1. Fixed point methods for solving nonlinear equations

We address the problem of solving an equation of the form

\[ r(x) = 0, \]

where \( F(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is a vector-function. Eq. (1) can be written as

\[
\begin{align*}
    r_1(x_1, x_2, \ldots, x_n) &= 0 \\
    r_2(x_1, x_2, \ldots, x_n) &= 0 \\
    &\vdots \\
    r_n(x_1, x_2, \ldots, x_n) &= 0.
\end{align*}
\]

If the components \( f_j \) of \( F \) are differentiable, we define the **Jacobian matrix**

\[
J(x) = \begin{bmatrix}
    \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\
    \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_2}{\partial x_n} \\
    \vdots & \vdots & \ddots & \vdots \\
    \frac{\partial r_n}{\partial x_1} & \frac{\partial r_n}{\partial x_2} & \cdots & \frac{\partial r_n}{\partial x_n}
\end{bmatrix}.
\]

Throughout the rest of the semester we will use the following theorem of calculus as working tool for proving convergence theorems.

**Theorem 1.** Let \( r : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be differentiable in an open domain \( \Omega \subset \mathbb{R}^n \) and let \( x^* \in \Omega \). Then for all \( x \in \Omega \) sufficiently close to \( x^* \)

\[
r(x) - r(x^*) = \int_0^1 J(x^* + t(x - x^*))(x - x^*)dt.
\]

For warming-up we start with methods for solving a single equation of one variable.

1.1. Some classic methods.

1.1.1. **Bisection method.** Suppose \( f(x) \) is continuous on \([a, b]\) and \( f(a)f(b) < 0 \). Then there exists at least one zero of \( f \) in \((a, b)\). The bisection method finds the root of

\[ f(x) = 0 \]

as follows.

Input: \( a, b, \) tolerance \( \epsilon \)

Output: \( x^* \), a root of \( f(x) \) in \((a, b)\)

Do

Set \( c = (a + b)/2 \)
If \( f(c) = 0 \) return \( c \)
Else if \( f(a)f(c) < 0 \) set \( b = c \)
Else set \( a = c \)
While \(|b - a| \geq \epsilon|\)
Set \( x^* = c \)
This algorithm is robust and globally convergent. It does not require more smoothness from \( f \) than continuity. Its shortcoming is that it is slow. The error bound, \(|x^* - c|\) reduces just by the factor of 2 with each iteration.

1.1.2. **Fixed point methods for a single equation.** The basic idea of the fixed point methods consists in finding an iteration function \( T(x) \) such that (i) the zero \( x^* \) of \( f(x) \) satisfies
\[
T(x^*) = x^*,
\]
and (ii) \( T(x) \) generates successive approximations to the solution
\[
x_{n+1} = T(x_n)
\]
starting from the provided initial approximation \( x_0 \).

**Theorem 2.** Let \( T(x) \) be a continuous and differentiable function in \([a, b]\) such that
\[
T([a, b]) \subset [a, b] \quad \text{and} \quad |T'(x)| \leq M < 1, \quad x \in [a, b].
\]
Then there exists a unique \( x^* \in [a, b] \) such that \( T(x^*) = x^* \) and for any \( x_0 \in [a, b] \) the sequence \( x_{n+1} = T(x_n) \), \( n = 0, 1, 2, \ldots \), converges to \( x^* \). The error after the \( n \)-th iteration is bounded by
\[
|x_n - x^*| \leq \frac{M^n}{1 - M}|x_1 - x_0|.
\]

Prior to proving this theorem let us see what sequences \( x_{n+1} = T(x_n) \) are generated in the case of linear map \( T \). For different cases are illustrated in Fig. 1. We see that if \( |T'| < 1 \) the sequence converges to the fixed point, while if \( |T'| > 1 \) is diverges. If \( T' > 0 \), the sequence is monotone, while if \( T' < 0 \) the sequence oscillates.

