Chapter 2

Nonlinear systems of equations

A general problem in mathematics: X, Y are normed vector spaces, and $f : X \mapsto Y$. Find $x \in X$ such that $f(x) = 0$.

Example 2.1 ① Find a non-zero $x \in R$ such that $x = \tan x$ (in wave diffraction).

② Find $(x, y, z) \in \mathbb{R}^3$ for which

$$\begin{aligned} z^2 - zy + 1 &= 0 \\ x^2 - 2 - y^2 - xyz &= 0 \\ e^y + 3 - e^x - 2 &= 0. \end{aligned}$$

③ Find a non-zero, twice differentiable function $y(t)$ for which

$$t\, y''(t) + (1 - t)y'(t) - y = 0.$$

Here $f : y \mapsto ty'' + (1 - t)y' - y$.

Comment:

① There are no general theorems of existence/uniqueness for nonlinear systems.

② Direct versus iterative methods.

③ Iterative algorithms: accuracy, efficiency, robustness, ease of implementation, tolerance, stopping criteria.

2.1 The bisection method

The bisection method applies for root finding in $\mathbb{R}$, and is based on the following elementary theorem:

Theorem 2.1 (Intermediate value theorem) Let $f \in C[a,b]$ such that (with no loss of generality) $f(a) < f(b)$. For any y such that $f(a) < y < f(b)$ there exists an $x \in (a,b)$ such that $f(x) = y$. In particular, if $f(a)f(b) < 0$, then there exists an $x \in (a,b)$ such that $f(x) = 0$.

The method of proof coincides with the root finding algorithm. Given a, b such that $f(a)f(b) < 0$, we set $c = \frac{1}{2}(a+b)$ to be the mid-point. If $f(a)f(c) < 0$ then we set $b := c$, otherwise we set $a := c$.

Stopping criteria:

① Number of iterations M.
② $|f(c)| < \epsilon$.
③ $|b-a| < \delta$.

Algorithm

Algorithm 2.1.1: BISECTION$(a, b, M, \delta, \epsilon)$

$f_a \leftarrow f(a)$
$f_b \leftarrow f(b)$
$\Delta \leftarrow b - a$
if $f_a f_b > 0$ **return** $(error)$
for $k \leftarrow 1$ **to** M
do $\begin{cases} \Delta \leftarrow \frac{1}{2}\Delta \\ c \leftarrow a + \Delta \\ f_c \leftarrow f(c) \\ \textbf{if } |\Delta| < \delta \text{ or } |f_c| < \epsilon \quad \textbf{return } (c) \\ \textbf{if } f_c f_a < 0 \\ \quad \textbf{then } b \leftarrow c, f_b \leftarrow f_c \\ \quad \textbf{else } a \leftarrow c, f_a \leftarrow f_c \end{cases}$
return $(error)$

Comments:

① There is one evaluation of f per iteration.
② There may be more than one root.

Error analysis Given (a, b) the initial guess is $x_0 = \frac{1}{2}(a+b)$. Let $e_n = x_n - r$ be the ***error***, where r is the/a root. Clearly,

$$|e_0| \le \frac{1}{2}|b-a| \equiv E_0.$$

After n steps we have

$$|e_n| \le \frac{1}{2^{n+1}}|b-a| \equiv E_n.$$

Note that we don't know what e_n is (if we knew the error, we would know the solution); we only have an ***error bound***, E_n. The sequence of error bounds satisfies,

$$E_{n+1} = \frac{1}{2}E_n,$$

so that the bisection method converges linearly.

Complexity Consider an application of the bisection method, where the stopping criterion is determined by δ. The number of steps needed is determined by the condition:

$$\frac{1}{2^{n+1}}|b-a| \le \delta,$$

i.e.,

$$n+1 \ge \log_2 \frac{|b-a|}{\delta}.$$

(If for example the initial interval is of length 1 and a tolerance of 10^{-16} is needed, then the number of steps exceeds $n = 50$.)

Advantages and disadvantages

Advantages	**Disvantages**
always works	systems in $\mathbb{R}^n$
easy to implement	slow convergence
requires only continuity	requires initial data a, b

✎ *Exercise 2.1* Find a positive root of

$$x^2 - 4x \sin x + (2 \sin x)^2 = 0$$

accurate to two significant digits. *Use a hand calculator!*

2.2 Iterative methods

We are looking for roots r of a function $f; X \mapsto Y$. Iterative methods generate an ***approximating sequence*** (x_n) by starting with an initial value x_0, and generating the sequence with an ***iteration function*** $\Phi : X \mapsto X$,

$$x_{n+1} = \Phi(x_n).$$

Suppose that each ***fixed point*** ζ of Φ corresponds to a root of f, and that Φ is continuous in a neighborhood of ζ, then ***if*** the sequence (x_n) converges, then by the continuity of Φ, it converges to a fixed point of Φ, i.e., to a root of f.

General questions (1) How to choose Φ? (2) Will the sequence (x_n) converge? How fast will it converge?

Example 2.2 Set $\Phi(x) = x - f(x)$ so that

$$x_{n+1} = x_n - f(x_n).$$

If the sequence converges and f is continuous, then it converges to a root of f.

