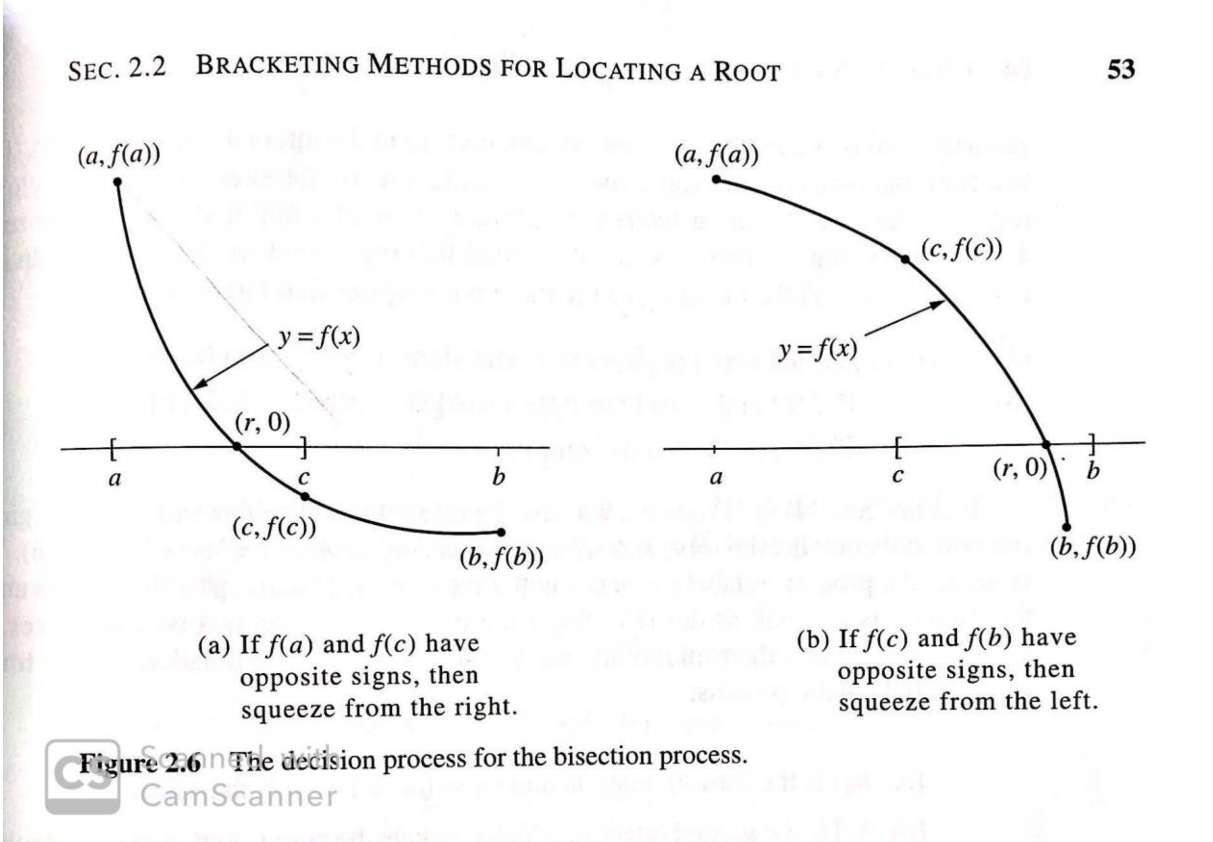
JOHN H. MATHEWS, KURTIS D. FINK, NUMERICAL METHODS USING MATLAB, FOURTH EDITION, PEARSON EDUCATION INTERNATIONAL, ISBN 0-13-191178-3, X+680 PAGES, 2004.

2 Solution of Nonlinear Equations $f(x)=0$

Definition 2.3. Assume that $f(x)$ is a continuous function. Any number r for which $f(r)=0$ is called a *root of the equation* $f(x)=0$. Also, we say that r is a zero of the function $f(x)$.

