Upon Reflection

Andrew Glassner

Microsoft Research

This time we'll look at the geometry of the classic law of mirror reflection. As anyone who's written a ray tracer knows, this law tells us that the angle of incidence equals the angle of reflection. It turns out that this little geometric truth can help solve another set of interesting problems called billiard problems, which describe the shortest path a billiard ball can take on a polygonal table.

We'll start by looking at two different derivations of the law of reflection, one geometric and one analytic. It's always nice to see the same result come from two very different approaches. Then we'll look at the problem of finding the shortest circuit of a billiard ball around a triangular table.

Before we get started, you might want to play around with pencil and paper yourself and see if you can rederive the law of specular reflection. You need only two laws of physics: In a vacuum, light will travel in a straight line unless interfered with, and it always seeks the shortest path.

The geometric approach

Figure 1 shows our basic setup. Light leaves a point P towards a mirror, represented by the line M, and ultimately arrives at point R. We want to find the point Q on M where the light does the bouncing; from that information we can deduce the law of reflection.

Ready to go? We know that three points determine a plane, so putting P, Q, and R all in the plane of the page seems reasonable. We'll use our problem statement from above to prevent paths directly from P to R, though that might happen if P radiates light directly along the line PR. Rather, we'll simply direct our attention to the path of light that actually does bounce off of M along the way.

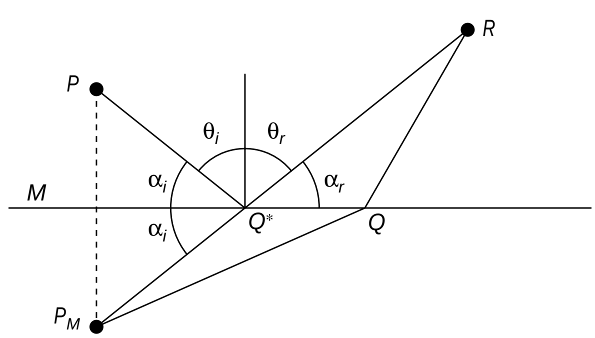
There are a few subtleties we'll deliberately ignore. First, we'll stick completely with geometric optics—no wave effects such as diffraction will be considered. Second, we'll assume we're in a vacuum, so the index of refraction is a constant 1 everywhere. Finally, we'll assume that we haven't heard of Einstein yet, so light travels in perfectly straight lines all the time.

Now we're ready to search high and low, putting our point Q anywhere on M. The goal is to find the placement of Q such that the time it takes the light to get from P to Q to R is minimized. So we want to find Q such that we minimize $|PQ|+|QR|$. In Figure 1, angles θ_i and θ_r refer to the angles of incidence and reflection, and α_i and α_r refer to their complements.

We'll begin by creating point P_M, which is simply P reflected through M. Thus for all Q on M, $|P_MQ|=|PQ|$, and therefore $|PQ|+|QR| = |P_MQ|+|QR|$. So if we minimize the right-hand side of this equality, we also minimize the left.

We know from Euclid that for any triangle with side lengths (a, b, c), $a+b \geq c$, achieving equality only when a, b, and c are collinear and in that order. So in ΔRQP_M, we can say $|P_MQ|+|QR| \geq |P_MR|$. As long as they're not collinear, the sum of the two steps from P_M to Q to R will be longer than the straight shot from P_M to R. Since P_M and R are both fixed, the only way to adjust the path is to move Q, and the shortest path is created when it's placed at Q^*, that point on M that intersects the line P_MR.

So we now have the intersection of two lines, M and P_MR, which meet at Q^*. Thus $\alpha_i = \alpha_r$, and therefore $\theta_i = \theta_r$—and we've proved the law of specular reflection.

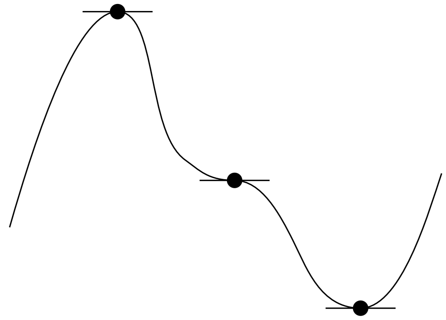


1 **Light from *P* to *R* bounces off of *M* at some point *Q*. The shortest path is given when $Q = Q^*$, located on the line from P_M to *R*.**

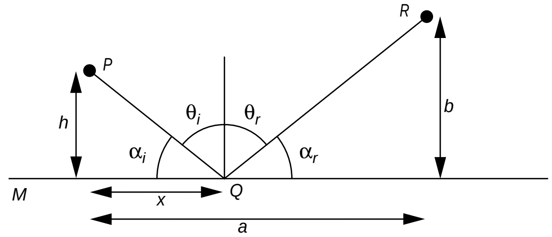
The algebraic approach

We begin the algebraic solution with two definitions.

