

2.4 Newton-Raphson and Secant Methods

Slope Methods for Finding Roots

If $f(x)$, $f'(x)$, and $f''(x)$ are continuous near a root p , then this extra information regarding the nature of $f(x)$ can be used to develop algorithms that will produce sequences $\{p_k\}$ that converge faster to p than either the bisection or false position method. The Newton-Raphson (or simply Newton's) method is one of the most useful and best known algorithms that relies on the continuity of $f'(x)$ and $f''(x)$. We shall introduce it graphically and then give a more rigorous treatment based on the Taylor polynomial.

Assume that the initial approximation p_0 is near the root p . Then the graph of $y = f(x)$ intersects the x -axis at the point $(p, 0)$, and the point $(p_0, f(p_0))$ lies on the curve near the point $(p, 0)$ (see Figure 2.13). Define p_1 to be the point of intersection of the x -axis and the line tangent to the curve at the point $(p_0, f(p_0))$. Then Figure 2.13 shows that p_1 will be closer to p than p_0 in this case. An equation relating p_1 and p_0 can be found if we write down two versions for the slope of the tangent line L :

$$(1) \quad m = \frac{0 - f(p_0)}{p_1 - p_0},$$

which is the slope of the line through $(p_1, 0)$ and $(p_0, f(p_0))$, and

$$(2) \quad m = f'(p_0),$$

which is the slope at the point $(p_0, f(p_0))$. Equating the values of the slope m in equations (1) and (2) and solving for p_1 results in

$$(3) \quad p_1 = p_0 - \frac{f(p_0)}{f'(p_0)}.$$

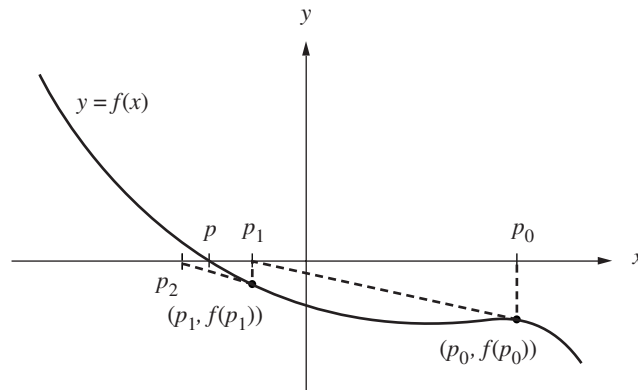


Figure 2.13 The geometric construction of p_1 and p_2 for the Newton-Raphson method.

The process above can be repeated to obtain a sequence $\{p_k\}$ that converges to p . We now make these ideas more precise.

Theorem 2.5 (Newton-Raphson Theorem). Assume that $f \in C^2[a, b]$ and there exists a number $p \in [a, b]$, where $f(p) = 0$. If $f'(p) \neq 0$, then there exists a $\delta > 0$ such that the sequence $\{p_k\}_{k=0}^{\infty}$ defined by the iteration

$$(4) \quad p_k = g(p_{k-1}) = p_{k-1} - \frac{f(p_{k-1})}{f'(p_{k-1})} \quad \text{for } k = 1, 2, \dots$$

will converge to p for any initial approximation $p_0 \in [p - \delta, p + \delta]$.

Remark. The function $g(x)$ defined by the formula

$$(5) \quad g(x) = x - \frac{f(x)}{f'(x)}$$

is called the *Newton-Raphson iteration function*. Since $f(p) = 0$, it is easy to see that $g(p) = p$. Thus the Newton-Raphson iteration for finding the root of the equation $f(x) = 0$ is accomplished by finding a fixed point of the function $g(x)$.

