Bellman's Escape Problem for Convex Polygons

Philip Gibbs philegibbs@gmail.com

Abstract: Bellman's challenge to find the shortest path to escape from a forest of known shape is notoriously difficult. Apart from a few of the simplest cases, there are not even many conjectures for likely solutions let alone proofs. In this work it is shown that when the forest is a convex polygon then at least one shortest escape path is a piecewise curve made from segments taking the form of either straight lines or circular arcs. The circular arcs are formed from the envelope of three sides of the polygon touching the escape path at three points. It is hoped that in future work these results could lead to a practical computational algorithm for finding the shortest escape path for any convex polygon.

Introduction

Sixty years ago Richard Bellman proposed the "Lost in a Forest" problem of finding the shortest path that ensures escape from a forest of known shape when the starting position and direction is unknown [1]. In geometric terms: Given a shape in the plane, what is the shortest continuous open curve which cannot be placed inside the shape using rotations and translations without intersecting the shapes boundary? An equivalent problem is to find the smallest shape similar to a given shape which is a universal cover (allowing rotations and translations but not reflections) for all continuous open curves of length one.

The latter formulation makes clear the relationship with Moser's famous "Worm Problem" which seeks to find the smallest convex area that is a universal cover for all open curves of length one [2]. Many of the shapes for which solutions to Bellman's problem are known come from known universal covers for Moser's problem. For example, a universal cover in the form of a rhombus can be embedded inside regular polygons with four or more sides to show that the optimal escape route for these shapes is simply the longest straight line path that fits inside. Triangles are a source of more interesting solutions.

Fortunately there are two very good reviews of Bellman's problem describing the details of prior work: Finch and Wetzel 2004 [3] and Ward 2008 [4]. Given the sparsity of solutions to the problem after 50 years who can blame Ward for suggesting that "a general solution appears elusive and likely unapproachable" However, in 2006 Coulton and Movshovich produced a breakthrough which indicates that there may be some hope [5]. They solved the problem for a range of isosceles triangles including the equilateral triangle for which Besicovitch forty years earlier had correctly conjectured a zig-zag shaped solution. The only other known solution which is not a simple straight line is a caliper shaped curve for shortest escape path from an infinite strip, first solved by Zalgaller in 1961. This may not seem like much of a repertoire but if these cases can be resolved why not others?

Escape Paths for Convex Polygons

Before turning to the specific cases it is useful to consider the general forms a shortest escape path can take for a polygonal shaped forest. For simplicity and without losing too much generality, consider a forest F as a region of the plane bounded by and including a convex polygon of which no two sides are parallel.

Let $\mathcal{C}$ be the set of all open continuous rectifiable plane curves of non-zero length. Let C be a curve in $\mathcal{C}$, and let $|C|$ denotes its length. C can be rotated through an angle θ without change of shape. For each θ there is a unique largest positive scale factor $s(F, C, \theta)$ such that the rotated curve scaled by this factor can be translated to a position where it is covered by F. Since F is convex with no parallel sides the position will be unique. Furthermore it can be shown that $s(F, C, \theta)$ is a continuous periodic function of θ and is therefore bounded and has a maximum value $s_{max}(F, C)$.

When the curve C is scaled by the factor $s_{max}(F, C)$ it forms a Shortest Escape Path (SEP) of length $s_{max}(F, C) \times |C|$. For any given orientation and any starting point in F the scaled curve must then intersect the boundary of F at some point. Furthermore any SEP can be formed in this way. There may be more than one SEP for a given F that are distinct under rotations and scalings, but they all have the same length.

Therefore Bellman's problem reduces to finding the curves C which minimises the function $|C| s_{max}(F, C)$. In other words the length of the shortest escape paths for F is given in terms of the scaling function by

$$L_{SEP}(F) = \min_{C \in \mathcal{C}} \left(|C| \max_{0 < \theta \leq 2\pi} s(F, C, \theta) \right)$$

To maximise the scaling factor $S(F,C,\theta)$ for a fixed curve C and a fixed angle θ the curve is translated and expanded giving three degrees of freedom. When it touches one of the edges of F, that will impose a constraint with one degree of freedom. This tells us that for the maximum case it will touch at least three sides of F. In some cases it may touch two adjacent sides at the vertex where they meet so that the path only touches the boundary of F at two points. Because we have specified that the boundary is a convex polygon with no parallel edges the position of the curve at maximum expansion will be uniquely determined.

