A Direct Approach to the Ladder Problem Using Envelopes

Dan Kalman
American University
Washington, D.C. 20016
kalman@american.edu

How long a ladder can you carry horizontally around a corner? This familiar problem can be found in the max/min sections of many calculus texts, as well as on numerous web sites. Sometimes the question concerns a ladder, sometimes a pipe, and sometimes a pole vaulter's pole, but the underlying problem is always the same. Two hallways of specified widths meet at a right angle corner, as illustrated in Figure 1. In the idealized geometry of the figure, the hallways form an L-shaped region bounded by the x and y axes, and the lines $x = a$ and $y = b$. We wish to know the longest line segment that can be maneuvered around the corner from one hall to the other.

Place figure about here.

Figure 1: Geometry of ladder problem.

Interestingly, the standard solution of this problem begins with an inversion. Rather than looking for a maximum (the longest line that goes through) we seek a minimum (the shortest line that gets stuck). That is, we look for the shortest line segment that touches both the outer walls (aka the positive x and y axes) and passes through the corner (a, b), observing that any shorter segment will turn the corner without getting stuck.

This is a clever idea, and it leads to a straightforward solution. Indeed, it illustrates an important principle in problem solving: when a direct approach fails, look for alternative formulations. But the ladder problem can also be solved by a beautiful direct approach, using envelopes of curves. This approach immediately gives a new insight about the problem. It also gives us an excuse to revisit a lovely topic that was once a standard part of the calculus curriculum, but which seems to be largely forgotten in the current generation of texts. A generalized ladder problem considered in [9] can also be analyzed using the direct approach.

The Direct Approach

Here is one way to try moving a segment around the corner. Begin with the segment along one of the outer walls, say with the left end at the origin and the right end at the point $(L, 0)$. Slide the left end up the y axis, all the while keeping the right end on the x axis. This is a conservative approach to the problem, in the sense that it keeps the line segment as far as possible from the corner point (a, b). Surely, if a segment of length L cannot get around the corner using this conservative approach, then it won't go around no matter what we do.

Now as you slide the segment along the walls, it sweeps out a region Ω, as illustrated in Figure 2. The outer boundary of Ω is part of curve called an *astroid*, with equation

$$x^{2/3} + y^{2/3} = L^{2/3}. \tag{1}$$

The full astroid is in the shape of a four pointed star (hence the name), but for the ladder problem, we are concerned only with the part of the curve that lies in the first quadrant. Our line segment will successfully turn the corner just in case Ω stays within the hallways. And that is true as long as the corner point (a, b) is outside Ω. The extreme case occurs when (a, b) lies on the boundary curve, whereupon

$$a^{2/3} + b^{2/3} = L^{2/3}.$$

This shows that the longest segment that can go around the corner has length $L = (a^{2/3} + b^{2/3})^{3/2}$.

Of course, this solution depends on knowing the boundary curve for Ω. Once you know that, the problem becomes transparent. We can easily visualize the region for a

Place figure about here.

Figure 2: Swept out region Ω.

short segment that will go around the corner, and just as easily see what happens if we increase the length of the segment. In contrast with the usual approach to this problem, we are led to a direct understanding of the maximization process. And in the context of the equation for the astroid, we understand *why* the formula for the extreme value of L takes the form that it does.

In fact, in this direct approach, the optimization part of the problem becomes trivial. It is akin to asking "What is the longest segment that can be contained within the unit interval?" This is nominally a max/min problem, but no analysis is needed to solve it. In the same way, the direct approach to the ladder problem renders the solution immediately transparent, once you have found the boundary curve for Ω. But it is not quite fair to claim that this approach eliminates the need for calculus. Rather, the point of application of the calculus is shifted from the optimization question to that of finding the boundary curve.

Envelopes of Families of Curves

So how is the boundary curve found? The answer involves the concept of the *envelope* of a family of curves. For the current case, observe that each successive position of the line segment can be identified with a linear equation. Let the angle between the segment and the positive x axis be α, as shown in Figure 3. Then the x and y intercepts of the line segment are $L\cos\alpha$ and $L\sin\alpha$, respectively, so the line is defined by the equation

$$\frac{x}{\cos\alpha} + \frac{y}{\sin\alpha} = L \tag{2}$$

This equation defines a family of lines in terms of the parameter α.

Place figure about here.

Figure 3: Parameter α.

The region Ω is the union of all the lines in the family. To be more precise, we restrict α to the interval $[0, \pi/2]$, and intersect all the lines with the first quadrant. Visually, it seems apparent that the boundary curve is tangent at each point to one of the lines. This observation, which will be proved presently, shows that the boundary curve is an envelope for the family of lines.

In general, an equation of the form

$$F(x, y, \alpha) = 0 \tag{3}$$

defines a family of plane curves with parameter α if for each value of α the equation defines a plane curve in x and y. An *envelope* for such a family is a curve every point of which is a point of tangency with one of the curves in the family.

There is a standard method for determining the envelope curve: Differentiate (3) with respect to α, and then use the original equation to eliminate the parameter. Technically, (x, y) is a point of the envelope curve only if it satisfies both (3) and

$$\frac{\partial}{\partial \alpha} F(x, y, \alpha) = 0 \tag{4}$$

for some α. We will refer to this procedure as the *envelope algorithm.*

Obviously, the envelope algorithm depends on certain assumptions about F, requiring at the very least differentiability with respect to α. Also, in the general case, the condition is necessary but not sufficient, so there may exist curves which satisfy (3) and (4), but which are not part of the envelope. For the moment, let us gloss over these issues, and move straight on to applying the algorithm for the ladder problem. A more careful discussion of the technicalities will follow.