**Proof.**
(1) **Uniqueness** Suppose there are two fixed points: \( y = T(y) \) and \( z = T(z) \).
Then applying the mean value theorem we obtain
\[
|y - z| = |T(y) - T(z)| = |T'(\zeta)||y - z| \leq M|y - z| < |y - z|,
\]
a contradiction. Hence there is at most one fixed point \( \zeta \in [a, b] \).

(2) **Existence** Let us show that the sequence \( x_{n+1} = T(x_n) \) is Cauchy. We have
\[
|x_{n+k} - x_n| = |T(x_{n+k-1}) - T(x_{n-1})| \leq M|x_{n+k-1} - x_{n-1}| \leq \ldots \leq M^n|x_k - x_0| \leq M^n|b - a|.
\]
Hence this difference can be made arbitrarily small as soon as \( n \) is large enough. Hence this sequence is Cauchy and hence it converges.

(3) **Error bound** Let \( x^* \) be the fixed point,
\[
|x_n - x^*| = |T(x_{n-1} - T(x_{ast})| \leq M|x_{n-1} - x^*| \leq \ldots \leq M^n|x_0 - x^*|.
\]
On the other hand,
\[
|x_0 - x^*| \leq |x_0 - x_1| + |x_1 - x^*| \leq |x_0 - x_1| + M|x_0 - x^*|.
\]
Hence,
\[
(1 - M)|x_0 - x^*| \leq |x_0 - x_1|.
\]
Figure 1. Examples of sequences $x_{n+1} = T(x_n)$.

Therefore we get the following error bound:

$$|x_n - x^*| \leq M^n |x_0 - x^*| \leq \frac{M^n}{1 - M} |x_0 - x_1|.$$

1.1.3. Newton-Raphson method. The idea of the Newton-Raphson method for solving $f(x) = 0$ is to plot a tangent line to the graph of $y = f(x)$ through the point $(x_n, f(x_n))$ where $x_n$ is the current approximation and set $x_{n+1}$ to its intersection with the $x$ axis. The equation
of the tangent line is $y = f(x_n) + f'(x_n)(x - x_n)$. Hence

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

Therefore the iteration function is

$$T(x) := x - \frac{f(x)}{f'(x)}.$$  

As is well-know, this method is not globally convergent or its behavior can be hard to predict if the initial guess is not close enough to the solution. However, it converges rapidly for a close enough initial approximation. To quantify this, we consider

$$T'(x) = 1 - \frac{f'(x)}{f'(x)} + \frac{f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{(f'(x))^2}.$$  

Therefore, if $f'(x^*) \neq 0$, $T'(x) \to 0$ as $x \to x^*$. Hence $|T'(x)| < 1$ in some neighborhood of $x^*$ which guarantees that the Newton-Raphson method converges.

1.2. Rates of convergence. The performance of any zero-finding or optimization algorithm is characterized by (i) its ability to find a solution (global convergence or local convergence) and by (ii) rate of convergence. The conceptually simplest definition of the rate of convergence is the following

**Definition 1.** Let $\{x_n\}_{n=0}^{\infty}$ be a sequence which converges to $x^*$ and such that $x_n \neq x^*$ for $n \in \mathbb{N}$. We will say that the sequence converges with $Q$-order $p \geq 1$ if there exists a constant $C > 0$ such that

$$\lim_{n \to \infty} \frac{\|e_{n+1}\|}{\|e_n\|^p} = C,$$

where $e_n = x_n - x^*$, $n \in \mathbb{N}$. $C$ is called the asymptotic error constant.

$Q$ stands for 'quotient'.

Now we determine the order of convergence for a fixed point method. We define

$$e_n := x_n - x^*$$

and consider the Taylor expansion

$$x_{n+1} = T(x_n) = T(x^* + e_n) = T(x^*) + T'(x^*) e_n + \frac{1}{2} T''(x^*) e_n^2 + \ldots.$$  

If $T^{(p)}(x^*)$ is the first non-vanishing derivative of $T$ at $x^*$ we have

$$e_{n+1} = \frac{T^{(p)}(x^*)}{p!} e_n^p.$$  

Hence

$$\lim_{n \to \infty} |e_{n+1}| e_n = \frac{T^{(p)}(x^*)}{p!}. $$

In this case we say that the method has $Q$-order of convergence $p$. $Q$ stands for quotient.
For the Newton-Raphson method, $T'(x^*) = 0$.

