7. Solution of Nonlinear Equations

We consider the numerical approximation of the roots of a single nonlinear equation $F(x) = 0$ (e.g., $F(x) = x - e^{-x} = 0$) and of a system of nonlinear equation $\boldsymbol{F}(\boldsymbol{x}) = 0$, where $\boldsymbol{F} = (F_1, \ldots, F_n)$ and $\boldsymbol{F}_i(\boldsymbol{x}) = \boldsymbol{F}_i(x_1, \ldots, x_n)$.

The methods will be iterative and we consider the issues (i) under what conditions does the iteration converge (to a root) and (ii) how fast does the iteration converge.

In some cases, we will be able to show that the iteration defining the method converges for all initial guesses in some specific range. In general, we will settle for a *local convergence result* which says that the iteration converges if the starting guess is *sufficiently close* to the root.

7.1. **Iterative methods for roots of a single nonlinear equation.** Suppose $F(x)$ is a continuous function on the interval $[a, b]$ and satisfies $F(a)F(b) < 0$. Then the Intermediate Value Theorem says there is at least one number s, with $a < s < b$, such that $F(s) = 0$. The simplest scheme to find a root s is to use the *method of bisection*.

Bisection Algorithm: Set $a_0 = a$ and $b_0 = b$. For $n = 0, 1, \ldots$,
Set $x_n = (a_n + b_n)/2$.
If $F(a_n)F(x_n) < 0$, set $a_{n+1} = a_n$, $b_{n+1} = x_n$.
If $F(x_n)F(b_n) < 0$, set $a_{n+1} = x_n$, $b_{n+1} = b_n$.
Then at each stage of the iteration, the root lies in the interval spanned by a_n and b_n. Hence

$$|s - x_n| \leq |b_n - a_n|/2.$$

Now since $|b_n - a_n| = |b_{n-1} - a_{n-1}|/2$, we easily get that

$$|s - x_n| \leq |b - a|/2^{n+1}, \quad n \geq 0.$$

Thus, we can achieve any desired accuracy by taking n sufficiently large.

Method of False Position: Instead of choosing x_n as the midpoint of the points bracketing the root, we choose it as the weighted average of these points, with the weights depending on the size of the function values. This can be done by choosing x_n as the point where the secant line joining the points $(a_n, F(a_n)$ and $(b_n, F(b_n))$ crosses the x axis. This line is given by

$$y - F(a_n) = \frac{F(b_n) - F(a_n)}{b_n - a_n}(x - a_n).$$

When $y = 0$, we get that

$$x \equiv x_n = a_n - \frac{b_n - a_n}{F(b_n) - F(a_n)} F(a_n).$$

To maintain the root bracketing property, we could then proceed as in the bisection algorithm, i.e.,
If $F(a_n)F(x_n) < 0$, set $a_{n+1} = a_n$, $b_{n+1} = x_n$.
If $F(x_n)F(b_n) < 0$, set $a_{n+1} = x_n$, $b_{n+1} = b_n$.
However, although the method of false position produces a point at which $|F(x)|$ is small

somewhat faster than the bisection method, it does not give a small interval in which the root is known to lie.

Consider the following example: $f(x) = x^2 - 1$, $a = 0$, $b = 2$.

Then all the iterates x_n lie to the left of the root s, so although $s \in [x_n, b]$, $|b - x_n| \geq |b - s|$ for all n, i.e., the size of the interval in which the root is known to lie is not converging to zero.

Secant method: This method is similar to the method of false position, except that we drop the requirement that the root be bracketed. Thus, starting from two values x_0 and x_1 which do bracket the root, we simply define the sequence $\{x_n\}$ by

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{F(x_n) - F(x_{n-1})} F(x_n).$$

When it converges, this method converges faster than the method of false position.

Newton's method: Geometrically, starting from an initial guess x_0, we define at each step a new approximation x_{n+1} as the position where the tangent line to the curve $y = F(x)$ at $x = x_n$ crosses the x-axis. Since the equation of this tangent line is given by

$$y - F(x_n) = F'(x_n)(x - x_n),$$

we get that when $y = 0$,

$$x = x_{n+1} = x_n - F(x_n)/F'(x_n).$$

Another way to think of Newton's method is that it is the approximation given by truncating the Taylor series expansion, i.e., we have

$$0 = F(s) = F(x_n) + F'(x_n)(s - x_n) + F''(\xi)(s - x_n)^2.$$

If x_n is close to s, then $s - x_n$ is small, so that $(s - x_n)^2$ is even smaller. Discarding this last term, we define x_{n+1} as the approximation to s which restores equality to this equation, i.e.,

$$F(x_n) + F'(x_n)(x_{n+1} - x_n) = 0.$$

Solving for x_{n+1}, we recover Newton's method.

