Computing the largest inscribed isothetic rectangle

Helmut Alt
Freie Universität Berlin

David Hsu
University of British Columbia

Jack Snoeyink*
University of British Columbia

Abstract

This paper describes an algorithm to compute, in $\Theta(\log n)$ time, a rectangle that is contained in a convex n-gon, has sides parallel to the coordinate axes, and has maximum area. With a slight modification it will compute the smallest perimeter. The algorithm uses a tentative prune-and-search approach, even though this problem does not appear to fit into the functional framework of Kirkpatrick and Snoeyink.

1 Introduction

In this paper, we give a logarithmic-time solution to the following problem: Given the list vertices of a convex polygon P in counterclockwise (ccw) order, stored in an array or balanced binary search tree, compute the rectangle $R \subset P$ with maximum area (or maximum perimeter) whose sides are parallel to the x and y coordinate axes.

Fischer and Höffgen [2] solved the maximum area problem by a nested binary search in $O(\log^2 n)$ time. To obtain a $\Theta(\log n)$ algorithm, we characterize the maximum rectangles, then use the tentative prune-and-search technique of Kirkpatrick and Snoeyink [4]. In some cases we are able to frame the search for a rectangle as a problem of computing a fixed-point and apply a theorem of [4]; in others we must use tentative prune-and-search directly.

In general, the *prune-and-search* technique for multiple lists looks at local information in $O(1)$ time to discard a fraction of some list. *Tentative prune-and-search* can sometimes be used when local information is insufficient to determine which fraction to discard. This technique makes *tentative* decisions that are later be *certified* or *revoked.*

Suppose that P is in general position: no two vertices on the same vertical or horizontal line and no two boundary edges parallel. (This can be simulated by perturbation methods if necessary [1].) We can decompose ∂P, the boundary of P, into four pieces by breaking at the horizontally and vertically extreme points. We name the pieces A, B, C, and D in counterclockwise order, starting from the southwest.

If a rectangle $R \subset P$ has only one corner on ∂P, or has only two corners from the same side of R on ∂P, then one can enlarge R by translating a side (figure 1a). Therefore a rectangle R with maximum area or perimeter either has two diagonally opposite corners, a and c, on ∂P, or has three corners a, b, and c, on ∂P. These two cases are illustrated in figure 1b and 1c, and characterized in the next two lemmas. We denote the slope of a line segment $\overline{ac}$ by m_{ac}.

Lemma 1 *Suppose that the maximum area or perimeter rectangle $R \subset P$ has exactly two corners on ∂P, namely $a \in A$ and $c \in C$. Then P has parallel tangents at a and c with slope m, where $m = 1$ in the perimeter case and $m = -m_{ac}$ in the area case.*

Proof: Fix a at the origin, and consider the curve where $c = (x, y)$ can lie for the rectangle with diagonal $\overline{ac}$ to have perimeter or area F. In the perimeter case $x + y = \pm F/2$ is a diamond with sides of slope 1 and -1. In the area case $xy = \pm F$ is a hyperbola.

*Supported in part by an NSERC Research Grant and a B.C. Advanced Systems Institute Fellowship.

