15 Solution of Nonlinear Equations

In this chapter, we depart from *linear algebra* problems and concentrate on the study of *nonlinear* problems. We will focus on methods for solving a nonlinear system of equations. In other words, given $m, n \in \mathbb{N}$ and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, we wish to find a point $\boldsymbol{\xi} \in \mathbb{R}^n$ such that

$$\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}. \tag{15.1}$$

Such a point $\boldsymbol{\xi}$, if it exists, is called a *root of* $\boldsymbol{f}$. Of course, if $m = n$ and $\boldsymbol{f}$ so happens to be affine then this problem reduces to (3.1), and can be treated either by the direct methods of Chapter 3 or by the iterative ones of Chapters 6 and 7. If $m \neq n$, but the function $\boldsymbol{f}$ is still affine, then the least squares methods of Chapter 5 apply. The function $\boldsymbol{f}$ in this chapter, however, is not assumed to be affine. The importance of (15.1) cannot be overstated. Similar to (3.1), many problems can be reduced to this. We already saw an instance of this in Chapter 8: although this may not be the preferred choice, finding eigenvalues can be reduced to solving a problem like (15.1) with $m = n = 1$ and the function is a polynomial. Other examples will be seen in future chapters.

We must immediately remark that any method that attempts to find a solution to (15.1) must be iterative, unless a very special structure is assumed on the function $\boldsymbol{f}$. The general strategy that we will follow can be simply stated as:

- Show that the problem has at least one solution.
- Isolate a root, i.e., find an open region $D \subset \mathbb{R}^n$ for which there is $\boldsymbol{\xi} \in D$ that solves (15.1) and $\boldsymbol{f}(\boldsymbol{x}) \neq \boldsymbol{0}$ for all $\boldsymbol{x} \in D \backslash \{\boldsymbol{\xi}\}$.
- Iterate.

Unfortunately, there is no general strategy to treat the first two points, as these usually require analytical methods or additional knowledge about the problem at hand.

Before we begin, we must give a word of caution regarding iterations. Much as in Chapter 6, starting from some $\boldsymbol{x}_0 \in \mathbb{R}^n$, we will construct sequences $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ which, hopefully, converge $\boldsymbol{x}_k \to \boldsymbol{\xi}$ as fast as possible. We will, as usual, stop the iteration when a prescribed tolerance $\varepsilon > 0$ is reached, i.e.,

$$\|\boldsymbol{x}_k - \boldsymbol{\xi}\| < \varepsilon.$$

How do we know when to stop the iterations? One might be tempted to say that, since $\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$, we can stop them whenever

$$\|\boldsymbol{f}(\boldsymbol{x}_k)\| < C\varepsilon$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

for some suitable constant C. The following two examples, however, show that this is not always a viable strategy.

Example 15.1 The function $f(x) = e^x$ does not have a root in $\mathbb{R}$. However, for any $\varepsilon > 0$, there is x_ε such that

$$0 < f(x_\varepsilon) = e^{x_\varepsilon} < \varepsilon.$$

Example 15.2 The previous example may be misleading in the sense that the problem did not have a solution, and we were considering a function defined on an unbounded interval. This can be easily fixed. Let $\delta \in (0, 1)$ and consider

$$f_\delta(x) = \begin{cases} \delta, & x \in [0, \frac{1}{2}), \\ 8(1-\delta)(x - \frac{1}{2}) + \delta, & x \in [\frac{1}{2}, \frac{5}{8}), \\ -\frac{16}{3}(x - \frac{5}{8}) + 1, & x \in [\frac{5}{8}, 1]. \end{cases}$$

This function is continuous and it has a unique root $\xi = \frac{13}{16} > \frac{1}{2}$. However, no matter what $\varepsilon > 0$ is, we can choose $\delta < \varepsilon$, so that any point $x \in [0, \frac{1}{2}]$ satisfies $0 < f_\delta(x) < \varepsilon$.

Let us now quickly and not very rigorously discuss the sensitivity of this problem to perturbations. To do so, we will assume that $m = n = 1$, and the function f is $k+1$ times continuously differentiable in a neighborhood of its unique root $\xi \in \mathbb{R}$. Moreover, we will assume that $f^{(p)}(\xi) = 0$ for $p < k$, but $f^{(k)}(\xi) \neq 0$. Then, given a perturbation δx, we let η be such that

$$f(\xi + \delta x) = \eta.$$

Taylor's Theorem then shows that

$$\begin{aligned} \eta &= f(\xi + \delta x) \\ &= f(\xi) + f'(\xi)\delta x + \frac{1}{2} f''(\xi)\delta x^2 + \cdots + \frac{1}{k!} f^{(k)}(\xi)\delta x^k + \mathcal{O}(|\delta x|^{k+1}) \\ &= \frac{1}{k!} f^{(k)}(\xi)\delta x^k + \mathcal{O}(|\delta x|^{k+1}). \end{aligned}$$

In other words, at least intuitively, the allowed relative size of the perturbation δx, to obtain an output of size η, is

$$\left| \frac{\delta x}{\eta} \right| \approx \left| \eta^{1-k} \frac{k!}{f^{(k)}(\xi)} \right|^{1/k}.$$

From this, we learn two things: the smaller the value of the first nonzero derivative, the larger δx can be; and, the higher the order of the first nonzero derivative, the larger δx can be. For this reason, of importance to us will be so-called *simple roots*.

Definition 15.1 (simple root)**.** Let $f : \mathbb{R} \to \mathbb{R}$ have a root $\xi \in \mathbb{R}$, and assume that f is differentiable at ξ. We say that ξ is a **simple root** if $f'(\xi) \neq 0$. If this is not the case, we say that the root is **nonsimple**.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

In this chapter, we will first present simple methods to tackle the case $m = n = 1$. As a rule, their convergence will be no better than linear (see Appendix B for definitions). We will then move on to Newton's method, for which we will show quadratic convergence. This method, and some of its variants, will be first presented in the one-dimensional case, and then for the multidimensional case, i.e., $m = n = d > 1$.

As always, we are barely scratching the surface of the subject. We refer the reader, for instance, to [28, 51, 68] for many more details.

15.1 Methods of Bisection and False Position

Here, we consider the case $m = n = 1$ and $f \in C([a, b])$ for some $-\infty < a < b < \infty$. The following result is a simple consequence of the Intermediate Value Theorem B.27.

Corollary 15.2 (existence)**.** *Let $-\infty < a < b < \infty$ and $f \in C([a, b])$. If $f(a)f(b) < 0$, then f has a (not necessarily unique) root in $\xi \in (a, b)$.*

Proof. From the condition $f(a)f(b) < 0$ we infer that

$$\inf_{x\in[a,b]} f(x) \le \min\{f(a), f(b)\} < 0 \quad \text{and} \quad \sup_{x\in[a,b]} f(x) \ge \max\{f(a), f(b)\} > 0.$$

The Intermediate Value Theorem B.27 then implies the result. □

The idea of the *bisection method*, as its name suggests, is that we will successively subdivide the interval $[a, b]$ into two subintervals of equal length and check on which of the subintervals there must be a root.

Definition 15.3 (bisection method)**.** Let $-\infty < a < b < \infty$ and $f \in C([a, b])$. Assume that $f(a)f(b) < 0$. The **bisection method** is an algorithm for generating an approximation sequence, $\{x_k\}_{k=0}^{\infty} \subset [a, b]$, via the following recursive procedure. Define $a_0 = a$, $b_0 = b$. For $k \ge 0$, set

$$x_k = \frac{1}{2}(a_k + b_k).$$

If $f(x_k) = 0$, the algorithm terminates. Otherwise,

$$(a_{k+1}, b_{k+1}) = \begin{cases} (x_k, b_k), & \text{if } f(x_k)f(b_k) < 0, \\ (a_k, x_k), & \text{if } f(x_k)f(b_k) > 0. \end{cases}$$

The convergence properties of this method are stated in the following result.

Theorem 15.4 (convergence)**.** *Let $-\infty < a < b < \infty$ and $f \in C([a, b])$ be such that $f(a)f(b) < 0$. Then the sequence $\{x_k\}_{k=0}^{\infty}$ generated by the bisection method converges to a point $\xi \in [a, b]$ such that $f(\xi) = 0$. Moreover, this method converges linearly, with the following rate of convergence:*

$$|x_k - \xi| \le \frac{1}{2^{k+1}}(b - a).$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Proof. The bisection method generates the sequences $\{a_k\}_{k=0}^{\infty}$, $\{b_k\}_{k=0}^{\infty}$, $\{x_k\}_{k=0}^{\infty}$, which satisfy

$$x_k \in (a_k, b_k), \quad [a_{k+1}, b_{k+1}] \subsetneq [a_k, b_k], \quad b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k).$$

Observe that $\{a_k\}_{k=0}^{\infty}$ is a bounded increasing sequence, and $\{b_k\}_{k=0}^{\infty}$ is a bounded decreasing sequence. By the Monotone Convergence Theorem (Theorem B.7), there exist limit points $\xi_a, \xi_b \in [a, b]$ such that

$$a_n \uparrow \xi_a \leq \xi_b \downarrow b_n.$$

But

$$b_n - a_n = \frac{1}{2^n}(b_0 - a_0) \downarrow 0,$$

which implies that $\xi_a = \xi_b$. Let us call the common point ξ. By the Squeeze Theorem B.6, since $x_k \in (a_k, b_k)$, we have also that $x_k \to \xi$. Moreover, since x_k is the midpoint of each interval, we observe that

$$|x_k - \xi| \leq \frac{1}{2}(b_k - a_k) = \cdots = \frac{1}{2^{k+1}}(b_0 - a_0) = \frac{1}{2^{k+1}}(b - a).$$

It remains then to show that $f(\xi) = 0$, and this follows from the continuity of f, since

$$0 \leq f(\xi)^2 = \lim_{k\to\infty} f(a_k) \lim_{k\to\infty} f(b_k) = \lim_{k\to\infty} f(a_k) f(b_k) \leq 0. \qquad \square$$

The bisection method requires very little from the function at hand, just continuity and that it takes values of different signs on the given interval. In this setting, it is always guaranteed to converge. The convergence, however, is only linear. Let us present a small modification, known as the *false position* method, which may improve the rate of convergence. The idea is simple. There is no reason why, in the bisection method, the approximation of the root must be the midpoint. Instead, suppose the update is chosen to be the zero of the line that connects $(a_k, f(a_k))$ and $(b_k, f(b_k))$.

Definition 15.5 (false position)**.** Let $-\infty < a < b < \infty$ and $f \in C([a, b])$. The sequence $\{x_k\}_{k=0}^{\infty} \subset [a, b]$ obtained by the following procedure defines the **false position method**. Define $a_0 = a$, $b_0 = b$. For $k \geq 0$, set

$$x_k = \frac{a_k f(b_k) - b_k f(a_k)}{f(b_k) - f(a_k)}.$$

If $f(x_k) = 0$, the algorithm terminates. Otherwise,

$$(a_{k+1}, b_{k+1}) = \begin{cases} (x_k, b_k), & \text{if } f(x_k) f(b_k) < 0, \\ (a_k, x_k), & \text{if } f(x_k) f(b_k) > 0. \end{cases}$$

The convergence of this method easily follows from that of the bisection method.

Theorem 15.6 (convergence)**.** *Let $-\infty < a < b < \infty$ and $f \in C([a, b])$ be such that $f(a) f(b) < 0$. Then the sequence $\{x_k\}_{k=0}^{\infty}$ generated by the false position method converges to a point $\xi \in [a, b]$ such that $f(\xi) = 0$.*

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Proof. The details of the proof are left to the reader as an exercise; see Problem 15.2. Here, we merely sketch why $x_k \in [a_k, b_k]$. Consider,

$$x_k - a_k = -f(a_k)\frac{b_k - a_k}{f(b_k) - f(a_k)}.$$

Now if $f(a_k) > 0$, then the numerator of this expression is negative. If this is the case, by construction, $f(b_k) < 0$ and the denominator of the expression above is negative. Thus, $x_k - a_k > 0$. The remaining cases are treated similarly.