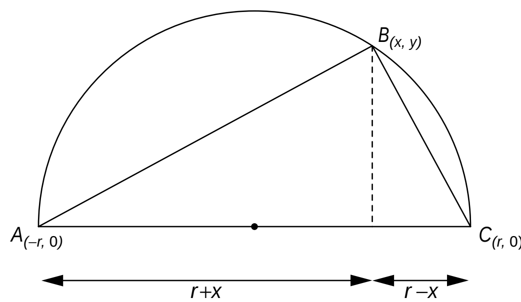
First, the *optical path length* (OPL) is the time it takes a photon to get from one point to another. Since (in our simplified world) we're in a vacuum, the OPL is simply the distance multiplied by the speed of light. For convenience, we'll ignore the speed of light, since it's just a constant scaling factor (or



2 **The black dots mark stationary points—locations where the function has a zero derivative.**



3 **Geometry for an analytic solution of specular reflection.**



4 **Because $\triangle ABC$ is inscribed in a semicircle, $\angle ABC$ is a right angle.**

we could say we're using a system of units where the speed of light is defined to be 1).

Second, a *stationary point* in a 1D function is a point where the function has a derivative of 0; that is, the tangent is parallel to the *X* axis. As shown in Figure 2, that can be at a local maximum, a local minimum, or a flat spot.

Among other things, Pierre de Fermat proposed that a ray of light follows a path corresponding to a stationary value in its OPL. In other words, if we look at a proposed candidate for the path taken by a ray of light, we should find that any other candidate path nearby takes the same amount of time or more. You might enjoy thinking of a physical situation where several different rays of light all take the same time to get from one point to another.

Figure 3 shows the same geometry as in Figure 1, but I've relabeled some of the distances to make the calculations easier. The OPL is simply the sum of the distances

$$OPL = |PQ| + |QR|$$
$$= \sqrt{h^2 + x^2} + \sqrt{b^2 + (a-x)^2}$$

To find the stationary points, we differentiate this with respect to *x* and set the result to 0:

$$\frac{d(OPL)}{dx} = \frac{x}{\sqrt{h^2 + x^2}} - \frac{(a-x)}{\sqrt{b^2 + (a-x)^2}} = 0$$

From Figure 3, we can see that

$$\sin(\theta_i) = \frac{x}{\sqrt{h^2 + x^2}} \text{ and } \sin(\theta_r) = \frac{a-x}{\sqrt{b^2 + (a-x)^2}}$$

Making these substitutions, we find

$$0 = \sin(\theta_i) - \sin(\theta_r)$$

or, more simply, $\theta_I = \theta_r$, and we've come to the same conclusion.

Now that we've pretty much settled the law of specular reflection, we can prove a little theorem about minimum inscribed polygons and billiard tables.

A trio of useful observations

Before we plunge into the land of billiard balls, three observations will make life easier later on. If you're geometrically inclined, these are probably old hat to you.

First, note that any triangle inscribed in a semicircle has a right angle. Figure 4 shows what I'm talking about, using a triangle $\triangle ABC$ in a semicircle of radius *r*, or diameter $d = 2r$. Points *A* and *C* are at opposite ends of the diameter, and *B* is on the circle.

There are lots of ways to prove this. Let's use an algebraic proof. Assume the circle is centered at (0, 0). Then $A = (-r, 0)$, $C = (r, 0)$, and $B = (x, y)$, where $r^2 = x^2 + y^2$, since *B* lies on the circle. Let's find the distance *AC* by assuming that angle $\angle ABC$ is a right angle. (For the rest of this article, when there's only one angle at a vertex, I'll write that vertex as an angle. So $\angle B$ here stands for $\angle ABC$). If $|AC|^2 = |AB|^2 + |BC|^2$, then Pythagoras is satisfied and $\angle B$ is a right angle.

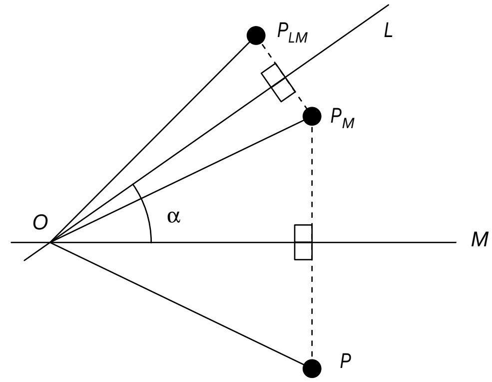
We can prove this just by writing everything out and simplifying:

$$|AC|^2 = |AB|^2 + |BC|^2$$
$$= (r+x)^2 + y^2 + (r-x)^2 + y^2$$
$$= 2((x^2 + y^2) + r^2)$$
$$= 4r^2$$