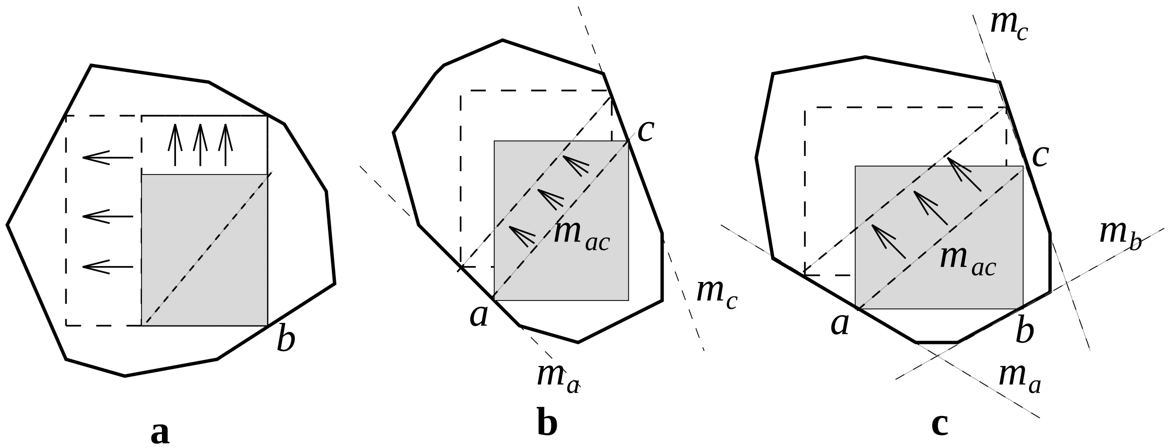

Figure 1: Cases for the maximum area or perimeter rectangle

For the maximum value of F, subject to c lying in P, the curve and P must have a parallel tangent at c; otherwise, R could be enlarged. In the case of maximum area, if $c = (x, y = F/x)$ then the tangent to $xy = F$ at c has slope $dF/dx = -F/x^2$, which is $-m_{ac}$.

Fixing c gives a parallel tangent at a. ■

For a maximum area or perimeter rectangle R with three corners on ∂P, we can assume, without loss of generality, that these corners are $a \in A$, $b \in B$, and $c \in C$.

Lemma 2 *Suppose the maximum rectangle $R \subset P$ has three corners on ∂P, as in figure 1c. Then there are tangents at $a \in A$, $b \in B$, and $c \in C$ with slopes $m_a < 0$, $m_b > 0$, $m_c < 0$, respectively, that satisfy $-m_a \geq m \geq -m_c > 0$, and $m = -m_a \frac{m_c - m_b}{m_a - m_b}$, where $m = 1$ if R has maximum perimeter and $m = m_{ac}$ if R has maximum area.*

Proof: To make the A, B, and C portions of ∂P into smooth functions of x, we construct, at each vertex v of P, a circle tangent to both edges incident to v. The radius of the tangent circle can be made infinitely small so that the area of the inscribed rectangle would not be changed by more than an infinitesimal amount. We cut this new boundary at its extreme points to form boundary chains A, B, and C on which a, b, and c lie, respectively.

Notice that slopes of tangents at a, b, and c satisfy $m_a, m_c < 0$ and $m_b > 0$. If a rectangle R with these corners has $-m_a < -m_c$, then one can enlarge R by translating $\overline{ac}$ in the direction that increases the length of $\overline{ac}$ while keeping m_{ac} fixed (figure 1c). Thus, for a maximum rectangle, $-m_a \geq -m_c$.

Now, let $b = (x, y)$ and take the derivatives of perimeter and area as functions of x.

$$\begin{aligned} \frac{d}{dx} perim &= \frac{d}{dx}(c_y - b_y) + \frac{d}{dx}(b_x - a_x). \\ \frac{d}{dx} area &= \frac{d}{dx}((b_x - a_x)(c_y - b_y)) \\ &= (b_x - a_x)\left(\frac{d}{dx}(c_y - b_y) + \frac{c_y - b_y}{b_x - a_x}\frac{d}{dx}(b_x - a_x)\right). \end{aligned}$$