Proof. The geometric construction of p_1 shown in Figure 2.13 does not help in understanding why p_0 needs to be close to p or why the continuity of $f''(x)$ is essential. Our analysis starts with the Taylor polynomial of degree $n = 1$ and its remainder term:

$$(6) \quad f(x) = f(p_0) + f'(p_0)(x - p_0) + \frac{f''(c)(x - p_0)^2}{2!},$$

where c lies somewhere between p_0 and x . Substituting $x = p$ into equation (6) and using the fact that $f(p) = 0$ produces

$$(7) \quad 0 = f(p_0) + f'(p_0)(p - p_0) + \frac{f''(c)(p - p_0)^2}{2!}.$$

If p_0 is close enough to p , the last term on the right side of (7) will be small compared to the sum of the first two terms. Hence it can be neglected and we can use the approximation

$$(8) \quad 0 \approx f(p_0) + f'(p_0)(p - p_0).$$

Solving for p in equation (8), we get $p \approx p_0 - f(p_0)/f'(p_0)$. This is used to define the next approximation p_1 to the root

$$(9) \quad p_1 = p_0 - \frac{f(p_0)}{f'(p_0)}.$$

When p_{k-1} is used in place of p_0 in equation (9), the general rule (4) is established. For most applications this is all that needs to be understood. However, to fully comprehend

what is happening, we need to consider the fixed-point iteration function and apply Theorem 2.2 in our situation. The key is in the analysis of $g'(x)$:

$$g'(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}.$$

By hypothesis, $f(p) = 0$; thus $g'(p) = 0$. Since $g'(p) = 0$ and $g'(x)$ is continuous, it is possible to find a $\delta > 0$ so that the hypothesis $|g'(x)| < 1$ of Theorem 2.2 is satisfied on $(p - \delta, p + \delta)$. Therefore, a sufficient condition for p_0 to initialize a convergent sequence $\{p_k\}_{k=0}^{\infty}$, which converges to a root of $f(x) = 0$, is that $p_0 \in (p - \delta, p + \delta)$ and that δ be chosen so that

$$(10) \quad \frac{|f(x)f''(x)|}{|f'(x)|^2} < 1 \quad \text{for all } x \in (p - \delta, p + \delta). \quad \bullet$$

Corollary 2.2 (Newton's Iteration for Finding Square Roots). Assume that $A > 0$ is a real number and let $p_0 > 0$ be an initial approximation to \sqrt{A} . Define the sequence $\{p_k\}_{k=0}^{\infty}$ using the recursive rule

$$(11) \quad p_k = \frac{p_{k-1} + \frac{A}{p_{k-1}}}{2} \quad \text{for } k = 1, 2, \dots$$

Then the sequence $\{p_k\}_{k=0}^{\infty}$ converges to \sqrt{A} ; that is, $\lim_{n \rightarrow \infty} p_k = \sqrt{A}$.

Outline of Proof. Start with the function $f(x) = x^2 - A$, and notice that the roots of the equation $x^2 - A = 0$ are $\pm\sqrt{A}$. Now use $f(x)$ and the derivative $f'(x)$ in formula (5) and write down the Newton-Raphson iteration formula

$$(12) \quad g(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 - A}{2x}.$$

This formula can be simplified to obtain

$$(13) \quad g(x) = \frac{x + \frac{A}{x}}{2}.$$

When $g(x)$ in (13) is used to define the recursive iteration in (4), the result is formula (11). It can be proved that the sequence that is generated in (11) will converge for any starting value $p_0 > 0$. The details are left for the exercises. •

An important point of Corollary 2.2 is the fact that the iteration function $g(x)$ involved only the arithmetic operations $+$, $-$, \times , and $/$. If $g(x)$ had involved the calculation of a square root, we would be caught in the circular reasoning that being able to calculate the square root would permit you to recursively define a sequence that will converge to \sqrt{A} . For this reason, $f(x) = x^2 - A$ was chosen, because it involved only the arithmetic operations.

Example 2.11. Use Newton's square-root algorithm to find $\sqrt{5}$.

