
Numerical methods for solving systems of linear and nonlinear equations

Keywords: Zero, fixed point, self mapping, Lipschitz-continuous, Lipschitz constant, contraction, Banach's fixed point theorem, Jacobi matrix, Newton method

Contents

1	The general problem	2
1.1	Zeros and fixed points	2
1.2	Systems of linear equations and fixed point iteration	3
2	Banach's fixed point theorem	4
3	The Newton method	7

1 The general problem

1.1 Zeros and fixed points

In this part of the course we will talk about numerical methods for the solution of systems of linear and nonlinear equations. Let

$$\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longmapsto \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{pmatrix}$$

be a map and we are looking for a zero or root $\mathbf{a} = (a_1, \dots, a_n)^T \in \mathbb{R}^n$, a point such that

$$\begin{aligned} f_1(\mathbf{a}) &= f_1(a_1, a_2, \dots, a_n) = 0 \\ f_2(\mathbf{a}) &= f_2(a_1, a_2, \dots, a_n) = 0 \\ &\vdots \\ f_n(\mathbf{a}) &= f_n(a_1, a_2, \dots, a_n) = 0 \end{aligned}$$

or shortly

$$\mathbf{f}(\mathbf{a}) = \mathbf{0}.$$

Many linear and nonlinear equations are naturally formulated as fixed point problem. Let $\mathbf{F} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a map and we are looking for a fixed point $\mathbf{a} = (a_1, \dots, a_n)^T \in \mathbb{R}^n$, a point such that

$$\begin{aligned} F_1(\mathbf{a}) &= F_1(a_1, a_2, \dots, a_n) = a_1 \\ F_2(\mathbf{a}) &= F_2(a_1, a_2, \dots, a_n) = a_2 \\ &\vdots \\ F_n(\mathbf{a}) &= F_n(a_1, a_2, \dots, a_n) = a_n \end{aligned}$$

or shortly

$$\mathbf{F}(\mathbf{a}) = \mathbf{a}.$$

Of course, there are many possibilities to transform a zero problem into an equivalent fixed-point problem and conversely.

The associated fixed point iteration is given by

$$\mathbf{x}^{(n+1)} = \mathbf{F}(\mathbf{x}^{(n)})$$

with an initial value $\mathbf{x}^{(0)}$. We analyze these processes and will see, that the iterates frequently converge to the fixed point \mathbf{a} .

1.2 Systems of linear equations and fixed point iteration

Example 1.1 We are often interested in solving systems of linear equations $\mathbf{Ax} = \mathbf{b}$ or $\mathbf{f}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b} = \mathbf{0}$ (zero form). If the matrix \mathbf{A} is invertible, then the unique solution of the problem can be written as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. But the computation of the inverse is difficult, time-consuming and complex.

There are many possibilities to transform this zero problem into a fixed point problem. Then we could try to find an approximation of the solution by fixed point iteration. Here are two important ways.

Richardson iteration

1. Write \mathbf{A} as $\mathbf{A} = \mathbf{A} + \mathbf{I} - \mathbf{I}$.

2.

$$\begin{aligned}\mathbf{Ax} = \mathbf{b} &\iff (\mathbf{A} + \mathbf{I} - \mathbf{I})\mathbf{x} = \mathbf{b} \\ &\iff \mathbf{Ax} + \mathbf{Ix} - \mathbf{Ix} = \mathbf{b} \\ &\iff \mathbf{x} = \mathbf{b} + (\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{F}(\mathbf{x}).\end{aligned}$$

3.

$$\mathbf{x}^{(n+1)} = \mathbf{b} + (\mathbf{I} - \mathbf{A})\mathbf{x}^{(n)}$$

Jacobi iteration

1. Write \mathbf{A} as the sum of three matrices \mathbf{L} , \mathbf{D} und \mathbf{R} :

$$A = \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & 0 \end{pmatrix}}{=: \mathbf{L}} + \underbrace{\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}}{=: \mathbf{D}} + \underbrace{\begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}}{=: \mathbf{R}}$$

This decomposition is called $\mathbf{L} - \mathbf{D} - \mathbf{R}$ -decomposition of \mathbf{A} .