It can also be seen that the edges of F touched by C can be extended to form a polygon that encloses F. Normally this polygon will be a triangle.

If a curve C is chosen at random it is reasonable to expect that the scaling function $s(F,C,\theta)$ would typically have a single maximum at a unique θ angle. However, when C is also varied to find the curve which minimises this maximum, it is possible that the optimum path will be a case where there is a switch between different local maxima so it is important to consider cases where there are multiple maxima. Indeed, the minimisation is over an infinite space of variables describing the curve C as a function so it may not be unreasonable to expect that the maxima will occur along continuous ranges of angles. We already know this to arise in the case of Zalgallar's escape path for an infinite strip.

In the general case where the shape of the curve C is unrestricted it will have a maximal scaling factor for fixed angle θ at a unique position where it touches at least three edges of the boundary of F. Define a subset T of C consisting of all points that touch the boundary of F at some angle where the scaling factor is maximal. T could consist of as few as just two discrete points of C but if C is an optimal escape path T may include several discrete points and even continuous segments of C. It will however by a closed set in the topological sense. What can be said about the sections of C that join up the points of T? The first step in demonstrating the claim that the shortest escape paths are composed of straight lines and circular arcs is to show that:

If C is an shortest escape path then sections of C that are not in T must be straight line segments. Furthermore the end points of C must be in T.

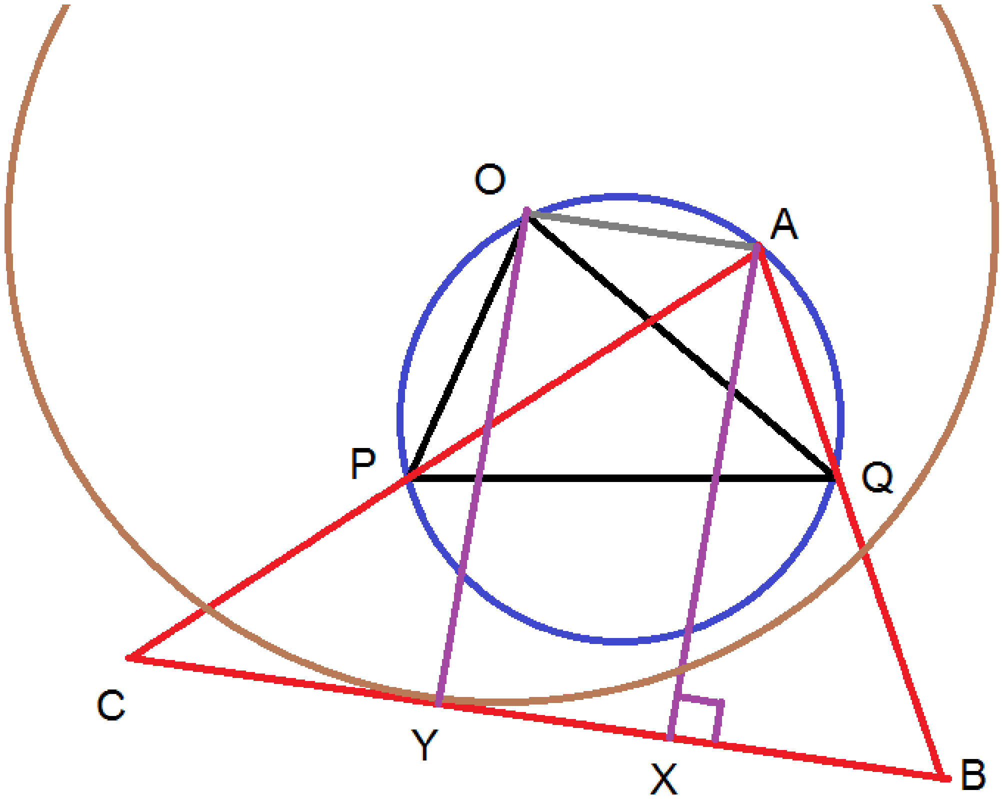
This can be proven by contradiction. If a segment of C that does not include points of T is not a straight line then it can be replaced by a straight line to give a shorter path, or if it is at one end of the curve it can be removed altogether. However, this change could also modify the scaling factor function $s(F, C, \theta)$ making its maximum bigger to counteract the shortening of the curve. This can only happen if there is an angle θ where a point X on the segment is touching the boundary of F. Since X is not in T the scaling factor for this angle must be smaller than the maximum. It then suffices to show that a small part of the segment near X can be replaced with a straight line segment without changing the scaling factor to make it greater than the maximum at any angle. This follows from the continuity of the scaling factor as a function of C.

