MA225 Differentiation

Thomas Reddington

May 6, 2013

Contents

Preface

These lecture notes are a projection of the MA225 Differentiation course 2012/2013, delivered by Dr Markus Kirkilionis at the University of Warwick. The up-to-date version of these notes should be found here:

`http://www.tomred.org/lecture-notes.html`

Markus' original handwritten script should be found here:

`http://lora.maths.warwick.ac.uk/groups/differentiation/wiki/39327/MA225_Lecture_Manuscript.html`

Students taking this course should also take a look at Alex Wendland's Dropbox notes:

`https://www.dropbox.com/sh/5m63moxv6csy8tn/HCmB8rY7va/Year%202/Differentiation`

These notes are, to my knowledge, complete (except from diagrams), but the tedious treasure hunt of errors will always be an open game. If you spot an error, or you want the source code to fiddle with the notes in your way, e-mail me at `me@tomred.org`. Writing these up has been a benefit to me (there aren't many foolproof ways to avoid proper work), but most of all I hope they're helpful, and good luck!
Tom ♡

The lecture will be split into two parts:

A Continuous functions from $\mathbb{R}^n \to \mathbb{R}^m$.

B Differentiable functions from $\mathbb{R}^n \to \mathbb{R}^m$

Let us start with A:

A Continuous functions from $\mathbb{R}^n \to \mathbb{R}^m$

A1 $\mathbb{R}^n$ as a normed vector space, convergence

The elements of $\mathbb{R}^n$ are ordered n-tuples of real numbers:

$$\mathbf{x} := (x_1, ..., x_n)$$

We have componentwise addition:

$$\mathbf{x} + \mathbf{y} := (x_n + y_n, ..., x_n + y_n)$$

and componentwise multiplication with scalars:

$$c\mathbf{x} := (cx_1, ..., cx_n)$$

This makes $\mathbb{R}^n$ a real-valued vector space of dimension n. The vectors:

$$\begin{aligned} \mathbf{e}_1 &:= (1, ..., 0) \\ &\vdots \\ \mathbf{e}_n &:= (0, ..., 1) \end{aligned}$$

form the *standard basis* of $\mathbb{R}^n$.

In one dimension concepts like convergence and continuity had to be discussed with the help of the *absolute value*, $|\cdot|$. In $\mathbb{R}^n$ we replace this with a *norm*, interpreted as a mapping from $\mathbb{R}^n$ to $\mathbb{R}$:

Definition **A1.1** A function $N : \mathbb{R}^n \to \mathbb{R}$ with:

1. $N(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$.
2. $N(c\mathbf{x}) = |c| \cdot N(\mathbf{x})$
3. $N(\mathbf{x} + \mathbf{y}) \leq N(\mathbf{x}) + N(\mathbf{y})$

is called a *norm* on $\mathbb{R}^n$.

For $n = 1$ we therefore must have $N(x) = a|x|$, with a some positive number. For $n > 1$ there are more possibilities.

Maximum norm:

$$\|\mathbf{x}\| = \max\{|x_1|, ..., |x_n|\}$$

Properties (1) and (2) are clear. To prove (3) we use the triangle inequality for the absolute value. For every i we have:

$$|x_i| \leq \|\mathbf{x}\| \text{ and } |y_i| \leq \|\mathbf{y}\|$$

and therefore:

$$|x_i + y_i| \leq |x_i| + |y_i| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

As there is at least one i for which $|x_i + y_i| = \|\mathbf{x} + \mathbf{y}\|$ we have:

$$\|x + y\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

Euclidean norm:

$$|\mathbf{x}| := \sqrt{x_1^2 + ... + x_n^2}$$

It can be expressed with the help of the *standard scalar product*:

$$\mathbf{x} \cdot \mathbf{y} := \langle \mathbf{x}, \mathbf{y} \rangle := x_1 y_1 + ... + x_n y_n$$

Therefore:

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

Again (1) and (2) are trivial. To prove (3) we use the *Schwarz inequality*:

$$(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})$$

This inequality can be proven by the fact that the quadratic polynomial:

$$(t\mathbf{x} + \mathbf{y}) \cdot (t\mathbf{x} + \mathbf{y}) = (\mathbf{x} \cdot \mathbf{x})t^2 + 2(\mathbf{x} \cdot \mathbf{y})t + (\mathbf{y} \cdot \mathbf{y})$$

cannot have two real roots, therefore the "discriminant" is:

$$(\mathbf{x} \cdot \mathbf{y})^2 - (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}) \leq 0$$

To *prove* (3), i.e.:

$$\sqrt{(\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y})} \leq \sqrt{\mathbf{x} \cdot \mathbf{x}} + \sqrt{\mathbf{y} \cdot \mathbf{y}}$$

we square and get:

$$(\mathbf{x} \cdot \mathbf{x}) + 2(\mathbf{x} \cdot \mathbf{y}) + (\mathbf{y} \cdot \mathbf{y}) \leq (\mathbf{x} \cdot \mathbf{x}) + 2\sqrt{(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})} + (\mathbf{y} \cdot \mathbf{y})$$

But this is equivalent to the Schwarz inequality!

The Euclidean norm can be generalised to the l_p-norm:

$$|\mathbf{x}|_p := (|x_1|^p + ... + |x_n|^p)^{\frac{1}{p}}$$

We have:

$$\begin{aligned} p = 1 &: \text{octaeder norm} \\ p = 2 &: \text{Euclidean norm} \\ &\vdots \\ p = \infty &: \text{maximum norm} \end{aligned}$$

We now look at convergence:

Definition **A1.2** The sequence $(\mathbf{x}_k)$ is called convergent against $\mathbf{a}$, if:

$$\lim_{k \to \infty} \|\mathbf{x}_k - \mathbf{a}\| = 0$$

One uses also the notation $\lim_{k \to \infty} \mathbf{x}_k = \mathbf{a}$. Like in one dimension one has to show that the limit $\mathbf{a}$ is unique. If there is such an $\mathbf{a}$, the sequence $(\mathbf{x}_k)$ is called convergent.

Theorem **A1.1** The sequence $(\mathbf{x}_k)$ is convergent if and only if all n component sequences are convergent.

Proof. We have the estimate:

$$|x_k^{(i)} - a^{(i)}| \leq \|\mathbf{x}_k - \mathbf{a}\| \leq |x_k^{(1)} - a^{(1)}| + ... + |x_k^{(n)} - a^{(n)}|$$

for every $i = 1, ..., n$. □

Like in $\mathbb{R}$ we have the *Cauchy criterion*:

The sequence $(\mathbf{x}_k)$ is convergent iff for every $\varepsilon > 0$ there exists k_0 such that for all $k \geq k_0$ and all $p > 0$ we have:

$$\|\mathbf{x}_{k+p} - \mathbf{x}_k\| < \varepsilon$$

Also the *theorem of Bolzano-Weierstrass* remains valid:

Theorem **A1.2** Every bounded sequence $(\mathbf{x}_k)$ (i.e. $\|\mathbf{x}_k\| < K$) has a convergent subsequence.

Proof. We look at all n components of $(\mathbf{x}_k)$. They are all bounded. Therefore from B-W in dimension 1 they have each a convergent subsequence. We look at component 1 and choose a subsequence of $(\mathbf{x}_k)$ such that $(x_k^{(1)})$ converges. Next we look at component 2 etc. until we reach n. □

We next need to make sure that instead of the maximum norm we could have chosen any norm N in all previous definitions. If we can show that:

$$N(\mathbf{x}) \leq a\|\mathbf{x}\| \text{ and } \|\mathbf{x}\| \leq bN(\mathbf{x})$$

then:

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{a}\| = 0 \Longleftrightarrow \lim_{k\to\infty} N(\mathbf{x}_k - \mathbf{a})$$

i.e. we are done and know independency on any specific norm.

Theorem **A1.3** There are positive numbers a and b such that for every norm N we have:

$$N(\mathbf{x}) \leq a\|\mathbf{x}\| \text{ and } \|\mathbf{x}\| \leq bN(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. The estimate left follows from (1), (2) and (3) and $|x_i| \leq \|\mathbf{x}\|$ from:

$$\begin{aligned} N(\mathbf{x}) &= N(x_1\mathbf{e}_1 + ... + x_n\mathbf{e}_n) \\ &\leq N(\mathbf{e}_1)|x_1| + ... + N(\mathbf{e}_n)|x_n| \\ &\leq (N(\mathbf{e}_1) + ... + N(\mathbf{e}_n))\|\mathbf{x}\| \end{aligned}$$

We can choose $a := N(\mathbf{e}_1) + ... + N(\mathbf{e}_n)$. The estimate right is proven indirect. Assume there is no such $b > 0$. Then one can find for each $b = 1, 2, 3, ..., k, ...$ a vector $\mathbf{x}_k$ such that:

$$\|\mathbf{x}_k\| > kN(\mathbf{x}_k)$$

holds.

With (2) we could have for $k = 1, 2, 3, ...$:

$$N\left(\frac{\mathbf{x}_k}{\|\mathbf{x}_k\|}\right) < \frac{1}{k}$$

For $\mathbf{y}_k := \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|}$ we then have $N(\mathbf{y}_k) < \frac{1}{k}$. This will create a contradiction. Because $\|\mathbf{y}_k\| = 1$ the sequence $(\mathbf{y}_k)$ is bounded. Therefore there is a subsequence $(\mathbf{z}_k)$ with limit z, i.e.:

$$\lim_{z\to\infty} \|\mathbf{z}_k - \mathbf{z}\| = 0$$

because $N(\mathbf{z} - \mathbf{z}_k) \leq a\|\mathbf{z} - \mathbf{z}_k\|$ we have:

$$\begin{aligned} N(\mathbf{z}) &= N(\mathbf{z} - \mathbf{z}_k + \mathbf{z}_k) \\ &\leq N(\mathbf{z} - \mathbf{z}_k) + N(\mathbf{z}_k) \\ &\leq a\|\mathbf{z} - \mathbf{z}_k\| + N(\mathbf{z}_k) \end{aligned}$$

Because $N(\mathbf{y}_k) < \frac{1}{k}$ we have for $(\mathbf{z}_k)$ $\lim_{k\to\infty} N(\mathbf{z}_k) = 0$. Therefore $N(\mathbf{z}) = 0$, i.e. we would conclude $\mathbf{z} = 0$. On the other hand we have:

$$\mathbf{z}_k = \mathbf{z} + (\mathbf{z}_k - \mathbf{z}) \text{ and } \|\mathbf{z}_k\| = 1$$

Therefore:

$$1 \leq \|\mathbf{z}\| + \|\mathbf{z}_k - \mathbf{z}\|$$

So $\|\mathbf{z}\| > 0$, and therefore $\mathbf{z} \neq 0$. This proves:

$$\|\mathbf{x}\| \leq bN(\mathbf{x})$$

for some positive b and for all $\mathbf{x} \in \mathbb{R}^n$. □

This theorem should be generalised:

Definition **A1.3** A norm N' is called equivalent to norm N if there are positive numbers a, b such that for all $\mathbf{x} \in \mathbb{R}^n$ we have:

$$aN(\mathbf{x}) \leq N'(\mathbf{x}) \leq bN(\mathbf{x})$$

We have:

Theorem **A1.4** Every pair of norms is equivalent on $\mathbb{R}^n$.