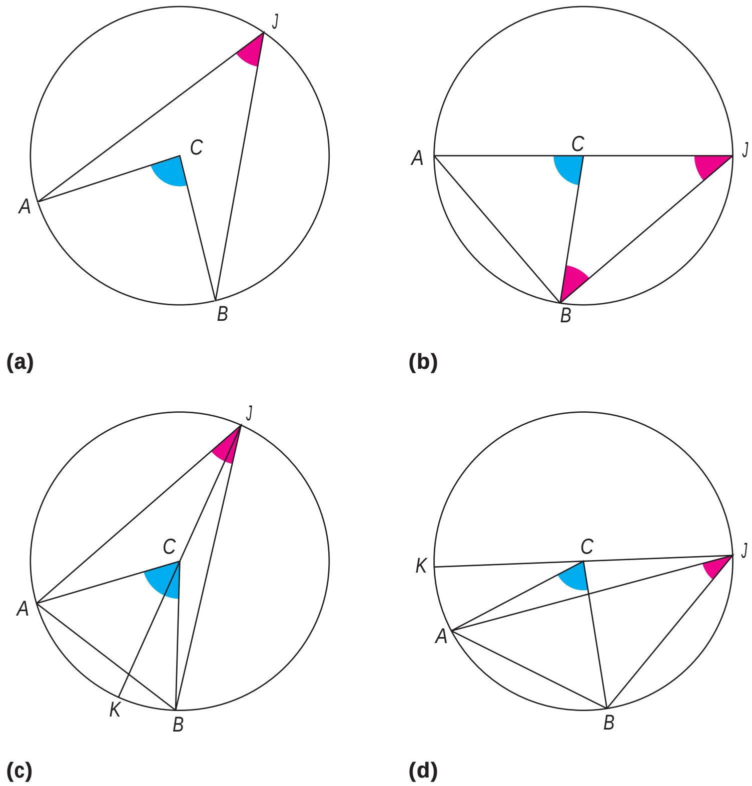
Taking the square root of both sides, $|AC| = 2r$, and therefore $\angle B$ is a right angle.

When several points lie on a circle, such as *A*, *B*, and *C* in this example, we call them *concyclic*.

Second, we'll need an observation that comes up when we reflect a point two times. If we reflect a point once, such as when we reflected *P* to P_M through the line *M* in Figure 1, we can say that P_M (the image of *P* as a result of reflecting through *M*) could have been created by simply translating *P* by the vector $P_M - P$. What happens if we create P_{LM} by reflecting *P* a second time



5 **The product of two reflections in intersecting lines is equivalent to a rotation through their point of intersection by twice the angle between the lines.**



6 **(a) The central angle (blue) is always twice the inscribed angle (red). (b) When the circle center *C* lies on the line *AJ*. (c) When *C* lies in $\triangle AJB$. (d) When *C* lies outside $\triangle AJB$.**

through a different line, say L, that is not parallel to M?

Figure 5 shows the result. If the angle from L to M is α, then P_{LM} can be created either by translation of P (which isn't too interesting) or by rotation of P through an angle 2α, using the intersection point of L and M (labeled O) as the center of rotation.

There are lots of nice proofs of this, which is only one of the many pretty results that come from looking at isometries in the plane. However, Figure 5 speaks eloquently for itself. Since there are two pairs of similar right triangles, and the sum of the angles of one of each at O is α, the sum of all four must be 2α. If you like, you can consider cases where L and M are parallel, or where P starts inside their acute angle, or where P_M appears on the other side of M, and so on. You'll find that when L and M are parallel, the point P is translated along a vector perpendicular to them; when L and M intersect, the result is always a rotation of 2α around their intersection point.

Third, we will want to know that all inscribed angles that contain a common chord are equal. An *inscribed angle* is just a pair of lines that intersect at a point on a circle and contain a chord on that circle, as in Figure 6a. By comparison, a *central angle* is created by a pair of radii that contain a chord. We want to show that in this figure, if we keep the A and B fixed, the inscribed angle $\angle AJB$ is the same no matter where on the circle we place J (as long as it's outside the arc AB).

Let's really nail this one down. The approach we'll take is to prove that an inscribed angle is always half as large as the central angle that inscribes the same arc (or chord). Since there's only one central angle for any arc, this means that all the inscribed angles that contain that arc have the same size. The special case comes when the chord is a diameter, and that's exactly the case we just handled above.

To be very careful, we have to consider all the places J can go. This turns out to demand three cases, based on where the circle's center C falls with respect to the triangle ΔJAB.

In Case 1, illustrated in Figure 6b, C falls on the line JA. Since CJ and CB are radii, ΔBCJ is isosceles, and $\angle CJB = \angle JBC$. Since ACJ is a diameter of the circle, $\angle ACB = \pi - \angle BCJ$. From ΔBCJ, we find that $\pi - \angle BCJ = \angle CJB + \angle JBC$. Putting these together, $\angle ACB = \angle CJB + \angle JBC = 2\angle CJB = 2\angle AJB$, just as we hoped. In words, the central angle $\angle ACB$ is twice the measure of the inscribed angle $\angle AJB$.

