Appendix K

ZEROS OF EQUATIONS

The *"zero"* of an equation is also called the *"root"* of the equation. Here we use the term "zero" so as not to confuse it with roots (e.g., square root) that were previously discussed. Consider the following equation

$$4x^2 - e^x = 0 \tag{K-1}$$

The zero is the x value that satisfies this equation (in this case there are three solutions, $x = 4.306585$, $x = 0.714806$, and $x = -0.407776$). Although it may seem that finding the zero is very limiting (after all, how many times does a practical problem have an equation that equals zero), in fact it is extremely powerful. **Any** equation can be rewritten so it is equal to zero. For example, the equation

$$4x^2 = e^x \tag{K-2}$$

is easily manipulated so that so it equals zero (see Equation K-1).

Finding zeros allows us to solve equations that cannot be solved *explicitly*. For example, the equation

$$4x^2 = 64 \tag{K-3}$$

$$x^2 = 16$$

$$x = 4$$

can be solved explicitly because algebraic manipulations can isolate x on one side of the equation. However, Equation K-2 **cannot** be solved explicitly; there are no algebraic manipulations that can isolate x on one side of the equation. Therefore, an *implicit* solution is required. By manipulating the equation so that it equals zero, the problem can be solved by finding the "zero" of the equation.

In summary, the solution to **any** equation with a single unknown can be solved with the following procedure:

1. Manipulate the equation so that it equals zero.
2. Find the zero of the equation.

The first step is easy, but the second step can be challenging. Here, we present *algorithms* (i.e., step-by-step methods) for finding zeros. Some methods are simple enough that a calculator is sufficient, whereas others require computer solutions. (*Note*: the distinction between a calculator and a computer is blurred; some sophisticated calculators have routines for finding zeros that can be accessed with a single key stroke.)

K.1 Linear Equation (First-Order Polynomial)

A linear equation has the form

$$ax+b=0 \tag{K-4}$$

The number of possible zeros is equal to the order of the polynomial, which in this case is 1. This equation is easily solved for x

$$x=-\frac{b}{a} \tag{K-5}$$

K.2 Quadratic Equation (Second-Order Polynomial)

A quadratic equation has the form

$$ax^2+bx+c=0 \tag{K-6}$$

The solution to this equation is very famous:

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{K-7}$$

The terms under the square root are called the *discriminant*, $D=b^2-4ac$. Because this is a second-order polynomial, there are two possible zeros, which are:

- Real and unequal if $D>0$

$$x_1=\frac{-b+\sqrt{b^2-4ac}}{2a} \quad \text{(real)} \tag{K-8}$$

$$x_2=\frac{-b-\sqrt{b^2-4ac}}{2a} \quad \text{(real)} \tag{K-9}$$

- Real and equal if $D=0$

$$x_1=x_2=\frac{-b}{2a} \quad \text{(real)} \tag{K-10}$$

- Complex if $D<0$.

$$x_1=-\frac{b}{2a}+\frac{\sqrt{4ac-b^2}}{2a}i \quad \text{(complex)} \tag{K-11}$$

$$x_2 = -\frac{b}{2a} - \frac{\sqrt{4ac - b^2}}{2a} i \quad \text{(complex)} \tag{K-12}$$

K.3 Cubic Equation (Third-Order Polynomial) (Advanced Topic)

A cubic equation has the form

$$ax^3 + bx^2 + cx + d = 0 \tag{K-13}$$

The *discriminant, Y,* is

$$Y = \frac{W^2}{4} + \frac{V^3}{27} \tag{K-14}$$

where

$$W = \frac{2\left(\frac{b}{a}\right)^3 - 9\left(\frac{b}{a}\right)\left(\frac{c}{a}\right) + 27\left(\frac{d}{a}\right)}{27} \tag{K-15}$$

$$V = \frac{3\left(\frac{c}{a}\right) - \left(\frac{b}{a}\right)^2}{3} \tag{K-16}$$

Because the cubic equation is third order, there are three potential zeros, which are:

