Chapter 2. Solution of a Single Nonlinear Equation

Section 1. Iteration Methods

1. Bisection method.(Burden & Faires, 2.1)
Consider the nonlinear equation

$$f(x) = 0, \quad a \le x \le b$$

The bisection method is the most simplest method to find a root of $f(x)$. Suppose $f(x)$ is a continuous function on $[a, b]$, with $f(a)$ and $f(b)$ of opposite sign. By the Intermediate Value Theorem, there exists a number $p \in (a, b)$ with $f(p) = 0$. For simplicity, we assume that the root p is unique. The algorithm of the bisection method is as follows,

- Let $a_1 = a$ and $b_1 = b$, and let p_1 be the midpoint of $[a_1, b_1]$,

$$p_1 = a_1 + \frac{b_1 - a_1}{2} = \frac{a_1 + b_1}{2}$$

- If $f(p_1) = 0$, then $p = p_1$. Otherwise, if $f(p_1) \neq 0$, then $f(p_1)$ has the same sign as either $f(a_1)$ or $f(b_1)$. When $f(p_1)$ and $f(a_1)$ has the same sign, $p \in (p_1, b_1)$, and we set $a_2 = p_1$, and $b_2 = b_1$. When $f(p_1)$ and $f(a_1)$ has the opposite sign, $p \in (a_1, p_1)$, and we set $a_2 = a_1$, and $b_2 = p_1$. We then set p_2 be the midpoint of $[a_2, b_2]$.

- This procedure is continued until the satisfactory approximation p_n is obtained.

Theorem 2.1. (convergence of bisection method) Suppose that $f \in C[a, b]$ and $f(a) \cdot f(b) < 0$. The bisection method generates a sequence $\{p_n\}_{n=1}^{\infty}$ approximating a zero p of f with

$$|p_n - p| \le \frac{b - a}{2^n}, \quad n \ge 1$$

Proof. For each $n \ge 1$, we have

$$b_n - a_n = \frac{1}{2^{n-1}} \text{ and } p \in (a_n, b_n)$$

Since $p_n = (a_n + b_n)/2$, then

$$|p_n - p| \le \frac{b_n - a_n}{2} = \frac{b - a}{2^n} \quad \blacksquare$$

Using this theorem we can determine the number of iterations n for a given accuracy ϵ. Set

$$|p_n - p| \le \frac{b - a}{2^n} \le \epsilon, \quad n \ge 1$$

or

$$2^n \geq \frac{b-a}{\epsilon}$$

Take logarithms we get

$$n \log_{10} 2 \geq \log_{10}(b-a) - \log_{10} \epsilon$$

or

$$n \geq \frac{\log_{10}(b-a) - \log_{10} \epsilon}{\log_{10} 2}$$

For example, if $a = 1$, $b = 2$, and $\epsilon = 10^{-3}$, we have

$$n \geq \frac{\log_{10} 1 - \log_{10} 10^{-3}}{\log_{10} 2} \geq \frac{3}{0.30103} \approx 9.97$$

2. Fixed point iteration.(Burden & Faires, 2.2)

A number p is called a **fixed point** for a given function g if $g(p) = p$.

Theorem 2.2.

a. If $g \in C[a,b]$ and $g(x) \in [a,b]$ for all $x \in [a,b]$, then g has a fixed point in $[a,b]$.

b. If, in addition, $g'(x)$ exists on (a,b) and a positive constant $k < 1$ exists with

$$|g'(x)| \leq k \quad \forall x \in (a,b)$$

then the fixed point in $[a,b]$ is unique.

Proof.

a. If $g(a) = a$ or $g(b) = b$, then g has a fixed point at an endpoint. If not, then $g(a) > a$ and $g(b) < b$. The function $h(x) = g(x) - x$ is continuous on $[a,b]$ with

$$h(a) = g(a) - a > 0 \text{ and } h(b) = g(b) - b < 0$$

The Intermediate Value Theorem implies that there exists $p \in (a,b)$ such that $h(p) = 0$, or $g(p) = p$.

b. If p and q, $p \neq q$, are both fixed points, then the Mean Value Theorem implies that a number ξ exists between p and q with

$$\frac{g(p) - g(q)}{p - q} = g'(\xi)$$

Thus,

$$|p - q| = |g(p) - g(q)| = |g'(\xi)| \leq k|p - q| < |p - q|$$

which is a contradiction. ■

To approximate the fixed point p, we take an initial guess p_0, and compute $p_n = g(p_{n-1})$. The following theorem shows that the p_n converges to p under some conditions.