How about when J is located somewhere else? In Case 2, illustrated in Figure 6c, J is not on line CA but is inside the triangle ΔAJB. We'll use Case 1 to figure things out, so we include the diameter from J through C; its other intersection with the circle is at K. The diameter JK cuts everything into two halves, each of which looks like Case 1. On the A side, we find from Case 1 that $\angle ACK = 2\angle AJK$. On the B side, $\angle KCB = 2\angle KJB$. Now the central angle $\angle ACB = \angle ACK + \angle KCB = 2(\angle AJK + \angle KJB) = 2\angle AJB$. So again we've found that $\angle ACB = 2\angle AJB$.

Finally, Case 3 is illustrated in Figure 6d, where J is not on line CA and is not in the triangle ΔAJB. As before, we'll draw the diameter from J through C, creating point K. We can see that $\angle ACB = \angle KCB - \angle KCA$. These two right-hand-side angles are easily found from Case 1. $\angle KCB = 2\angle KJB$ (where the enclosed arc is BK), and $\angle KCA = 2\angle KJA$. Subtracting these,

$\angle ACB = 2(\angle KJB - \angle KJA) = 2\angle AJB$.

That wraps up the proof. No matter where we put the vertex of an inscribed angle, it always has the same size, which is twice that of the corresponding central angle.

Playing billiards

When you bounce a billiard ball off a bumper, you expect the same sort of result as the ray of light we studied above—the reflection should be a perfect mirror reflection.

Let's ask the following question: What is the shortest path a billiard ball can take such that it bounces off each table wall and returns to its starting point traveling in the same direction from which it started? If you're willing to discount friction, we could say that we want the ball to come back with the same position and velocity as when it started, so that it could repeat the same circuit over and over. For simplicity, we'll use an acute triangular table in this discussion. At the end I'll mention what happens when we generalize our results to tables with more sides or obtuse angles.

Our answer comes from studying the reflections of the ball as it bounces around. H. Schwarz (1843-1921) developed the technique of the proof; we'll follow Dan Pedoe's presentation in *Geometry* (Dover, 1988).

Let's begin with ΔABC as shown in Figure 7. Since we want one complete cycle, we'll say that the ball follows the inscribed triangle ΔPQR. The point P lies strictly within edge BC—that is, it is on the line between those points and not right on top of either vertex. Similarly, Q is inscribed in edge AB and R in edge AC. Just where should these points lie?

To get started, let's assume that we've already found the ideal locations for P and Q. We might expect that the shortest path from P to Q via line AC would come from reasoning similar to that we used to find the shortest path of light. This would come from placing R so that it forms a perfect mirror reflection from P to Q. We would expect the same property from P and Q as well.

That's a nice intuitive start, but hardly a proof. If we're willing to commit to this line of reasoning, we have to answer two questions:

1. Can we construct a triangle that creates a perfect specular reflection at each edge?
2. Is this the shortest path that forms a repeating circuit?

We'll take these two questions—existence and optimality—in turn. First, we'll analyze the construction of ΔPQR to see if we can in fact make the beast. If we can, we'll have created what is called a *light polygon* (in this case, a light triangle), because of its close relation to the reflection of light.

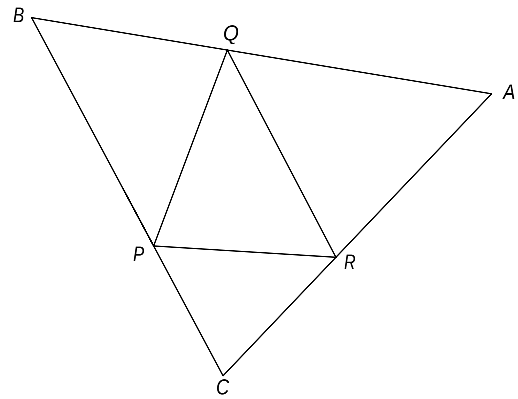
Questions of existence

If you'd like, you can take a break here and try to devise a construction scheme for making a light triangle ΔPQR. I'll cut to the chase and propose a method, and then show that it works.

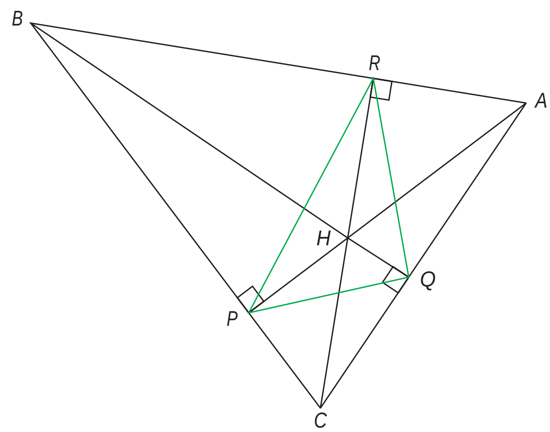
The proposed technique is pretty simple: P, Q, and R are the feet of the altitudes of ΔABC. Recall that the altitudes are the perpendiculars of each edge that pass through the opposite vertex, as shown in Figure 8. The three altitudes meet at point H, called the orthocenter.

