On the Newton method

December 1, 2021

1 Banach contraction principle

Discuss the Banach contraction principle, continuous dependence on parameters of the fixed point for the family of contractions.

Definition 1 *Assume (X, d) is a metric space. We say $f : X \to X$ is a contraction with a constant $L < 1$, iff*

$$d(f(x), f(y)) \leq Ld(x, y) \quad \forall x, y \in X \tag{1}$$

Theorem 1 *Assume (X, d) is a complete metric space and $f : X \to X$ is a contraction with a constant $L < 1$. Then*

There exists a unique fix point x_0 for f.

Moreover, for every $y \in X$ the sequence $f^n(y)$ converges geometrically to x_0,

$$d(f^n(y), x_0) \leq \frac{L^n}{1-L} d(y, f(y)), \tag{2}$$

$$d(f^n(y), x_0) \leq L^n d(y, x_0). \tag{3}$$

Proof: Let $y \in X$. Then

$$d(f^n(y), f^{n+1}(y)) \leq Ld(f^{n-1}(y), f^n(y)) \leq \cdots \leq L^n d(y, f(y)). \tag{4}$$

We show that the sequence $\{f^n(y)\}$ satisfies the Cauchy condition.
We have

$$\begin{aligned} d(f^n(y), f^{n+k}(y)) &\leq d(f^n(y), f^{n+1}(y)) + \cdots + d(f^{n+k-1}(y), f^{n+k}(y)) \\ &\leq (L^n + \cdots + L^{n+k-1}) d(y, f(y)) \leq \frac{L^n}{1-L} d(y, f(y)) \end{aligned}$$

Hence the Cauchy condition is satisfied and there exists $x_0 = \lim_{n\to\infty} f^n(y)$.
From continuity of f it follows that

$$f(x_0) = f(\lim_{n\to\infty} f^n(y)) = \lim_{n\to\infty} f^{n+1}(y) = x_0.$$

The uniqueness is of x_0 is obvious.
We have

$$d(f^n(y), x_0) = d(f^n(y), f^n(x_0)) \leq L^n d(y, x_0).$$

∎

Continuous dependence on parameters.

Theorem 2 *Assume that (X, d) is a complete metric space and (Λ, ρ) is a matric space.*

Let

$$f : \Lambda \times X \to X,$$

satisfies the following

- *there exists $L < 1$ such that for all $\lambda \in \Lambda$ the map $f_\lambda : X \to X$ given by $f_\lambda(x) = f(\lambda, x)$ is a contraction with the constant L*
- *there exists C such that for all $x \in X$*

$$d(f(\lambda_1, x), f(\lambda_2, x)) \leq C\rho(\lambda_1, \lambda_2) \tag{5}$$

Let x_λ be the unique fixed point of f_λ. Then

$$d(x_{\lambda_1}, x_{\lambda_2}) \leq \frac{C\rho(\lambda_1, \lambda_2)}{1 - L}. \tag{6}$$

Proof: We have

$$\begin{aligned} d(x_{\lambda_1}, x_{\lambda_2}) &= d(f(\lambda_1, x_{\lambda_1}), f(\lambda_2, x_{\lambda_2})) \\ &\leq d(f(\lambda_1, x_{\lambda_1}), f(\lambda_1, x_{\lambda_2})) + d(f(\lambda_1, x_{\lambda_2}), f(\lambda_2, x_{\lambda_2})) \\ &\leq Ld(x_{\lambda_1}, x_{\lambda_2}) + C\rho(\lambda_1, \lambda_2). \end{aligned}$$

Hence

$$(1 - L)d(x_{\lambda_1}, x_{\lambda_2}) \leq C\rho(\lambda_1, \lambda_2) \tag{7}$$

∎

2 Smoothness

Theorem 3 *Assume that X is closed and convex subset of Banach space $\tilde{X}$ and Λ is closed subset of the Banach space $\tilde{\Lambda}$.*

Let

$$f : \Lambda \times X \to X,$$

be C^2 map, such that there exist $L < 1$ and constants $M_1, \ldots$ such that

$$\left\| \frac{\partial f}{\partial x}(\lambda, x) \right\| \leq L, \tag{8}$$

$$\left\| \frac{\partial f}{\partial \lambda}(\lambda, x) \right\| \leq M_1, \tag{9}$$

$$\left\| \frac{\partial^2 f}{\partial \lambda \partial x}(\lambda, x) \right\| \leq M_2, \tag{10}$$

$$\left\| \frac{\partial^2 f}{\partial x^2}(\lambda, x) \right\| \leq M_3. \tag{11}$$

Let $\bar{x}(\lambda)$ be the unique fixed point of f_λ. Then $\bar{x}(\lambda)$ is C^1.

