1 Fixed Point Iteration and Contraction Mapping Theorem

Notation: For two sets A, B we write $A \subset B$ iff $x \in A \implies x \in B$. So $A \subset A$ is true. Some people use the notation "$\subseteq$" instead.

1.1 Introduction

Consider a function $y = g(x)$ where $x, y \in \mathbb{R}^n$:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} g_1(x_1, \dots, x_n) \\ \vdots \\ g_n(x_1, \dots, x_n) \end{bmatrix}$$

We assume that $g(x)$ is defined for $x \in D$ where D is a subset of $\mathbb{R}^n$.

The goal is to find a solution x^* of the **fixed point equation**

$$g(x) = x.$$

A method to find x^* is the **fixed point iteration**: Pick an initial guess $x^{(0)} \in D$ and define for $k = 0, 1, 2, \dots$

$$x^{(k+1)} := g(x^{(k)})$$

Note that this may not converge. But if the sequence $x^{(k)}$ converges, and the function g is continuous, the limit x^* must be a solution of the fixed point equation.

1.2 Contraction Mapping Theorem

The following theorem is called **Contraction Mapping Theorem** or **Banach Fixed Point Theorem**.

Theorem 1. *Consider a set $D \subset \mathbb{R}^n$ and a function $g \colon D \to \mathbb{R}^n$. Assume*

1. *D is closed (i.e., it contains all limit points of sequences in D)*
2. $x \in D \implies g(x) \in D$
3. *The mapping g is a* contraction *on D: There exists $q < 1$ such that*

$$\forall x, y \in D: \qquad \|g(x) - g(y)\| \le q \|x - y\| \tag{1}$$

Then

1. *there exists a unique $x^* \in D$ with $g(x^*) = x^*$*
2. *for any $x^{(0)} \in D$ the fixed point iterates given by $x^{(k+1)} := g(x^{(k)})$ converge to x^* as $k \to \infty$*
3. *$x^{(k)}$ satisfies the **a-priori error estimate***

$$\left\|x^{(k)} - x^*\right\| \le \frac{q^k}{1-q} \left\|x^{(1)} - x^{(0)}\right\| \tag{2}$$

*and the **a-posteriori error estimate***

$$\left\|x^{(k)} - x^*\right\| \le \frac{q}{1-q} \left\|x^{(k)} - x^{(k-1)}\right\| \tag{3}$$

Proof. Pick $x^{(0)} \in D$ and define $x^{(k)}$ for $k = 1, 2, \dots$ by $x^{(k)} := g(x^{(k-1)})$. We have from the contraction property (1)

$$\left\| x^{(k+1)} - x^{(k)} \right\| = \left\| g(x^{(k)}) - g(x^{(k-1)}) \right\| \le q \left\| x^{(k)} - x^{(k-1)} \right\| \tag{4}$$

and hence

$$\left\| x^{(k+1)} - x^{(k)} \right\| \le q^k \left\| x^{(1)} - x^{(0)} \right\| \tag{5}$$

Let $d := \left\| x^{(1)} - x^{(0)} \right\|$. We have from the triangle inequality and (5)

$$\begin{aligned} \left\| x^{(k)} - x^{(k+\ell)} \right\| &\le \left\| x^{(k)} - x^{(k+1)} \right\| + \cdots + \left\| x^{(k+\ell-1)} - x^{(k+\ell)} \right\| \\ &\le q^k d + \cdots + q^{k+\ell-1} d = q^k d \, (1 + q + \cdots + q^{\ell-1}) \\ \left\| x^{(k)} - x^{(k+\ell)} \right\| &\le q^k d \frac{1}{1-q} \end{aligned} \tag{6}$$

using the sum of the geometric series $\sum_{j=0}^{\ell-1} q^j \le \sum_{j=0}^{\infty} q^j = 1/(1-q)$. Note that (6) shows that the sequence $x^{(k)}$ is a *Cauchy sequence*. Therefore it must converge to a limit $x^* \in \mathbb{R}^n$ (since the space $\mathbb{R}^n$ is complete). As D is closed, we must have $x^* \in D$.