The first step is to differentiate (2) with respect to α. That gives

$$\frac{x \sin\alpha}{\cos^2\alpha} - \frac{y \cos\alpha}{\sin^2\alpha} = 0$$

and after rearrangement we obtain

$$x \sin^3\alpha = y \cos^3\alpha. \tag{5}$$

By combining this equation with (2), we wish to eliminate the parameter α. With that in mind, rewrite (5) in the form

$$\tan\alpha = \frac{y^{1/3}}{x^{1/3}}.$$

This leads to

$$\cos\alpha = \frac{x^{1/3}}{\sqrt{x^{2/3} + y^{2/3}}} \quad \text{and} \quad \sin\alpha = \frac{y^{1/3}}{\sqrt{x^{2/3} + y^{2/3}}}.$$

Now we can substitute these expressions in (2), and so derive the following equation in x and y alone.

$$x^{2/3}\sqrt{x^{2/3} + y^{2/3}} + y^{2/3}\sqrt{x^{2/3} + y^{2/3}} = L$$

Simplifying, we have

$$(x^{2/3} + y^{2/3})\sqrt{x^{2/3} + y^{2/3}} = (x^{2/3} + y^{2/3})^{3/2} = L$$

and so we arrive at (1).

We can also derive a parameterization of the envelope. In the equations above for $\cos\alpha$ and $\sin\alpha$, replace $\sqrt{x^{2/3} + y^{2/3}}$ with $L^{1/3}$. Solving for x and y produces

$$\begin{aligned} x(\alpha) &= L\cos^3\alpha \\ y(\alpha) &= L\sin^3\alpha \end{aligned} \tag{6}$$

In this parameterization, $(x(\alpha), y(\alpha))$ is the point of the envelope that lies on the line corresponding to parameter value α.

History and Pedagogy

The foregoing computation is an intriguing way to deduce the boundary curve for the region Ω. From that curve we can immediately find the solution to the ladder problem as discussed earlier. As elegant as this solution is, it may be inaccessible to today's calculus students. Interestingly, there is some evidence to suggest that the computation of envelopes via the method above was once a standard topic in calculus. This is certainly the impression left by [3, 4, 6], all of which date to the 1940's and 1950's. On the other hand, anecdotal reports by colleagues who were students and teachers of calculus during that time are inconsistent on this point.

In today's calculus texts (or more precisely, in their indices), one finds no mention of envelopes. The topic is covered in older treatments of calculus [2, 7] and advanced calculus [15] and the expositions in these sources tend to be very similar. Was the topic of envelopes common enough in the calculus curriculum in the first half of the twentieth century to be considered standard? If so, when and why did this topic fall out of favor? These are interesting historical questions.

If the topic of envelopes has been forgotten in calculus texts, it has not disappeared from the mathematical literature. Indeed, in expository publications like this Magazine, one readily finds recent mention of envelopes and the envelope algorithm. See, for example, [5, 8, 11, 13, 14]. Nevertheless, I have a feeling that this topic is not as widely known among college mathematics faculty as it should be. Accordingly, a rather detailed discussion of envelopes is presented in the next section.

Outside of calculus courses, where might envelopes by found? The topic appears in works on properties of plane curves (see [10, 16]), another subject that seems to have been much more common in an earlier era. To a previous generation of mathematicians who were well acquainted with such terms as *involute*, *evolute*, and *caustic*, the boundary curve (1) would be familiar indeed. It is known not only as an astroid, but more generally as an instance of hypocychloid, the locus of a point on a circle rolling within a larger circle. We obtained it as the envelope of the family of lines (2), identified in [16] as the *Trammel of Archimedes*. The same curve can also be obtained as the envelope of a family of ellipses, the sum of whose axes is equal to L. ([16, p. 2]). See Figure 4.

Place figure about here.

Figure 4: The astroid as envelope of a family of ellipses.

The treatment of envelopes in [16] implies that this topic is properly a part of the study of differential equations. Perhaps it is in this context that envelopes once were considered a standard calculus topic, although that is certainly not the case in [2, 7, 15].

However the historical questions are answered, it is something of a shame that envelopes are not included in modern curricula, even for enrichment. The topic has obvious visual appeal, and the method is an attractive application of differentiation. In addition, the consideration of why and how the method works leads to interesting insights. And of course, if our students knew about envelopes, the solution of the ladder problem would be much simplified. Still, taking everything into consideration, this topic is probably too great a digression for most calculus classes. No doubt, embarking on such a digression just to reach an elegant solution to the ladder problem would be (dare I say it) *pushing the envelope.*

As a compromise, it might be reasonable to guide students through a construction of the boundary curve of Ω, without using the general method of envelopes. Here is one approach. Consider sweeping out the region Ω using a segment of length L. For each value of α, there will be one position of the line segment, given by (2). Now for a fixed value of x_0, consider the points $(x_0, y_0(\alpha)))$ that lie on the various line segments. Evidently, the maximum value $y_0(\alpha)$ defines the point of the boundary curve corresponding to x_0. From (2), we have

$$y_0(\alpha) = L \sin \alpha - x_0 \tan \alpha$$

and the maximum value for $0 \le \alpha \le \pi/2$ is easily found to be

$$y_0 = (L^{2/3} - x_0^{2/3})^{3/2}.$$

In this way, the boundary curve is obtained. But to solve the ladder problem, we do not really need the entire boundary curve. All we need to know is where the point (a, b) lies relative to the boundary curve. So, in the preceding argument, simply take $x_0 = a$. This provides another approach to the ladder problem.