Proof: We see that for each $\lambda \in \Lambda$ f_λ is a contraction on X, hence for any function $x_0(\lambda)$ the sequence

$$x_{n+1}(\lambda) = f_\lambda(x_n(\lambda)) \tag{12}$$

is uniformly converging to $\bar{x}(\lambda)$.

Observe that if $x_0(\lambda)$ is C^1, then $x_n(\lambda)$ is also C^1. The question is wether $Dx_n(\lambda)$ converges uniformly to $D\bar{x}(\lambda)$. We will show that this is the case.

Observe that

$$Dx_{n+1}(\lambda) = \frac{\partial f}{\partial \lambda}(\lambda, x_n(\lambda)) + \frac{\partial f}{\partial x}(\lambda, x_n(\lambda))Dx_n(\lambda). \tag{13}$$

Now we have two ways to proceed.

1. take any $x_0(\lambda)$ and knowing that $x_n(\lambda)$ converges uniformly work with (13)
2. work on the space of functions $\lambda \to (x(\lambda), A(\lambda))$, where $A \in \mathrm{Lin}(\tilde{\Lambda}, \tilde{X})$ and prove that we have a contraction,...

Let us chose the second method. We define a map on $\{g : \Lambda \to X \times \mathrm{Lin}(\tilde{\Lambda}, \tilde{X})\}$ as follows

$$\mathcal{F}(x, A)(\lambda) = (f(\lambda, x(\lambda)), \frac{\partial f}{\partial \lambda}(\lambda, x(\lambda)) + \frac{\partial f}{\partial x}(\lambda, x(\lambda))A(\lambda)) \tag{14}$$

Let us find first a set Z such that $\mathcal{F}(Z) \subset Z$. It is given as follows

$$Z = X \times \{A\,, \|A\| \leq G\} \tag{15}$$

where G is a suitable constant computed below.

$$\|\mathcal{F}_2(Z)(\lambda)\| \leq \left\| \frac{\partial f}{\partial \lambda}(\lambda, x(\lambda)) \right\| + LG \leq M_1 + LG$$

hence if

$$G \geq \frac{M_1}{(1-L)} \tag{16}$$

then

$$\mathcal{F}(Z) \subset Z. \tag{17}$$

In order to show that $\mathcal{F}$ is a contraction on Z we introduce the following norm

$$\|(x, A)\|_1 = \sup_{\lambda \in \Lambda} \|x(\lambda)\| + C \sup_{\lambda \in \Lambda} \|A(\lambda)\| \tag{18}$$

for some $C > 0$ to be fixed later.

We have

$$\begin{aligned}
\|\mathcal{F}_2(x,A)(\lambda)-\mathcal{F}_2(y,B)(\lambda)\| &\le \left\|\frac{\partial f}{\partial\lambda}(\lambda,x(\lambda))-\frac{\partial f}{\partial\lambda}(\lambda,y(\lambda))\right\|+\\
&\quad+\left\|\frac{\partial f}{\partial x}(\lambda,x(\lambda))A(\lambda)-\frac{\partial f}{\partial x}(\lambda,y(\lambda))B(\lambda))\right\|\\
&\le \left\|\frac{\partial^2 f}{\partial\lambda\partial x}\right\|\cdot\|x(\lambda)-y(\lambda)\|+\left\|\frac{\partial f}{\partial x}(\lambda,x(\lambda))A(\lambda)-\frac{\partial f}{\partial x}(\lambda,x(\lambda))B(\lambda)\right\|\\
&\quad+\left\|\frac{\partial f}{\partial x}(\lambda,x(\lambda))B(\lambda)-\frac{\partial f}{\partial x}(\lambda,y(\lambda))B(\lambda)\right\|\\
&\le M_2\|x(\lambda)-y(\lambda)\|+L\|A(\lambda)-B(\lambda)\|+\left\|\frac{\partial^2 f}{\partial x^2}\right\|\cdot\|B(\lambda)\|\cdot\|x(\lambda)-y(\lambda)\|\\
&\le (M_2+M_3G)\|x(\lambda)-y(\lambda)\|+L\|A(\lambda)-B(\lambda)\|.
\end{aligned}$$

From the above we obtain

$$\begin{aligned}
\|\mathcal{F}(x,A)-\mathcal{F}(y,B)\|_1 &\le L\|x-y\|+C((M_2+M_3G)\|x-y\|+L\|A-B\|)\\
&= (L+C(M_2+M_3G))\|x-y\|+CL\|A-B\|
\end{aligned}$$

We need that

$$\|\mathcal{F}(x,A)-\mathcal{F}(y,B)\|_1 \le \tilde{L}\|(x,A)-(y,B)\|_1,\quad \tilde{L}<1.$$

For this need that

$$(L+C(M_2+M_3G))\|x-y\|+CL\|A-B\| \le \tilde{L}\|x-y\|+C\tilde{L}\|A-B\|. \quad (19)$$

It is easy to see that this is achieved for C sufficiently small, such that

$$L+C(M_2+M_3G)=\tilde{L}<1. \quad (20)$$

To finish the proof observe that if we start the iteration of $\mathcal{F}$ with $(x_0(\lambda),Dx_0(\lambda))$, then our uniformly converging sequence has the form (x_n,Dx_n) and Dx_n are uniformly convergent, then $\bar{x}(\lambda)$ is C^1 and this limit must be $D\bar{x}$. ∎

Now we will try the first approach mentioned in the above proof.

