1. Minimums, Maximums, Derivatives, and Computers

1.1 Introduction

This book has been written from the practical point of view of the engineer, and so you'll see few rigorous proofs on any of the pages that follow. As important as such proofs are in modern mathematics, I make no claims for rigor in this book (plausibility and/or direct computation are the themes here), and if absolute rigor is what you are after, well, you have the wrong book. Sorry!

Why, you may ask, are *engineers* interested in minimums? That question could be given a very long answer, but instead I'll limit myself to just two illustrations (one serious and one not, perhaps, quite as serious). Consider first the problem of how to construct a gadget that has a fairly short operational lifetime and which, during that lifetime, must perform flawlessly. Short lifetime and low failure probability are, as is often the case in engineering problems, potentially conflicting specifications: the first suggests using low-cost material(s) since the gadget doesn't last very long, but using cheap construction may result in an unacceptable failure rate. (An example from everyday life is the ordinary plastic trash bag—how thick should it be? The bag is soon thrown away, but we definitely will be unhappy if it fails too soon!) The trash bag engineer needs to calculate the minimum thickness that still gives acceptable performance.

For my second example, let me take you back to May 1961, to the morning the astronaut Alan Shepard lay on his back atop the rocket that would make him America's first man in space. He was very brave to be there, as previous unmanned launches of the same type of rocket had shown a disturbing tendency to explode into stupendous fireballs. When asked what he had been thinking just before blastoff, he replied "I was thinking that the whole damn thing had been built by the lowest bidder."

This book is a math history book, and the history of minimums starts centuries before the time of Christ. So, soon, I will be starting at the beginning of our story, thousands of years in the past. But before we climb into our time machine and travel back to those ancient days, there are a few modern technical issues I want to address first.

First, to write a book on minimums might seem to be a bit narrow; why not include maximums, too? Why not write a history of *extremas*, instead? Well, of course minimums and maximums are indeed certainly intimately connected, since a maximum of $y(x)$ is a minimum of $-y(x)$. To be honest, the reason for the book's title is simply that I couldn't think of one I could use with extrema as catchy as is "When Least Is Best." I did briefly toy with "When Extrema Are xxx" with the xxx replaced with *exotic*, *exciting,* and even (for a while, in a temporary fit of marketing madness that I hoped would attract Oprah's attention), *erotic*. Or even "Minimums Are from Venus, Maximums Are from Mars." But all of those (certainly the last one) are dumb, and so it stayed "When Least Is Best." There will be times, however, when I will discuss maximums, too. And now and then we'll use a computer as well.

For example, consider the problem of finding the maximum value of the rather benign-looking function

$$y(x) = 3\cos(4\pi x - 1.3) + 5\cos(2\pi x + 0.5).$$

Some students answer too quickly and declare the maximum value is 8, believing that for some value of x the individual maximums of the two cosine terms will add. That is not the case, however, since it is equivalent to saying that there is some $x = \hat{x}$ such that

$$4\pi\hat{x} - 1.3 = 2\pi n$$
$$2\pi\hat{x} + 0.5 = 2\pi k,$$

where n and k are integers. That is, those students are assuming there is an $\hat{x}$ such that

$$\hat{x} = \frac{2\pi n + 1.3}{4\pi} = \frac{2\pi k - 0.5}{2\pi}, \qquad n \text{ and } k \text{ integers.}$$

Thus,

$$2n\pi + 1.3 = 4\pi k - 1,$$

or

$$2.3 = 4\pi k - 2\pi n = 2\pi(2k - n),$$

or

$$\pi = \frac{2.3}{2(2k - n)} = \frac{23}{20(2k - n)}.$$

But if this is actually so, then as n and k are integers we would have π as the ratio of integers, i.e., π would be a rational number. Since 1761, however, π has been known to be irrational and so there are no integers n and k. And that means there is no $\hat{x}$ such that $y(\hat{x}) = 8$, and so $y_{\text{max}}(x) < 8$.

Well, then, what *is* $y_{\text{max}}(x)$? Is it perhaps *close* to 8? You might try setting the derivative of $y(x)$ to zero to find $\hat{x}$, but that quickly leads to a mess. (Try it.) The best approach, I think, is to just numerically study $y(x)$ and watch what it does. The result is that $y_{\text{max}}(x) = 5.7811$, significantly less than 8. My point in showing you this is twofold. First, a computer is often quite useful in minimum studies (and we will use computers a lot in this book). Second, taking the derivative of something and setting it equal to zero is *not* always what you have to do when finding the extrema of a function.

An amusing (and perhaps, for people who like to camp, even useful) example of this is provided by the following little puzzle. Imagine that you have been driving for a long time along a straight road that borders an immense, densely wooded area. It looks enticing, and so you park your car on the side of the road and hike into the woods for a mile along a straight line perpendicular to the road. The woods are very dense (you instantly lose sight of the road when you are just one step into the woods), and after a mile you are exhausted.

You call it a day and camp overnight. When you get up the next morning, however, you've completely lost your bearings and don't know which direction to go to get back to your car. You could, if you panic, wander around in the woods indefinitely! But there *is* a way to travel that absolutely guarantees that you will arrive back at your car's *precise location* after walking a certain maximum distance (it might take even less). How do you walk out of the woods, and what is the maximum distance you would have to walk? The answer requires only simple geometry—if you are stumped the answer is at the end of this chapter.

1.2 When Derivatives Don't Work

Here's another example of a minimization problem for which calculus is not only *not* required, but in fact seems not to be able to solve. Suppose we have the real line before us (labeled as the x-axis), stretching from $-\infty$ to $+\infty$. On this line there are marked n points, labeled in increasing value as $x_1 < x_2 < \cdots < x_n$. Let's assume all the x_i are finite (in particular x_1 and x_n), and so the interval of the x-axis that contains all n points is finite in length. Now, somewhere (anywhere) on the finite x-axis we mark one more point (let's call it x). We wish to pick x so that the sum of the distances between x and all of the original points is minimized. That is, we wish to pick x so that

$$S = |x - x_1| + |x - x_2| + \cdots + |x - x_n|$$

is minimized. I've used absolute-value signs on each term to insure each *distance* is non-negative, independent of where x is, either to the left or to the right of a given x_i. Those absolute-value signs may seem to badly complicate matters, but that's not so. Here's why.

First, focus your attention on the two points that mark the ends of the interval, x_1 and x_n. The sum of the distances between x and x_1, and between x and x_n, is

$$|x - x_1| + |x - x_n|$$

and this is *at least* $|x_1 - x_n|$. If $x > x_n$, or if $x < x_1$ (i.e., if x is outside the interval), then strict *in*equality holds, but if x is *anywhere* inside the interval (i.e., $x_1 \le x \le x_n$) then equality holds. Thus, the minimum value of $|x-x_1|+|x-x_n|$ is achieved by placing x anywhere between x_1 and x_n.

Next, shift your attention to the two points x_2 and x_{n-1}. We can repeat the above argument, without modification, to conclude that the minimum value of $|x - x_2| + |x - x_{n-1}|$ is achieved when x is *anywhere* between x_2 and x_{n-1}. Note that this automatically satisfies the condition for minimizing the value of $|x - x_1| + |x - x_n|$, i.e., placing x anywhere between x_2 and x_{n-1} minimizes $|x - x_1| + |x - x_2| + |x - x_{n-1}| + |x - x_n|$. You can now see that we can repeat this line of reasoning, over and over, to conclude

$|x - x_3| + |x - x_{n-2}|$ is minimized by placing x *anywhere* between x_3 and x_{n-2},

$|x - x_4| + |x - x_{n-3}|$ is minimized by placing x *anywhere* between x_4 and x_{n-3},

$\vdots$

and finally, if we suppose that n is an *even* number of points, then

$|x - x_{\frac{n}{2}}| + |x - x_{\frac{n}{2}+1}|$ is minimized by placing x *anywhere* between $x_{\frac{n}{2}}$ and $x_{\frac{n}{2}+1}$.

So, we simultaneously satisfy all of these individual minimizations by placing x anywhere between $x_{n/2}$ and $x_{(n/2)+1}$ (*if* n is even), and this of course minimizes S.

But what if n is odd? Then the same reasoning as for even n still works, *until* the final step; then there is no second point to pair with $x_{(n+1)/2}$. Thus, simply let $x = x_{(n+1)/2}$ and so $|x - x_{(n+1)/2}| = 0$, which is certainly the minimum value for a *distance*. Thus, we have the somewhat unexpected, noncalculus solution that, for n even, S is minimized by placing x *anywhere* in an *interval*, but for n odd there is just *one*, *unique* value for x (the *middle* x_i) that minimizes S.