From this, it follows that we are constructing, again, a sequence $\{[a_k, b_k]\}_{k=0}^{\infty}$ of nested intervals whose lengths are strictly decreasing. □

We will not provide an analysis of the false position method. We will just mention that, in general, it is possible (Problem 15.3) for one of the endpoints of the interval to "*get stuck*," i.e., we can have, for all $k \geq 0$,

$$a_k = a_0 = a \qquad \text{or} \qquad b_k = b_0 = b.$$

Because of this, the rate of convergence of this method is no better than linear. In practice, however, this method seems to perform better than bisection. A partial explanation for this fact will be given in the analysis of the secant method provided in Section 15.4.4.

15.2 Fixed Points and Contraction Mappings

In this section, we will relate root finding to contraction mappings and fixed point iteration schemes. The reader should examine Appendix C and Theorem C.4 for the more general setting.

Definition 15.7 (fixed point iteration)**.** Suppose that $-\infty < a < b < \infty$, $g \in C([a, b])$, and $g(x) \in [a, b]$ for all $x \in [a, b]$. In other words, $g([a, b]) \subseteq [a, b]$. Given $x_0 \in [a, b]$, the algorithm for constructing the recursive sequence $\{x_k\}_{k=0}^{\infty}$ via

$$x_{k+1} = g(x_k), \quad k \geq 0 \tag{15.2}$$

is called a **simple iteration scheme** or, sometimes, a **fixed point iteration scheme**.

The following fact follows easily from continuity.

Proposition 15.8 (fixed point)**.** *Suppose that $-\infty < a < b < \infty$, $g \in C([a, b])$, and $g([a, b]) \subseteq [a, b]$. Assume that the sequence $\{x_k\}_{k=0}^{\infty}$ obtained by a simple iteration scheme converges to a limit $\xi \in [a, b]$. Then ξ is a* fixed point *of g, i.e., $g(\xi) = \xi$.*

Proof. Indeed, by continuity,

$$g(\xi) = g\left(\lim_{k\to\infty} x_k\right) = \lim_{k\to\infty} g(x_k) = \lim_{k\to\infty} x_{k+1} = \xi. \qquad \square$$

The following result provides sufficient conditions for the existence of a fixed point.

Theorem 15.9 (existence)**.** *Suppose that $-\infty < a < b < \infty$, $g \in C([a, b])$, and $g([a, b]) \subseteq [a, b]$. Then there exists at least one fixed point $\xi \in [a, b]$ of g.*

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Proof. Define

$$f(x) = x - g(x), \quad \forall x \in [a, b].$$

Then, since $g([a, b]) \subseteq [a, b]$,

$$f(b) = b - g(b) \geq 0 \quad \text{and} \quad f(a) = a - g(a) \leq 0.$$

By the Intermediate Value Theorem B.27, since $f(a) \leq 0 \leq f(b)$, there is a point $\xi \in [a, b]$ such that $f(\xi) = 0$. For this point, $\xi = g(\xi)$. □

In practice, it is very difficult to verify the condition that $g([a, b]) \subseteq [a, b]$. The following definition provides a sufficiently large class of functions for which this condition is almost automatically satisfied.

Definition 15.10 (contraction)**.** Suppose that $-\infty < a < b < \infty$ and $g \in C([a, b])$. We say that g is **Lipschitz continuous**[1] on $[a, b]$ if and only if there exists a constant $L > 0$ such that

$$|g(x) - g(y)| \leq L|x - y|, \quad \forall x, y \in [a, b],$$

and the associated L is called the **Lipschitz constant**. The function g is called a **contraction** on $[a, b]$ if and only if it is Lipschitz on $[a, b]$ and its associated Lipschitz constant, L, satisfies $L \in (0, 1)$.

Proposition 15.11 (translation)**.** *Let $-\infty < a < b < \infty$ and $g \in C([a, b])$ be a contraction on $[a, b]$. There is a constant $m \in \mathbb{R}$ such that the function $\tilde{g}\colon [a, b] \to \mathbb{R}$, where*

$$\tilde{g}(x) = g(x) + m, \quad \forall x \in [a, b]$$

is a contraction on $[a, b]$; moreover, $\tilde{g}([a, b]) \subseteq [a, b]$.

Proof. See Problem 15.5. □

We see also that, although Theorem 15.9 provides existence of fixed points, it says nothing about uniqueness. It turns out that, for contractions, fixed points must be unique.

Theorem 15.12 (uniqueness)**.** *Suppose that $-\infty < a < b < \infty$, $g \in C([a, b])$, and $g([a, b]) \subseteq [a, b]$. If g is a contraction on $[a, b]$, then g has a unique fixed point $\xi \in [a, b]$. Furthermore, the sequence $\{x_k\}_{k=0}^{\infty}$ generated by* (15.2) *converges to ξ for any starting value $x_0 \in [a, b]$.*

Proof. Theorem 15.9 guarantees the existence of at least one fixed point $\xi \in [a, b]$. Suppose that $\eta \in [a, b]$ is another fixed point. Since g is a contraction,

$$|\xi - \eta| = |g(\xi) - g(\eta)| \leq L|\xi - \eta|.$$

Therefore,

$$0 \leq (1 - L)|\xi - \eta| \leq 0,$$

which proves that $\xi = \eta$. Hence, the fixed point $\xi \in [a, b]$ is unique.

[1] Named in honor of the German mathematician Rudolf Otto Sigismund Lipschitz (1832–1903).

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Now suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (15.2) for $x_0 \in [a, b]$. Then

$$|\xi - x_k| = |g(\xi) - g(x_{k-1})| \leq L|\xi - x_{k-1}|.$$

By induction, for any $k \in \mathbb{N}$,

$$|\xi - x_k| \leq L^k |\xi - x_0|.$$

By the Squeeze Theorem B.6, since $L \in (0, 1)$, we must have $x_k \to \xi$. □

Theorem 15.13 (local [at least linear] convergence)**.** *Suppose that $-\infty < a < b < \infty$, $g \in C([a, b])$, and $g([a, b]) \subseteq [a, b]$. Let $\xi \in [a, b]$ be a fixed point of g, i.e., $\xi = g(\xi)$. Suppose that there is a constant $\delta > 0$ such that $I_\delta = (\xi - \delta, \xi + \delta) \subset [a, b]$, $g \in C^1(I_\delta)$, and $|g'(\xi)| < 1$. Then the sequence $\{x_k\}_{k=0}^{\infty}$ generated by* (15.2) *converges at least linearly to ξ, provided that x_0 is sufficiently close to ξ.*

Proof. Suppose that $\xi \in (a, b)$, i.e., ξ is in the interior of the set $[a, b]$. The reader can examine the other cases. Since $|g'(\xi)| < 1$, there is an $h \in (0, \delta)$ and an $L \in (0, 1)$ such that, for all $x \in I_h = (\xi - h, \xi + h)$,

$$|g'(x)| \leq L < 1.$$

The proof of this last fact is left to the reader as an exercise; see Problem 15.6.

Suppose that $x_k \in I_h$. Then, using the Mean Value Theorem B.30,

$$|\xi - x_{k+1}| = |g(\xi) - g(x_k)| = |g'(\eta_k)| \cdot |\xi - x_k| \leq L|\xi - x_k|$$

for some $\eta_k \in I_h$ between ξ and x_k. This proves that if $x_k \in I_h$, then $x_{k+1} \in I_h$. Using induction, if $x_0 \in I_h$,

$$|\xi - x_k| \leq L^k |\xi - x_0|.$$

By the Squeeze Theorem, since $L^k \to 0$ as $k \to \infty$, we have $x_k \to \xi$ as $k \to \infty$. Furthermore, the convergence is at least linear; see Definition B.10. □

We will now consider how simple fixed point iterations can be used to find roots. A general strategy is to find some function α that does not vanish and to consider a fixed point iteration scheme for

$$g(x) = x - \alpha(x) f(x).$$

The particular choice of α gives rise to the various methods we now consider.

15.2.1 Relaxation Method

Definition 15.14 (relaxation)**.** Let $I \subseteq \mathbb{R}$ be an interval, $f \in C(I)$, and $x_0 \in I$ be given. The **relaxation method** is an algorithm for computing the terms of the sequence $\{x_k\}_{k=0}^{\infty}$ via the recursive formula

$$x_{k+1} = x_k - \lambda f(x_k), \tag{15.3}$$

where $\lambda \neq 0$. The method is **well defined** if and only if $x_k \in I$ for all $k = 1, 2, 3, \ldots$, and the relaxation method **converges** if and only if there is a $\xi \in I$, with $f(\xi) = 0$, such that $x_k \to \xi$ as $k \to \infty$.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Notice that the relaxation method is a simple fixed point iteration scheme with $g(x) = x - \lambda f(x)$. From this definition, it necessarily follows that a fixed point of g must be a root of f. Thus, the following result is just a translation of the theory of fixed point iterations.

Theorem 15.15 (convergence)**.** *Let $I \subset \mathbb{R}$ be an interval. Suppose that $f : I \to \mathbb{R}$ and, for some $\xi \in I$, $f(\xi) = 0$, but $f'(\xi) \neq 0$. Assume that, for some $\delta > 0$, $f \in C^1(I_\delta)$, where $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Then there exists positive real numbers λ and $h \in (0, \delta)$ such that the sequence $\{x_k\}_{k=0}^{\infty}$ defined by the relaxation scheme* (15.3) *converges to ξ for any $x_0 \in I_h = [\xi - h, \xi + h]$.*

Proof. Suppose that $f'(\xi) = \alpha > 0$. The case $\alpha < 0$ is analogous and left to the reader. By continuity, we may assume that

$$0 < \frac{1}{2}\alpha \leq f'(x) \leq \frac{3}{2}\alpha, \quad \forall x \in I_\delta.$$

If this is not the case, we can just choose a smaller δ and redefine I_δ. Set

$$M = \max_{x \in I_\delta} f'(x).$$

Thus, $\frac{1}{2}\alpha \leq M \leq \frac{3}{2}\alpha$. For any $\lambda > 0$, it follows that

$$1 - \lambda M \leq 1 - \lambda f'(x) \leq 1 - \frac{1}{2}\lambda\alpha, \quad \forall x \in I_\delta.$$

We now choose, if possible, $\lambda > 0$ such that

$$1 - \lambda M = -\theta \quad \text{and} \quad 1 - \frac{1}{2}\lambda\alpha = \theta.$$

These equations are satisfied if and only if

$$\lambda M - 1 = 1 - \frac{1}{2}\lambda\alpha, \quad \theta = 1 - \frac{1}{2}\lambda\alpha$$

if and only if

$$\lambda = \frac{4}{2M + \alpha}, \quad \theta = \frac{2M - \alpha}{2M + \alpha}.$$

Now define the iteration function g via

$$g(x) = x - \lambda f(x) = x - \frac{4f(x)}{2M + \alpha}.$$

The rest of the details are left to the reader as an exercise; see Problem 15.7. Use Theorem 15.13 to conclude that $x_k \to \xi$, provided that x_0 is sufficiently close to ξ. □

15.2.2 Stationary Slope Approximation Methods

Let $I \subset \mathbb{R}$ be an interval and $f \in C^1(I)$. Assume that there is $\xi \in I$, which is a simple root of f. Then, by Taylor's Theorem B.31, we obtain that, for any $x \in I$,

$$0 = f(\xi) = f(x) + f'(\theta)(\xi - x)$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

for some θ between x and ξ. Thus, if $f'(\theta) \neq 0$, the root must satisfy

$$\xi = x - [f'(\theta)]^{-1} f(x).$$

This motivates the construction of a family of schemes. Let $s_0 \neq 0$ be a *slope approximation* and $x_0 \in I$ an initial guess. Then, for $k \geq 0$,

$$x_{k+1} = x_k - s_k^{-1} f(x_k), \tag{15.4}$$

with some rule to compute $s_{k+1} \neq 0$. Notice that if $s_k = s_0 = \frac{1}{\lambda}$, then this reduces to the relaxation method of Definition 15.14. Let us consider two particular examples.