- One real and two complex if $Y > 0$

$$x_1 = S + T - \frac{b}{3a} \quad \text{(real)} \tag{K-17}$$

$$x_2 = -\left(\frac{S+T}{2} + \frac{b}{3a}\right) + \left(\frac{\sqrt{3}}{2}(S-T)\right) i \quad \text{(complex)} \tag{K-18}$$

$$x_3 = -\left(\frac{S+T}{2} + \frac{b}{3a}\right) - \left(\frac{\sqrt{3}}{2}(S-T)\right) i \quad \text{(complex)} \tag{K-19}$$

where

$$S = \sqrt[3]{\left|-0.5W + \sqrt{Y}\right|} \quad (\text{if } -0.5W + \sqrt{Y} \geq 0) \tag{K-20}$$

$$S = -\sqrt[3]{\left|-0.5W + \sqrt{Y}\right|} \quad (\text{if } -0.5W + \sqrt{Y} < 0) \tag{K-21}$$

$$T = \sqrt[3]{\left|-0.5W - \sqrt{Y}\right|} \quad (\text{if } -0.5W - \sqrt{Y} \geq 0) \tag{K-22}$$

$$T = -\sqrt[3]{\left|-0.5W - \sqrt{Y}\right|} \quad (\text{if } -0.5W - \sqrt{Y} < 0) \tag{K-23}$$

- Three unequal reals if $Y < 0$

$$x_1 = 2\sqrt{\frac{|V|}{3}} \cos\left(\frac{\theta}{3}\right) - \frac{b}{3a} \quad \text{(real)} \tag{K-24}$$

$$x_2 = 2\sqrt{\frac{|V|}{3}}\cos\left(\frac{2\pi+\theta}{3}\right) - \frac{b}{3a} \quad \text{(real)} \tag{K-25}$$

$$x_3 = 2\sqrt{\frac{|V|}{3}}\cos\left(\frac{4\pi+\theta}{3}\right) - \frac{b}{3a} \quad \text{(real)} \tag{K-26}$$

where

$$Z = -\frac{W}{2}\sqrt{\frac{27}{|V^3|}} \tag{K-27}$$

$$\theta = \frac{\pi}{2} \text{ (if } Z = 0) \tag{K-28}$$

and

$$\theta = \arctan\left(\frac{\sqrt{1-Z^2}}{Z}\right) \text{ (if } Z \neq 0) \tag{K-29}$$

(*Note*: If $\theta < 0$, then increase θ by π.)

- Three reals (two identical) if $Y = 0$

$$x_1 = 2S - \frac{b}{3a} \quad \text{(real)} \tag{K-30}$$

$$x_2 = x_3 = -S - \frac{b}{3a} \quad \text{(real)} \tag{K-31}$$

where

$$S = \sqrt[3]{\frac{|W|}{2}} \text{ (if } W \leq 0) \tag{K-32}$$

$$S = -\sqrt[3]{\frac{|W|}{2}} \text{ (if } W > 0) \tag{K-33}$$

K.4 Quartic Equations (Fourth-Order Polynomial) (Advanced Topic)

A quartic equation has the form

$$ax^4 + bx^3 + cx^2 + dx + e = 0 \tag{K-34}$$

The four zeros of this equation are found by first finding a real zero, y_1, of the following cubic equation

$$y^3 - \frac{c}{a}y^2 + \left(\frac{bd}{a^2} - 4\frac{e}{a}\right)y + \left[4\frac{ce}{a^2} - \left(\frac{d}{a}\right)^2 - \frac{b^2e}{a^3}\right] = 0 \tag{K-35}$$