Bisection Method of Bolzano

In this section we develop our first bracketing method for finding a zero of a continuous function. We must start with an initial interval $[a,b]$, where $f(a)$ and $f(b)$ have opposite signs. Since the graph $y=f(x)$ of a continuous function is unbroken, it will cross the $x-$ axis at a zero $x=r$ that lies somewhere in the interval (see Figure 2.6).



SEC. 2.2 BRACKETING METHODS FOR LOCATING A ROOT 53

(a) If $f(a)$ and $f(c)$ have opposite signs, then squeeze from the right.

(b) If $f(c)$ and $f(b)$ have opposite signs, then squeeze from the left.

Figure 2.6 The decision process for the bisection process.

The bisection method systematically moves the endpoints of the interval closer and closer together until we obtain an interval of arbitrarily small width that brackets the zero. The decision step for this process of interval halving is to choose the midpoint $c=(a+b)/2$ and then to analyze the three possibilities that might arise:

(4) If $f(a)$ and $f(c)$ have opposite signs, a zero lies in $[a,c]$.

(5) If $f(c)$ and $f(b)$ have opposite signs, a zero lies in $[c,b]$.

(6) If $f(c)=0$, then the zero is c.

If either case (4) or (5) occurs, we have found an interval half as wide as the original interval that contains the root, and we are "sequeezing down on it" (see Figure 2.6). To continue the process, relabel the new smaller interval $[a,b]$ and repeat the process until the interval is as small as desired. Since the bisection process involves sequences of nested intervals and their midpoints, we will use the following notation to keep track of the details in the process.

$[a_0,b_0]$ is the starting interval and $c_0=(a_0+b_0)/2$ is the midpoint.

(7) $[a_1,b_1]$ is the second interval, which brackets the zero r, and c_1 is its midpoint; the interval

$[a_1,b_1]$ is half as wide as $[a_0,b_0]$.

After arriving at the $n-$ th interval $[a_n,b_n]$, which brackets r and has midpoint c_n, the interval $[a_{n+1},b_{n+1}]$ is constructed, which also brackets r and is half as wide as $[a_n,b_n]$.

The sequence of the left endpoints is increasing and the sequence of the right endpoints is decreasing; that is,

(8) $$a_0\le a_1\le\cdots\le a_n\le\cdots\le r\le\cdots\le b_n\le\cdots\le b_1\le b_0,$$

where $c_n=(a_n+b_n)/2$, and if $f(a_{n+1})f(b_{n+1})<0$, then

(9) $$[a_{n+1},b_{n+1}]=[a_n,c_n] \text{ or } [a_{n+1},b_{n+1}]=[c_n,b_n] \text{ for all } n.$$

Theorem 2.4 (Bisection Theorem). Assume that $f\in C[a,b]$ and that there exists a number $r\in[a,b]$ such that $f(r)=0$. If $f(a)$ and $f(b)$ have opposite signs, and $\{c_n\}_{n=0}^{\infty}$ represents the sequence of midpoints generated by the bisection process of (8) and (9), then

(10) $\left|r-c_n\right|\le\dfrac{b-a}{2^{n+1}}$ for $n=0,1,\cdots,$

and therefore the sequence $\left\{c_n\right\}_{n=0}^{\infty}$ converges to the zero $x=r$, that is,

(11) $\lim\limits_{n\to\infty}c_n=r$.

Example 2.7. The function $h\left(x\right)=x\sin x$ occurs in the study of undamped forced oscillations. Find the value of x that lies in the interval $\left[0,2\right]$, where the function takes on the value $h\left(x\right)=1$ (the function $\sin x$ is evaluated in radians).

