Chapter2: Solving Nonlinear Equations

2.1 BACKGROUND

Equations need to be solved in all areas of science and engineering. An equation of one variable can be written in the form:

$$f(x) = 0 \qquad (2.1)$$

A solution to the equation (also called a root of the equation) is a numerical value of x that satisfies the equation. Graphically, as shown in Fig. 2-1, the solution is the point where the function *f(x)* crosses or touches the x-axis. An equation might have no solution or can have one or several (possibly many) roots. When the equation is simple, the value of *x* can be determined analytically. This is the case when x can be written explicitly by applying mathematical operations, or when a known formula (such as the formula for solving a quadratic equation) can be used to determine the exact value of x. In many situations, however, it is impossible to determine the root of an equation analytically.

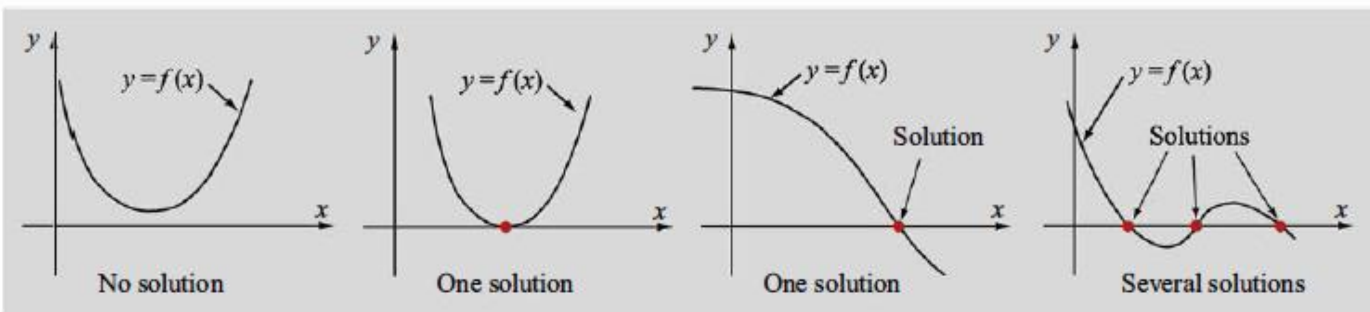


Figure 2-1: Illustration of equations with no, one, or several solutions.

Overview of approaches in solving equations numerically

The process of solving an equation numerically is different from the procedure used to find an analytical solution. An analytical solution is obtained by deriving an expression that has an exact numerical value. A numerical solution is obtained in a process that starts by finding an approximate solution and is followed by a numerical procedure in which a better (more accurate) solution is determined.

An initial numerical solution of an equation *f(x) = 0* can be estimated by plotting *f(x)* versus x and looking for the point where the graph crosses the *x*-axis.

It is also possible to write and execute a computer program that looks for a domain that contains a solution. Such a program looks for a solution by evaluating *f(x)* at different values of x. It starts at one value of x and then changes the value of *x* in small increments. A change in the sign of *f(x)* indicates that there is a root within the last increment. In most cases, when the equation that is solved is related to an application in science or engineering, the range of *x* that includes the solution can be estimated and used in the initial plot of *f(x)*, or for a numerical search of a small domain that contains a solution. When an equation has more than one root, a numerical solution is obtained one root at a time.

The methods used for solving equations numerically can be divided into two groups: bracketing methods and open methods.

In bracketing methods, illustrated in Fig. 2-2, an interval that includes the solution is identified. By definition, the endpoints of the interval are the upper bound and lower bound of the solution. Then, by using a numerical scheme, the size of the interval is successively reduced until the distance between the endpoints is less than the desired accuracy of the solution. In open methods, illustrated in Fig. 2-3, an initial estimate (one point) for the solution is assumed. The value of this initial guess for the solution should be close to the actual solution. Then, by using a numerical scheme, better (more accurate) values for the solution are calculated. Bracketing methods always converge to the solution. Open methods are usually more efficient but sometimes might not yield the solution.

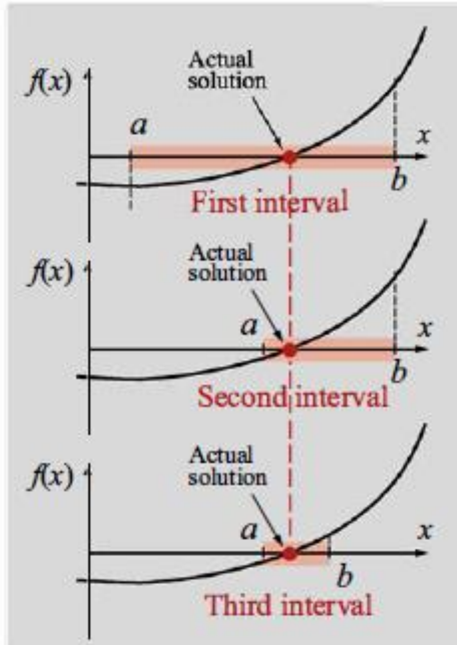


Figure 2-2: Illustration of a bracketing method.

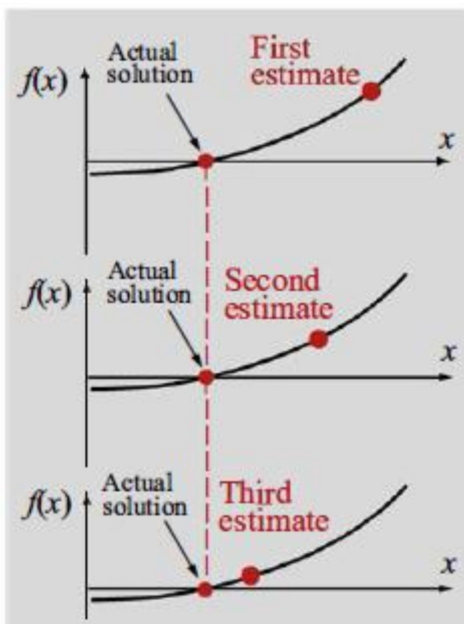


Figure 2-3: Illustration of an open method.

2.2 ESTIMATION OF ERRORS IN NUMERICAL SOLUTIONS

Since numerical solutions are not exact, some criterion has to be applied in order to determine whether an estimated solution is accurate enough. Several measures can be used to estimate the accuracy of an approximate solution. The decision as to which measure to use depends on the application and has to be made by the person solving the equation. Let x_{rs} be the true (exact) solution such that $f(x_{rs}) = 0$, and let x_{Ns} be a numerically approximated solution such that $f(x_{Ns}) = E$ (where E is a small number). Four measures that can be considered for estimating the error are:

2.2.1 True error

The true error is the difference between the true solution, X_{rs} and a numerical solution, X_{Ns}:

$$TrueError = X_{rs} - X_{Ns} \qquad (2.2)$$

Unfortunately, however, the true error cannot be calculated because the true solution is generally not known.

2.2.2 Tolerance in *f(x)*:

Instead of considering the error in the solution, it is possible to consider the deviation of $f(x_{Ns})$ from zero (the value of $f(x)$ at x_{rs} is obviously zero). The tolerance in $f(x)$ is defined as the absolute value of the difference between $f(x_{rs})$ and $f(x_{Ns})$:

$$ToleranceInf = |f(x_{rs}) - f(x_{Ns})| = |0 - \varepsilon| = |\varepsilon| \qquad (2.3)$$

The tolerance in $f(x)$ then is the absolute value of the function at x_{Ns}.