Depending on the value of the discriminant, this cubic equation could produce one, two, or three real zeros. Select the zero, y_1, that produces only real coefficients, i.e.,

$$4y_1 > \left(\frac{b}{a}\right)^2 - 4\frac{c}{a} \qquad \text{and} \qquad y_1^2 > 4\frac{e}{a}$$

in the following two quadratics:

$$x^2+0.5\left(\frac{b}{a}+\sqrt{\left(\frac{b}{a}\right)^2-4\frac{c}{a}+4y_1}\right)x+0.5\left(y_1-\sqrt{y_1^2-4\frac{e}{a}}\right)=0 \qquad \text{(K-36)}$$

$$x^2+0.5\left(\frac{b}{a}-\sqrt{\left(\frac{b}{a}\right)^2-4\frac{c}{a}+4y_1}\right)x+0.5\left(y_1+\sqrt{y_1^2-4\frac{e}{a}}\right)=0 \qquad \text{(K-37)}$$

The four zeros of these two quadratics are the zeros of the quartic equation K-34 provided you selected the correct zero y_1 from the cubic equation K-35. To verify that you selected the correct zero y_1 your four zeros must satisfy the following criteria:

$$x_1+x_2+x_3+x_4=-\frac{b}{a} \qquad \text{(K-38)}$$

$$x_1x_2+x_2x_3+x_3x_4+x_4x_1+x_1x_3+x_2x_4=\frac{c}{a} \qquad \text{(K-39)}$$

$$x_1x_2x_3+x_2x_3x_4+x_1x_2x_4+x_1x_3x_4=-\frac{d}{a} \qquad \text{(K-40)}$$

$$x_1x_2x_3x_4=\frac{e}{a} \qquad \text{(K-41)}$$

K.5 High-Order Polynomials (Advanced Topic)

For polynomials greater than fourth order, there are no general formulas for finding the zeros. One approach is to factor the original polynomial into lower-order polynomials. The zeros of those lower order polynomials will also be the zeros of the original polynomial. For example, the polynomial

$$x^3-6x^2+11x-6=0 \qquad \text{(K-42)}$$

may be factored into

$$(x-3)(x^2-3x+2)=0 \qquad \text{(K-43)}$$

The zero of the first term is obviously 3, and the quadratic formula gives 1 and 2 as the zeros of the other term. These three zeros are also the zeros of the original equation.

Factoring a large polynomial can be a difficult (or impossible) task, so a more general approach is required. The following generalized computer algorithms are able to find the zeros of large polynomials, or any desired equation.

K.6 Bisection Method

The *bisection method* may be used for any general equation

$$y=f(x) \qquad \text{(K-44)}$$

Two x values are selected; the smaller x is called x_{low} and the larger x is called x_{high}. If the x values have been properly selected, one produces a negative y and the other produces a positive y (see Figure K.1).

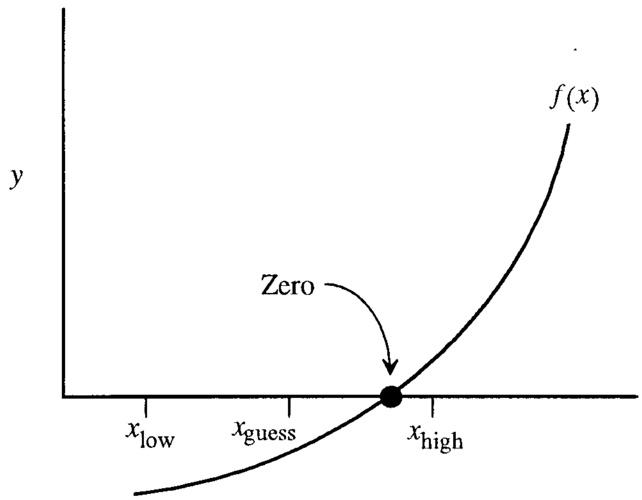

Figure K.1 Bisection method.

Clearly, the zero must exist between these two x values because y changed sign. The next step is to close in on the zero by narrowing the range of the x values. This is done by splitting the difference between the x values in half, i.e., bisecting them. The value of this midpoint x_{guess} is easily calculated as

$$x_{\text{guess}} = \frac{x_{\text{high}} + x_{\text{low}}}{2} \tag{K-45}$$

One of the halves will have a sign change, so it must have the zero. This segment with the zero is then divided in half. This process may be repeated until the zero x_1 is identified to arbitrary precision; that is, $x_{\text{guess}} \approx x_1$.