This means that if T included only discrete points on a shortest escape path then the optimal curve would be a polyline. What can be said about the shape of continuous sections of the curve which are in T?

As the angle θ covers the range $0 < \theta \leq 2\pi$ a curve C expanded by the scaling factor $s(F, C, \theta)$ will touch three sides of F except at discrete points where it may touch 4 or more sides. The range can be subdivided into intervals $\theta_i < \theta \leq \theta_{i-1}$ over which the curve touches the same three sides and the points in T where is touches move continuously in one direction of remain at the same point. This subdivision of the angular ranges also induces the curve C to be subdivided into segments which are either straight lines joining points of T or continuous sections within T which touch the same straight edge of F over one of the angular intervals. The remaining task is to establish the possible form of curve that one of these segments can take.

To visualise the situation it is better to take the expanded curve C as static and allow the polygon F to rotate through an angle $-\theta$ for $\theta_i < \theta \leq \theta_{i-1}$ while being translated by a variable vector so as to touch the curve. Over the whole segment the scaling factor of the curve is fixed at its maximum value so the size of F and C will be fixed. Since only the same three aides of F touch the curve it is sufficient to look at just those three sides and extend them to form a triangle enclosing the curve. The triangle will touch C along three subsets of T. One or two of these could be single points but in the most general case they are continuous segments and are formed by the envelope of the edges of the triangle as it moves and rotates.

To understand how these envelopes form, first consider the special case where two of three segments of T on the curve collapse to points. Let the triangle formed by extending three sides of the forest boundary be ΔABC. The two points on the curve are P and Q with Q touching the side AB and P touching the side AC. The third segment of the curve is then the envelope of the side BC as the triangle rotates under these constraints. It can be demonstrated that the solution is a circular arc by the following construction.



Define a new point O to construct a triangle ΔOPQ on the given side PQ so that it is similar to the reflection of the moving triangle ΔABC. I.e. $\angle OPQ = \angle ABC$ and $\angle OQP = \angle ACB$. Draw the circumscribed circle of triangle ΔOPQ. Because $\angle POQ = \angle CAB = \angle PAQ$ it follows that A must also lie on this circle. Also, $\angle OAP = \angle OQP = \angle ACB$, therefore OA is parallel to CB. Therefore the distance OY from the point O to the side BC is equal to the distance h_A from the point A to the side BC and this is the fixed height of the triangle. Therefore the side BC is tangent to the circle with centre O and radius $r_A = h_A$ and the envelope of the edge must be an arc of that circle.

Although this is just a specific special case it does appear to be one that arises on the shortest path for some triangular shaped forests (future work.) It also generalises easily to the case where the triangle touch circular arcs at all three points. To see this draw a circle of radius r_C centred on Q and a circle of radius r_B centred on P. The radii must be sufficiently small that the two circles can fit simultaneously inside the triangle ΔABC so that the side AB touches the circle tangentially and the side AC touches the second circle tangentially. What then is the envelope of the side BC?

To find the answer, first reduce the triangle ΔABC to a smaller similar triangle $\Delta A'B'C'$ by drawing the edge $A'B'$ parallel to AB but passing through , the side $A'C'$ parallel to AC but passing through P , and the side $B'C'$ lying on the side BC. The triangle $\Delta A'B'C'$ is then similar to ΔABC with linear dimensions reduced by a factor $f = 1 - \frac{r_B}{h_B} - \frac{r_C}{h_C}$. Where $h_B$is the height of the original triangle with AC as the base and $h_B$is the height with AB as the base. The earlier result can now be applied to the motion of the triangle $\Delta A'B'C'$ to show that the envelope of the side BC is also a circular arc with radius $r_A = fh_A$. The equation relating the three radii can also be written in the more symmetrical form $\frac{r_A}{h_A} + \frac{r_B}{h_B} + \frac{r_C}{h_C} = 1$