Proof. We have already shown that every norm N is equivalent to the maximum norm. We have to check equivalence of norms forms and equivalence relationship:

1. Obviously every norm is equivalent to itself. [Reflexivity]
2. From $aN(\mathbf{x}) \leq N'(\mathbf{x}) \leq bN(\mathbf{x})$ follows:

$$\frac{1}{b}N'(\mathbf{x}) \leq N(\mathbf{x}) \leq \frac{1}{a}N'(\mathbf{x}) \text{ [Symmetry]}$$

3. Given $aN(\mathbf{x}) \leq N'(\mathbf{x}) \leq bN(\mathbf{x})$ and $a'N'(\mathbf{x}) \leq N''(\mathbf{x}) \leq b'N'(\mathbf{x})$, we get:

$$aa'N(\mathbf{x}) \leq N''(\mathbf{x}) \leq bb'N(\mathbf{x}) \text{ [Transitivity]}$$

For example the maximum norm and Euclidean norm are equivalent. It holds:

$$\|\mathbf{x}\| \leq |\mathbf{x}| \leq \sqrt{n}\|\mathbf{x}\| \qquad \text{(check!)}$$

□

Infinite sums:

Like in $\mathbb{R}$ an infinite sum (row) is defined as a sequence $(\mathbf{s}_k)$ with:

$$\mathbf{s}_k := \sum_{j=1}^{k} \mathbf{x}_j$$

If $\lim_{k\to\infty} \mathbf{s}_k$ exists, the limit is called the *sum* $\mathbf{s}$:

$$\mathbf{s} := \sum_{k=1}^{\infty} \mathbf{s}_k$$

The infinite sum $(\mathbf{s}_k)$ is called *absolutely convergent* if:

$$\sum_{k=1}^{\infty} \|\mathbf{x}_k\|$$

converges. Every infinite sum which is *absolutely convergent* is also convergent (uses Cauchy criterion):

$$\left\| \sum_{j=k+1}^{k+p} \mathbf{x}_j \right\| \leq \sum_{j=k+1}^{k+p} \|\mathbf{x}_j\|$$

As in $\mathbb{R}$ absolutely convergent sums (rows) have the same limit independent of the order of the partial sums.

A2 Topology of $\mathbb{R}^n$

We will define neighbourhoods of $\mathbb{R}^n$ with the help of norms.

Definition A2.1 Let $\varepsilon > 0$ and $\mathbf{a} \in \mathbb{R}^n$. The set

$$U(\mathbf{a}, \varepsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{a}\| < \varepsilon\}$$

is called *ε-neighbourhood of* $\mathbf{a}$ or *ball around* $\mathbf{a}$ *with diameter* ε. Every set containing $U(\mathbf{a}, \varepsilon)$ is called *neighbourhood of* $\mathbf{a}$ (with respect to the maximum norm).
Visualisation:

If we take the Euclidean norm we get regular balls around $\mathbf{a}$. The equivalence of norms is geometrically equivalent to the ability to nest ε-neighbourhoods:

The neighbourhood definition is therefore independent of the choice of any specific norm. Like in $\mathbb{R}$ we need to define some basic topological properties:

- $\mathbf{x} \in \mathbb{R}^n$ is called an *interior point* of $M \subset \mathbb{R}^n$, if there is a neighbourhood of $\mathbf{x}$ contained in M.
- $\mathbf{x} \in \mathbb{R}^n$ is called a *limit point* of $M \subset \mathbb{R}^n$, if in every neighbourhood of $\mathbf{x}$ there is a point in M different to $\mathbf{x}$.
- $\mathbf{x} \in \mathbb{R}^n$ is called a *boundary point* of $M \subset \mathbb{R}^n$, if in every neighbourhood of $\mathbf{x}$ there is at least one point in M and at least one point in $\complement M$, the complement of M.

Remark: The definition of convergence can be made now with the help of neighbourhoods. The fact that a sequence can have only one limit point follows from the property that in $\mathbb{R}^n$ two different points have disjunct neighbourhoods. $\mathbb{R}^n$ is a *Hausdorff space.*

- A set M consisting only of interior points is called *open.*
- If M incorporates all its limit points it is called *closed.*

We have:

$$\begin{aligned} M \text{ open } &\implies \complement M \text{ closed} \\ M \text{ closed } &\implies \complement M \text{ open} \end{aligned}$$

Only $\emptyset$ and $\mathbb{R}^n$ are at the same time open and closed sets in $\mathbb{R}^n$.

Theorem A2.1 The union of open sets in $\mathbb{R}^n$ is open. The dissection of *finitely* many open sets is open.

Proof. Follows directly from definition. □

Theorem A2.2 The dissection of closed sets is closed. The union of fintely many closed sets is closed.

Example of a closed set:

$$[\mathbf{a}, \mathbf{b}] := \{x : a_i \leq x_i \leq b_i, i = 1, ..., n\}$$

A complex example of an open set in $\mathbb{R}^2$:

- Cut out the open cube of the unit cube Q from the centre with boundary interval length $\frac{1}{3}$ (see figure).
- Do the same construction again in all remaining cubes not yet covered on the respective local scale.
- Proceed to infinity.

Let us denote all area of Q not covered during the construction by M. Then $Q \setminus M = C$, with C closed. C is called the *Cantor* set.

- C is the set of boundary points of M.
- Every point in Q is a limit point of M (check!).

Compact sets:

Theorem **A2.3** Let $M \subset \mathbb{R}^n$. The following statements are equivalent:

1. M is bounded and closed.
2. Every cover of M with open sets allows the choice of finitely many of them, such that M is covered (Heine-Borel property).
3. Every infinite subset of M has a limit point in M.

Proof. We show (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1).

To show (1) $\Longrightarrow$ (2) we use an interaction. Because M is bounded there exists a closed cube W which contains M. Assume there is an open cover C of M, which does *not* have the Heine-Borel property. Next we can divide W into 2^n closed cubes which all have half of the length of edges of previous iterations on W (see figure for dim = 2).
With our assumption there is at least one such cube such that its dissection with M cannot be covered with finitely many sets from C, the cover. We choose such a cube and call it $W^{(1)}$. Following the construction we can derive a sequence $(W^{(k)})$ of closed cubes such that:

(i) $W \supset W^{(1)} \supset ... \supset W^{(k)} \supset ...$,

(ii) $\lim_{k\to\infty} \delta(W^{(k)}) = 0$ (with $\delta(\cdot)$ being the diameter of the cube),

(iii) $M \cap W^{(k)}$ cannot be covered with finitely many sets from C.

The remainder of the proof will be given after the following theorem. □

The sequence $M \cap W^{(k)}$ satisfies the condition of the theorem of Cantor:

Theorem **A2.4** (Cantor) Let (A_k) be a decreasing sequence of bounded closed sets, which diameters $\delta(A_k)$ converge to zero, i.e.

$$A_1 \supset A_2 \supset ... \supset A_k \supset ... \text{ and } \lim_{k\to\infty} \delta(A_k) = 0$$

Then there is exactly one point $\mathbf{x}$ which belongs to all sets A_k:

$$\bigcap_{k=1}^{\infty} A_k = \{\mathbf{x}\}$$

Proof (Cantor). Choose an arbitrary point $\mathbf{x}_k$ from each set A_k. One obtains a Cauchy-sequence because $A_{k+p\subset A_k}$ for every $p \geq 0$ and:

$$\|\mathbf{x}_{k+p} - \mathbf{x}_k\| \leq \delta(A_k)$$

Therefore $(\mathbf{x}_k)$ converges. Let us call the limit $\mathbf{x}$. Every other possibly chosen sequence $(\mathbf{y}_k)$ also converges to $\mathbf{x}$ because:

$$\|\mathbf{x}_k - \mathbf{y}_k\| \leq \delta(A_k)$$

$\mathbf{x}$ is therefore *unique*, but it is also an *element* of *all* A_k. If not a k_0 would exist such that $\mathbf{x} \notin A_{k_0}$, and $x \notin A_{\mathbf{k}_0+p}$, with $p \geq 0$. There would be a neighbourhood of $\mathbf{x}$ not contained in A_{k_0+p}, $p \geq 0$. This *contradicts* that we proved $(\mathbf{x}_k)$ converges to $\mathbf{x}$. □

Proof of A2.3 *(continued).* We know now there is exactly one point $\mathbf{x}$ belonging to all sets $M \cap W^{(k)}$, and therefore $\mathbf{x} \in M$ as well. By assumption (W has a nonempty open interior!) there is an open set $O \in C$ with $\mathbf{x} \in O$. For all sufficiently large k the cubes satisfy $W^{(k)} \subset O$, as $\mathbf{x}$ is an interior point of O. To cover $M \cap W^{(k)}$ we need only a single open set $O \in C$, whereas by (iii) we would need infinitely many sets from C. This is a contradiction, so (1) $\Longrightarrow$ (2) is proven.

We next prove (2) $\Longrightarrow$ (3), again indirectly. Assume M has an infinite subset (a set with infinitely many points) A, which has *no* limit points in M ("$\neg$(3)"). Every point in A possesses an open ball which has no further elements in A. Also, every point in $M \setminus A$ possesses an open ball around it with no elements in A. The set of all these open balls forms a cover of M. Because M has the Heine-Borel property finitely many open balls are sufficient to cover M. Therefore they also form a cover of A. Because each of these balls contains only one element of A, A must be finite (only contains finitely many points). This contradicts the assimption that A is infinite, so (2) $\Longrightarrow$ (3) is proven.

To prove (3) $\Longrightarrow$ (1) we assume M is *not* bounded. Then for every $k \in \mathbb{N}$ there exists $\mathbf{x}_k \in M$ with:

$$\|\mathbf{x}_k\| \geq k$$

Assemble all $\mathbf{x}_k$, $k = 1, 2, ...$ in a set A. This set contains no limit point $\mathbf{y}$, otherwise infinitely many elements of A would be in a 1-neighbourhood (ε=1!) of $\mathbf{y}$, in contradiction to the construction. Therefore M is *bounded*. Assume M is not closed. Then there would be a limit point $\mathbf{x}$ of M which does not belong to M. For this $\mathbf{x}$ and every natural number $k \geq 1$ we would have a point $\mathbf{x}_k \in M$ such that:

$$\|\mathbf{x}_k - \mathbf{x}\| < \frac{1}{k}$$

The infinite but *countable* set A has exactly one limit point in $\mathbb{R}^n$, and this is $\mathbf{x}$. By assumption (3) this limit point of $A \subset M$ belongs to M: contradiction! Therefore (3) $\Longrightarrow$ (1) is proven. □

Definition **A2.1** A set $M \subset \mathbb{R}^n$ is called *compact* if it satisfies either (1), (2) or (3).

Remark: In infinite-dimensional vector spaces (1) is *not* equivalent to (2) or (3)!

Another property of M equivalent to (1) is:

4. Let $M \subset \mathbb{R}^n$. Every sequence with values in M possesses a convergent subsequence with a limit (point) in M.

We show equivalence with (3):

Proof. If (3) holds consider first a sequence with finitely many different values in M. Then most of the values of this sequence are *constant*, i.e. convergent. If this sequence has infinitely many different values in M then by (3) it has a limit point in M. Now let (4) be satisfied and let B be an arbitrary infinite subset of M. First choose a countable infinite subset A of B. Let the values of A be assembled as a sequence. As these values belong to M this sequence by property (4) has a convergent subsequence with limit in M. This limit is a limit point of A, and therefore of M. □

As a simple consequence we get a generalised version of the theorem of Bolzano-Weierstrass.

Theorem **A2.5** (Bolzano-Weierstrass) Every infinite bounded subset M of $\mathbb{R}^n$ possesses at least one limit point in $\mathbb{R}^n$.

Proof. M is contained in a compact cube. Because of (3) M has a limit point in this cube. □

A3 Continuous functions

We already know some examples of functions or mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$, or $\mathbb{N}$ to $\mathbb{R}^n$:

$$\begin{aligned} \|\cdot\| &: \mathbb{R}^n \to \mathbb{R} \\ (\mathbf{x}_k) &: \mathbb{N} \to \mathbb{R}^n \end{aligned}$$

Let D be some subset of $\mathbb{R}^n$. The "D" denotes "domain of definition". We write:

$$\begin{aligned} f &: D \to \mathbb{R}^m \\ \mathbf{x} &\mapsto f(\mathbf{x}) \qquad (\mathbf{x} \in D) \end{aligned}$$

We can expand this short compact notation:

$$\begin{aligned} y_1 &= f_1(x_1, ..., x_n) \\ &\vdots \qquad\qquad (\mathbf{x} \in D, \mathbf{y} \in \mathbb{R}^m) \\ y_m &= f_m(x_1, ..., x_n) \end{aligned}$$

Especially simple are linear mappings:

1. $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$
2. $f(c\mathbf{x}) = cf(\mathbf{x})$ $(c \in \mathbb{R})$

Linear mappings from $\mathbb{R}^n \to \mathbb{R}^m$ will become very important as the key idea of *differentiation* is to approximate a non-linear function by a linear function or mapping. For linear mappings we have:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{aligned}$$

As a short notation we can therefore use the matrix $A = (a_{ij})$, so instead of f use the symbol A. A linear map A is always defined on all of $\mathbb{R}^n$. The *image* of A is always a sub-vector space of $\mathbb{R}^m$. A is *injective* if and only if the image of A has dimension n. We therefore must have $n \le m$ in this case.

A linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ can only be *bijective* if $n = m$.

An algebraic condition for A being bijective is that $\det(A) \neq 0$, with $\det(\cdot)$ being the *determinant* of a matrix. We have:

$$\det(A) := \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

Here σ "runs" through all permutations of $\{1, ..., n\}$, the symmetric group S_n. $\operatorname{sign}(\sigma)$ is either 1 or -1 depending on whether σ is an even or odd permutation. Instead of $\det(A)$ we often write:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

We have:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

With $\det(\cdot)$ we have another example of a map:

$$\det(\cdot) : \mathbb{R}^{n^2} \to \mathbb{R}^1$$

Let now $m = 1$ and $n = 2$. Then for example:

$$y = \sqrt{1 - x_1^2 - x_2^2} \qquad (x_1^2 + x_2^2 \leq 1 \text{ defines } D!)$$

maps the unit disk $D \subset \mathbb{R}^2$ to $\mathbb{R}^1$, the result is the upper half of the unit ball in $\mathbb{R}^3$.

So $D = \{\mathbf{x} \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ and I, the image is:

$$I = [0, 1],$$
$$f : D \to I$$

The set:

$$L_c := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\},$$

With $f : \mathbb{R}^n \to \mathbb{R}$ is called the *level set* (for value c). In our upper ball example the set L_1 consists of any single point i.e. the origin $\{\mathbf{0} = (0, 0)\}$. For $0 < c < 1$ the level sets are circles. An important case of mappings $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m = n$ are "new coordinates". For example:

$$x = f(r, \varphi) = r \cdot \cos \varphi$$
$$y = g(r, \varphi) = r \cdot \sin \varphi$$

describes the transition to polar coordinates in $\mathbb{R}^2$. Even more important are *vector fields*: for every point $x \in D \subset \mathbb{R}$, $f(x) \in \mathbb{R}^n$ defines a vector of length $\|f(x)\|$ located at x:

Continuity:

Definition **A3.1** Let $D \subset \mathbb{R}^n$ and $f : D \to \mathbb{R}^m$. Then f is called *continuous* in $\mathbf{a} \in D$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $\mathbf{x} \in D$ with $\|\mathbf{x} - \mathbf{a}\| < \delta$ we have:

$$\|f(\mathbf{x}) - f(\mathbf{a})\| < \varepsilon$$

Remark: It is not important which norms we use, of course both norms appearing in the definition could be different.

Alternative definition of continuity:

Definition **A3.2** f is called continuous in $\mathbf{a} \in D$ if for every neighbourhood V of $f(\mathbf{a})$ there is a neighbourhood U of $\mathbf{a}$ such that:

$$f(U \cap D) \subset V$$

If f is continuous it maps a sequence $(\mathbf{x}_k) \in D$ (with $(\mathbf{x}_k)$ convergent to $\mathbf{a}$) to a sequence $(f(\mathbf{x}_k))$, convergent to $f(\mathbf{a})$ (check!).

Let $\mathbf{a}$ now be a limit point of D. We write:

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = \mathbf{b}$$

If there is for every $\varepsilon > 0$ a $\delta > 0$ such that:

$$\mathbf{x} \in D, \mathbf{x} \neq \mathbf{a}, \|\mathbf{x} - \mathbf{a}\| < \delta \implies \|f(\mathbf{x}) - \mathbf{b}\| < \varepsilon$$

Remark: $\mathbf{a}$ needs not to be an element of D_1. If $\mathbf{a} \in D$ then $f(\mathbf{a})$ does play no role in the definition.

We have therefore another definition of continuity:

$$f \ \textit{continuous} \iff \lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$$

This is often written as:

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f\left(\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{x}\right)$$

All the previous examples of functions f have been continuous. Example of f being *not* continuous: Divide $\mathbb{R}^n$ into rational and non-rational points. If $x_1, ..., x_n \in \mathbb{Q}$ then define $f(\mathbf{x}) = 0$ otherwise $f(\mathbf{x}) = 1$ (i.e. for $x_1, ..., x_n \notin \mathbb{Q}$).

Consider:

$$f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2} & \text{for } x^2 + y^2 > 0 \\ 0 & \text{for } x = 0, y = 0 \end{cases}$$

Then f is continuous on $\mathbb{R}^2$ with the exception of $(0, 0)$. Consider first the x and y axis, here f is 0 (so could be continuous...), but in a neighbourhood of $(0, 0)$ we find:

$$f(x, x) = \frac{2x^2}{x^2 + x^2} = 1 \qquad (x \neq 0)$$

This example shows we can have "partial continuity".

B Differentiable functions from $\mathbb{R}^n \to \mathbb{R}^m$

B1 Definition of the derivative

We consider a general function $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ *open*. That D is open will be crucial for the following.

Definition B1.1 The mapping $f : D \to \mathbb{R}^m, D \subset \mathbb{R}^n$ open, is called *differentiable at point* $\mathbf{p} \in D$ if there is a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ and a mapping $R : D \to \mathbb{R}^m$, R continuous at $\mathbf{p}$, such that for all $\mathbf{x} \in D$ it holds that:

$$f(\mathbf{x}) = f(\mathbf{p}) + A(\mathbf{x}-\mathbf{p}) + R(\mathbf{x})\|\mathbf{x}-\mathbf{p}\| \text{ and } R(\mathbf{p}) = 0$$

Remark: If a proof is required whether a certain function f is differentiable at $\mathbf{p}$ one has first to find a linear map A. Next one has to show that the mapping $R(\mathbf{x})$ defined on a neighbourhood of $\mathbf{x}$ by:

$$R(\mathbf{x}) = \frac{f(\mathbf{x}) - f(\mathbf{p}) - A(\mathbf{x}-\mathbf{p})}{\|\mathbf{x}-\mathbf{p}\|}$$

satisfies $\lim_{\mathbf{x}\to\mathbf{p}} R(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x} \neq \mathbf{p}$.

Remark: The linear map A is defined *uniquely*. Assume there would be two such mappings A_1 and A_2, associated to R_1 and R_2 respectively. Then:

$$f(\mathbf{x}) = f(\mathbf{p}) + A_1(\mathbf{x}-\mathbf{p}) + R_1(\mathbf{x})\|\mathbf{x}-\mathbf{p}\|,$$
$$f(\mathbf{x}) = f(\mathbf{p}) + A_2(\mathbf{x}-\mathbf{p}) + R_2(\mathbf{x})\|\mathbf{x}-\mathbf{p}\|,$$

also:

$$(A_1 - A_2)(\mathbf{x}-\mathbf{p}) = (R_2(\mathbf{x}) - R_1(\mathbf{x}))\|\mathbf{x}-\mathbf{a}\|$$

and:

$$\lim_{\mathbf{x}\to\mathbf{a}} R_1(\mathbf{x}) = \mathbf{0} \text{ and } \lim_{\mathbf{x}\to\mathbf{a}} R_2(\mathbf{x}) = \mathbf{0}$$

Consider a vector $\mathbf{h}$ and a small $t > 0$ such that $t\mathbf{h} = \mathbf{x}-\mathbf{p}$, i.e.

$$(A_1 - A_2)\mathbf{h} = (R_2(\mathbf{p}+t\mathbf{h}) - R_1(\mathbf{p}+t\mathbf{h}))\|\mathbf{h}\|$$

Because $\lim_{t\to 0}(R_2(\mathbf{p}+t\mathbf{h}) - R_1(\mathbf{p}+t\mathbf{h})) = \mathbf{0}$ we have:

$$(A_1 - A_2)\mathbf{h} = \mathbf{0}. \qquad (\mathbf{h} \in \mathbb{R}^n)$$

Therefore $A_1 = A_2$!

Definition B1.2 If f is differentiable in $\mathbf{p} \in D$ (which means $f(\mathbf{x}) = f(\mathbf{p}) + A(\mathbf{x}-\mathbf{p}) + R(\mathbf{x})\|\mathbf{x}-\mathbf{p}\|$ with $\lim_{\mathbf{x}\to\mathbf{p}} R(\mathbf{x}) = 0$), then the linear mapping A is called the *derivative* of f at $\mathbf{p}$. We write $\boxed{A = f'(\mathbf{p})}$.

The vector $f'(\mathbf{p})\mathbf{h} \in \mathbb{R}^m$ $(\mathbf{h} \in \mathbb{R}^n)$ is sometimes written in the form $f'(\mathbf{p}, \mathbf{h})$.

The derivative for $f : \mathbb{R}^n \to \mathbb{R}$ (i.e. $m = 1$)

If f is real-valued, then f' is a *linear form*. In this case the derivative is also called the *gradient* and denoted either:

$$\text{grad} f(\mathbf{p}) \text{ or } \nabla f(\mathbf{p})$$

We then obtain:

$$f'(\mathbf{p}, \mathbf{h}) = \langle \nabla f(\mathbf{p}), \mathbf{h} \rangle \qquad (\forall \mathbf{h} \in \mathbb{R}^n)$$

Because by definition:

$$\begin{cases} f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) = \langle \nabla f(\mathbf{p}), \mathbf{h} \rangle, \\ \lim_{\mathbf{h}\to\mathbf{0}} R(\mathbf{p}+\mathbf{h}) = \mathbf{0}, \end{cases}$$

the increment $f(\mathbf{p}+\mathbf{h})-f(\mathbf{p})$ for sufficiently small $\mathbf{h}$ is well approximated by the scalar product $\langle \nabla f(\mathbf{p}), \mathbf{h}\rangle$. We obtain the following interpretation:

$\nabla f(\mathbf{b})$ points in direction of the largest increment of f at location $\mathbf{p}$. The scalar product $\langle \nabla f(\mathbf{p}), \mathbf{h}\rangle$ (as a function of $\mathbf{h}$) has its largest value if $\mathbf{h}$ is a positive multiple of $f(\mathbf{p})$.

The function values of f can be illustrated if one looks at the level sets:

$$L_c = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c, c \in \mathbb{R}\}$$

Vectors $\mathbf{h}$ orthogonal to $\nabla f(\mathbf{p})$ lead to points in the neighbourhood of $\mathbf{p}$ which function values only differ minimally from $f(\mathbf{p})$.

