Parabola Volume 31, Issue 1 (1995)

GOLDEN SECTION SEARCH

Geoffrey Coombs[1]

The problem of finding the minimum (or the maximum) of a function is a familiar application of calculus. However, if the derivative of the function in question is not known, one must resort to numerical methods. In this article I intend to describe a simple numerical procedure for estimating the location of the minimum of a function. The novel aspect of this procedure is the unexpected emergence of the Golden Section. This number, which is usually represented by the letter ϕ, arises most simply as the location of the point P in an interval AB that subdivides the interval in such a way that $AB : AP = AP : PB$. The Golden Section is defined to be just this ratio so that

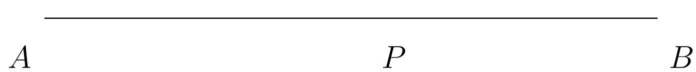

Figure 1

$$
\begin{aligned}
\phi = \frac{AB}{AP} &= \frac{AP}{PB}, \\
&= \frac{AP}{AB - AP}, \\
&= \frac{1}{AB/AP - 1}, \\
&= \frac{1}{\phi - 1}.
\end{aligned}
$$

It can thus be seen that ϕ satisfies the quadratic equation $\phi^2 - \phi - 1 = 0$ which has just one positive root and so $\phi = \frac{1}{2}(1 + \sqrt{5})$. This quantity that was the subject of a recent article in **Parabola** by Milan Pahor and Tony van Ravenstein (Vol.28, No.2, 1992).

Ideally a numerical procedure should return an estimate of the location of the minimum to some specified accuracy subject to course to the limitations inevitably placed on such a calculation by the limited precision of the calculator or computer used. We shall consider the case of a function that has a single minimum between two points and is increasing on the left of the point and also increasing on the right. How do we go about estimating the location of this point? The method I shall use is based on the simple idea of reducing the size of the interval in which the minimum is known to lie. We will develop a method of systematically reducing the size of the interval until the desired precision is obtained.

[1]Geoff is a member of the Department of Applied Mathematics at UNSW

Let us begin with a naive approach. Suppose we calculate the value of the function in the mid-point of the interval. On the basis of this value are we able to ascertain on which side of the mid-point the minimum lies? The obvious answer is that, in the absence of any other information, we cannot. Will the calculation of the function at another point be any help? In the interests of symmetry let us instead start again and calculate our function values at two points which trisect the interval as is done in Figure 2. **Now if we compare the function values at these two points it is clear that the minimum must lie between the point with the greater value and the more remote end-point.** In the Figure we can conclude that the minimum must lie between c and b. If the value at c is less than that at d, the new interval would be ad. In the event of equality, either interval may be chosen. And thus we have reduced the size of the interval in which the minimum is known to lie to two-thirds of its original size. We can then apply the method again to obtain a further reduction of this interval and so on until the length of the interval is less than the desired precision, and we have our result. Now it is evident that in any such process the most time consuming element of

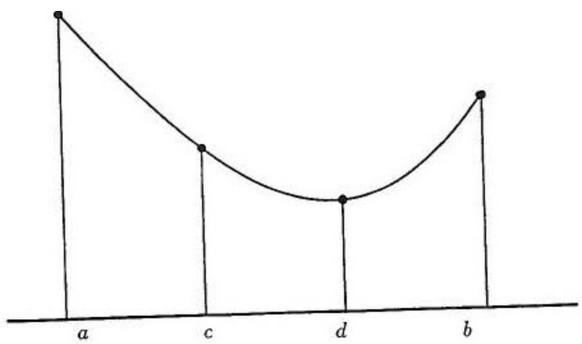

Figure 2

the calculation is the evaluation of the function. In the method proposed it costs two function evaluations for each two-thirds reduction of the interval. So after n iterations of the method we will have reduced the interval to $(2/3)^n$ of its original size. Another way of thinking about this is that we achieve a $\sqrt{2/3} \approx 0.82$ reduction for each function evaluation. Can we do better than this?

We can! And the reason we can follows from the following two observations. It was not necessary to choose the positions of the two points in any special way. And secondly, at each step we discard the two function values after having determined which is the greater. Would it not be possible to re-use at least one of these values at the next step.

Consider Figure 3.

The original interval has end-points a and b and the function has been calculated at the two new points c and d. In this case it has been found that the value at c is the greater and thus the new interval is $[c, b]$. We relabel this as $[a', b']$. The two points in the new interval we label as c' and d'. Here we have not as yet specified the method for placing the points. We would like to place these points so that we are able to re-use the function value at at least one of the points. Looking at the figure, it appears we

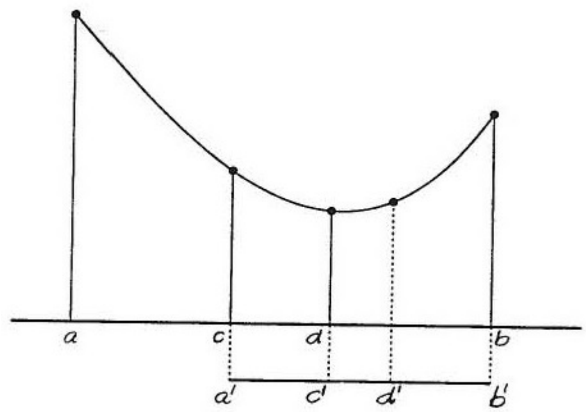

Figure 3

might be able to make the point c' coincide with d. If we do this and make the quite reasonable demand that the ratios of the subintervals to the whole interval remain the same. In which case we require that $cb/ab = c'b'/a'b'$. Let us call this ratio r. From the Figure we can see that $c'b' = db$ and $a'b' = cb$, so that $r = cb/ab = c'b/a'b' = db/cb$. We also require that the positions of the points c and d be symmetric so that $ac = db$, with the result that $db = ac = ab - cb$.

We are now in a position to obtain an expression for the ratio r.

$$
\begin{aligned}
r = \frac{cb}{ab} &= \frac{c'b'}{a'b'}, \\
&= \frac{db}{cb}, \\
&= \frac{ab - cb}{cb}, \\
&= \frac{ab}{cb} - 1, \\
&= \frac{1}{r} - 1.
\end{aligned}
$$

With a little rearrangement, we see that the ratio satisfies the quadratic equation $r^2 + r - 1 = 0$ which has the roots $\frac{1}{2}(-1 \pm \sqrt{5})$. Now the ratio must be positive so we reject the negative root and conclude that $r = \frac{1}{2}(\sqrt{5} - 1) \approx 0.618$. It follows that each new interval will be less than the previous one by a factor of $r \approx 0.62$, giving a reduction factor of 62% per function evaluation to compare with 82% obtained by using the naive method. We may also express the change of the length of the interval as the ratio of the old interval to the new, which is $1/(1 - r) = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$. This is just the Golden Ratio and it is from this quantity that the method gets its title: *The Golden Section Search.*

So this method proceeds as for the naive method discussed above. For each iteration a new function value is calculated at a point determined as described and compared with the appropriate function value retained from the previous iteration and the new interval is thus chosen.