We need to show that $x^* = g(x^*)$: We have $x^{(k+1)} = g(x^{(k)})$, hence

$$\lim_{k\to\infty} x^{(k+1)} = \lim_{k\to\infty} g(x^{(k)})$$

The limit of the left hand side is x^*. Note that because of (1) the function g must be continuous. Therefore

$$\lim_{k\to\infty} g(x^{(k)}) = g(\lim_{k\to\infty} x^{(k)}) = g(x^*).$$

Next we need to show that the fixed point x^* is unique. Assume that we have fixed points $x^* = g(x^*)$ and $y^* = g(y^*)$. Then we obtain using the contraction property (1)

$$\left\| x^* - y^* \right\| = \left\| g(x^*) - g(y^*) \right\| \le q \left\| x^* - y^* \right\|$$

implying $(1-q)\left\| x^* - y^* \right\| \le 0$ and therefore $\left\| x^* - y^* \right\| = 0$, i.e., $x^* = y^*$.

The a-priori estimate (2) follows from (6) by letting ℓ tend to infinity. For the a-posteriori estimate use (2) with $k = 1$ for $\tilde{x}^{(0)} := x^{(k)}$, $\tilde{x}^{(1)} = x^{(k+1)}$. □

1.3 Proving the Contraction Property

The contraction property is related to the Jacobian $g'(x)$ which is an $n \times n$ matrix for each point $x \in D$. If the matrix norm satisfies $\|g'(x)\| \le q < 1$ then the mapping g must be a contraction:

Theorem 2. *Assume the set $D \subset \mathbb{R}^n$ is convex and the function $g\colon D \to \mathbb{R}^n$ has continuous partial derivatives $\dfrac{\partial g_j}{\partial k}$ in D. If for $q < 1$ the matrix norm of the Jacobian satisfies*

$$\forall x \in D: \qquad \left\| g'(x) \right\| \le q \tag{7}$$

the mapping g is a contraction in D and satisfies (1).

Proof. Let $x, y \in D$. Then the points on the straight line from x to y are given by $x + t(y - x)$ for $t \in [0, 1]$. As D is convex all these points are contained in D. Let $G(t) := g\big(x + t(y-x)\big)$, then by the chain rule we have $G'(t) = g'\big(x + t(y-x)\big)\,(y-x)$ and

$$g(y) - g(x) = G(1) - G(0) = \int_0^1 G'(t)dt = \int_0^1 g'\big(x + t(y-x)\big)\,(y-x)\,dt$$

As an integral of a continuous function is a limit of Riemann sums the triangle inequality implies $\left\| \int_a^b F(t)dt \right\| \le \int_a^b \|F(t)\|\,dt$:

$$\|g(y) - g(x)\| \le \int_0^1 \left\| g'\big(x + t(y-x)\big)\,(y-x)dt \right\| \le \int_0^1 \underbrace{\left\| g'\big(x + t(y-x)\big) \right\|}_{\le q} \|y - x\|\,dt \le q\,\|y - x\|$$

□

This is usually the easiest method to prove that a given mapping g is a contraction, see the examples in sections 1.5, 1.6.

1.4 A-priori and a-posteriori error estimates

The error estimates (2), (3) are useful for figuring out how many iterations we need. For this we need to know the contraction constant q (typically we get this from (7)).

A-priori estimate: For an initial guess $x^{(0)}$ we can find $x^{(1)}$. Without computing anything else we then have the error bound $\left\|x^{(k)}-x^*\right\| \leq \frac{q^k}{1-q}\left\|x^{(1)}-x^{(0)}\right\|$ for all future iterates $x^{(k)}$, before ("a-priori") we actually compute them. We can e.g. use this to find a value k such that $\left\|x^{(k)}-x^*\right\|$ is below a given tolerance.

A-posteriori estimate: After we have actually computed $x^{(k)}$ ("a-posteriori") we would like to know where the true solution x^* is located. Let

$$\delta_k := \frac{q}{1-q}\left\|x^{(k)}-x^{(k-1)}\right\|, \qquad D_k := \{x \mid \left\|x-x^{(k)}\right\| \leq \delta_k\}$$

The a-posteriori estimate states that x^* is contained in the set D_k. Note:

- the "radius" δ_k of D_k decreases at least by a factor of q with each iteration: $\delta_{k+1} \leq q\delta_k$
- the sets D_k are nested: $D_1 \supset D_2 \supset D_3 \supset \cdots$