We use the bisection method to find a zero of the function $f\left(x\right)=x\sin x-1$. Starting with $a_0=0$ and $b_0=2$, we compute

$$f\left(0\right)=-1.000000 \text{ and } f\left(2\right)=0.818595,$$

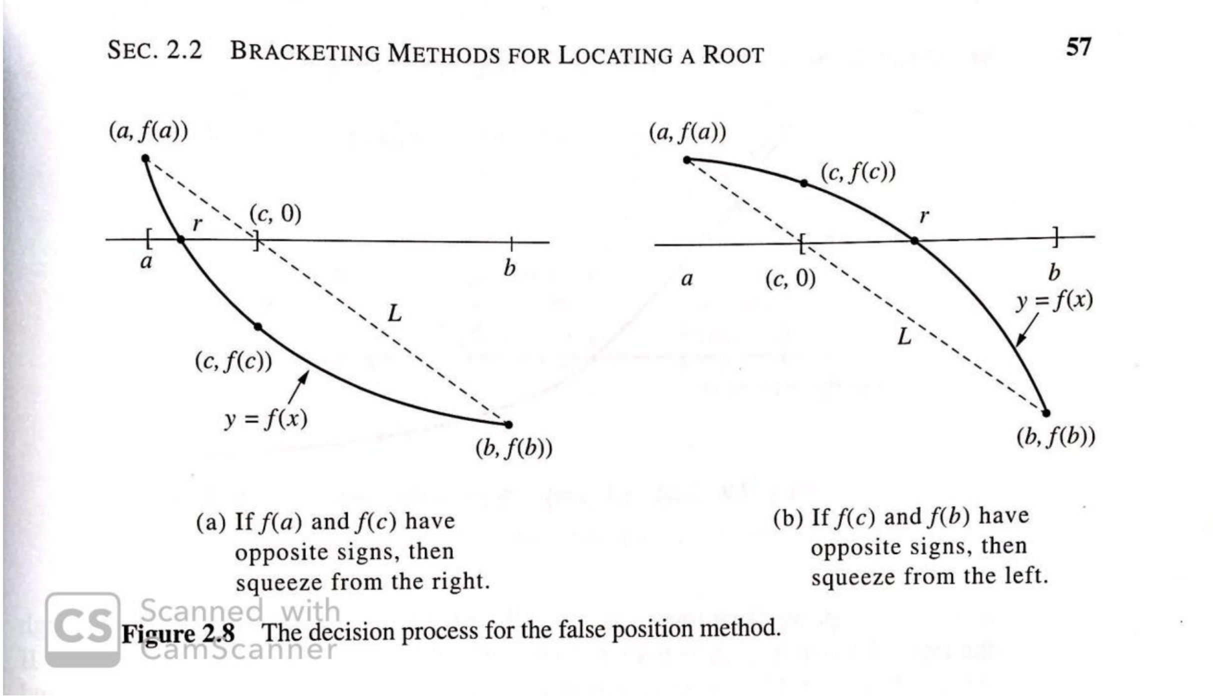
so a root of $f\left(x\right)=0$ lies in the interval $\left[0,2\right]$. At the midpoint $c_0=1$, we find that $f\left(1\right)=-0.158529$. Hence the function changes sign on $\left[c_0,b_0\right]=\left[1,2\right]$.

To continue, we squeeze from the left and set $a_1=c_0$ and $b_1=b_0$. The midpoint is $c_1=1.5$ and $f\left(c_1\right)=0.496242$. Now, $f\left(1\right)=-0.158529$ and $f\left(1.5\right)=0.496242$ imply that the root lies in the interval $\left[a_1,c_1\right]=\left[1.0,1.5\right]$. The next decision is to sequeeze from the right and set $a_2=a_1$ and $b_2=c_1$. In this manner we obtain a sequence $\left\{c_k\right\}$ that converges to $r\approx1.114157141$. A sample calculation is given in Table 2.1. □

Table 2.1 Bisection Method Solution of $x\sin x-1=0$

k	Left endpoint, a_k	Midpoint, c_k	Right endpoint, b_k	Function value, $f\left(c_k\right)$
0	0	1.	2.	–0.158529
1	1.0	1.5	2.0	L 0.496242
2	1.00	1.25	1.50	0.186231
3	1.000	1.125	1.250	0.015051
4	1.000	1.0625	1.1250	–0.071827
5	1.06250	1.09375	1.12500	–0.028362
6	1.093750	1.109375	1.125000	–0.006643
7	1.1093750	1.1171875	1.1250000	0.004208
8	1.10937500	1.11328125	1.11718750	–0.001216
⋮	⋮	⋮	⋮	⋮

Another popular algorithm is the method of false position or the regula falsi method. It was developed because the bisection method converges at a fairly slow speed. As before, we assume that $f(a)$ and $f(b)$ have opposite signs. The bisection method used the miidpoint of the interval $[a,b]$ as the next iterate. A better approximation is obtained if we find the point $(c,0)$ where the secant line L joining the points $(a, f(a))$ and $(b, f(b))$ crosses the $x-$ axis (see Figure 2.8).