Technicalities

In deriving the envelope algorithm, one generally assumes that locally the envelope is a curve smoothly parameterized by α. By definition, each point P of the envelope is tangent to some member of the family of curves, and each member of the family is tangent to the envelope at some point P. This suggests that P can be defined as a function of α. However, some caution is necessary. If a curve in the family touches the envelope in multiple points, there will be ambiguity in defining $P(\alpha)$. This is the situation when we generate the astroid as the envelope of a family of ellipses, as in Figure 4. In this case there are many functions $P(\alpha)$ that map the parameter domain to the envelope, not all of which are continuous. Of course in this example it is possible to choose $P(\alpha)$ consistently to obtain a smooth parameterization of the envelope. But it is not clear how this can be done in general. Accordingly, the existence of a smooth parameterization $P(\alpha) = (x(\alpha), y(\alpha))$ is assumed.

With that assumption, observe that the equation $F(x(\alpha), y(\alpha), \alpha) = 0$ holds identically, so the derivative with respect to α is zero. Viewing F as a function of three variables, the chain rule gives

$$\frac{\partial F}{\partial x}\frac{dx}{d\alpha} + \frac{\partial F}{\partial y}\frac{dy}{d\alpha} + \frac{\partial F}{\partial \alpha} = 0 \tag{7}$$

But we can also view F as a function of two variables, thinking of α as a fixed parameter. In this view, the xy gradient of F is normal to the curve $F(x, y, \alpha) = 0$ at each point. Meanwhile, the parameterization of the envelope provides a tangent vector $(\frac{dx}{d\alpha}, \frac{dy}{d\alpha})$ at each point of that curve. At the point $(x(\alpha), y(\alpha))$, the two curves are tangent, so the normal vector $\nabla_{xy}F = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y})$ is orthogonal to the velocity vector $(\frac{dx}{d\alpha}, \frac{dy}{d\alpha})$. This shows that the first two terms on the left side of (7) add to 0, and hence $\frac{\partial F}{\partial \alpha} = 0$.

The preceding argument is the basis for the envelope algorithm. It shows that at each point (x, y) of the envelope there is a value of α for which both (3) and (4) hold. This is a necessary condition, and it can be satisfied by points which are not on the envelope. Indeed, we can construct an example of this phenomenon by reparameterizing the family of curves. The general idea behind this construction will be clear from the specific case of the family of lines for the ladder problem.

Let $s(\alpha)$ be any differentiable function from $(0, \pi/2)$ onto itself. Then the equation

$$\frac{x}{\cos s(\alpha)} + \frac{y}{\sin s(\alpha)} = L$$

parameterizes the same family of lines as (2), and so has the same envelope. Applying the envelope algorithm, we compute the partial derivative with respect to α, obtaining

$$\frac{x \sin s(\alpha)\ s'(\alpha)}{\cos^2 s(\alpha)} - \frac{y \cos s(\alpha)\ s'(\alpha)}{\sin^2 s(\alpha)} = 0$$

Clearly, this equation will be satisfied for any value of α where $s'(\alpha) = 0$. If $\alpha*$ is such a point, then every point of the corresponding line

$$\frac{x}{\cos s(\alpha*)} + \frac{y}{\sin s(\alpha*)} = L$$

satisfies the two conditions of the envelope algorithm. That is, the entire line segment corresponding to $\alpha*$ will be produced by the envelope algorithm. No such line is actually included in the boundary of Ω, nor can any such line be tangent to all the lines in the family. This illustrates how the envelope algorithm can produce extraneous results.

A more complete discussion of these technical points can be found in [2, 10]. In particular, conditions that can give rise to extraneous results from the envelope algorithm are characterized. As Courant remarks, once the envelope algorithm produces a curve, "it is still necessary to make a futher investigation in each case, in order to discover whether it is really an envelope, or to what extent it fails to be one." In practice, graphing software can often give a clear picture of the envelope of a family of curves, and so guide our understanding of the results of the envelope algorithm.

A technical point of a slightly different nature concerns the relationship between the envelope of a family of curves, and the boundary of the region that family encompasses.

By definition, the envelope is a curve which is tangent at each of its points to some member of the family. This is the definition used to justify the envelope algorithm. But the curve we are interested in for the ladder problem is defined as a boundary curve. How are these two concepts related? Visually, it appears obvious that at each point of the boundary of Ω, the tangent line is a member of the family of lines defining Ω. We substantiate this appearance as follows.

Since each line segment is contained within the region Ω, none of the lines can cross the boundary curve. On the other hand, each point of the boundary must lie on one of the lines. To see this, consider a boundary point P, and a sequence of points P_j in Ω converging to P. Each point P_j is on a line for some parameter value α_j, and these values all lie in the interval $[0, \pi/2]$. So there is a convergent subsequence α_{j_k} with limit $\alpha*$. Now by the continuity of (2), $P = \lim P_{j_k}$ is a point on the line with parameter $\alpha*$. Since this line cannot cross the boundary curve at P, it must be tangent there.

This suggests as a general principle that the boundary of a region swept out by a family of curves lies on the envelope for that family. As in the earlier discussion, some caution is necessary. Here, it is sufficient to assume that the boundary curve is smoothly parameterized by α, the parameter defining the family of curves. On any arc where this is true, the boundary curve will indeed fall along the envelope. On the other hand, consider the following:

$$F(x, y, \alpha) = x^2 + y^2 - \sin^2 \alpha$$

This describes a family of circles centered at the origin, with radius varying smoothly between 0 and 1. The region swept out by the family of circles is the closed unit disk $x^2 + y^2 \leq 1$, and the unit circle is the boundary curve. But the unit circle is not an envelope for the family of circles. What went wrong? Arguing as above, we can again assign a value of α to each point of the boundary curve. But to do this continuously, we have to take α to be constant, say $\alpha = \pi/2$. Then the entire boundary is one of the curves in the family, but it is not parameterized by α. Interestingly, in this example, the curve for $\alpha = \pi/2$ also satisfies the equation $\frac{\partial F}{\partial \alpha} = 0$, so this is an example where the envelope algorithm does locate the boundary of the region, but does not produce the envelope.