Our goal is to show that these points form a light triangle with respect to the outer triangle ΔABC. This means that each pair of edges supports the mirror-reflection law. In Figure 9, I've indicated this with colored dots—each pair of angles that shares a similarly colored dot should be the same size. Look at point R, where $\theta_i = \angle QRH$ and $\theta_r = \angle HRP$. So showing $\theta_i = \theta_r$ requires showing $\angle QRH = \angle HRP$.

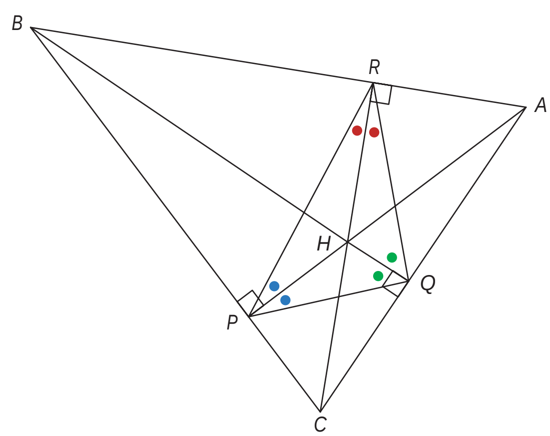
No problem. We'll do this in three steps, as shown in



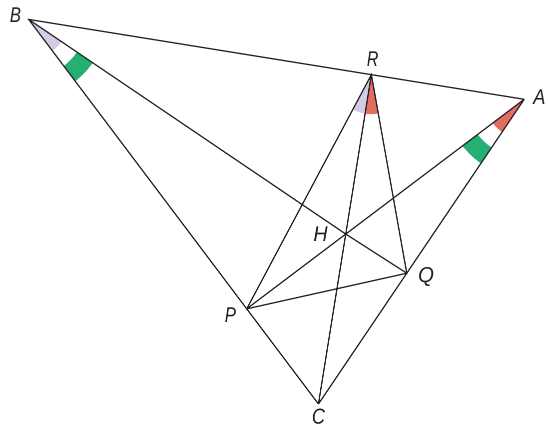
7 **A triangle ΔABC and an inscribed triangle ΔPQR.**



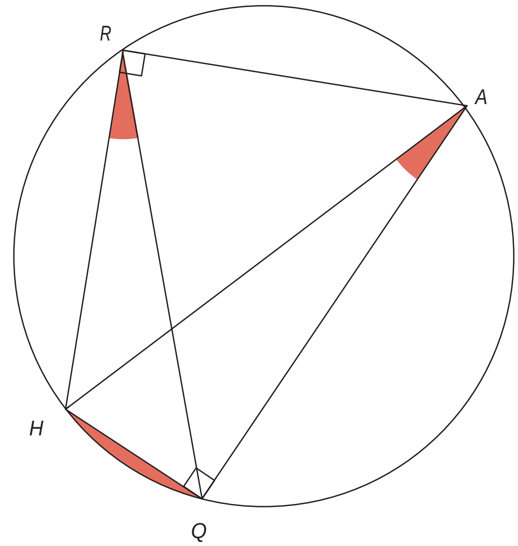
8 **The green triangle ΔPQR is formed by the altitudes of ΔABC.**



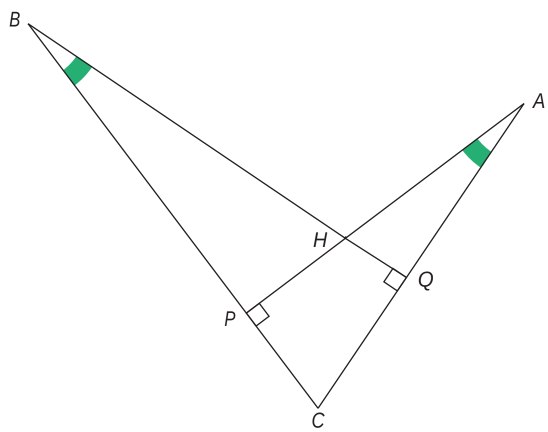
9 **If ΔPQR is a light triangle, then each pair of like-colored angles is equal.**



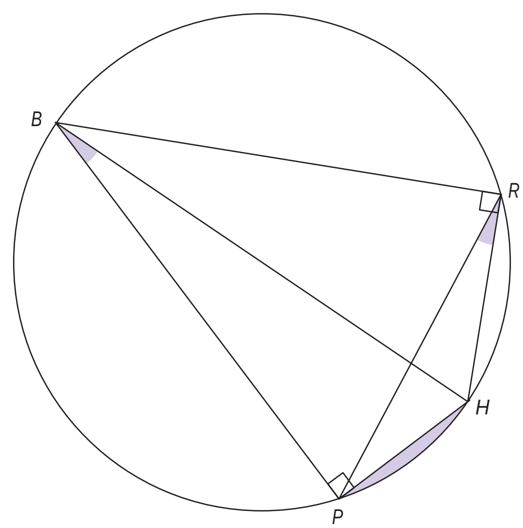
10 **The proof strategy for showing that ∆*PQR* is a light triangle: Step 1 (red wedges): ∠*QRH* = ∠*QAH*. Step 2 (green bands): ∠*QAH* = ∠*HBP*. Step 3 (lavender wedges): ∠*HBP* = ∠*HRP*.**