1.3 Using Algebra to Find Minimums

As another elementary but certainly not a trivial example of the claim that derivatives are not always what you want to calculate, consider the fact that ancient mathematicians knew that of all rectangles with a given perimeter it is the square that has the largest area. (This is a special result from a general class of maximum/minimum questions of great historical interest and practical value called *isoperimetric problems*, and I'll have more to say about them in the next chapter.) Ask most modern students to show this and you will almost surely get back something like the following. Define P to be the given perimeter of a rectangle, with x denoting one of the two side lengths. The other side length is then $(P - 2x)/2$, and so the area of the rectangle is

$$A(x) = x\left(\frac{P-2x}{2}\right) = \frac{1}{2}\,Px - x^2.$$

$A(x)$ is maximized by setting $dA/dx = \frac{1}{2}P - 2x$ equal to zero, and so $x = \frac{1}{4}P$, which completes the proof. Using only algebra, however, an ancient mathematician could have argued that

$$\begin{aligned} A = x\left(\frac{P-2x}{2}\right) &= \frac{1}{2}\,Px - x^2 = \frac{P^2}{16} - \frac{P^2}{16} + \frac{1}{2}\,Px - x^2 \\ &= \frac{P^2}{16} - \left(x^2 - \frac{1}{2}\,Px + \frac{P^2}{16}\right) = \frac{P^2}{16} - \left(x - \frac{P}{4}\right)^2 \le \frac{P^2}{16} \end{aligned}$$

since $(x - (P/4))^2 \ge 0$ for all x. That is, A is never larger than the constant $P^2/16$ and is equal to $P^2/16$ if and only if (a useful phrase I will henceforth write as simply iff) $x = P/4$, which completes the ancient, noncalculus proof.

As a final comment on this result, which again illustrates the intimate connection between minimum and maximum problems, we can restate matters as follows: of all rectangles with a given area, the square has the smallest perimeter. This is the so-called *dual* of our original problem and, indeed, all isoperimetric problems come in such pairs. I'll prove this particular dual in section 1.5. Another useful isoperimetric result that seems much like the one just established—one also known to the precalculus, ancient mathematicians—is not so easy to prove: of all the triangles with the same area, the equilateral has the smallest perimeter. See if you can show this (or its dual) before I do it later in this chapter.

We can use the previous result—of all rectangles with a fixed perimeter, the square has the maximum area—to solve *without* calculus a somewhat more complicated appearing problem found in all calculus textbooks. Suppose we wish to enclose a rectangular plot of land with a fixed length of fencing, with the *side of a barn* forming one side of the enclosure. How should the fencing now be used? We could, of course, use calculus as follows: let x be the length of each of the two sides perpendicular to the barn wall, and $\ell - 2x$ be the length of the side parallel to the barn wall (ℓ is the fixed, total length of the fencing). Then the enclosed area is

$$A = x(\ell - 2x) = x\ell - 2x^2$$

and so

$$\frac{dA}{dx} = \ell - 4x,$$

which, when set equal to zero, gives $x = \frac{1}{4}\ell$. Thus, $\ell - 2x = \frac{1}{2}\ell$, which says the enclosed area is maximized when it is twice as long as it is wide. But this solution is far more sophisticated than required. Simply imagine that we enclose another rectangular area *on the other side* of the barn wall. We already know that, together, the two rectangular plots should form a square, and so each of the two rectangular plots are *half* of the square, i.e., twice as long in one dimension as in the other.

Our ancient mathematician's trick of completing the square is a very old one, and some historians claim that it can be found implicit in Euclid's *Elements* (Book 6, Proposition 27), circa 300 B.C. There, the problem discussed is equivalent to that of dividing a constant into two parts so that their product is maximum. So, if the constant is C, then the two parts are x and $C - x$, with the product

$$\begin{aligned} M &= x(C - x) = Cx - x^2 = -(x^2 - Cx) \\ &= -\left(x^2 - Cx + \frac{C^2}{4} - \frac{C^2}{4}\right) = -\left(x - \frac{C}{2}\right)^2 + \frac{C^2}{4}. \end{aligned}$$

Thus, as $(x - (C/2))^2 \geq 0$ for all x, then M is never larger than $C^2/4$ and is equal to $C^2/4$ iff $x = C/2$.

Stated this way, Euclid's problem surely seems rather abstract, but in 1573 the Dutch mathematical physicist Christiaan Huygens gave a nice physical setting to the calculation. Suppose we have a line and

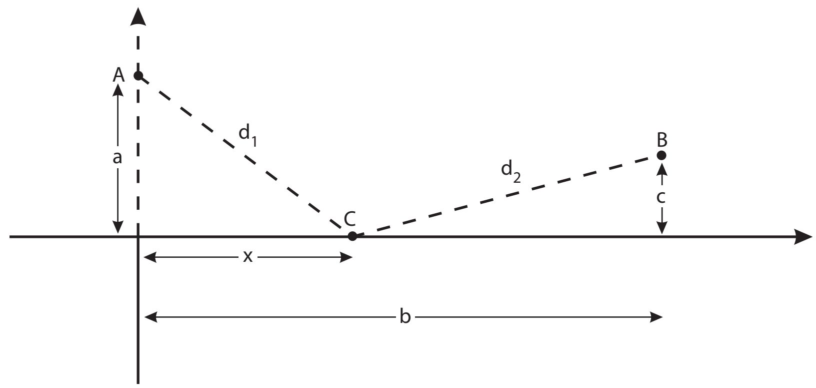

FIGURE 1.1. Huygen's problem.

two points (A and B) *not* on the line. Where should the point C be located *on* the line so that the sum of the squares of the distances from C to A and C to B, $(AC)^2 + (BC)^2$, is minimum? With no loss in generality we can draw the geometry of this problem as shown in figure 1.1, with A on the y-axis. The figure shows A and B on the same side of the line, and places C between A and B, but as the analysis continues you'll see that these assumptions in no way affect the result.

In the notation of the figure we are to find the value of x that, with a, b, and c constants, minimizes $d_1^2 + d_2^2$. Now,

$$\begin{aligned} d_1^2 + d_2^2 &= \left\{x^2 + a^2\right\} + \left\{(b - x)^2 + c^2\right\} \\ &= a^2 + b^2 + c^2 + 2x\,(x - b)\,. \end{aligned}$$

Thus, we need to minimize the product $x(x - b)$; but we already know from Euclid how to do that—set $x = \frac{1}{2}b$. That is, C is midway between A and B. If you redraw figure 1.1 so that either $x > b$ or $x < 0$, and then write the expression for $d_1^2 + d_2^2$, you'll see that the result is unchanged.

An elementary example of an extremal problem in which there is (by the very nature of the problem) nothing to differentiate comes from discrete probability theory. Then the independent variable does not vary continuously but, rather, in discontinuous jumps. In such cases, taking a derivative simply has no meaning. So, suppose we toss four fair die, i.e., each one of the six faces on each die has probability $\frac{1}{6}$ of showing. What is the most likely number of die that will show a 3? The answer can only be one of five numbers, of course, the integers zero through four. If we define the value of the random variable X as the number of die that show a 3, then elementary probability theory tells us that $P(X = k)$ = probability that $X = k$ is given by

$$P(X = k) = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k},$$

where n is the number of die and $\binom{n}{k} = n!/(k!(n-k)!)$. So, with $n = 4$,

$$P(X = 0) = \frac{625}{1296}$$
$$P(X = 1) = \frac{500}{1296}$$
$$P(X = 2) = \frac{150}{1296}$$
$$P(X = 3) = \frac{20}{1296}$$
$$P(X = 4) = \frac{1}{1296}.$$

Thus, the most likely number of 3's to show is zero. But even more likely to happen is that *at least one* 3 shows, as

$$P(X \geq 1) = \sum_{k=1}^{4} P(X = k) = \frac{671}{1296} > P(X = 0).$$

This strikes many as a paradoxical result, but that is part of the inexhaustible charm of probability!

1.4 A Civil Engineering Problem

As a more sophisticated example of how minimization problems can sometimes be attacked with noncalculus approaches, consider the following. We have two towns, A and B, on opposite sides of a river

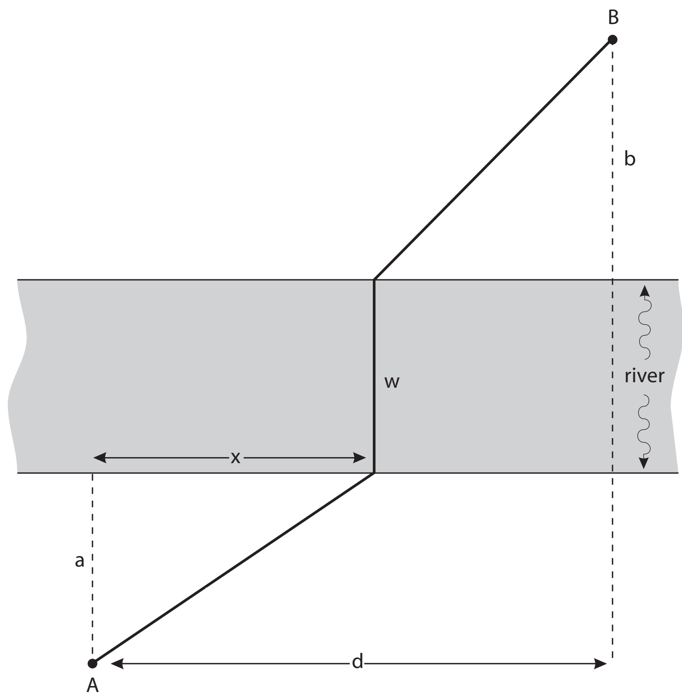

FIGURE 1.2. Minimum-distance bridge placement problem.

with constant width w. As shown in figure 1.2, A is distance a from the river, B is distance b, and the lateral separation of the two towns is d. Our problem is to determine where we should build a bridge over the river (perpendicular to the river's banks) so as to make the journey between A and B as short as possible. That is, what is x?