Relating the envelope to the boundary also gives a different insight about why the envelope technique works. View (3) as defining a level surface S of the function F in $xy\alpha$ space. The family of curves is then defined as level curves of this surface. At the same time, the region Ω swept out by the family of curves is the projection of S on the xy plane. Now suppose A is a point on S that projects to a point P on the boundary of Ω. The tangent plane to S at A must be vertical and project to a line in the xy plane. Otherwise, there is an open neighborhood of A on the tangent plane that projects to an open neighborhood of P in the xy plane, and that puts P in the interior of Ω. So the tangent plane is vertical. That implies a horizontal normal vector to S at A. But that means that the gradient of F, which is normal to the surface at each point, must be horizontal at A. This shows that the partial derivative of F with respect to α vanishes at A, which is the derivative condition of the envelope algorithm.

As a final topic in this section, I cannot resist mentioning one more way to think about the envelope of a family of curves. The idea is to consider a point on the envelope as the intersection point of two *neighboring* members of the family. That is, define a curve C_α by choosing one value of α, and intersect it with the curve for the *next* value of α. Of course, that is not literally possible, but we can implement this idea using limits. Just express the intersection of the curves C_α and $C_{\alpha+h}$ as a function of h and α, and take the limit as h goes to 0. It is instructive to carry this procedure out for the example of the ladder problem. It once again yields the envelope as the astroid (1). In the process, one can observe differentiation with respect to α implicitly occuring in the calculation of the limit as h goes to 0. Indeed Courant [2] uses this idea to provide a heuristic derivation of the envelope algorithm, before developing a more rigorous justification. In contrast, Rutter [10] terms this the *limiting position* definition of *envelope*, one of three closely related but distinct definitions that he considers. He also provides the following interesting example of a family of circles with an envelope, but for which neighboring circles in the family are disjoint.

Begin with the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

and at each point, compute the osculating circle. This is the circle whose curvature matches that of the ellipse at the specified point, and whose center and radius are the center and radius of curvature of the ellipse. The family of osculating circles for all the points of the ellipse is the focus for this example. This situation is illustrated for an ellipse with $a = 8$ and $b = 4$ in Figure 5, which shows the ellipse and several members of the family of circles.

Place figure about here.

Figure 5: Circles of curvature for an elliptical arc.

It is apparent from the construction of this example that the original ellipse is tangent to each of the circles in the family, and so is an envelope for the family. But the ellipse can not be obtained as the limiting points of intersections of neighboring circles. In fact, the neighboring circles are disjoint! This surprising state of affairs is shown in Figure 6, with an enlarged view in Figure 7.

Place figure about here.

Figure 6: Neighboring circles are disjoint.

Place figure about here.

Figure 7: Closeup view of neighboring circles.

This same example also exhibits some of the other exceptional behaviors that have been discussed above. One can show that the original ellipse does satisfy the two conditions specified in the envelope algorithm, and so the algorithm would properly identify the envelope for this family of circles. But the entire circles of curvature for each of the vertices $((\pm a, 0), (0, \pm b))$ also satisfy the conditions of the envelope algorithm, although these circles are *not* part of the envelope. And these circles also contain the boundary of the region swept out by the family of circles. These properties can be observed in the next two figures, defined by two different ellipses. Note also how visually striking these figures are.

Place figure about here.

Figure 8: Family of circles for an ellipse.

Place figure about here.

Figure 9: Family of circles for another ellipse.

Experimenting with figures of this sort can convey a good deal of understanding of the properties of envelopes discussed above, and I highly recommend it. Modern graphical software is ideally suited to this purpose. For the family of circles in the preceding example, graphical exploration is abetted by the following formulae ([10, p. 192]). Identify one point of the ellipse as $(a\cos\alpha, b\sin\alpha)$. Then the radius of curvature at that point is given by

$$\frac{(a^2\sin^2\alpha + b^2\cos^2\alpha)^{3/2}}{ab}$$

and the center of the circle of curvature is

$$\left(\frac{a^2-b^2}{a}\cos^3\alpha, \frac{b^2-a^2}{b}\sin^3\alpha\right).$$

Extending the ladder problem

A slight variation on the ladder problem is illustrated in Figure 10, with a rectangular alcove in the corner where the two hallways meet. The same configuration might occur if there is some sort of obstruction, say a table or a counter, in one hallway near the corner. As before, the problem is to find how long a line segment will go around this corner.

Place figure about here.

Figure 10: Family of circles for another ellipse.

Here, the envelope method again provides an immediate solution. Consider again the region Ω swept out by a family of lines of fixed length L. If this region avoids both points (a, b) and (c, d), then the segment can be moved around the corner. As L increases, the envelope (1) expands out from the origin. The maximal feasible L occurs when the envelope first touches one of the corner points (a, b) and (c, d). This shows that the maximal value of L is given by

$$L_{\max} = \min\left\{(a^{2/3} + b^{2/3})^{3/2}, (c^{2/3} + d^{2/3})^{3/2}\right\}$$

Going a bit further in this direction, we might replace the inside corner with any sort of curve C (see Figure 11). The ladder problem can then be solved by seeking the point (x, y) of C for which $f(x, y) = x^{2/3} + y^{2/3}$ is minimized. Unfortunately, this plan is not so easy to execute. For example, an elliptical arc is a natural choice for the curve C. But even for that simple case the analytic determination of the minimal value of f appears quite formidable, if not impossible. On the other hand, if C is a polygonal path, we need only find the minimum value of f at the vertices.

The Couch Problem. Extending the problem in a different direction, we can make the situation a bit more faithful to the real world by recognizing that a ladder actually

Place figure about here.