From (13) we have

$$\begin{aligned}
\|Dx_{n+1} - Dx_n\| &\le \left\| \frac{\partial f}{\partial \lambda}(\lambda, x_n(\lambda)) + \frac{\partial f}{\partial x}(\lambda, x_n(\lambda)) Dx_n(\lambda) \right. \\
&\quad \left. - \left(\frac{\partial f}{\partial \lambda}(\lambda, x_{n-1}(\lambda)) + \frac{\partial f}{\partial x}(\lambda, x_{n-1}(\lambda)) Dx_{n-1}(\lambda) \right) \right\| \\
&\le M_2 \|x_n - x_{n-1}\| + \left\| \frac{\partial f}{\partial x}(\lambda, x_n(\lambda)) Dx_n(\lambda) - \frac{\partial f}{\partial x}(\lambda, x_{n-1}(\lambda)) Dx_{n-1}(\lambda) \right\| \\
&\le M_2 \|x_n - x_{n-1}\| + \left\| \frac{\partial f}{\partial x}(\lambda, x_n(\lambda)) Dx_n(\lambda) - \frac{\partial f}{\partial x}(\lambda, x_n(\lambda)) Dx_{n-1}(\lambda) \right\| \\
&\quad + \left\| \frac{\partial f}{\partial x}(\lambda, x_n(\lambda)) Dx_{n-1}(\lambda) - \frac{\partial f}{\partial x}(\lambda, x_{n-1}(\lambda)) Dx_{n-1}(\lambda) \right\| \\
&\le M_2 \|x_n - x_{n-1}\| + L \|Dx_n - Dx_{n-1}\| + M_3 \|Dx_{n-1}\| \cdot \|x_n - x_{n-1}\| \\
&\le (M_2 + M_3 G) \|x_n - x_{n-1}\| + L \|Dx_n - Dx_{n-1}\|
\end{aligned}$$

where G is an a priori bound for $\|Dx_n\|$.

Observe that we have the following relation

$$\begin{bmatrix} \|x_{n+1} - x_n\| \\ \|Dx_{n+1} - Dx_n\| \end{bmatrix} \le P \cdot \begin{bmatrix} \|x_n - x_{n-1}\| \\ \|Dx_n - Dx_{n-1}\| \end{bmatrix}$$

where

$$P = \begin{bmatrix} L, & 0 \\ M_2 + M_3 G, & L \end{bmatrix}.$$

Since

$$\begin{bmatrix} \|x_{n+1} - x_n\| \\ \|Dx_{n+1} - Dx_n\| \end{bmatrix} \le P^n \cdot \begin{bmatrix} \|x_1 - x_0\| \\ \|Dx_1 - Dx_0\| \end{bmatrix}$$

and $P^n \to 0$ we obtain the uniform convergence of Dx_n.

3 Derivation of Newton method

Let $F : \mathbb{R}^d \to \mathbb{R}^d$ be a C^2 map. We want to solve

$$F(x) = 0. \tag{21}$$

We have

$$F(x_n + h_n) = F(x_n) + DF(x_n) h_n + R(x_n, h_n) \tag{22}$$

where $R(x_n, h_n)$ is the remainder term, which can be estimated as follows

$$|R(x_n, h_n)| \le \frac{1}{2} \sup_{x \in \overline{B(x_n, |h_n|)}} \|D^2 F(x)\| \cdot |h_n|^2. \tag{23}$$

We drop the remainder term and solve for h_n equation

$$F(x_n) + DF(x_n)h_n = 0$$

to obtain

$$h_n = -DF(x_n)^{-1}F(x_n). \tag{24}$$

We define the Newton operator by

$$N(x) = x - DF(x)^{-1}F(x) \tag{25}$$

and then

$$x_{n+1} = N(x_n). \tag{26}$$

Lemma 4 *Assume that $DF(\bar{x})$ is an isomorphism. Then $F(\bar{x}) = 0$ iff $N(\bar{x}) = \bar{x}$.*

Proof: We have the following sequence of equivalent statements

$$\begin{aligned} F(\bar{x}) &= 0, \\ DF(\bar{x})^{-1}F(\bar{x}) &= 0, \\ \bar{x} - DF(\bar{x})^{-1}F(\bar{x}) &= \bar{x}, \\ N(\bar{x}) &= \bar{x} \end{aligned}$$

∎

3.1 Newton method for fixed point

We look fixed point of P.