If we set $k = m = 1$ in the perimeter case and $k = b_x - a_x$ and $m = m_{ac} = (c_y - b_y)/(b_x - a_x)$ in the area case, then we can write both derivatives in terms of slopes

$$\frac{dF}{dx} = k((m_c - m_b) + m(1 - \frac{m_b}{m_a}))$$

$$= \frac{k}{m_a}(m_a(m_c + m)) - m_b(m_a + m)) \quad (1)$$

Area and perimeter are both smooth functions of x, so the extreme values of F occur at extremes of x and when $dF/dx = 0$. The extreme cases, $b_x = a_x$ or $b_y = c_y$ cannot be solutions of the maximum area problem (these rectangles have zero area) and can be handled as special cases for the maximum perimeter problem. Thus, to see when the derivative of F is zero, we can assume that $k \neq 0$ and $m_a \neq 0$ in 1 and derive equivalent conditions

$$m_a(m + m_c) - m_b(m + m_a) = 0 \quad (2)$$

$$m = -m_a \tfrac{m_c - m_b}{m_a - m_b}. \quad (3)$$

Since m_a and $-m_b$ are both negative, the terms $m + m_c$ and $m + m_a$ must have opposite signs in a solution to equation 2 and m_{ac} lies between $-m_c$ and $-m_a$. If we move b counter-clockwise along B, then m_b and $b_x - a_x$ increase monotonically, and m_a, m_c, and m_{ac} decrease monotonically. Equation 1 then implies that dF/dx decreases monotonically, and therefore the zero of the derivative is a maximum. ■

With these characterizations of maximal rectangles there are now two diagonal rectangles and four three-corner rectangles to test. In the next two sections we show how to locate the rectangle corners on ∂P—it is then easy to check that the other corners are inside of P. We'll treat each test independently, but results of, say, the diagonal tests will indicate which three-corner tests are needed.

2 Computing a two-corner rectangle

In this section we look for chords of polygon P that could be diagonals of a maximum-area or maximum-perimeter inscribed rectangle—chords that satisfy lemma 1. By casting the search as a fixed-point problem, we show that there is at most one such chord for each diagonal, and we find it in $O(\log n)$ time.

Suppose that we consider the SW to NE diagonal, $\overline{ac}$. Let A and C be the boundary chains of P on which a and c can lie. (We can conceptually round the corners again so that every point has a unique tangent, and also bow the edges infinitesimally so that every tangent has a unique point of tangency. Then not only are A and C be continuous *trails* [3] where a point moving along a line alternates with a line rotating about a point, they are also differentiable.)

Lemma 3 *We can find the two-corner rectangle with maximum area or perimeter in* $O(\log n)$ *time.*

Proof: Let f map a point $a \in A$ to the point $f(a) \in C$ such that tangents to P at a and $f(a)$ are parallel. Let $g\colon C \to A$ map a point c with tangent to P of slope m to the point $a \in A$ that is intersected by the line through c of slope $-m$. Then a fixed-point of the composition, $a = g(f(a))$, gives a chord $\overline{af(a)}$ that satisfies lemma 1.

If we parameterize A and C counterclockwise, then f is monotone increasing and g is monotone decreasing. Their composition has a unique fixed-point.

Given an $a \in A$ and a $c \in C$, one can compare $f(a)$ against c and $g(c)$ against a in constant time by comparing tangents and slopes. If A and C are both reduced to single segments or vertices, then the computation can be completed by simple algebra. Therefore, by a lemma of Kirkpatrick and Snoeyink [4], we can compute the fixed-point in $O(\log n)$ steps. It is then a simple matter to check if the rectangle R with this diagonal is indeed inscribed in P. ■

3 Finding a three-corner rectangle

Suppose we want to look for a maximum inscribed rectangle R that has three corners on the boundary of P: $a \in A$, $b \in B$ and $c \in C$. In order to fit in the fixed-point framework, we would need three functions whose composition gives a fixed-point. Two functions are natural to define: Let $f(a) \in B$ have the same y coordinate as a and let $g(b) \in C$ have the same x coordinate as b. The third function, however, involves slopes at all three corners: Let $h(a, c)$ be the point in B with tangent slope determined by lemma 2. If R' is maximum, $f(a') = b'$, $g(b') = c'$ and $h(a', c') = b'$ for corners $a' \in A$, $b' \in B$ and $c' \in C$, by lemma 2. Furthermore, the conditions are satisfied only if R' is maximum. To see this, notice that dF/dx decreases monotonically, as proved in lemma 2, and therefore attains zero at most once. It follows that there exist unique points a', b' and c' that satisfy the given conditions.