Figure 11: Ladder problem with a curve in the corner.

has some positive width. Thus, in the idealized geometry of the problem statement, perhaps we should try to manuever a rectangle rather than a line segment around the corner. If the width of the rectangle is fixed at w, what is the greatest length L that permits the rectangle to go around the corner?

This version of the problem also provides a reasonable model for moving bulkier objects than ladders. For example, trying to push a desk or a couch around a corner in a corridor is naturally idealized to the problem of moving a rectangle around the corner in Figure 1. This is the motivation given by Moretti [9] in his analysis of the rectangle version of the ladder problem. In honor of his work, we refer to the rectangular version hereafter as the couch problem.

Moretti's analysis mimics the standard solution to the ladder problem. Thus, rather than looking for the longest couch that will go around the corner, he seeks the shortest couch that will get stuck. This occurs when the outer corners of the rectangle touch the outer walls of the corridor, and the inner edge touches the inside corner point, as illustrated in Figure 12. Using the slope as a parameter, Moretti reduces the problem to finding a particular root of a sixth degree polynomial.

Place figure about here.

Figure 12: This couch is stuck.

For the couch problem, as for the ladder problem, the direct approach using envelopes is illuminating. Indeed, we again make use of the astroid, and one of its *parallel* curves.

Here, a *parallel* curve means one whose points are all at a fixed distance from a given curve. From each point P on the given curve C, move a fixed distance w along the normal vector to locate a point Q, being careful to choose the direction of the normal vector consistently. The locus of all such Q is a curve parallel to C at distance w. Parallel curves are discussed in [10].

For the couch problem, the envelope we need is parallel to the envelope we found for the ladder problem. This leads to the following appealing geometric interpretation. Let L be the longest rectangle of width w that can be moved around a corner as in the original ladder problem. Then the boundary of Ω (as in Figure 2) must be tangent to the circle of radius w centered at the point (a, b). That is, w must be the distance from the corner point (a, b) to the astroid (1). Algebraically, this approach has an appealing simplicity, up to the point of actually finding a solution. Unfortunately, that requires solving a sixth degree equation, which is essentially equivalent to the one considered by Moretti.

On the other hand, the geometric setting of the envelope approach provides a simple method for parameterizing a family of rational solutions to the couch problem. That is, we can specify an infinite set of triples (a, b, w) such that the couch problem has an exact rational solution L. This partially answers one of Moretti's questions. In fairness, though, a similar parameterization can be developed using Moretti's method.

To apply the envelope method to the couch problem, we adopt the same strategy as for the ladder problem. Consider the following process for moving a rectangle around a corner. Initially, the rectangle is aligned with the walls of the corridor, so that the bottom of the rectangle is on the x axis and the left side is on the y axis. Slide the rectangle in such a way that the lower left-hand corner follows the y axis, while keeping the lower right-hand corner on the x axis. Thus, the bottom edge of the rectangle follows the exact trajectory of the segment in the ladder problem, sweeping out the region Ω, as before. But now we want to look at region swept out by the entire rectangle. The boundary of this region is the envelope of the family of lines corresponding to the motion of the *top* edge of the rectangle. For a couch with length L and width w, these lines are characterized as follows. Begin with a line in the family for the original ladder problem, whose intersection with the first quadrant has length L. Construct the parallel line at distance w (and in the direction away from the origin). We seek the envelope of the family of all of these parallel lines.

As before, we parameterize the lines in this new family in terms of the angle α between such a line and the (negatively directed) x axis. The parallel unit vector is given by $\mathbf{m} = (-\cos\alpha, \sin\alpha)$, and the normal unit vector (pointing into the first quadrant) is $\mathbf{n} = (\sin\alpha, \cos\alpha)$. These vectors provide a simple way to define a line at a specified distance d from the origin: begin with the line through the origin parallel to $\mathbf{m}$, and translate by $d\mathbf{n}$. That defines a point on the line as

$$(x, y) = t\mathbf{m} + d\mathbf{n}$$

Taking the dot product of both sides of this equation with $\mathbf{n}$ thus gives

$$\sin\alpha\, x + \cos\alpha\, y = d.$$

Lines in the original family are described by (2), which we rewrite as

$$\sin\alpha\, x + \cos\alpha\, y = L\,\sin\alpha\,\cos\alpha.$$

This line is at a distance $L\,\sin\alpha\,\cos\alpha$ from the origin. Now we want the parallel line that is w units away. The equation for that line is evidently

$$\sin\alpha\, x + \cos\alpha\, y = L\,\sin\alpha\,\cos\alpha + w.$$

To make use of the envelope algorithm, let us define the function

$$G(x, y, \alpha) = \sin\alpha\, x + \cos\alpha\, y - L\,\sin\alpha\,\cos\alpha - w.$$

Then, thinking of α as a fixed value, the equation $G(x, y, \alpha) = 0$ defines one line in the family. Similarly, with

$$F(x, y, \alpha) = \sin\alpha\, x + \cos\alpha\, y - L\,\sin\alpha\,\cos\alpha$$

we obtain the lines in the original family by setting $F(x, y, \alpha) = 0$. It will be convenient in what follows to express these functions in the form

$$\begin{aligned} F(x, y, \alpha) &= \mathbf{n} \cdot (x, y) - L \sin\alpha \cos\alpha \\ G(x, y, \alpha) &= \mathbf{n} \cdot (x, y) - L \sin\alpha \cos\alpha - w \end{aligned}$$

Our goal is to find the envelope for the lines defined by G, (hereafter, the envelope for G). According to the envelope algorithm, we should eliminate α from the equations

$$\begin{aligned} G(x, y, \alpha) &= 0 \\ \frac{\partial}{\partial \alpha} G(x, y, \alpha) &= 0 \end{aligned}$$

But rather than apply this directly, we can use the fact that we know the envelope for F. In fact, since each line in $G's$ family is parallel to a corresponding line in $F's$ family, and at a uniform distance w, it is not surprising that the envelope of G is parallel to the envelope of F, and at the same distance. That is, if (x, y) is on the envelope of F, then the corresponding point of the envelope of G is w units away in the normal direction.