With calculus, this question is not hard to answer. We simply write the total distance as

$$T = \sqrt{a^2 + x^2} + w + \sqrt{b^2 + (d - x)^2}$$

and then set $dT/dx = 0$. Thus,

$$\frac{dT}{dx} = \frac{1}{2}\left[\frac{2x}{\sqrt{a^2 + x^2}} - \frac{2(d - x)}{\sqrt{b^2 + (d - x)^2}}\right]$$

and setting this equal to zero gives

$$x = \frac{ad}{a+b}.$$

Ancient mathematicians could also have solved this problem, however, long before the invention of the calculus, using just elementary geometry. To see how, let me first make a fundamental, exceedingly important and useful mathematical observation called the *triangle inequality*. The triangle inequality asserts that, given any triangle, the sum of any two of its sides is at least as large as the third side. It is really just a statement of the fact that the shortest path connecting two points in a plane is the straight line passing through the two points. Thinking of the triangle's sides as *directed* line segments with both magnitude and direction (i.e., as vectors), we can write $\vec{u}$ and $\vec{v}$ as two of the sides and $\vec{u}+\vec{v}$ as the third side, as shown in figure 1.3.

The triangle inequality says that $|\vec{u}|+|\vec{v}| \geq |\vec{u}+\vec{v}|$, where the absolute value signs denote the length of the vector. It is obvious that the inequality becomes an equality iff $\vec{u}$ and $\vec{v}$ point in the same direction (and so the triangle collapses to the "trivial triangle" with zero area). We can, in fact, now drop the imagery of the triangle itself, and simply think of $\vec{u}$ and $\vec{v}$ as *any* two vectors not necessarily associated with a triangle (although in many problems there will be a triangle).

Now, redraw figure 1.2 as figure 1.4 and label the various path segments as vectors. Notice that no matter what $\vec{x}$ is, the sum $(\vec{a}+\vec{x})+(\vec{d}-\vec{x}+\vec{b})$ is constant. Mathematically this is trivial (the two $\vec{x}$'s

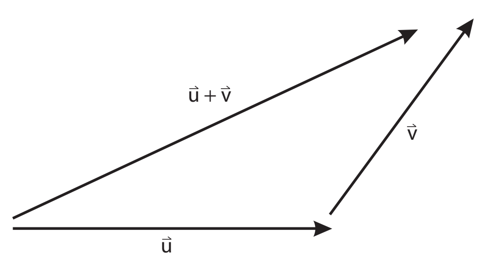

FIGURE 1.3. Vector addition.

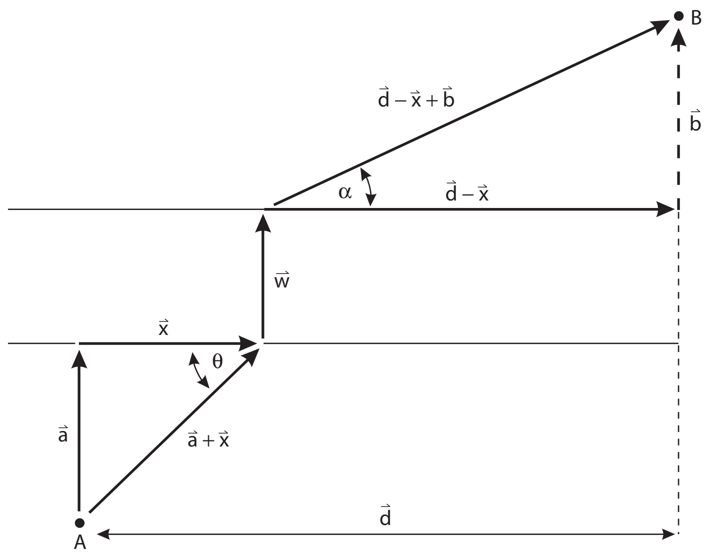

FIGURE 1.4. Bridge geometry in vector notation.

cancel), but physically this is because of the important observation that every vector sum (plus a constant $\vec{w}$ term to account for the bridge) starts at A and ends at B, no matter what $\vec{x}$ may be. By the triangle inequality $|\vec{a}+\vec{x}|+|\vec{d}-\vec{x}+\vec{b}| \geq |\vec{a}+\vec{d}+\vec{b}|$; an equality (which is the minimum sum) is achieved only when $\vec{a}+\vec{x}$ and $\vec{d}-\vec{x}+\vec{b}$ are in the same direction. That is, when $\theta = \alpha$ in the notation of figure 1.4.

Since the two triangles in figure 1.4 are right triangles with their other two angles equal, they are similar triangles. Thus, dropping the vector notation, we have

$$\frac{a}{x} = \frac{b}{d-x},$$

which is easily solved to give the location of the bridge at

$$x = \frac{ad}{a+b},$$

just as before. But this time no derivative was required. And, in fact, our ancient mathematician's solution actually provides some immediate extra physical insight that the calculus one does not; since $\theta = \alpha$, the path segments connecting each town to its respective river bank are *parallel*.

1.5 The AM-GM Inequality

There are yet other methods the mathematicians of old, in the days before calculus, could have used to solve many problems that seemingly require the calculation of derivatives. One of the most elegant of these methods is what is called the AM-GM inequality (the arithmetic mean-geometric mean inequality). It is easy to state:

If $x_1, x_2, \cdots, x_n$ are any n positive numbers, $n \geq 1$, and
if $A = (1/n)\,(x_1 + x_2 + \cdots + x_n)$ is the arithmetic mean of the x's
and if $G = (x_1 x_2 \cdots x_n)^{1/n}$ is the geometric mean of the x's,
then $A \geq G$ with equality iff $x_1 = x_2 = \cdots = x_n$.

New demonstrations of this famous and remarkably useful inequality appear on a regular basis to this day, but one of the easiest to understand (as well as one of the most elegant) is the 1954 proof by a mathematician named G. Ehlers. I know nothing more about Ehlers, but his proof of the AM-GM inequality is a gem and you can find it in appendix A. That proof uses just simple algebra and induction, but *no calculus*, which is appropriate since the whole point here is to show how we can solve many minimum/maximum problems without the techniques of calculus.

For example, recall the isoperimetric dual problem mentioned at the start of section 1.3: show that of all rectangles with a given area it is the square that has the smallest perimeter. This is actually quite easy to demonstrate with the AM-GM inequality. If we call the sides of the rectangle x and y, then the problem is to determine x and y so that we minimize

$$P = 2x + 2y = 2(x + y),$$

given that

$$A = xy$$

is a constant. From the AM-GM inequality with $n = 2$ we immediately have

$$\frac{1}{2}(x + y) \geq \sqrt{xy} = \sqrt{A}$$

with equality iff $x = y$. That is,

$$P = 2(x + y) \geq 4\sqrt{A},$$

which says P is never smaller than the *constant* $4\sqrt{A}$ and is equal to that constant iff $x = y$ (iff the rectangle is a square).

Closely related to this result is one concerning right triangles. Imagine all possible right triangles with perpendicular sides of lengths x and y that sum to a constant, i.e.,

$$x + y = k.$$

If we write A to denote the areas of the triangles, then

$$A = \frac{1}{2}\,xy.$$

Now, the AM-GM inequality for $n = 2$ says

$$\frac{x + y}{2} \geq \sqrt{xy} = \sqrt{2A}$$

with equality iff $x = y$. Thus,

$$\frac{k}{2} \geq \sqrt{2A},$$

or

$$A \leq \frac{k^2}{8}$$

with equality iff $x = y$. This shows that of all right triangles with perpendicular sides that sum to a constant, it is the *isosceles* right triangle that has the largest area (a result known since ancient times).

For another elegant illustration of the power of the AM-GM inequality, think back a bit to a question I asked you to ponder: of all triangles with a given area, show that it is the equilateral that has the smallest perimeter. Did you have any success doing that? It's not trivial! I'll do it here with the aid of the AM-GM inequality by showing the dual theorem: of all triangles with a given perimeter P, the equilateral has the largest area. As a prelude, recall the amazing formula for the area A of any triangle in terms of the lengths of its sides (a, b, and c). This formula is named after the Egyptian mathematician Heron of Alexandria, who is thought to have lived in the first century A.D. Some historians have speculated that the formula was known by Archimedes three centuries earlier, but there is no real evidence of that (other than Archimedes' genius, which makes it probable that he *did* know it), while the formula does appear in Heron's *Metrica*. It is not an easy formula to derive [see William Dunham, *Journey through Genius: The Great Theorems of Mathematics* (John Wiley 1990, pp. 118–27)], but it is easy to state:

$$A = \sqrt{s(s-a)(s-b)(s-c)},$$

where $s = \frac{1}{2}(a+b+c) = \frac{1}{2}P$, the so-called semiperimeter of the triangle. Since P is given, then so is s and Heron's formula tells us that to maximize A we must maximize the product $(s-a)(s-b)(s-c)$.