The computer algorithm that performs the bisection must handle four possible scenarios:

- Positive sloping curve with zero in right segment
- Positive sloping curve with zero in left segment
- Negative sloping curve with zero in right segment
- Negative sloping curve with zero in left segment

The four possibilities are presented in Figure K.2.

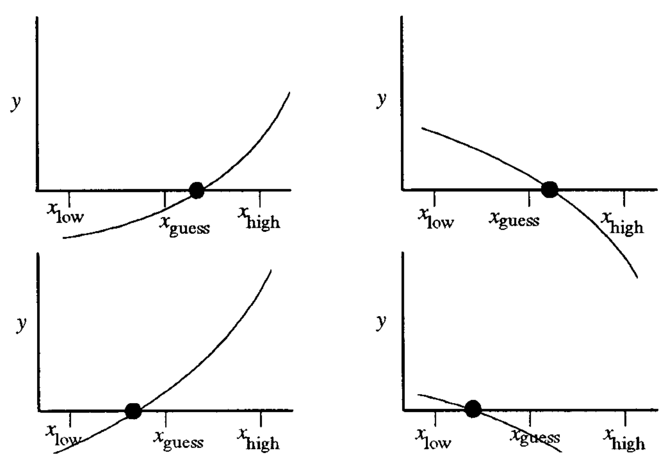

Figure K.2 The four possibilities in the bisection method.

The computer algorithm must decide if the midpoint x_{guess} should become the new x_{high} or x_{low}. The following table presents the proper actions:

	Positive slope: $f(x_{low})<0$	Negative slope: $f(x_{low})>0$
$f(x_{guess}) < 0$	$x_{low} = x_{guess}$	$x_{high} = x_{guess}$
$f(x_{guess}) > 0$	$x_{high} = x_{guess}$	$x_{low} = x_{guess}$
$f(x_{guess}) \approx 0$	$x_1 \approx x_{guess}$	$x_1 \approx x_{guess}$

Now all that is required is a method to decide whether the curve has a positive or negative slope. This is accomplished by observing that $f(x_{low})$ is negative for the positively sloping curve and $f(x_{low})$ is positive for the negatively sloping curve. (An alternative criteria is that $f(x_{high})$ is positive for positively sloping curves and $f(x_{high})$ is negative for negatively sloping curves.)

The *bisection method* has some potential pitfalls if you are not careful. Some mathematical functions have multiple zeros. If your initial guesses are widely spaced, there may be multiple zeros between them, but the method will find only one.

K.7 Inverse Linear Interpolation

The *inverse linear interpolation* method of finding zeros (also known as the *regula-falsi* method) is similar to the bisection method. However, it uses a more sophisticated approach to finding x_{guess}. Instead of simply dividing the range from x_{low} to x_{high} in half, it approximates the curve with a straight line, and uses the x intercept as x_{guess}. This is better understood by considering Figure K.3.

The straight line is defined by the two points (x_{low}, $f(x_{low})$) and (x_{high}, $f(x_{high})$). x_{guess} is equal to the x intercept a, which is known from the slope m and the y intercept b