In general the envelopes formed from the sides of a triangle as it moves need not be arcs. In general the equation above is the relation between the radii of curvature of the three segments of curve at three points touched simultaneously by the triangle. The three segments defined as the points on the curve touched by the boundary of F as it turns through an angle $-\theta$ for $\theta_i < \theta \leq \theta_{i-1}$. Fix the six end points of the segments and consider the set of possible arcs that can join them. In fact we have to include possible straight line segments at each end of each segment at fixed angles. The curve must be convex along these segments so the longest the curve can be is when the two straight lines meet at a point touched by the triangle over the full range of angles. This corresponds to a radius of curvature of zero. Let L_0 be the length of the full curve if all three segments have a zero radius of curvature.

Assume that the arcs have a constant curvatures. If the curvatures are non-zero then the length of the curve will be reduced according to the formula $L = L_0 - (r_A + r_B + r_C)(2\ \tan\frac{\Delta\theta}{2} - \Delta\theta)$. The factor multiplying the sum of the curvatures is positive so to minimise the path length we need to maximise $(r_A + r_B + r_C)$ under the constraint $\frac{r_A}{h_A} + \frac{r_B}{h_B} + \frac{r_C}{h_C} = 1$

Each radius of curvature must be non-negative. From the constraint given above, this means that each radius of curvature must be less than the corresponding altitude of the triangle. There can be stricter upper constraints from the positioning the fixed end-points of the arcs. In the most extreme case when two end points coincide the corresponding radius of curvature would have to be zero. In general we get $0 \le r_A \le R_A \le h_A$ and similarly for the other two cases.

Suppose that $h_A > h_B > h_C$. Then to maximise the sum of curvatures we need to take $r_A = R_A,\ r_B = \min\left(R_B,\ h_B\left(1 - \frac{R_A}{h_A}\right)\right)$. The third radius can then be computed from the constraint.

In general the curvature could vary with angle of rotation but must remain within the same limits. A full analysis of the general case should be given but it is not hard to guess that the shortest path is given when the curvature is constant and given by the same formulae.

However, we neglected to consider the special case where two or three of the triangle's altitudes are equal. In this case the corresponding curvatures can be varied with no change in the length of the curve. This means that unless the constraints force one or two of the curvatures to be zero, there will be an infinite set of shortest escape paths with the same minimum length. The curves would then not necessarily have to be composed of circular arcs, but there will always be cases where they are. Note that it is not currently known if this situation arises in actual solutions of Bellman's problem, but in a variant of the problem where the paths are closed (i.e. the end of the path must return to the start point) this does seem to be the situation for the equilateral triangle shaped forest.

In conclusion, we have sketched the outline of a proof that for the solutions of Bellman's problem for any forest bounded by a convex polygon, there is at least one shortest escape path which is constructed piecewise from straight lines and circular arcs.

Future Work

The results in this paper are just a small step towards resolving Bellman's "Lost in a Forest" problem. There is still a long way to go but this progress suggests that rather than being unapproachable there is in fact a possibility that it could be resolved for convex polygons. Knowing that there is a shortest escape path composed of straight lines and circular arcs comes close to reducing the problem to a combinatorial one. It is plausible that an algorithmic solution could be found that could be implemented in software. If this is too ambitious then the case of triangular forests may be more tractable.

Any solution to Bellman's problem for a given shape of forest also sets an upper bound for the more famous Moser's worm problem. Current upper bounds are impressive accomplishments, but it is likely that they can be bettered even using just triangular covers. If a proven algorithmic solution is found for Bellman's problem it will therefore be a big step forward for Moser's problem too.

References

[1] R. Bellman, "Minimization problem", Bull. Amer. Math. Soc. 62 (1956) 270.

[2] L. Moser, "Poorly formulated unsolved problems in combinatorial geometry", mimeographed (1966)

[3] S.R. Finch, J. E. Wetzel, "Lost in a Forest", Amer. Math. Monthly 111 (2004), 645–654

[4] J. W. Ward, "Exploring the Bellman Forest Problem" 2008

[5] P. Coulton, Y. Movshovich, "Besicovitch triangles cover unit Arcs". Geom. Dedicata 123 79–88 (2006)