Example 2.3 (Newton's method in $\mathbb{R}$) If f is differentiable,

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

Show the geometrical construction. Also,

$$0 = f(r) = f(x_n) + (r - x_n) f'(x_n) + \frac{1}{2}(r - x_n)^2 f''(x_n + \theta(r - x_n)),$$

for some $\theta \in (0, 1)$. If we neglect the remainder we obtain

$$r \approx x_n - \frac{f(x_n)}{f'(x_n)}.$$

✎ *Exercise 2.2 (Computer exercise)* Write a Matlab function which gets for input the name of a real-valued function f, an initial value x_0, a maximum number of iterations M, and a tolerance ϵ. Let your function then perform iterations based on Newton's method for finding roots of f, until either the maximum of number iterations has been exceeded, or the convergence criterion $|f(x)| \le \epsilon$ has been reached. Experiment your program on the function $f(x) = \tan^{-1} x$, whose only root is $x = 0$. Try to characterize those initial values x_0 for which the iteration method converges.

Example 2.4 (Newton's method in $\mathbb{R}^n$) Now we're looking for the root $r = (r_1, \dots, r_n)$ of a function $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, which means

$$\begin{aligned} f_1(x_1, \dots, x_n) &= 0 \\ f_2(x_1, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \dots, x_n) &= 0 \end{aligned}$$

Using the same ***linear approximation***:

$$0 = f(r) \approx f(x_n) + Df(x_n) \cdot (r - x_n),$$

where Df is the differential of f, from which we obtain

$$r \approx x_n - [Df(x_n)]^{-1} \cdot f(x_n) \equiv x_{n+1}.$$

Example 2.5 (Secant method in $\mathbb{R}$) Slightly different format. The secant line is

$$y = f(x_n) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}(x - x_n).$$

We define x_{n+1} to be the interaction with the x-axis:

$$x_{n+1} = x_n - \frac{f(x_n)}{[f(x_n) - f(x_{n-1})]/(x_n - x_{n-1})}.$$

Think of it as an iteration

$$\begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix} = \Phi \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix}.$$

Definition 2.1 (Local and global convergence) *Let Φ be an iteration function on a complete normed vector space $(X, \|\cdot\|)$, and let ζ be a fixed point of Φ. The iterative method defined by Φ is said to be* ***locally convergent*** *if there exists a neighbourhood $\mathcal{N}(\zeta)$ of ζ, such that for all $x_0 \in \mathcal{N}(\zeta)$, the sequence (x_n) generated by Φ converges to ζ. The method is called* ***globally convergent*** *if $\mathcal{N}(\zeta)$ can be extended to the whole space X.*

Definition 2.2 (Order of an iteration method) *Let Φ be an iteration function on a complete normed vector space $(X, \|\cdot\|)$, and let ζ be a fixed point of Φ. If there exists a neighbourhood $\mathcal{N}(\zeta)$ of ζ, such that*

$$\|\Phi(x) - \zeta\| \leq C\|x - \zeta\|^p, \qquad \forall x \in \mathcal{N}(\zeta),$$

for some $C > 0$ and $p > 1$, or $0 < C < 1$ and $p = 1$, then the iteration method is said to be of order (at least) p at the point ζ.

Theorem 2.2 *Every iterative method Φ of order at least p at ζ is locally convergent at that point.*

Proof: Let $\mathcal{N}(\zeta)$ be the neighbourhood of ζ where the iteration has order at least p. Consider first the case $C < 1$, $p = 1$, and take any open ball

$$B_r(\zeta) = \{x \in X : \|x - \zeta\| < r\} \subseteq \mathcal{N}(\zeta).$$

If $x \in B_r(\zeta)$ then

$$\|\Phi(x) - \zeta\| \leq C\|x - \zeta\| < \|x - \zeta\| < r,$$

hence $\Phi(x) \in B_r(\zeta)$ and the entire sequence lies in $B_r(\zeta)$. By induction,

$$\|x_n - \zeta\| \leq C^n \|x_0 - \zeta\| \to 0,$$

hence the sequence converges to ζ.

If $p > 1$, take $B_r(\zeta) \subseteq \mathcal{N}(\zeta)$, with r sufficiently small so that $Cr^{p-1} < 1$. If $x \in B_r(\zeta)$ then

$$\|\Phi(x) - \zeta\| \leq C\|x - \zeta\|^{p-1}\|x - \zeta\| < Cr^{p-1}\|x - \zeta\| < \|x - \zeta\|,$$

hence $\Phi(x) \in B_r(\zeta)$ and the entire sequence lies in $B_r(\zeta)$. By induction,

$$\|x_n - \zeta\| \leq (Cr^{p-1})^n \|x_0 - \zeta\| \to 0,$$

hence the sequence converges to ζ.

■

One dimensional cases Consider the simplest case where $(X, \|\cdot\|) = (\mathbb{R}, |\cdot|)$. If Φ is differentiable in a neighbourhood $\mathcal{N}(\zeta)$ of a fixed point ζ, with $|\Phi'(x)| \leq C < 1$ for all $x \in \mathcal{N}(\zeta)$, then

$$\Phi(x) = \Phi(\zeta) + \Phi'(\zeta + \theta(x - \zeta))(x - \zeta),$$

from which we obtain

$$|\Phi(x) - \zeta| \leq C|x - \zeta|,$$

i.e., the iteration method is at least first order and therefore converges locally. [Show geometrically the cases $\Phi'(x) \in (-1, 0)$ and $\Phi'(x) \in (0, 1)$.]