SEC. 2.2 BRACKETING METHODS FOR LOCATING A ROOT 57

(a) If $f(a)$ and $f(c)$ have opposite signs, then squeeze from the right.

(b) If $f(c)$ and $f(b)$ have opposite signs, then squeeze from the left.

Figure 2.8 The decision process for the false position method.

To find the value c, we write down two versions of the slope m of the line L:

(16) $$m = \frac{f(b) - f(a)}{b - a},$$

where the points $(a, f(a))$ and $(b, f(b))$ are used, and

(17) $$m = \frac{0 - f(b)}{c - b},$$

where the points $(c,0)$ and $(b, f(b))$ are used.

Equating the slopes in (16) and (17), we have

$$\frac{f(b) - f(a)}{b - a} = \frac{0 - f(b)}{c - b},$$

which is easily solved for c to get

(18) $$c = b - \frac{f(b)(b-a)}{f(b) - f(a)}.$$

The three possibilities are the same as before:

(19) If $f(a)$ and $f(c)$ have opposite signs, a zero lies in $[a,c]$.

(20) If $f(c)$ and $f(b)$ have opposite signs, a zero lies in $[c,b]$.

(21) If $f(c) = 0$, then the zero is c.

Convergence of the False Position Method

The decision process implied by (19) and (20) along with (18) is used to construct a sequence of intervals $\{[a_n, b_n]\}$ each of which brackets the zero. At each step the approximation of the zero r is

(22) $$c_n = b_n - \frac{f(b_n)(b_n - a_n)}{f(b_n) - f(a_n)},$$

and it can be proved that the sequence $\{c_n\}$ will converge to r. But beware; although the interval width $b_n - a_n$ is getting smaller, it is possible that it may not go to zero. If the graph of $y = f(x)$ is concave near $(r,0)$, one of the endpoints becomes fixed and the other one marches into the solution (see Figure 2.9).

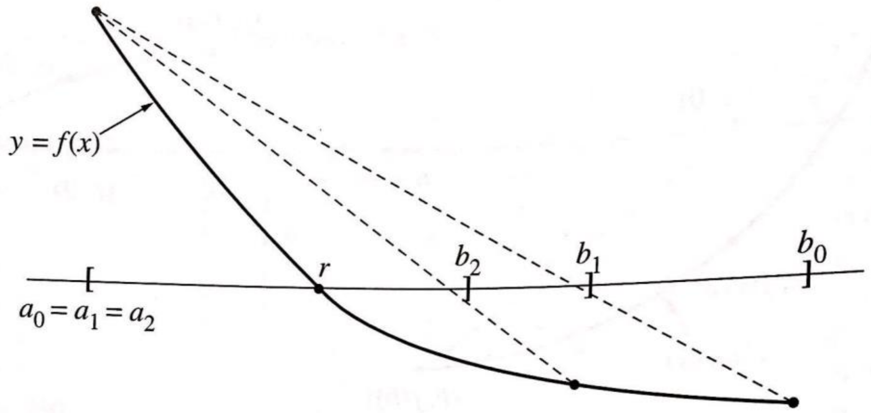


Figure 2.9 The stationary endpoint for the false position method.

Now we rework the solution to $x\sin x - 1 = 0$ using the method of false position and observe that it converges faster than the bisection method. Also, notice that $\{b_n - a_n\}_{n=0}^{\infty}$ does not go to zero.

Example 2.8. Use the false position method to find the root of $x\sin x - 1 = 0$ that is located in the interval $[0,2]$ (the function $\sin x$ is evaluated in radians).