We next look at the *graph* of a function $f : D \to \mathbb{R}$, $D \in \mathbb{R}^n$. The graph of f builds a *hypersurface* in $\mathbb{R}^{n+1}$, at each $\mathbf{x} = (x_1, ..., x_n) \in D$ there is a $\mathbf{y} = f(x_1, ..., x_n)$. In a neighbourhood of $\mathbf{p} \in D$ this hypersurface (graph of f) can be approximated by the hyperplane:

$$\mathbf{y} = f(\mathbf{p}) + \langle \nabla f(\mathbf{p}, \mathbf{x} - \mathbf{p})\rangle$$

We will call this hyperplane the *tangential hyperplane* of f at location p. The vector:

$$(-\nabla f(\mathbf{p}), 1) \in \mathbb{R}^{n+1}$$

is a *normal vector* on the tangential hyperplane (and therefore at the hypersurface) at location $(\mathbf{p}, f(\mathbf{p}))$

Partial derivatives

How can we determine $f'(\mathbf{p})$ of f differentiable at point $\mathbf{p} \in D$? The linear mapping $f'(\mathbf{p})$ is determined if we fix a basis $\mathbf{v}_1, ..., \mathbf{v}_n$ of $\mathbb{R}^n$. Let us consider $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$. We like to compute $f'(\mathbf{p}, \mathbf{v})$, and set $\mathbf{x} = \mathbf{p} + t\mathbf{v}$, $t \in \mathbb{R}$. So we get:

$$f(\mathbf{p} + t\mathbf{v}) = f(\mathbf{p}) + f'(\mathbf{p}, t\mathbf{v}) + R(\mathbf{p} + t\mathbf{v})|t| \cdot \|\mathbf{v}\|$$

For $t \neq 0$ we get:

$$f'(\mathbf{p}, \mathbf{v}) = \frac{f(\mathbf{p} + t\mathbf{v}) - f(\mathbf{p})}{t} - R(\mathbf{p} + t\mathbf{v})\frac{|t|}{t}\|\mathbf{v}\|$$

and:

$$f'(\mathbf{p}, \mathbf{v}) = \lim_{t \to 0} \frac{f(\mathbf{p} + t\mathbf{v}) - f(\mathbf{p})}{t}$$

The expression right of the limit is called *directional derivative*. $D_{\mathbf{v}} f(\mathbf{p})$ of f at location $\mathbf{p}$ with respect of $\mathbf{v}$. If we choose for $\mathbf{v}$ one of the standard basis vectors $\mathbf{e}_1, ..., \mathbf{e}_n$ of $\mathbb{R}^n$, then the respective directional derivatives are called *partial derivatives*. There are different notations for partial derivatives in the literature. The i-th partial derivative can be written:

$$D_i f(\mathbf{x}), \frac{\partial f}{\partial x_i}(\mathbf{x}), \text{ or } f_{x_i}(\mathbf{x})$$

although the last one should be abandoned. Instead one sometimes uses just $f_i(\mathbf{x})$, but this can lead to misunderstandings if f is not real-valued anymore. The task to compute partial derivatives can be solved by means of differential calculus in one variable. All arguments besides the i-th argument remain fixed. We have to compute:

$$\lim_{x_i \to p_i} \frac{f(p_1, ..., x_i, .., p_n) - f(p_1, ..., p_i, ..., p_n)}{x_i - p_i}$$

This limit can exist even if our function is not differentiable in P according to our definition. If some function $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ is differentiable at $\mathbf{p}$ then *all* partial derivatives exist. We have for $\mathbf{h} = \sum_{i=1}^n h_i \mathbf{e}_i$:

$$f'(\mathbf{p}, \mathbf{h}) = \sum_{i=1}^n f'(\mathbf{p}, \mathbf{e}_i) \cdot h_i = \sum_{i=1}^n D_i f(\mathbf{p}) \cdot h_i$$

If the vectors $D_i f(\mathbf{p})$ are determined with the help of the standard basis of $\mathbb{R}^m$, i.e. one dissects f into the components $f_1, ..., f_m$ then $f'(\mathbf{p})$ can be written as the $m \times n$ matrix:

$$Jf(\mathbf{p}) := f'(\mathbf{p}) = \begin{pmatrix} D_1 f_1(\mathbf{p}) & D_2 f_1(\mathbf{p}) & \cdots & D_n f_1(\mathbf{p}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathbf{p}) & D_2 f_m(\mathbf{p}) & \cdots & D_n f_m(\mathbf{p}) \end{pmatrix}$$

This matrix is also called the *Jacobian*-matrix. Now differentiability can be checked with partial derivatives. First the existence of all partial derivatives $\frac{\partial f_j}{\partial x_i}(\mathbf{p}) = D_i f_j(\mathbf{p})$ must be assured. Then one checks that the matrix $Jf(\mathbf{p})$ satisfies:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - Jf(\mathbf{p})\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$

Important: The existence of the partial derivatives of f at $\mathbf{p}$ alone does *not* imply differentiability of f at $\mathbf{p}$!

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, with:

$$f_1(x_1, x_2) = x_1^2 - x_2^2$$
$$f_2(x_2, x_2) = 2x_1 x_2$$

Then $Jf(\mathbf{p})$ becomes:

$$Jf(\mathbf{p}) = \begin{pmatrix} 2p_1 & -2p_2 \\ 2p_2 & 2p_1 \end{pmatrix}$$

We compute $R(\mathbf{h})$ related to $Jf(\mathbf{p})$ component-wise:

$$R_1(\mathbf{h}) = \frac{(p_1+h_1)^2 - (p_2+h_2)^2 - p_1^2 - p_2^2 - 2p_1h_1 + 2p_2h_2}{\|\mathbf{h}\|} = \frac{h_1^2 - h_2^2}{\|\mathbf{h}\|}$$
$$R_2(\mathbf{h}) = \frac{2(p_1+h_1)(p_2-h_2) - 2p_1p_2 - 2p_2h_1 - 2p_1h_2}{\|\mathbf{h}\|} = \frac{2h_1h_2}{\|\mathbf{h}\|}$$

Using the maximum norm we get:

$$|\tau_1(\mathbf{h})| \leq 2\|\mathbf{h}\|$$
$$|\tau_2(\mathbf{h})| \leq 2\|\mathbf{h}\|$$

i.e.

$$\lim_{h\to 0} R_1(\mathbf{h}) = \lim_{h\to 0} R_2(\mathbf{h}) = 0$$

It follows f is differentiable at $\mathbf{p}$. The derivative is:

$$f'(\mathbf{p}) = \begin{pmatrix} 2p_1 & -2p_2 \\ 2p_2 & 2p_1 \end{pmatrix} = Jf(\mathbf{p})$$

Definition **B1.3** If $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ is differentiable for all $\mathbf{p} \in D$, D open, then f is called *differentiable in D*. If the derivative $f'(\mathbf{p})$ is continuous for all $p \in D$ then f is called *continuously differentiable in D*. We write:

$$f \in \mathcal{C}^1(D, \mathbb{R}^m)$$

i.e. $\mathcal{C}^1(D, \mathbb{R}^m)$ is the set of all functions $f : D \to \mathbb{R}^m$ which are continuously differentiable. $\mathcal{C}^1(D, \mathbb{R}^m)$ is a *function space* (of infinite dimension!).

B2 Compositions of differentiable functions

We first show that for $f : D \to \mathbb{R}$ there is a linear map $L(\mathbf{h})$ such that:

$$R(\mathbf{p}+\mathbf{h})\|\mathbf{h}\| = \langle L(\mathbf{h}), \mathbf{h}\rangle \text{ and } \lim_{\mathbf{h}\to\mathbf{0}} L(\mathbf{h}) = 0 \tag{B2.1}$$

The definition of differentiability gives us:

$$\begin{cases} f(\mathbf{p}+\mathbf{h}) = f(\mathbf{p}) + f'(\mathbf{p},\mathbf{h}) + R(\mathbf{p}+\mathbf{h})\|\mathbf{h}\| \\ \lim_{\mathbf{h}\to\mathbf{0}} R(\mathbf{p}+\mathbf{h}) = 0 \end{cases}$$

Let us try (ansatz):

$$L(\mathbf{h}) = \lambda(\mathbf{h})\cdot\mathbf{h} = \begin{pmatrix} \lambda h_1 \\ \vdots \\ \lambda h_n \end{pmatrix}\cdot\mathbf{h} \qquad (\lambda \in \mathbb{R})$$

Assume WLOG that $\|\cdot\|$ is the *Euclidean norm.* In this case $\|\mathbf{h}\|^2 = \mathbf{h}\cdot\mathbf{h}$. For $\mathbf{h}\neq\mathbf{0}$ we get:

$$L(\mathbf{h}) = \frac{R(\mathbf{p}+\mathbf{h})}{\|\mathbf{h}\|}\mathbf{h}$$

and from this:

$$\lim_{\mathbf{h}\to\mathbf{0}} L(\mathbf{h}) = 0$$

Therefore $L(\mathbf{h})$ indeed satisfies (B2.1). If now $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ then we obtain a generalised result for $L(\mathbf{h})$:

$$\begin{cases} R(\mathbf{p}+\mathbf{h})\|\mathbf{h}\| = L(\mathbf{h})\mathbf{h} \\ \lim_{\mathbf{h}\to\mathbf{0}} L(\mathbf{h}) = 0 \end{cases}$$

We only have to make above reasoning for each component of f.

As a next final step we combine $f'(\mathbf{p})$ and $L(\mathbf{h})$:

$$F(\mathbf{h}) := f'(\mathbf{p}) + L(\mathbf{h})$$

Then obviously $F(\mathbf{h})$ is again *linear.*

We inherit the properties:

$$\begin{cases} f(\mathbf{a}+\mathbf{p}) = f(\mathbf{p}) + F(\mathbf{h})\mathbf{h} \\ \lim_{\mathbf{h}\to\mathbf{0}} F(\mathbf{h}) = f'(\mathbf{p}) \end{cases}$$

Remark: F is of course depending on location $\mathbf{p}\in D$. We suppress this in the notation as long as we keep $\mathbf{p}$ fixed.

Theorem B2.1 Let $f : D \to \mathbb{R}^m$ and $g : D \to \mathbb{R}^m$ be both differentiable at $\mathbf{p}\in D$, $D\subset\mathbb{R}^n$. Then also $f+g$ and cf $(c\in\mathbb{R})$ are differentiable at $\mathbf{p}$ and we have:

$$(f+g)'(\mathbf{p}) = f'(\mathbf{p}) + g'(\mathbf{p})$$
$$(cf)'(\mathbf{p}) = cf'(\mathbf{p})$$

Proof. We have:

$$\begin{cases} f(\mathbf{a}+\mathbf{p}) = f(\mathbf{p}) + f'(\mathbf{p},\mathbf{h}) + R_1(\mathbf{a}+\mathbf{p})\|\mathbf{h}\| \\ \lim_{\mathbf{h}\to\mathbf{0}} R_1(\mathbf{p}+\mathbf{h}) = 0 \end{cases}$$

and

$$\begin{cases} g(\mathbf{a}+\mathbf{p}) = g(\mathbf{p}) + g'(\mathbf{p},\mathbf{h}) + R_2(\mathbf{p}+\mathbf{h})\|\mathbf{h}\| \\ \lim_{\mathbf{h}\to\mathbf{0}} R_2(\mathbf{p}+\mathbf{h}) = \mathbf{0} \end{cases}$$

By adding both equation families (brackets) we obtain the desired result. Using the first bracket and multiplying with c proves the second claim. □

Theorem B2.2 Let $f : D \to \mathbb{R}$ and $g : D \to R$ be differentiable at $\mathbf{p} \in D$, $D \subset \mathbb{R}^n$. Then $f \cdot g$ is differentiable at $\mathbf{p} \in D$ with:

$$(f \cdot g)'(\mathbf{p}, \mathbf{h}) = f(\mathbf{p}) \cdot g'(\mathbf{p}, \mathbf{h}) + g(\mathbf{p}) \cdot f'(\mathbf{p}, \mathbf{h})$$

Proof. After multiplication and re-arrangement we get:

$$\begin{aligned}(fg)(\mathbf{p}+\mathbf{h}) = (fg)(\mathbf{p}) &+ f(\mathbf{p}) \cdot g'(\mathbf{p}, \mathbf{h}) + g(\mathbf{p}) \cdot f'(\mathbf{p}, \mathbf{h}) \\ &+ f'(\mathbf{p}, \mathbf{h})g'(\mathbf{p}, \mathbf{h}) + (g(\mathbf{p}) + g'(\mathbf{p}, \mathbf{h}))R_1(\mathbf{p}+\mathbf{h})\|\mathbf{h}\| \\ &+ (f(\mathbf{p}) + f'(\mathbf{p}, \mathbf{h}))R_2(\mathbf{p}+\mathbf{h})\|\mathbf{h}\|\end{aligned}$$

We can estimate $|f'(\mathbf{p}, \mathbf{h}) \cdot g'(\mathbf{p}, \mathbf{h})|$ by $c\|\mathbf{h}\|^2$ from above. It follows that the last four terms of the sum (establishing the error term) can be written in the form:

$$\begin{cases} R(\mathbf{h})\|\mathbf{h}\| \\ \lim_{\mathbf{h}\to\mathbf{0}} R(\mathbf{h}) = \mathbf{0} \end{cases}$$

□

Remarks: The product rule can be written in the form:

$$\nabla fg(\mathbf{p}) = f(\mathbf{p})\nabla g(\mathbf{p}) + g(\mathbf{p})\nabla f(\mathbf{p})$$

If $f(\mathbf{p}) \neq 0$, $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, f differentiable in $\mathbf{p}$, then:

$$\nabla \frac{1}{f}(\mathbf{p}) = -\frac{1}{(f(\mathbf{p}))^2}\nabla f(\mathbf{p})$$

If $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, $g : D \to \mathbb{R}^m$, f and g differentiable in $\mathbf{p}$, then:

$$(fg)'(\mathbf{p}, \mathbf{h}) = f(\mathbf{p})g'(\mathbf{p}, \mathbf{h}) + f'(\mathbf{p}, \mathbf{h})g(\mathbf{p})$$

Let $[\cdot, \cdot]$ be a *bilinear product* (not necessarily symmetric), with:

$$[\cdot, \cdot] : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^m$$

Let $f : D \to \mathbb{R}^p$, $g : D \to \mathbb{R}^q$, $D \subset \mathbb{R}^n$, f and g differentiable in $\mathbf{p}$. Then:

$$[f, g]'(\mathbf{p}, \mathbf{h}) = [f(\mathbf{p}), g'(\mathbf{p}, \mathbf{h})] + [f'(\mathbf{p}, \mathbf{h}), g(\mathbf{p})]$$

The order of terms is important as long as $[\cdot, \cdot]$ is not symmetric. The proof of this can be made directly or with Theorem B2.2 component-wise. Check!