$$T''(x^*) = \frac{f'(x^*)f''(x^*)}{(f'(x^*))^2} = \frac{f''(x^*)}{f'(x^*)}.$$  

Hence, if $f''(x^*) \neq 0$, the Newton-Raphson method converges $Q$-quadratically, and the asymptotic error constant is

$$C = \frac{f''(x^*)}{f'(x^*)}.$$

**Exercise** Show that if $f(x^*) = 0$ and $f'(x^*) = 0$ then the Newton-Raphson method converges $Q$-linearly and the asymptotic error constant is $\frac{1}{2}$.

Definition 1 is somewhat restrictive. What if our method makes a more significant progress every second step or each time when we do some kind of reinitialization? To characterize these cases, Definition 1 has been generalized to

**Definition 2.** Let $\{x_n\}_{n=0}^{\infty}$ be a sequence which converges to $x^*$ and such that $x_n \neq x^*$ for $n \in \mathbb{N}$. We will say that the sequence converges with $Q$-order $p \geq 1$ if there exists a constant $C > 0$ such that

$$\|x_{n+1} - x^*\| \leq C\|x_n - x^*\|^p$$

for sufficiently large $n$. In particular, $x_n \to x^*$ $Q$-linearly if

$$\|x_{n+1} - x^*\| \leq C\|x_n - x^*\|$$

for sufficiently large $n$.

We say that $x_n \to x^*$ $Q$-superlinearly

$$\lim_{n \to \infty} \frac{\|x_{n+1} - x^*\|}{\|x_n - x^*\|} = 0.$$

We say that $x_n \to x^*$ $Q$-quadratically if there exists a constant $C > 0$ such that

$$\|x_{n+1} - x^*\| \leq C\|x_n - x^*\|^2$$

for sufficiently large $n$.

The prefix $Q$ is often omitted. It has been introduced to distinguish the convergence described in Definition 2 from a weaker form of convergence called $R$-convergence. $R$ stands for ‘root’. This kind of rate is concerned with overall rate of decrease of the error rather than the decrease over a single step of the algorithm.

**Definition 3.** We say that the sequence converges $R$-linearly to $x^*$ if there exists a sequence of nonnegative numbers $\{\nu_n\}_{n=0}^{\infty}$ such that

$$\|x_n - x^*\| \leq \nu_n, \quad n \in \mathbb{N}$$

and $\nu_n \to 0$ $Q$-linearly.

Similarly we can define $R$-superlinear convergence and $R$-convergence of order $r$. This definition will be useful for the case if the convergence is non-monotone, or from time to time the method fails to make progress over a step.
Example The sequence $x_n = 2^{-n}$, $n$ is even, $x_n = 2^{-n-1}$, $n$ is odd, converges $R$-linearly to zero. This sequence is

$$1, \frac{1}{4}, \frac{1}{4}, \frac{1}{16}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \ldots$$

As you see, every second step we fail to make progress toward zero.

1.3. Fixed point iteration in $\mathbb{R}^n$.

Definition 4. Let $\Omega \subset \mathbb{R}^n$ and let $T : \Omega \to \mathbb{R}^n$. $T$ is Lipschitz-continuous on $\Omega$ with constant $M$ if

$$\|T(x) - T(y)\| \leq M\|x - y\|.$$

$T$ is a contraction if $M < 1$.

The standard result for the fixed point iteration is the Contraction Mapping Theorem.

Theorem 3. Let $\Omega$ be a closed subset of $\mathbb{R}^n$ and let $T$ be a contraction with constant $M < 1$ such that $T(\Omega) \subset \Omega$. Then there exists a unique fixed point $x^* \in \Omega$. Then the sequence of iterates $x_{n+1} = T(x_n)$ converges $Q$-linearly with asymptotic error constant $M$ to $x^*$ for all initial iterates $x_0$.

The proof of this theorem repeats the proof to Theorem 2.

References