Definition 15.16 (chord method)**.** Let $I = [a, b]$ be an interval and $x_0 \in I$. The sequence $\{x_k\}_{k=0}^{\infty}$ is obtained by the **chord method** if is constructed via (15.4) with

$$s_k = \frac{f(b) - f(a)}{b - a}.$$

Theorem 15.17 (convergence)**.** *Let $I = [a, b] \subset \mathbb{R}$ be an interval and $f \in C^1([a, b])$ is such that it has a unique simple root $\xi \in [a, b]$. If $b - a$ is sufficiently small, then the sequence $\{x_k\}_{k=1}^{\infty}$ obtained by the chord method of Definition 15.16 converges linearly to ξ as $k \to \infty$.*

Proof. See Problem 15.8. □

Definition 15.18 (simplified Newton)**.** Let I be an interval and $x_0 \in I$ be such that $f'(x_0) \neq 0$. The sequence $\{x_k\}_{k=0}^{\infty}$ is obtained by the **simplified Newton method**[2] if it is constructed via (15.4) with

$$s_k = f'(x_0).$$

This method can be analyzed in two different ways. Here, we analyze it using the theory of fixed point iterations.

Proposition 15.19 (convergence)**.** *Let $I \subset \mathbb{R}$ be an interval and $f \in C(I)$ be such that there is $\xi \in I$ for which $f(\xi) = 0$. Define, for $\delta > 0$, $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Assume that there is $\delta > 0$ such that $f \in C^1(I_\delta)$. Finally, assume that $f'(\xi) = \alpha > 0$. If x_0 is sufficiently close to ξ, then the simplified Newton method of Definition 15.18 converges linearly to ξ.*

Proof. See Problem 15.9. □

15.2.3 Fixed Point Iterations in Several Dimensions

Let us quickly comment that the idea of fixed point iterations can be generalized to several dimensions. For this, we consider a closed ball $B \subset \mathbb{R}^d$ and $\boldsymbol{g} \colon B \to \mathbb{R}^d$. The *fixed point iteration scheme*, in this setting, starts from $\boldsymbol{x}_0 \in B$ and proceeds, for $k \geq 0$, as

$$\boldsymbol{x}_{k+1} = \boldsymbol{g}(\boldsymbol{x}_k).$$

[2] Named in honor of the English mathematician, physicist, astronomer, theologian, and natural philosopher Sir Isaac Newton (1643–1726/27).

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

As in the one-dimensional case, we say that the fixed point iteration scheme is *well defined* if $\boldsymbol{x}_k \in B$ for all k. The convergence theory of this approach and its application to the solution of systems of nonlinear equations follow the same ideas we have presented here. We leave the details to the reader.

15.3 Newton's Method in One Space Dimension

We have now arrived at our preferred method of choice: Newton's method. Throughout our discussion, we will assume that the function f, for which we are trying to find a root, is at least twice continuously differentiable. One can prove these results using less regularity than this, but this assumption will greatly simplify our arguments.

Definition 15.20 (Newton's method[3])**.** Let $I \subseteq \mathbb{R}$ be an interval, $f \in C^1(I)$, and $x_0 \in I$, with $f'(x_0) \neq 0$, be given. **Newton's method** is an algorithm for computing the terms of the sequence $\{x_k\}_{k=0}^{\infty}$ via the recursive formula

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \tag{15.5}$$

We say that the method is **well defined** if and only if $x_k \in I$ and $f'(x_k) \neq 0$ for all $k = 1, 2, 3, \ldots$. We say that Newton's method **converges** if and only if there is a $\xi \in I$, with $f(\xi) = 0$, such that $x_k \to \xi$ as $k \to \infty$.

Newton's method can be studied with the theory of fixed point iterations presented in Section 15.2. Indeed, from the definition, it is clear that Newton's method is a fixed point iteration for the function

$$g(x) = x - \frac{f(x)}{f'(x)}. \tag{15.6}$$

Thus, for a simple root, one immediately obtains a linear convergence for Newton's method provided the initial guess, x_0, is sufficiently close to the root.

Proposition 15.21 (linear convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval and $f \in C^2(I)$ with $|f'(x)| \geq \alpha > 0$ for all $x \in I$. Assume that there is $\xi \in I$ for which $f(\xi) = 0$. There is a constant $h > 0$ such that, if $x_0 \in I$ and $|x_0 - \xi| < h$, Newton's method converges at least linearly to ξ.*

Proof. For this proof, let us apply the theory of fixed points, as detailed in Section 15.2, to the function g defined in (15.6). Notice that since $f \in C^2(I)$, then

$$g'(x) = \frac{f(x)f''(x)}{[f'(x)]^2} \quad \Longrightarrow \quad g'(\xi) = 0.$$

Thus, by continuity, there is a constant $\delta > 0$ such that if $x \in I_\delta = [\xi - \delta, \xi + \delta]$, then

$$|g'(x)| < 1.$$

[3] Named in honor of the British mathematician, physicist, astronomer, theologian, and natural philosopher Sir Isaac Newton (1642–1726/27).

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

In addition, notice that, by Taylor's Theorem B.32, we have, for any $x \in I$,

$$0 = f(\xi) = f(x) + f'(x)(\xi - x) + \frac{1}{2} f''(\eta)(\xi - x)^2$$

for some η between x and ξ. Let us define

$$A = \frac{\max_{x \in I} |f''(x)|}{\alpha}, \quad h = \min\left\{\delta\,,\ \frac{1}{A}\right\}.$$

Then, for any $x \in I_h = [\xi - h, \xi + h]$, we have $|g'(x)| < 1$ and

$$\begin{aligned} |g(x) - \xi| &= \left| x - \xi - \frac{f(x)}{f'(x)} \right| \\ &= \frac{1}{2} \left| \frac{f''(\eta)}{f'(x)} (x - \xi)^2 \right| \\ &\leq \frac{1}{2} A h^2 \\ &\leq \frac{1}{2} h, \end{aligned}$$

so that $g(x) \in I_h$. Theorem 15.13 then allows us to conclude the (at least linear) convergence of the fixed point iteration sequence defined in (15.2). □

It turns out that, under normal, reasonable circumstances, Newton's method converges quadratically.

Theorem 15.22 (quadratic convergence)**.** *Let $I \subset \mathbb{R}$ be an interval. Suppose that $f : I \to \mathbb{R}$ and, for some $\xi \in I$, $f(\xi) = 0$, but $f'(\xi) \neq 0$ and $f''(\xi) \neq 0$. Assume that, for some $\delta > 0$, $f \in C^2(I_\delta)$, where $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$, and $0 < \alpha \leq |f'(x)|$ for all $x \in I_\delta$. Set*

$$A = \frac{\max_{x \in I_\delta} |f''(x)|}{\alpha}\,, \quad h = \min\left\{\delta\,,\ \frac{1}{A}\right\}. \tag{15.7}$$

If $|\xi - x_0| \leq h$, then the sequence $\{x_k\}_{k=0}^{\infty}$ defined by Newton's method (15.5) *converges quadratically, as $k \to \infty$, to the root ξ.*

Proof. We could use the result of Proposition 15.21 as our starting point for this proof. To give some variety, let us repeat the proof of linear convergence, but this time using a direct approach.

(Well-posedness) Suppose that $x_k \in I_\delta$. Then, by Taylor's Theorem B.32,

$$0 = f(\xi) = f(x_k) + (\xi - x_k) f'(x_k) + \frac{(\xi - x_k)^2}{2} f''(\eta_k)$$

for some η_k between x_k and ξ. Note that $f'(x_k) \neq 0$, and we have, using (15.5) (the definition of Newton's method),

$$x_{k+1} - \xi = \frac{(\xi - x_k)^2 f''(\eta_k)}{2 f'(x_k)}. \tag{15.8}$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Now, if $x_k \in I_h = [\xi - h, \xi + h]$,

$$|\xi - x_{k+1}| = \frac{1}{2}\frac{|f''(\eta_k)|}{|f'(x_k)|} \cdot |\xi - x_k| \cdot |\xi - x_k| \leq \frac{A}{2} \cdot h \cdot |\xi - x_k| = \frac{1}{2}|\xi - x_k|,$$

and $x_{k+1} \in I_h$ as well. The algorithm is, therefore, well defined, since $x_{k+1} \in I_h$ and $f'(x_{k+1}) \neq 0$.

(Linear convergence) By induction, it is clear from the contraction estimate that

$$|\xi - x_k| \leq \frac{h}{2^k},$$

which proves that the sequence converges to the root ξ as $k \to \infty$, at least linearly; see Definition B.10.

(Quadratic convergence) From (15.8) we obtain

$$\frac{|\xi - x_{k+1}|}{|\xi - x_k|^2} = \frac{1}{2}\frac{|f''(\eta_k)|}{|f'(x_k)|}.$$

Since $x_k \to \xi$, by the Squeeze Theorem B.6, $\eta_k \to \xi$ as $k \to \infty$ as well. Passing to limits, we have

$$\lim_{k\to\infty} \frac{|\xi - x_{k+1}|}{|\xi - x_k|^2} = \frac{1}{2}\frac{|f''(\xi)|}{|f'(\xi)|} = \sigma \in (0, \infty),$$

which establishes quadratic convergence; see Definition B.10. □

Remark 15.23 (faster convergence)**.** We see immediately that it is possible for the convergence to be faster than quadratic if and only if $f''(\xi) = 0$.

One glaring fact about the last result for Newton's method is that it guarantees convergence only in a local region. To improve this to a more global result, we need additional assumptions.

Theorem 15.24 (global convergence)**.** *Let $[a, b] \subset \mathbb{R}$ be an interval and $f \in C^2([a, b])$ be such that, for some $\xi \in [a, b]$, $f(\xi) = 0$. Assume further that f' and f'' are strictly positive on the interval $[a, b]$. For any starting value $x_0 \in (\xi, b]$, the sequence $\{x_k\}_{k=0}^{\infty}$ defined by Newton's method* (15.5) *converges quadratically to the root ξ as $k \to \infty$. Moreover, $x_k > \xi$ for all $k \in \mathbb{N}$.*

Proof. Since f is monotonically increasing on $[a, b]$, ξ is the only root in $[a, b]$. Otherwise, by Rolle's Theorem B.29, one could find a point where f' is zero, contradicting the assumptions. Also note that $f(x) > 0$ for all $x \in (\xi, b]$; likewise, $f(x) < 0$ for all $x \in [a, \xi)$.

Assume that $x_k > \xi$. Employing Newton's method (15.5) and using the positivity of $f(x_k)$ and $f'(x_k)$, we immediately obtain that

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} < x_k.$$

Furthermore, from the error equation (15.8), using the positivity of $f''(\eta_k)$ and $f'(x_k)$, we find out that $\xi - x_{k+1} < 0$. In other words, $\xi < x_{k+1} < x_k$. Thus, $\{x_k\}_{k=0}^{\infty}$ is a bounded, monotonically decreasing sequence in $[a, b]$. By Theorem B.7, it must,

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

therefore, have a limit point in $[a, b]$, call it η, as $k \to \infty$. But this limit must be a fixed point of the function g, defined in (15.6). Therefore, $f(\eta) = 0$. But since ξ is the unique root of f in $[a, b]$, it must be that $\xi = \eta$. This shows that $x_k \to \xi$ as $k \to \infty$. Quadratic convergence follows as in the proof of Theorem 15.22. □

Remark 15.25 (other initial guesses)**.** If in the setting of the last theorem one assumes that $x_0 < \xi$ to begin, one encounters some problems. One will conclude from (15.5) that $x_1 > x_0$. But from (15.8) one will conclude that $x_1 > \xi$. Thus, there is no "squeezing" action as before. However, one might recover if, by chance, $x_1 \leq b$. For then we could restart the argument in the last proof to guarantee convergence. Of course, if the interval in question is $(-\infty, +\infty)$ there is no problem at all.