Notice first that each of the factors in that product is indeed positive, e.g.,

$$s - a = \frac{a+b+c}{2} - a = \frac{-a+b+c}{2} > 0$$

because from the triangle inequality for nontrivial triangles (triangles with nonzero area) we have $b + c > a$. Now, from the AM-GM inequality, we have

$$\frac{(s-a)+(s-b)+(s-c)}{3} = \frac{3s-(a+b+c)}{3} = \frac{3s-2s}{3}$$
$$= \frac{s}{3} \geq [(s-a)(s-b)(s-c)]^{1/3}$$

with equality iff $(s-a) = (s-b) = (s-c)$, i.e., iff $a = b = c$. The term $s/3$ is a *constant upper*-bound to the inequality and so the area is maximized if $a = b = c$, and that's the entire proof!

As a third example of the AM-GM inequality solving a problem ordinarily thought to require calculus, consider the following question that probably appears in every calculus textbook ever written. A food can (with both ends sealed, of course) with the given volume V is to have the shape of a right circular cylinder. What are the dimensions of the can (the radius r and the height h) so that the surface area is minimum? The "calculus way" to answer this is to write the surface area S and the volume as

$$S = 2\pi r^2 + 2\pi rh$$
$$V = \pi r^2 h$$

and then to eliminate h. Thus, $h = V/\pi r^2$, and so

$$S = 2\pi r^2 + 2\pi r \frac{V}{\pi r^2} = 2\pi r^2 + \frac{2V}{r}.$$

We minimize S (as we'll see in chapter 4) by setting dS/dr to zero, i.e.,

$$\frac{dS}{dr} = 4\pi r - \frac{2V}{r^2} = 0,$$

which gives the solution for r. Thus, $V = 2\pi r^3$, or

$$\frac{V}{\pi r^2} = h = 2r.$$

That is, the height of the can with minimum surface area is equal to the diameter of the can.

Here's how the AM-GM inequality answers the same question. As before,

$$S = 2\pi\left(r^2 + rh\right) = 2\pi\left(r^2 + \frac{V}{\pi r}\right) = 2\pi\left(r^2 + \frac{V}{2\pi r} + \frac{V}{2\pi r}\right).$$

Or

$$\frac{S}{6\pi} = \frac{1}{3}\left(r^2 + \frac{V}{2\pi r} + \frac{V}{2\pi r}\right).$$

From the AM-GM inequality, we have

$$\frac{1}{3}\left(r^2 + \frac{V}{2\pi r} + \frac{V}{2\pi r}\right) \geq \left(r^2 \cdot \frac{V}{2\pi r} \cdot \frac{V}{2\pi r}\right)^{1/3} = \left(\frac{V^2}{4\pi^2}\right)^{1/3},$$

and so

$$\frac{S}{6\pi} \geq \left(\frac{V^2}{4\pi^2}\right)^{1/3} \quad \text{or} \quad S \geq 6\pi\left(\frac{V^2}{4\pi^2}\right)^{1/3}.$$

Thus, the surface area is never less than the *constant* $6\pi\left(\frac{V^2}{4\pi^2}\right)^{1/3}$, and is equal to that minimum value when $r^2 = \frac{V}{2\pi r} = \frac{V}{2\pi r}$, i.e., when $V = 2\pi r^3$ just as we found before (but before we had to know how to calculate a derivative).

Now, here's a little variation for you to play with: in the example just done, both ends of the can were sealed. Suppose instead that only the bottom end is sealed. For the same volume as before, what now is the relationship between r and h to minimize the surface area, and what is the ratio of the new minimized surface area to the one just calculated? It should be obvious that the ratio is less than one, but *how much* less than one? Remember, *no calculus*! There are *two* ways for you to attack this problem. You can start over and use the AM-GM inequality, of course. More clever, however, is to use our previous result, by noticing that if we take two cans, each with only one end sealed, and butt the unsealed ends together, we get a can with *both* ends sealed! Either way, you should get the same answers. (The answers are at the end of this section.)

We can use the AM-GM inequality to prove the following curious, and I think unobvious, fact: given two food cans of equal volume *and equal height*, one cylindrical and the other rectangular in shape, the cylindrical can will *always* have the smaller total surface area. To see this, observe that if V is the common volume, then, for either shape, we can write

$$V = (\text{area of bottom}) \times (\text{height}).$$

So, since the heights are also equal, then the areas of the bottoms (and tops) of the two shapes are equal, too. Thus, to decide which can shape has the smaller total surface area we need only to compare the vertical surface areas. To do that, let's make the following definitions:

S_c = vertical surface area of a cylindrical can of radius r and height h, i.e., $S_c = 2\pi rh$,

S_r = vertical surface area of a rectangular can with dimensions $a \times b \times h$, i.e., $S_r = 2ha + 2hb = 2h(a + b)$.

This means

$$S_r - S_c = 2h(a+b) - 2\pi rh = 2h[(a+b) - \pi r].$$

From the AM-GM inequality we have $(a+b) \geq 2\sqrt{ab}$, and so

$$S_r - S_c \geq 2h\left[2\sqrt{ab} - \pi r\right]$$

because I've replaced $(a+b)$ with a smaller quantity. Now, since the volumes of the two cans are equal we can also write

$$V = \pi r^2 h = abh,$$

and so

$$\sqrt{ab} = \sqrt{\frac{V}{h}}$$

and

$$\pi r = \pi\sqrt{\frac{V}{\pi h}} = \sqrt{\pi}\sqrt{\frac{V}{h}}.$$

This gives us

$$S_r - S_c \geq 2h\left[2\sqrt{\frac{V}{h}} - \sqrt{\pi}\sqrt{\frac{V}{h}}\right] = 2h\sqrt{\frac{V}{h}}\left[2 - \sqrt{\pi}\right] > 0$$

because it is clear that $2 > \sqrt{\pi}$ (i.e., $4 > \pi$). So, no matter how you choose the various dimensions of the two cans, if they have equal volume *and* equal height then the cylindrical can will *always* have the smaller total surface area.

If we don't require the two can shapes to have the same height, then it is no longer true that the cylindrical can will have the smaller surface area no matter what the dimensions may be. For example, suppose the rectangular can has dimensions $1 \times 1 \times \pi$, for a volume of π. Its total surface area is then $2 + 4\pi = 14.57$. If the cylindrical can has a radius of r and height h, then for the same volume we have $\pi r^2 h = \pi$, or

$$h = \frac{1}{r^2}.$$

Its total surface area is

$$T = 2\pi r^2 + 2\pi rh = 2\pi r^2 + 2\pi r\frac{1}{r^2}$$
$$= 2\pi\left(r^2 + \frac{1}{r}\right).$$

It is clear that we could pick r to make T arbitrarily larger than $2+4\pi$.

But it is also true that, if we pick r to give the minimum surface area for the cylindrical can, that area *will* be smaller than $2 + 4\pi$. That is, differentiating T gives

$$\frac{dT}{dr} = 2\pi\left(2r - \frac{1}{r^2}\right)$$

which is zero when $r = \left(\frac{1}{2}\right)^{1/3}$, which gives

$$T = 2\pi\left[\left(\frac{1}{2}\right)^{2/3} + \frac{1}{\left(\frac{1}{2}\right)^{1/3}}\right] = 2\pi\frac{\frac{1}{2}+1}{\left(\frac{1}{2}\right)^{1/3}} = 2\pi 2^{1/3}\cdot\frac{3}{2}$$
$$= 3\pi 2^{1/3} = 11.87,$$

nearly 19% less than the surface area of the rectangular can.

As the final example of this section, let me show you how mathematicians of old could have solved yet another maximum problem. As shown in appendix B, using nothing but algebra (no calculus), a consequence of the AM-GM inequality is yet another inequality called the arithmetic mean-quadratic mean inequality (the AM-QM inequality): if $x_1, x_2, \cdots, x_n$ are n numbers, then

$$\frac{x_1 + x_2 + \cdots + n_n}{n} \leq \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}, \qquad n \geq 1$$

with equality iff $x_1 = x_2 = \cdots = x_n$. But the AM-GM inequality itself tells us that

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}$$

with equality iff $x_1 = x_2 = \cdots = x_n$, and so

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

with equality iff $x_1 = x_2 = \cdots = x_n$.

This general result has a very pretty geometric interpretation for $n = 2$, i.e., for

$$\sqrt{x_1 x_2} \leq \sqrt{\frac{x_1^2 + x_2^2}{2}}.$$

Suppose that $x_1^2 + x_2^2 = R^2$ (a constant). The equation $x_1^2 + x_2^2 = R^2$ is a circle (centered on the origin of the x_1, x_2 coordinate system) with radius R, and so $\sqrt{x_1 x_2}$ is bounded from above by the constant $R/\sqrt{2}$. And since $4x_1x_2$ is the area of a rectangle inscribed in that circle, then that area is bounded from above by the constant $2R^2$ and that area is equal to $2R^2$ iff $x_1 = x_2$. That is, the inscribed rectangle of maximum area is the inscribed square.

