cdots+X_m$, *Siam J. on Computing*, 7 , 1978, pp. 147-153 .
5. N. Megiddo, Applying Parallel Computation Algorithms in the Design of Serial Algorithms , *J. ACM* , 30 (4), 1983, pp. 852-865 .
6. J. S. Salowe, *Selection Problems in Computational Geometry* , PhD Thesis, Rutgers University , 1987.