Example 2.6 Suppose we want to find a root ζ of the function $f \in C^1(\mathbb{R})$ with the iteration

$$x_{n+1} = x_n + \alpha f(x_n),$$

i.e., $\Phi(x) = x + \alpha f(x)$. Suppose furthermore that $f'(\zeta) = M$. Then, for every $\epsilon > 0$ there exists a neighbourhood $\mathcal{N}(\zeta) = (\zeta - \delta, \zeta + \delta)$ such that

$$|f'(x) - M| \leq \epsilon, \qquad \forall x \in \mathcal{N}(\zeta).$$

In this neighbourhood,

$$|\Phi'(x)| = |1 + \alpha f'(x)|,$$

which is less than one provided that

$$-2 + |\alpha|\epsilon < \alpha M < -|\alpha|\epsilon.$$

Thus, the iteration method has order at least linear provided that α has sign opposite to that of $f'(\zeta)$, and is sufficiently small in absolute value.

If Φ is sufficiently often differentiable in a neighbourhood $\mathcal{N}(\zeta)$ of a fixed point ζ, with

$$\Phi'(\zeta) = \Phi''(\zeta) = \cdots = \Phi^{(p-1)}(\zeta) = 0,$$

then for all $x \in \mathcal{N}(\zeta)$,

$$\Phi(x) = \Phi(\zeta) + \Phi'(\zeta)(x - \zeta) + \cdots + \frac{\Phi^{(p)}(\zeta + \theta(x - \zeta))}{p!}(x - \zeta)^p,$$

i.e.,

$$|\Phi(x) - \zeta| = \frac{|\Phi^{(p)}(\zeta + \theta(x - \zeta))|}{p!}|x - \zeta|^p.$$

If $\Phi^{(p)}$ is bounded in some neighbourhood of ζ, say $|\Phi^{(p)}(x)| \leq M$, then

$$|\Phi(x) - \zeta| \leq \frac{M}{p!}|x - \zeta|^p,$$

so that the iteration method is at least of order p, and therefore locally convergent. Moreover,

$$\lim_{n\to\infty} \frac{|\Phi(x) - \zeta|}{|x - \zeta|^p} = \frac{|\Phi^{(p)}(\zeta)|}{p!},$$

i.e., the method is precisely of order p.

Example 2.7 Consider Newton's method in $\mathbb{R}$,

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

and assume that f has a simple zero at ζ, i.e., $f'(\zeta) \neq 0$. Then,

$$\Phi'(\zeta) = \left.\frac{f(x)f''(x)}{[f'(x)]^2}\right|_{x=\zeta} = 0,$$

and

$$\Phi''(\zeta) = \frac{f''(\zeta)}{f'(\zeta)},$$

the latter being in general different than zero. Thus, Newton's method is of second order and therefore locally convergent.

✎ *Exercise 2.3* The two following sequences constitute iterative procedures to approximate the number $\sqrt{2}$:

$$x_{n+1} = x_n - \frac{1}{2}(x_n^2 - 2), \qquad x_0 = 2,$$

and

$$x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}, \qquad x_0 = 2.$$

① Calculate the first six elements of both sequences.

② Calculate (numerically) the error, $e_n = x_n - \sqrt{2}$, and try to estimate the order of convergence.

③ Estimate the order of convergence by Taylor expansion.

✎ *Exercise 2.4* Let a sequence x_n be defined inductively by

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

Suppose that $x_n \to x$ as $n \to \infty$ and that $F'(x) = 0$. Show that $x_{n+2} - x_{n+1} = o(x_{n+1} - x_n)$. (Hint: assume that F is continuously differentiable and use the mean value theorem.)

✎ *Exercise 2.5* Analyze the following iterative method,

$$x_{n+1} = x_n - \frac{f^2(x_n)}{f(x_n + f(x_n)) - f(x_n)},$$

designed for the calculation of the roots of $f(x)$ (this method is known as Steffensen's method). Prove that this method converges quadratically (order 2) under certain assumptions.

✎ *Exercise 2.6* Kepler's equation in astronomy is $x = y - \epsilon \sin y$, with $0 < \epsilon < 1$. Show that for every $x \in [0, \pi]$, there is a y satisfying this equation. (Hint: Interpret this as a fixed-point problem.)

Contractive mapping theorems General theorems on the convergence of iterative methods are based on a fundamental property of mapping: contraction.

Theorem 2.3 (Contractive mapping theorem) Let K be a closed set in a complete normed space $(X, \|\cdot\|)$, and let Φ be a continuous mapping on X such that (i) $\Phi(K) \subseteq K$, and there exists a $C < 1$ such that for every $x, y \in K$,

$$\|\Phi(x) - \Phi(y)\| \leq C\|x - y\|.$$

Then,

① *The mapping Φ has a unique fixed point ζ in K.*

② *For every $x_0 \in K$, the sequence (x_n) generated by Φ converges to ζ.*

Proof: Since $\Phi(K) \subseteq K$, $x_0 \in K$ implies that $x_n \in K$ for all n. From the contractive property of Φ we have

$$\|x_n - x_{n-1}\| \leq C\|x_{n-1} - x_{n-1}\| \leq C^{n-1}\|x_1 - x_0\|.$$

Now, write x_n as

$$x_n = x_0 + \sum_{j=1}^{n} (x_j - x_{j-1}).$$

For any $m < n$,

$$\begin{aligned} \|x_n - x_m\| &\le \sum_{j=m+1}^{n} \|x_j - x_{j-1}\| \le \sum_{j=m+1}^{n} C^{j-1} \|x_1 - x_0\| \\ &\le \sum_{j=m+1}^{\infty} C^{j-1} \|x_1 - x_0\| \le \frac{C^m}{1-C} \|x_1 - x_0\|, \end{aligned}$$

which converges to zero as $m, n \to \infty$. Thus (x_n) is a Cauchy sequence, and since X is complete it converges to a limit ζ, which must reside in K since K is closed. The limit point must on the other hand be a fixed point of Φ.