Starting with $p_0 = 2$ and using formula (11), we compute

$$\begin{aligned} p_1 &= \frac{2 + 5/2}{2} = 2.25 \\ p_2 &= \frac{2.25 + 5/2.25}{2} = 2.236111111 \\ p_3 &= \frac{2.236111111 + 5/2.236111111}{2} = 2.236067978 \\ p_4 &= \frac{2.236067978 + 5/2.236067978}{2} = 2.236067978. \end{aligned}$$

Further iterations produce $p_k \approx 2.236067978$ for $k > 4$, so we see that convergence accurate to nine decimal places has been achieved. ■

Now let us turn to a familiar problem from elementary physics and see why determining the location of a root is an important task. Suppose that a projectile is fired from the origin with an angle of elevation b_0 and initial velocity v_0 . In elementary courses, air resistance is neglected and we learn that the height $y = y(t)$ and the distance traveled $x = x(t)$, measured in feet, obey the rules

$$(14) \quad y = v_y t - 16t^2 \quad \text{and} \quad x = v_x t,$$

where the horizontal and vertical components of the initial velocity are $v_x = v_0 \cos(b_0)$ and $v_y = v_0 \sin(b_0)$, respectively. The mathematical model expressed by the rules in (14) is easy to work with, but tends to give too high an altitude and too long a range for the projectile's path. If we make the additional assumption that the air resistance is proportional to the velocity, the equations of motion become

$$(15) \quad y = f(t) = (Cv_y + 32C^2) \left(1 - e^{-t/C}\right) - 32Ct$$

and

$$(16) \quad x = r(t) = Cv_x \left(1 - e^{-t/C}\right),$$

where $C = m/k$ and k is the coefficient of air resistance and m is the mass of the projectile. A larger value of C will result in a higher maximum altitude and a longer range for the projectile. The graph of a flight path of a projectile when air resistance is considered is shown in Figure 2.14. This improved model is more realistic, but requires the use of a root-finding algorithm for solving $f(t) = 0$ to determine the elapsed time until the projectile hits the ground. The elementary model in (14) does not require a sophisticated procedure to find the elapsed time.

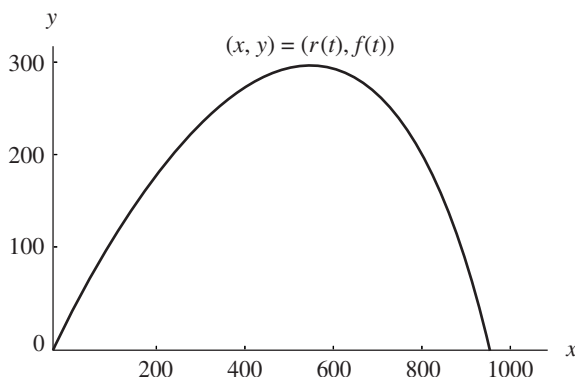


Figure 2.14 The path of a projectile with air resistance considered.

Table 2.4 Finding the Time When the Height $f(t)$ Is Zero

k	Time, p_k	$p_{k+1} - p_k$	Height, $f(p_k)$
0	8.00000000	0.79773101	83.22097200
1	8.79773101	-0.05530160	-6.68369700
2	8.74242941	-0.00025475	-0.03050700
3	8.74217467	-0.00000001	-0.00000100
4	8.74217466	0.00000000	0.00000000

Example 2.12. A projectile is fired with an angle of elevation $b_0 = 45^\circ$, $v_y = v_x = 160$ ft/sec, and $C = 10$. Find the elapsed time until impact and find the range.

Using formulas (15) and (16), the equations of motion are $y = f(t) = 4800(1 - e^{-t/10}) - 320t$ and $x = r(t) = 1600(1 - e^{-t/10})$. Since $f(8) = 83.220972$ and $f(9) = -31.534367$, we will use the initial guess $p_0 = 8$. The derivative is $f'(t) = 480e^{-t/10} - 320$, and its value $f'(p_0) = f'(8) = -104.3220972$ is used in formula (4) to get

$$p_1 = 8 - \frac{83.22097200}{-104.3220972} = 8.797731010.$$

A summary of the calculation is given in Table 2.4.