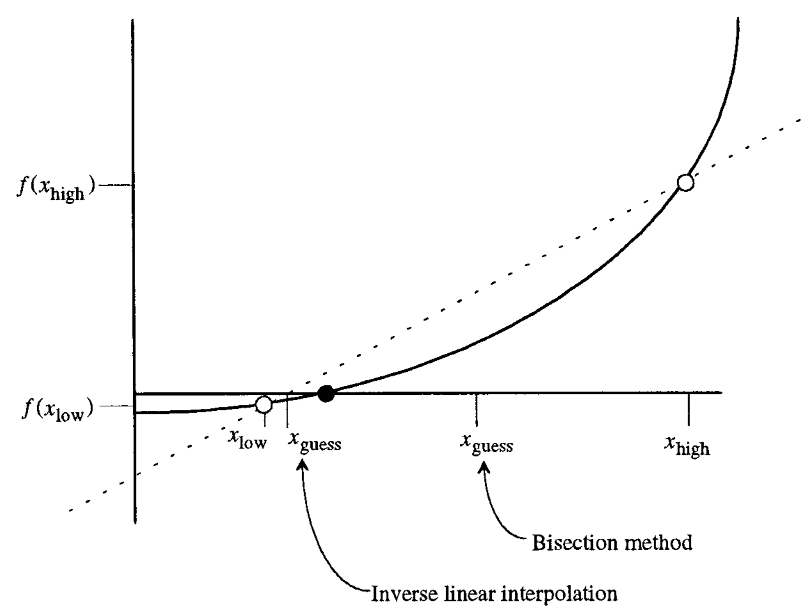

Figure K.3 Inverse linear interpolation.

$$x_{\text{guess}} = a = -\frac{b}{m} \qquad \text{(K-46)}$$

The slope is

$$m = \frac{f(x_{\text{high}}) - f(x_{\text{low}})}{x_{\text{high}} - x_{\text{low}}} \qquad \text{(K-47)}$$

and the y intercept is

$$b = f(x_{\text{high}}) - m x_{\text{high}} \qquad \text{(K-48)}$$

Substituting Equation K-48 for b in Equation K-46 gives

$$x_{\text{guess}} = -\frac{f(x_{\text{high}}) - m x_{\text{high}}}{m} = -\frac{f(x_{\text{high}})}{m} + x_{\text{high}} \qquad \text{(K-49)}$$

Substituting Equation K-47 for m gives

$$x_{\text{guess}} = -\left(\frac{x_{\text{high}} - x_{\text{low}}}{f(x_{\text{high}}) - f(x_{\text{low}})}\right) f(x_{\text{high}}) + x_{\text{high}} \qquad \text{(K-50)}$$

$$x_{\text{guess}} = \frac{f(x_{\text{high}})x_{\text{low}} - f(x_{\text{high}})x_{\text{high}} + f(x_{\text{high}})x_{\text{high}} - f(x_{\text{low}})x_{\text{high}}}{f(x_{\text{high}}) - f(x_{\text{low}})} \qquad \text{(K-51)}$$

$$x_{\text{guess}} = \frac{f(x_{\text{high}})x_{\text{low}} - f(x_{\text{low}})x_{\text{high}}}{f(x_{\text{high}}) - f(x_{\text{low}})} \qquad \text{(K-52)}$$

The inverse linear interpolation method uses all the procedures from the bisection method, except Equation K-45 is replaced with Equation K-52.

K.8 Summary

The zero of the equation is the x value that causes the equation to equal zero. In some cases, the zero may be found explicitly by algebraically manipulating the equation and solving for x. However, with many equations, no algebraic manipulations are able to isolate x so that implicit methods for finding the zero are necessary. Depending on the equation, there may be multiple values of x that cause the equation to equal zero.

There are algorithms for finding the zeros of polynomials up to fourth order. However, higher-order polynomials and other functions may require a numerical method to find the zero. Here, we presented two numerical methods for finding zeros: the bisection method and the inverse linear interpolation method.

Further Readings

M. S. Spiegel, *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series in Mathematics, McGraw-Hill, New York, 1968.

Handbook of Mathematical Formulas, Tables, Graphs, Functions, Transforms, Research and Education Association, New York, 1980.

Problems

K.1 Find the zeros of the following equations:

a. $3x+5=0$
b. $8x-3=0$

K.2 Write a computer program that finds the zeros of a linear equation. Use the two equations given in Problem K.1 as test cases.

K.3 Find the zeros of the following equations:

a. $x^2+10x-4=0$
b. $3x^2+6x+3=0$
c. $3x^2+6x+5=0$

K.4 Write a computer program that finds the zeros of a quadratic equation. Use the three equations given in Problem K.3 as test cases.