Uniqueness is immediate for if ζ, ξ are distinct fixed point in K, then

$$\|\zeta - \xi\| = \|\Phi(\zeta) - \Phi(\xi)\| \le C \|\zeta - \xi\| < \|\zeta - \xi\|,$$

which is a contradiction. ■

Example 2.8 Consider for example the mapping

$$x_{n+1} = 3 - \frac{1}{2}|x_n|$$

on $\mathbb{R}$. Then,

$$|x_{n+1} - x_n| = \frac{1}{2}\big||x_n| - |x_{n-1}|\big| \le \frac{1}{2}|x_n - x_{n-1}|.$$

Hence, for every x_0 the sequence (x_n) converges to the unique fixed point $\zeta = 2$.

✎ *Exercise 2.7* Let p be a positive number. What is the value of the following expression:

$$x = \sqrt{p + \sqrt{p + \sqrt{p + \cdots}}}.$$

By that, I mean the sequence $x_0 = p$, $x_{k+1} = \sqrt{p + x_k}$. (Interpret this as a fixed-point problem.)

✎ *Exercise 2.8* Show that the function

$$F(x) = 2 + x - \tan^{-1} x$$

satisfies $|F'(x)| < 1$. Show then that $F(x)$ doesn't have fixed points. Why doesn't this contradict the contractive mapping theorem?

✎ *Exercise 2.9* Bailey's iteration for calculating $\sqrt{a}$ is obtained by the iterative scheme:

$$x_{n+1} = g(x_n) \qquad g(x) = \frac{x(x^2+3a)}{3x^2+a}.$$

Show that this iteration is of order at least three.

✎ *Exercise 2.10* (Here is an exercise which tests whether you *really* understand what root finding is about.) One wants to solve the equation $x + \ln x = 0$, whose root is $x \sim 0.5$, using one or more of the following iterative methods:

$$(i) \quad x_{k+1} = -\ln x_k \qquad (ii) \quad x_{k+1} = e^{-x_k} \qquad (iii) \quad x_{k+1} = \frac{x_k + e^{-x_k}}{2}.$$

① Which of the three methods *can* be used?

② Which method *should* be used?

③ Give an even better iterative formula; explain.

2.3 Newton's method in $\mathbb{R}$

We have already seen that Newton's method is of order two, provided that $f'(\zeta) \neq 0$, therefore locally convergent. Let's first formulate the algorithm

Algorithm 2.3.1: NEWTON(x_0, M, ϵ)

$y \leftarrow f(x_0)$
if $|y| < \epsilon$ **return** (x_0)
for $k \leftarrow 1$ **to** M
do $\begin{cases} x \leftarrow x_0 - f(x_0)/f'(x_0) \\ y \leftarrow f(x_0) \\ \textbf{if } |y| < \epsilon \quad \textbf{return } (x) \\ x_0 \leftarrow x \end{cases}$
return $(error)$

Note that in every iteration we need to evaluate both f and f'.

Newton's method does not, in general, converge globally [show graphically the example of $f(x) = x - \tan^{-1}$.] The following theorem characterizes a class of functions f for which Newton's method converges globally:

Theorem 2.4 *Let $f \in C^2(R)$ be monotonic, convex and assume it has a root. Then the root is unique and Newton's method converges globally.*

Proof: The uniqueness of the root is obvious. It is given that $f''(x) > 0$, and assume, without loss of generality, that $f'(x) > 0$. If $e_n = x_n - \zeta$, then

$$0 = f(\zeta) = f(x_n) - e_n f'(x_n) + \frac{1}{2} e_n^2 f''(x_n - \theta e_n),$$

hence

$$e_{n+1} = e_n - \frac{f(x_n)}{f'(x_n)} = \frac{1}{2} \frac{f''(x_n - \theta e_n)}{f'(x_n)} e_n^2 > 0.$$

Thus, the iterates starting from e_1 are always to the right of the root. On the other hand, since

$$x_{n+1} - x_n = -\frac{f(x_n)}{f'(x_n)} < 0,$$

it follows that (x_n) is a monotonically decreasing sequence bounded below by ζ hence it converges. The limit must coincide with ζ by continuity. ■

Newton's method when f has a double root We now examine the local convergence of Newton's method when ζ is a double root, i.e., $f(\zeta) = f'(\zeta) = 0$. We assume that $f''(\zeta) \neq 0$, so that there exists a neighbourhood of ζ where $f'(x) \neq 0$. As above, we start with the relation

$$e_{n+1} = e_n - \frac{f(x_n)}{f'(x_n)}.$$

Using Taylor's expansion we have

$$0 = f(\zeta) = f(x_n) - e_n f'(x_n) + \frac{1}{2} e_n^2 f''(x_n - \theta e_n),$$

from which we extract $f(x_n)$ and substitute above to get

$$e_{n+1} = \frac{1}{2} e_n^2 \frac{f''(x_n - \theta e_n)}{f'(x_n)}.$$

The problem is that the denominator is not bounded away from zero. We use Taylor's expansion for f':

$$0 = f'(\zeta) = f'(x_n) - e_n f''(x_n - \theta_1 e_n),$$

from which we extract $f'(x_n)$ and finally obtain

$$e_{n+1} = \frac{1}{2} e_n \frac{f''(x_n - \theta e_n)}{f''(x_n - \theta_1 e_n)}.$$

Thus, Newton's method is locally convergent, but the order of convergence reduces to first order. In particular, if the sequence (x_n) converges then

$$\lim_{n\to\infty} \frac{e_{n+1}}{e_n} = \frac{1}{2}.$$

The same result can be derived from an examination of the iteration function Φ. The method is at least second order if $\Phi'(\zeta) = 0$ and at least first order if $|\Phi'(\zeta)| < 1$. Now,

