Chapter 8

Nonlinear algebraic equations

8.1 Continuous functions

A function $f : I \to \mathbb{R}$, is said to be *continuous* on the interval $I = [a, b]$, if for each $\epsilon > 0$ we can find a $\delta > 0$, such that

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon, \quad \forall x, y \in I, \tag{8.1}$$

and *Lipschitz continuous* on the interval I, if there exists a real number $L_f > 0$, the *Lipschitz constant* of f, such that

$$|f(x) - f(y)| \leq L_f |x - y|, \quad \forall x, y \in I. \tag{8.2}$$

The vector space of real valued continuous functions on the interval I is denoted by $\mathcal{C}^0(I)$, or $\mathcal{C}(I)$, which is closed under the basic operations of pointwise addition and scalar multiplication, defined by,

$$f + g = f(x) + g(x), \qquad \forall x \in I, \tag{8.3}$$

$$\alpha f = \alpha f(x), \qquad \forall x \in I, \tag{8.4}$$

for $f, g \in \mathcal{C}(I)$ and $\alpha \in \mathbb{R}$. Similarly, we let $Lip(I)$ denote the vector space of Lipschitz continuous functions together with the same basic operations, and we note that any Lipschitz continuous function is also continuous, that is $Lip(I) \subset \mathcal{C}(I)$.

Further, we denote the vector space of continuous functions with also continuous derivatives up to the order k by $\mathcal{C}^k(I)$, with $\mathcal{C}^\infty(I)$ the vector space of continuous functions with continuous derivatives of arbitrary order.

Local approximation of continuous functions

By *Taylor's theorem* we can construct a local approximation of a function $f \in \mathcal{C}^k(I)$ near any point $y \in I$, in terms of $f(y)$ and the derivatives $f^{(i)}(y)$, for $i \leq k$. In particular, we can approximate $f(x)$ by a linear function,

$$f(x) \approx f(y) + f'(y)(x - y), \tag{8.5}$$

corresponding to the *tangent* of the function at $x = y$, with the approximation error decreasing quadratically with the distance $|x - y|$.

Theorem 13 (Tayor's theorem). *For $f \in \mathcal{C}^2(I)$, we have that*

$$f(x) = f(y) + f'(y)(x - y) + \frac{1}{2} f''(\xi)(x - y)^2, \tag{8.6}$$

for $y \in I$ and $\xi \in [x, y]$.

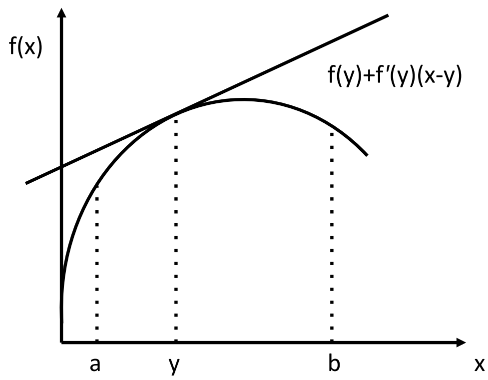

Figure 8.1: The tangent line $f(y) + f'(y)(x - y)$ at $y \in [a, b]$.

8.2 Nonlinear scalar equations

Fixed point iteration

For a nonlinear function $f : \mathbb{R} \to \mathbb{R}$, we seek a solution x to the equation

$$f(x) = 0, \tag{8.7}$$

for which we can formulate a fixed point iteration $x^{(k+1)} = g(x^{(k)})$, as

$$x^{(k+1)} = g(x^{(k)}) = x^{(k)} + \alpha f(x^{(k)}), \tag{8.8}$$

where α is a parameter to be chosen. The fixed point iteration (8.8) converges to a unique fixed point $x = g(x)$ that satisfies equation (8.7), under the condition that the function $g \in Lip(\mathbb{R})$ with $L_g < 1$, which we can prove similar to the case of a linear system of equations (7.17).

For any $k > 1$, we have that

$$|x^{(k+1)} - x^{(k)}| = |g(x^{(k)}) - g(x^{(k-1)})| \leq L_g|x^{(k)} - x^{(k-1)}| \leq L_g^k|x^{(1)} - x^{(0)}|,$$

so that for $m > n$,

$$\begin{aligned} |x^{(m)} - x^{(n)}| &= |x^{(m)} - x^{(m-1)}| + ... + |x^{(n+1)} - x^{(n)}| \\ &\leq (L_g^{m-1} + ... + L_g^n)\,|x^{(1)} - x^{(0)}|, \end{aligned}$$

by the triangle inequality, and with $L_g < 1$, we have that $\{x^{(n)}\}_{n=1}^{\infty}$ is a Cauchy sequence,

$$\lim_{n\to\infty} |x^{(m)} - x^{(n)}| = 0, \tag{8.9}$$

which implies that there exists an $x \in \mathbb{R}$, such that

$$x = \lim_{n\to\infty} x^{(n)}, \tag{8.10}$$

since $\mathbb{R}$ is a Banach space.