2.2.3 Tolerance in the solution:

Tolerance is the maximum amount by which the true solution can deviate from an approximate numerical solution. A tolerance is useful for estimating the error when bracketing methods are used for determining the numerical solution. In this case, if it is known that the solution is within the domain [a, b] , then the numerical solution can be taken as the midpoint between a and b:

$$x_{Ns} = \frac{a+b}{2} \qquad (2.4)$$

plus or minus a tolerance that is equal to half the distance between a and b:

$$\text{Tolerance} = \frac{b-a}{2} \qquad (2.5)$$

2.2.4 Relative error:

If x_{Ns} is an estimated numerical solution, then the True Relative Error is given by:

$$\text{TrueRelativeError} = \left|\frac{x_{rs} - x_{Ns}}{x_{Ns}}\right| \qquad (2.6)$$

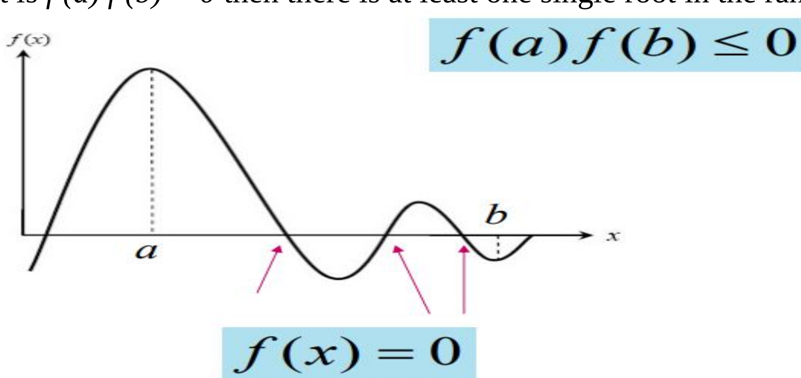
This True Relative Error cannot be calculated since the true solution x_{rs} is not known. Instead, it is possible to calculate an Estimated Relative Error when two numerical estimates for the solution are known. This is the case when numerical solutions are calculated iteratively, wherein each new iteration a more accurate solution is calculated. If $x_{Ns}^{(n)}$ is the estimated numerical solution in the last iteration and $x_{Ns}^{(n-1)}$ is the estimated numerical solution in the preceding iteration, then an Estimated Relative Error can be defined by:

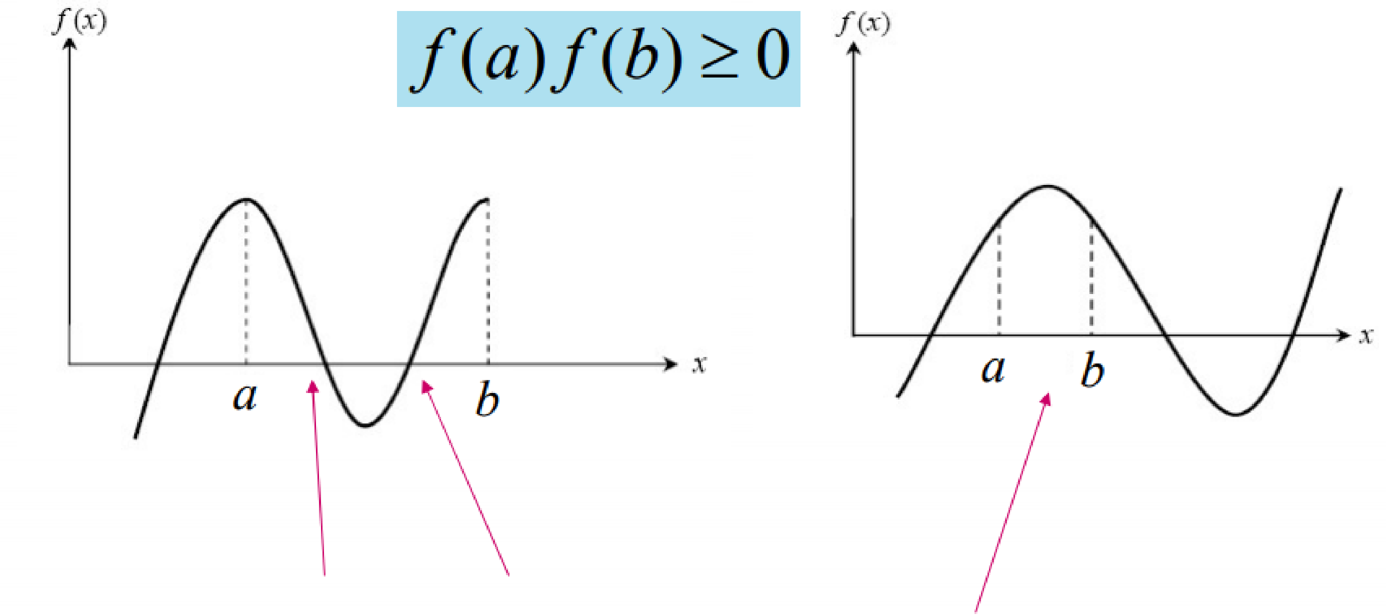
$$\text{Estimated Relative Error} = \left|\frac{x_{Ns}^{(n)} - x_{Ns}^{(n-1)}}{x_{Ns}^{(n-1)}}\right| \qquad (2.7)$$

When the estimated numerical solutions are close to the true solution it is anticipated that the difference $x_{Ns}^{(n)} - x_{Ns}^{(n-1)}$ is small compared to the value of $x_{Ns}^{(n)}$, and the Estimated Relative Error is approximately the same as the True Relative Error.

2.3 Root-finding algorithms

Theorem: If the function $f(x)$ is defined and continuous in the range [a,b] and function changes sign at the ends of the interval that is $f(a)\, f(b) < 0$ then there is at least one single root in the range [a,b].





If the function does not change the sign between two points, there may not be or there may exist roots for this equation between the two points.

Root-finding strategy

- Plot the function (the plot provides an initial guess, and indication of potential problems).
- Isolate single roots in separate intervals (bracketing).
- Select an initial guess.
- Iteratively refine the initial guess with a root-finding algorithm, i.e. generate the sequence :

$$\{x_i\}_{i=0}^{n} : \lim_{n\to\infty}(x_n - \alpha) = 0$$

<u>EXAMPLE 2.1</u>

Find the largest root of $f(x) = x^6 - x - 1 = 0$.

x	-2	-1	0	1	2	3	4
$f(x)$	65	1	-1	-1	61	725	4091

It is obvious that the largest root of this equation is in the interval [1.2].

2.4 BISECTION METHOD

The bisection method is a bracketing method for finding a numerical solution of an equation of the form $f(x) = 0$ when it is known that within a given interval [a, b], $f(x)$ is continuous and the equation has a solution. When this is the case, $f(x)$ will have opposite signs at the endpoints of the interval. As shown in Fig. 2-4, if $f(x)$ is continuous and has a solution between the points x = a and x = b , then either $f(a) > 0$ and $f(b) < 0$ or $f(a) < 0$ and $f(b) > 0$. In other words, if there is a solution between x=a and x = b, then $f(a)f(b) < 0$.

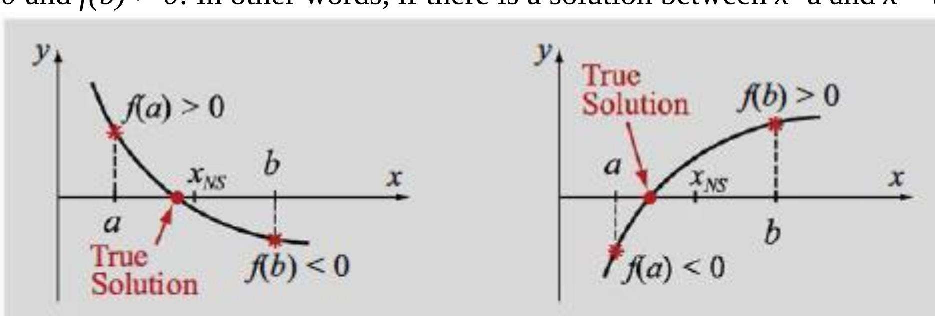


Figure 2-4: Solution of *f(x) = 0* between *x* =a and *x* = b.

Algorithm for the bisection method

1. Choose the first interval by finding points a and b such that a solution exists between them. This means that *f(a)* and *f(b)* have different signs such that *f(a)f(b) < 0*. The points can be determined by examining the plot of f(x) versus x.
2. Calculate the first estimate of the numerical solution x_{Ns1} by:

$$x_{Ns1} = \frac{a+b}{2}$$

3. Determine whether the true solution is between a and x_{NS1} or between x_{Ns1} and b. This is done by checking the sign of the product $f(a) \cdot f(x_{Ns1})$:
If $f(a) \cdot f(x_{Ns1}) < 0$, the true solution is between a and x_{Ns1}.
If $f(a) \cdot f(x_{Ns1}) > 0$, the true solution is between x_{Ns1} and b.
4. Select the subinterval that contains the true solution (a to x_{Ns1}, or x_{Ns1} to b) as the new interval [a, b], and go back to step 2.
Steps 2 through 4 are repeated until a specified tolerance or error bound is attained.

When should the bisection process be stopped?

Ideally, the bisection process should be stopped when the true solution is obtained. This means that the value of x_{Ns} is such that $f(x_{Ns}) = 0$. In reality, as discussed in Section 2.1, this true solution generally cannot be found computationally. In practice, therefore, the process is stopped when the estimated error, according to one of the measures listed in Section 2.2, is smaller than some predetermined value. The choice of termination criteria may depend on the problem that is actually solved.