To make this more precise, let us consider a point (x, y) on the envelope of F. There is a corresponding α such that (x, y, α) is a zero of both F and $\frac{\partial F}{\partial \alpha}$. Then (x, y) is on the line with parameter α, which is tangent to the envelope of F at (x, y). Thus, at this point, the line and the envelope share the same normal direction. As observed earlier, the unit normal is given by $\mathbf{n} = (\sin\alpha, \cos\alpha)$. We will now consider a new point $(x', y') = (x, y) + w\mathbf{n}$. We wish to show that (x', y') is on the envelope of G.

To that end, observe that $F(x, y, \alpha) = G(x', y', \alpha)$ and $\frac{\partial F}{\partial \alpha}(x, y, \alpha) = \frac{\partial G}{\partial \alpha}(x', y', \alpha)$. To justify the first of these equations,

$$\begin{aligned} G(x', y', \alpha) &= \mathbf{n} \cdot (x', y') - L \sin\alpha \cos\alpha - w \\ &= \mathbf{n} \cdot [(x, y) + w\mathbf{n}] - L \sin\alpha \cos\alpha - w \\ &= \mathbf{n} \cdot (x, y) + w - L \sin\alpha \cos\alpha - w \\ &= F(x, y, \alpha) \end{aligned}$$

To justify the second equation, we observe first that since F and G differ by a constant, they have the same derivative. Also, note that $\frac{\partial \mathbf{n}}{\partial \alpha} = -\mathbf{m}$. This gives $\frac{\partial G}{\partial \alpha}(x, y, \alpha) = \frac{\partial F}{\partial \alpha}(x, y, \alpha) = -\mathbf{m} \cdot (x, y) - L(\cos^2 \alpha - \sin^2 \alpha)$. Now we can write

$$\begin{aligned} \frac{\partial G}{\partial \alpha}(x', y', \alpha) &= -\mathbf{m} \cdot [(x, y) + w\mathbf{n}] - L(\cos^2 \alpha - \sin^2 \alpha) \\ &= -\mathbf{m} \cdot (x, y) - L(\cos^2 \alpha - \sin^2 \alpha) \\ &= \frac{\partial F}{\partial \alpha}(x, y, \alpha). \end{aligned}$$

Together, these results show that (x, y) is on the envelope of F if and only if (x', y') is on the envelope of G, and that in each case the points (x, y) and (x', y') correspond to the same value of α. In fact, this result reflects a more general situation: If the family G consists of parallels of the curves in F, all at a fixed distance w, then the envelope for G is the parallel of the envelope of F, at the same distance w. In the context of the couch problem, we can find the needed envelope of G as a parallel to the known envelope of F.

Based on earlier work, we know that the envelope of F is parameterized by the equations

$$\begin{aligned} x &= L \cos^3 \alpha \\ y &= L \sin^3 \alpha. \end{aligned}$$

That leads immediately to the following parametric description of the envelope of G :

$$\begin{aligned} x &= L \cos^3 \alpha + w \sin \alpha \\ y &= L \sin^3 \alpha + w \cos \alpha. \end{aligned}$$

For the solution L of the couch problem, the point (a, b) must lie on the envelope of G. Therefore, we can find L (and also find the critical value of α) by solving the system

$$\begin{aligned} a &= L \cos^3 \alpha + w \sin \alpha \\ b &= L \sin^3 \alpha + w \cos \alpha. \end{aligned}$$

This leads readily enough to an equation in α alone:

$$a \sin^3 \alpha - b \cos^3 \alpha = w(\sin^2 \alpha - \cos^2 \alpha).$$

At this point, finding α appears to depend on solving a sixth degree polynomial equation, which is easy to do numerically (given values for a, b, and w), and very likely impossible to do symbolically.

As mentioned, this analysis leads to a nice geometric interpretation for the solution. If (a, b) is on the envelope of G, then there is a corresponding point (x, y) on the envelope of F. We know that (x, y) is w units away from (a, b), and that the vector between these two points is normal to the envelope of F. This shows that the circle centered at (a, b) of radius w is tangent to the envelope of F at (x, y).

Visually, we can see how to find the maximum value of L. Start with a small enough L so that the astroid (1) stays well clear of the circle about (a, b) of radius w. Now increase L, expanding the astroid out from the origin, until the curve just touches the circle. When that happens, the corresponding value of L is the solution to the couch problem. See Figure 13.

Place figure about here.

Figure 13: Maximizing L geometrically.

The visual image of solving the couch problem in this way is reminiscent of Lagrange Multipliers. Indeed, what we have is the dual of a fairly typical constrained optimization problem: find the point on the curve (1) that is closest to (a, b). The visual image for that problem is to expand circles centered at (a, b) until one just touches the astroid. Our dual problem is to hold the circle fixed and look at level curves for increasing values of the function $f(x, y) = x^{2/3} + y^{2/3}$. We increase the value of f until the corresponding level curve just touches the fixed circle. This geometric conceptualization is associated with the following optimization problem: Find the minimum value of $f(x, y)$ where (x, y) is constrained to lie on the circle of radius w centered at (a, b). We will return to the idea of dual problems in the last section of the paper.