Starting with $a_0 = 0$ and $b_0 = 2$, we have $f(0) = -1.00000000$ and $f(2) = 0.81859485$, so a root lies in the interval $[0,2]$. Using formula (22), we get

$$c_0 = 2 - \frac{0.81859485(2-0)}{0.81859485-(-1)} = 1.09975017 \text{ and } f(c_0) = -0.02001921.$$

The function changes sign on the interval $[c_0, b_0] = [1.09975017, 2]$, so we squeeze from the left and set $a_1 = c_0$ and $b_1 = b_0$. Formula (22) produces the next approximation:

$$c_1 = 2 - \frac{0.81859485(2-1.09975017)}{0.81859485-(-0.02001921)} = 1.12124074$$

and

$$f(c_1) = 0.00983461.$$

Next $f(x)$ changes sign on $[a_1,c_1]=[1.09975017,1.12124074]$, and the next decision is to sequeeze from the right and set $a_2=a_1$ and $b_2=c_1$. A summary of the calculations is given in Table 2.2. □

Table 2.2 False Position Method Solution of $x\sin x-1=0$

k	Left endpoint, a_k	Midpoint, c_k	Right endpoint, b_k	Function value, $f(c_k)$
0	0.00000000	1.09975017	2.00000000	–0.02001921
1	1.09975017	1.12124074	2.00000000	0.00983461
2	1.09975017	1.11416120	1.12124074	0.00000563
3	1.09975017	1.11415714	1.11416120	0.00000000

Program 2.2 (Bisection Method) To approximate a root of the equation $f(x)=0$ in the interval $[a,b]$. Proceed with the method only if $f(x)$ is continuous and $f(a)$ and $f(b)$ have opposite signs.

```
function [c,err,yc]=bisec(f,a,b,delta)
%Input   – f is the function input as a string ‘f’
%          – a and b are the left and right endpoints
%          – delta is the tolerance
% Output – a and b are the left and right endpoints
%          –yc=f(c)
%          –err is the error estimate for c
ya=feval(f,a);
yb=feval(f,b);
if ya*yb>0,break,end
max1=1+round(log(b-a)-log(delta))/log(2));
for k=1:max1
    c=(a+b)/2;
    yc=feval(f,c);
    if yc==0
      a=c;
      b=c;
elseif yb*yc>0
      b=c;
      yb=yc;
  else
      a=c;
      ya=yc;
  end
  if b-a<delta, break, end
end
c=(a+b)/2;
err=abs(b-a);
yc=feval(f,c);
```

Program 2.3 (False Position or Regula Falsi Method) To approximate a root of the equation $f(x)=0$ in the interval $[a,b]$. Proceed with the method only if $f(x)$ is continuous and $f(a)$ and $f(b)$ have opposite signs.

```
function [c,err,yc]=regula(f,a,b,delta,epsilon,max1)
%Input    – f is the function input as a string ‘f’
%           – a and b are the left and right endpoints
%           – delta is the tolerance for the zero
%           – epsilon is the tolerance for the value of f at the zero
%           – max1 is the maximum number of iterations
% Output – c is the zero
%           –yc=f(c)
%           –err is the error estimate for c
ya=feval(f,a);
yb=feval(f,b);
    disp(‘Note: f(a)*f(b)>0’),
    break,
end
for k=1:max1
    dx=yb*(b-a)/(yb-ya);
    c=b-dx;
    ac=c-a;
    yc=feval(f,c);
    if yc==0, break;
    elseif yb*yc>0
      b=c;
      yb=yc;
    else
      a=c;
      ya=yc;
    end
    dx=min(abs(dx),ac);
    if abs(dx)<delta,break,end
    if abs(yc)<epsilon,break,end
end
c;
err=abs(b-a);
yc=feval(f,c);
```

ASSIGNMENT

Modify Programs 2.2 and 2.3 to output a matrix analogous to Tables 2.1 and 2.2, respectively (i.e., the first row of the matrix would be [$0\ \ a_0\ \ c_0\ \ b_0\ \ f(c_0)$]).