Example 15.3 Let us use Newton's method to compute the square root of a positive real number. Suppose that we want to compute $\sqrt{5}$. Define $f(x) = x^2 - 5$. There are two solutions to $f(x) = 0$, namely $\xi_\pm = \pm\sqrt{5}$. Let us pick $x_0 = 5$. Theorem 15.24 guarantees that this is a suitable choice if we want to compute the zero $\xi_+ = \sqrt{5}$. The sequence of approximations for Newton's method is defined by

$$x_{k+1} = x_k - \frac{x_k^2 - 5}{2x_k}, \quad k = 0, 1, \ldots. \tag{15.9}$$

Below, we show the result of using the code presented in Listing 15.1. The correct digits are indicated using boldface.

k	x_k
0	5.000 000 000 000 000
1	3.000 000 000 000 000
2	**2**.333 333 333 333 333
3	**2.23**8 095 238 095 238
4	**2.236 06**8 895 643 363
5	**2.236 067 977 499** 978
6	**2.236 067 977 499 790**

This example illustrates an empirical fact that is a consequence of quadratic convergence: Newton's method doubles the number of correct digits with each iteration. A partial explanation for this fact is as follows: from the proof of Theorem 15.22, we see that

$$|\xi - x_{k+1}| \leq C|\xi - x_k|^2,$$

so that, by taking base-10 logarithms (which essentially counts the number of correct digits), we have

$$\log_{10} |\xi - x_{k+1}| \leq 2 \log_{10} |\xi - x_k| + \log_{10} C.$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Example 15.4 Define $f\colon [1,5] \to \mathbb{R}$ via $f(x) = (x-3)^3$. Observe that, for this simple example, $\xi = 3$ is a nonsimple root, i.e., $f(3) = f'(3) = 0$, which is something that the theory (up to this point) cannot handle. Nevertheless, Newton's method will still work. In particular, if Newton's method is employed with the starting point $x_0 = 4$ to approximate the root $\xi = 3$, then one can show directly that the convergence is exactly linear; see Problem 15.16.

15.3.1 Nonsimple roots

The next result describes what may happen when a certain number of derivatives vanish at the zero of interest, as in the Example 15.4.

Theorem 15.26 (nonsimple roots)**.** *Let m be a positive integer and I be a closed and bounded interval. Suppose that $f \in C^m(I)$ is such that there is $\xi \in I$, for which $f(\xi) = f'(\xi) = \cdots = f^{(m-1)}(\xi) = 0$, but $f^{(m)}(\xi) \neq 0$. If $|\xi - x_0|$ is sufficiently small, the sequence $\{x_k\}_{k=0}^{\infty}$ defined by Newton's method* (15.5) *is well defined and converges to ξ exactly linearly with*

$$\lim_{k\to\infty} \frac{|x_{k+1}-\xi|}{|x_k-\xi|} = \frac{m-1}{m} = \sigma \in (0,1).$$

Proof. We give a sketch of the proof. The details of well-definedness and convergence, in particular, are left to the reader as an exercise; see Problems 15.17 and 15.18. By Taylor's Theorem,

$$\begin{aligned} f(x_k) &= f(\xi) + f'(\xi)(x_k-\xi) + \cdots + f^{(m-1)}(\xi)\frac{(x_k-\xi)^{m-1}}{(m-1)!} + f^{(m)}(\eta_k)\frac{(x_k-\xi)^m}{m!} \\ &= f^{(m)}(\eta_k)\frac{(x_k-\xi)^m}{m!} \end{aligned}$$

for some η_k between x_k and ξ. Another application of Taylor's Theorem gives

$$\begin{aligned} f'(x_k) &= f'(\xi) + f''(\xi)(x_k-\xi) + \cdots + f^{(m-1)}(\xi)\frac{(x_k-\xi)^{m-2}}{(m-2)!} \\ &\quad + f^{(m)}(\zeta_k)\frac{(x_k-\xi)^{m-1}}{(m-1)!} = f^{(m)}(\zeta_k)\frac{(x_k-\xi)^{m-1}}{(m-1)!} \end{aligned}$$

for some ζ_k between x_k and ξ. Thus, assuming $x_k \neq \xi$,

$$x_{k+1} - \xi = x_k - \xi - \frac{f(x_k)}{f'(x_k)} = x_k - \xi - \frac{f^{(m)}(\eta_k)\frac{(x_k-\xi)^m}{m!}}{f^{(m)}(\zeta_k)\frac{(x_k-\xi)^{m-1}}{(m-1)!}}$$

or

$$\frac{x_{k+1}-\xi}{x_k-\xi} = 1 - \frac{f^{(m)}(\eta_k)}{m\cdot f^{(m)}(\zeta_k)}.$$

Since $\eta_k, \zeta_k \to \xi$ as $k \to \infty$,

$$\lim_{k\to\infty} \frac{x_{k+1}-\xi}{x_k-\xi} = 1 - \frac{1}{m} = \frac{m-1}{m}.$$

□

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

The fact that, in Theorem 15.26, the constant σ depends only on the multiplicity m of the root hints at the fact that quadratic convergence for Newton's method can be recovered by a small modification. This is explored in Problem 15.19.

15.4 Quasi-Newton Methods

In the previous section, we developed the analysis of Newton's method of (15.5). We showed that, under suitable assumptions, this method converges quadratically. There are, however, two major drawbacks to Newton's method: it requires a sufficiently close initial approximation to the root, and it requires evaluating the derivative at every iteration. Requiring a good initial guess is mostly unavoidable, but such an approximation may be obtained by some other method. The evaluation of the derivative, on the other hand, may be an issue. In applications, this may be very costly, or not even at all possible. For this reason, here we propose several important variants.

The common feature of all these methods is that they take the form (15.4), where the slope approximation s_k will, in general, change on every iteration. Its construction will depend for instance, for $r \geq 0$, on $x_{k-r}, \ldots, x_k$, and the values of the function at these points, but not require evaluation of the derivative at these points. Another possibility is that we require fewer derivative evaluations than Newton does, i.e., not at every iteration.

15.4.1 Simplified Newton's Method

We already saw the simplified Newton method in Definition 15.18. Here, we provide an analysis for it. We recall that this method has the form (15.4) with $s_k = s = f'(x_0)$.

Theorem 15.27 (convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval. Assume that $f \in C(I)$ is such that there is an $\xi \in I$ for which $f(\xi) = 0$, but $f'(\xi) \neq 0$ and $f''(\xi) \neq 0$. Set, for $\delta > 0$, $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Assume that, for some $\delta > 0$, $f \in C^2(I_\delta)$ and $0 < \alpha \leq |f'(x)|$ for all $x \in I_\delta$. Set*

$$A = \frac{\max_{x \in I_\delta} |f''(x)|}{\alpha}, \quad h = \min\left\{\delta, \ \frac{1}{3A}\right\}. \tag{15.10}$$

If $|\xi - x_0| \leq h$, then the sequence $\{x_k\}_{k=0}^{\infty}$ defined by the simplified Newton method of Definition 15.18 converges linearly to the zero ξ of f as $k \to \infty$.

Proof. Suppose that $x_0, x_k \in [\xi - h, \xi + h]$. Then we have

$$\begin{aligned} x_{k+1} - \xi &= \frac{1}{f'(x_0)}\Big[f'(x_0)\,(x_k - \xi) - f(x_k)\Big] \\ &= \frac{1}{f'(x_0)}\Big[\big(f'(x_0) - f'(\xi)\big)\,(x_k - \xi) - f(x_k) + f'(\xi)\,(x_k - \xi)\Big]. \end{aligned}$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

By the Mean Value Theorem B.30, there is $\beta \in [\xi - h, \xi + h]$ between x_0 and ξ, for which

$$f''(\beta)(x_0 - \xi) = f'(x_0) - f'(\xi).$$

Furthermore, for some $\eta_k \in [\xi - h, \xi + h]$ between ξ and x_k, we have

$$f(x_k) = f(\xi) + f'(\xi)(x_k - \xi) + \frac{f''(\eta_k)}{2}(x_k - \xi)^2$$

from Taylor's Theorem. Rearranging terms and using $f(\xi) = 0$ yields

$$-f(x_k) + f'(\xi)(x_k - \xi) = -\frac{f''(\eta_k)}{2}(x_k - \xi)^2.$$

Putting things together, we find

$$\begin{aligned} x_{k+1} - \xi &= \frac{1}{f'(x_0)} \left[f''(\beta)(x_0 - \xi)\,(x_k - \xi) - \frac{f''(\eta_k)}{2}(x_k - \xi)^2 \right] \\ &= \frac{1}{f'(x_0)} \left[f''(\beta)(x_0 - \xi) - \frac{f''(\eta_k)}{2}(x_k - \xi) \right] (x_k - \xi). \end{aligned}$$

Taking absolute values,

$$\begin{aligned} |x_{k+1} - \xi| &= \frac{1}{|f'(x_0)|} \left| f''(\beta)(x_0 - \xi) - \frac{f''(\eta_k)}{2}(x_k - \xi) \right| \cdot |x_k - \xi| \\ &\le \frac{1}{|f'(x_0)|} \left(|f''(\beta)| \cdot |x_0 - \xi| + \frac{1}{2}|f''(\eta_k)| \cdot |x_k - \xi| \right) |x_k - \xi| \\ &\le \left(A|x_0 - \xi| + \frac{1}{2}|x_k - \xi|A \right) |x_k - \xi| \\ &\le \left(A \cdot \frac{1}{3A} + \frac{1}{2}\frac{1}{3A}A \right) |x_k - \xi| = \frac{1}{2}|x_k - \xi|. \end{aligned}$$

Hence, $x_{k+1} \in [\xi - h, \xi + h]$ for any $k \in \mathbb{N}$, as long as $x_0, x_k \in [\xi - h, \xi + h]$. The simplified Newton algorithm is well defined. Furthermore, it is clear that

$$|x_k - \xi| \le \frac{h}{2^k},$$

which proves that $x_k \to \xi$ as $k \to \infty$.

The convergence is exactly linear, as can be seen from the error equation:

$$\lim_{k \to \infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|} = \frac{|f''(\beta)| \cdot |x_0 - \xi|}{|f'(x_0)|} = \mu.$$

By our assumptions, $0 < \mu \le 1/3 < 1$. □

15.4.2 Steffensen's Method

In the simplified Newton method, only one derivative evaluation is required. The trade-off is that the order of convergence is reduced. In the following, we can eliminate derivative evaluation altogether and still retain quadratic convergence.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Definition 15.28 (Steffensen's method[4])**.** Let $I \subseteq \mathbb{R}$ be an interval, $f \in C(I)$, and $x_0 \in I$. **Steffensen's method** is an algorithm for computing the terms of the sequence $\{x_k\}_{k=0}^{\infty}$ via (15.4) with

$$s_k = \frac{f(x_k + f(x_k)) - f(x_k)}{f(x_k)}.$$

We say that this method is **well defined** if and only if $x_0 \in I$ implies that $x_k \in I$ for all $k = 1, 2, \ldots$. We say that this method **converges** if and only if there is $\xi \in I$, with $f(\xi) = 0$, such that $x_k \to \xi$ as $k \to \infty$.

Before studying the convergence of this method, let us provide some intuition behind this slope approximation. If the method is to converge, then $h = f(x_k) \to f(\xi) = 0$, so that

$$s_k = \frac{f(x_k + h) - f(x_k)}{h}$$

is indeed a good approximation of the derivative $f'(x_k)$.

Let us now study the convergence of this method.