Additional notes on the bisection method

- The method always converges to an answer, provided a root was trapped in the interval [a, b] to begin with.
- The method may fail when the function is tangent to the axis and does not cross the x-axis at f(x) = 0.
- The method converges slowly relative to other methods.

EXAMPLE 2.2

Find the largest root of $f(x) = x^6 - x - 1 = 0$ accurate to within $\in= 0.001$.

Solution With a graph, it is easy to check that $1 < \alpha < 2$. We choose a = 1, b =2; then $f(a) = -1, f(b) = 61$, and the requirement $f(a) f(b) < 0$ is satisfied. The results from Bisect are shown in the table. The entry n indicates the iteration number n.

n	a	b	c	$b-c$	$f(c)$
1	1.0000	2.0000	1.5000	0.5000	8.8906
2	1.0000	1.5000	1.2500	0.2500	1.5647
3	1.0000	1.2500	1.1250	0.1250	−0.0977
4	1.1250	1.2500	1.1875	0.0625	0.6167
5	1.1250	1.1875	1.1562	0.0312	0.2333
6	1.1250	1.1562	1.1406	0.0156	0.0616
7	1.1250	1.1406	1.1328	0.0078	−0.0196
8	1.1328	1.1406	1.1367	0.0039	0.0206
9	1.1328	1.1367	1.1348	0.0020	0.0004
10	1.1328	1.1348	1.1338	0.00098	−0.0096

Example 2.3 Show that $f(x) = x^3 + 4x^2 - 10 = 0$ has a root in [1, 2], and use the Bisection method to determine an approximation to the root that is accurate to at least within10^{-4}.

Solution Because $f(1) = -5$ and $f(2) = 14$, the Intermediate Value Theorem ensures that this continuous function has a root in [1, 2].

For the first iteration of the Bisection method we use the fact that at the midpoint of [1,2] we have $f(1.5) = 2.375 > 0$. This indicates that we should select the interval [1,1.5] for our second iteration. Then we find that $f(1.25) = -1.796875$ so our new interval becomes [1.25, 1.5], whose midpoint is 1.375. Continuing in this

manner gives the values in the following table. After 13 iterations, p_{13} = 1.365112305 approximates the root p with an error:
$|p - p_{13}| < |b_{14} - a_{14}| = |1.365234375 - 1.365112305| = 0.000122070$.
Since $|a_{14}| < |p|$, we have $|p - p_{13}|/|p| < |b_{14} - a_{14}|/|a_{14}| \leq 9.0 \times 10^{-5}$,

n	a_n	b_n	p_n	$f(p_n)$
1	1.0	2.0	1.5	2.375
2	1.0	1.5	1.25	−1.79687
3	1.25	1.5	1.375	0.16211
4	1.25	1.375	1.3125	−0.84839
5	1.3125	1.375	1.34375	−0.35098
6	1.34375	1.375	1.359375	−0.09641
7	1.359375	1.375	1.3671875	0.03236
8	1.359375	1.3671875	1.36328125	−0.03215
9	1.36328125	1.3671875	1.365234375	0.000072
10	1.36328125	1.365234375	1.364257813	−0.01605
11	1.364257813	1.365234375	1.364746094	−0.00799
12	1.364746094	1.365234375	1.364990235	−0.00396
13	1.364990235	1.365234375	1.365112305	−0.00194

so the approximation is correct to at least within 10^{-4}. The correct value of p to nine decimal places is p = 1.365230013. Note that p_9 is closer to p than is the final approximation p_{13}. You might suspect this is true because $|f(p_9)| < |f(p_{13})|$, but we cannot be sure of this unless the true answer is known.

Example 2.4 Use the Bisection method to find a root of the equation $x^3 - 4x - 8.95 = 0$ accurate to three decimal places using the Bisection Method.

Solution

Here, $f(x) = x^3 - 4x - 8.95 = 0$
$f(2) = 2^3 - 4(2) - 8.95 = -8.95 < 0$
$f(3) = 3^3 - 4(3) - 8.95 = 6.05 > 0$
Hence, a root lies between 2 and 3.
Hence, we have a = 2 and b = 3. The results of the algorithm for Bisection method are shown in Table.

n	a	b	x_{S_1}	$b-x_{S_1}$	$f(x_{S_1})$
0	2	3	2.5	0.5	-3.324999999999999
1	2.5	3	2.75	0.25	0.846875000000001
2	2.5	2.75	2.625	0.125	-1.362109374999999
3	2.625	2.75	2.6875	0.0625	-0.289111328124999
4			2.71875	0.03125	0.270916748046876
5			2.703125	0.015625	-0.011077117919921
6			2.7109375	0.007812	0.129423427581788
7			2.7070312	0.003906	0.059049236774445
8			2.7050781	0.001953	0.023955102264882
9			2.7041016	0.000976	0.006431255675853
10			2.7036133	0.000488	-0.002324864896945
11			2.7038574	0.000244	0.002052711902071
12			2.7037354	0.000122	-0.000136197363826
13			2.7037964	0.000061	0.000958227051843

Hence the root is 2.703 accurate to three decimal places.

2.5 FALS POSITION METHOD

The false position method (also called regula falsi and linear interpolation methods) is a bracketing method for finding a numerical solution of an equation of the form *f(x) = 0* when it is known that, within a given interval [a, b], *f(x)* is continuous and the equation has a solution. As illustrated in Fig. 2-5, the solution starts by finding an initial interval [a_1, b_1] that brackets the solution. The values of the function at the endpoints are $f(a_1)$ and $f(b_1)$. The endpoints are then connected by a straight line, and the first estimate of the numerical solution, x_{Ns1}, is the point where the straight line crosses the *x*-axis. This is in contrast to the bisection method, where the midpoint of the interval was taken as the solution. For the second iteration a new interval, [a_2, b_2] is defined. The new interval is a subsection of the first interval that contains the solution. It is either [a_1, x_{Ns1}] (a_1 is assigned to a_2, and x_{Ns1} to b_2) or [x_{Ns1}, b_1] (x_{Ns1} is assigned to a_2, and b_1 to b_2). The endpoints of the second interval are next connected with a straight line, and the point where this new line crosses the *x*-axis is the second estimate of the solution, x_{Ns2}. For the third iteration, a new subinterval [a_3, b_3] is selected, and the iterations continue in the same way until the numerical solution is deemed accurate enough.

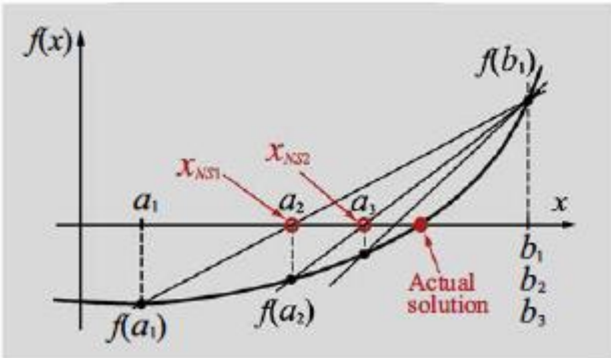


Figure 2-5: False position method

For a given interval [a, b], the equation of a straight line that connects point *(b, f(b))* to point *(a, f(a))* is given by:

$$y = \frac{f(b)-f(a)}{b-a}(x-b) + f(a) \qquad (2.8)$$

The point x_{NS} where the line intersects the *x*-axis is determined by substituting *y=0* in Eq. (2.8), and solving the equation for *x* :

$$x_{Ns} = \frac{af(b)-bf(a)}{f(b)-f(a)} \qquad (2.9)$$

The procedure (or algorithm) for finding a solution with the regula falsi method is almost the same as that for the bisection method**.**

Algorithm for the regula falsi method

1. Choose the first interval by finding points a and b such that a solution exists between them. This means that *f(a)* and *f(b)* have different signs such that *f(a)f(b) < 0*. The points can be determined by looking at a plot of *f(x)* versus *x*.
2. Calculate the first estimate of the numerical solution x_{Ns1} by using Eq. (2.9).
3. Determine whether the actual solution is between a and x_{Ns1} or between x_{Ns1} and b. This is done by checking the sign of the product $f(a) \cdot f(x_{Ns1})$:
If $f(a) \cdot f(x_{Ns1}) < 0$, the solution is between a and x_{Ns1}.
If $f(a) \cdot f(x_{Ns1}) > 0$, the solution is between x_{Ns1} and b.
4. Select the subinterval that contains the solution (a to x_{Ns1}, or x_{Ns1} to b) as the new interval [a, b], and go back to step 2.