To show $D_{k+1} \subset D_k$ assume $x \in D_{k+1}$. Then

$$\left\|x-x^{(k)}\right\| \leq \underbrace{\left\|x-x^{(k+1)}\right\|}_{\leq \delta_{k+1}} + \left\|x^{(k+1)}-x^{(k)}\right\| \leq \left(\frac{q}{1-q}+1\right)\left\|x^{(k+1)}-x^{(k)}\right\| \overset{(4)}{\leq} \frac{1}{1-q}q\left\|x^{(k)}-x^{(k-1)}\right\| = \delta_k \tag{8}$$

If we use the ∞-norm: $\left\|x^{(k)}-x^*\right\|_\infty \leq \delta_k$ means that for each component x_j^* we have a bracket

$$x_j^* \in [x_j^{(k)}-\delta_k, x_j^{(k)}+\delta_k],$$

i.e., the set D_k is a square/cube/hypercube with side length $2\delta_k$ centered in $x^{(k)}$.

1.5 Example

We want to solve the nonlinear system

$$x_1 = \tfrac{1}{10}\left[1 - x_2 - \sin(x_1+x_2)\right]$$
$$x_2 = \tfrac{1}{10}\left[2 + x_1 + \cos(x_1-x_2)\right]$$

where we have $g(x) = \frac{1}{10}\begin{bmatrix} 1-x_2-\sin(x_1+x_2) \\ 2+x_1+\cos(x_1-x_2) \end{bmatrix}$.

First we want to show that g is a contraction using Theorem 2. Therefore we first have to find the Jacobian $g'(x)$:

$$g'(x) = \frac{1}{10}\begin{bmatrix} -\cos(x_1+x_2) & -1-\cos(x_1+x_2) \\ 1-\sin(x_1-x_2) & \sin(x_1-x_2) \end{bmatrix}$$

Let $A := g'(x)$. Let us use the ∞-norm. We need to find an upper bound for $\|A\|_\infty = \max\{|a_{11}|+|a_{12}|, |a_{21}|+|a_{22}|\}$. We obtain for any $x_1, x_2 \in \mathbb{R}$

$$|a_{11}| = \tfrac{1}{10}\left|-\cos(x_1+x_2)\right| \leq \tfrac{1}{10}, \qquad |a_{12}| = \tfrac{1}{10}\left|-1-\cos(x_1+x_2)\right| \leq \tfrac{1}{10}(1+1)$$
$$|a_{21}| = \tfrac{1}{10}\left|1-\sin(x_1-x_2)\right| \leq \tfrac{1}{10}(1+1), \qquad |a_{22}| \leq \tfrac{1}{10}\left|\sin(x_1-x_2)\right| \leq \tfrac{1}{10}$$

Therefore for any $x \in \mathbb{R}^2$ we have

$$\left\|g'(x)\right\|_\infty \leq \frac{3}{10} = q < 1.$$

By Theorem 2 we therefore obtain that g is a contraction for all of $\mathbb{R}^2$.

We now want to use Theorem 1. We need to pick a set D such that the three assumptions of the theorem are satisfied. We consider two choices:

First choice $D = \mathbb{R}^2$: We can use the set $D = \mathbb{R}^2$. This set is closed. For any $x \in \mathbb{R}^2$ we certainly have that $g(x) \in \mathbb{R}^2$. We have also shown that g is a contraction for all of $\mathbb{R}^2$. Therefore we obtain from Theorem 1 that the nonlinear system $g(x) = x$ has exactly one solution x^* in all of $\mathbb{R}^2$.

Second choice $D = [-1,1] \times [-1,1]$: We can use for D the square with $-1 \le x_1 \le 1$ and $-1 \le x_2 \le 1$. This is a closed set (the boundary of the square is included). We now have to check that for $x \in D$ we have that $y = g(x) \in D$: We have using $-1 \le \sin\alpha \le 1$, $-1 \le \cos\alpha \le 1$

$$\begin{aligned} -\tfrac{2}{10} = \tfrac{1}{10}(1-1-1) \le y_1 = \tfrac{1}{10}[1 - x_2 - \sin(x_1+x_2)] \le \tfrac{1}{10}(1+1+1) = \tfrac{3}{10} \\ 0 = \tfrac{1}{10}(2-1-1) \le y_2 = \tfrac{1}{10}[2 + x_1 + \cos(x_1-x_2)] \le \tfrac{1}{10}(2+1+1) = \tfrac{4}{10} \end{aligned}$$

therefore $y \in D$ and the second assumption of the theorem is satisfied. We already showed that g is a contraction for all of $\mathbb{R}^2$, so the third assumption definitely holds for $x, y \in D$. We can now apply Theorem 1 and obtain that the nonlinear system has exactly one solution x^* which is located in the square $D = [-1,1] \times [-1,1]$.