Theorem 15.29 (convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval and $f \in C(I)$ be such that, for some $\xi \in I$, $f(\xi) = 0$. Define, for $\delta > 0$, $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Assume that there is $\delta > 0$ for which $f \in C^2(I_\delta)$, $f'(\xi) \neq 0$, and $f''(\xi) \neq 0$. If $|\xi - x_0|$ is sufficiently small, then the sequence $\{x_k\}_{k=0}^{\infty}$ defined by Steffensen's method of Definition 15.28 is well defined and converges quadratically to the zero ξ as $k \to \infty$.*

Proof. First observe that, using Taylor's Theorem B.32, there are points η_k and γ_k between x_k and $x_k + f(x_k)$ such that

$$s_k = \frac{f'(x_k)f(x_k) + \frac{1}{2}f''(\eta_k)f^2(x_k)}{f(x_k)} = f'(x_k) + \frac{1}{2}f''(\eta_k)f(x_k) = f'(\gamma_k). \quad (15.11)$$

From the definition of the scheme, we have

$$\begin{aligned} x_{k+1} - \xi &= x_k - \xi - \frac{f(x_k)}{s_k} \\ &= \frac{(x_k - \xi)\left(f'(x_k) + \frac{1}{2}f''(\eta_k)f(x_k)\right) - f(x_k)}{f'(x_k) + \frac{1}{2}f''(\eta_k)f(x_k)} \\ &= \frac{-\left[f(x_k) + f'(x_k)(\xi - x_k)\right] + \frac{1}{2}f''(\eta_k)f(x_k)(x_k - \xi)}{f'(x_k) + \frac{1}{2}f''(\eta_k)f(x_k)}. \end{aligned}$$

By Taylor's Theorem B.32, there is a point β_k between x_k and ξ such that

$$0 = f(\xi) = f(x_k) + f'(x_k)(\xi - x_k) + \frac{1}{2}f''(\beta_k)(\xi - x_k)^2,$$

so that

$$\frac{1}{2}f''(\beta_k)(\xi - x_k)^2 = -\left[f(x_k) + f'(x_k)(\xi - x_k)\right].$$

[4] Named in honor of the Danish mathematician and statistician Johan Frederik Steffensen (1873–1961).

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Hence,

$$\begin{aligned} x_{k+1}-\xi &= \frac{\frac{1}{2}f''(\beta_k)(\xi-x_k)^2+\frac{1}{2}f''(\eta_k)f(x_k)(x_k-\xi)}{f'(x_k)+\frac{1}{2}f''(\eta_k)f(x_k)} \\ &= \frac{\frac{1}{2}f''(\beta_k)(\xi-x_k)^2+\frac{1}{2}f''(\eta_k)(f(x_k)-f(\xi))(x_k-\xi)}{f'(x_k)+\frac{1}{2}f''(\eta_k)f(x_k)} \\ &= \frac{\frac{1}{2}f''(\beta_k)(\xi-x_k)^2+\frac{1}{2}f''(\eta_k)f'(\alpha_k)(x_k-\xi)^2}{f'(x_k)+\frac{1}{2}f''(\eta_k)f(x_k)} \\ &= \left[\frac{\frac{1}{2}f''(\beta_k)+\frac{1}{2}f''(\eta_k)f'(\alpha_k)}{f'(x_k)+\frac{1}{2}f''(\eta_k)f(x_k)}\right](x_k-\xi)^2, \end{aligned}$$

where α_k is some point between x_k and ξ.

Taking absolute values, using the triangle inequality, and using (15.11), we get

$$\begin{aligned} |x_{k+1}-\xi| &= \left|\frac{\frac{1}{2}f''(\beta_k)+\frac{1}{2}f''(\eta_k)f'(\alpha_k)}{f'(\gamma_k)}\right|\cdot|x_k-\xi|\cdot|x_k-\xi| \\ &\le \frac{\frac{1}{2}|f''(\beta_k)|+\frac{1}{2}|f''(\eta_k)|\cdot|f'(\alpha_k)|}{|f'(\gamma_k)|}\cdot|x_k-\xi|\cdot|x_k-\xi|. \end{aligned}$$

Notice now that the points η_k and γ_k are between x_k and $x_k+f(x_k)$. By continuity, there is an $h\in(0,\delta/2)$ such that

$$|f(x)|\le\delta/2,\quad \forall x\in I_h=[\xi-h,\xi+h].$$

Therefore, if $x_k\in I_h$, it easily follows that

$$|\xi-\eta_k|\le\delta,\qquad |\xi-\gamma_k|\le\delta.$$

We assume, as usual, that there are constants $m_2\ge m_1>0$ such that

$$m_1\le|f'(x)|\le m_2,\quad \forall x\in I_\delta,$$

and constants $m_4\ge m_3>0$ such that

$$m_3\le|f''(x)|\le m_4,\quad \forall x\in I_\delta.$$

Therefore, if $x_k\in I_s=[\xi-s,\xi+s]$, where

$$s<\min\left\{\frac{m_1}{m_4+m_4m_2},h\right\},$$

$$\begin{aligned} |x_{k+1}-\xi| &\le \frac{\frac{1}{2}|f''(\beta_k)|+\frac{1}{2}|f''(\eta_k)|\cdot|f'(\alpha_k)|}{|f'(\gamma_k)|}\cdot|x_k-\xi|\cdot|x_k-\xi| \\ &\le \frac{\frac{1}{2}m_4+\frac{1}{2}m_4m_2}{m_1}\cdot\frac{m_1}{m_4+m_4m_2}\cdot|x_k-\xi|\le\frac{1}{2}|x_k-\xi|. \end{aligned}$$

Thus, the method is well defined and it converges at least linearly.

The proof of quadratic convergence follows from the fact that

$$\lim_{k\to\infty}\frac{|x_{k+1}-\xi|}{|x_k-\xi|^2}=\frac{1}{2}\left|\frac{f''(\xi)+f''(\xi)f'(\xi)}{f'(\xi)}\right|\neq 0,$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

where we used the fact that

$$\alpha_k, \beta_k, \gamma_k, \eta_k \to \xi, \qquad k \to \infty. \qquad \square$$

15.4.3 Two-Step Newton's Method

The next method is a variant of Newton's method which exhibits convergence that may be faster than quadratic.

Definition 15.30 (two-step Newton)**.** Let $I \subseteq \mathbb{R}$ be an interval and $f \in C^1(I)$. For $x_0 \in I$, with $f'(x_0) \neq 0$, the sequence $\{x_k\}_{k=0}^{\infty}$ defined by

$$y_k = x_k - \frac{f(x_k)}{f'(x_k)}, \qquad x_{k+1} = y_k - \frac{f(y_k)}{f'(x_k)} \tag{15.12}$$

is called the **two-step Newton** method. We say that the method is **well defined** if $x_k \in I$ and $f'(x_k) \neq 0$ for all $k \geq 0$. We say that the method **converges** if there is $\xi \in I$ such that $f(\xi) = 0$ and $x_k \to \xi$ as $k \to \infty$.

Theorem 15.31 (convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval, $f \in C(I)$ is such that there is $\xi \in I$ for which $f(\xi) = 0$, but $f'(\xi) \neq 0$, and $f''(\xi) \neq 0$. Set, for $\delta > 0$, $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Assume that is $\delta > 0$ for which $f \in C^2(I_\delta)$ and $0 < \alpha \leq |f'(x)|$ for all $x \in I_\delta$. Set*

$$A = \frac{\max_{x \in I_\delta} |f''(x)|}{\alpha}, \quad h = \min\left\{\delta, \frac{1}{A}\right\}.$$

If $|\xi - x_0| \leq h$, then the sequence $\{x_k\}_{k=0}^{\infty}$ defined by the two-step Newton method converges exactly cubically to the zero ξ as $k \to \infty$.

Proof. Suppose that $x_k \in [\xi - h, \xi + h] \subseteq I_\delta$. Then, by Taylor's Theorem,

$$0 = f(\xi) = f(x_k) + f'(x_k)(\xi - x_k) + \frac{f''(\eta_k)}{2}(\xi - x_k)^2$$

for some η_k between x_k and ξ. Note that $f'(x_k) \neq 0$ and, using the first equation in (15.12), we have that

$$\xi - y_k = -\frac{(\xi - x_k)^2}{2}\frac{f''(\eta_k)}{f'(x_k)}. \tag{15.13}$$

Hence,

$$|\xi - y_k| = \frac{1}{2}\frac{|f''(\eta_k)|}{|f'(x_k)|} \cdot |\xi - x_k| \cdot |\xi - x_k| \leq \frac{A}{2} \cdot h \cdot |\xi - x_k| \leq \frac{1}{2}|\xi - x_k| \ \leq \frac{h}{2}.$$

We can conclude that if $x_k \in [\xi - h, \xi + h]$, then $y_k \in [\xi - h, \xi + h]$ as well.

Now, using the second equation in (15.12), we have

$$\begin{aligned} x_{k+1} - \xi &= \frac{1}{f'(x_k)}\Big[f'(x_k)\,(y_k - \xi) - f(y_k)\Big] \\ &= \frac{1}{f'(x_k)}\Big[\big(f'(x_k) - f'(\xi)\big)\,(y_k - \xi) - f(y_k) + f'(\xi)\,(y_k - \xi)\Big]. \end{aligned}$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

By the Mean Value Theorem B.30, there is $\beta_k \in [\xi - h, \xi + h]$ between x_k and ξ, for which

$$f''(\beta_k)(x_k - \xi) = f'(x_k) - f'(\xi).$$

Furthermore, for some $\gamma_k \in [\xi - h, \xi + h]$ between ξ and y_k, we have

$$f(y_k) = f(\xi) + f'(\xi)(y_k - \xi) + \frac{f''(\gamma_k)}{2}(y_k - \xi)^2$$

from Taylor's Theorem B.32. Rearranging terms and using $f(\xi) = 0$ yields

$$-f(y_k) + f'(\xi)(y_k - \xi) = -\frac{f''(\gamma_k)}{2}(y_k - \xi)^2.$$

Putting things together, we find

$$x_{k+1} - \xi = \frac{1}{f'(x_k)}\left[f''(\beta_k)(x_k - \xi)\,(y_k - \xi) - \frac{f''(\gamma_k)}{2}(y_k - \xi)^2\right]. \tag{15.14}$$

Taking absolute values,

$$\begin{aligned}
|x_{k+1} - \xi| &= \frac{1}{|f'(x_k)|}\left|f''(\beta_k)\,(x_k - \xi)\,(y_k - \xi) - \frac{f''(\gamma_k)}{2}(y_k - \xi)^2\right| \\
&\le A|x_k - \xi| \cdot |y_k - \xi| + \frac{1}{2}|y_k - \xi| \cdot |y_k - \xi| A \\
&\le A|x_k - \xi|\frac{h}{2} + \frac{1}{2} \cdot \frac{h}{2} \cdot \frac{1}{2}|x_k - \xi| A \\
&\le \frac{1}{2}|x_k - \xi| + \frac{1}{8}|x_k - \xi| \\
&= \frac{5}{8}|x_k - \xi|.
\end{aligned}$$

We can conclude that if $x_k \in [\xi - h, \xi + h]$, then $x_{k+1} \in [\xi - h, \xi + h]$ as well. More importantly, we see by induction that

$$|\xi - x_k| \le \left(\frac{5}{8}\right)^k h,$$

which proves that $x_k \to \xi$ as $k \to \infty$. Using this fact, it is easy to see that y_k, β_k, γ_k, $\eta_k \to \xi$ as $k \to \infty$ as well.

Now, using (15.14), we see that

$$\frac{x_{k+1} - \xi}{(x_k - \xi)\,(y_k - \xi)} = \frac{f''(\beta_k)}{f'(x_k)} - \frac{(y_k - \xi)}{2(x_k - \xi)}\frac{f''(\gamma_k)}{f'(x_k)}.$$

Making use of (15.13),

$$\frac{(x_{k+1} - \xi)}{(x_k - \xi)^3} = \frac{f''(\beta_k)f''(\eta_k)}{2(f'(x_k))^2} - \frac{1}{8}(x_k - \xi)\frac{(f''(\eta_k))^2}{(f'(x_k))^2}\frac{f''(\gamma_k)}{f'(x_k)}.$$

Taking limits, we have

$$\lim_{k\to\infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|^3} = \frac{1}{2}\left|\frac{f''(\xi)}{f'(\xi)}\right|^2 = \sigma \in (0, \infty).$$

This shows that the convergence is exactly cubic. □

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

15.4.4 The Secant Method

Definition 15.32 (secant method)**.** Let $I \subseteq \mathbb{R}$ be an interval, $f \in C(I)$, and $x_0 \in I$. The **secant method** is an algorithm for computing the terms of the sequence $\{x_k\}_{k=0}^{\infty}$ via (15.4) with

$$s_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}, \quad k \geq 1. \tag{15.15}$$

We say that this method is **well defined** if and only if $x_0, x_1 \in I$ implies that $x_k \in I$ for all $k = 2, 3, \ldots$. We say that this method **converges** if and only if there is $\xi \in I$, with $f(\xi) = 0$, such that $x_k \to \xi$ as $k \to \infty$.