Let us examine more closely the Lagrangian-esque version of the couch problem. We wish to find a point of tangency between the following two curves

$$\begin{aligned} x^{2/3} + y^{2/3} &= L^{2/3} \\ (x-a)^2 + (y-b)^2 &= w^2. \end{aligned}$$

For each curve, we can compute a normal vector as the gradient of the function on the left side of the curve's equation. Insisting that these gradients be parallel leads to the following additional condition

$$x^{1/3}(x-a) = y^{1/3}(y-b)$$

In principle, solving these three equations for x, y, and L would produce the desired solution L to the couch problem. Or, solving the second and third for x and y, and then substituting those values in the first equation, also would lead to the value of L. Unfortunately, every approach seems to lead inevitably to a sixth degree equation.

While the envelope method does not seem to provide a symbolic solution to the couch problem, it does provide a nice procedure for generating solvable examples. Here, we will begin with a value of L and produce a triple (a, b, w) so that L is a solution to the (a, b, w) couch problem. To begin, we generate some *nice* points on the astroid $x^{2/3} + y^{2/3} = L^{2/3}$ using Pythagorean triples. Specifically, if $r^2 + s^2 = t^2$, we can take $x = r^3$, $y = s^3$ and $L = t^3$ to define a point on an astroid. In particular, we can generate an abundance of rational points on astroids. Notice that the orginal Pythagorean triple need not be rational. For example, if $(r, s, t) = (3, 4, 5)/\sqrt[3]{5}$, we find $(27/5, 64/5)$ as a rational point on the astroid curve for $L = 25$.

Now the equations $x = r^3$, $y = s^3$ are closely related to the parameterization

$$x = L\cos^3\alpha \quad y = L\sin^3\alpha$$

of the astroid. As a result, we can recover α from the equations

$$\cos\alpha = \frac{r}{t} \quad \sin\alpha = \frac{s}{t}$$

This in turn gives us the normal vector $\mathbf{n} = (\frac{s}{t}, \frac{r}{t})$, and hence, for any value of w, leads to the point (x', y'). Define that point to be (a, b). It necessarily lies on the envelope for G. This shows that the L for the astroid constructed at the outset solves the (a, b, w) couch problem. We formalize these arguments in the following theorem.

Theorem: *For any positive pythagorean triple (r,s,t) and any positive w define*

$$\begin{aligned} a &= r^3 + w\frac{s}{t} \\ b &= s^3 + w\frac{r}{t} \\ L &= t^3 \end{aligned}$$

Then L is the solution to the (a, b, w) *couch problem.*

For example, with $(r, s, t) = (3, 4, 5)/\sqrt[3]{5}$ and $w = 2$, the equations above give $(a, b) = (7, 14)$ and $L = 25$. So for a rectangle of width 2, 25 is the maximum length that will fit around the corner defined by the point $(7, 14)$. In general, if (r', s', t') is a rational Pythagorean triple, and if u^3 is rational, then taking $(r, s, t) = u(r', s', t')$ and rational w produces rational values of a, b, and L, as well as a rational point (x, y) where the astroid meets the circle centered at (a, b) of radius w.

The preceding example, where $L = 25$ is the solution of the $(7, 14, 2)$ couch problem, was given by Moretti. He mentioned that such examples are relatively rare, and asked for conditions on a, b, and w that make the (a, b, w) couch problem exactly solvable. The theorem above provides a partial answer to Moretti's question, by providing an infinite family of such triples. It would be nice to know whether every rational (a, b, w) with rational solution L to the couch problem arises in this way. If the critical value of α corresponds to a rational point (x, y) on the astroid (1), then a, b, w, and L are related as in the theorem. But there might be rational (a, b, w) for which the solution to the couch problem is also rational, but which does not correspond to a rational point (x, y).

The envelope approach leads in a natural way to the theorem, and provides a nice geometric interpretation of the couch problem solution. But it should be observed that Moretti's approach can also lead to an equivalent method for parameterizing triples (a, b, w) with rational solution L. He formulates the problem in terms of a variable m

(corresponding to $\cot \alpha$ in this paper) and derives a sixth degree equation in m with coefficients that depend on a, b, and w. If that equation is solved for w, one can again parameterize solutions in terms of Pythagorean triples. From this standpoint, the envelope method does not seem to hold any advantage over Moretti's earlier analysis.

Duality in the Ladder Problem

In discussing the tangency condition for an astroid and an ellipse, the idea of dual optimization problems was briefly mentioned. As a concluding topic, we will look at this idea again.

Segalla and Watson [12] discuss what they call the flip side of a constrained optimization problem in the context of Lagrange multipliers. For example, in seeking to maximize the area of a rectangle with a specified perimeter, we have an objective function (the area) and a constraint (the perimeter). At the solution point, the level curve for the extreme value of the objective function is tangent to the given level curve of the constraint function. Here, the roles of the objective and constraint are symmetric, and can be interchanged. Given the maximal area, we can ask what is the minimal perimeter that can enclose a rectangle having this area. The solution corresponds to the same point of tangency between level curves of the objective and constraint functions. Thus, we see that the problem of maximizing area with a fixed perimeter, and minimizing the perimeter with a fixed area are linked.

Maximizing area with fixed perimeter is the famous isoperimetric problem (see Blåsjö [1] for a beautiful discussion), and in that context minimizing the perimeter with fixed area is referred to as the *dual* problem. Duality in this sense corresponds to Segalla and Watson's idea of flip side symmetry. It is also reminiscent of the idea of duality in linear and non-linear programming. There, although the primal and dual optimization problems occur in different spaces, one again finds the idea of linked problems whose solutions somehow coincide.

Duality permits information about one problem to be inferred from information about its dual problem. This property is important in both the isoperimetric problem and in linear programming. As Segalla and Watson point out, a solution to an initial optimization problem immediately leads to a corresponding statement and solution of a dual problem. Thus, discovering that a rectangle with perimeter 40 has maximal area 100, also tells us at once that a rectangle with area 100 has minimal perimeter 40. But there is another way to use duality. If you are unable to solve an optimization problem, try to solve the dual.