We are now turning to the *chain rule*:
Theorem B2.3 Let $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$, f differentiable in $\mathbf{p} \in D$, and $g : E \to \mathbb{R}^k$, $E \subset \mathbb{R}^m$, g differentiable in $\mathbf{q} \in E$, with $f(D) \subset E$, and $\mathbf{q} = f(\mathbf{p})$. Then $g \circ f$ is differentiable at $\mathbf{p} \in D$ and it holds:

$$(g \circ f)'(\mathbf{p}) = g'(f(\mathbf{p})) \circ f'(\mathbf{p})$$

Proof. By assumption, and using the compact notation introduced earlier in B2 we have:

$$f(\mathbf{p}+\mathbf{h}) = f(\mathbf{p}) + F(\mathbf{h})\mathbf{h}, \qquad \lim_{\mathbf{h}\to\mathbf{0}} F(\mathbf{h}) = f'(\mathbf{p})$$

$$g(\mathbf{q}+\mathbf{l}) = g(\mathbf{q}) + G(\mathbf{l})\mathbf{l}, \qquad \lim_{\mathbf{l}\to\mathbf{0}} G(\mathbf{l}) = g'(\mathbf{q})$$

$\mathbf{h} \in \mathbb{R}^n, \mathbf{l} \in \mathbb{R}^m$.
WLOG let $\mathbf{p} = \mathbf{0}$, $f(\mathbf{p}) = \mathbf{0}$, $g(\mathbf{q}) = \mathbf{0}$. In this case:

$$f(\mathbf{h}) = F(\mathbf{h})\mathbf{h}, \qquad \lim_{\mathbf{h}\to\mathbf{0}} F(\mathbf{h}) = f'(\mathbf{0})$$
$$g(\mathbf{l}) = G(\mathbf{l})\mathbf{l}, \qquad \lim_{\mathbf{l}\to\mathbf{0}} G(\mathbf{l}) = g'(\mathbf{0})$$

We obtain for $\mathbf{l} = f(\mathbf{h})$:

$$g(f(\mathbf{h})) = (G(F(\mathbf{h})\mathbf{h}) \circ F(\mathbf{h}))\mathbf{h}.$$

This shows the chain rule, as we have:

$$\lim_{\mathbf{h}\to\mathbf{0}} G(F(\mathbf{h})\mathbf{h}) \circ F(\mathbf{h}) = g'(\mathbf{0}) \circ f'(\mathbf{0}).$$

This means $g \circ f$ is differentiable at $\mathbf{0}$, and the derivative of the chain is equal to the chained derivatives. If we go back to an arbitrary location $\mathbf{p}$ and $\mathbf{q} = f(\mathbf{p})$ then indeed:

$$(g \circ f)'(\mathbf{p}) = g'(f(\mathbf{p})) \circ f'(\mathbf{p})$$

□

Remark: Using the notation $f'(\mathbf{p}, \mathbf{h})$ and $g'(\mathbf{q}, \mathbf{l})$ the chain rule becomes:

$$(g \circ f)'(\mathbf{p}, \mathbf{h}) = g'(f(\mathbf{p}), f'(\mathbf{p}, \mathbf{h})).$$

The chain rule implies the following structure of the Jacobian matrices Jg, Jf and $J(g \circ f) =: JH$, i.e. $H := g \circ f$:

$$\underbrace{\begin{pmatrix} D_1H_1 & \cdots & D_nH_1 \\ \vdots & \ddots & \vdots \\ D_1H_k & \cdots & D_nH_k \end{pmatrix}}_{(k\times n)\text{-matrix}} = \underbrace{\begin{pmatrix} D_1g_1 & \cdots & D_mg_1 \\ \vdots & \ddots & \vdots \\ D_1g_k & \cdots & D_mg_k \end{pmatrix}}_{(k\times m)\text{-matrix}} \underbrace{\begin{pmatrix} D_1f_1 & \cdots & D_nf_1 \\ \vdots & \ddots & \vdots \\ D_1f_m & \cdots & D_nf_m \end{pmatrix}}_{(m\times n)\text{-matrix}}$$

We have $JH = JH(\mathbf{p})$ and $Jf = Jf(\mathbf{p})$, but $Jg = Jg(f(\mathbf{p}))$. For the partial derivatives of $H_1, ..., H_k$ we get:

$$D_iH_j = \sum_{l=1}^{m} D_lg_j \cdot D_if_l \qquad (1 \le i \le n,\ 1 \le j \le k)$$

The vector ∇H at $\mathbf{p} \in D \subset \mathbb{R}^n$ is the image of $\nabla g \in \mathbb{R}^m$ at $\mathbf{q} = f(\mathbf{p})$ under the linear mapping $f'(\mathbf{p}) = Jf(\mathbf{p})$.

Diffeomorphisms

Consider $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$ possesses the inverse mapping $g : E \to \mathbb{R}^n$, $E \subset \mathbb{R}^n$ open. Moreover f is differentiable at $\mathbf{p} \in D$ and g is differentiable at $\mathbf{q} = f(\mathbf{p}) \in E$. Then using the chain rule it is possible to express $g'(\mathbf{q})$ by $f'(\mathbf{p})$. We have $g \circ f = I_D$, $(g \circ f)'(\mathbf{p}) = I_{\mathbb{R}^n}$ (I_X: identity mapping on X), so:

$$g'(\mathbf{q}) \circ f'(\mathbf{p}) = I_{\mathbb{R}^n},$$
$$g'(f(\mathbf{p})) = (f'(\mathbf{p}))^{-1}.$$

We even have the following theorem with slightly weaker assumptions:
Theorem **B2.4** Let $f : D \to \mathbb{R}^n$ ($D \subset \mathbb{R}^n$) be invertible (i.e. bijective) and differentiable at $\mathbf{p} \in D$ with $Jf(\mathbf{p}) \neq 0$. The inverse map $g : f(D) \to \mathbb{R}^n$ is continuous at $\mathbf{q} = f(\mathbf{p})$. Then g is differentiable at $\mathbf{p}$ and we have:

$$g'(\mathbf{q}) = (f'(\mathbf{p}))^{-1}$$

Proof. WLOG $\mathbf{p} = \mathbf{0}$, $f(\mathbf{p}) = \mathbf{0}$. From the definition of differentiability we get:

$$\begin{cases} f(\mathbf{x}) = F(\mathbf{x})\mathbf{x} \\ \lim_{\mathbf{x}\to\mathbf{0}} F(\mathbf{x}) = f'(\mathbf{0}) \end{cases}$$

With $F(\mathbf{x})$ a linear (affine, but linear after the shift in $\mathbb{R}^n$) mapping depending on $\mathbf{x}$, and continuous in $\mathbf{x} = \mathbf{0}$. By assumption $\det F(\mathbf{0}) = \det F'(\mathbf{0}) \neq 0$, so in a neighbourhood of $\mathbf{0}$ the inverse linear mappings $(F(\mathbf{x}))^{-1}$ exist! We also have:

$$\lim_{\mathbf{x}\to\mathbf{0}} (F(\mathbf{x}))^{-1} = (f'(\mathbf{0}))^{-1}.$$

For example use Cramer's rule. From:

$$\mathbf{y} = F(\mathbf{x})\mathbf{x}$$

we get the equivalent equation:

$$\mathbf{x} = (F(\mathbf{x}))^{-1}\mathbf{y}.$$

Set $\mathbf{x} = g(\mathbf{y})$ and define $G(\mathbf{y})$ by:

$$G(\mathbf{y}) := (F(g(\mathbf{y})))^{-1}.$$

Then the inverse map g has the form:

$$g(\mathbf{y}) = G(\mathbf{y})\mathbf{y} \text{ and } \lim_{\mathbf{y}\to\mathbf{0}} G(\mathbf{y}) = (f'(\mathbf{0}))^{-1}$$

Here the continuity of g at $\mathbf{0}$ was used. It follows that g is differentiable at $\mathbf{y} = \mathbf{0}$ and that:

$$g'(\mathbf{0}) = (f'(\mathbf{0}))^{-1}.$$

□

Example: Polar coordinate transformation:

$$x_1 = f_1(r, \varphi) = r\cos\varphi$$
$$x_2 = f_2(r, \varphi) = r\sin\varphi$$

This mapping is uniquely invertible, differentiable, and its Jacobian matrix is different from 0. The Jacobian of f is given by:

$$\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

We therefore get as derivative for the inverse mapping g:

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x_1}{x_1^2+x_2} & \frac{x_2}{x_1^2+x_2^2} \\ \frac{x_2}{x_1^2+x_2^2} & \frac{x_1}{x_1^2+x_2^2} \end{pmatrix}$$

To define and understand differomorphisms let us have a fresh look at the interpretation of the derivative. If $f : D \to \mathbb{R}^m, D \subset \mathbb{R}^n$ is differentiable for all $\mathbf{x} \in D$ then we can define a map:

$$\mathbf{x} \mapsto f'(\mathbf{x}),$$

i.e. we map D into $L(\mathbb{R}^n, \mathbb{R}^m)$, the set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e.:

$$f' : D \to L(\mathbb{R}^n, \mathbb{R}^m)$$

If $f'(\mathbf{x})$ is applied to $\mathbf{h} \in \mathbb{R}^n$ the result is a vector $f'(\mathbf{x}, \mathbf{h}) \in \mathbb{R}^m$, i.e. one can also interpret f' as a mapping in the following sense:

$$f' : D \times \mathbb{R}^n \to \mathbb{R}^m$$

This mapping is linear in the second argument. Both interpretations are valid and should be taken when convenient.

If $f : D \to \mathbb{R}^n, D \subset \mathbb{R}^n$, continuously differentiable and invertible (for all $\mathbf{x} \in D$), then $g := f^{-1}$ is not necessarily differentiable. But if g is even continuously differentiable, then f becomes a *diffeomorphism*:
Definition **B2.1** Let $f : D \to \mathbb{R}^n, D \subset \mathbb{R}^n$ be continuously differentiable and invertible on D, and $I_f = f(D) \subset \mathbb{R}^n$ open. If the inverse mapping:

$$f^{-1} : I_f \to \mathbb{R}^n$$

is again continuously differentiable, then f and f^{-1} are called *differomorphisms*.
For example the polar coordinate transformation is a diffeomorphism.