Numerical Computation: We start with the initial guess $x^{(0)} = (0,0)^\top$. After each iteration we find δ_k and the square D_k containing x^*:

k	$x^{(k)}$	δ_k	D_k
1	$(.1,\ .3)^\top$	$1.3 \cdot 10^{-1}$	$[-.02857,\ .2286] \times [.1714,\ .4286]$
2	$(.03106,\ .3080)^\top$	$3.0 \cdot 10^{-2}$	$[.00151,\ .06060] \times [.2785,\ .3376]$
3	$(.03594,\ .2993)^\top$	$3.7 \cdot 10^{-3}$	$[.03221,\ .03967] \times [.2956,\ .3030]$
4	$(.03717,\ .3001)^\top$	$5.3 \cdot 10^{-4}$	$[.03664,\ .03770] \times [.2996,\ .3007]$
5	$(.03689,\ .3003)^\top$	$1.2 \cdot 10^{-4}$	$[.03677,\ .03701] \times [.3001,\ .3004]$

Note: (i) δ_k decreases at least by a factor of $q = 0.3$ with each iteration.

(ii) The sets D_k are nested: $D_1 \supset D_2 \supset D_3 \supset \cdots$

1.6 Using the Fixed Point Theorem *without* the Assumption $g(D) \subset D$

The tricky part in using the contraction mapping theorem is to find a set D for which *both* the 2nd and 3rd assumption of the fixed point theorem hold:

- $x \in D \implies g(x) \in D$
- g is a contraction on D

Typically we can prove that $\|g'(x)\| \le q < 1$ for x in some convex region $\tilde{D}$. We suspect that there is a solution x^* of the fixed point equation in $\tilde{D}$. But it may not be true that $g(x) \in \tilde{D}$ for all $x \in \tilde{D}$.

In this case we may be able to prove a result by computing a few iterates $x^{(k)}$: Start with $k = 0$ and an initial guess $x^{(0)} \in \tilde{D}$. Then repeat

- let $k := k+1$ and compute $x^{(k)} := g(x^{(k-1)})$
- compute $\delta_k := \dfrac{q}{1-q}\left\|x^{(k)} - x^{(k-1)}\right\|$, let $D_k := \{x \mid \left\|x - x^{(k)}\right\| \le \delta_k\}$

until either $D_k \subset \tilde{D}$ or $x^{(k)} \notin \tilde{D}$.

If the iterates exit from the set $\tilde{D}$ we cannot conclude anything. But as long as the points $x^{(k)}$ stay inside $\tilde{D}$ we have $\delta_{k+1} \le q\delta_k$ and $D_{k+1} \subset D_k$. So we expect that for some k the condition $D_k \subset \tilde{D}$ will be satisfied (if $x^{(k)}$converges to a limit in the interior of $\tilde{D}$ the loop must terminate with $D_k \subset \tilde{D}$; but in general it is possible that the loop never terminates). If the loop does terminate with $D_k \subset \tilde{D}$ for $k = K$ we have the following result:

Theorem 3. *Let $\tilde{D} \subset \mathbb{R}^n$ and assume that the function $g\colon \tilde{D} \to \mathbb{R}^n$ satisfies for $q < 1$*

$$\forall x, y \in \tilde{D}: \qquad \|g(x) - g(y)\| \le q\,\|x - y\|$$

Let $x^{(0)} \in \tilde{D}$ and define for $k = 0, 1, 2, \ldots$

$$x^{(k+1)} := g(x^{(k)}), \qquad \delta_k := \frac{q}{1-q}\left\|x^{(k)} - x^{(k-1)}\right\|, \qquad D_k := \{x \mid \left\|x - x^{(k)}\right\| \le \delta_k\}$$

If for some K we have $x^{(K-1)} \in \tilde{D}$ and $D_K \subset \tilde{D}$ there holds

- *the equation $g(x) = x$ has a unique solution x^* in $\tilde{D}$*
- *this solution satisfies $x^* \in D_k$ for all $k \ge K$*

Proof. Let $x \in D_K$. We want to show that $g(x) \in D_K$: As $D_K \subset \tilde{D}$ the contraction property gives using the definition of D_k and δ_k

$$\left\|g(x) - x^{(K)}\right\| \le q\left\|x - x^{(K-1)}\right\| \le q\left\|x - x^{(K)}\right\| + q\left\|x^{(K)} - x^{(K-1)}\right\| \le q\delta_K + (1-q)\delta_K = \delta_K$$