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

In the limit $x \to \zeta$ we have, by our assumptions, $f(x) \sim a(x - \zeta)^2$, to that

$$\lim_{x\to\zeta} \Phi'(x) = \frac{1}{2}.$$

How can second order convergence be restored? The iteration method has to be modified into

$$x_{n+1} = x_n - 2\frac{f(x_n)}{f'(x_n)}.$$

If is easily verified then that

$$\lim_{x\to\zeta} \Phi'(x) = 0.$$

✎ *Exercise 2.11* Your dog chewed your calculator and damaged the division key! To compute reciprocals (i.e., one-over a given number R) without division, we can solve $x = 1/R$ by finding a root of a certain function f with Newton's method. Design such an algorithm (that, of course, does not rely on division).

✎ *Exercise 2.12* Prove that if r is a root of multiplicity k (i.e., $f(r) = f'(r) = \cdots = f^{(k-1)}(r) = 0$ but $f^{(k)}(r) \neq 0$), then the quadratic convergence of Newton's method will be restored by making the following modification to the method:

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

✎ *Exercise 2.13* Similarly to Newton's method (in one variable), derive a method for solving $f(x)$ given the functions $f(x)$, $f'(x)$ and $f''(x)$. What is the rate of convergence?

✎ *Exercise 2.14* What special properties must a function f have if Newton's method applied to f converges cubically?

2.4 The secant method in $\mathbb{R}$

Error analysis The secant method is

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

If we want to analyze this method within our formalism of iterative methods we have to consider an iteration of a couple of numbers. To obtain the local convergence properties of the secant method we can resort to an explicit calculation.

Subtracting ζ from both side we get

$$
\begin{aligned}
e_{n+1} &= e_n - (e_n - e_{n-1}) \frac{f(x_n)}{f(x_n) - f(x_{n-1})} \\
&= -\frac{f(x_{n-1})}{f(x_n) - f(x_{n-1})}\, e_n + \frac{f(x_n)}{f(x_n) - f(x_{n-1})}\, e_{n-1} \\
&= \frac{f(x_n)/e_n - f(x_{n-1})/e_{n-1}}{f(x_n) - f(x_{n-1})}\, e_{n-1} e_n \\
&= \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \frac{f(x_n)/e_n - f(x_{n-1})/e_{n-1}}{x_n - x_{n-1}}\, e_{n-1} e_n
\end{aligned}
$$

The first term can be written as

$$
\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} = \frac{1}{f'(x_{n-1} + \theta(x_n - x_{n-1}))}.
$$

The second term can be written as

$$
\frac{g(x_n) - g(x_{n-1})}{x_n - x_{n-1}} = g'(x_{n-1} + \theta_1(x_n - x_{n-1})),
$$

where

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

Here comes a useful trick. We can write

$$
f(x) - f(\zeta) = \int_\zeta^x f'(s)\, ds = (x - \zeta) \int_0^1 f'(s\zeta + (1-s)x)\, ds,
$$

so that

$$
g(x) = \int_0^1 f'(s\zeta + (1-s)x)\, ds.
$$

We can then differentiate under the integral sign so get

$$
g'(x) = \int_0^1 (1-s)\, f''(s\zeta + (1-s)x)\, ds,
$$

and by the integral mean value theorem, there exists a point ξ between x and ζ such that

$$
g'(x) = f''(\xi) \int_0^1 (1-s)\, ds = \frac{1}{2} f''(\xi).
$$

Combining together, there are two intermediate points so that

$$e_{n+1} = \frac{f''(\xi)}{2\, f'(\xi_1)} e_n e_{n-1},$$

and sufficiently close to the root,

$$e_{n+1} \approx C\, e_{n-1} e_n.$$

What is then the order of convergence? Guess the ansatz $e_n = a\, e_{n-1}^\alpha$, then

$$a\, e_n^\alpha = C\, (a^{-1} e_n)^{1/\alpha} e_n,$$

which implies that $\alpha^2 = \alpha + 1$, or $\alpha = \frac{1}{2}(1 + \sqrt{5}) \approx 1.62$ (the golden ratio). Thus, the order of convergence is super-linear but less that second order. On the other hand, each iteration require only one function evaluation (compared to two for Newton)!

✎ *Exercise 2.15* The method of "false position" for solving $f(x) = 0$ starts with two initial values, x_0 and x_1, chosen such that $f(x_0)$ and $f(x_1)$ have opposite signs. The next guess is then calculated by

$$x_2 = \frac{x_1 f(x_0) - x_0 f(x_1)}{f(x_0) - f(x_1)}.$$

Interpret this method geometrically in terms of the graph of $f(x)$.

2.5 Newton's method in $\mathbb{R}^n$

In the first part of this section we establish the local convergence property of the multi-dimensional Newton method.

Definition 2.3 (Differentiability) Let $f : \mathbb{R}^n \mapsto \mathbb{R}^n$. f is said to be differentiable at the point $x \in \mathbb{R}^n$, if there exists a linear operator on $\mathbb{R}^n$ (i.e., an $n \times n$ matrix) A, such that

$$\lim_{y \to x} \frac{\|f(y) - f(x) - A(y - x)\|}{\|y - x\|} = 0.$$

We call the matrix A the differential of f at the point x and denote it by $Df(x)$.

Comment: While the choice of norm of $\mathbb{R}^n$ is not unique, convergence in one norm implies convergence in all norm for finite dimensional spaces. We will typically use here the Euclidean norm.