To study the convergence of some of these methods, we next consider a scheme called fixed point iteration. In this method, instead of seeking a root of $F(x) = 0$, we look for a fixed point of a function $f(x)$, i.e., a value of x satisfying $x = f(x)$. Note that if we define $f(x) = x - F(x)/F'(x)$, then a simple root x^* of F (i.e., $F'(x^*) \neq 0$) will be a fixed point of $f(x)$.

Fixed point iteration: Given a starting guess x_0, we define the iteration $x_{n+1} = f(x_n)$. We then have the following convergence result for this iteration scheme.

Theorem 23. *Let $I = [a, b]$, where a and b are finite and assume that f satisfies the following conditions: (i) f is continuous on I, (ii) $f(x) \in I$ for all $x \in I$, and (iii) $|f(x_2) - f(x_1)| \leq L|x_2 - x_1|$, with $L < 1$ for all $x_1, x_2 \in I$. Then there is a unique fixed point s of f (i.e., $s = f(s)$) in the interval I and for any choice of $x_0 \in I$, the sequence $\{x_n\}$ defined by the iteration $x_{n+1} = f(x_n)$ converges to s.*

Proof. To prove existence of a fixed point, we set $g(x) = x - f(x)$. Since by (ii), $a \le f(a) \le b$ and $a \le f(b) \le b$, $g(a) = a - f(a) \le 0$ and $g(b) = b - f(b) \ge 0$. Since f is continuous on I, so is g. Hence, by the Intermediate Value Theorem, there exists at least one point s in $[a, b]$ such that $g(s) = 0$, i.e., $s = f(s)$. To see there can be only one such point, we suppose there are two fixed points s_1 and s_2. Then using (iii),

$$|s_2 - s_1| = |f(s_2) - f(s_1) \le L|s_2 - s_1|.$$

Since $L < 1$, we must have $s_2 = s_1$. To establish convergence, we note that

$$|s - x_n| = |f(s) - f(x_{n-1})| \le L|s - x_{n-1}| \le L^2|s - x_{n-2}| \le \ldots \le L^n|s - x_0|.$$

Since $L < 1$, $\lim_{n\to\infty} = 0$ and so $\lim_{n\to\infty} |s - x_n| = 0$, i.e., $\lim_{n\to\infty} x_n = s$. □

We can also derive error bounds on the approximation that do not depend on the unknown solution.

Corollary 5.

$$|s - x_n| \le L^n \max\{b - x_0, x_0 - a\}.$$

From the proof of the theorem, we know that $|s - x_n| \le L^n|s - x_0|$. Since both x_0 and s belong to I, either $s \in [a, x_0]$ or $s \in [x_0, b]$. Hence, $|s - x_0| \le L^n \max\{b - x_0, x_0 - a\}$.

It is also possible to establish the following result.

Corollary 6.

$$|s - x_n| \le \frac{L^n}{1 - L}|x_1 - x_0|.$$

Note that if f' is also continuous and $|f'(\xi)| \le L < 1$ for all $\xi \in [a, b]$, then by the Mean Value Theorem, we have for all $x_1, x_2 \in [a, b]$, there exists $\xi \in [a, b]$ such that

$$|f(x_2) - f(x_1)| = |f'(\xi)||x_2 - x_1| \le L|x_2 - x_1|,$$

so this is a simple way to show that (iii) is satisfied.

Example: Use the theorem on fixed point iteration to prove the convergence of the iteration scheme: $x_{n+1} = g(x_n)$ for any $x_0 \in [-1, 1]$, when $g(x) = (x^2 - 1)/3$. (i) Since g is a polynomial, it is continuous and differentiable everywhere. (ii) We find the maximum and minimum of $g(x)$ on $I = [-1, 1]$. Now $g'(x) = 2x/3 = 0$ only for $x = 0$. Hence the max and min can occur only at $x = -1, 0, 1$. Since $g(-1) = 0$, $g(1) = 0$ and $g(0) = -1/3$, we get $-1/3 \le g(x) \le 0$. Hence $g(x) \in [-1, 1]$ for all $x \in [-1, 1]$ and (ii) is satisfied. (iii) is also satisfied since $|g'(x)| = |2x/3| \le 2/3 = L < 1$ for $x \in [-1, 1]$. Hence, the iteration $x_{n+1} = (x_n^2 - 1)/3$ converges to the unique fixed point of g in $[-1, 1]$.