The value p_4 has eight decimal places of accuracy, and the time until impact is $t \approx 8.74217466$ seconds. The range can now be computed using $r(t)$, and we get

$$r(8.74217466) = 1600 \left(1 - e^{-0.874217466} \right) = 932.4986302 \text{ ft.} \quad \blacksquare$$

Division-by-Zero Error

One obvious pitfall of the Newton-Raphson method is the possibility of division by zero in formula (4), which would occur if $f'(p_{k-1}) = 0$. Program 2.5 has a procedure

to check for this situation, but what use is the last calculated approximation p_{k-1} in this case? It is quite possible that $f(p_{k-1})$ is sufficiently close to zero and that p_{k-1} is an acceptable approximation to the root. We now investigate this situation and will uncover an interesting fact, that is, how fast the iteration converges.

Definition 2.4. Assume that $f(x)$ and its derivatives $f'(x), \dots, f^{(M)}(x)$ are defined and continuous on an interval about $x = p$. We say that $f(x) = 0$ has a **root of order M** at $x = p$ if and only if

$$(17) \quad f(p) = 0, \quad f'(p) = 0, \quad \dots, \quad f^{(M-1)}(p) = 0, \quad \text{and} \quad f^{(M)}(p) \neq 0.$$

A root of order $M = 1$ is often called a *simple root*, and if $M > 1$, it is called a *multiple root*. A root of order $M = 2$ is sometimes called a *double root*, and so on. The next result will illuminate these concepts. \blacktriangle

Lemma 2.1. If the equation $f(x) = 0$ has a root of order M at $x = p$, then there exists a continuous function $h(x)$ so that $f(x)$ can be expressed as the product

$$(18) \quad f(x) = (x - p)^M h(x), \quad \text{where } h(p) \neq 0.$$

Example 2.13. The function $f(x) = x^3 - 3x + 2$ has a simple root at $p = -2$ and a double root at $p = 1$. This can be verified by considering the derivatives $f'(x) = 3x^2 - 3$ and $f''(x) = 6x$. At the value $p = -2$, we have $f(-2) = 0$ and $f'(-2) = 9$, so $M = 1$ in Definition 2.4; hence $p = -2$ is a simple root. For the value $p = 1$, we have $f(1) = 0$, $f'(1) = 0$, and $f''(1) = 6$, so $M = 2$ in Definition 2.4; hence $p = 1$ is a double root. Also, notice that $f(x)$ has the factorization $f(x) = (x + 2)(x - 1)^2$. \blacksquare

Speed of Convergence

The distinguishing property we seek is the following. If p is a simple root of $f(x) = 0$, Newton's method will converge rapidly, and the number of accurate decimal places (roughly) doubles with each iteration. On the other hand, if p is a multiple root, the error in each successive approximation is a fraction of the previous error. To make this precise, we define the **order of convergence**. This is a measure of how rapidly a sequence converges.

Definition 2.5. Assume that $\{p_n\}_{n=0}^{\infty}$ converges to p and set $E_n = p - p_n$ for $n \geq 0$. If two positive constants $A \neq 0$ and $R > 0$ exist, and

$$(19) \quad \lim_{n \rightarrow \infty} \frac{|p - p_{n+1}|}{|p - p_n|^R} = \lim_{n \rightarrow \infty} \frac{|E_{n+1}|}{|E_n|^R} = A,$$

then the sequence is said to converge to p with **order of convergence R** . The number A is called the *asymptotic error constant*. The cases $R = 1, 2$ are given special

Table 2.5 Newton's Method Converges Quadratically at a Simple Root

k	p_k	$p_{k+1} - p_k$	$E_k = p - p_k$	$\frac{ E_{k+1} }{ E_k ^2}$
0	-2.400000000	0.323809524	0.400000000	0.476190475
1	-2.076190476	0.072594465	0.076190476	0.619469086
2	-2.003596011	0.003587422	0.003596011	0.664202613
3	-2.000008589	0.000008589	0.000008589	
4	-2.000000000	0.000000000	0.000000000	

consideration.

(20) If $R = 1$, the convergence of $\{p_n\}_{n=0}^{\infty}$ is called **linear**.