$F(x) = x - P(x)$

4 Contraction in the vicinity of the fixed point

$$N(x) = x - DF(x)^{-1}F(x) \tag{27}$$

Theorem 5 *Assume that $F(\bar{x}) = 0$ and $DF(\bar{x})$ is an isomorphism. Then there exists U, a neighborhood of $\bar{x}$, such that for all $x_0 \in U$ $DN(x_0)$ is defined and $x_n = N^n(x_0) \in U$ for $n \neq 1$ and $x_n \to \bar{x}$.*

Proof: Observe that $N(\bar{x}) = \bar{x}$.

It is enough to show that $\|DN(x)\|$ is small for x close to $\bar{x}$.

We have

$$\begin{aligned} DN(\bar{x}) &= I - D(DF^{-1}(x)F(x))_{x=\bar{x}} \\ &= I - D(DF^{-1})_{x=\bar{x}}F(\bar{x}) - DF^{-1}(\bar{x})DF(\bar{x}) = I - I = 0. \end{aligned}$$

For any $\epsilon > 0$ (we take $\epsilon < 1$) we can find $r = r(\epsilon) > 0$, such that

- $DF(x)$ is invertible for $|x - \bar{x}| \leq r$, i.e. such x are in the domain DN
- $\|DN(x)\| \leq \epsilon$ for $|x - \bar{x}| \leq r$

Then we define $U = \bar{B}(\bar{x}, r)$. We have

$$\|N(x) - \bar{x}\| = \|N(x) - N(\bar{x})\| \leq \epsilon r.$$

∎

5 Heuristics about convergence

$$N(x) = x - DF(x)^{-1}F(x) \tag{28}$$

Assume that we have fixed convex set $D \subset \mathbb{R}^d$ and positive constants M and B such that

- $\sup_{x \in D} \|D^2F(x)\| \leq 2M$
- $DF(x)^{-1}$ exists for all $x \in D$. Moreover,

$$\|DF^{-1}(x)\| \leq B, \quad \forall x \in D. \tag{29}$$

Consider a sequence $x_{n+1} = N(x_n)$. Let us denote the "error" e_n and the "step" $h_n = x_{n+1} - x_n$ by

$$\begin{aligned} e_n &= |F(x_n)|, \\ h_n &= -DF^{-1}F(x_n) \end{aligned}$$

Then as long as our sequence is in D we have

$$\begin{aligned} |h_n| &\leq Be_n \\ e_{n+1} &= |F(x_{n+1})| \leq M|h_n|^2. \end{aligned}$$

Therefore

$$e_{n+1} \leq MB^2 e_n^2.$$

We see that the error is decaying at least quadratically (if the initial error e_0 is small). Then $|h_n|$ rapidly approach zero and $\sum h_n$ is convergent. The solution is then given as $\bar{x} = \lim_{n\to\infty} x_n = x_0 + \sum_{n=0}^{\infty} h_n$.

5.1 More formally

Lemma 6 *We assume that $x_n \in D$ for all $n = 0, 1, \ldots, K$. Then*

$$e_n \leq \frac{\left(MB^2 e_0\right)^{2^n}}{MB^2}, \quad |h_n| \leq \frac{\left(MB^2 e_0\right)^{2^n}}{MB}, \qquad 1 \leq n \leq K.$$

Proof: For a given e_0 we have

$$\begin{aligned} e_0 &, \quad |h_0| \le Be_0, \\ e_1 &\le MB^2e_0^2, \quad |h_1| \le MB^3e_0^2, \\ e_2 &\le (MB^2)\left(MB^2e_0^2\right)^2 = M^3B^6e_0^4, \quad |h_2| \le M^3B^7e_0^4 \end{aligned}$$

We see that the power of B in $|h_n|$ is by one larger than it is in e_n.

The formal proof by induction. We see that for $n = 1$ it is ok. The induction step

$$e_{n+1} \le M|h_n|^2 \le M\left(\frac{\left(MB^2e_0\right)^{2^n}}{MB}\right)^2 = \frac{\left(MB^2e_0\right)^{2^{n+1}}}{MB^2},$$

$$|h_{n+1}| \le Be_{n+1} \le \frac{\left(MB^2e_0\right)^{2^{n+1}}}{MB}$$

∎

Now in order for the iteration to be continued forever and converging to a solution of (21) we need that we stay in D.