Definition 2.4 (Norm of an operator) *Let $(X, \|\cdot\|)$ be a normed linear space and $\mathcal{B}(X)$ be the space of continuous linear transformations on X. Then, $\mathcal{B}(X)$ is a linear space which can be endowed with a norm,*

$$\|A\| = \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}, \qquad A \in \mathcal{B}(X).$$

In particular, every vector norm induces a subordinate matrix norm.

Comments:

① By definition, for all $x \in X$ and $A \in \mathcal{B}(X)$,

$$\|Ax\| \leq \|A\| \|x\|.$$

② We will return to subordinate matrix norms in depth in the next chapter.

Lemma 2.1 *Suppose that $Df(x)$ exists in a convex set K, and there exists a constant $C > 0$, such that*

$$\|Df(x) - Df(y)\| \leq C\|x - y\| \qquad \forall x, y \in K,$$

then

$$\|f(x) - f(y) - Df(y)(x - y)\| \leq \frac{C}{2}\|x - y\|^2 \qquad \forall x, y \in K.$$

Proof: Consider the function

$$\varphi(t) = f(y + t(x - y))$$

defined on $t \in [0, 1]$. Since K is convex then $\varphi(t)$ is differentiable on the unit segment, with

$$\varphi'(t) = Df(y + t(x - y)) \cdot (x - y),$$

and

$$\|\varphi'(t) - \varphi'(0)\| \leq \|Df(y + t(x-y)) - Df(y)\| \|x - y\| \leq Ct\|x-y\|^2. \quad (2.1)$$

On the other hand,

$$\begin{aligned}\Delta \equiv f(x) - f(y) - Df(y)(x-y) &= \varphi(1) - \varphi(0) - \varphi'(0) \\ &= \int_0^1 [\varphi'(t) - \varphi'(0)]\, dt,\end{aligned}$$

from which follows, upon substitution of (2.1),

$$\|\Delta\| \leq \int_0^1 \|\varphi'(t) - \varphi'(0)\|\, dt \leq \frac{C}{2}\|x-y\|^2.$$

■

With this lemma, we are in measure to prove the local quadratic convergence of Newton's method.

Theorem 2.5 *Let $K \subseteq \mathbb{R}^n$ be an open set, and K_0 be a convex set, $\overline{K_0} \subset K$. Suppose that $f : K \mapsto \mathbb{R}^n$ is differentiable in K_0 and continuous in K. Let $x_0 \in K_0$, and assume the existence of positive constants α, β, γ so that*

① $\|Df(x) - Df(y)\| \leq \gamma\|x-y\|$ *in* K_0.
② $[Df(x)]^{-1}$ *exists and* $\|[Df(x)]^{-1}\| \leq \beta$ *in* K_0.
③ $\|Df(x_0)]^{-1}f(x_0)\| \leq \alpha$,

with

$$h \equiv \frac{\alpha\beta\gamma}{2} < 1,$$

and

$$B_r(x_0) \subseteq K_0,$$

where

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

Then,

① *The Newton sequence (x_n) defined by*

$$x_{n+1} = x_n - [Df(x_n)]^{-1}f(x_n)$$

is well defined and contained in $B_r(x_0)$.

② *The sequence (x_n) converges in the closure of $B_r(x_0)$ to a root ζ of f.*

③ *For all n,*

$$\|x_n - \zeta\| \le \alpha \frac{h^{2^n-1}}{1-h^{2^n}},$$

i.e., the convergence is at least quadratic.

Proof: We first show that the sequence remains in $B_r(x_0)$. The third assumption implies

$$\|x_1 - x_0\| = \|Df(x_0)]^{-1} f(x_0)\| \le \alpha < r,$$

i.e., $x_1 \in B_r(x_0)$. Suppose that the sequence remains in $B_r(x_0)$ up to the k-th element. Then x_{k+1} is well defined (by the second assumption), and

$$\begin{aligned}\|x_{k+1} - x_k\| &= \|[Df(x_k)]^{-1} f(x_k)\| \le \beta \|f(x_k)\| \\ &= \beta \|f(x_k) - f(x_{k-1}) - Df(x_{k-1})(x_k - x_{k-1})\|,\end{aligned}$$

where we have used the fact that $f(x_{k-1}) + Df(x_{k-1})(x_k - x_{k-1}) = 0$. Now, by the first assumption and the previous lemma,

$$\|x_{k+1} - x_k\| \le \frac{\beta\gamma}{2} \|x_k - x_{k-1}\|^2.$$

From this, we can show inductively that

$$\|x_{k+1} - x_k\| \le \alpha h^{2^k-1}, \tag{2.2}$$

since it is true for $k = 0$ and if it is true up to k, then

$$\|x_{k+1} - x_k\| \le \frac{\beta\gamma}{2} \alpha^2 (h^{2^{k-1}-1})^2 = \alpha \frac{\alpha\beta\gamma}{2} h^{2^k-2} < \alpha h^{2^k-1}.$$

From this we have

$$\begin{aligned}\|x_{k+1} - x_0\| &\le \|x_{k+1} - x_k\| + \cdots + \|x_1 - x_0\| \\ &\le \alpha(1 + h + h^3 + \cdot + h^{2^k-1}) < \frac{\alpha}{1-h} = r,\end{aligned}$$

i.e., $x_{k+1} \in S_r(x_0)$, hence the entire sequence.