If $R = 2$, the convergence of $\{p_n\}_{n=0}^{\infty}$ is called **quadratic**. ▲

If R is large, the sequence $\{p_n\}$ converges rapidly to p ; that is, relation (19) implies that for large values of n we have the approximation $|E_{n+1}| \approx A|E_n|^R$. For example, suppose that $R = 2$ and $|E_n| \approx 10^{-2}$; then we would expect that $|E_{n+1}| \approx A \times 10^{-4}$.

Some sequences converge at a rate that is not an integer, and we will see that the order of convergence of the secant method is $R = (1 + \sqrt{5})/2 \approx 1.618033989$.

Example 2.14 (Quadratic Convergence at a Simple Root). Start with $p_0 = -2.4$ and use Newton-Raphson iteration to find the root $p = -2$ of the polynomial $f(x) = x^3 - 3x + 2$. The iteration formula for computing $\{p_k\}$ is

$$(21) \quad p_k = g(p_{k-1}) = \frac{2p_{k-1}^3 - 2}{3p_{k-1}^2 - 3}.$$

Using formula (19) with $R = 2$ to check for quadratic convergence, we get the values in Table 2.5. ■

A detailed look at the rate of convergence in Example 2.14 will reveal that the error in each successive iteration is proportional to the square of the error in the previous iteration. That is,

$$|p - p_{k+1}| \approx A|p - p_k|^2,$$

where $A \approx 2/3$. To check this, we use

$$|p - p_3| = 0.000008589 \quad \text{and} \quad |p - p_2|^2 = |0.003596011|^2 = 0.000012931$$

and it is easy to see that

$$|p - p_3| = 0.000008589 \approx 0.000008621 = \frac{2}{3}|p - p_2|^2.$$

Table 2.6 Newton's Method Converges Linearly at a Double Root

k	p_k	$p_{k+1} - p_k$	$E_k = p - p_k$	$\frac{ E_{k+1} }{ E_k }$
0	1.200000000	-0.096969697	-0.200000000	0.515151515
1	1.103030303	-0.050673883	-0.103030303	0.508165253
2	1.052356420	-0.025955609	-0.052356420	0.496751115
3	1.026400811	-0.013143081	-0.026400811	0.509753688
4	1.013257730	-0.006614311	-0.013257730	0.501097775
5	1.006643419	-0.003318055	-0.006643419	0.500550093
\vdots	\vdots	\vdots	\vdots	\vdots

Example 2.15 (Linear Convergence at a Double Root). Start with $p_0 = 1.2$ and use Newton-Raphson iteration to find the double root $p = 1$ of the polynomial $f(x) = x^3 - 3x + 2$. Using formula (20) to check for linear convergence, we get the values in Table 2.6. ■

Notice that the Newton-Raphson method is converging to the double root, but at a slow rate. The values of $f(p_k)$ in Example 2.15 go to zero faster than the values of $f'(p_k)$, so the quotient $f(p_k)/f'(p_k)$ in formula (4) is defined when $p_k \neq p$. The sequence is converging linearly, and the error is decreasing by a factor of approximately $1/2$ with each successive iteration. The following theorem summarizes the performance of Newton's method on simple and double roots.

Theorem 2.6 (Convergence Rate for Newton-Raphson Iteration). Assume that Newton-Raphson iteration produces a sequence $\{p_n\}_{n=0}^{\infty}$ that converges to the root p of the function $f(x)$. If p is a simple root, convergence is quadratic and

$$(22) \quad |E_{n+1}| \approx \frac{|f''(p)|}{2|f'(p)|} |E_n|^2 \quad \text{for } n \text{ sufficiently large.}$$

If p is a multiple root of order M , convergence is linear and

$$(23) \quad |E_{n+1}| \approx \frac{M-1}{M} |E_n| \quad \text{for } n \text{ sufficiently large.}$$

Pitfalls

The division-by-zero error was easy to anticipate, but there are other difficulties that are not so easy to spot. Suppose that the function is $f(x) = x^2 - 4x + 5$; then the sequence $\{p_k\}$ of real numbers generated by formula (4) will wander back and forth from left to right and not converge. A simple analysis of the situation reveals that $f(x) > 0$ and has no real roots.