Theorem 7 *Assume that we have the set $D \subset \mathbb{R}^d$ and constants B, M as above.*
Let $x_0 \in \mathbb{R}^d$ and $R > 0$ be such that $B(x_0, R) \subset D$.
Let $e_0 = |F(x_0)|$.
Assume that

$$\begin{aligned} \max(M,1)\max(B^2,1)e_0 &\le \epsilon \le 1/2 \\ 2\epsilon &\le R. \end{aligned}$$

Then iteration (26) is converging to $\bar{x} \in D$ a solution of (21). Moreover, it holds

$$|x_n - \bar{x}| \le 2\epsilon^{2^n}. \tag{30}$$

Proof:

By increasing B and M if necessary we can assume that

$$M \ge 1, \quad B \ge 1.$$

From our assumptions it follows that

$$|h_0| \le Be_0 \le MB^2e_0 \le \epsilon$$

and

$$|h_n| \le \frac{(MB^2e_0)^{2^n}}{MB} \le \epsilon^{2^n}.$$

The series $\sum_{n=0}^{\infty} h_n$ is converging very fast we have the following estimate

$$\left|\sum_{n=0}^{\infty} h_n\right| \le \sum_{n=0}^{\infty} |h_n| \le \sum_{n=0}^{\infty} \epsilon^{2^n} < \sum_{n=0}^{\infty} \epsilon^{n+1} = \epsilon\frac{1}{1-\epsilon} \le 2\epsilon.$$

Since by our assumption

$$2\epsilon \leq R$$

so we can iterate forever.

Observe that

$$|x_n - \bar{x}| = \left| \sum_{k=n}^{\infty} h_k \right| \leq \sum_{k=n}^{\infty} |h_k| \leq \sum_{k=n}^{\infty} \epsilon^{2^k} = \sum_{k=n}^{\infty} \epsilon^{2^n \cdot 2^{k-n}} =$$

$$\sum_{k=n}^{\infty} \left(\epsilon^{2^n}\right)^{2^{k-n}} \leq \sum_{j=0}^{\infty} \left(\epsilon^{2^n}\right)^{2^j} < \sum_{j=0}^{\infty} \left(\epsilon^{2^n}\right)^{j+1} = \epsilon^{2^n} \frac{1}{1-\epsilon^{2^n}} < 2\epsilon^{2^n}.$$

∎

6 Modifications of Newton method

We can try to do not compute $DF(x)^{-1}$ but instead use a constant matrix A which is somehow similar to $DF(x)^{-1}$. The scheme will take the form

$$x_{n+1} = N_m(x_n) = x_n - Af(x_n). \tag{31}$$

It is easy to see that if A is an isomorphism, then $f(\bar{x}) = 0$ iff $N_m(\bar{x}) = \bar{x}$.

We have the following theorem (compare Thm. 5)

Theorem 8 *Consider the modified Newton-method (31) Assume that $F(\bar{x}) = 0$ and $\|I - ADF(\bar{x})\| < 1$. Then there exists U, a neighborhood of $\bar{x}$, such that for all $x_0 \in U$ $x_n \in U$ for $n \geq 1$ and $x_n \to \bar{x}$.*

Moreover, $N_m : U \to U$ is a contraction.

Proof: Since

$$DN_m(x) = I - A \cdot DF(x), \tag{32}$$

in $U = B(\bar{x}, r)$ for $r > 0$ sufficiently small we will have

$$\|DN_m(x)\| \leq L < 1$$

Hence we will have a contraction in the ball $U = B(\bar{x}, r)$, some r small enough. ∎

Observe that the condition $\|I - ADF(\bar{x})\| < 1$ implies that both A and $DF(\bar{x})$ are isomorphisms.

Lemma 9 *Assume that $A, B \in L(\mathbb{R}^n, \mathbb{R}^n$ and*

$$\|I - AB\| < 1, \tag{33}$$

then A and B are isomorphisms.

Proof: First observe that $\ker AB = \{0\}$. Indeed if $v \in \ker AB$, then

$$v = (I - AB)v$$
$$\|v\| \leq \|I - AB\| \cdot \|v\|$$

since $\|I - AB\| < 1$ we obtain $\|v\| = 0$.

Since the dimension is finite, this implies that AB is an isomorphism, hence also A and B are isomorphisms. ∎

In infinite dimensions the above lemma is not true. We need to assume that B is an isomorphism. We have for some D with $\|D\| < 1$

$$AB = I + D$$
$$A = (I + D)B^{-1}$$

Now it is easy to see that $I+D$ is invertible with $(I+D)^{-1} = I-D+D^2-D^3+\ldots$, hence $A^{-1} = B^{-1}(I + D)^{-1}$.

Analogous approach works if we assume that A is an isomorphism, then B is an isomorphism.

7 How to find a good candidate for the starting point

- uniform grid - good for low dimension only
- random choice
- homotopy method

8 The local diffeomorhism theorem

The goal of this section is to prove the local diffeomorphism theorem using our experience and knowledge of the Newton method.