Uniqueness of x follows from assuming that there exists another solution $y \in \mathbb{R}$ such that $y = g(y)$, for which we have that

$$|x - y| = |g(x) - g(y)| \leq L_g|x - y| \Rightarrow (1 - L_g)|x - y| \leq 0 \Rightarrow |x - y| = 0,$$

and thus $x = y$ is the unique solution to the equation $x = g(x)$.

Newton's method

The analysis above suggests that (8.8) converges linearly, since for the error $e^{(k)} = x - x^{(k)}$ we have that $|e^{(k+1)}| \leq L_g|e^{(k)}|$, by

$$|x - x^{(k+1)}| = |g(x) - g(x^{(k)})| \leq L_g|x - x^{(k)}|. \tag{8.11}$$

Although, for the choice $\alpha = -f'(x^{(k)})^{-1}$, which we refer to as *Newton's method*, the fixed point iteration (8.8) exhibits quadratic convergence. The geometric interpretation of Newton's method is that $x^{(k+1)}$ is determined from the tangent of the function f at $x^{(k)}$.

The quadratic convergence rate of Newton's method follows from *Taylor's formula*, with $f(x)$ expanded around $x^{(k)}$,

$$0 = f(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2}f''(\xi)(x - x^{(k)})^2, \tag{8.12}$$

Algorithm 11: Newton's method

Given initial approximation $x^{(0)} \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$
while $|f(x^{(k)})| \geq TOL$ **do**
  Compute $f'(x^{(k)})$
  $x^{(k+1)} = x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})$
end

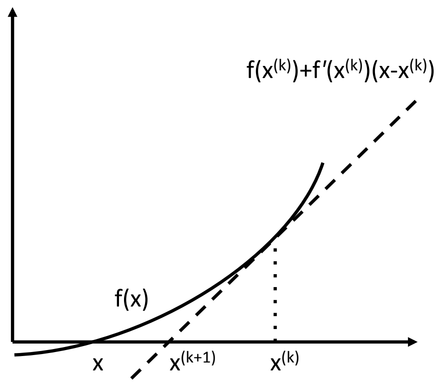

Figure 8.2: Geometric interpretation of Newtons method with the approximation $x^{(k+1)}$ obtained as the zero value of the tangent at $x^{(k)}$.

with $\xi \in [x, x^{(k)}]$. We divide by $f'(x^{(k)})$ to get

$$x - (x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})) = -\frac{1}{2} f'(x^{(k)})^{-1} f''(\xi)(x - x^{(k)})^2, \qquad (8.13)$$

so that with $x^{(k+1)} = x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})$, we get

$$|e^{(k+1)}| = \frac{1}{2} |f'(x^{(k)})^{-1} f''(\xi)| \, |e^{(k)}|^2, \qquad (8.14)$$

which displays the quadratic convergence of the sequence $x^{(k)}$ close to x.

8.3 Systems of nonlinear equations

Continuous functions in $\mathbb{R}^n$

In $\mathbb{R}^n$ we define continuity in terms of norms, so that a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is *continuous*, denoted $f \in \mathcal{C}(\mathbb{R}^n)$, if for each $\epsilon > 0$ we can find a $\delta > 0$, such that

$$\|x-y\| < \delta \Rightarrow \|f(x)-f(y)\| < \epsilon, \quad \forall x,y \in \mathbb{R}^n, \tag{8.15}$$

and *Lipschitz continuous*, denoted $f \in Lip(\mathbb{R}^n)$, if there exists a real number $L_f > 0$, such that

$$\|f(x)-f(y)\| \leq L_f \|x-y\|, \quad \forall x,y \in \mathbb{R}^n. \tag{8.16}$$

Similarly, the vector space of continuous functions with also continuous derivatives up to the order k is denoted by $\mathcal{C}^k(I)$, with $\mathcal{C}^\infty(I)$ the vector space of continuous functions with continuous derivatives of arbitrary order.

Fixed point iteration for nonlinear systems

Now consider a system of nonlinear equations: find $x \in \mathbb{R}^n$, such that

$$f(x) = 0, \tag{8.17}$$

with $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, for which we can formulate a fixed point iteration $x^{(k+1)} = g(x^{(k)})$, with $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, just as in the case of the scalar problem.

Algorithm 12: Fixed point iteration for solving the system $f(x) = 0$

Given initial approximation $x^{(0)} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$
while $\|f(x^{(k)})\| \geq TOL$ **do**
  $x^{(k+1)} = x^{(k)} + \alpha f(x^{(k)})$
end

Existence of a unique solution to the fixed point iteration follows by the Banach fixed point theorem.

