STALKING THE SOLUTION TO AN EQUATION

Frederick S. Gass
Miami University
Oxford, Ohio

This article shows a method of solving equations numerically, using ideas that resemble Newton's method but requiring no familiarity with calculus. The method is useful for attacking equations that don't yield to factoring or other familiar algebraic techniques, and it can be implemented on any scientific calculator. Students who like working with computers will find it interesting and worthwhile to construct a program for this method to facilitate its use.

Here's the basic idea: You want to solve an equation in one variable, say x, and have moved all terms to one side so that the equation looks like $f(x) = 0$ for some formula $f(x)$. Solving this equation is equivalent to finding the x–intercepts of the graph of $y = f(x)$, and that is how we will view the problem from now on. (See Figure 1.) The title of this article is prompted by the fact that one starts with two or three initial estimates of the solution and then follows a sequence of better and better estimates that hunt down the solution like big game.

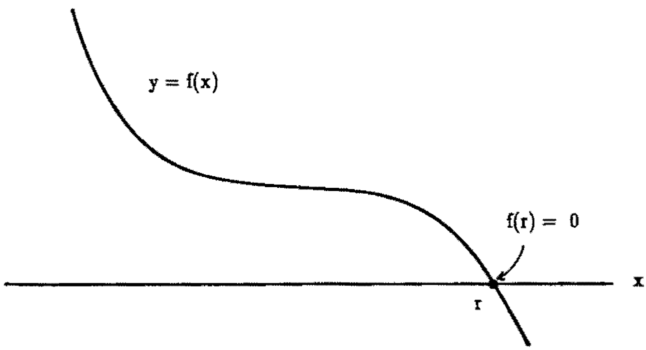

Fig. 1

Now on with the hunt. After looking over your equation $f(x) = 0$, you choose a value of x that comes reasonably close to satisfying the equation, and then you choose another value that, if possible, comes even closer (but that's not crucial). For example, faced with $x^3 - x - 10 = 0$, you might let 2 be the initial estimate and take 2.5 as the next one (after trying out 3 and seeing how f(x) jumps from $f(2) = -4$ to $f(3) = 14$). This is where the "method" takes over and generates additional estimates, one at a time.

At any moment in the process, you have a "current" estimate for x (e.g. 2.5 in the example) and a "previous" one (e.g. 2). Let's symbolize them respectively as "x_c" and "x_p". Figure 2 shows how to use points $(x_p,y_p)$and (x_c,y_c) to obtain a "next" estimate, x_n, which should be closer to the solution we're after. Namely, to get x_n you locate the point where the dotted secant line meets the x–axis. Because that line plays a central role in the solution, numerical analysts call this "the secant method" of solving equations.

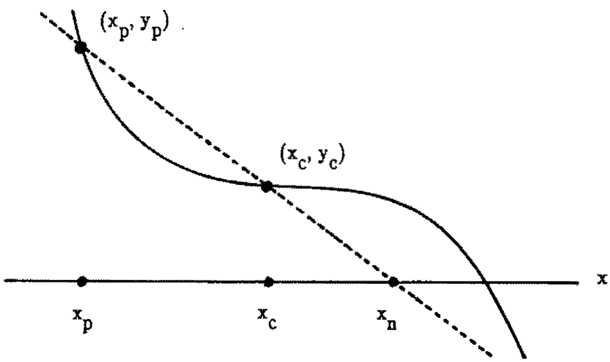

Fig. 2

As you might expect, there is a simple formula that gives x_n in terms of x_p, x_c, y_p and y_c, and it is obtained easily by looking at Figure 2 and calculating the slope of the line using two different pairs of points:

[slope using (x_c,y_c) and (x_p,y_p)] = [slope using $(x_n, 0)$ and (x_c,y_c)].
If you fill in the slope formulas and solve for x_n, you can get

$$(*) \qquad x_n = x_c - y_c\left(\frac{x_c - x_p}{y_c - y_p}\right).$$

(Notice that the fractional part is just the reciprocal of the secant line's slope.)

The overall strategy is to use the current and previous estimates for x to generate the next estimate via (*). That next estimate then becomes the current one, the "old" current estimate becomes the previous one, and we're ready to apply (*) again with a slightly different pair of numbers x_p and x_c. Let's see how this goes by working through the example mentioned earlier.

EXAMPLE. Find the unique real solution to $x^3 - x - 10 = 0$, correct to four decimal places.

SOLUTION. We'll round off to four decimal places at the end but carry lots of digits in the meanwhile. Whenever y–values are needed, we'll be using the function $y = x^3 - x - 10$ to calculate them. Here are the steps to follow in tracking our quarry, with successive estimates underlined to help you compare them.

STEP 1. As suggested earlier, we start by letting $x_p = 2$ and $x_c = 2.5$. Then $y_p = -4$, $y_c = 3.125$ and by (*), $x_n = \underline{2.280701754}$.

STEP 2. Now we regard 2.280701754 as the current x estimate and 2.5 as the previous one, so $x_p = 2.5$ and $x_c = 2.280701754$. Then $y_p = 3.125$, $y_c = -.4174023913$ and $x_n = \underline{2.306541736}$.

STEP 3. In the next up–date, $x_p = 2.280701754$ and $x_c = 2.306541736$, so $y_p = -.4174023913$, $y_c = -.03542882499$ and $x_n = \underline{2.308938447}$.

STEPS 4, 5 AND 6. The successive estimates we get for x are
2.308907286, 2.30890732 and 2.30890732.

Since those last two estimates are the same, we might expect that the method has carried us as far as it can, given our initial estimates. (In fact, if we try another step, everything – and in particular formula (*) – falls apart. Do you see why?) In the process of taking step 6, we find that $x_c = 2.30890732$ yields $y_c = 3.52 \times 10^{-9}$, so that 2.30890732 nearly satisfies our original equation. Our answer, then, is 2.3089. [Editors' note: Rounding–off errors may produce different values along the way, but the final result should be the same.]

The secant method is quite reliable when used with smooth, continuous graphs, although formula (*) is subject to roundoff error when y_c and y_p become very close. Also, the shape of the graph near solutions and estimated solutions can make the initial choice of x_p and x_c especially crucial. For example, Figure 3 shows an arrangement that would lead one on a wild goose chase away from the solution.

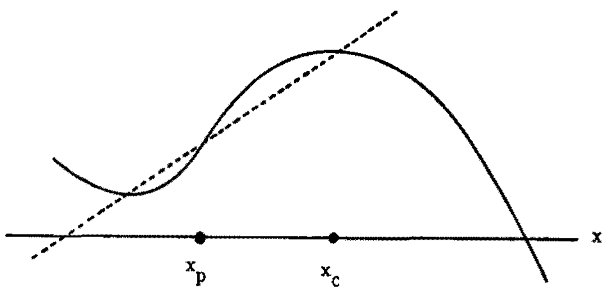

Fig. 3

Since the solution of equations is a frequently–occurring task in mathematics, it pays to mechanize procedures such as the secant method to make them more convenient. If you enjoy programming, then try your hand at a program for this method. Notice that the values of x_c, y_c and x_n at each step are just passed along as the x_p, y_p and x_c for the next step, so only y_c and x_n need to be figured. Happy hunting!