Steps 2 through 4 are repeated until a specified tolerance or error bound is attained.

When should the iterations be stopped?

The iterations are stopped when the estimated error, according to one of the measures listed in Section 2 .2, is smaller than some predetermined value.

Additional notes on the regula falsi method

• The method always converges to an answer, provided a root is initially trapped in the interval [a, b].

• Frequently, as in the case shown in Fig. 2-5, the function in the interval [a, b] is either concave up or concave down. In this case, one of the endpoints of the interval stays the same in all the iterations, while the other endpoint advances toward the root. In other words, the numerical solution advances toward the root only from one side. The convergence toward the solution could be faster if the other endpoint would also "move" toward the root. Several modifications have been introduced to the regula falsi method that makes the subinterval in successive iterations approach the root from both sides.

Example 2.5

Using the False Position method, find a root of the function $f(x) = e^x - 3x^2$ to an accuracy of 5 digits. The root is known to lie between 0.5 and 1.0.

Solution

We apply the method of False Position with $a = 0.5$ and $b = 1.0$ and equation (2.2) which is:

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

The calculations based on the method of False Position are shown in the following table:

n	a	b	f(a)	f(b)	x_{s_1}	$f(x_{s_1})$	ξ Relative error
1	0.5	1	0.89872	–0.28172	0.88067	0.08577	—
2	0.88067	1	0.08577	–0.28172	0.90852	0.00441	0.03065
3	0.90852	1	0.00441	–0.28172	0.90993	0.00022	0.00155
4	0.90993	1	0.00022	–0.28172	0.91000	0.00001	0.00008
5	0.91000	1	0.00001	–0.28172	0.91001	0	3.7952×10^{-6}

The relative error after the fifth step is

$$\left(\frac{0.91001 - 0.91}{0.91001}\right) = 3.7952 \times 10^{-6}.$$

The root is 0.91 accurate to five digits.

Example 2.6

Using the method of False Position, find a real root of the equation $x^4 - 11x + 8 = 0$ accurate to four decimal places.

Solution

Here $f(x) = x^4 - 11x + 8 = 0$

$f(1) = 1^4 - 11(1) + 8 = -2 < 0$

$f(2) = 2^4 - 11(2) + 8 = 2 > 0$

Therefore, a root of $f(x) = 0$ lies between 1 and 2.We apply the method of False Position with $a = 1$and $b = 2$.

The calculations based on the method of False Position are summarized in the following Table :

n	a	b	f(a)	f(b)	x_{s_1}	$f(x_{s_1})$	ξ
1	1	2	−2	2	1.5	−3.4375	—
2	1.5	2	−3.4375	2	1.81609	−1.9895	0.17405
3	1.81609	2	−1.09895	2	1.88131	−0.16758	3.4666×10^{-2}
4	1.88131	2	−0.16758	2	1.89049	−0.02232	4.85383×10^{-3}
5	1.89049	2	−0.02232	2	1.89169	−0.00292	6.3902×10^{-4}
6	1.89169	2	−0.00292	2	1.89185	−0.00038	8.34227×10^{-5}
7	1.89185	2	−0.00038	2	1.89187	−0.00005	1.08786×10^{-5}

The relative error after the seventh step is

$$\xi = \frac{1.89187 - 1.89185}{1.89187} = 1.08786 \times 10^{-5}$$

Hence, the root is 1.8918 accurate to four decimal places.

2.6 NEWTON'S METHOD

Newton's method (also called the Newton-Raphson method) is a scheme for finding a numerical solution of an equation of the form *f(x) = 0* where *f(x)* is continuous and differentiable and the equation is known to have a solution near a given point. The method is illustrated in Fig. 2.6.

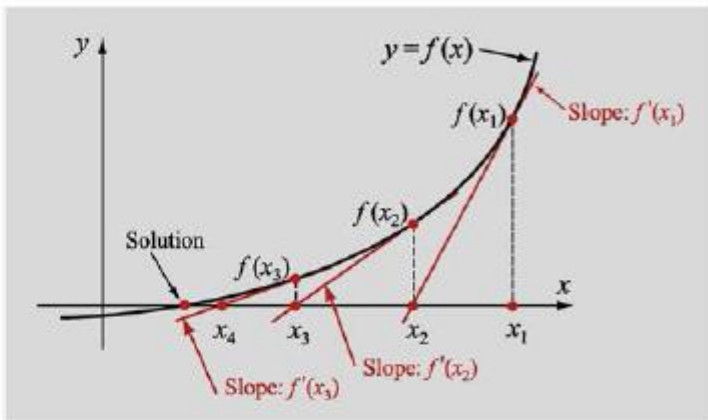


Figure 2-6: Newton's method.

The solution process starts by choosing point x_1 as the first estimate of the solution. The second estimate x_2 is obtained by taking the tangent line to *f(x)* at the point $(x_1, f(x_1))$ and finding the intersection point of the tangent line with the *x*-axis. The next estimate x_3 is the intersection of the tangent line to *f(x)* at the point $(x_2, f(x_2))$ with the *x*-axis, and so on. Mathematically, for the first iteration, the slope, $f'(x_1)$, of the tangent at point $(x_1, f(x_1))$ is given by:

$$f'(x_1) = \frac{f(x_1) - 0}{x_1 - x_2} \qquad (2.10)$$

Solving Eq. (2.10) for x_2 gives:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \qquad (2.11)$$

Equation (2 .11) can be generalized for determining the "next" solution x_{i+1} from the present solution x_i:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad (2.12)$$

Equation (2.12) is the general iteration formula for Newton's method. It is called an iteration formula because the solution is found by repeated application of Eq. (2.12) for each successive value of *i*.

Algorithm for Newton's method

1. Choose a point x_i as an initial guess of the solution.
2. For *i* = 1, 2, . . . , until the error is smaller than a specified value, calculate x_{i+1} by using Eq. (2.12).

When are the iterations stopped?

Ideally, the iterations should be stopped when an exact solution is obtained. This means that the value of x is such that $f(x) = 0$. Generally, as discussed in Section 2.1, this exact solution cannot be found computationally. In practice, therefore, the iterations are stopped when an estimated error is smaller than some predetermined value. Tolerance in the solution, as in the bisection method, cannot be calculated since bounds are not known. Two error estimates that are typically used with Newton's method are:

Estimated relative error: The iterations are stopped when the estimated relative error is smaller than a specified value ε:

$$\left|\frac{x_{i+1} - x_i}{x_i}\right| \le \boldsymbol{\varepsilon}$$

Tolerance in f(x): The iterations are stopped when the absolute value of f(x;) is smaller than some number δ:

$$|f(x_i)| \le \delta$$

Notes on Newton's method

• The method, when successful, works well and converges fast. When it does not converge, it is usually because the starting point is not close enough to the solution. Convergence problems typically occur when the value of $f'(x)$ is close to zero in the vicinity of the solution (where $f(x) = 0$). It is possible to show that Newton's method converges if the function $f(x)$ and its first and second derivatives $f'(x)$ and $f''(x)$ are all continuous, if $f'(x)$ is not zero at the solution, and if the starting value x_1 is near the actual solution. Illustrations of two cases where Newton's method does not converge (i.e., diverges) are shown in Fig. 2-7.

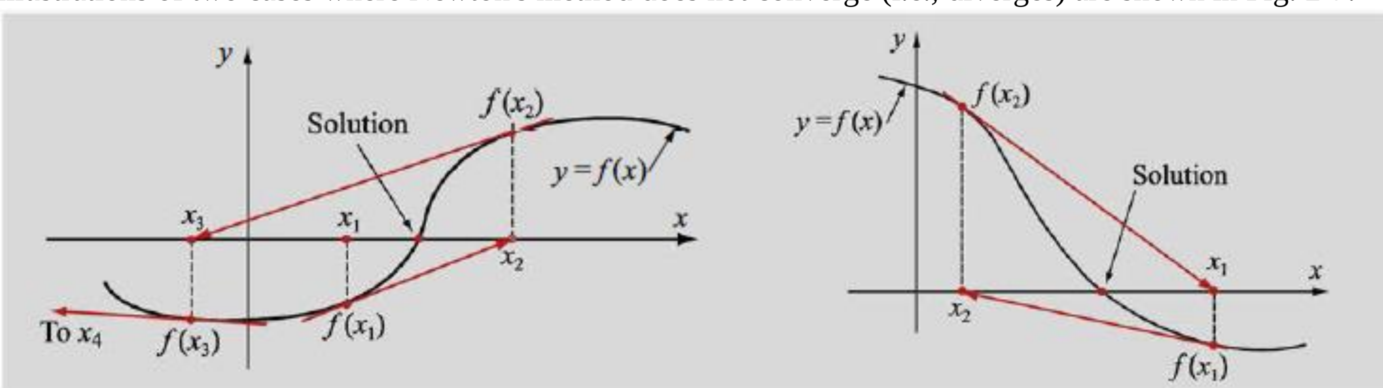


Figure 2-7: Cases where Newton's method diverges.