2.

$$\begin{aligned}\mathbf{Ax} = \mathbf{b} &\iff (\mathbf{L} + \mathbf{D} + \mathbf{R})\mathbf{x} = \mathbf{b} \\ &\iff \mathbf{Lx} + \mathbf{Dx} + \mathbf{Rx} = \mathbf{b} \\ &\iff \mathbf{Dx} = -(\mathbf{L} + \mathbf{R})\mathbf{x} + \mathbf{b} \\ &\iff \mathbf{x} = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R})\mathbf{x} + \mathbf{D}^{-1}\mathbf{b} = \mathbf{F}(\mathbf{x}).\end{aligned}$$

if \mathbf{D} is regular.

3. Iteration:

$$\mathbf{x}^{(n+1)} = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{R})\mathbf{x}^{(n)} + \mathbf{D}^{-1}\mathbf{b}$$

2 Banach's fixed point theorem

Before discussing convergence of fixed-point iteration we need two definitions.

Definition 2.1 Let $D \subset \mathbb{R}^n$, $\|\cdot\|$ any norm on \mathbb{R}^n and $\mathbf{F} : D \rightarrow \mathbb{R}^n$. The map \mathbf{F} is called Lipschitz-continuous on D with Lipschitz constant γ if

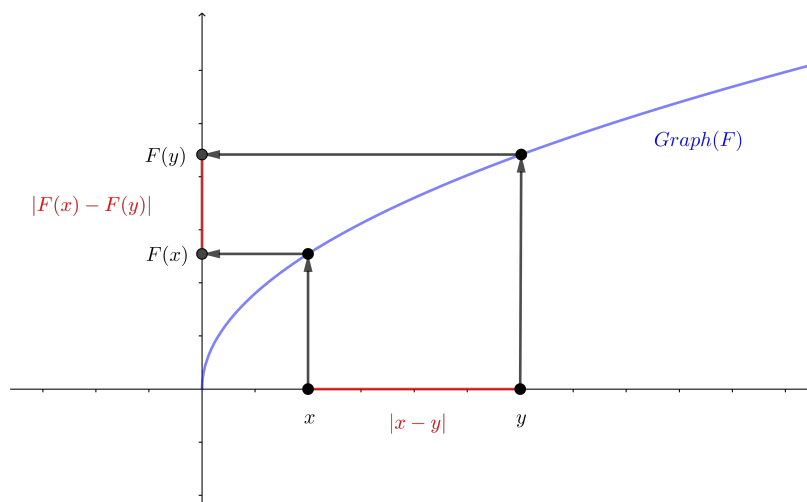
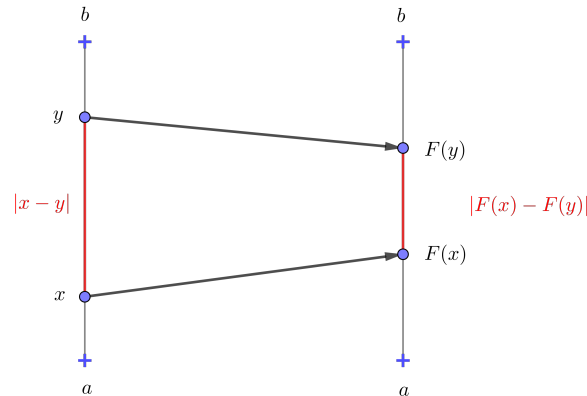
$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}, \mathbf{y} \in D$.

\mathbf{F} is called a contraction mapping on D if \mathbf{F} is Lipschitz continuous on D with Lipschitz constant $\gamma < 1$. This means, that

$$\frac{\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\|}{\|\mathbf{x} - \mathbf{y}\|} \leq \gamma < 1$$

for all $\mathbf{x}, \mathbf{y} \in D$, or that the distance between the images of two points is always smaller than the distance between the two points.