The program must test the value of the discriminant D to determine if it is positive, negative, or exactly zero. Highly positive or negative discriminants will cause no problems, but what if the calculated discriminant is close to zero? If the computer were able to represent numbers with an infinite number of digits, we could be assured that the calculated discriminant really was different from zero. However, real computers represent numbers with a finite number of digits. Therefore, the calculated discriminant might actually be zero, but the computer failed to report it as such because of roundoff error. This problem is solved by giving the computer a "bigger target" for zero. We can define positive, zero, and negative in the following way:

$$D > \varepsilon \text{ (positive)}$$

$$-\varepsilon < D < \varepsilon \text{ (zero)}$$

$$D < -\varepsilon \text{ (negative)}$$

By giving ε (epsilon) a very small value (say 10^{-7}), then we will not allow truly positive or negative discriminants to be called zero.

All the variables in your program should be declared as reals. What happens if the discriminant is negative, producing two complex zeros? In this case, the real part and the imaginary part of the complex number are calculated by using real numbers. In your computer output, you should separately report the real part of x_1, the imaginary part of x_1, the real part of x_2, and the imaginary part of x_2.

K.5 Write a computer program that finds the zeros of a cubic equation. Use the following three equations as test cases

$$x^3-x^2+x+3=0$$

$$x^3-7x^2+16x-12=0$$

$$x^3-6x^2+11x-6=0$$

The program must test the value of the discriminant Y to determine if it is positive, negative, or exactly zero. Highly positive or negative discriminants will cause no problems, but what if the calculated discriminant is close to zero? If the computer were able to represent numbers with an infinite number of digits, we could be assured that the calculated discriminant really was different from zero. However, real computers represent numbers with a finite number of digits. Therefore, the calculated discriminant might actually be zero, but the computer failed to report it as such because of roundoff error. This problem is solved by giving the computer a "bigger target" for zero. We can define positive, zero, and negative in the following way:

$$Y > \varepsilon \text{ (positive)}$$

$$-\varepsilon < Y < \varepsilon \text{ (zero)}$$

$$Y < -\varepsilon \text{ (negative)}$$

By giving ε (epsilon) a very small value (say 10^{-7}), then we will not allow truly positive or negative discriminants to be called zero. (*Note*: You will also need to define $Z = 0$ in the same way.)

All the variables in your program should be declared as reals. What happens if the discriminant is positive, producing two complex zeros? In this case, the real part and the imaginary part of the complex number are calculated by using real numbers. In your computer output, you should separately report the real part of x_2, the imaginary part of x_2, the real part of x_3, and the imaginary part of x_3.

K.6 Write a computer program that calculates the zero(s) of the following equations using the bisection method:

a. $f(x) = e^x - 10x = 0$
b. $f(x) = 3x^2 + \ln x = 0$
c. $f(x) = x^3 - 7x^2 + 16x - 12 = 0$
d. $f(x) = x^3 - x^2 + x + 3 = 0$
e. $f(x) = x^3 - 6x^2 + 11x - 6 = 0$
f. $f(x) = 2x + \sin 3x = 0$

To determine the number of zeros, plot these functions by hand or using a spreadsheet. For each zero, choose your initial guess to be on either side so one initial guess is a positive $f(x)$ and the other is a negative $f(x)$. Stop the program when a successive iteration (i.e., cycle through the loop) changes the estimate of the zero by less than 0.0001% (1 part in a million). Report your initial guesses and the number of iterations it takes to find the zero.

K.7 Write a computer program that calculates the zero(s) of the equations listed in Problem K.6 using the inverse linear interpolation (regula falsi) method. To determine the number of zeros, plot these functions by hand or using a spreadsheet. For each zero, choose your initial guess to be on either side so one initial guess is a positive $f(x)$ and the other is a negative $f(x)$. Stop the program when a successive iteration (i.e., cycle through the loop) changes the estimate of the zero by less than 0.0001% (1 part in a million). Report your initial guesses and the number of iterations it takes to find the zero. If you also did Problem K.6, compare the number of iterations required by each technique.