Theorem 15.33 (convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval and $f \in C(I)$ be such that there is $\xi \in I$ for which $f(\xi) = 0$. Set, for $\delta > 0$, $I_\delta = [\xi - \delta, \xi + \delta] \subseteq I$. Assume that there is $\delta > 0$ for which $f \in C^1(I_\delta)$ and, for simplicity, $f'(\xi) > 0$. The sequence $\{x_k\}_{k=0}^{\infty}$ defined by the secant method converges (at least) linearly to the root ξ as $k \to \infty$, provided that x_0 and x_1 are sufficiently close to ξ.*

Proof. Set $f'(\xi) = \alpha > 0$. By continuity, there is no loss in generality in assuming that, for all $x \in I_\delta$,

$$0 < \frac{3\alpha}{4} \leq f'(x) \leq \frac{5\alpha}{4}.$$

Suppose now that $x_k, x_{k-1} \in I_\delta$. By the Mean Value Theorem B.30, there is η_k between x_k and x_{k-1} such that $s_k = f'(\eta_k)$. Then

$$x_{k+1} - \xi = x_k - \xi - \frac{f(x_k)}{f'(\eta_k)}.$$

By Taylor's Theorem, there is a γ_k between x_k and ξ such that

$$f(x_k) = f(\xi) + f'(\gamma_k)(x_k - \xi) = f'(\gamma_k)(x_k - \xi).$$

Thus,

$$x_{k+1} - \xi = x_k - \xi - \frac{f'(\gamma_k)(x_k - \xi)}{f'(\eta_k)} = (x_k - \xi)\left[1 - \frac{f'(\gamma_k)}{f'(\eta_k)}\right] \leq \frac{2}{5}(x_k - \xi).$$

If $|x_0 - \xi| \leq \delta$ and $|x_1 - \xi| \leq \delta$, then, by induction, we see that, for $k \geq 2$,

$$|x_k - \xi| \leq \left(\frac{2}{5}\right)^{k-1} \delta.$$

This proves that the method is well defined and that $x_k \to \xi$ at least linearly. □

It turns out that the convergence of the secant method is super-linear.

Theorem 15.34 (super-linear convergence)**.** *Let $I \subseteq \mathbb{R}$ be an interval and $f \in C(I)$. In the setting of Theorem 15.33, assume, in addition, that $f \in C^2(I_\delta)$ and $f''(\xi) > 0$. Then the sequence $\{x_k\}_{k=0}^{\infty}$ generated by the secant method converges to ξ at the rate $q = \frac{1+\sqrt{5}}{2}$.*

Proof. See Problem 15.21. □

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

15.5 Newton's Method in Several Dimensions

In this section, we develop and analyze Newton's method for the solution of (15.1) in the case that $n = m = d > 1$. We will need several facts about basic calculus in several variables, and we refer the reader to Appendix B for a review.

Definition 15.35 (Newton's method)**.** Suppose that $d \in \mathbb{N}$, $\Omega \subseteq \mathbb{R}^d$ is an open, convex set, $\boldsymbol{x}_0 \in \Omega$ is given, and $\boldsymbol{f} \in C^1(\Omega; \mathbb{R}^d)$. **Newton's method in d-dimensions** is an algorithm for computing the terms of the sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ via the recursive iteration

$$\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)\,(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = -\boldsymbol{f}(\boldsymbol{x}_k), \tag{15.16}$$

where $\mathsf{J}_{\boldsymbol{f}}$ is the Jacobian matrix of $\boldsymbol{f}$. We say that the method is **well defined** if and only if $\boldsymbol{x}_k \in \Omega$ and $\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)$ is nonsingular, for all $k \in \mathbb{N}$. We say that Newton's method **converges** if and only if there is a $\boldsymbol{\xi} \in \Omega$, with $\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$, such that $\boldsymbol{x}_k \to \boldsymbol{\xi}$ as $k \to \infty$.

As in the one-dimensional case, Newton's method converges quadratically to the root.

Theorem 15.36 (convergence)**.** *Let $\boldsymbol{\xi} \in \mathbb{R}^d$ and $r > 0$ be given. Suppose that*

$$\boldsymbol{f} \in C^2(\overline{B}(\boldsymbol{\xi}, r); \mathbb{R}^d),$$

$\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$, and for every $\boldsymbol{x} \in \overline{B}(\boldsymbol{\xi}, r)$ the Jacobian matrix $\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x})$ is invertible, with the estimate

$$\left\| [\mathsf{J}(\boldsymbol{x})]^{-1} \right\|_2 \leq \beta.$$

Then the sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ defined by Newton's method (15.16) *converges (at least) quadratically to the root $\boldsymbol{\xi}$ as $k \to \infty$, provided that $\boldsymbol{x}_0$ is sufficiently close to $\boldsymbol{\xi}$.*

Proof. By Taylor's Theorem B.51, for each $i = 1, \ldots, d$, there is a point $\boldsymbol{\eta}_{k,i} \in B(\boldsymbol{\xi}, r)$ such that

$$0 = \boldsymbol{f}_i(\boldsymbol{\xi}) = \boldsymbol{f}_i(\boldsymbol{x}_k) + \nabla \boldsymbol{f}_i(\boldsymbol{x}_k)^{\mathsf{T}} (\boldsymbol{\xi} - \boldsymbol{x}_k) + c_{k,i},$$

where $c_{k,i} = \frac{1}{2} (\boldsymbol{\xi} - \boldsymbol{x}_k)^{\mathsf{T}} \mathsf{H}_i(\boldsymbol{\eta}_{k,i}) (\boldsymbol{\xi} - \boldsymbol{x}_k)$ and H_i is the Hessian matrix of $\boldsymbol{f}_i$. To simplify notation in the proof, let us set $\mathbf{c}_k = [c_{k,1}, \ldots, c_{k,d}]^{\mathsf{T}}$, $\mathsf{J}_k = \mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)$, and $\mathsf{H}^{(k,i)} = \mathsf{H}_i(\boldsymbol{\eta}_{k,i})$. Using the definition of Newton's method together with our Taylor expansion, we get

$$\nabla \boldsymbol{f}_i(\boldsymbol{x}_k)^{\mathsf{T}} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = -\boldsymbol{f}_i(\boldsymbol{x}_k) = \nabla \boldsymbol{f}_i(\boldsymbol{x}_k)^{\mathsf{T}} (\boldsymbol{\xi} - \boldsymbol{x}_k) + c_{k,i},$$

which simplifies to

$$\boldsymbol{\xi} - \boldsymbol{x}_{k+1} = -\mathsf{J}_k^{-1} \mathbf{c}_k.$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Using the Cauchy–Schwarz and other basic inequalities,

$$\begin{aligned}
\left\|\mathsf{J}_k^{-1}\mathbf{c}_k\right\|_2 &\le \left\|\mathsf{J}_k^{-1}\right\|_2 \|\mathbf{c}_k\|_2 \\
&\le \beta\sqrt{\sum_{i=1}^{d} c_{k,i}^2} \\
&= \frac{\beta}{2}\sqrt{\sum_{i=1}^{d} \left|(\boldsymbol{\xi}-\boldsymbol{x}_k)^{\mathsf{T}}\,\mathsf{H}^{(k,i)}\,(\boldsymbol{\xi}-\boldsymbol{x}_k)\right|^2} \\
&\le \frac{\beta}{2}\sqrt{\sum_{i=1}^{d} \|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2^2 \left\|\mathsf{H}^{(k,i)}\,(\boldsymbol{\xi}-\boldsymbol{x}_k)\right\|_2^2} \\
&= \frac{\beta}{2}\,\|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2 \sqrt{\sum_{i=1}^{d} \left\|\mathsf{H}^{(k,i)}\,(\boldsymbol{\xi}-\boldsymbol{x}_k)\right\|_2^2}.
\end{aligned}$$

Another application of Cauchy–Schwarz gives

$$\begin{aligned}
\left\|\mathsf{H}^{(k,i)}\,(\boldsymbol{\xi}-\boldsymbol{x}_k)\right\|_2^2 &= \sum_{j=1}^{d}\left|\sum_{m=1}^{d}\frac{\partial^2 \boldsymbol{f}_i}{\partial x_j \partial x_m}(\boldsymbol{\eta}_{k,i})(\xi_m - x_{k,m})\right|^2 \\
&\le \sum_{j=1}^{d}\left[\sum_{m=1}^{d}\left|\frac{\partial^2 \boldsymbol{f}_i}{\partial x_j \partial x_m}(\boldsymbol{\eta}_{k,i})\right|^2 \|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2^2\right] \\
&\le \|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2^2 \sum_{j=1}^{d}\left[\sum_{m=1}^{d} A^2\right] \\
&= \|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2^2\, A^2 d^2,
\end{aligned}$$

where A is an upper bound on the absolute values of the second derivatives of $\boldsymbol{f}$, which is available because $\boldsymbol{f} \in C^2\left(\overline{B}(\boldsymbol{\xi}, r); \mathbb{R}^d\right)$. We finally get the fundamental error estimate:

$$\|\boldsymbol{\xi}-\boldsymbol{x}_{k+1}\|_2 \le \frac{\beta A d^{3/2}}{2}\,\|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2^2\,.$$

Therefore, if $\|\boldsymbol{\xi}-\boldsymbol{x}_0\|_2 \le \frac{1}{\beta A d^{3/2}} = h$, then $\|\boldsymbol{\xi}-\boldsymbol{x}_1\|_2 \le \frac{1}{2}\,\|\boldsymbol{\xi}-\boldsymbol{x}_0\|_2$. By induction, it follows that

$$\|\boldsymbol{\xi}-\boldsymbol{x}_k\|_2 \le h\left(\frac{1}{2}\right)^{2^k-1} = \varepsilon_k.$$

Thus, $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ is well defined, $\boldsymbol{x}_k \to \boldsymbol{\xi}$, and the order is at least quadratic. □

The next result, due to Kantorovich, is interesting in that it requires less regularity than what we have assumed in the previous results. Additionally, the existence of the zero point is not a required assumption, but is a consequence of the convergence. The following proof is similar to that in [86].

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Theorem 15.37 (Kantorovich[5])**.** *Let $d \in \mathbb{N}$. Suppose that $\Omega \subset \mathbb{R}^d$ is an open, bounded, convex set, $\boldsymbol{x}_0 \in \Omega$, and $\boldsymbol{f} \in C^1\left(\overline{\Omega}; \mathbb{R}^d\right)$. Assume, additionally, with $\mathrm{J}_{\boldsymbol{f}}$ denoting the Jacobian matrix of $\boldsymbol{f}$, that there is $\gamma > 0$ such that*

$$\left\|\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}) - \mathrm{J}_{\boldsymbol{f}}(\boldsymbol{y})\right\|_2 \leq \gamma \left\|\boldsymbol{x} - \boldsymbol{y}\right\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \Omega.$$

Furthermore, let us assume the following.

a. *For all $\boldsymbol{x} \in \Omega$, the Jacobian matrix $\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x})$ is invertible and there is $\beta > 0$ such that*

$$\left\|\left[\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x})\right]^{-1}\right\|_2 \leq \beta, \quad \forall \boldsymbol{x} \in \Omega.$$

b. *The initial iterate, $\boldsymbol{x}_0 \in \Omega$, satisfies*

$$\left\|\left[\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)\right]^{-1} \boldsymbol{f}(\boldsymbol{x}_0)\right\|_2 \leq \alpha.$$

c. *The parameters satisfy*

$$h = \frac{\alpha\beta\gamma}{2} < 1.$$

d. *The initial iterate is well inside Ω, in the sense that*

$$\overline{B}(\boldsymbol{x}_0, r) \subseteq \Omega,$$

where $r = \frac{\alpha}{1-h}$.