The standard result of fixed point iteration is the following.

Theorem 2.1 (Banach's fixed point theorem) *Let D be a closed subset of \mathbb{R}^n and let $\mathbf{F} : D \rightarrow \mathbb{R}^n$ be a contraction mapping on D with Lipschitz constant $\gamma (< 1)$ such that $\mathbf{F}(\mathbf{x}) \in D$ for all $\mathbf{x} \in D$ (\mathbf{F} is a self mapping).*

Then there is a unique fixed point $\mathbf{a} \in D$ of \mathbf{F} and the iteration $\mathbf{x}^{(n+1)} = \mathbf{F}(\mathbf{x}^{(n)})$ converges to \mathbf{a} for all initial values $\mathbf{x}^{(0)} \in D$.

Proof:

1. Let $\mathbf{x}^{(0)} \in D$ and $\mathbf{x}^{(n+1)} = \mathbf{F}(\mathbf{x}^{(n)})$. Note that all $\mathbf{x}^{(n)} \in D$, because \mathbf{F} is a self mapping. We have

$$\begin{aligned}
 \|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\| &= \|\mathbf{F}(\mathbf{x}^{(j)}) - \mathbf{F}(\mathbf{x}^{(j-1)})\| \\
 &\leq \gamma \|\mathbf{x}^{(j)} - \mathbf{x}^{(j-1)}\| \\
 &= \gamma \|\mathbf{F}(\mathbf{x}^{(j-1)}) - \mathbf{F}(\mathbf{x}^{(j-2)})\| \\
 &\leq \gamma^2 \|\mathbf{x}^{(j-1)} - \mathbf{x}^{(j-2)}\| \\
 &\dots \\
 &\leq \gamma^n \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|
 \end{aligned}$$

$$\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\| \leq \gamma^n \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|$$

2. For the distance between $\mathbf{x}^{(m)}$ and $\mathbf{x}^{(k)}$ ($m > k$) we get:

$$\begin{aligned}
 \|\mathbf{x}^{(m)} - \mathbf{x}^{(k)}\| &= \|(\mathbf{x}^{(m)} - \mathbf{x}^{(m-1)}) + (\mathbf{x}^{(m-1)} - \mathbf{x}^{(m-2)}) + \dots + (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})\| \\
 &\leq \|(\mathbf{x}^{(m)} - \mathbf{x}^{(m-1)})\| + \|(\mathbf{x}^{(m-1)} - \mathbf{x}^{(m-2)})\| + \dots + \|(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})\| \\
 &\leq (\gamma^m + \gamma^{m-1} + \dots + \gamma^k) \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| \\
 &\leq \sum_{j=k}^{\infty} \gamma^j \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = \gamma^k \sum_{j=0}^{\infty} \gamma^j \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = \underbrace{\frac{\gamma^k}{1-\gamma}}_{\text{small}} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|
 \end{aligned}$$

$$\|\mathbf{x}^{(m)} - \mathbf{x}^{(k)}\| \leq \frac{\gamma^k}{1-\gamma} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|$$

We see

$$\lim_{k \rightarrow \infty} \|\mathbf{x}^{(m)} - \mathbf{x}^{(k)}\| = \frac{\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|}{1 - \gamma} \lim_{k \rightarrow \infty} \gamma^k = 0$$

and the sequence of iterations $\{\mathbf{x}^{(n)}\}$ is a Cauchy sequence and has a limit $\mathbf{a} = \lim_{n \rightarrow \infty} \mathbf{x}^{(n)}$.