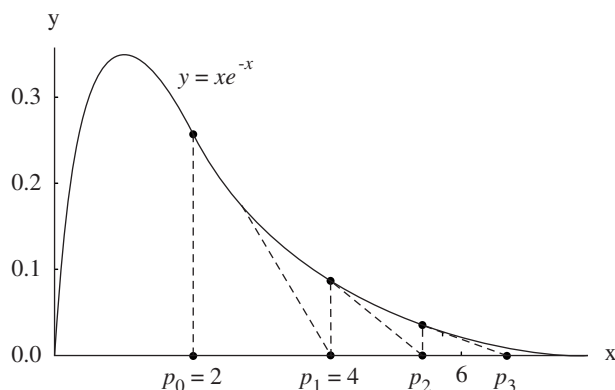


Figure 2.15 (a) Newton-Raphson iteration for $f(x) = xe^{-x}$ can produce a divergent sequence.

Sometimes the initial approximation p_0 is too far away from the desired root and the sequence $\{p_k\}$ converges to some other root. This usually happens when the slope $f'(p_0)$ is small and the tangent line to the curve $y = f(x)$ is nearly horizontal. For example, if $f(x) = \cos(x)$ and we seek the root $p = \pi/2$ and start with $p_0 = 3$, calculation reveals that $p_1 = -4.01525255$, $p_2 = -4.85265757$, \dots , and $\{p_k\}$ will converge to a different root $-3\pi/2 \approx -4.71238898$.

Suppose that $f(x)$ is positive and monotone decreasing on the unbounded interval $[a, \infty)$ and $p_0 > a$; then the sequence $\{p_k\}$ might diverge to $+\infty$. For example, if $f(x) = xe^{-x}$ and $p_0 = 2.0$, then

$$p_1 = 4.0, \quad p_2 = 5.333333333, \quad \dots, \quad p_{15} = 19.723549434, \quad \dots,$$

and $\{p_k\}$ diverges slowly to $+\infty$ (see Figure 2.15(a)). This particular function has another surprising problem. The value of $f(x)$ goes to zero rapidly as x gets large, for example, $f(p_{15}) = 0.0000000536$, and it is possible that p_{15} could be mistaken for a root. For this reason we designed the stopping criterion in Program 2.5 to involve the relative error $2|p_{k+1} - p_k|/(|p_k| + 10^{-6})$, and when $k = 15$, this value is 0.106817, so the tolerance $\delta = 10^{-6}$ will help guard against reporting a false root.

Another phenomenon, **cycling**, occurs when the terms in the sequence $\{p_k\}$ tend to repeat or almost repeat. For example, if $f(x) = x^3 - x - 3$ and the initial approximation is $p_0 = 0$, then the sequence is

$$\begin{aligned} p_1 &= -3.000000, & p_2 &= -1.961538, & p_3 &= -1.147176, & p_4 &= -0.006579, \\ p_5 &= -3.000389, & p_6 &= -1.961818, & p_7 &= -1.147430, & \dots & \end{aligned}$$

and we are stuck in a cycle where $p_{k+4} \approx p_k$ for $k = 0, 1, \dots$ (see Figure 2.15(b)). But if the starting value p_0 is sufficiently close to the root $p \approx 1.671699881$, then $\{p_k\}$

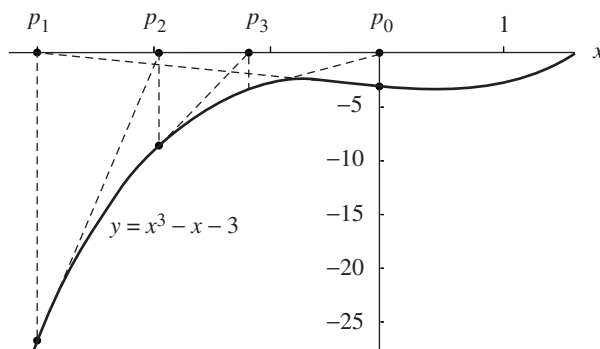


Figure 2.15 (b) Newton-Raphson iteration for $f(x) = x^3 - x - 3$ can produce a cyclic sequence.