• A function $f'(x)$, which is the derivative of the function $f(x)$, has to be substituted in the iteration formula, Eq. (2.12). In many cases, it is simple to write the derivative, but sometimes it can be difficult to determine. When an expression for the derivative is not available, it might be possible to determine the slope numerically or to find a solution by using the secant method (Section 2.7), which is somewhat similar to Newton's method but does not require an expression for the derivative.

Example 2.7

Find the solution of the equation $8-4.5(x- \sin(x)) = 0$ by using Newton's method in the following two ways:

(a) Using a nonprogrammable calculator, calculate the first two iterations on paper using six significant figures.

(b) Use MATLAB with 0.0001 for the maximum relative error and 10 for the maximum number of iterations.

In both parts, use $x = 2$ as the initial guess of the solution.

Solution:

In the present problem, $f(x) = 8 - 4.5(x-\sin x)$ and $f'(x) = -4.5(1 -\cos x)$.

(a) To start the iterations, $f(x)$ and $f'(x)$ are substituted in Eq. (2.12):

$$x_{i+1} = x_i - \frac{8-4.5(x_i-\sin x_i)}{-4.5(1-\cos x_i)} \qquad (2.13)$$

In the first iteration, $i = 1$ and $x_1 = 2$, and Eq. (2.13) gives:

$$x_2 = 2 - \frac{8-4.5(2-\sin 2)}{-4.5(1-\cos 2)} = 2.485172$$

for the second iteration, $i = 2$ and x_2 = 2.485172, and Eq. (2.13) gives:

$$x_3 = 2.485172 - \frac{8-4.5(2.485172-\sin 2.485172)}{-4.5(1-\cos 2.485172)} = 2.430987$$

(b) (Exc)

Example2.8 Consider $f(x) \equiv x^6 - x - 1 = 0$ for its positive root α. An initial guess x_0 can be generated from a graph of $y = f(x)$. The iteration is given by:

$$x_{n+1} = x_n - \frac{x_n^6 - x_n - 1}{6x_n^5 - 1} \quad , n \geq 0$$

We use an initial guess of $x_0 = 1.5$.
The column "$x_n - x_{n-1}$" is an estimate of the error $\alpha - x_{n-1}$; justification is given later.

n	x_n	$f(x_n)$	$x_n - x_{n-1}$
0	1.5	8.89E + 1	
1	1.30049088	2.54E + 1	−2.00E − 1
2	1.18148042	5.38E − 1	−1.19E − 1
3	1.13945559	4.92E − 2	−4.20E − 2
4	1.13477763	5.50E − 4	−4.68E − 3
5	1.13472415	7.11E − 8	−5.35E − 5
6	1.13472414	1.55E − 15	−6.91E − 9

As seen from the output, the convergence is very rapid. The iterate x_6 is accurate to the machine precision of around 16 decimal digits. This is the typical behaviour seen with Newton's method for most problems, but not all.

Example 2.9:

Use the Newton-Raphson method to find the real root near 2 of the equation $x^4 - 11x + 8 = 0$ accurate to five decimal places.

Solution:

Here $f(x) = x^4 - 11x + 8$
$f'(x) = 4x^3 - 11$
$x_0 = 2$
and $f(x_0) = f(2) = 2^4 - 11(2) + 8 = 2$
$f'(x_0) = f'(2) = 4(2)^3 - 11 = 21$
Therefore,

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2 - \frac{2}{21} = 1.90476$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 1.90476 - \frac{(1.90476)^4 - 11(1.90476) + 8}{4(1.90476)^3 - 11} = 1.89209$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 1.89209 - \frac{(1.89209)^4 - 11(1.89209) + 8}{4(1.89209)^3 - 11} = 1.89188$$

$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = 1.89188 - \frac{(1.89188)^4 - 11(1.89188) + 8}{4(1.89188)^3 - 11} = 1.89188$$

Hence the root of the equation is 1.89188.

Example 2.10

Using Newton-Raphson method, find a root of the function $f(x) = e^x - 3x^2$ to an accuracy of 5 digits. The root is known to lie between 0.5 and 1.0. Take the starting value of x as $x_0 = 1.0$.

Solution:

Start at $x_0 = 1.0$ and prepare a table as shown in Table 2.8, where $f(x) = e^x - 3x^2$ and $f'(x) = e^x - 6x$. The relative error

$$\zeta = \left|\frac{x_{i+1} - x_i}{x_{i+1}}\right|$$

The Newton-Raphson iteration method is given by

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

i	x_i	$f(x_i)$	$f'(x_i)$	x_{i+1}	ξ
0	1.0	−0.28172	−3.28172	0.91416	0.09391
1	0.91416	−0.01237	−2.99026	0.91002	0.00455
2	0.91002	−0.00003	−2.97574	0.91001	0.00001
3	0.91001	0	−2.97570	0.91001	6.613×10^{-11}

Example 2.11:

Evaluate $\sqrt{29}$ to five decimal places by Newton-Raphson iterative method.

Solution:

Let $x = \sqrt{29}$ then $x^2 - 29 = 0$.
We consider $f(x) = x^2 - 29 = 0$ and $f'(x) = 2x$
The Newton-Raphson iteration formula gives

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^2 - 29}{2x_i} = \frac{1}{2}\left(x_i + \frac{29}{x_i}\right) \qquad \text{(E.1)}$$

Now $f(5) = 25 - 29 = -4 < 0$ and $f(6) = 36 - 29 = 7 > 0$.
Hence, a root of $f(x) = 0$ lies between 5 and 6.
Taking $x_0 = 5.3$, Equation (E.1) gives

$$x_1 = \frac{1}{2}\left(5.3 + \frac{29}{5.3}\right) = 5.38585$$

$$x_2 = \frac{1}{2}\left(5.38585 + \frac{29}{5.38585}\right) = 5.38516$$

$$x_3 = \frac{1}{2}\left(5.38516 + \frac{29}{5.38516}\right) = 5.38516$$

Since $x_2 = x_3$ up to five decimal places, $\sqrt{29} = 5.38516$.

2.7 SECANT METHOD

The secant method is a scheme for finding a numerical solution of an equation of the form *f(x) = 0*. The method uses two points in the neighbourhood of the solution to determine a new estimate for the solution
(Fig. 2-8). The two points (marked as x_1 and x_2 in the figure) are used to define a straight line (secant line), and the point where the line intersects the *x*-axis (marked as x_3 in the figure) is the new estimate for the solution. As shown, the two points can be on one side of the solution

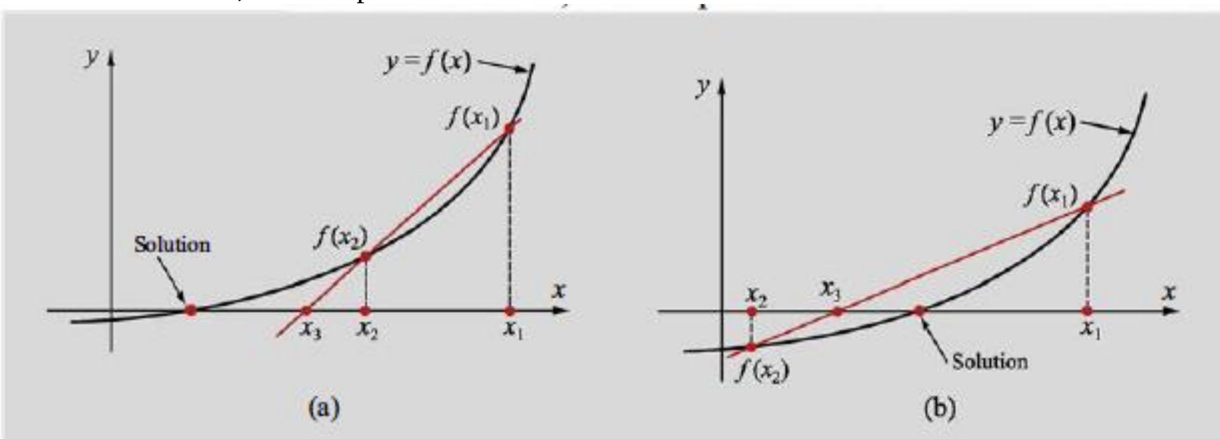


Figure 2-8: The secant method.