We can still use the technique of tentative prune-and-search to find a three-corner rectangle with maximum area or perimeter.

Lemma 4 *A maximum three-corner rectangle satisfying lemma 2 can be found in* $O(\log n)$ *time.*

Proof: Let R' be the maximum rectangle that we seek and let $a' \in A$, $b' \in B$ and $c' \in C$ be the corners of R' on the boundary of P. We search for these corners by looking at the positions of (and tangents at) "middle" vertices in A, B, and C, then using constant-time local computations to discard (or tentatively discard) half of some list that cannot contain one of the corners of R'.

Let $a \in A$, $b \in B$, and $c \in C$ be given. Define boolean variables $a\triangleright$ (after-a) and $\triangleleft a$ (before-a): $a\triangleright$ is true iff a' is at or after a in counter-clockwise order around P. Figure 2 lists conclusions from all tests using this notation.

From function f we can test if $a.y \geq b.y$. If so, as illustrated in figure 3, either a' must be below the line $y = a.y$ or b' must be above $y = b.y$, so we can conclude $f_T = (a\triangleright \vee b\triangleright)$. If $a.y \leq b.y$ then $f_F = (\triangleleft a \vee \triangleleft b)$. Function g gives a similar test: $b.x \leq c.x$ implies $g_T = (b\triangleright \vee c\triangleright)$ and $b.x \geq c.x$ implies $g_F = (\triangleleft b \vee \triangleleft c)$.

For function h, consider violations of the three conditions $-m_a \geq m \geq -m_c$ of lemma 2. First, if $-m_a < m$ then m must decrease (impossible for the perimeter problem where $m = 1$, but possible for area where $m = m_{ac}$) or $-m_a$ must increase. Thus, $h_1 = (\triangleleft a \vee \triangleleft c)$. Second, if $m < -m_c$ then $h_0 = (a\triangleright \vee c\triangleright)$. Third, if $-m_a < -m_c$, then $(\triangleleft a \vee c\triangleright)$.

When this third constraint ($-m_a \geq -m_c$) is violated, then necessarily either the first or second is also violated. If the first, then we can conclude that $\triangleleft a$. This means that a' is clockwise of a on A, so we can discard the counter-clockwise portion of A. If the second, we conclude $c\triangleright$. (Notice that both violations can occur, allowing us to eliminate half of each of A and C.)

Assume for the moment that these three constraints are satisfied, but the main constraint of lemma 2 is violated. Suppose that "equality" is replaced by "less than" in equation 3. We can rewrite to obtain

$$-(m + m_c) + m_b \left(1 - \frac{m}{-m_a}\right) < 0.$$

If we move a and c counter-clockwise, then $m = m_{ac}$ increases and $-m_a$ and $-m_c$ stay the same or decrease. Because $-m_c \leq m$, we know $(m + m_c)$ is non-negative and increasing. The term $1 - m/(-m_a)$ is non-negative and decreasing. Thus, to restore equality in 3, we must have m_b increase by b moving counter-clockwise. We can conclude $h_T = (\triangleleft a \vee b\triangleright \vee \triangleleft c)$. If "equality" in equation 3 becomes "greater than," then we conclude $h_F = (a\triangleright \vee \triangleleft b \vee c\triangleright)$.

Figure 2 displays the tests associated with the functions f, g, and h and their conclusions.