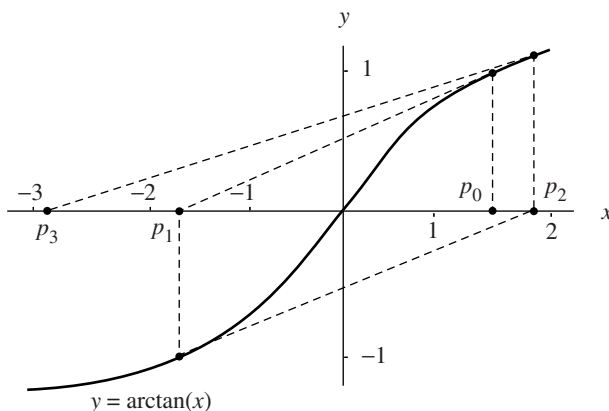


Figure 2.15 (c) Newton-Raphson iteration for $f(x) = \arctan(x)$ can produce a divergent oscillating sequence.

converges. If $p_0 = 2$, the sequence converges: $p_1 = 1.72727272$, $p_2 = 1.67369173$, $p_3 = 1.671702570$, and $p_4 = 1.671699881$.

When $|g'(x)| \geq 1$ on an interval containing the root p , there is a chance of divergent oscillation. For example, let $f(x) = \arctan(x)$; then the Newton-Raphson iteration function is $g(x) = x - (1 + x^2) \arctan(x)$, and $g'(x) = -2x \arctan(x)$. If the starting value $p_0 = 1.45$ is chosen, then

$$p_1 = -1.550263297, \quad p_2 = 1.845931751, \quad p_3 = -2.889109054,$$

etc. (see Figure 2.15(c)). But if the starting value is sufficiently close to the root $p = 0$,

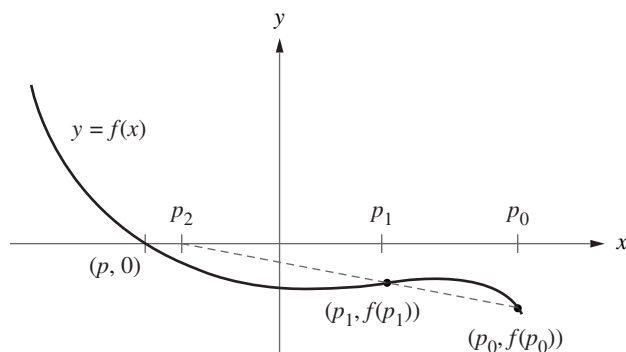


Figure 2.16 The geometric construction of p_2 for the secant method.

a convergent sequence results. If $p_0 = 0.5$, then

$$p_1 = -0.079559511, \quad p_2 = 0.000335302, \quad p_3 = 0.000000000.$$

The situations above point to the fact that we must be honest in reporting an answer. Sometimes the sequence does not converge. It is not always the case that after N iterations a solution is found. The user of a root-finding algorithm needs to be warned of the situation when a root is not found. If there is other information concerning the context of the problem, then it is less likely that an erroneous root will be found. Sometimes $f(x)$ has a definite interval in which a root is meaningful. If knowledge of the behavior of the function or an “accurate” graph is available, then it is easier to choose p_0 .

Numerical Methods Using Matlab, 4th Edition, 2004

John H. Mathews and Kurtis K. Fink

ISBN: 0-13-065248-2

Prentice-Hall Inc.

Upper Saddle River, New Jersey, USA

<http://vig.prenhall.com/>

NUMERICAL METHODS USING MATLAB

FOURTH EDITION



JOHN H. MATHEWS • KURTIS D. FINK