Theorem 14 (Banach fixed point theorem in $\mathbb{R}^n$). *The fixed point iteration of Algorithm 12 converges to a unique solution if $L_g < 1$, with L_g the Lipschitz constant of the function $g(x) = x + \alpha f(x)$.*

Proof. For $k > 1$ we have that

$$\begin{aligned} \|x^{(k+1)} - x^{(k)}\| &= \|x^{(k)} - x^{(k-1)} + \alpha(f(x^{(k)}) - f(x^{(k-1)}))\| \\ &= \|g(x^{(k)}) - g(x^{(k-1)})\| \leq L_g^k \|x^{(1)} - x^{(0)}\|, \end{aligned}$$

and for $m > n$,

$$\begin{aligned} \|x^{(m)} - x^{(n)}\| &= \|x^{(m)} - x^{(m-1)}\| + \ldots + \|x^{(n+1)} - x^{(n)}\| \\ &\leq (L_g^{m-1} + \ldots + L_g^n)\|x^{(1)} - x^{(0)}\|. \end{aligned}$$

Since $L_g < 1$, $\{x^{(n)}\}_{n=1}^{\infty}$ is a Cauchy sequence, which implies that there exists an $x \in \mathbb{R}^n$ such that

$$x = \lim_{n\to\infty} x^{(n)},$$

since $\mathbb{R}^n$ is a Banach space. Uniqueness follows from assuming that there exists another solution $y \in \mathbb{R}^n$ such that $f(y) = 0$, so that

$$\|x - y\| = \|g(x) - g(y)\| \leq L_g\|x - y\| \Rightarrow (1 - L_g)\|x - y\| \leq 0,$$

and thus $x = y$ is the unique solution to the equation $f(x) = 0$. □

Newton's method for nonlinear systems

The *Jacobian matrix* $f'(x) \in \mathbb{R}^{n\times n}$, at a point $x \in \mathbb{R}^n$, is defined as

$$f'(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{bmatrix} = \begin{bmatrix} (\nabla f_1(x))^T \\ \vdots \\ (\nabla f_n(x))^T \end{bmatrix}, \tag{8.18}$$

with the *gradient* $\nabla f(x) \in \mathbb{R}^n$, defined by

$$\nabla f_i(x) = (\frac{\partial f_i}{\partial x_1}(x), ..., \frac{\partial f_i}{\partial x_n}(x))^T. \tag{8.19}$$

Newton's method for a system in $\mathbb{R}^n$ is analogous to the method for scalar equations, but with the inverse of the derivative of the function replaced by the inverse of the jacobian matrix $(f'(x^{(k)}))^{-1}$. The inverse is not constructed explicitly, instead we solve a linear system of equations $Ax = b$, with $A = f'(x^{(k)})$ the jacobian, $x = \Delta x^{(k+1)} = x^{(k+1)} - x^{(k)}$ the increment, and $b = f(x^{(k)})$ the residual.

Depending on how the jacobian is computed, and the linear system solved, we get different versions of Newton's method. If the system is large and sparse, we use iterative methods for solving the linear system $Ax = b$, and if the jacobian $f'(\cdot)$ is not available we use an approximation, obtained, for example, by a difference approximation based on the function $f(\cdot)$.

Quadratic convergence rate of Newton's method for systems follows from *Taylor's formula* in $\mathbb{R}^n$, which states that

$$f(x) - (f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)})) = \mathcal{O}(\|x - x^{(k)}\|^2),$$

so that for $f(x) = 0$, we have that,

$$x - (x^{(k)} - f'(x^{(k)})^{-1}f(x^{(k)})) = \mathcal{O}(\|x - x^{(k)}\|^2),$$

Algorithm 13: Newton's method for systems of nonlinear equations

Given initial approximation $x^{(0)} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$
while $\|f(x^{(k)})\| \geq TOL$ **do**
  Compute $f'(x^{(k)})$ ▷ `compute jacobian`
  $f'(x^{(k)})\Delta x^{(k+1)} = -f(x^{(k)})$ ▷ `solve for` $\Delta x^{(k+1)}$
  $x^{(k+1)} = x^{(k)} + \Delta x^{(k+1)}$ ▷ `update by` $\Delta x^{(k+1)}$
end

which gives for the error $e^{(k)} = x - x^{(k)}$, that

$$\frac{\|e^{(k+1)}\|}{\|e^{(k)}\|^2} = \mathcal{O}(1), \tag{8.20}$$

since $x^{(k+1)} = x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})$. The quadratic convergence rate for $x^{(k)}$ close to x then follows for $x^{(k)}$ close to x.