As D_K is closed and $D_K \subset \tilde{D}$ the set $D := D_K$ satisfies all three assumptions of the fixed point theorem Theorem 1. Hence there is a unique solution $x^* \in D$. The a-posteriori estimate (3) states that $x^* \in D_k$ for all iterates $x^{(k)}$ with $k \ge K$. Assume that there is another fixed point $y^* \in \tilde{D}$ with $g(y^*) = y^*$. Then

$$\|y^* - x^*\| = \|g(y^*) - g(x^*)\| \le q\,\|y^* - x^*\|$$

As $q < 1$ we must have $\|y^* - x^*\| = 0$. □

Summary:

- Find a convex set $\tilde{D}$ for which you suspect $x^* \in \tilde{D}$ and where you can show $\|g'(x)\| \le q < 1$
- Pick $x^{(0)} \in \tilde{D}$ and perform the fixed point iteration:
  for each iteration:
  - find $x^{(k)}$ and D_k
  - if $x^{(k)} \notin \tilde{D}$: stop (we can't conclude anything)
  - if $D_k \subset \tilde{D}$: success: there is a unique solution $x^* \in \tilde{D}$, and there holds $x^* \in D_k$ for this and all following iterations

Example: Let $g(x) := \frac{1}{3}\begin{bmatrix} x_1 - x_1x_2 + 1 \\ x_2 + x_1x_2^2 + 1 \end{bmatrix}$. Then the Jacobian is $g'(x) = \frac{1}{3}\begin{bmatrix} 1 - x_2 & -x_1 \\ x_2^2 & 1 + 2x_1x_2 \end{bmatrix}$.
Let us try to use $\tilde{D} = [0,a] \times [0.a]$ with $a \le 1$ and the ∞-norm. We then obtain for $x \in \tilde{D}$ that

$$\left\|g'(x)\right\|_\infty \le \tfrac{1}{3}\max\{1 + a, a^2 + 1 + 2a^2\}$$

For $a = 1$ we get $\|g'(x)\|_\infty \le \frac{4}{3}$ which is too large. So we try $a = 0.6$ which gives $\|g'(x)\|_\infty \le \frac{2.08}{3} =: q < 1$. Therefore g is a contraction on $\tilde{D} = [0,.6] \times [0,.6]$. Note that $g(\begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix}) = \begin{bmatrix} 0.41333 \\ 0.60533 \end{bmatrix} \notin \tilde{D}$, so $\tilde{D}$ does *not* satisfy all three assumptions of Theorem 1.

For $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ we obtain

$$\begin{aligned}
x^{(1)} &= (.33333,\ .33333)^\top \in \tilde{D}, & D_1 &= [-0.42029,\ 1.08696] \times [-0.42029,\ 1.08696] \not\subset \tilde{D}\\
x^{(2)} &= (.40741,\ .45679)^\top \in \tilde{D}, & D_2 &= [0.12829,\ 0.68653] \times [0.17767,\ 0.73591] \not\subset \tilde{D}\\
x^{(3)} &= (.40710,\ .51393)^\top \in \tilde{D}, & D_3 &= [0.27791,\ 0.53629] \times [0.38474,\ 0.64313] \not\subset \tilde{D}\\
x^{(4)} &= (.39929,\ .54049)^\top \in \tilde{D}, & D_4 &= [0.33926,\ 0.45933] \times [0.48045,\ 0.60052] \not\subset \tilde{D}\\
x^{(5)} &= (.39449,\ .55238)^\top \in \tilde{D}, & D_5 &= [0.36761,\ 0.42138] \times [0.52549,\ 0.57926] \subset \tilde{D}
\end{aligned}$$

Therefore we can conclude from Theorem 3 that there exists a unique solution $x^* \in \tilde{D} = [0, 0.6] \times [0, 0.6]$. This solution x^* is located in the smaller square D_5. For $k = 5, 6, 7, \ldots$ we obtain $x^* \in D_k$ where D_k is a square with side length $2\delta_k$. As $\delta_k \le q^{k-5}\delta_5 \le \left(\frac{2.08}{3}\right)^{k-5} 0.027$ we can obtain arbitrarily small squares containing the solution if we choose k sufficiently large.