APPROACHES TO AN OPTIMIZATION PROBLEM.

A mathematical vignette

Ed Barbeau, University of Toronto

1. The problem. Consider the following problem in elementary calculus: *Minimize for $x > 0$ the function*

$$f(x) = x^2 + \frac{2}{x}.$$

We should begin by getting an intuitive feel for the sitution. If we were to graph the function, what does it look like? There are two terms in the expression defining it, and we should note that the first term is the controlling one for large values of x and the second the controlling term for small values of x. In other words, the value of $f(x)$ is very close to x^2 when x is large and close to $2/x$ when x is small. To get a rough picture, begin by graphing $y = 2/x$ and $y = x^2$ for $x > 0$ on the same set of axes. The graph of $f(x)$ lies above both of these curves, and looks as though it is bowl-shaped, reaching its lowest point somewhere near to where the graphs of x^2 and $2/x$ cross. (*Question:* Do you think this means that the minimum of $f(x)$ occurs when $x^2 = 2/x$ or $x = 2^{1/3}$?) Thus, we would expect to find a unique minimum value for $f(x)$. We shall see whether this is indeed true.

There are several ways of approaching an optmization problem, and we shall look at the use of algebra, the use of standard inequalities and calculus.

2. Algebra. Let us assume for the moment that the graph of $f(x)$ is bowl shaped. Then a line in the first quadrant parallel to the $x-$axis will either not intersect the graph at all, intersect it in two points or, at the bottom of the bowl, be tangent to it. The tangent line can be thought of intersecting the graph at two points which coincide. Algebraically, this means that when we solve the equation $f(x) = k$, we will get an equation for x that has two positive roots when k is large, none when k is small and a double root when the line $y = k$ is tangent.

Before we proceed further, recall the *Factor Theorem*, which provides that a number a is a root of the polynomial $p(x)$ (*i.e.* satisfies $p(a) = 0$) if and only if the polynomial can be factored $p(x) = (x - a)q(x)$ for some polynomial $q(x)$.

Begin with an arbitrary number $u > 0$ and consider the equation $f(x) = f(u)$. This will clearly have the solution $x = u$, and it will have other solutions corresponding to other values of x for which the function takes the value $f(u)$. We can write the equation out as

$$x^2 + \frac{2}{x} = u^2 + \frac{2}{u}.$$

This is equivalent to

$$ux^3 - (u^3 + 2)x + 2u = 0.$$

The left side is a polynomial of degree 3 in x. Since u is a root, we expect the $x-u$ is a factor of the left side. Indeed,

$$ux^3 - (u^3+2)x + 2u = (x-u)(ux^2 + u^2x - 2).$$

When $x = u$, minimizes $f(x)$, then $x-u$ should divide the cubic to degree at least 2, so that it is a factor of the quadratic $ux^2 + u^2x - 2$. Thus, the quadratic will vanish when $x = u$: $2u^3 - 2 = 0$. This happens when $u = 1$, so we can conclude that $f(1) = 3$ is a candidate for the minimum value of $f(x)$.

We can check this directly using algebra:

$$\begin{aligned} f(x) - f(1) &= x^2 + \frac{2}{x} - 3 = \frac{1}{x}[x^3 - 3x + 2] \\ &= \frac{1}{x}[(x-1)(x^2+x-2)] = \frac{1}{x}(x-1)^2(x+2) \geq 0, \end{aligned}$$

when $x > 0$, with equality if and only if $x = 1$.

3. The arithemtic-geometric means inequality. The simplest version of the arithmetic-geometric means inequality is that, when $a, b > 0$, then

$$\frac{1}{2}(a+b) \geq \sqrt{ab}.$$

The simplest way to see this is to take twice the difference between the two sides and note that we get a square:

$$(a+b) - 2\sqrt{ab} = (\sqrt{a} - \sqrt{b})^2 \geq 0.$$

Equality occurs if and only if $a = b$.

An alternative argument exploits the theory of the quadratic. For each positive real pair a, b, the quadratic equation

$$(x-a)(x-b) = x^2 - (a+b)x + ab = 0$$

has real roots. Therefore, its discriminant must be nonnegative:

$$(a+b)^2 - 4ab \geq 0.$$

Rearranging terms and taking square roots yields the arithmetic-geometric means inequality.