B3 Intermediate value theorem

Differentiable curves and lines in $\mathbb{R}^n$

Consider a mapping:

$$\gamma : D \to \mathbb{R}^n, \text{ with } D \subset \mathbb{R}$$

i.e. $\gamma = \gamma(t) \in \mathbb{R}^n$, $t \in D \subset \mathbb{R}$. Then γ is called a curve in $\mathbb{R}^n$ with *parameterisation* t. We assume now γ is differentiable on D. Then at $t_0 \in D$ we have:

$$\gamma'(t_0) = \lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0},$$

and each component $\gamma_1'(t_0), ..., \gamma_n'(t_0)$ is differentiable in t_0 in the one-dimensional sense. The derivative $\gamma'(t_0)$ has the following interpretation. Define $g(t)$ by:

$$g(t) := \gamma(t_0) + \gamma'(t_0)(t - t_0).$$

Then $g(t)$ is defining a line in $\mathbb{R}^n$ parameterised by t. This line approximates the curve $\gamma(t)$ at t_0 very well, i.e.:

$$\lim_{t \to t_0} \frac{\|\gamma(t) - g(t)\|}{t - t_0} = 0$$

We also have a *kinematic* interpretation. The limit vector $\gamma'(t_0)$ is the current *velocity vector* at t_0, and $|\gamma'(t_0)|$ the current *velocity* at t_0, with $|\cdot|$ being the Euclidean norm. Also:

$$\frac{\gamma'(t_0)}{|\gamma'(t_0)|}$$

is called the *unit tangential vector* or *normal vector* of γ at t_0. If $\gamma' : [a, b] =: D \to \mathbb{R}^n$ is continuous, then γ is called *smooth.* γ is called *piecewise smooth* if it is composed of finitely many(!) smooth curves.

Definition **B3.1** Let $\gamma : [a, b] \to \mathbb{R}^n$ be a piecewise smooth curve. Then:

$$L(\gamma) := \int_a^b |\gamma'(t)| \, dt$$

is called the *length* of curve γ.

Examples: Let:

$$\gamma(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \qquad (t \in [0, 2\pi])$$

be the unit circle (in $\mathbb{R}^2$). The velocity vectors:

$$\gamma'(t) = \begin{pmatrix} \gamma_1'(t) \\ \gamma_2'(t) \end{pmatrix} = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

has Euclidean norm 1, so $L(\gamma) = \int_0^{2\pi} dt = 2\pi$. Let now $a, b \in \mathbb{R}^n$ and:

$$\gamma(t) = (1 - t)a + tb \qquad (t \in [0, 1])$$

Then:

$$\gamma'(t) = b - a,\ |\gamma'(t)| = |b - a| \text{ and } L(\gamma) = \int_0^1 |b - a|\, dt = |b - a|.$$

$\gamma(t)$ is the *line segment* between a and b.

Theorem $\boxed{\textbf{B3.1}}$ Let $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, be differentiable, and let the line between points $\mathbf{a}, \mathbf{b} \in D$ be contained in D. Then there is a point $\mathbf{c}$ on the line between $\mathbf{a}$ and $\mathbf{b}$ such that:

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\mathbf{c}, \mathbf{b} - \mathbf{a})$$

Proof. We can describe the line between $\mathbf{a}$ and $\mathbf{b}$ by:

$$\gamma(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) \qquad (0 \le t \le 1)$$

Then γ is clearly differentiable, and we have:

$$\gamma'(t) = \mathbf{b} - \mathbf{a}$$

which means γ is constant. Let $F : [0, 1] \to \mathbb{R}$ be defined by:

$$F(t) = f(\gamma(t)).$$

We can apply the one-dimensional intermediate value theorem. Therefore there is a $\nu \in]0, 1[$ such that:

$$F(1) - F(0) = F'(\nu)$$

Using the chain rule we get:

$$F'(t) = f'(\gamma(t), \gamma'(t))$$

Because $F(0) = f(\mathbf{a})$, $F(1) = f(\mathbf{b})$, $\gamma(\nu) =: c$ is a point on the line between $\mathbf{a}$ and $\mathbf{b}$, moreover $\gamma'(\nu) = \mathbf{b} - \mathbf{a}$, the theorem is proven. □

Remark: It is possible to replace the line between $\mathbf{a}$ and $\mathbf{b}$ by a differentiable curve γ connecting $\mathbf{a}$ and $\mathbf{b}$. In this case the "increment" $f(\mathbf{b}) - f(\mathbf{a})$ can be expressed as:

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\gamma(\nu), \gamma'(\nu)) \qquad (0 < \nu < 1)$$

Using $\mathbf{x} + \mathbf{h}$ instead of $\mathbf{a}$ and $\mathbf{b}$ one obtains another formulation of the intermediate value theorem:

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = f'(\mathbf{x} + \nu\mathbf{h}, \mathbf{h}) \qquad (0 < \nu < 1)$$

With the help of the intermediate value theorem we can characterise constant functions:

Theorem $\boxed{\textbf{B3.2}}$ Let $D \subset \mathbb{R}^n$ be open, and let there exist a differentiable curve connecting any two points of D such that the curve is entirely contained in D (D is called *path connected*). Let $f : D \to \mathbb{R}$ be differentiable. Then:

$$\underbrace{f \text{ is constant}}_{(a)} \iff \underbrace{f'(\mathbf{x}, \mathbf{h}) = 0 \text{ for all } \mathbf{x} \in D, \text{ and all } \mathbf{h} \in \mathbb{R}^n.}_{(b)}$$

Proof. $(a) \implies (b)$ is trivial, even in case D is not path connected. Now consider $(b) \implies (a)$. Choose two points in D, say $\mathbf{a}$ and $\mathbf{b}$. Connect $\mathbf{a}$ and $\mathbf{b}$ by finitely many line segments, such that the sequence of vertices connected by lines becomes $\mathbf{a} = \mathbf{a}_0, \mathbf{a}_1, ..., \mathbf{a}_k = \mathbf{b}$. Apply the mean-value theorem to the line segment joining $\mathbf{a}_j, \mathbf{a}_{j+1}$. Because $f'(\mathbf{c}_j, \mathbf{a}_{j+1} - \mathbf{a}_j) = 0$ for some $\mathbf{c}_j$ on the line segment we have:

$$f(\mathbf{a}_{j+1}) = f(\mathbf{a}_j) \text{ for } j = 0, 1, ..., k - 1$$

This implies $f(\mathbf{b}) = f(\mathbf{a})$. □

Remarks: The equation $f'(\mathbf{x},\mathbf{h}) = 0$ is equivalent to:

$$D_1 f(\mathbf{x}) = 0, ..., D_n f(\mathbf{x}) = 0$$

or:

$$\nabla f(\mathbf{x}) = \mathbf{0}.$$

Instead of using line segments to prove Theorem B3.2 we could have used a differentiable curve or path connecting $\mathbf{a}$ and $\mathbf{b}$. In this case we get:

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\gamma(\nu), \gamma'(\nu)) = 0$$

with some ν along the curve.

We now turn to an important theorem making a connection between differentiability and continuity of functions:

Theorem B3.3 Let $f : D \to \mathbb{R}, D \subset \mathbb{R}^n$. Assume all partial derivatives $D_1 f, ..., D_n f$ exist and are continuous at some point $\mathbf{p} \in D$. Then f is differentiable at $\mathbf{p}$.

Proof. We have to show:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - f'(\mathbf{a},\mathbf{h})}{\|\mathbf{h}\|} = 0.$$

We already know that if this would hold then necessarily:

$$f'(\mathbf{a},\mathbf{h}) = D_1 f(\mathbf{a}) \cdot h_1 + ... + D_n f(\mathbf{a}) \cdot h_n$$

Next we set $\mathbf{a}_0 = \mathbf{p}$, $\mathbf{a}_n = \mathbf{p} + \mathbf{h}$ and:

$$\begin{aligned}
\mathbf{a}_1 - \mathbf{a}_0 &= h_1 \mathbf{e}_1, \\
\mathbf{a}_2 - \mathbf{a}_1 &= h_2 \mathbf{e}_2, \\
&\vdots \\
\mathbf{a}_n - \mathbf{a}_{n-1} &= h_n \mathbf{e}_n.
\end{aligned}$$

Then $f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p})$ can be written as:

$$(f(\mathbf{p}+\mathbf{h}) - f(\mathbf{a}_{n-1})) + (f(\mathbf{a}_{n-1}) - f(\mathbf{a}_{n-2})) + ... + (f(\mathbf{a}_1) - f(\mathbf{p}))$$

We apply the mean value theorem, and denote the intermediate value on the line segment between $\mathbf{a}_{j-1}$ and $\mathbf{a}_j$ by $\mathbf{c}_j$. Then we have:

$$\begin{aligned}
f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) &= \sum_{j=1}^{n} f'(\mathbf{c}_j, \mathbf{a}_j - \mathbf{a}_{j-1}) \\
&= \sum_{j=1}^{n} f'(\mathbf{c}_j, \mathbf{e}_j)\mathbf{h}_j \\
&= \sum_{j=1}^{n} D_j f(\mathbf{c}_j)\mathbf{h}_j
\end{aligned}$$

Therefore we get:

$$f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - f'(\mathbf{p},\mathbf{h}) = \sum_{j=1}^{n} (D_j f(\mathbf{c}_j) - D_j f(\mathbf{p})) \cdot \mathbf{h}_j$$

and the estimate:

$$|f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - f'(\mathbf{p},\mathbf{h})| \le \sum_{j=1}^{n} |D_j f(\mathbf{c}_j) - D_j f(\mathbf{p})||\mathbf{h}_j|$$

$$\le \|\mathbf{h}\| \cdot \sum_{j=1}^{n} |D_j f(\mathbf{c}_j) - d_j f(\mathbf{p})|.$$

Because $\|\mathbf{c}_j - \mathbf{p}\| \le n\|\mathbf{h}\|$, and $D_1 f, ..., D_n f$ continuous in $\mathbf{p} \in D$, the theorem follows. □

Theorem B3.3 gives us an often used possibility to prove differentiability of a function at some point in its domain of definition.

B4 Higher derivatives and Taylor's formula

Consider again $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ differentiable in D. Then as discussed either:

$$f' : D \to L(\mathbb{R}^n, \mathbb{R}^m) \text{ or}$$
$$f' : D \times \mathbb{R}^n \to \mathbb{R}^m$$

Assume now f' is differentiable in D. Then $f'' := (f')'$, and f'' can be considered as:

$$f'' : D \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)),$$

i.e. f'' is *multi-linear* for fixed $\mathbf{p} \in D \subset \mathbb{R}^n$. We set:

$$g_{\mathbf{h}} : D \to \mathbb{R}^n$$

by defining $g_{\mathbf{h}}(\mathbf{x}) := f'(\mathbf{x}, \mathbf{h})$. If $g_{\mathbf{h}}$ is differentiable then:

$$g'_{\mathbf{h}}(\mathbf{x}, \mathbf{k}) =: f''(\mathbf{x}, \mathbf{h}, \mathbf{k})$$

In this case we can interpret f'' as:

$$f'' : D \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m,$$

and for every $\mathbf{x} \in D$ f'' is bilinear (in the last two arguments). We have:

$$g_{\mathbf{h}_1+\mathbf{h}_2} = g_{\mathbf{h}_1} + g_{\mathbf{h}_2}$$
$$g_{c\mathbf{h}} = c g_{\mathbf{h}}$$

and therefore (by the composition rules):

$$g'_{\mathbf{h}_1+\mathbf{h}_2}(\mathbf{x}, \mathbf{k}) = g'_{\mathbf{h}_1}(\mathbf{x}, \mathbf{k}) + g'_{\mathbf{h}_2}(\mathbf{x}, \mathbf{k})$$
$$g_{c\mathbf{h}}(\mathbf{x}, \mathbf{k}) = c g'_{\mathbf{h}}(\mathbf{x}, \mathbf{k}),$$

which means:

$$f''(\mathbf{x}, \mathbf{h}_1 + \mathbf{h}_2, \mathbf{k}) = f'(\mathbf{x}, \mathbf{h}_1, \mathbf{k}) + f''(\mathbf{x}, \mathbf{h}_2, \mathbf{k}),$$
$$f''(\mathbf{x}, c\mathbf{h}, \mathbf{k}) = c f''(\mathbf{x}, \mathbf{h}, \mathbf{k}).$$