In this setting, the sequence $\boldsymbol{x}_k$ defined by Newton's method (15.16) *is well defined; in particular, $\boldsymbol{x}_k \in B(\boldsymbol{x}_0, r)$ for each $k \in \mathbb{N}$. Moreover, there exists a point $\boldsymbol{\xi} \in \overline{B}(\boldsymbol{x}_0, r)$ such that $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{\xi}$, with the convergence estimate*

$$\left\|\boldsymbol{x}_k - \boldsymbol{\xi}\right\|_2 \leq \alpha \frac{h^{2^k - 1}}{1 - h^{2^k}}, \quad \forall k \in \mathbb{N}.$$

Since $0 < h < 1$, convergence is at least quadratic. Finally, the point $\boldsymbol{\xi}$ is a zero of the function $\boldsymbol{f}$, i.e., $\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$.

Proof. We split the proof into several steps.

1. Since $\left[\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x})\right]^{-1}$ exists for all $\boldsymbol{x} \in \Omega$, we will have that $\boldsymbol{x}_{k+1}$ is defined if $\boldsymbol{x}_k \in B(\boldsymbol{x}_0, r)$. Suppose that, for all $j = 0, 1, \ldots, k$, $\boldsymbol{x}_j \in B(\boldsymbol{x}_0, r)$. Then

$$\begin{aligned}
\left\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\right\|_2 &= \left\|\left[\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)\right]^{-1} \boldsymbol{f}(\boldsymbol{x}_k)\right\|_2 \\
&\leq \left\|\left[\mathrm{J}(\boldsymbol{x}_k)\right]^{-1}\right\|_2 \left\|\boldsymbol{f}(\boldsymbol{x}_k)\right\|_2 \\
&\leq \beta \left\|\boldsymbol{f}(\boldsymbol{x}_k)\right\|_2 \\
&= \beta \left\|\boldsymbol{f}(\boldsymbol{x}_k) - \boldsymbol{f}(\boldsymbol{x}_{k-1}) - \mathrm{J}(\boldsymbol{x}_{k-1})(\boldsymbol{x}_k - \boldsymbol{x}_{k-1})\right\|_2 \\
&\leq \frac{\beta\gamma}{2} \left\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\right\|_2^2,
\end{aligned} \tag{15.17}$$

[5] Named in honor of the Soviet mathematician Leonid Vitalyevich Kantorovich (1912–1986).

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

using the result of Theorem B.56 in the last step. We claim that (15.17) implies that, for all $k \geq 0$,

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_2 \leq \alpha h^{2^k - 1}. \tag{15.18}$$

The proof is by induction. The case $k = 0$ holds because of assumption b:

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2 = \left\| [\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)]^{-1} \boldsymbol{f}(\boldsymbol{x}_0) \right\|_2 \leq \alpha.$$

For the induction step, we suppose that (15.18) is valid for $k = j - 1$:

$$\|\boldsymbol{x}_j - \boldsymbol{x}_{j-1}\|_2 \leq \alpha h^{2^{j-1} - 1}.$$

Let $k = j$ now. Using (15.17) and the induction hypothesis,

$$\begin{aligned} \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\|_2 &\leq \frac{\beta\gamma}{2} \|\boldsymbol{x}_j - \boldsymbol{x}_{j-1}\|_2^2 \leq \frac{\beta\gamma}{2} \alpha^2 \left(h^{2^{j-1}-1} \right)^2 = \frac{\beta\gamma}{2} \alpha^2 h^{2^j - 2} \\ &= \frac{\alpha\beta\gamma}{2} \alpha h^{2^j - 2} = \alpha h^{2^j - 1}. \end{aligned}$$

Hence, estimate (15.18) follows by induction.

Now, by the triangle inequality,

$$\begin{aligned} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_0\|_2 &\leq \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_2 + \cdots + \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2 \\ &\leq \alpha \left(1 + h + h^3 + h^7 + \cdots + h^{2^k - 1} \right) \\ &< \alpha \left(1 + h + h^2 + \cdots \right) \\ &= \frac{\alpha}{1 - h} \\ &= r. \end{aligned}$$

Thus, $\boldsymbol{x}_{k+1} \in B(\boldsymbol{x}_0, r)$. By induction, $\boldsymbol{x}_k \in B(\boldsymbol{x}_0, r)$ for all $k \in \mathbb{N}$.

2. Using (15.18), we can prove that $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ is a Cauchy sequence. Suppose that $m > n \geq 0$. Then

$$\begin{aligned} \|\boldsymbol{x}_m - \boldsymbol{x}_n\|_2 &\leq \|\boldsymbol{x}_m - \boldsymbol{x}_{m-1}\|_2 + \cdots + \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\|_2 \\ &\leq \alpha h^{2^n - 1} \left(1 + h^{2^n} + h^{3 \cdot 2^n} + h^{5 \cdot 2^n} + \cdots \right) \\ &< \frac{\alpha h^{2^n - 1}}{1 - h^{2^n}} \\ &< \varepsilon, \end{aligned} \tag{15.19}$$

provided that n is sufficiently large. Since $\boldsymbol{x}_k$ is Cauchy, it converges to a unique limit point $\boldsymbol{\xi} \in \overline{B}(\boldsymbol{x}_0, r)$, appealing to Theorem B.8 and the fact that $\overline{B}(\boldsymbol{x}_0, r)$ is closed. It follows on taking $m \to \infty$ in (15.19) that

$$\|\boldsymbol{\xi} - \boldsymbol{x}_n\|_2 < \frac{\alpha h^{2^n - 1}}{1 - h^{2^n}}.$$

From this estimate, it follows that convergence is at least quadratic.

3. Finally, we prove that $\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$. Since $\boldsymbol{x}_k \in B(\boldsymbol{x}_0, r)$,

$$\|\mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_k) - \mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)\|_2 \leq \gamma \|\boldsymbol{x}_0 - \boldsymbol{x}_k\|_2 \leq \gamma r.$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

Thus,

$$\|\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)\|_2 = \|\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k) - \mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_0) + \mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)\|_2 \leq \gamma r + \|\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)\|_2 = R.$$

As a consequence,

$$\|\boldsymbol{f}(\boldsymbol{x}_k)\|_2 = \|-\mathsf{J}_{\boldsymbol{f}}(\boldsymbol{x}_k)\,(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)\|_2 \leq R\,\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_2\,,$$

which implies that $\lim_{k\to\infty} \|\boldsymbol{f}(\boldsymbol{x}_k)\|_2 = 0$. It follows that $\boldsymbol{f}(\boldsymbol{\xi}) = \boldsymbol{0}$.

The proof is complete. □

Problems

15.1 Suppose that, for every $k \geq 0$,

$$|\xi - x_k| \leq \alpha \frac{h^{2^k - 1}}{1 - h^{2^k}},$$

where $\alpha > 0$ and $0 < h < 1$. Show that the sequence $\{x_k\}_{k=0}^{\infty}$ converges to ξ at least quadratically.

15.2 Complete the proof of Theorem 15.6.

15.3 Let

$$f(x) = e^x - 2x - 1, \quad a = 1, \quad b = 2.$$

Show that the false position method in this setting will converge to a root of f. Show, in addition, that, for all k, we will have $b_k = 2$.

15.4 Can you generalize the example of the previous problem? In other words, let $-\infty < a < b < \infty$ and $f \in C^2([a, b])$ with $f(a)f(b) < 0$. Can you provide sufficient conditions on f, f', and f'', so that $b_k = b$ for all k?

15.5 Prove Proposition 15.11.

15.6 Complete the proof of Theorem 15.13.

15.7 Complete the proof of Theorem 15.15.

15.8 Prove Theorem 15.17.

15.9 Prove Proposition 15.19.

15.10 Let $\{x_k\}_{k=0}^{\infty}$ be the sequence generated, for some $g \in C([a, b])$, by the fixed point iteration scheme. In the setting of Theorem 15.12, assume, in addition, that $g'(x) < 0$ for all $x \in [a, b]$ and that $x_0 < \xi$, where ξ is the (unique) fixed point of g. Show that, for any $k \geq 0$,

$$x_{2k} < \xi < x_{2k+1}.$$

15.11 Let $a \in \mathbb{R}$ and, for some $r > 0$, $I = [a - r, a + r]$ be an interval. Let $g \in C(I)$ be such that there is $q \in (0, 1)$ for which

$$|g'(x)| \leq q, \quad \forall x \in I,$$

and

$$|g(a) - a| \leq (1 - q)r.$$

Show that g has a unique fixed point $\xi \in I$ and that the fixed point iteration scheme converges for any starting value x_0. Moreover,

$$|x_k - \xi| \leq q^k |x_0 - \xi|.$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

15.12 Consider the relaxation method (15.3). Show that if $f:\mathbb{R}\to\mathbb{R}$ is such that $f'(x)<0$ and $|f'(x)|\in[m,M]\subset(0,\infty)$ for all $x\in\mathbb{R}$, then the choice

$$\lambda=\frac{2}{m+M}$$

is optimal for the relaxation parameter.

15.13 Let $\boldsymbol{f}:\mathbb{R}^d\to\mathbb{R}^d$ be given by

$$\boldsymbol{f}(\boldsymbol{x})=\mathrm{A}\boldsymbol{x}+\boldsymbol{g}(\boldsymbol{x}),$$

where $\mathrm{A}\in\mathbb{R}^{d\times d}$ is invertible and $\boldsymbol{g}\in C^1(\mathbb{R}^d;\mathbb{R}^d)$. Assume that there is $\boldsymbol{\xi}\in\mathbb{R}^d$ for which $\boldsymbol{f}(\boldsymbol{\xi})=\boldsymbol{0}$. To approximate it, consider the following *Picard-like* iteration method: given $\boldsymbol{x}_0\in\mathbb{R}^d$, find $\boldsymbol{x}_{k+1}$, for $k\geq 0$, via

$$\mathrm{A}\boldsymbol{x}_{k+1}+\boldsymbol{g}(\boldsymbol{x}_k)=\boldsymbol{0}.$$

Provide sufficient conditions for the convergence of this approach.
Hint: Find an expression for the error $\boldsymbol{e}_k=\boldsymbol{\xi}-\boldsymbol{x}_k$. Then, use a version of the Mean Value Theorem in multiple dimensions.

15.14 Assume that $f\in C^2(\mathbb{R})$ with $f'(x)>0$ and $f''(x)>0$ for all $x\in\mathbb{R}$.

a) Exhibit a function that satisfies these assumptions, but has no root.

b) Show that, if a root $\xi\in\mathbb{R}$ exists, it is unique.

c) Prove that, for any starting guess $x_0\in\mathbb{R}$, Newton's method converges and the convergence is quadratic.

15.15 Show that (15.26) will converge, for any $x_0>0$, to $\sqrt{5}$.

15.16 Define $f:[1,5]\to\mathbb{R}$ via $f(x)=(x-3)^3$. Use Newton's method with the starting point $x_0=4$ to approximate the root $\xi=3$. Show directly that convergence is linear.

Hint: Show that, for $k\geq 0$,

$$x_k=3+\left(\frac{2}{3}\right)^k.$$

15.17 Suppose that $f\in C^2(I)$, where I is an interval. Let $\xi\in I$ be such that $f(\xi)=f'(\xi)=0$, but $f''(\xi)\neq 0$.

a) Show that the sequence $\{x_k\}_{k=0}^{\infty}$ defined by Newton's method satisfies the relation

$$\xi-x_{k+1}=-\frac{1}{2}\frac{(\xi-x_k)^2f''(\eta^k)}{f'(x_k)}=\frac{1}{2}(\xi-x_k)\frac{f''(\eta_k)}{f''(\chi_k)},$$

where η_k and χ_k lie between ξ and x_k.

b) Suppose that $0<m\leq|f''(x)|\leq M$ for all $x\in[\xi-\delta,\xi+\delta]\subset I$ for some $\delta>0$, where $0<M<2m$. Prove that if $x_0\in[\xi-\delta,\xi+\delta]$, then $x_k\to\xi$.