Inequality (2.2) implies also that (x_n) is a Cauchy sequence, for

$$\begin{aligned}\|x_{n+1} - x_m\| &\le \|x_{n+1} - x_n\| + \cdots + \|x_{m+1} - x_m\| \\ &\le \alpha \left(h^{2^m-1} + \cdots + h^{2^n-1}\right) \\ &< \alpha h^{2^m-1}\left(1 + h^{2^m} + (h^{2^m})^3 + \dots\right) < \alpha \frac{h^{2^m-1}}{1-h^{2^m}}.\end{aligned}$$

which tends to zero as $m, n \to \infty$. Thus the sequence (x_n) converges to a limit $\zeta \in \overline{S_r(x_0)}$. As a side results we obtain that

$$\|\zeta - x_m\| \le \alpha \frac{h^{2^m-1}}{1-h^{2^m}}.$$

It remains to show that ζ is indeed a root of f. The first condition implies the continuity of the differential of f, so that taking limits:

$$\zeta = \zeta - [Df(\zeta)]^{-1} f(\zeta),$$

and since by assumption, Df is invertible, it follows that $f(\zeta) = 0$. ■

✎ *Exercise 2.16 (Computer exercise)* Use Newton's method to solve the system of equations

$$\begin{aligned} xy^2 + x^2y + x^4 &= 3 \\ x^3y^5 - 2x^5y - x^2 &= -2. \end{aligned}$$

Start with various initial values and try to characterize the "basin of convergence" (the set of initial conditions for which the iterations converge).

Now, Matlab has a built-in root finder `fsolve()`. Try to solve the same problem using this functions, and evaluate whether it performs better or worse than your own program in terms of both speed and robustness.

✎ *Exercise 2.17* Go the the following site and enjoy the nice pictures:

`http://aleph0.clarku.edu/~djoyce/newton/newton.html`

(Read the explanations, of course....)

2.6 A modified Newton's method in $\mathbb{R}^n$

Newton's method is of the form

$$x_{k+1} = x_k - d_k,$$

where

$$d_k = [Df(x_k)]^{-1} f(x_k).$$

When this method converges, it does so quadratically, however, the convergence is only guaranteed locally. A modification to Newton's method, which converges under much wider conditions is of the following form:

$$x_{k+1} = x_k - \lambda_k d_k,$$

where the coefficients λ_k are chosen such that the sequence $(h(x_k))$, where

$$h(x) = f^T(x) f(x) = \|f(x)\|^2,$$

is strictly monotonically decreasing (here $\|\cdot\|$ stands for the Euclidean norm in $\mathbb{R}^n$). Clearly, $h(x_k) \geq 0$, and if the sequence (x_k) converges to a point ζ, where $h(\zeta) = 0$ (i.e., a global minimum of $h(x)$), then $f(\zeta) = 0$. *The modified Newton method aims to minimize $h(x)$ rather than finding a root of $f(x)$.*

Definition 2.5 *Let $h : \mathbb{R}^n \mapsto \mathbb{R}$ and $\|\cdot\|$ be the Euclidean norm in $\mathbb{R}^n$. For $0 < \gamma \leq 1$ we define*

$$D(\gamma, x) = \left\{ s \in \mathbb{R}^n : \ \|s\| = 1, \ \frac{Dh(x)}{\|Dh(x)\|} \cdot s \geq \gamma \right\},$$

which is the set of all directions s which form with the gradient of h a not-too-accute angle.

Lemma 2.2 *Let $h : \mathbb{R}^n \mapsto \mathbb{R}$ be in C^1 in a neighbourhood $V(\zeta)$ of a point ζ. Suppose that $Dh(\zeta) \neq 0$ and let $0 < \gamma \leq 1$. Then there exist a neighbourhood $U(\zeta) \subseteq V(\zeta)$ and a number $\lambda > 0$, such that*

$$h(x - \mu s) \leq h(x) - \frac{\mu\gamma}{4} \|Dh(\zeta)\|$$

for all $x \in U(\zeta)$, $s \in D(\gamma, x)$, and $0 \leq \mu \leq \lambda$.

Proof: Consider first the set

$$U_1(\zeta) = \left\{ x \in V(\zeta) : \ \|Dh(x) - Dh(\zeta)\| \leq \frac{\gamma}{4} \|Dh(\zeta)\| \right\},$$

which by the continuity of Dh and the non-vanishing of $Dh(\zeta)$ is a non-empty set and a neighbourhood of ζ. Let also

$$U_2(\zeta) = \left\{ x \in V(\zeta) : \ D(\gamma, x) \subseteq D(\tfrac{\gamma}{2}, \zeta) \right\},$$

which again is a non-empty neighbourhood of ζ. Indeed, it consists of all $x \in V(\zeta)$ for which

$$\left\{ s : \ \frac{Dh(x)}{\|Dh(x)\|} \cdot s \geq \gamma \right\} \subseteq \left\{ s : \ \frac{Dh(\zeta)}{\|Dh(\zeta)\|} \cdot s \geq \frac{\gamma}{2} \right\}.$$

Choose now a λ such that

$$\overline{B_{2\lambda}(\zeta)} \subseteq U_1(\zeta) \cap U_2(\zeta),$$

and finally set

$$U(\zeta) = \overline{B_\lambda(\zeta)}.$$

Now, for all $x \in U(\zeta)$, $s \in D(\gamma, x)$ and $0 \leq \mu \leq \lambda$, there exists a $\theta \in (0, 1)$ such that

$$\begin{aligned} h(x) - h(x - \mu s) &= \mu \, Dh(x - \theta\mu s) \cdot s \\ &= \mu \left\{ (Dh(x - \theta\mu s) - Dh(\zeta)) \cdot s + Dh(\zeta) \cdot s \right\}. \end{aligned}$$