The answers to the problem of the cylindrical can with minimum surface area, with just one end sealed, are

a. $r = h$

b. ratio of surface areas $= \frac{1}{2}\sqrt[3]{4} = 0.7937$.

1.6 Derivatives from Physics

There are minimum/maximum problems of great interest that *do* contain derivatives, but *not* because we are going to set them equal to zero. They are present because, for example, the *physics* of the problem requires them. The actual determination of a minimum (or a maximum) of something in such problems, however, depends on other sorts of arguments. So, for the penultimate section of this introductory chapter, let me take you through the details of one such problem that has derivatives aplenty because of the *physics* and not because of the mathematics.

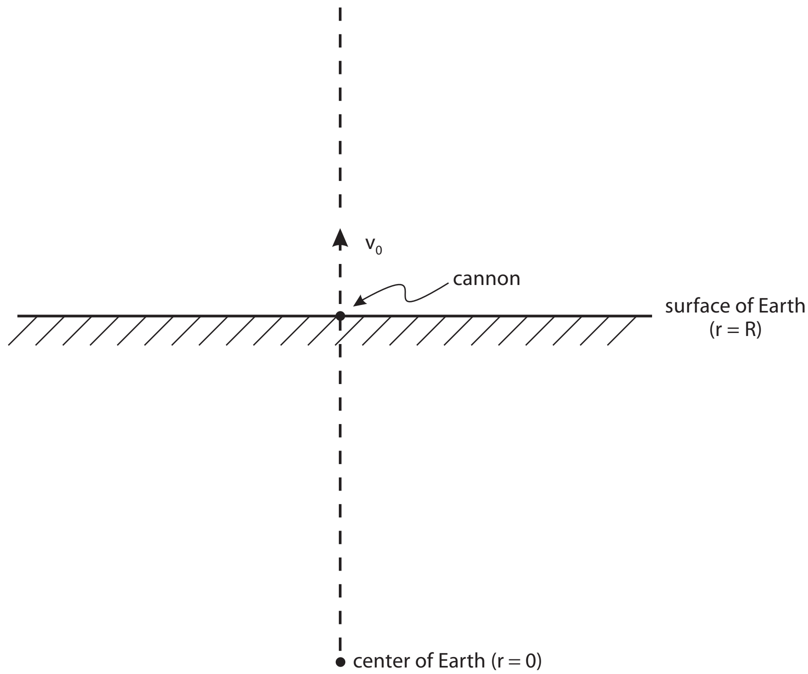

FIGURE 1.5. Vertical cannon shot.

Consider figure 1.5. There we have a cannon pointing straight up, directly away from the center of the earth (not drawn to scale!). If we fire the cannon a shell is ejected with initial velocity v_0, it rises upward to some maximum height, stops, and then falls back down to the ground. It is clear that the larger v_0, the higher the shell will go before gravity brings its upward motion to a halt. We can show, in fact, that if v_0 has a certain critical minimum value, then the shell will *not* return to earth. That minimum value for v_0 is called the *escape* velocity.

If we measure distance from the center of the earth as r ($r = 0$ *is* the center, and $r = R$ is the surface of the earth), then Newton's second law of motion (force equals mass times acceleration) and his inverse-square law of gravity tells us that if we ignore air-drag on the shell, then

$$m\frac{d^2r}{dt^2} = -G\frac{Mm}{r^2}, \qquad r \geq R,$$

where: m = mass of the shell,
M = mass of the earth,
G = universal constant of gravitation.

The minus sign on the right side of the differential equation is present because increasing r is directed upward, while the gravitational force on the shell is in the opposite direction, downward toward the center of the earth.

We can solve this second-order differential equation with the help of a powerful result from differential calculus called the chain rule (discussed in chapter 4): if we write $v(r)$ as the velocity of the shell at distance r from the center of the earth, then by definition

$$v = \frac{dr}{dt},$$

and so the acceleration of the shell is

$$\frac{d^2r}{dt^2} = \frac{dv}{dt} = \frac{dr}{dt} \cdot \frac{dv}{dr} = v\frac{dv}{dr}.$$

This reduces our original differential equation to the more tractable (with m canceled on both sides) equation

$$v\frac{dv}{dr} = -GM\,\frac{1}{r^2}, \qquad r \geq R.$$

We can "separate the variables" in this equation and write

$$v\,dv = -GM\,\frac{dr}{r^2},$$

which is easily integrated to give

$$\frac{1}{2}\,v^2 = GM\,\frac{1}{r} + C,$$

where C is the so-called "constant of indefinite integration." Now, since $v = v_0$ when $r = R$, then

$$\frac{1}{2}\,v_0^2 = GM\,\frac{1}{R} + C,$$

or

$$C = \frac{1}{2}\,v_0^2 - GM\,\frac{1}{R},$$

and thus

$$\frac{1}{2}\,v^2 = GM\,\frac{1}{r} + \frac{1}{2}\,v_0^2 - GM\,\frac{1}{R}.$$

If we define H as the shell's maximum distance from the center of the earth, then, as by definition $v = 0$ when $r = H$, we have

$$0 = \frac{GM}{H} + \frac{1}{2}\,v_0^2 - \frac{GM}{R},$$

or

$$H = \frac{GM}{\dfrac{GM}{R} - \dfrac{1}{2}\,v_0^2}.$$

If $v_0 = 0$ then $H = R$, which is simply the obvious; if the shell "leaves" the cannon with zero initial velocity, then it doesn't go anywhere! But as v_0 increases from zero, then H increases from R and, obviously, as $\frac{1}{2}v_0^2$ approaches GM/R we see that H diverges to infinity, i.e., the shell does not return to earth. So, the minimum escape velocity is the initial velocity given by

$$v_0 = \sqrt{\frac{2GM}{R}}.$$

Any velocity greater than this also means the shell isn't coming back, of course.

We can express this result in the following interesting alternative way. When $r = R$, the gravitational force on the shell is simply what we call its weight at the surface of the earth, which is mg, where g is the acceleration of gravity *at the surface*. Thus,

$$mg = G\,\frac{Mm}{R^2},$$

and so $GM = g\,R^2$. This gives the escape velocity as

$$v_0 = \sqrt{\frac{2gR^2}{R}} = \sqrt{2gR}.$$

Taking the earth's radius as 3,950 miles, and g as 32.2 ft/sec^2, we have the escape velocity as

$$v_0 = \sqrt{2 \times 32.2 \times 3{,}950 \times 5{,}280}\ \text{ft/sec}$$
$$= 36{,}649\ \text{ft/sec} = 6.94\ \text{miles/sec}.$$

This is not the way we send people into space, of course, as the initial acceleration of the shell (spaceship) from zero to almost seven miles per second over the length of a cannon barrel would be unsurvivable. (But see Jules Verne's *From the Earth to the Moon*. In his 1865 novel, he proposed getting around the problem of shooting men to the moon using a fantastic 900-foot-long cannon. It wouldn't work, but it *is* clever.) But, serious proposals *have* been made to put nonhuman payloads into orbit or on the moon, using super-high acceleration up to the escape velocity. Such accelerations would be achieved not with a cannon but, rather, with the far more exotic technology of electromagnetic launchers, which are in actual use today at several sophisticated rollercoaster rides around the world.

1.7 Minimizing with a Computer

For the final two examples of this chapter, which return to the theme of the computer as a useful tool in extremal problems, suppose first that a man can walk n times faster than he can swim (it seems reasonable that $n \geq 1$, but I'll not use that assumption in what follows). He wants to travel from A, on the edge of a circular lake with radius R (centered on point O) to C, also on the edge of the lake. C's location is specified by the given angle β (measured from the diameter AOD), as shown in figure 1.6. His general strategy is to first swim along the chord AB, and then to walk the rest of the way along the lake's edge from B to C. If his total travel time is T, then where should B be to minimize T?