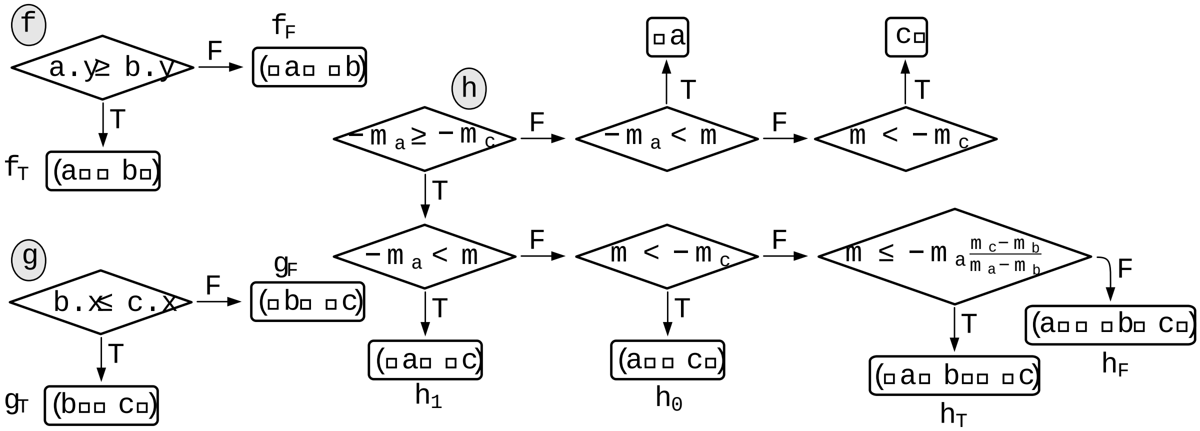

Figure 2: Tests associated with functions f, g and h. For area $m = m_{ac}$; for perimeter $m = 1$

By combining tests we can derive further conclusions. For example, $f_T = (a\triangleright \vee b\triangleright)$ and $h_1 = (\triangleleft a \vee \triangleleft c)$ imply $(b\triangleright \vee \triangleleft c)$. If we also have $g_T = (b\triangleright \vee c\triangleright)$, then we know $(b\triangleright)$. Table 1 lists the conclusions from all test combinations. In six cases, our constant-time tests are sufficient to discard half of one of the boundary chains A, B, or C. In the remaining two cases, we can use Kirkpatrick and Snoeyink's technique [4] of making *tentative* discards with the assurance that we are making at most one mistake. Notice that these two cases are complements—all test outcomes are opposite.

Values			Conclusion
f_T	g_T	h_1 or h_T	$(b\triangleright)$
f_F	g_T	h_1 or h_T	$(\triangleleft a)$
f_F	g_F	h_1 or h_T	$(\triangleleft a \wedge \triangleleft b) \vee (\triangleleft b \wedge \triangleleft c) \vee (\triangleleft a \wedge \triangleleft c)$
f_T	g_F	h_1 or h_T	$(\triangleleft c)$
f_T	g_F	h_0 or h_F	$(a\triangleright)$
f_F	g_F	h_0 or h_F	$(\triangleleft b)$
f_F	g_T	h_0 or h_F	$(c\triangleright)$
f_T	g_T	h_0 or h_F	$(a\triangleright \wedge b\triangleright) \vee (b\triangleright \wedge c\triangleright) \vee (a\triangleright \wedge c\triangleright)$

Table 1: Conclusions from combining tests.

Suppose that we have f_T, g_T, and h_F. We can *tentatively* discard boundary chains clockwise of a, b, and c as shown in figure 3. We can then refine the remaining portions in a round-robin fashion: choosing a middle vertex from what remains of A and re-evaluating the tests associated with functions f and h, then of C and re-evaluating g and h, then of B and re-evaluating f, g, and h.

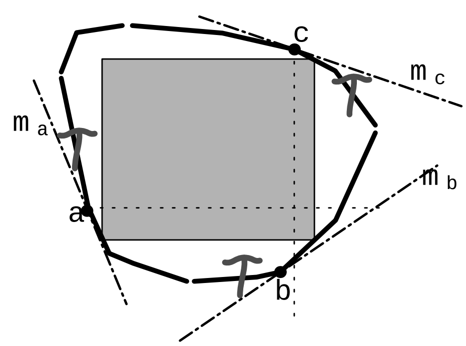

Figure 3: Tentative discards caused by f_T, g_T, and h_F

Refining and re-evaluating A (or C) can change the outcomes of the two associated tests. One can check in table 1 that from the new outcomes we can derive one of three conclusions: (1) we remain in the tentative case (i.e., no test outcomes change) and extend the tentative discard on A (or C), (2) we permanently discard half of what remains of A (or C), or (3) we certify that all previous tentative discards done to one chain were correct.