3. The limit $\mathbf{a} = \lim_{n \rightarrow \infty} \mathbf{x}^{(n)}$ is a fixed point of \mathbf{F} because

$$\begin{aligned} 0 &\leq \|\mathbf{a} - \mathbf{F}(\mathbf{a})\| \\ &\leq \|\mathbf{a} - \mathbf{x}^{(k)}\| + \|\mathbf{x}^{(k)} - \mathbf{F}(\mathbf{a})\| \\ &= \|\mathbf{a} - \mathbf{x}^{(k)}\| + \|\mathbf{F}(\mathbf{x}^{(k-1)}) - \mathbf{F}(\mathbf{a})\| \\ &\leq \|\mathbf{a} - \mathbf{x}^{(k)}\| + \gamma \|\mathbf{x}^{(k-1)} - \mathbf{a}\| \longrightarrow 0 \quad \text{if } k \rightarrow \infty \end{aligned}$$

4. If \mathbf{F} has two fixed points $\mathbf{a}, \mathbf{b} \in D$, then

$$\|\mathbf{a} - \mathbf{b}\| = \|\mathbf{F}(\mathbf{a}) - \mathbf{F}(\mathbf{b})\| \leq \gamma \|\mathbf{a} - \mathbf{b}\| < \|\mathbf{a} - \mathbf{b}\|$$

and we see that $\|\mathbf{a} - \mathbf{b}\| = 0$ and $\mathbf{a} = \mathbf{b}$. Hence the fixed point is unique.

□

How can we decide if a map is a contraction? In the case of differentiable mappings we have the following result.

Theorem 2.2 *Let $D \subset \mathbb{R}^n$ be a compact set, $\|\cdot\|$ any matrix norm on $\mathbb{R}^{n \times n}$, $\mathbf{F} : D \rightarrow \mathbb{R}^n$ continuously differentiable and*

$$\lambda = \max_{\mathbf{x} \in D} \|\mathbf{DF}(\mathbf{x})\|.$$

If $\lambda < 1$, then \mathbf{F} is a contraction with Lipschitz constant λ .

3 The Newton method

Let

$$\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longmapsto \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{pmatrix}$$

be a map and we are looking for a root \mathbf{a} with $\mathbf{f}(\mathbf{a}) = \mathbf{0}$.

If all the functions f_i are differentiable, then we can approximate f_i by the tangent hyperplane in an arbitrary point $\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_n^{(n)}) \in \mathbb{R}^n$:

$$f_i(\mathbf{x}) \approx f_i(\mathbf{x}^{(n)}) + \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(\mathbf{x}^{(n)}) \cdot (x_j - x_j^{(n)})$$

By collecting all partial derivatives of all the functions f_i and rearranging all these expressions in the Jacobi matrix

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

we get

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}^{(n)}) + D\mathbf{f}(\mathbf{x}^{(n)}) \cdot (\mathbf{x} - \mathbf{x}^{(n)}).$$

Now we are looking for \mathbf{x} such that

$$\mathbf{f}(\mathbf{x}^{(n)}) + D\mathbf{f}(\mathbf{x}^{(n)}) \cdot (\mathbf{x} - \mathbf{x}^{(n)}) = \mathbf{0}$$

and we may hope that the point $\mathbf{x} =: \mathbf{x}^{(n+1)}$ is a better approximation of the exact zero \mathbf{a} of \mathbf{f} than $\mathbf{x}^{(n)}$.

If the matrix $D\mathbf{f}(\mathbf{x}^{(n)})$ is regular (invertible) then we can solve this equation for $\mathbf{x}^{(n+1)}$ and get

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - [D\mathbf{f}(\mathbf{x}^{(n)})]^{-1} \cdot \mathbf{f}(\mathbf{x}^{(n)}).$$

Theorem 3.1 *Let \mathbf{a} be a zero of \mathbf{f} . The Newton method*

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - [D\mathbf{f}(\mathbf{x}^{(n)})]^{-1} \cdot \mathbf{f}(\mathbf{x}^{(n)}) \quad (\star)$$

converges for all initial values $\mathbf{x}^{(0)}$, **sufficiently near to \mathbf{a}** , if

- $D\mathbf{f}(\mathbf{a})$ is regular and
- all f_i are three times partially differentiable.