Here is how this works for the perimeter-area problem. Imagine that we do not know how to maximize the area of a rectangle with perimeter 40. The dual problem is to minimize the perimeter subject to a given area, but of course we do not know what that fixed area should be. So we solve the general problem for a fixed area of A. That is, we prove that area A occurs with a minimal perimeter of $4\sqrt{A}$. Now relate this to the original problem by insisting on a perimeter of 40. That forces $A = 100$, and tells us this: for area 100, the minimal perimeter is 40. The dual statement now solves the original problem.

Segalla and Watson give several examples of pairs of dual problems. In addition to the area-perimeter example mentioned above, they discuss the *milkmaid problem*: find the minimum distance the milkmaid must walk from her home to fetch water from a river and take it to the barn. In this case the dual problem fixes the length of the milkmaid's hike, and minimally shifts the river to accommodate her. They also give the example of the ladder problem, but do not describe the dual. Indeed, they ask for an interpretation of the dual ladder problem. Let us answer this question using envelopes, and see how the dual version leads to another solution of the ladder problem.

Segalla and Watson use the standard approach to the ladder problem – finding the minimal length segment that will get stuck in the corner. We formulate this as a constrained optimization problem in terms of variables u and v, interpreted as intercepts on the x and y axes of a line segment in the first quadrant. The objective function, $f(u, v) = \sqrt{u^2 + v^2}$, is the distance between the intercepts. The goal is minimize this distance subject to the constraint that the line must pass through (a, b).

For the dual problem, if we hold f fixed, and look at varying values of the constraint function, what does that mean? An answer will depend, naturally, on how the constraint g is formulated. Here is one approach. A line with intercepts u and v satisfies the equation

$$\frac{x}{u} + \frac{y}{v} = 1.$$

From this equation, the condition that (a, b) lie on the line is

$$\frac{a}{u} + \frac{b}{v} = 1.$$

Accordingly, define $g(u, v) = a/u + b/v$. Now observe that $g(u, v) = t$ means that $(a/t, b/t)$ lies on a line with intercepts u and v. That is, g measures the (reciprocal of the) distance from the origin to the line for u and v along the ray through (a, b). This gives the following meaning to the dual problem: Look at all the lines with intercepts u and v, where fixing the value of f means that the distance between these intercepts is constant. Among all these lines, find the one whose distance from the origin, measured in the direction of (a, b) is a maximum.

The by now familiar astroid appears once again as the envelope of a family of lines. Holding f constant with value L, we are again considering the family of line segments of length L with ends on the positive x and y axes, filling up the region Ω. The point that maximizes g will now be the furthest point you can reach in Ω traveling on the ray from the origin to (a, b). That is, the solution occurs at the intersection of the ray with the envelope (1).

The solution point will be of the form $(a/t, b/t)$ where t is the optimal value of g. Substituting in (1), we find $t = ((a/L)^{2/3} + (b/L)^{2/3})^{3/2}$. This gives us the optimal value of g for the dual problem, in terms of L. To return to the primal problem, we have to choose the value of L that gives us the original constraint value for g, namely $g = 1$. So with $t = ((a/L)^{2/3} + (b/L)^{2/3})^{3/2} = 1$ we again find $L = (a^{2/3} + b^{2/3})^{3/2}$.

It is interesting that the ladder problem has so many formulations. The usual approach is to reverse the original problem, so that we seek a minimal line that cannot go around the corner rather than a maximal line that will go around the corner. The envelope approach presented here deals directly with the problem as stated, finding the maximum line that will fit around the corner. A third approach is to take the dual of the reversed version, viewed as an example of constrained optimization. Although all of these approaches are closely related, each contributes a slightly different understanding of the problem.

References

[1] Viktor Blåsjö, The Isoperimetric Problem, *American Mathematical Monthly* 112 (2005), 526–566

[2] Richard Courant, *Differential and Integral Calculus, Volume 2,* tranlated by E. J. McShane, Interscience, New York, 1949

[3] L. M. Court, Envelopes of Plane Curves, *American Mathematical Monthly* 57 (1950), 168–169

[4] Howard Eves, A Note on Envelopes, *American Mathematical Monthly* 51 (1944), 344

[5] Peter J. Giblin, Zigzags, *Mathematics Magazine* 74 (2001), 259–271

[6] J. W. Green, On the Envelope of Curves Given in Parametric Form, *American Mathematical Monthly* 59 (1952), 626–628

[7] Morris Kline, *Advanced Calculus,* 2nd ed., Wiley, New York, 1972.

[8] Brian J. Loe and Nathanial Beagley, The Coffee Cup Caustic for Calculus Students, *College Mathematics Journal* 28 (1997), 277–284

[9] Christopher Moretti, Moving a Couch Around a Corner, *College Mathematics Journal* 33 (2002), 196–201

[10] John W. Rutter, *Geometry of Curves,* Chapman & Hall / CRC, Boca Raton, 2000

[11] Mark Schwartz, The Chair, the Area Rug, and the Astroid, *College Mathematics Journal* 26 (1995), 229–231

[12] Angelo Segalla and Saleem Watson, The Flip-Side of a Lagrange Multiplier Problem, *College Mathematics Journal* 36 (2005), 232–235

[13] Andrew Simoson, An Envelope for a Spirograph, *College Mathematics Journal* 28 (1997), 134–139

[14] Andrew Simoson, The Trochoid as a Tack in a Bungee Cord, *Mathematics Magazine* 73 (2000), 171–184

[15] Angus E. Taylor and W. Robert Mann, *Calculus: an Intuitive and Physical Approach,* 2nd ed., Wiley, New York, 1977.

[16] Robert C. Yates, *Curves and their Properties,* National Council of Teachers of Mathematics, Reston, VA, 1974