If we denote by θ the central angle subtended by the man's walk, then the isosceles triangle OAB (with the chord AB as its base) has equal base angles of α and a third angle of $\gamma = \pi - \theta - \beta$. Thus,

$$(2\alpha) + (\pi - \theta - \beta) = \pi \text{ radians},$$

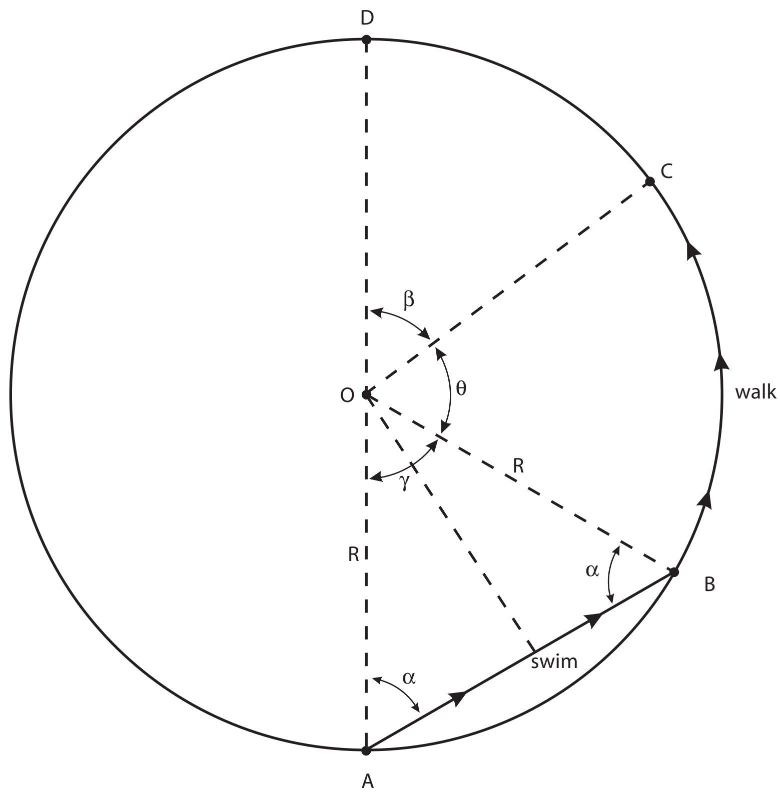

FIGURE 1.6. Crossing a circular lake in minimum time.

or

$$\alpha = \frac{1}{2}\,(\theta + \beta).$$

It is clear from figure 1.6 that the man's swimming and walking distances are, respectively, $2R\cos\left\{\frac{1}{2}\,(\theta+\beta)\right\}$ and $R\theta$. So, if we call his swimming speed unity (in arbitrary units) then his walking speed is n and we have the total travel time as

$$\begin{aligned} T &= 2R\cos\left\{\frac{1}{2}\,(\theta+\beta)\right\} + \frac{R\theta}{n} \\ &= R\left[2\cos\left\{\frac{1}{2}\,(\theta+\beta)\right\} + \frac{\theta}{n}\right]. \end{aligned}$$

As a quick, partial check on this expression, notice that if $\beta = \pi$ radians ($C = A$) then we also have $\theta = 0$ and $T = 0$, just as we should have (it doesn't take any time to travel from where you are to where you are!).

Our problem then is simply this: given a value of β in the interval 0 to π (thus locating C), what θ minimizes T (thus locating B)? This is an easy question to study with the aid of a computer. Figure 1.7 shows how T varies with θ, for five values of n, with $\beta = 0$ (C is directly across the lake from A) and figure 1.8 assumes $\beta = 90°$. In both figures the constant scale factor of R in the expression for T has been ignored since it has no affect on the value for θ that gives an *extrema* in T.

The plots in the two figures contain a wealth of information. In figure 1.7, for example, the $n = 1$ and $n = 1.5$ curves have their minimum values at $\theta = 0$ (the man should *swim*, all the way, from A to C), while the $n = 2$, $n = 2.5$, and $n = 3$ curves have their

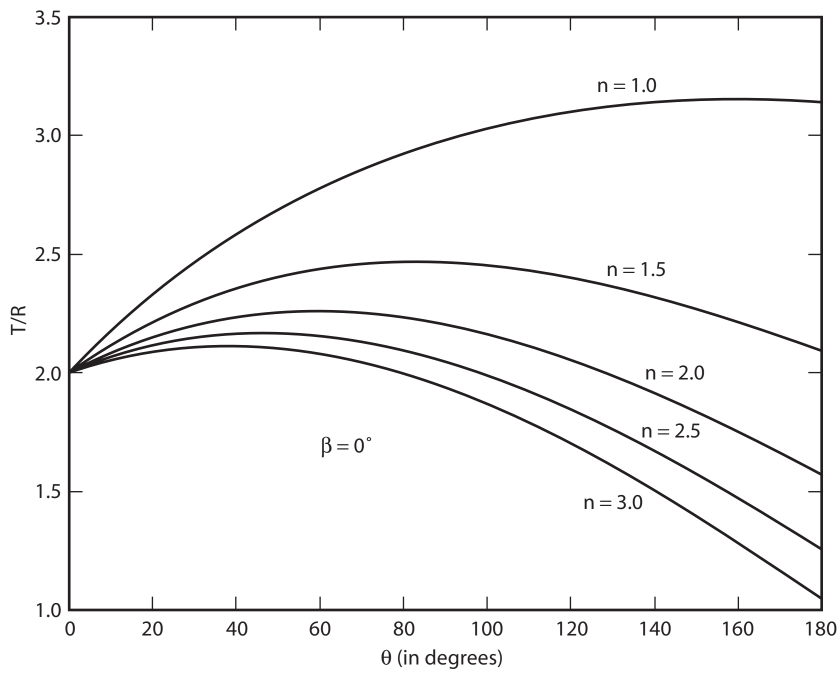

FIGURE 1.7. Total travel time across the lake, $\beta = 0°$.

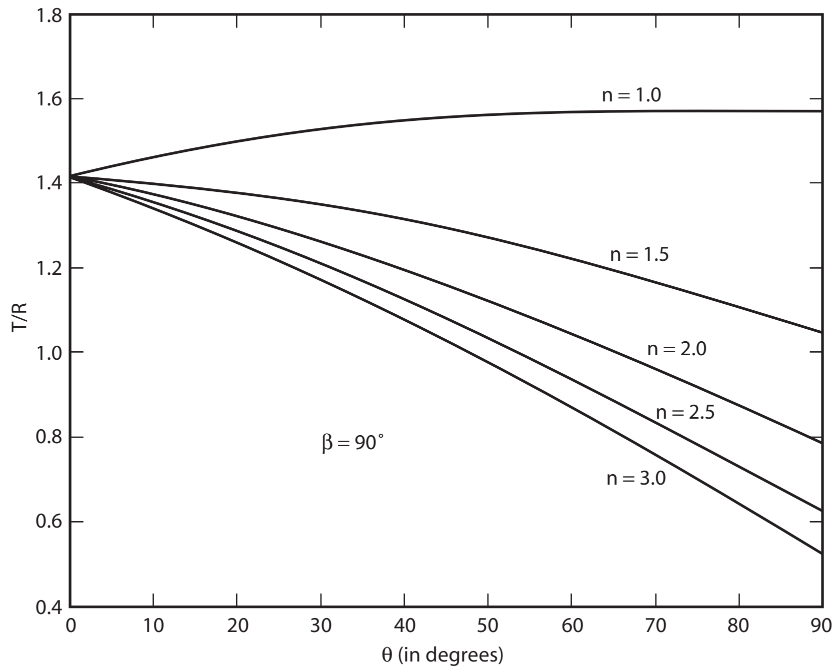

FIGURE 1.8. Total travel time across the lake, $\beta = 90°$.

minimum values at $\theta = 180°$ (the man should *walk*, all the way, from A to C). The curves suggest that there is some value of n between 1.5 and 2 where *either* of the pure walk-only and swim-only strategies would give the minimum travel time. What is that critical value of n? A little thought should convince you it is $n = \frac{1}{2}\pi = 1.57$. The curves of figure 1.8 suggest the same general conclusion for $\beta > 0$, i.e., as n increases from unity the strategy for minimizing the total travel time begins as the pure strategy of swimming all the way and then switches to the pure strategy of walking all the way. Is this always true? That is, for any value of β, is it true that there is never a mixed strategy of walking *and* swimming that minimizes T? I'll leave that for you to think about!

For my last example in this chapter, consider the following problem that is superficially similar to the one just treated, but which offers some surprising complications. But *not* so much complication that we can no longer make a fruitful computer analysis. So,

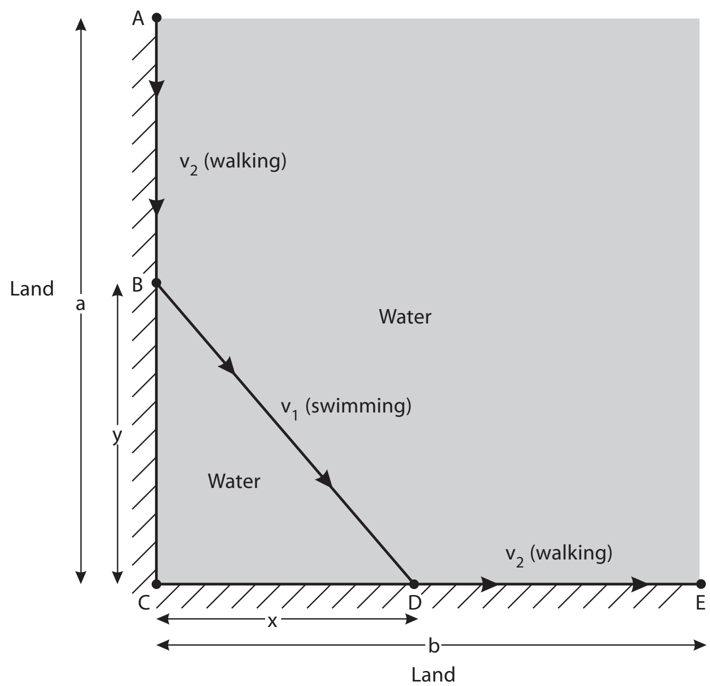

FIGURE 1.9. Another water-crossing problem.

suppose now that the man is initially at point A on a beach with a right-angle bend, as shown in figure 1.9. The man wishes to travel from A to E in minimum time; at any point B, as he walks along the first section of beach toward C, he can enter the water and swim to D, where he exits the water and continues walking on the second section of beach to E. That is, he can "cut a corner" from one section of beach to the other. The lengths of the two sections of beach are a and b, as shown in figure 1.9.