Using the standard basis in $\mathbb{R}^n$ we can write:

$$f'(\mathbf{x}, \mathbf{h}) = \sum_{i=1}^{n} D_i f(\mathbf{x}) h_i$$

$$f''(\mathbf{x}, \mathbf{h}, \mathbf{k}) = \sum_{j=1}^{n} D_j \left(\sum_{i=1}^{n} D_i f(\mathbf{x}) h_i \right) k_j$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} D_j D_i f(\mathbf{x}) h_i k_j$$

If f'' exists then also the partial derivatives of *second order* D_jD_if exist, created by differentiating D_if partially with respect to the j-th variable. We write:

$$D_jD_if, \text{ or } \frac{\partial \frac{\partial f}{\partial x_i}}{\partial x_j} =: \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Often used are also: $f_{x_ix_j}$ or f_{ij}. We next show that the bilinear map:

$$(\mathbf{h}, \mathbf{k}) \mapsto f''(\mathbf{x}, \mathbf{h}, \mathbf{k})$$

is *symmetric*:

Theorem **B4.1** [Theorem of H.A. Schwarz] Let $f : D \to \mathbb{R}^m$ be twice differentiable. Then for all $\mathbf{h}, \mathbf{k} \in \mathbb{R}^n$ we have:

$$f''(\mathbf{x}, \mathbf{h}, \mathbf{k}) = f''(\mathbf{x}, \mathbf{k}, \mathbf{h})$$

Proof. We start by considering the expression:

$$f(\mathbf{x} + \mathbf{h} + \mathbf{k}) - f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x} + \mathbf{k}) + f(\mathbf{x}).$$

We would like to know whether this is a good approximation for:

$$f''(\mathbf{x}, \mathbf{h}, \mathbf{k}).$$

Indeed we will prove that:

$$\lim_{s \to 0} \frac{f(\mathbf{x} + s\mathbf{h} + s\mathbf{k}) - f(\mathbf{x} + s\mathbf{h}) - f(\mathbf{x} + s\mathbf{k}) + f(\mathbf{x})}{s^2} = f''(\mathbf{x}, \mathbf{h}, \mathbf{k}) \qquad (s > 0)$$

If the last equation holds the theorem follows, as the value of the limit does not change when $\mathbf{h}$ and $\mathbf{k}$ are interchanged. We restrict ourselves to a real-valued f (otherwise we would make the proof component-wise). First we show the following equality:

$$f(\mathbf{x} + \mathbf{h} + \mathbf{k}) - f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x} + \mathbf{k}) + f(\mathbf{x}) = f''(\mathbf{x}, \mathbf{h}, \mathbf{k}) + R_x(\mathbf{h}, \nu\mathbf{h} + \mathbf{k})\|\nu\mathbf{h} + \mathbf{k}\| - R_x(\mathbf{h}, \nu\mathbf{h})\|\nu\mathbf{h}\|,$$

for some $0 < \nu < 1$.
Define:

$$F(t) := f(\mathbf{x} + t\mathbf{h} + \mathbf{k}) - f(\mathbf{x} + t\mathbf{h})$$

for $0 \le t \le 1$. Then:

$$F'(t) := f'(\mathbf{x} + t\mathbf{h} + \mathbf{k}, \mathbf{h}) - f'(\mathbf{x} + t\mathbf{h}, \mathbf{h})$$

and:

$$F(1) - F(0) = f(\mathbf{x} + \mathbf{h} + \mathbf{k}) - f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x} + \mathbf{k}) + f(\mathbf{x})$$

We apply the mean-value theorem in one dimension and get:

$$F(1) - F(0) = F'(\nu)$$

for some $0 < \nu < 1$. Therefore:

$$f(\mathbf{x} + \mathbf{h} + \mathbf{k}) - f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x} + \mathbf{k}) + f(\mathbf{x}) = (f'(\mathbf{x} + \nu\mathbf{h} + \mathbf{k}, \mathbf{h}) - f'(\mathbf{x}, \mathbf{h})) - (f'(\mathbf{x} + \nu\mathbf{h}, \mathbf{h}) - f'(\mathbf{x}, \mathbf{h}))$$

For the two differences on the RHS we can find by definition:

$$f'(\mathbf{x} + \nu\mathbf{h} + \mathbf{k}, \mathbf{h}) - f'(\mathbf{x}, \mathbf{h}) = f''(\mathbf{x}, \mathbf{h}, \nu\mathbf{h} + \mathbf{k}) + R_x(\mathbf{h}, \nu\mathbf{h} + \mathbf{k})\|\nu\mathbf{h} + \mathbf{k}\|,$$
$$f'(\mathbf{x} + \nu\mathbf{h}, \mathbf{h}) - f'(\mathbf{x}, \mathbf{h}) = f''(\mathbf{x}, \mathbf{h}, \nu\mathbf{h}) + R_x(\mathbf{h}, \nu\mathbf{h})\|\nu\mathbf{h}\|.$$

But because $f''(\mathbf{x}, \mathbf{h}, \nu\mathbf{h} + \mathbf{k}) - f''(\mathbf{x}, \mathbf{h}, \nu\mathbf{h}) = f''(\mathbf{x}, \mathbf{h}, \mathbf{k})$ the theorem follows. □

From Theorem B4.1 follows:

$$D_i D_j f(\mathbf{x}) = D_j D_i f(\mathbf{x})$$

It is therefore not important in which order we form higher order derivatives.

Definition B4.2 Using Definition B4.1 iteratively we call:

$$f^{(p)} : D \times \underbrace{\mathbb{R}^n \times ... \times \mathbb{R}^n}_{p\text{-times}} \to \mathbb{R}^m$$

the *p-th derivative of f on D*. For fixed $\mathbf{x} \in D$, $f^{(p)}$ is p-multi-linear. Moreover $f^{(p)}$ is totally symmetric, i.e. its values do change when $\mathbf{x} \in D$ is fixed, and any two arguments are interchanged. Also $f^{(p)}$ is called *p-times continuously differentiable at* $\mathbf{x} \in D$ if $f^{(p)}$ is continuous in $\mathbf{x}$. If $f^{(p)}$ exists for all $p \in \mathbb{N}$ then f is called *infinitely often differentiable* in $\mathbf{x} \in D$.

Theorem B4.2 (Taylor's Formula) Let $f : D \to \mathbb{R}$ be $(p+1)$ times differentiable in D, and let the line segment between $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ be contained in D. Then we have:

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + f'(\mathbf{x}, \mathbf{h}) + \frac{1}{2!} f''(\mathbf{x}, \mathbf{h}, \mathbf{h}) + ... + \frac{1}{p!} f^{(p)}(\mathbf{x}, \underbrace{\mathbf{h}, ..., \mathbf{h}}_{p\text{-times}}) + R_p(\mathbf{x}, \mathbf{h}).$$

Here $R_p(\mathbf{x}, \mathbf{h})$ is given by:

$$R_p(\mathbf{x}, \mathbf{h}) = \frac{1}{(p+1)!} f^{(p+1)}(\mathbf{x} - \nu\mathbf{h}, \underbrace{\mathbf{h}, ..., \mathbf{h}}_{p\text{-times}}) \text{ and } 0 < \nu < 1$$

Proof. We define a line segment:

$$\gamma(t) = \mathbf{x} + t\mathbf{h}, \qquad (t \in [0, 1])$$

together with $F : f \circ \gamma$. Now we can apply the one-dimensional Taylor's formula to get:

$$F(1) = F(0) + F'(0) + \frac{1}{2!} F''(0) + ... + \frac{1}{p!} F^{(p)}(0) + \frac{1}{(p+1)!} F^{(p+1)}(\nu).$$

For the derivatives of F we find:

$$\begin{aligned} F'(t) &= f'(\gamma(t), \gamma(t)) = f'(\gamma(t), \mathbf{h}) \\ F''(t) &= f''(\gamma(t), \mathbf{h}, \gamma'(t)) = f''(\gamma(t), \mathbf{h}, \mathbf{h}) \\ &\vdots \\ F^p(t) &= f^{(p)}(\gamma(t), \underbrace{\mathbf{h}, ..., \mathbf{h}}_{p\text{-times}}) \\ F^{(p+1)}(t) &= f^{(p)}(\gamma(t), \underbrace{\mathbf{h}, ..., \mathbf{h}}_{(p+1)\text{-times}}) \end{aligned}$$

Because $\gamma(0) = \mathbf{x}$ and $\gamma(1) = \mathbf{x} + \mathbf{h}$ the theorem follows. □

Of course we can generalise Taylor's formula to vector functions.

If f is infinitely often differentiable one obtains the series:

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(\mathbf{a}, \mathbf{x} - \mathbf{a}, ..., \mathbf{x} - \mathbf{a})$$

This expression is called Taylor series developed at $\mathbf{a} \in \mathbb{R}^n$. The terms are homogenous polynomials in the variables $(\mathbf{x}_1 - \mathbf{a}_1), ..., (\mathbf{x}_n - \mathbf{a}_n)$. The Taylor series is convergent iff:

$$\lim_{p \to \infty} R_p(\mathbf{a}, \mathbf{x} - \mathbf{a}) = 0$$

If there is an open neighbourhood of $\mathbf{a}$ in which f can be expressed by its Taylor-series developed at $\mathbf{a}$, then f is called (real) *analytical*.

Remark: If we are not using the coordinate free notation but partial derivatives one obtains:

$$\sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \frac{1}{k!} D_{i_1} D_{i_k} ... f(\mathbf{a})(\mathbf{x}_{1_1} - \mathbf{a}_{i_1})...(\mathbf{x}_{i_k} - \mathbf{a}_{i_k})$$

Because the order of partial derivatives can be changed we can rearrange these polynomials. We introduce *multi-indices* $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)$ with:

$$\begin{aligned} D_1^{\alpha_1} ... D_n^{\alpha_n} f(a) &=: D^{\alpha} f(a) \\ (x_1 - a_1)^{\alpha_1} ... (x_n - a_n)^{\alpha_n} &=: (x - a)^{\alpha} \\ \alpha_1! \cdot ... \cdot \alpha_n! &=: \alpha! \\ \alpha_1 + ... + \alpha_n &=: |\alpha|. \end{aligned}$$

If a multi-index of order k is given ($|\alpha| = k$), then there are $\frac{k!}{\alpha_1!...\alpha_n!}$ terms in the sum which all contribute the same value:

$$D_1^{\alpha_1} ... D_n^{\alpha_n} f(a)(x_1 - a_1)_1^{\alpha} ... (x_n - a_n)_n^{\alpha}.$$

We can therefore write:

$$\frac{1}{k!} f^{(k)}(a, \underbrace{x - a, ..., x - a}_{k\text{-times}}) = \sum_{|\alpha| = k} \frac{1}{\alpha!} D^{\alpha} f(a)(x - a)^{\alpha}.$$

Definition **B4.3** The symmetric bilinear form:

$$f''(\mathbf{x}, \mathbf{h}, \mathbf{h}), \ \mathbf{h} \in \mathbb{R}^n$$

is called Hesse-form at location $\mathbf{x} \in D, D \subset \mathbb{R}^n$, $f : D \to \mathbb{R}$. Using the standard basis in $\mathbb{R}^n$ we have:

$$f''(\mathbf{x}, \mathbf{h}, \mathbf{h}) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_j D_i f(x) h_i h_j$$

The corresponding summetric matrix of the second order partial derivatives is called *Hesse-matrix*.

Local extrema:

Definition **B4.4** Let $f : D \to \mathbb{R}$ be differentiable in D, and let $\mathbf{c} \in D$. If:

$$f'(\mathbf{c}, \mathbf{h}) = 0 \text{ for all } \mathbf{h} \in \mathbb{R}^n,$$

then $\mathbf{c}$ is called *critical point* of f.
Of course $f'(\mathbf{c}, \mathbf{h}) = 0 \ \forall \mathbf{h} \in \mathbb{R}^n$ implies:

$$D_1 f(\mathbf{c}) = ... = D_n f(\mathbf{c}) = 0.$$

We can also write $\nabla f(\mathbf{c}) = 0$.