The slope of the secant line is given by:

$$\frac{f(x_1)-f(x_2)}{x_1-x_2} = \frac{f(x_2)-0}{x_2-x_3} \qquad (2.14)$$

which can be solved for x_3 :

$$x_3 = x_2 - \frac{f(x_2)(x_1-x_2)}{f(x_1)-f(x_2)} \qquad (2.15)$$

Once point x_3 is determined, it is used together with point x_2 to calculate the next estimate of the solution, x_4. Equation (2.15) can be generalized to an iteration formula in which a new estimate of the solution x_{i+1} is determined from the previous two solutions x_i and x_{i-1}.

$$x_{i+1} = x_i - \frac{f(x_i)(x_{i-1}-x_i)}{f(x_{i-1})-f(x_i)} \qquad (2.16)$$

Figure 2-9 illustrates the iteration process with the secant method.

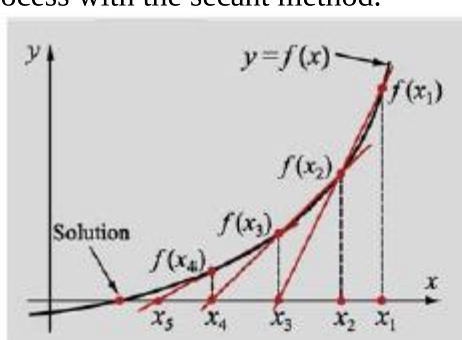


Figure 2-9: Secant method.

Example 2.12

Find a root of the equation $x^3 - 8x - 5 = 0$ using the secant method.

Solution:

$f(x) = x^3 - 8x - 5 = 0$

$f(3) = 3^3 - 8(3) - 5 = -2$

$f(4) = 4^3 - 8(4) - 5 = 27$

Therefore one root lies between 3 and 4. Let the initial approximations be $x_0 = 3$, and $x_1 = 3.5$. Then, x_2 is given by:

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

The calculations are summarized in the above Table

x_0	$f(x_0)$	x_1	$f(x_1)$	x_2	$f(x_2)$
3	−2	3.5	9.875	3.08421	−0.33558
3.5	9.875	3.08421	−0.33558	3.09788	−0.05320
3.08421	−0.33558	3.09788	−0.05320	3.10045	0.00039
3.09788	−0.05320	3.10045	0.00039	3.10043	0
3.10045	0.00039	3.10043	0	3.10043	0

Hence, a root is 3.1004 correct up to five significant figures.

Example 2.13

Determine a root of the equation $sin(x) + 3\ cos(x) - 2 = 0$ using the secant method. The initial approximations x_0 and x_1 are 0 and 1.5.

Solution:

The formula for x_2 is given by:

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

The calculations are summarized in the above Table.

x_0	$f(x_0)$	x_1	$f(x_1)$	x_2	$f(x_2)$
0	-2.33914	1.5	-0.79029	0.83785149	0.75039082
1.5	-0.79029	0.83785149	0.75039082	1.160351166	0.113995951
0.83785149	0.75039082	1.160351166	0.113995951	1.2181197917	-0.025315908
1.160351166	0.113995951	1.2181197917	-0.025315908	1.2076220119	0.000503735
1.2181197917	-0.025315908	1.2076220119	0.000503735	1.2078268211	0.000002099
1.2076220119	0.000503735	1.2078268211	0.000002099	1.2078276783	-0.000000000
1.2078268211	0.000002099	1.2078276783	-0.000000000		

Hence, a root is 1.2078 correct up to five significant figures.

2.8 FIXED-POINT ITERATION METHOD

Fixed-point iteration is a method for solving an equation of the form $f(x) = 0$. The method is carried out by rewriting the equation in the form:

$$x = g(x) \qquad (2.17)$$

Obviously, when x is the solution of $f(x) = 0$, the left side and the right side of Eq. (2.17) are equal. This is illustrated graphically by plotting y = x and y = g(x), as shown in Fig. 2-10.

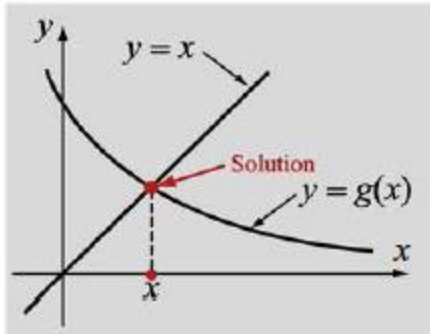


Figure 2-10: Fixed-point iteration method.

The point of intersection of the two plots, called the fixed point, is the solution. The numerical value of the solution is determined by an iterative process. It starts by taking a value of x near the fixed point as the first guess for the solution and substituting it in *g(x)*. The value of *g(x)* that is obtained is the new (second) estimate for the solution. The second value is then substituted back in *g(x)*, which then gives the third estimate of the solution. The iteration formula is thus given by:

$$x_{i+1} = g(x_i) \qquad (2.18)$$

The function g(x) is called ***the iteration function***.

• When the method works, the values of x that are obtained are successive iterations that progressively converge toward the solution. Two such cases are illustrated graphically in Fig. 2-11. The solution process starts by choosing point x_1 on the x-axis and drawing a vertical line that intersects the curve y = g(x) at point $g(x_1)$. Since $x_2 = g(x_1)$, a horizontal line is drawn from point $(x_1, g(x_1))$ toward the line y = x. The intersection point gives the location of x_2 . From x_2 a vertical line is drawn toward the curve y = g(x). The intersection point is now $(x_2, g(x_2))$, and $g(x_2)$ is also the value of x_3. From point $(x_2, g(x_2))$ a horizontal line is drawn again toward y = x, and the intersection point gives the location of x3 . As the process continues the intersection points converge toward the fixed point or the true solution x_{rs}.

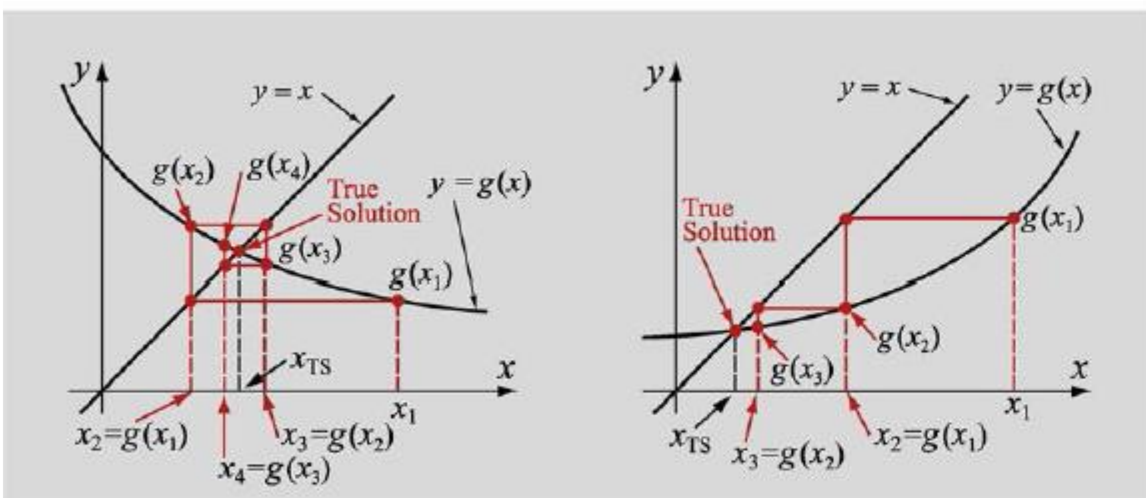


Figure 2-11: Convergence of the fixed-point iteration method.

• It is possible, however, that the iterations will not converge toward the fixed point, but rather diverge away. This is shown in Fig. 2-12. The figure shows that even though the starting point is close to the solution, the subsequent points are moving farther away from the solution.