This proof works also for the Banach spaces.

Theorem 10 *Let $W \subset \mathbb{R}^n$ be an open set and $f : W \to \mathbb{R}^n$ is C^k, with $k \geq 1$.*

Assume that for some $x_0 \in W$ $Df(x_0)$ is an isomorphism. Then there exists $U \subset W$ - an open set with $x_0 \in U$ such that $f : U \to f(U)$ is diffemorphism, $f(U)$ is open and $f^{-1} : f(U) \to U$ is also of class C^k.

Proof:

The idea of the proof:

In order to invert f we need to solve the following equation

$$f(x) = y. \tag{34}$$

We know that we can solve $f(x) = y_0$ with x_0 being the solution. Moreover, we know that if we set-up a Newton method for $F(x) = y_0 - f(x)$, then $DF(x) =$

$-df(x)$, hence $DF(x_0)$ is an isomorphism, it will be an attracting fixed point for the (modified) Newton method for function F.

Observe now that will have a ball $B(x_0, R)$ which will be mapped into itself and the Newton method will be a contraction. Imagine now that we change the parameter $y_0 \to y$, then we will have a contraction in the same ball $B(x_0, R)$ and the obtained fixed point will be a solution of (34) and will depend continuously on y.

Details:

For $y \in \mathbb{R}^n$ let us define a function

$$F_y(x) = y - f(x) \tag{35}$$

and the modified Newton operator for equation $F_y(x) = 0$

$$N_y(x) = x - (-Df(x_0))^{-1}(y - f(x)) = x + (Df(x_0))^{-1}(y - f(x)). \tag{36}$$

In view of the above heuristics we would like to find $R > 0$, $r > 0$ and $0 \leq L < 1$, such that

$$\begin{aligned} \overline{B(x_0, R)} &\subset W, && (37)\\ N_y(\overline{B(x_0, R)}) &\subset B(x_0, R), \quad \forall y \in \overline{B(y_0, r)}, && (38)\\ \left\| \frac{\partial}{\partial x} N_y(x) \right\| &\leq L < 1, \quad \forall y \in \overline{B(y_0, r)}, \forall x \in \overline{B(x_0, R)} && (39) \end{aligned}$$

From (36) it follows that

$$\frac{\partial}{\partial x} N_y(x) = I - (Df(x_0))^{-1} Df(x)), \tag{40}$$

hence $\frac{\partial}{\partial x} N_y(x_0) = 0$ and for $L = 1/2$ we will find and $R > 0$ such that

$$\overline{B(x_0, R)} \subset W, \quad \left\| \frac{\partial}{\partial x} N_y(x) \right\| \leq L, \forall y \forall x \in \overline{B(x_0, R)}. \tag{41}$$

and $Df(x)$ is an isomorphism for all $x \in \overline{B(x_0, R)}$.

It remains to establish (38). First let us notice that

$$N_{y_1}(x) - N_y(x) = (Df(x_0))^{-1}(y_1 - y).$$

We have for $x \in \overline{B(x_0, R)}$ and $y \in \overline{B(y_0, r)}$

$$\begin{aligned} \|N_y(x) - x_0\| = \|N_y(x) - N_{y_0}(x_0)\| \leq \\ \|N_y(x) - N_y(x_0)\| + \|N_y(x_0) - N_{y_0}(x_0)\| \leq \\ L\|x - x_0\| + \|(Df(x_0))^{-1}(y - y_0)\| \leq LR + \|Df(x_0)^{-1}\| r. \end{aligned}$$

Hence for (38) it is enough to have the following inequality

$$LR + \|Df(x_0)^{-1}\| r \leq R, \tag{42}$$

therefore we obtain

$$r < \frac{(1-L)R}{\|Df(x_0)^{-1}\|}. \tag{43}$$

We fix such an r. From the Banach contraction theorem we obtain that for each $y \in B(y_0, r)$ there is a unique $x \in B(x_0, R)$, such that $N_y(x) = x$ which implies that $f(x) = y$.

We set $U = f^{-1}(B(y_0, r)) \cap B(x_0, R)$. This is an open set and $f(U) = B(y_0, r)$.

Now we need to establish the continuity and the differentiability of f^{-1}.

Continuity: This is a consequence of the fact that $f^{-1}(y)$ is obtained as the fixed point of the parameterized family of contractions. This is realized as follows.