But if you like to perform the Newton method then you should avoid the difficult computation of the inverse Jacobi matrix. Rewrite the equation (\star) as

$$D\mathbf{f}(\mathbf{x}^{(n)}) \cdot \underbrace{(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)})}_{=: \boldsymbol{\delta}^{(n)}} = -\mathbf{f}(\mathbf{x}^{(n)})$$

with

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \boldsymbol{\delta}^{(n)}.$$

Concrete implementation of the Newton-method

Let $\mathbf{x}^{(0)}$ be an initial value near to the zero \mathbf{a} of \mathbf{f} . For $n = 0, 1, 2, \dots$ do:

1. Compute $\boldsymbol{\delta}^{(n)}$ as the solution of the following system of linear equations

$$D\mathbf{f}(\mathbf{x}^{(n)}) \cdot \boldsymbol{\delta}^{(n)} = -\mathbf{f}(\mathbf{x}^{(n)}).$$

2. $\mathbf{x}^{(n+1)} := \mathbf{x}^{(n)} + \boldsymbol{\delta}^{(n)}$.

Example 3.1 *We will use the Newton-method for the system*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} 2x_1 + 3x_2 \\ 5x_1 + 4x_2^3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

with initial value $\mathbf{x}^{(0)} = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$

Of course, the exact solutions are

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{3}{8}\sqrt{30} \\ -\frac{1}{4}\sqrt{30} \end{pmatrix} \approx \begin{pmatrix} 2.0539 \\ -1.3693 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -\frac{3}{8}\sqrt{30} \\ \frac{1}{4}\sqrt{30} \end{pmatrix} \approx \begin{pmatrix} -2.0539 \\ 1.3693 \end{pmatrix}.$$

We have $D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} 2 & 3 \\ 5 & 12x_2^2 \end{pmatrix}$ and this matrix is regular for all $x_2 \neq \pm\sqrt{15/24}$.

For $n = 0$ and $n = 1$ we get by direct calculations

- $n = 0$

$$\begin{aligned}
 D\mathbf{f}(\mathbf{x}^{(0)}) \boldsymbol{\delta}^{(0)} &= -\mathbf{f}(\mathbf{x}^{(0)}) \\
 \Leftrightarrow \begin{pmatrix} 2 & 3 \\ 5 & 48 \end{pmatrix} \boldsymbol{\delta}^{(0)} &= \begin{pmatrix} -2 \\ 12 \end{pmatrix} \\
 \Leftrightarrow \boldsymbol{\delta}^{(0)} &= \begin{pmatrix} -44/27 \\ 34/81 \end{pmatrix} \\
 \Leftrightarrow \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} + \boldsymbol{\delta}^{(0)} \\
 &= \begin{pmatrix} 4 \\ -2 \end{pmatrix} + \begin{pmatrix} -44/27 \\ 34/81 \end{pmatrix} = \begin{pmatrix} 64/27 \\ -128/81 \end{pmatrix} \approx \begin{pmatrix} 2.3704 \\ -1.5802 \end{pmatrix}
 \end{aligned}$$

- $n = 1$

$$\begin{aligned}
 D\mathbf{f}(\mathbf{x}^{(1)}) \boldsymbol{\delta}^{(1)} &= -\mathbf{f}(\mathbf{x}^{(1)}) \\
 \Leftrightarrow \begin{pmatrix} 2 & 3 \\ 5 & 65'536/2'187 \end{pmatrix} \boldsymbol{\delta}^{(1)} &= \begin{pmatrix} 0 \\ 2'090'048/531'441 \end{pmatrix} \\
 \Leftrightarrow \boldsymbol{\delta}^{(1)} &= \begin{pmatrix} -2'090'048/7'959'627 \\ 4'180'096/23'878'881 \end{pmatrix} \\
 \Leftrightarrow \mathbf{x}^{(2)} &= \mathbf{x}^{(1)} + \boldsymbol{\delta}^{(1)} \\
 &= \begin{pmatrix} 16'777'216/7'959'627 \\ -33'554'432/23'878'881 \end{pmatrix} \approx \begin{pmatrix} 2.1078 \\ -1.4052 \end{pmatrix}
 \end{aligned}$$