15.18 Complete the proof of Theorem 15.26.

15.19 Suppose that $f\in C^2(\mathbb{R})$ with f'' Lipschitz continuous and $f(\zeta)=f'(\zeta)=0$, but $f''(\zeta)\neq 0$.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

a) Prove that the iterative method

$$x_{k+1} = x_k - 2\frac{f(x_k)}{f'(x_k)}$$

converges at least quadratically to ζ provided that x_0 is sufficiently near, but not equal, to ζ.

b) Can one extend the last result for the case

$$f(\zeta) = f'(\zeta) = f''(\zeta) = 0, \quad \text{but} \quad f'''(\zeta) \neq 0?$$

What method, if any, would still give quadratic convergence?

15.20 Use the secant method to show that the sequence, whose recursive definition is given below, converges to $\sqrt{Q}$, where $Q > 0$, given "good" starting values x_0 and x_1:

$$x_{k+1} = \frac{x_k x_{k-1} + Q}{x_k + x_{k-1}}.$$

Come up with a similar recursion for approximating $Q^{1/3}$ using the secant method.

15.21 Prove Theorem 15.34. To do so, proceed as follows.

a) Prove the iterations are well defined and converge.

b) Show that the secant method may be written in the equivalent form,

$$x_{k+1} = \frac{x_k f(x_{k-1}) - x_{k-1} f(x_k)}{f(x_{k-1}) - f(x_k)}, \quad k \geq 1.$$

c) Define

$$\phi(x_k, x_{k-1}) = \frac{x_{k+1} - \xi}{(x_k - \xi)(x_{k-1} - \xi)},$$

where x_{k+1} is expressed in terms of x_k and x_{k-1} through the recursive formula above. Find an expression for

$$\psi(x_{k-1}) = \lim_{x_k \to \xi} \phi(x_k, x_{k-1})$$

and then determine the value of

$$\lim_{x_{k-1} \to \xi} \psi(x_{k-1}).$$

d) Deduce that

$$\lim_{x_k, x_{k-1} \to \xi} \phi(x_k, x_{k-1}) = \frac{f''(\xi)}{2f'(\xi)}.$$

e) Next, suppose that

$$\lim_{k \to \infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|^q} = A > 0.$$

Prove that it must be that $q - 1 - 1/q = 0$ and, therefore, $q = (1 + \sqrt{5})/2$.

f) Finally, deduce that

$$\lim_{k \to \infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|^q} = \left| \frac{f''(\xi)}{2f'(\xi)} \right|^{q/(1+q)}.$$

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

15.22 Consider the nonlinear equation $\exp(x) = \sin(x)$.

a) With pencil and paper only, argue that there is one, and only one, solution $\xi \in \left(-\frac{3}{2}\pi, -\pi\right)$.

b) Show, in fact, that there is one, and only one, solution $\xi \in \left(-\frac{5}{4}\pi, -\pi\right)$.

c) Consider the following iterative methods:

$$x_{k+1} = \ln(\sin(x_k))$$

and

$$x_{k+1} = \sin^{-1}(\exp(x_k)),$$

where the inverse of the sine function is appropriately, and carefully, defined. What can you say about the local convergence of each of these methods to ξ and their convergence orders?

d) For estimating ξ, provide a method that is quadratically convergent. Will the method converge for any starting value in the interval $x_0 \in \left(-\frac{5}{4}\pi, -\pi\right)$? Why or why not?

15.23 In this method, we will explore a variant of the simplified Newton method. Let $f \in C(\mathbb{R})$. Let $\xi \in \mathbb{R}$ be such that $f(\xi) = 0$ and $f'(\xi) > 0$. Show that there is an $\varepsilon > 0$ such that if $|x_0 - \xi| < \varepsilon$, then the following iteration converges to ξ:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_m)},$$

where $m \leq k$ is chosen such that $|f'(x_m)| = \max_{j \leq k} |f'(x_j)|$.

15.24 In this problem, we will construct iterative schemes to find a solution of $f(\xi) = 0$ that converges with orders $q = 2$ and $q = 3$, respectively. Assume that $f(\xi) = 0$ can be rewritten as the fixed point of g, i.e., $\xi = g(\xi)$, so that we consider the iterative scheme

$$x_{k+1} = g(x_k).$$

a) Define the error $e_k = \xi - x_k$. Show that

$$|g(\xi - e_{k-1}) - g'(\xi)e_{k-1}| \leq C|e_{k-1}|^2$$

for some constant $C > 0$.

b) Show that if $g'(\xi) \neq 0$, then the method converges linearly.

c) Show that if $g'(\xi) = 0$, but $g''(\xi) \neq 0$, the method converges quadratically.

d) Consider, for a_1 and a_2, nonvanishing functions

$$g(x) = x + a_1(x)f(x) + a_2(x)[f(x)]^2. \tag{15.20}$$

Show that $x = g(x)$ if and only if $f(x) = 0$.

e) Let g be given by (15.20). Evaluate the first and second derivatives of g with respect to x. From them, show that if $q = 2$ (the method converges quadratically), then we obtain Newton's method.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

f) Let g be given by (15.20). Show that if the method converges cubically, i.e., $q = 3$, then

$$a_1(x) = -\frac{1}{f'(x)}, \qquad a_2(x) = -\frac{f''(x)}{2[f'(x)]^3}.$$

Unfortunately, it turns out that this method is unstable. It is only used to accelerate the convergence once a good guess is already available.

15.25 Let $\boldsymbol{f}\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ be twice continuously differentiable. Suppose that $\boldsymbol{\xi} \in \Omega$ is a solution of $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ and the Jacobian matrix of $\boldsymbol{f}$, denoted $\mathrm{J}_{\boldsymbol{f}}$, is invertible at $\boldsymbol{\xi}$. Prove that if $\boldsymbol{x}_0 \in \Omega$ is sufficiently close to $\boldsymbol{\xi}$, then the following iteration converges to $\boldsymbol{\xi}$:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mathrm{J}_{\boldsymbol{f}}(\boldsymbol{x}_0)^{-1}\, \boldsymbol{f}(\boldsymbol{x}_k).$$

15.26 Suppose that the function $\boldsymbol{f}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined via

$$\boldsymbol{f}(x_1, x_2) = \begin{bmatrix} 16 - x_1^2 - x_2^2 \\ x_1^2 - 1 \end{bmatrix}.$$

How many real-valued (vector) solutions does the system $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ have? Use Newton's method to obtain the approximations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ when $\boldsymbol{x}_0 = [1, 1]^\intercal$.

15.27 Suppose that the function $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined via

$$\boldsymbol{f}(x_1, x_2) = \begin{bmatrix} x_1^2 - 2x_1 + x_2^2 \\ x_1^2 + x_2^2 - 1 \end{bmatrix}.$$

How many real-valued (vector) solutions does the system $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ have? Use Newton's method to obtain the approximations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ when $\boldsymbol{x}_0 = [0, -1]^\intercal$.

15.28 By $B(\boldsymbol{x}, r) \subset \mathbb{R}^2$, denote the open ball of radius $r > 0$ centered at $\boldsymbol{x}$. Suppose that, for some $r > 0$, f, $g\colon B(\boldsymbol{\xi}, r) \to \mathbb{R}$ are nonlinear, twice continuously differentiable functions with

$$f(\boldsymbol{\xi}) = 0, \quad g(\boldsymbol{\xi}) = 0.$$

Consider the Gauss–Seidel-like iterative scheme: given

$$\boldsymbol{x}_k = [x_{1,k}, x_{2,k}]^\intercal \in B(\boldsymbol{\xi}, r),$$

find $\boldsymbol{x}_{k+1} = [x_{1,k+1}, x_{2,k+1}]^\intercal \in \mathbb{R}^2$ such that

$$f(x_{1,k+1}, x_{2,k}) = 0, \quad g(x_{1,k+1}, x_{2,k+1}) = 0.$$

a) Let $\boldsymbol{e}_k = \boldsymbol{\xi} - \boldsymbol{x}_k$ be the error.

b) Establish an iteration error equation of the form

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{\xi}) & \frac{\partial f}{\partial x_2}(\boldsymbol{\xi}) \\ \frac{\partial g}{\partial x_1}(\boldsymbol{\xi}) & \frac{\partial g}{\partial x_2}(\boldsymbol{\xi}) \end{bmatrix} \boldsymbol{e}_{k+1} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{\xi})\boldsymbol{e}_{1,k+1} + \frac{\partial f}{\partial x_2}(\boldsymbol{\xi})\boldsymbol{e}_{2,k+1} \\ \frac{\partial g}{\partial x_1}(\boldsymbol{\xi})\boldsymbol{e}_{1,k+1} + \frac{\partial g}{\partial x_2}(\boldsymbol{\xi})\boldsymbol{e}_{2,k+1} \end{bmatrix} = \boldsymbol{r}_{k+1}.$$

Give a precise expression for the remainder term, $\boldsymbol{r}_{k+1}$.

c) Give sufficient conditions for the convergence of the scheme.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

15.29 Let $\boldsymbol{f}: \mathbb{R}^2 \to \mathbb{R}^2$ be defined via

$$\boldsymbol{f}(x_1, x_2) = \begin{bmatrix} x_1^2 - 2x_1 + x_2 \\ 2x_1 - x_2^2 - 1 \end{bmatrix}.$$

Observe that $\boldsymbol{f}$ has the zero

$$\boldsymbol{\xi} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Consider the iteration

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \mathrm{A}\boldsymbol{f}(\boldsymbol{x}_n), \qquad \mathrm{A} = \begin{bmatrix} 1 & 1/2 \\ 1 & 0 \end{bmatrix}. \tag{15.21}$$

a) Prove that $\boldsymbol{x}_n \to \boldsymbol{\xi}$, provided that $\boldsymbol{x}_0$ is sufficiently close to $\boldsymbol{\xi}$.
b) Show that the convergence is at least quadratic.
c) Is the iteration (15.21) equivalent to Newton's method?

Listings

```
function [root,count,err] = NewtonRoot(xin, p, q, tol, maxits)
%
% This function calculates the pth root of q ,
%
%   q^(1/p)
%
% using Newton's method. For simplicity, we assume that p is a
% positive integer and q is a positive real number
%
% Input:
%   xin : initial guess
%   p : the positive integer degree of the root
%   q : the positive number whose pth root is to estimated
%   tol : stopping tolerance
%   maxits : the maximal number of iterations
%
% Output:
%   root : approximation of the root
%   count : number of newton iterations required to compute the
%           root
%   err: = 0, if the algorithm proceeded to completion
%        = 1, if an error was encountered
%
  root = NaN;
  count = 0;
  err = 0;
  if int32(p) ~= p || p < 0
    disp('Error: p must be a positive integer');
    err = 1;
  return
  end
  if q < 0
```

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press

```
33      disp('Error: q must be positive');
34      err = 1;
35    return
36    end
37    diff = 1.0;
38    x = xin;
39    while diff > tol && count < maxits
40      xo = x;
41      x = xo - fn(xo,p,q)/dfn(xo,p,q);
42      diff = abs(x-xo);
43      count = count+1;
44    end
45    if count >= maxits
46      err = 1;
47    end
48    root = x;
49  end
50
51  function y = fn(x,p,q)
52    y = x^p-q;
53  end
54
55  function y = dfn(x,p,q)
56    y = p*x^(p-1);
57  end
```

Listing 15.1 Newton's method for computing $\sqrt[p]{q}$.

https://doi.org/10.1017/9781108942607.020 Published online by Cambridge University Press