11 **Proving ∠*QRH* = ∠*QAH*, since both are inscribed in the same circle and both contain arc *QH*.**



12 **Proving ∠*QAH* = ∠*HBP*, since both ∆*CAP* and ∆*CBQ* contain ∠*C* and a right angle.**



13 **Proving ∠*HBP* = ∠*HRP*, since both are inscribed in the same circle and both contain arc *HP*.**

Figure 10. First, we'll show that ∠*QRH* = ∠*QAH* (the red wedges), then ∠*QAH* = ∠*HBP* (the green bands), and then ∠*HBP* = ∠*HRP* (the lavender wedges). Putting these together, ∠*QRH* = ∠*QAH* = ∠*HBP* = ∠*HRP*, which will conclude the proof.

For the first step, we note that *Q*, *H*, *R*, and *A* are concyclic—they all lie on a circle, as shown in Figure 11. Think of *AH* as the diameter of a circle; since ∠*AQH* is a right angle, we know that *Q* lies on that circle. Similarly, *R* lies on the same circle, since it has the same diameter and ∠*ARH* is also a right angle. Note that ∠*QRH* forms an inscribed angle enclosing the chord *QH*, and ∠*QAH* does the same. As we saw above, this means the angles are equal: ∠*QRH* = ∠*QAH*.

Figure 12 shows the second step, involving ∆*ACP* and ∆BQC. In ∆*ACP*, since ∠*P* is a right angle, ∠*CAP* = ∠*QAH* is equal to (π/2) – ∠*C*. In ∆*BQC*, ∠*Q* is a right angle, so ∠*QBC* = ∠*HBP* is equal to (π/2) – ∠*C*; that is, ∠*QBP* is also the complement of ∠*C*. Thus ∠*QAH* = ∠*HBP*.

In step three, we'll play the same game as in step 1. Note from Figure 13 that points *P*, *H*, *R*, and *B* are concyclic, since they share the common diameter *BH*, and ∠*BRH* = ∠*HPB* = (π/2). And as before, note that ∠*HRP* = ∠*HBP*, since they both contain the common chord *HP*. Therefore ∠*HBP* = ∠*HRP*.

That's it! As promised, this chain of reasoning has led to ∠*QRH* = ∠*HRP*, which was our goal.

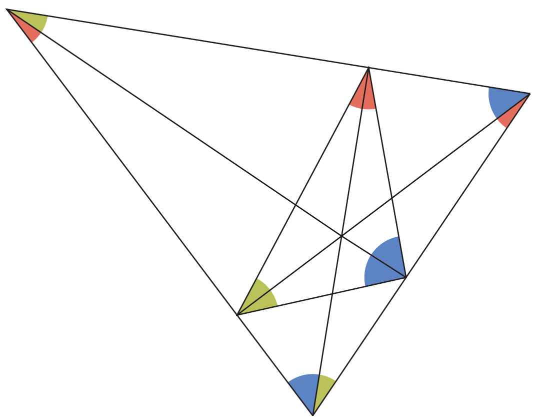
We have shown that if a billiard ball leaves *Q* and strikes *R* as we've constructed it, and then bounces just like light bounces off a mirror, the ball will travel to *P*. Since there was nothing special about our choice of *R*, the same reasoning holds for the other two points. Figure 14 shows the complete set of relationships.

Thus ∆*PQR* is a light triangle for ∆*ABC*—we have proven that such a triangle indeed exists.

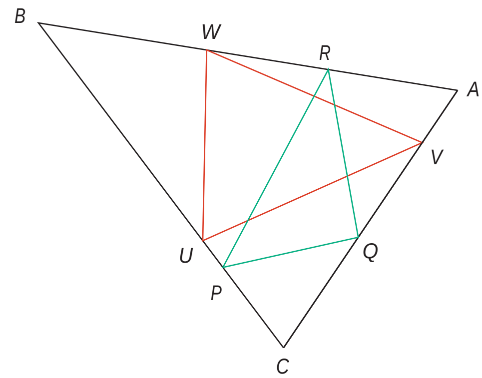
Proving that it's the fastest route for the ball is another matter, but reflection will be our ally again in that proof.