Theorem 2.3. (fixed point iteration) Let $g \in C[a,b]$ be such that $g(x) \in [a,b]$, for all $x \in [a,b]$. Suppose, in addition, that g' exists on (a,b) and that a constant $0 < k < 1$ exists with

$$|g'(x)| \leq k \quad \forall x \in (a,b)$$

Then, for any $p_0 \in [a,b]$, the sequence defined by

$$p_n = g(p_{n-1}), \quad n \geq 1$$

converges to the unique fixed point $p \in [a,b]$.

Proof. Using the Mean Value Theorem we have

$$|p_n - p| = |g(p_{n-1}) - g(p)| = |g'(\xi_n)||p_n - p| \leq k|p_{n-1} - p|$$

Applying this repeatedly we get

$$|p_n - p| =\leq k|p_{n-1} - p| \leq k^2|p_{n-2} - p| \leq \cdots \leq k^n|p_0 - p|$$

Therefore,

$$\lim_{n\to\infty} |p_n - p| = 0 \quad \blacksquare$$

Corollary 2.4. If g satisfies the hypotheses of Theorem 2.3, then the error is bounded by

$$|p_n - p| =\leq k^n \max\{p_0 - a, b - p_0\}$$

and

$$|p_n - p| =\leq \frac{k^n}{1-k}|p_1 - p_0| \quad n \geq 1$$

3. Newton's method.(Burden & Faires, 2.3)

Newton's is a very efficient method for finding the roots of $f(x) = 0$. Newton's method can be derived by using Taylor's theorem. Suppose that $f \in C^2[a,b]$. Let $p_0 \in [a,b]$ be an good approximation to the root p such that $f'(p_0) \neq 0$. Then we have

$$f(p) = f(p_0) + (p - p_0)f'(p_0) + \frac{(p-p_0)^2}{2}f''(\xi)$$

Dropping the second order term and using $f(p) = 0$ we find a new approximation to p,

$$p \approx p_1 = p_0 - \frac{f(p_0)}{f'(p_0)}$$

In general, we have the iteration algorithm,

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})} \text{ for } n \geq 1$$

Theorem 2.5. (local convergence of Newton's method) Let $f \in C^2[a,b]$, if $p \in [a,b]$ is such that $f(p) = 0$ and $f'(p) \neq 0$, then there exists a $\delta > 0$ such that Newton's method converges for any initial approximation $p_0 \in [p-\delta, p+\delta]$.

Proof. Let

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

then Newton's method is equivalent to the fixed point iteration $p_n = g(p_{n-1})$. Thus, we only need to check the conditions in the convergence theorem for the fixed point iteration. Take the derivative

$$g'(x) = \frac{f(x)f''(x)}{[f'(x)]^2}$$

Since $g'(p) = 0$, there exists a $\delta > 0$, such that

$$|g'(x)| \leq k < 1 \quad \forall x \in [p-\delta, p+\delta]$$

Also, since

$$|g(x) - p| = |g(x) - g(p)| = |g'(\xi)||x-p| \leq k|x-p| < |x-p| < \delta$$

g maps $[p-\delta, p+\delta]$ into $[p-\delta, p+\delta]$. Therefore, from the convergence theorem for the fixed point iteration, Newton's method converges to p for any $p_0 \in [p-\delta, p+\delta]$. ■

Secant method. Newton's method is a very powerful method. However, in some application problems, $f'(p_{n-1})$ is not easy to find. To avoid this difficulty, we use an approximation to replace $f'(p_{n-1})$

$$f'(p_{n-1}) \approx \frac{f(p_{n-1}) - f(p_{n-2})}{p_{n-1} - p_{n-2}}$$

Then Newton's method is modified to

$$p_n = p_{n-1} - \frac{f(p_{n-1})(p_{n-1} - p_{n-2})}{f(p_{n-1}) - f(p_{n-2})}$$

This method is called the Secant Method.

Brent's method. Brent's method is a modification of the secant method, which combines the bisection method and the secant method. At each step, if the new approximation from the secant method is not inside the new subinterval, then the new approximation is replaced by the bisection point, and the new subinterval is chosen to guarantee that the root is inside.

Section 2. Convergence Analysis

Convergence order. If

$$\lim_{n\to\infty} \frac{|p_{n+1}-p|}{|p_n-p|^\alpha} = \lambda$$

for some constant λ, then we say that the sequence $\{p_n\}_{n=0}^{\infty}$ converges to p of order α. When $\alpha = 1$ ($\lambda < 1$), the sequence is linearly convergent. When $\alpha = 2$, the sequence is quadratically convergent.