Now $x \in \overline{B_\lambda(\zeta)}$ and $\mu \leq \lambda$ implies that

$$x - \mu s, x - \theta\mu s \in \overline{B_{2\lambda}(\zeta)} \subseteq U_1(\zeta) \cap U_2(\zeta),$$

and by the membership in $U_1(\zeta)$,

$$(Dh(x - \theta\mu s) - Dh(\zeta)) \cdot s \geq -\|Dh(x - \theta\mu s) - Dh(\zeta)\| \geq -\frac{\gamma}{4} \|Dh(\zeta)\|,$$

whereas by the membership in $U_2(\zeta)$, $s \in D(\frac{\gamma}{2}, \zeta)$, hence

$$Dh(\zeta) \cdot s \geq \frac{\gamma}{2} \|Dh(\zeta)\|,$$

and combining the two,

$$h(x) - h(x - \mu s) \geq -\mu \frac{\gamma}{4} \|Dh(\zeta)\| + \mu \frac{\gamma}{2} \|Dh(\zeta)\| = \frac{\mu\gamma}{4} \|Dh(\zeta)\|.$$

This completes the proof. ■

Minimization algorithm Next, we describe an algorithm for the minimization of a function $h(x)$ via the construction of a sequence (x_k).

① Choose sequences (γ_k), (σ_k), satisfying the constraints

$$\sup_k \gamma_k \le 1, \quad \gamma \equiv \inf_k \gamma_k > 0, \quad \sigma \equiv \inf_k \sigma_k > 0,$$

as well as a starting point x_0.

② For every k, choose a ***search direction*** $s_k \in D(\gamma_k, x_k)$ and set

$$x_{k+1} = x_k - \lambda_k s_k,$$

where $\lambda_k \in [0, \sigma_k \|Dh(x_k)\|]$ is chosen such to minimize $h(x_k - \lambda_k s_k)$.

Theorem 2.6 *Let $h : \mathbb{R}^n \mapsto \mathbb{R}$ and $x_0 \in \mathbb{R}^n$ be such that*

① *The set $K = \{x : \; h(x) \le h(x_0)\}$ is compact.*

② *$h \in C^1$ in an open set containing K.*

Then,

① *The sequence (x_k) is in K and has at least one accumulation point ζ.*

② *Each accumulation point ζ is a critical point of h, $Dh(\zeta) = 0$.*

Proof: Since, by construction, the sequence $(h(x_k))$ is monotonically decreasing then the $\{h(x_k)\}$ are all in K. Since K is compact, then the set $\{x_k\}$ has at least one accumulation point ζ.

Without loss of generality we can assume that $x_k \to \zeta$, otherwise we consider a converging sub-sequence. Assume that ζ is not a critical point, $Dh(\zeta) \ne 0$. From the previous lemma, we know that there exist a neighbourhood $U(\zeta)$ and a number $\lambda > 0$, such that

$$h(x - \mu s) \le h(x) - \frac{\mu\gamma}{4} \|Dh(\zeta)\| \tag{2.3}$$

for all $x \in U(\zeta)$, $s \in D(\gamma, x)$, and $0 \le \mu \le \lambda$. Since $x_k \to \zeta$ and because Dh is continuous, it follows that for sufficiently large k,

① $x_k \in U(\zeta)$.

② $\|Dh(x_k)\| \geq \frac{1}{2}\|Dh(\zeta)\|$.

Set now

$$\Lambda = \min\left(\lambda, \frac{1}{2}\sigma\|Dh(\zeta)\|\right), \qquad \epsilon = \Lambda\frac{\gamma}{4}\|Dh(\zeta)\| > 0.$$

Since $\sigma_k \geq \sigma$ it follows that for sufficiently large k,

$$[0, \Lambda] \subseteq [0, \sigma_k\frac{1}{2}\|Dh(\zeta)\|] \subseteq [0, \sigma_k\|Dh(x_k)\|],$$

the latter being the set containing λ_k in the minimization algorithm. Thus, by the definition of x_{k+1},

$$h(x_{k+1}) \leq h(x_k - \mu s_k),$$

for every $0 \leq \mu \leq \Lambda$. Since $\Lambda \leq \lambda$, $x_k \in U(\zeta)$, and $s_k \in D(\gamma_k, x_k) \subseteq D(\gamma, x_k)$, it follows from (2.3) that

$$h(x_{k+1}) \leq h(x_k) - \frac{\Lambda\gamma}{4}\|Dh(\zeta)\| = h(x_h) - \epsilon.$$

This means that $h(x_k) \to -\infty$ which contradicts its lower-boundedness by $h(\zeta)$. ■

The modified Newton algorithm The modified Newton algorithm works as follows: at each step

$$x_{k+1} = x_k - \lambda_k d_k, \qquad d_k = [Df(x_k)]^{-1}f(x_k),$$

where $\lambda_k \in (0, 1]$ is chosen such to minimize $h(x_k - \lambda_k d_k)$, where $h(x) = f^T(x)f(x)$.

Theorem 2.7 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$ satisfy the following properties:*

① *The set $K = \{x : h(x) \leq h(x_0)\}$ with $h(x) = f^T(x)f(x)$ is compact.*
② *$f \in C^1$ in some open set containing K.*
③ *$[Df(x)]^{-1}$ exists in K.*

Then, the sequence x_k defined by the modified Newton method is well-defined, and

① *The sequence (x_k) is in K and has at least one accumulation point.*
② *Every such accumulation point is a zero of f.*