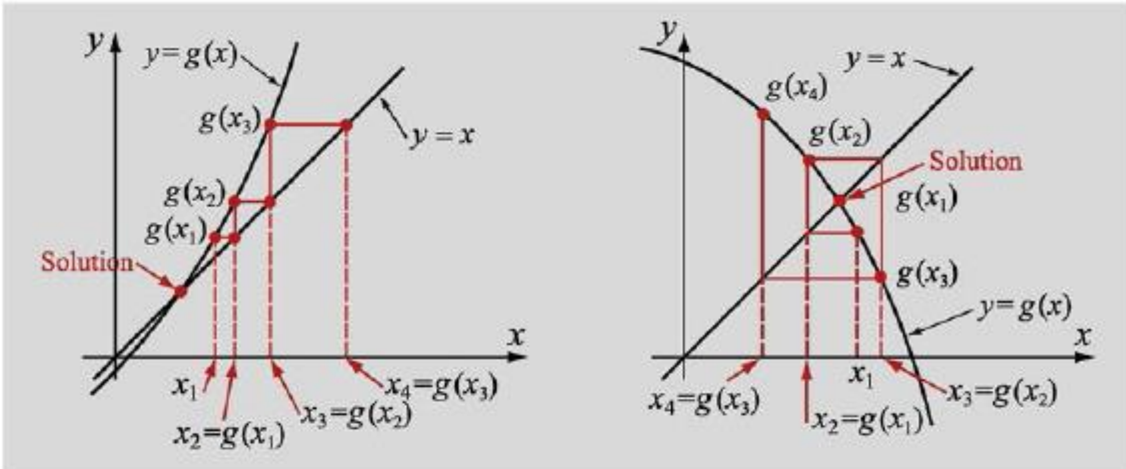


Figure 2-12: Divergence of the fixed-point iteration method.

• Sometimes, the form $f(x) = 0$ does not lend itself to deriving an iteration formula of the form $x = g(x)$. In such a case, one can always add and subtract x to $f(x)$ to obtain $x + f(x) - x = 0$. The last equation can be rewritten in the form that can be used in the fixed-point iteration method:

$$x = x + f(x) = g(x)$$

Choosing the appropriate iteration function g(x)

For a given equation $f(x) = 0$, the iteration function is not unique since it is possible to change the equation into the form $x = g(x)$ in different ways. This means that several iteration functions g(x) can be written for the same equation. A $g(x)$ that should be used in Eq. (2.18) for the iteration process is one for which the iterations converge toward the solution. There might be more than one form that can be used, or it may be that none of the forms is appropriate so that the fixed-point iteration method cannot be used to solve the equation. In cases where there are multiple solutions, one iteration function may yield one root, while a different function yields other roots. Actually, it is possible to determine ahead of time if the iterations converge or diverge for a specific g(x).

The fixed-point iteration method converges if, in the neighbourhood of the fixed point, the derivative of g(x) has an absolute value that is smaller than 1:

$$|g'(x)| < 1 \qquad \textbf{(2.19)}$$

As an example, consider the equation:

$$xe^{0.5x} + 1.2x - 5 = 0 \qquad (2.20)$$

A plot of the function $f(x) = xe^{0.5x} + 1.2x - 5$ (see Fig. 2-13) shows that the equation has a solution between $x = 1$ and $x = 2$.

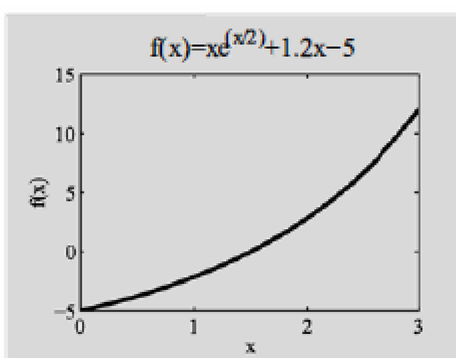


Figure 2-13: A plot of $f(x) = xe^{x/2} + 1.2x - 5$.

Equation (2 .20) can be rewritten in the form $x = g(x)$ in different ways. Three possibilities are discussed next.

Case a: $x = \frac{5 - xe^{x/2}}{1.2}$

In this case $g(x) = \frac{5 - xe^{x/2}}{1.2}$ and $g'(x) = \frac{-(e^{\frac{x}{2}} + 0.5xe^{\frac{x}{2}})}{1.2}$

The values of g'(x) at points *x= 1* and *x= 2* , which are in the neighborhood of the solution, are:

$g'(1) = \frac{-(e^{\frac{1}{2}} + 0.5(1)e^{\frac{1}{2}})}{1.2}$= -2.0609

$g'(2) = \frac{-(e^{\frac{2}{2}} + 0.5(2)e^{\frac{2}{2}})}{1.2}$= -4.5305

Case b: $x = \frac{5}{e^{0.5x} + 1.2}$

In this case $g(x) = \frac{5}{e^{0.5x} + 1.2}$ and $g'(x) = \frac{-5e^{0.5x}}{2(e^{0.5x} + 1.2)^2}$

The value of g'(x) at points *x= 1* and *x= 2* , which are in the neighborhood of the solution, are:

$g'(1) = \frac{-5e^{0.5(1)}}{2(e^{0.5(1)} + 1.2)^2}$= -0.5079

$g'(2) = \frac{-5e^{0.51(2)}}{2(e^{0.5(2)} + 1.2)^2}$= -0.4426

Case c: $x = \frac{5 - 1.2x}{e^{0.5x}}$

In this case, $g(x) = \frac{5 - 1.2x}{e^{0.5x}}$ and $g'(x) = \frac{-3.7 + 0.6x}{e^{0.5x}}$

The value of g'(x) at points *x= 1* and *x= 2* , which are in the neighborhood of the solution, are:

$g'(1) = \frac{5 - 1.2(1)}{e^{0.5(1)}}$= -1.8802

$g'(2) = \frac{5 - 1.2(2)}{e^{0.5(2)}}$= -0.9197

These results show that the iteration function *g(x)* from Case b is the one that should be used since, in this case, $|g'(1)| < 1$and$|g'(2)| < 1$.

Substituting g(x) from Case b in Eq. (2.18) gives:

$$x_{i+1} = \frac{5}{e^{0.5x_i} + 1.2}$$

Starting with x_1 = 1, the first few iterations are:

$x_2 = \frac{5}{e^{0.5(1)} + 1.2}$ = 1.755173 , $x_3 = \frac{5}{e^{0.5(1.755173)} + 1.2}$ = 1.386928

$x_4 = \frac{5}{e^{0.5(1.386928)} + 1.2}$ = 1.56219 , $x_5 = \frac{5}{e^{0.5(1.56219)} + 1.2}$ = 1.477601

$x_6 = \frac{5}{e^{0.5(1.477601)} + 1.2}$ = 1.518177 , $x_7 = \frac{5}{e^{0.5(1.518177)} + 1.2}$ = 1.498654

As expected, the values calculated in the iterations are converging toward the actual solution, which is **x = 1.5050**. On the contrary, if the function g(x) from Case a is used in the iteration, the first few iterations are:

$x_2 = \frac{5 - e^{1/2}}{1.2}$ = 2.792732

$x_3 = \frac{5 - 2.792732 * e^{2.792732/2}}{1.2}$ = -5.23667

$x_4 = \frac{5 + 5.23667 * e^{-5.23667/2}}{1.2}$ = 4.4849

$x_5 = \frac{5 - 4.4849 * e^{4.4849/2}}{1.2}$ = -31.0262

In this case, the iterations give values that diverge from the solution.

When should the iterations be stopped?
The true error (the difference between the true solution and the estimated solution) cannot be calculated since the true solution, in general, is not known. As with Newton's method, the iterations can be stopped either when the relative error or the tolerance in *f(x)* is smaller than some predetermined value.

Example 2.14
Find a real root of $x^3 - 2x - 3 = 0$, correct to three decimal places using the Successive Approximation method.

Solution:
Here $f(x) = x^3 - 2x - 3 = 0$ (E.1)
Also $f(1) = 1^3 - 2(1) - 3 = -4 < 0$
and $f(2) = 2^3 - 2(2) - 3 = 1 > 0$
Therefore, root of Eq.(E.1) lies between 1 and 2. Since $f(1) < f(2)$, we can take the initial approximation $x_0 = 1$. Now, Eq. (E.1) can be rewritten as
$x^3 = 2x + 3$
or $x = (2x + 3)^{1/3} = \varphi(x)$
The successive approximations of the root are given by
$x_1 = \varphi(x_0) = (2x_0 + 3)^{1/3} = [2(1) + 3]^{1/3} = 1.709975947$
$x_2 = \varphi(x_1) = (2x_1 + 3)^{1/3} = [2(1.709975947) + 3]^{1/3} = 1.858562875$
$x_3 = \varphi(x_2) = (2x_2 + 3)^{1/3} = [2(1.858562875) + 3]^{1/3} = 1.88680851$
$x_4 = \varphi(x_3) = (2x_3 + 3)^{1/3} = [2(1.88680851) + 3]^{1/3} = 1.892083126$
$x_5 = \varphi(x_4) = (2x_4 + 3)^{1/3} = [2(1.892083126) + 3]^{1/3} = 1.89306486$
Hence, the real roots of $f(x) = 0$ is 1.893 correct to three decimal places.