Smaller is better

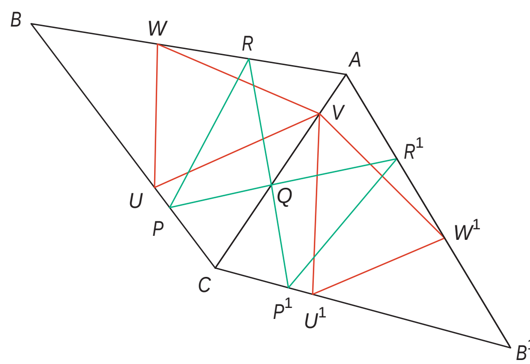
Now we want to show that ∆*PQR* is the smallest (or fastest) light triangle. Writing *p*(∆*PQR*) to denote the perimeter of a triangle, we want to show that



14 **The light triangle and its relationship to the triangle in which it's inscribed. All angles with the same color are equal.**



15 **Our original triangle ΔABC, the light triangle ΔPQR in green, and a possible shorter inscribed triangle ΔUVW in red.**



16 **Reflecting ΔABC around edge AC results in two reflected inscribed triangles.**

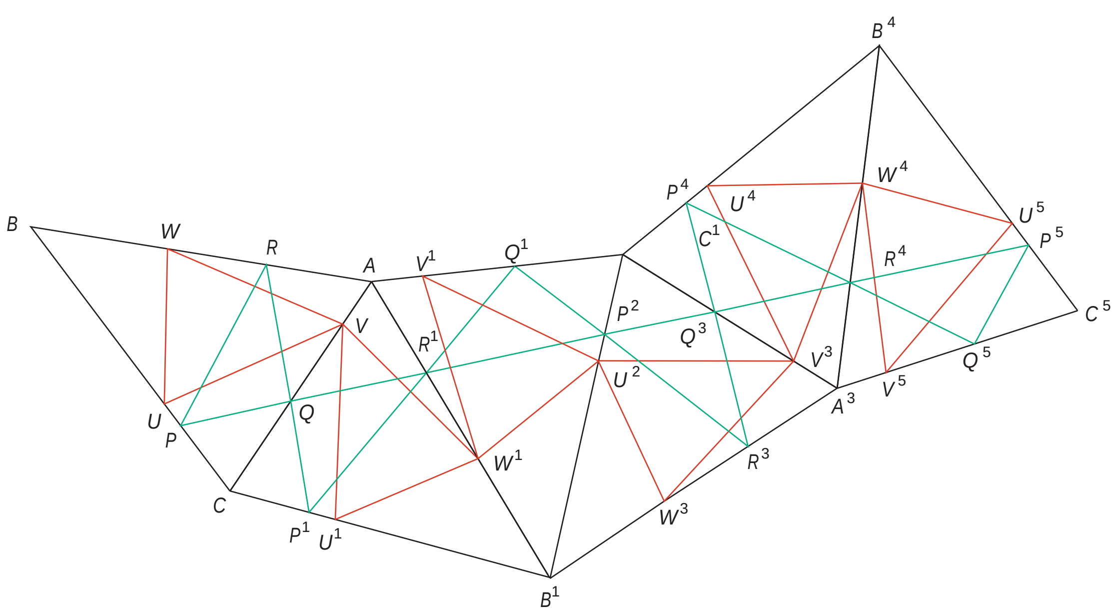
$p(\Delta PQR) \leq p(\Delta UVW)$ for any other light triangle ΔUVW. In fact, we can prove something even stronger: ΔPQR has the smallest perimeter of any triangle inscribed in ΔABC, whether it's a light triangle or not. This is actually even a little easier to prove, since we don't have to prove that our contender ΔUVW is a light triangle.

Let's begin with our original triangle ΔABC and our light triangle ΔPQR, as shown in Figure 15. Here we've also drawn some other triangle ΔUVW. We won't say anything specific about ΔUVW except that it's inscribed in ΔABC. We'll now see that the perimeter of ΔUVW will always be larger than the perimeter of ΔPQR, unless they're the same triangle.

First, we'll reflect ΔABC through side AC, as in Figure 16. Superscripts refer to the images of points after reflection. So A and C stay fixed, but B flips to B^1. Note that P, Q, and R^1 are collinear, as are R, Q, and P^1, because of equal angles at AC. In general, U, V, and W^1 will not be collinear, nor will W, V, or U^1.

Now we'll reflect the new triangle ΔAB^1C through edge B^1C, creating ΔAB^1C^2. Continuing, we'll reflect through B^1C^2, then C^2A^3, and finally through A^3B^4, giving us the chain shown in Figure 17.

We'll stop here after five reflections, because now edge B^4C^5 is parallel to its original position BC. To see this, consider that on the first reflection, edge BC is rotated around point C by a clockwise angle $2\angle C$. On the second reflection, it's rotated around point B^1 by the same



17 **Reflecting ΔABC five times about different edges creates the straight polyline (P, P^5), and B^4C^5 is parallel to BC. Polyline (U, U^5) will always be longer than (P, P^5).**

amount. The third reflection is in B^1C^2 itself, so that edge stays fixed. Then that edge gets rotated by $-2\angle C$ around $C2$ (that is, it's rotated counterclockwise), and then again by another $-2\angle C$ around B^4. Adding these up, we find $\angle C(2 + 2 + 0 - 2 - 2) = 0$, so the total angle of rotation is 0 and B^4C^5 is again parallel to BC.