Refining and re-evaluating B can potentially change all three test outcomes. Thus, two new conclusions could apply, in addition to 1–3 above: (4) the remaining portion of A or C is discarded, and we certify that the tentative discards on that chain were in error and that the tentative discards on other chains were correct, or (5) all three test outcomes change and we are in the opposite tentative case. In this last case, however, we have tentatively discarded all of A and all of C, clearly making one mistake on each chain. Therefore, all tentative discards done on B can be certified as correct and those on A and C revoked.

To evaluate the total cost of the algorithm, we can use the potential function of Kirkpatrick and Snoeyink [4]. For chain A, let A_T denote the number of segments tentatively discarded and A_R denote the number remaining. Define the chain potential $\Phi_A = 2\log A_R + 4\log(A_R + A_T)$. Using Iverson's notation, where a boolean evaluates to 1 or 0 depending on whether it is true or false respectively, the global potential is the sum of chain potentials plus 5 in tentative mode:

$$\Phi = \Phi_A + \Phi_B + \Phi_C + 5(A_T + B_T + C_T > 0).$$

Notice that the initial potential is $O(\log n)$ and that Φ cannot be negative. We can conclude our theorem by showing that Φ decreases by a constant at each step.

When there are no tentative discards, we discard half of some chain and decrease Φ by 6 or tentatively discard from each chain and decrease Φ by 1. When portions have been tentatively discarded, we either extend the discard on the refined chain (a permanent discard can be considered tentative for the analysis) or else certify all tentative discards made to one chain and revoke the rest. In the former case Φ decreases by 2. In the latter case, suppose that the certified chain participated in t tentative steps, including the first. Then the potential of the certified list decreases by $4t$. The other two chains each gained 2 for every tentative step that they participated in—since the chains were considered in round robin order, this is at most $t+1$ steps each. Since we leave tentative mode, Φ's total decrease is at least $4t - 4(t+1) + 5 = 1$. ∎

4 Conclusion

We have shown how to compute, in $\Theta(\log n)$ time, the maximum area or perimeter rectangle that has sides parallel to the coordinate axes and is inscribed in a convex n-gon. Our algorithm used a tentative prune-and-search approach, even though this problem did not fit into the fixed-point framework of Kirkpatrick and Snoeyink [4]. We applied constant-time tests to discover boolean predicates on the locations of corners of the maximum rectangle. Sometimes predicates combined to eliminate half of a chain that contained a corner. If we were not so lucky, we could still tentatively eliminate portions of chains in a round robin fashion, maintaining predicates that implied that we were doing the right thing on at least one of the chains. The analysis of running time is by a potential argument. It would be interesting to find natural classes of problems that can be solved by maintaining predicates.

References

[1] H. Edelsbrunner and E. P. Mücke. Simulation of simplicity: A technique to cope with degenerate cases in geometric algorithms. *ACM Trans. Graph.*, 9(1):66–104, 1990.

[2] P. Fischer and K.-U. Höffgen. Computing a maximum axis-aligned rectangle in a convex polygon. *Info. Proc. Let.*, 51:189–194, 1994.

[3] L. Guibas, L. Ramshaw, and J. Stolfi. A kinetic framework for computational geometry. In *Proc. 24th FOCS*, pages 100–111, 1983.

[4] D. Kirkpatrick and J. Snoeyink. Tentative prune-and-search for computing fixed-points with applications to geometric computation. *Fund. Infor.*, 1994.