A third argument uses geometry and a sketch goes like this. Let AB be a diameter of a circle, and suppose that C is a point on it so that the length of AC is a and the length of CB is b. Let D be a point on the circumference of the circle for which $CD \perp AB$. Then the length of CD is $\sqrt{ab}$. This is evidently not greater than the radius $\frac{1}{2}(a+b)$ of the circle, and equality occurs if and only if C is the centre of the circle (*i.e.* $a = b$).

This can be generalized to any number of positive reals. Let n be a positive integer and let $a_1, a_2, \ldots, a_n > 0$; then

$$\frac{1}{n}(a_1 + a_2 + \cdots + a_n) \geq (a_1 a_2 \cdots a_n)^{1/n},$$

with equality if and only if all the a_i are equal. We will need this only for $n = 3$ and $n = 4$. Let us deal with the $n = 4$ case first. Suppose that a, b, c, d are positive reals. Then, by repeated use of the inequality for two variables, we obtain that

$$\frac{1}{4}(a+b+c+d) = \frac{1}{2}\left(\frac{a+b}{2} + \frac{c+d}{2}\right) \geq \frac{1}{2}\left(\sqrt{ab} + \sqrt{cd}\right)$$
$$= \sqrt{\sqrt{ab}\cdot\sqrt{cd}} = (abcd)^{1/4}.$$

To get the case for $n = 3$, we now let $d = (abc)^{1/3}$. Then

$$\frac{1}{4}\left(a+b+c+(abc)^{1/3}\right) \geq [(abc)(abc)^{1/3}]^{1/4} = (abc)^{1/3}.$$

Therefore

$$\frac{1}{4}(a+b+c) \geq \frac{1}{4}[4(abc)^{1/3} - (abc)^{1/3}] = \frac{3}{4}(abc)^{1/3},$$

so that

$$\frac{1}{3}(a+b+c) \geq (abc)^{1.3}$$

as desired.

A more direct proof of the $n = 3$ case can be had by taking $(a, b, c) = (u^3, v^3, w^3)$ and noting that

$$u^3 + v^3 + w^2 - 3uvw = (u+v+w)(u^2+v^2+w^2-uv-vw-wu)$$
$$= \frac{1}{2}(u+v+w)[(u-v)^2 + (v-w)^2 + (w-u)^2] \geq 0.$$

Equality occurs if and only if $a = b = c$.

To return to the original problem, we take $a = x^2$ and $b = c = 1/x$, so that $abc = 1$. Applying the arithmetic-geometric means inequlity, we find that

$$x^2 + \frac{2}{x} = x^2 + \frac{1}{x} + \frac{1}{x} \geq 3,$$

with equality if and only if $x^2 = 1/x$ or $x = 1$.

4. Calculus. Finally, we resort to calculus. First, recall that the derivative of the function $1/x$ is equal to

$$\lim_{h\to 0}\frac{1}{h}\left(\frac{1}{x+h} - \frac{1}{x}\right) = \lim_{h\to 0}\frac{-1}{x(x+h)} = \frac{1}{x^2}.$$

Then

$$f'(x) = 2x - \frac{2}{x^2} = \frac{2(x^3-1)}{x^2}.$$

From here, there are three ways to proceed.

(i) We know that where the derivative vanishes produces a candidate for the minimizing value of x. Since $f'(1) = 0$, we can then check whether $f(x) - f(1) \geq 0$ (which we did in section 2).

(ii) We know that $f(x)$ decreases or increases according to the sign of the derivative. Since $f'(x) < 0$ for $0 < x < 1$ and $f'(x) > 0$ for $x > 0$, we deduce that

$f(x)$ decreases as x increases to 1, and then increases for $x > 1$. Thus, it reaches its minimum value when $x = 1$.

(iii) Finally, we can apply the second derivative test. Since $f''(x) = 2 + 4/x^3 > 0$ for $x > 0$, we know that $f(x)$ is a convex function. Since $f''(1) > 0$, $f(x)$ must have a minimum at $x = 1$. (Actually, all the positivity of the second derivative at the critical point gives you is that there is a local minimum; it leave open the possibility the function my take a lower value elsewhere. This is not an issue with the present example.)

For a given function to minimize, these three aproaches are available. Which one you choose will depend on the severity of the computations that might occur and how easy it is to read off the information you need.