We have

$$\begin{aligned} \|f^{-1}(y_1) - f^{-1}(y_2)\| = \|N_{y_1}(f^{-1}(y_1)) - N_{y_2}(f^{-1}(y_2)\| \leq \\ \|N_{y_1}(f^{-1}(y_1)) - N_{y_2}(f^{-1}(y_1))\| + \|N_{y_2}(f^{-1}(y_1)) - N_{y_2}(f^{-1}(y_2)\| \leq \\ \|Df(x_0)^{-1}\| \cdot \|y_1 - y_2\| + L\|f^{-1}(y_1) - f^{-1}(y_2)\|, \end{aligned}$$

hence

$$\|f^{-1}(y_1) - f^{-1}(y_2)\| \leq \frac{\|Df(x_0)^{-1}\|}{1-L}\|y_1 - y_2\|. \tag{44}$$

From (44) we see that $y_1 \to y_2$ iff $x_1 = f^{-1}(y_1) \to x_2 = f^{-1}(y_2)$. (The other direction is from the Lipshitz property of f.)

Differentiability.

Let us fix $x_2 \in U$.

From the differentiability of f at x_2 we have

$$f(x_1) - f(x_2) = Df(x_2)(x_1 - x_2) + g(x_1), \tag{45}$$

where

$$\lim_{x_1 \to x_2} \frac{\|g(x_1)\|}{\|x_2 - x_1\|} = 0. \tag{46}$$

Equation (45) can be written as follows

$$y_1 - y_2 = Df(x_2) \cdot (f^{-1}(y_1) - f^{-1}(y_2)) + g(x_1), \tag{47}$$

hence

$$f^{-1}(y_1) - f^{-1}(y_2) = Df(x_2)^{-1}(y_1 - y_2) - Df(x_2)^{-1}g(x_1). \tag{48}$$

It remains to show that

$$\lim_{y_1 \to y_2} \frac{\|Df(x_2)^{-1}g(x_1)\|}{\|y_1 - y_2\|} = 0 \tag{49}$$

We have

$$\frac{\|Df(x_2)^{-1}g(x_1)\|}{\|y_1-y_2\|} \le \|Df(x_2)^{-1}\| \cdot \frac{\|g(x_1)\|}{\|y_1-y_2\|} = \|Df(x_2)^{-1}\| \cdot \frac{\|g(x_1)\|}{\|x_1-x_2\|} \cdot \frac{\|x_1-x_2\|}{\|y_1-y_2\|}.$$

Since (44) can be written as

$$\|x_1-x_2\| \le \frac{\|Df(x_0)^{-1}\|}{1-L}\|y_1-y_2\|$$

we obtain

$$\frac{\|Df(x_2)^{-1}g(x_1)\|}{\|y_1-y_2\|} \le \|Df(x_2)^{-1}\| \frac{\|Df(x_0)^{-1}\|}{1-L} \frac{g(x_1)}{\|x_1-x_2\|} \to 0$$

This proves that

$$Df^{-1}(y_2) = \left(Df(f^{-1}(y_2))\right)^{-1}. \tag{50}$$

The higher derivatives are established by induction by differentiation of both sides of (50). ∎

9 Implicit function theorem

Theorem 11 *Let $f : \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}^n$ be C^k with $k \ge 1$.*

Assume that (x_0, y_0) are such that $f(x_0, y_0) = 0$ and $\frac{\partial F}{\partial y}(x_0, y_0)$ is an isomorphism.

Then there exist open sets $U_x \subset \mathbb{R}^l$ and $U_y \subset \mathbb{R}^n$, $(x_0, y_0) \in U_x \times U_y$, and function $h : U_x \to U_y$ in C^k, such that

$$f(x, h(x)) = 0, \quad \forall x \in U_x, \tag{51}$$

and if $(x_1, y_1) \in U_x \times U_y$ are such that $f(x_1, y_1) = 0$, then $y_1 = h(x_1)$.

9.1 Proof by reduction to Theorem 10

A proof can be done by reduction to the local diffeomorphism theorem (Thm. 10) as follows. Consider a map $F(x, y) = (x, f(x, y))$. It is easy to see that for (x_0, y_0), $(x_0, 0)$ assumptions of Theorem 10 are satisfied, hence there exists the local inverse $G(x, z) = (x, g(x, z))$ and then $h(x) = g(x, 0)$.

9.2 Direct proof based on the Newton method

We want to solve equation

$$f(x, y) = 0 \tag{52}$$

for y with x as the parameter.

We set-up a modified Newton method which we know works in the neighborhood of $x = x_0$ and $y = y_0$. We expect that the contraction so obtained will work also for nearby values of x.

Let

$$N_x(y) = y - \frac{\partial f}{\partial y}(x_0, y_0)^{-1} f(x, y). \tag{53}$$

We would like to find $r > 0$, $R > 0$ and $0 \leq L < 1$, such that for all $(x, y) \in \overline{B}(x_0, r) \times \overline{B}(y_0, R)$ holds

$$\begin{aligned} N_x(\overline{B}(y_0, R)) &\subset B(y_0, R), & (54)\\ \left\| \frac{\partial}{\partial y} N_x(y) \right\| &\leq L < 1. & (55) \end{aligned}$$

We have

$$\frac{\partial}{\partial y} N_x(y) = I - \left(\frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} \frac{\partial f}{\partial y}(x, y),$$

let $R > 0$ and $r > 0$ be such that (55) holds for $L = 1/2$ and $\frac{\partial f}{\partial y}(x, y)$ is an isomorphism.