It is not difficult to express the problem mathematically. If we write v_1 and v_2 for the man's speeds while swimming and walking, respectively, and if x and y are the distances of points D and B from the corner of the beach (C), respectively, then the total travel time is a function of *two* variables:

$$T(x, y) = \frac{a - y}{v_2} + \frac{\sqrt{x^2 + y^2}}{v_1} + \frac{b - x}{v_2}$$

$$= \frac{(a+b)-(x+y)}{v_2} + \frac{\sqrt{x^2+y^2}}{v_1}.$$

Our problem, then, is to determine the values of x and y that minimize T for given values of a, b, v_1, and v_2.

The answer for $v_1 > v_2$, for *any* a and b, is physically obvious: $x = b$ and $y = a$, i.e., the man *swims* the entire trip because then he travels the straight line path (shortest possible path) from A to E at the greater speed. As argued before, swimming faster than he can walk isn't very plausible, however, and the case of $v_1 < v_2$ is far more interesting (both physically *and* mathematically). Before continuing with the analysis of $T(x, y)$, it is important to notice that, with a *single* exception, the values of x and y are independent, subject only to the constraints of $0 \leq x \leq b$, $0 \leq y \leq a$. The single exception is that if either x or y is zero then so must be the other; this is because of the physically required *continuous* nature of a path from A to E.

Now, we *could* attack the problem of minimizing $T(x, y)$ with the aid of rather sophisticated calculus, but that isn't attractive for several reasons. First, that would be out of place so early in this book and, second, there is a very pretty *geometric* interpretation of the problem. Indeed, you'll see the same approach used later, when we get to linear programming in chapter 7. And third, the approach I'll show you now makes great use of the sheer computational power of a computer.

To begin, all pairs of points (x, y) that satisfy the constraints $0 \leq x \leq b$, $0 \leq y \leq a$ form what is called the set of *feasible* solutions. For our problem, this set is the rectangle shown in figure 1.10, with the understanding that the bottom edge ($x > 0$, $y = 0$) and the left vertical edge ($x = 0$, $y > 0$) are *not* included in the feasible solution set; the corner *point* $(0, 0)$ *is*, however, in the feasible solution set. We want to find the point in the feasible solution set that minimizes $T(x, y)$. Now, notice that we can write

$$v_1 v_2\, T = v_1(a+b) - v_1(x+y) + v_2\sqrt{x^2+y^2},$$

or

$$\sqrt{x^2+y^2} - \left(\frac{v_1}{v_2}\right)(x+y) = U,$$

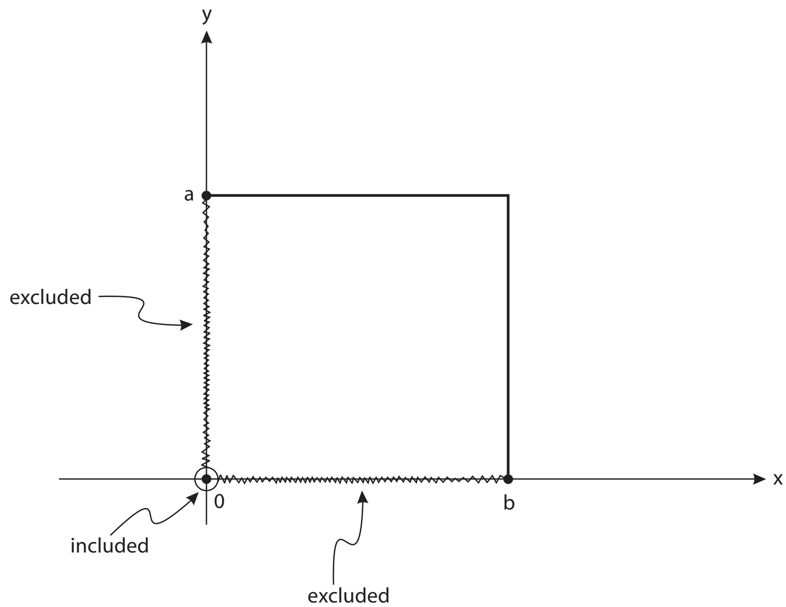

FIGURE 1.10. Feasible solution set for the geometry of figure 1.7.

where

$$U = v_1 T - \left(\frac{v_1}{v_2}\right)(a + b).$$

Since v_1, v_2, a, and b are given positive constants, then it is clear that the minimization of T is equivalent to the minimization of U. This simple observation turns out to be the key observation in the following analysis.

The equation

$$\sqrt{x^2 + y^2} = \left(\frac{v_1}{v_2}\right)(x + y) + U$$

defines a curve $y = y(x)$ for any given U; as we vary U we will also vary the curve $y = y(x)$. We wish to determine the minimum U that results in a curve that still passes through at least one point of the feasible solution set. Using a computer to draw these curves will give us all the insight we need to determine the minimizing $U(= U_{\min})$ and, hence, the minimized $T(= T_{\min})$:

$$T_{\min} = \frac{1}{v_2}(a+b) + \frac{1}{v_1}U_{\min}.$$

To plot $y = y(x)$ as a function of U, it is convenient to change to polar coordinates:

$$x = r\cos(\theta)$$
$$y = r\sin(\theta),$$

and so

$$\sqrt{r^2\cos^2(\theta) + r^2\sin^2(\theta)} = \left(\frac{v_1}{v_2}\right)[r\cos(\theta) + r\sin(\theta)] + U.$$

This is easily reduced to

$$r = \frac{U}{1 - \left(\frac{v_1}{v_2}\right)[\sin(\theta) + \cos(\theta)]},$$

where, of course, it is understood that the radius vector r (at polar angle θ from the origin to the arbitrary point (x, y) on the $y(x)$ curve) is always nonnegative, i.e., $r \geq 0$. That is, the numerator and the denominator must have the same sign.

For the remainder of this analysis, let's assume that both the numerator and denominator are nonnegative, i.e., that

$$U \geq 0$$
$$1 - \left(\frac{v_1}{v_2}\right)[\sin(\theta) + \cos(\theta)] \geq 0.$$

Since $f(\theta) = \sin(\theta) + \cos(\theta)$ achieves a maximum value of $\sqrt{2}$ at $\theta = 45°$ (easily verified by either setting $df/d\theta = 0$ or by simply plotting $f(\theta)$), then as long as

$$\left(\frac{v_1}{v_2}\right) \leq \frac{1}{\sqrt{2}},$$

we will have $r \geq 0$ for any $U \geq 0$ for all values of the polar angle θ. That is, we are now dealing with a restrictive case of $v_1 < v_2$, i.e., with $v_1 \leq 1/\sqrt{2}v_2$.

Returning to the original x, y coordinate system, we have the result we are after: the $y = y(x)$ curve is the curve defined by

$$x = \frac{U\cos(\theta)}{1 - \left(\frac{v_1}{v_2}\right)[\sin(\theta) + \cos(\theta)]},$$

$$y = \frac{U\sin(\theta)}{1 - \left(\frac{v_1}{v_2}\right)[\sin(\theta) + \cos(\theta)]}.$$

We can see now that all U "does" is *scale* the plot. Indeed, in figure 1.11 you'll find the curve $y = y(x)$ for $v_2 = 5$ with four different values of v_1 (all satisfy the condition $v_1 \le (1/\sqrt{2})v_2$), for two values of U (solid for $U = 1$, dashes for $U = 0.4$). It is clear from these plots that $y = y(x)$ is elliptical, and that as U decreases toward zero, the curves shrink inward to around the lower-left-corner point of the feasible solution set (in solid).

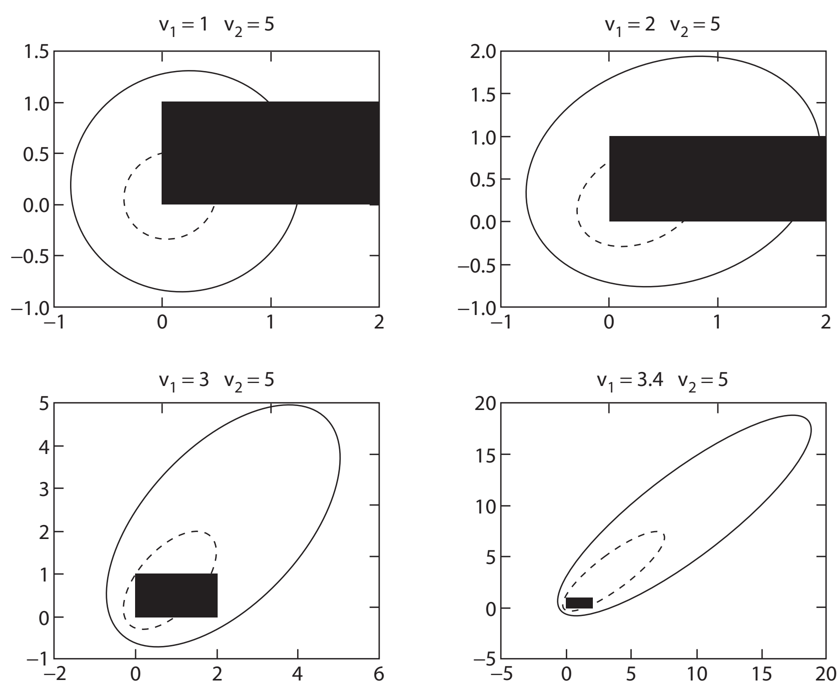

FIGURE 1.11. Converging to the minimum-time solution as U vanishes.