Theorem 2.7. Let $g \in C[a,b]$ be such that $g(x) \in [a,b]$, for all $x \in [a,b]$. Suppose, in addition, that g' is continuous on (a,b) and

$$|g'(x)| \le k \quad \forall x \in (a,b)$$

with $k < 1$. If $g'(p) \neq 0$, then for $p_0 \in [a,b]$, the sequence

$$p_n = g(p_{n-1}) \quad \forall n \ge 1$$

converges only linearly to the unique fixed point $p \in [a,b]$.

Proof. Since

$$p_{n+1} - p = g(p_n) - g(p) = g'(\xi_n)(p_n - p)$$

then

$$\lim_{n\to\infty} \frac{|p_{n+1}-p|}{|p_n-p|} = \lim_{n\to\infty} |g'(\xi_n)| = |g'(p)| \quad \blacksquare$$

Theorem 2.8. Let p be a solution of the equation $x = g(x)$. Suppose that $g'(p) = 0$ and g'' is continuous with $|g''(x)| < M$ on an open interval I containing p. Then there exists a number $\delta > 0$ such that, for $p_0 \in [p-\delta, p+\delta]$, the sequence defined by $p_n = g(p_{n-1})$, $n \ge 1$, converges at least quadratically to p. Moreover, for sufficiently large value of n,

$$|p_{n+1} - p| < \frac{M}{2}|p_n - p|^2$$

Proof. Using Taylor expansion we have

$$\begin{aligned} g(x) &= g(p) + g'(p)(x-p) + \frac{g''(\xi)}{2}(x-p)^2 \\ &= p + \frac{g''(\xi)}{2}(x-p)^2 \end{aligned}$$

Let $x = p_n$ we get

$$p_{n+1} = g(p_n) = p + \frac{g''(\xi_n)}{2}(p_n - p)^2$$

or

$$p_{n+1} - p = \frac{g''(\xi_n)}{2}(p_n - p)^2$$

Thus, we have

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^2} = \frac{|g''(p)|}{2} < \frac{M}{2} \quad \blacksquare$$

For Newton's method,

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

It is easy to check that $g'(p) = 0$ and g'' is continuous with $|g''(x)| < M$. Thus, Newton's method converges quadratically.

Section 3. Zeros of Polynomials

There are different ways to find the zeros of polynomials, here we only introduce one simple method, **Bairstow's Method**. Consider the polynomial

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_n = 0 \tag{1}$$

The idea is to find a quadratic factor of the form

$$x^2 + ux + v$$

where u and v are constants to be determined. Once this quadratic factor is obtained, then the degree of the polynomial is reduced by 2, and we continue to find other quadratic factors. In this way, we find all quadratic factors, and thus find all the roots. Let

$$\begin{aligned} &a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_n = \\ &(x^2 + ux + v)(b_0x^{n-2} + b_1x^{n-3} + b_2x^{n-4} + \cdots + b_{n-2}) + R(x) \end{aligned}$$

where

$$R(x) = rx + s$$

To ensure that $x^2 + ux + v$ is a factor of the polynomial (1), the remainder $R(x)$ must be zero. Since r and s depend on u and v, we have

$$\begin{aligned} r(u, v) &= 0 \\ s(u, v) &= 0 \end{aligned}$$

Suppose initial approximations u_0 and v_0 are given, let $u_1 = u_0 + \Delta u_0$ and $v_1 = v_0 + \Delta v_0$. We require

$$\begin{aligned} r(u_1, v_1) &= 0 \\ s(u_1, v_1) &= 0 \end{aligned}$$

or,

$$r(u_0 + \Delta u_0, v_0 + \Delta v_0) = 0$$
$$s(u_0 + \Delta u_0, v_0 + \Delta v_0) = 0$$

Use Taylor's theorem we obtain

$$r(u_0, v_0) + \Delta u_0 \frac{\partial r(u_0, v_0)}{\partial u} + \Delta v_0 \frac{\partial r(u_0, v_0)}{\partial v} \approx 0$$
$$s(u_0, v_0) + \Delta u_0 \frac{\partial s(u_0, v_0)}{\partial u} + \Delta v_0 \frac{\partial s(u_0, v_0)}{\partial v} \approx 0$$

Solving this we obtain Δu_0 and Δv_0, and therefore we have u_1 and v_1. This procedure is repeated until r and s are sufficiently small.