Observe that

$$\begin{aligned} N_{x_1}(y) - N_x(y) &= \left(\frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} (f(x, y) - f(x_1, y)), \\ \|N_{x_1}(y) - N_x(y)\| &\leq \left\| \left(\frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} \right\| \cdot M \cdot \|x - x_1\| \end{aligned}$$

where $M = \sup_{(x,y) \in \overline{B}(x_0, r) \times \overline{B}(y_0, R)} \left\| \frac{\partial f}{\partial x}(x, y) \right\|$.

Let us set

$$A = \left\| \left(\frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} \right\| \tag{56}$$

In order to obtain (54) we compute for $(x, y) \in \overline{B}(x_0, r) \times \overline{B}(y_0, R)$

$$\begin{aligned} \|N_x(y) - y_0\| = \|N_x(y) - N_{x_0}(y_0)\| \leq \\ \|N_x(y) - N_{x_0}(y)\| + \|N_{x_0}(y) - N_{x_0}(y_0)\| \leq \\ AM\|x - x_0\| + L\|y - y_0\| \leq AMr + LR. \end{aligned}$$

We need

$$AMr + LR < R, \tag{57}$$

hence $r < \frac{R(1-L)}{AM}$.

We obtain a solution of (52) for $x \in \overline{B}(x_0, R)$ which is unique for $y \in \overline{B}(y_0, r)$. This is our function $h(x)$.

Continuity: We apply the same approach as before. We have a continuous family of contractions, hence the fixed point should depend continuously on the parameter

$$\begin{aligned}\|h(x_1)-h(x_2)\| = \|N_{x_1}(h(x_1))-N_{x_2}(h(x_2))\| \le \\ \|N_{x_1}(h(x_1))-N_{x_1}(h(x_2))\| + \|N_{x_1}(h(x_2))-N_{x_2}(h(x_2))\| \le \\ L\|h(x_1)-h(x_2)\| + AM\|x_1-x_2\|\end{aligned}$$

hence

$$\|h(x_1)-h(x_2)\| \le \frac{AM\|x_1-x_2\|}{1-L}. \tag{58}$$

Differentiability:

Let us fix x_2.

We have from the differentiablity of f at $(x_2, h(x_2))$

$$0 = f(x_1,h(x_1)) - f(x_2,h(x_2)) = \tag{59}$$

$$\frac{\partial f}{\partial x}(x_2,h(x_2))(x_1-x_2) + \frac{\partial f}{\partial y}(x_2,h(x_2))(h(x_1)-h(x_2)) + g(x_1,h(x_1)),$$

$$\frac{\|g(x_1,h(x_1))\|}{\|x_1-x_2\|+\|h(x_1)-h(x_2)\|} \to 0, \quad (x_1,h(x_1)) \to (x_2,h(x_2)). \tag{60}$$

From this we obtain

$$\begin{aligned}h(x_1)-h(x_2) = -\left(\frac{\partial f}{\partial y}(x_2,h(x_2))\right)^{-1}\frac{\partial f}{\partial x}(x_2,h(x_2))(x_1-x_2) + \\ -\left(\frac{\partial f}{\partial y}(x_2,h(x_2))\right)^{-1} g(x_1,h(x_1)).\end{aligned}$$

Observe that

$$\begin{aligned}\frac{\left\|\left(\frac{\partial f}{\partial y}(x_2,h(x_2))\right)^{-1} g(x_1,h(x_1))\right\|}{\|x_1-x_2\|} \le \\ \left\|\left(\frac{\partial f}{\partial y}(x_2,h(x_2))\right)^{-1}\right\| \cdot \frac{\|g(x_1,h(x_1))\|}{\|x_1-x_2\|+\|h(x_1)-h(x_2)\|} \cdot \\ \cdot \frac{\|x_1-x_2\|+\|h(x_1)-h(x_2)\|}{\|x_1-x_2\|}\end{aligned}$$

in view of (60) and (58) implies that

$$\frac{\left\|\left(\frac{\partial f}{\partial y}(x_2,h(x_2))\right)^{-1} g(x_1,h(x_1))\right\|}{\|x_1-x_2\|} \to 0, \quad x_1 \to x_2$$

This means that

$$Dh(x) = -\left(\frac{\partial f}{\partial y}(x,h(x))\right)^{-1}\frac{\partial f}{\partial x}(x,h(x)). \tag{61}$$

Now the induction argument allow us to increase the regularity up to C^k.

∎