Example 2.15
Find a real root of $\cos x - 3x + 5 = 0$. Correct to four decimal places using the fixed point method.

Solution:
Here, we have
$f(x) = \cos x - 3x + 5 = 0$ (E.1)
$f(0) = \cos(0) - 3(0) + 5 = 5 > 0$
$f(\pi/2) = \cos(\pi/2) - 3(\pi/2) + 5 = -3\pi/2 + 5 < 0$
Also $f(0)\, f(\pi/2) < 0$
Hence, a root of $f(x) = 0$ lies between 0 and $\pi/2$.
The given Eq. (E.1) can be written as:

$$x = \frac{1}{3}[5 + \cos x]$$

Here $$\phi(x) = \frac{1}{3}[5 + \cos x] \text{ and } \phi'(x) = -\frac{\sin x}{3}$$

$$|\phi'(x)| = \left|\frac{\sin x}{3}\right| < 1 \text{ in } (0, \pi/2)$$

Hence, the successive approximation method applies.

Let

$$x_0 = 0$$
$$x_1 = \phi(x_0) = \frac{1}{3}[5 + \cos 0] = 2$$
$$x_2 = \phi(x_1) = \frac{1}{3}[5 + \cos(2)] = 1.52795$$
$$x_3 = \phi(x_2) = \frac{1}{3}[5 + \cos(1.52795)] = 1.68094$$
$$x_4 = \phi(x_3) = \frac{1}{3}[5 + \cos(1.68094)] = 1.63002$$
$$x_5 = \phi(x_4) = \frac{1}{3}[5 + \cos(1.63002)] = 1.64694$$
$$x_6 = \phi(x_5) = \frac{1}{3}[5 + \cos(1.64694)] = 1.64131$$
$$x_7 = \phi(x_6) = \frac{1}{3}[5 + \cos(1.64131)] = 1.64318$$
$$x_8 = \phi(x_7) = \frac{1}{3}[5 + \cos(1.64318)] = 1.64256$$
$$x_9 = \phi(x_8) = \frac{1}{3}[5 + \cos(1.64256)] = 1.64277$$
$$x_{10} = \phi(x_9) = \frac{1}{3}[5 + \cos(1.64277)] = 1.64270$$

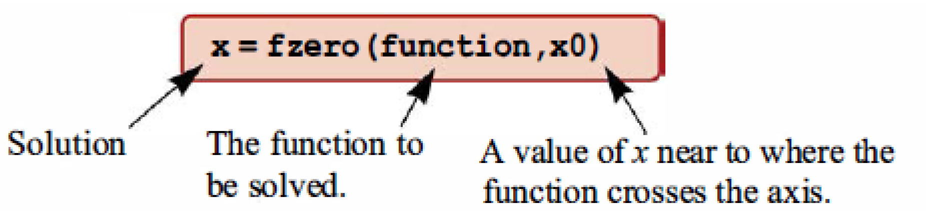
Hence, the root of the equation is 1.6427 correct to four decimal places.

2.9 Use of MATLAB Built-in Functions for Solving NONLINEAR EQUATIONS

MATLAB has two built-in functions for solving equations with one variable. The ***fzero command*** can be used to find a root of any equation, and the ***roots command*** can be used for finding the roots of a polynomial.

2.9.1 The *fzero* Command

The *fzero* command can be used to solve an equation (in the form f(x) = 0) with one variable. The user needs to know approximately where the solution is, or if there are multiple solutions, which one is desired. The form of the command is:



• **x** is the solution, which is a scalar. A value of x near to where the function crosses the axis.
• **function** is the function whose root is desired. It can be entered in three different ways:
1. The simplest way is to enter the mathematical expression as a string.
2. The function is first written as a user-defined function, and then the function handle is entered.
3. The function is written as an anonymous function, and then its name (which is the name of the handle) is entered.
• The function has to be written in a standard form. For example, if the function to be solved is xe -x = 0.2, it has to be written as
$f(x) = xe^{-x}-0.2 = 0$. If this function is entered into the fzero commands as a string, it is typed as:

*'x*exp (-x) -0. 2'.*

• When a function is entered as a string, it cannot include predefined variables. For example, if the function to be entered is $f(x) = xe^{-x} -0.2$, it is not possible to first define b=0. 2 and then enter:

*'x*exp (-x) -b'.*

• x_0 can be a scalar or a two-element vector. If it is entered as a scalar, it has to be a value of x near the point where the function crosses the x-axis. If x_0 is entered as a vector, the two elements have to be points on opposite sides of the solution. When a function has more than one solution, each solution can be determined separately by using the *fzero* function and entering values for x_0 that are near each of the solutions. Usage of the *fzero* command is illustrated next for solving equation $8 - 4.5(x - \sin(x))$. The function $f(x) = 8 - 4.5(x - \sin(x))$ is first defined as an anonymous function named FUN. Then the name FUN is entered as an input argument in the function *fzero*.

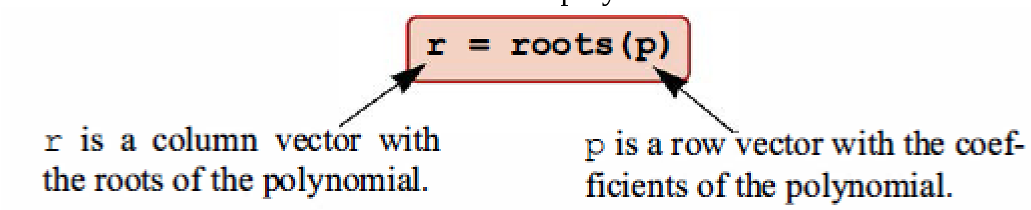
```
>> FUN = @ (x) 8-4.5*(x-sin(x))
FUN =
    @(x)8-4.5*(x-sin(x))
>> sol=fzero(FUN,2)
sol =
   2.430465741723630
```

$f(x)$ is written as an anonymous function.

The name FUN of the anonymous function is entered in fzero.

2.9.2 The *roots* Command

The *roots* command can be used to find the roots of a polynomial. The form of the command is:



2.10 PROBLEMS

1. Determine the root of $f(x) = x - 2e^{-x}$ by:

(a) Using the bisection method. Start with a= 0 and b= 1, and carry out the first three iterations.

(b) Using the secant method. Start with the two points, $x_1 = 0$ and $x_2 = 1$, and carry out the first three iterations.

(c) Using Newton's method. Start at $x_1 = 1$ and carry out the first three iterations.

2. Determine the fourth root of 200 by finding the numerical solution of the equation $x^4 - 200 = 0$. Use Newton's method. Start at $x = 8$ and carry out the first five iterations.

3. Determine the positive root of the polynomial $x^3 + 0.6x^2 + 5.6x - 4.8$.

(a) Plot the polynomial and choose a point near the root for the first estimate of the solution. Using Newton's method, determine the approximate solution in the first four iterations.

(b) From the plot in part (a), choose two points near the root to start the solution process with the secant method. Determine the approximate solution in the first four iterations.

4. The equation $x^3 - x - e^x - 2 = 0$ has a root between $x = 2$ and $x = 3$.

(a) Write four different iteration functions for solving the equation using the fixed-point iteration method.

(b) Determine which g(x) from part (a) could be used according to the condition in Eq. (2.19).

(c) Carry out the first five iterations using the g(x) determined in part (b), starting with $x = 2$.

5. Use the Bisection method to compute the root of $e^x - 3x = 0$ correct to three decimal places in the interval (1.5, 1.6).

6. Use the Bisection method to find a root of the equation $x^3 - 4x - 9 = 0$ in the interval (2, 3), accurate to four decimal places.

7. Use the method of False Position to find a root correct to three decimal places of the function $x^3 - 4x - 9 = 0$.

8. A root of $f(x) = x^3 - 10x^2 + 5 = 0$ lies close to $x = 0.7$. Determine this root with the Newton-Raphson method to five decimal accuracy.

9. A root of $f(x) = x^3 - x^2 - 5 = 0$ lies in the interval (2, 3). Determine this root with the Newton-Raphson method for four decimal places.

10. Use the fixed point method to find a root of the equation $e^x - 3x = 0$ in the interval (0, 1) accurate to four decimal places.

11. Use the method of Successive Approximation to determine a solution accurate to within 10^{-2} for $x^4 - 3x^2 - 3 = 0$ on [1, 2]. Use $x_0 = 1$.