Consider the polyline P,Q,R^1,P^2,Q^3,R^4,P^5, which we will denote as simply (P, P^5). This is equivalent to two copies of ΔPQR unfolded and straightened out. Each edge in ΔPQR is accounted for twice, so the length of this polyline is $|(P, P^5)| = 2p(\Delta PQR)$. Note that this polyline is in fact a straight line because of construction.

Now consider the polyline created by our potentially shorter triangle ΔUVW. This polyline is given by U,V,W^1,U^2,V^3,W^4,U^5. Again we find that we have the three sides of ΔUVW, each counted twice, so $|(U, U^5)| = 2p(\Delta UVW)$. Since BC and B^4C^5 are in the same orientation, U^5 is in the same relative position to P^5 as U is to P, so the straight line from U to U^5 is parallel to the straight line from P to P^5. Now we can ask if it's possible that $p(\Delta UVW) < p(\Delta PQR)$. Visually, the answer is clear from Figure 17: There's no path from U to U^5 shorter than the straight line from P to P^5. Writing $|P, P^5|$ for the length of the straight line from P to P^5, and similarly for $|U, U^5|$, we summarize this as

$$\left|\left(U,U^5\right)\right| = 2p\left(\Delta UVW\right) \geq \left|U,U^5\right| = \left|\left(P,P^5\right)\right| = 2p\left(\Delta QRP\right)$$

So the perimeter of ΔUVW will always be greater than that of ΔPQR, except when they're the same triangle (and then of course they have the same lengths).

We can see from Figure 17 that ΔPQR is the only triangle that can be reflected five times and have its pieces lie on a straight line from P to P^5. Any other triangle will necessarily have kinks in the path, and this will make it longer.

Thus we have proven that ΔPQR is a light triangle and that this light triangle has the smallest perimeter of any inscribed triangle.

More of everything

The next step is to think about generalizing this approach to polygons with obtuse angles, or more than three sides. It turns out things get rather more complicated. I'll summarize the results here; you can find proofs in *Geometry I* by M. Berger (Springer-Verlag, 1987).

Let's begin with obtuse angles. Suppose ΔABC has an obtuse angle at A (that is, $\angle A$ is more than 90 degrees). Clearly the other two angles must be acute. The minimum-perimeter inscribed polygon has one vertex at the foot of the altitude from A through BC, and the other two vertices at A itself.

This is a little unfortunate. First, this degenerate triangle hardly seems like a triangle at all. Secondly, it's not strictly inscribed, since at least one (and in fact two) of the vertices do not lie strictly on an edge of the original triangle. And finally it brings up the question of what happens when a ball strikes the vertex of a polygon—after all, there's no well-defined normal or tangent plane there, even though we fake one all the time in graphics to do smooth shading on polygons.

Mathematicians shrug off this problem of reflection at a vertex by basically defining it away—it happens infrequently enough that most feel it can be ignored when studying the larger problem of general billiard trajectories. After all, to hit the vertex exactly right, you have to hit that point exactly, which is likely to be pretty infrequent. Even if you think of real balls on real tables, rarely will a ball strike two sides of the table at precisely the same moment. Of course, that doesn't answer the question of what happens at such an occurrence, but it allows you to look the other way for a while without feeling that you've left a big hole in your theory. The usual fallback is just to say that in such a case the ball simply reverses direction at the time of the hit.

What if you have more than three sides to your polygon? A few things are true of billiard trajectories in general. Any convex polygon has a least-perimeter inscribed polygon, which may contain vertices of the original polygon. Any strictly inscribed least-perimeter polygon is a light polygon. Conversely, any light polygon has the minimum perimeter of any strictly inscribed polygon.

If your convex polygon has an odd number of sides, then you have a unique minimum-perimeter inscribed polygon, and if that polygon is strictly inscribed, it is a light polygon. So break out the heptagonal tables!

On the other hand, if your convex polygon has an even number of sides, and it has at least one light polygon, then it will have an infinite number of light polygons. There are some necessary and sufficient conditions that you can check to see if you have any (and thus an infinite number of) light polygons, but they're too complicated for us to cover here.

The billiard-ball problem is not completely solved in general. Suppose you're playing billiards on a circular or elliptical table, or one shaped like a lima bean or the letter R. Can you always find a starting point and trajectory such that the ball will return to the same point and trajectory? When such starting conditions exist, can you actually find them?

The simple geometry of perfect specular reflection turns out to be equivalent to one of the most basic theorems in plane geometry, which states that when two straight lines intersect, opposite angles are equal. That's the essential step in Figure 1 that completes the proof. Any time you apply this equal-angles theorem, you might want to think about that geometry and see if that reflection sheds more light on your topic. ■