I have, to be specific, used $a = 1$ and $b = 2$ to define the feasible solution rectangle, but it should be obvious that the actual size of the rectangle doesn't matter. That is, for any choice of the a and b values, the smallest nonnegative U is $U = 0$, which collapses the $y = y(x)$ curve to the single point $x = 0$, $y = 0$. Since the smallest U gives the smallest T, then $T(0, 0)$ is the minimum journey time.

Thus, somewhat surprisingly I think, if $v_1 \leq (1/\sqrt{2})v_2$ then the man should *walk* the *entire* way, and that is so no matter what are the dimensions of the beach. So, we have solved the problem for the two cases of $v_1 \geq v_2$ (swim all the way) and $v_1 \leq (1/\sqrt{2})v_2$ (walk all the way). What if $(1/\sqrt{2})v_2 < v_1 < v_2$? I'll leave that case for *you* to ponder!

How to Walk Out of the Woods

Our lost hiker doesn't know which way to go to walk directly back to his car, but he does know that the car is *somewhere* on the circumference of the circle, with a one-mile radius, centered on his present location. So, to insure he returns to his car, he should first walk one mile in a randomly selected direction—if he is *very* lucky he'll walk straight back along the radius that was his original path—and then walk along the circular (one-mile-radius) path centered on his starting point. *Somewhere* along that circular path is his car. The absolute maximum distance he'll have to walk is the initial one-mile radius plus the 2π-mile circumference, i.e., $1 + 2\pi = 7.2832$ miles.

This is a mathematician's solution, of course, as it ignores the practical detail of just *how* one manages to walk along a circular arc in a densely wooded forest. Another setting for this problem, that avoids that objection, is to have our lost soul be a fisherman in a rowboat one mile off shore, in a dense fog. *Rowing* in a circle is now "easy"; all the fisherman need do is to take one end of a rope, drop it overboard with a heavy anchor, measure the depth of the water, and then (with due regard for the depth) row away until enough rope has played out to put him a mile away. He can then, keeping the rope taunt, swing in a circular path about his original position.

Now, here's a new twist on this puzzle for you to think about. Is this solution the best one can do, where *best* means

(continued)
having the minimum maximum path length? The answer is *no*, there are paths that require smaller maximum travel distances that, with certainty, return the fisherman back to shore (this is not quite the same as getting back to the *car* itself, of course, but for both the hiker and fisherman it is probably good enough!)

To see this, imagine our lost fisherman first picks some angle $\theta > 0$, and then at random picks a direction that he assumes is the direct one-mile path to the shore. He then rows at angle θ to this line for a distance of $\sqrt{1 + \tan^2(\theta)}$, as shown in figure

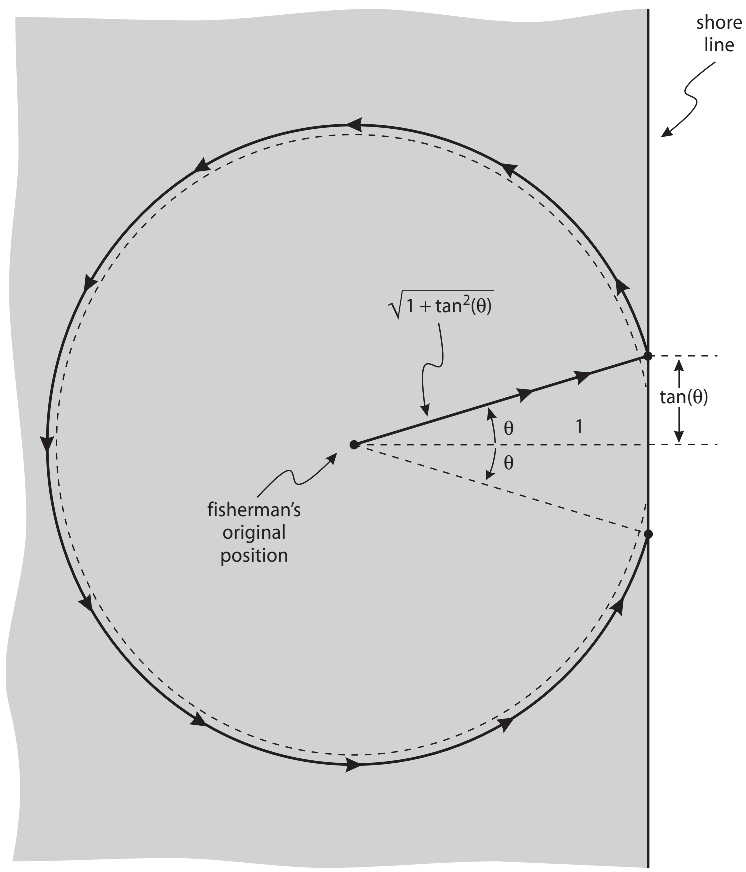

FIGURE 1.12. Geometry of the lost fisherman problem.

(continued)

1.12. That is, the triangle formed by his initial position, his new position, and the end of the one-mile path in the *assumed* direction to the shore, is a right triangle. If the assumed direction to the shore happens to be correct, then his journey is over. Otherwise, he next rows along a circular path with radius $\sqrt{1+\tan^2(\theta)}$ until the line from his original position to his present position is once again θ with respect to the assumed one-mile path. That is, he rows along a circular path through an angle of $2\pi - 2\theta$ radians. Since the original solution was sure to eventually return him to shore, it is clear from figure 1.12 that this new path will also eventually reach the shore as well (since the original solution path lies entirely *inside* the new path). The maximum total length of this new path is

$$L(\theta) = \sqrt{1+\tan^2(\theta)} + 2\pi\sqrt{1+\tan^2(\theta)}\left(\frac{2\pi - 2\theta}{2\pi}\right)$$
$$= [1 + 2\pi - 2\theta]\sqrt{1+\tan^2(\theta)}.$$

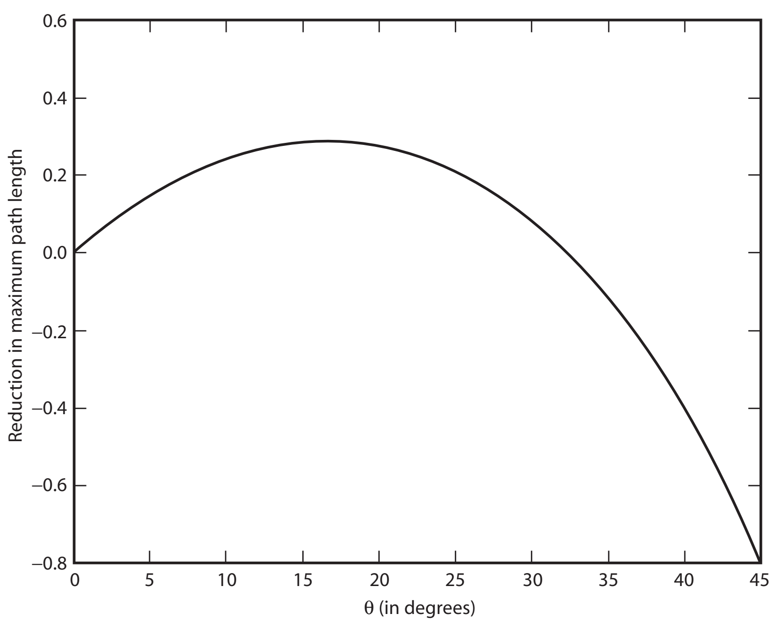

FIGURE 1.13. Proof that $2\pi + 1$ is not the minimum path length.

(continued)

Notice that $L(0) = 1+2\pi$, the maximum length of the original solution. The astonishing result is that there are values for $\theta > 0$ that *do* result in $L(\theta) < L(0)$!

This claim is easily established by simply plotting the quantity $L(0)-L(\theta)$ versus θ, as shown in figure 1.13. (We might try setting the derivative of $L(\theta)$ to zero, of course, to calculate the value of θ that minimizes $L(\theta)$, but if you do that you'll find you are led to a transcendental equation in θ, i.e., you will *still* need to use a computer—see section 4.5.) Figure 1.13 shows that, for $\theta = 16.61°$, $L(\theta)$ is 0.2879 miles *less* than 7.2832 miles.

Now, one final question for you—can our fisherman do *even better*? Is there a rowing path that has an even smaller maximum length that is still certain to get him to shore? The answer is again *yes*, and an analysis demonstrating that is given in appendix H—but don't look until you've made an honest try.