Theorem **B4.3** If $\mathbf{c} \in D$ is a local extremum of the differentiable function f, then $\mathbf{c}$ must be a critical point of f.

Proof. By assumption there is a neighbourhood N of $\mathbf{c}$ such that either:

$$\underbrace{f(\mathbf{x}) \leq f(\mathbf{c})}_{\text{maximum}} \text{ or } \underbrace{f(\mathbf{x}) \geq f(\mathbf{c})}_{\text{minimum}} \text{ for all } \mathbf{x} \in N.$$

We choose $\mathbf{h}$ such that $\mathbf{c} + \mathbf{h}$ and $\mathbf{c} - \mathbf{h}$ are contained in N. Consider:

$$F(t) := f(\mathbf{c} + t\mathbf{h}), \quad -1 \leq t \leq 1.$$

It holds $F(t) \leq F(0)$ (maximum) or $F(t) \geq F(0)$ (minimum). Because F is differentiable:

$$F'(0) = f'(\mathbf{c}, \mathbf{h}) = 0$$

This equality holds for all $\mathbf{h}$ in a neighbourhood of $\mathbf{0} \in \mathbb{R}^n$. Because $f'(\mathbf{c}, \mathbf{h})$ is linear in $\mathbf{h}$ we have $f'(\mathbf{c}, \mathbf{h}) = 0$ for all $\mathbf{h} \in \mathbb{R}^n$. □

Theorem B4.3 implies a strategy for finding extrema: if f is differentiable *first* find the *critical points* in D. Then *secondly* check if those points are *extremal points* (i.e. extrema). How can we decide critical points are extrema?

Theorem B4.4 Let $f : D \to \mathbb{R}$ be twice differentiable, and let $\mathbf{c} \in D$ be a critical point of f. If the quadratic form $f''(\mathbf{c}, \mathbf{h}, \mathbf{h})$ is:

(i) positive-definite, then $\mathbf{c}$ is a minimum of f,

(ii) negative-definite, then $\mathbf{c}$ is a maximum of f,

(iii) indefinite, then $\mathbf{c}$ is no extremum of f.

Remark: $f''(\mathbf{c}, \mathbf{h}, \mathbf{h})$ can also be positive semi-definite or negative semi-definite. If we have for fixed $\mathbf{c} \in D$:

(i) $f''(\mathbf{c}, \mathbf{h}, \mathbf{h}) > 0 \ \forall \mathbf{h} \in \mathbb{R}^n$ then $f''(\mathbf{c}, \mathbf{h}, \mathbf{h})$ is *positive definite* and if $f''(\mathbf{c}, \mathbf{h}, \mathbf{h}) \geq 0 \ \forall \mathbf{h} \in \mathbb{R}^n$ then f'' is *positive semi-definite.*

(ii) $f''(\mathbf{c}, \mathbf{h}, \mathbf{h}) < (\leq) 0 \ \forall \mathbf{h} \in \mathbb{R}^n \iff f''$ negative-definite (*negative semi-definite*).

(iii) There exist $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^n$ such that $f''(\mathbf{c}, \mathbf{h}_1, \mathbf{h}_1) > 0$ and $f''(\mathbf{c}, \mathbf{h}_2, \mathbf{h}_2) < 0 \iff f''$ is *indefinite.*

In case $f'(\mathbf{c}, \mathbf{h}, \mathbf{h})$ is semi-definite no conclusion can be drawn.

Proof. We use Taylor's formula with $\mathbf{h} =: \mathbf{x} - \mathbf{c}$ for some $\mathbf{x} \in D$ in a neighbourhood of $\mathbf{c}$. Let ν be 0.

$$\begin{cases} f(\mathbf{x}) = f(\mathbf{c}) + f'(\mathbf{c}, \mathbf{x} - \mathbf{c}) + \frac{1}{2} f''(\mathbf{c} + \nu(\mathbf{x} - \mathbf{c}), \mathbf{x} - \mathbf{c}, \mathbf{x} - \mathbf{c}) + R(\mathbf{x}) \\ \lim_{\mathbf{x} \to \mathbf{c}} R(\mathbf{x}) = 0 \end{cases}$$

Because $f'(\mathbf{c}, \mathbf{x} - \mathbf{c}) = 0$ we need to look at the properties of f'' above.

Let $f''(\mathbf{c}, \mathbf{h}, \mathbf{h})$ be positive-definite. Then there is a constant $K > 0$ such that:

$$f''(\mathbf{c}, \mathbf{h}, \mathbf{h}) \geq K \|\mathbf{h}\|^2$$

(because the minimum K of the quadratic form is positive on the unit ball $\|h\| = 1$)

Choose $\|\mathbf{x} - \mathbf{c}\|$ small enough such that $|R(\mathbf{x})| < \frac{K}{2}$. Then:

$$f(\mathbf{x}) - f(\mathbf{c}) = \frac{1}{2} f''(\mathbf{c}, \mathbf{x} - \mathbf{c}, \mathbf{x} - \mathbf{c}) + \frac{1}{2} R(\mathbf{x}) \|\mathbf{x} - \mathbf{c}\|^2$$

and therefore:

$$f(\mathbf{x}) - f(\mathbf{c}) \geq \frac{K}{4} \|\mathbf{x} - \mathbf{c}\|^2 \geq 0.$$

This means $\mathbf{c}$ is a local minimum. The proof of a maximum is equivalent.

Let now $f''(\mathbf{c}, \mathbf{h}, \mathbf{h})$ be indefinite. Then there is a vector $\mathbf{h}_1$ with $\|\mathbf{h}_1\| = 1$ such that:

$$f''(\mathbf{c}, \mathbf{h}_1, \mathbf{h}_1) = K_1 > 0,$$

and a vector $\mathbf{h}_2$, $\|\mathbf{h}_2\| = 1$ such that:

$$f''(\mathbf{c}, \mathbf{h}_2, \mathbf{h}_2) = K_2 < 0.$$

If we go along the line segments from $\mathbf{c}$ to $\mathbf{h}_1$, and $\mathbf{c}$ to $\mathbf{h}_2$ the difference of $f(\mathbf{x}) - f(\mathbf{c})$ is positive respectively negative for $\|\mathbf{x} - \mathbf{c}\|$ sufficiently small. Therefore $\mathbf{c}$ cannot be an extremum of f.

□

Example: Let $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^2$, $f = f(x_1, x_2)$. Let $\mathbf{c} = (c_1, c_2)$ be a critical point of f. Define:

$$\frac{\partial^2 f}{\partial x_1^2}(c_1, c_2) = a, \ \frac{\partial^2 f}{\partial x_1 \partial x_2}(c_1, c_2) = b, \ \frac{\partial^2 f}{\partial x_2^2}(c_1, c_2) = c.$$

Then c is a:

(i) minimum if $a > 0$, $ac - b^2 > 0$.

(ii) maximum if $a < 0$, $ac - b^2 > 0$.

(iii) potential minimum if $a \geq 0$, $ac - b^2 \geq 0$.

(iv) potential maximum if $a \leq 0$, $ac - b^2 \geq 0$.

This means (i) and (ii) are *necessary and sufficient*, (iii) and (iv) only *necessary* conditions.

If $ac - b^2 < 0$ then c cannot be on extremum. The surface described by $x_3 = f(x_1, x_2)$ has a saddle-point at c.

Remark: A quadratic form:

$$Q(\mathbf{h}, \mathbf{h}) := \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} h_i h_j$$

is *positive definite* if:

$$c_{11} > 0, \ \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} > 0, \dots, \ \det(c_{ij}) > 0.$$

B5 Banach Fixed Point Theorem

We now like to find fixed points of "high-dimensional" mappings with some structure, i.e. where distances are defined.

Definition B5.1 A metric space (M, d) is a couple, with M an arbitrary set, and d a mapping:

$$d : M \times M \to \mathbb{R}$$

satisfying:

(i) $d(x, y) > 0$ for $x \neq y$.

(ii) $d(x, y) = d(y, x)$.

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ [triangle inequality].

Clearly $d(x, y) = 0$ if $x = y$. Every normed vector space is a metric space by setting:

$$d(x, y) = \|y - x\|$$

Also any arbitrary subset of a normed vector space is a metric space. One can easily define sequences and converge in metric spaces. However the limit of a convergent sequence with elements in M need not be an element of M (take $M = \mathbb{Q}$, a sequence might have a limit in $\mathbb{R} \supset \mathbb{Q}$).

Definition **B5.2** A metric space (M, d) is called *complete* is every limit of a convergent sequence $(x_k) \in M$ is an element of M.

Definition **B5.3** A mapping $\Phi : M \to M$ is a *contraction* if there is a constant C with $0 \leq C < 1$ such that:

$$d(\Phi(x), \Phi(y)) \leq C \cdot d(x, y)$$

for all $x, y \in M$.

Theorem **B5.1** Every contraction Φ defined on a complete metric space (M, d) has exactly one fixed point, i.e. there is an $m \in M$ with:

$$\Phi(m) = m$$

(Banach fixed point theorem)

Proof. We first show the existence of m. We choose an arbitrary $x_0 \in M$ and define recursively:

$$x_{k+1} = \Phi(x_k),\ k = 0, 1, 2, ...$$

We next show (x_k) is a Cauchy sequence. We have by assumption:

$$d(x_{j+1}, x_j) \leq C d(x_j, x_{j-1}) \text{ for } j \geq 1, 0 \leq C < 1$$

Using the triangle inequality and incorporating the points $x_{k+p-1}, x_{k+p-2}, ..., x_{k-1}$ we get:

$$d(x_{k+p}, x_k) \leq d(x_{k+p}, x_{k+p-1}) + ... + d(x_{k+1}, x_k).$$

Using Φ is a contraction we get the estimate:

$$d(x_{k+p}, x_k) \leq (C^{p-1} + C^{p-2} + ... + C + 1) d(x_{k+1}, x_k)$$

Because $0 \leq C < 1$ we get the following estimate for arbitrary indexes k and p:

$$d(x_{k+p}, x_k) \leq \frac{C^k}{1 - C} d(x_1, x_0)$$

Because $\lim_{k\to\infty} C^k = 0$ (x_k) clearly is a Cauchy sequence. By assumption (M, d) is complete. Therefore there exists $z \in M$ with $\lim_{k\to\infty}(x_k) = z$. Clearly z is a fixed point because by the continuity of Φ we have:

$$z = \lim_{k\to\infty} x_{k+1} = \lim_{k\to\infty} \Phi(x_k) = \Phi\left(\lim_{k\to\infty} x_k\right) = \Phi(z)$$

We need to show the uniqueness of z, i.e. $z = m$. Assume $m \neq z$, but $z = \Phi(z)$ and $m = \Phi(m)$. Then $d(z, m) = d(\Phi(z), \Phi(m)) \leq C \cdot d(z, m)$. Because $0 \leq C < 1$ we have $d(z, m) = 0$. Therefore $z = m$. □

B6 Implicit Function Theorem

Theorem B6.1 (Local Invertibility Theorem) Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^n$ be continuously differentiable on D. Assume that for $\mathbf{p} \in D$ fixed, $f'(\mathbf{p}) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, i.e. $\det(f'(p)) \neq 0$.

Then there is an open neighbourhood N of p such that:

1. $f|_N$ is injective,
2. $f(N)$ is open,
3. $f^{-1}|_{f(N)}$ is continuously differentiable, i.e. $f|_N$ is a diffeomorphism with $f|_N : N \to f(N)$.

Proof. Assume $\mathbf{p} = 0$ and $f(\mathbf{p}) = 0$ (as otherwise instead of $x \mapsto f(x)$ we consider $x \mapsto f(\mathbf{x}+\mathbf{p}) - f(\mathbf{p})$). We like to apply the Banach Fixed Point Theorem. To do so we first make a normation of $f'(\mathbf{0})$. Indeed $f'(\mathbf{0})$ is invertible by assumption. Instead of f consider the mapping $(f'(\mathbf{0}))^{-1} \circ f$. We have $((f'(\mathbf{0}))^{-1} \circ f)'(\mathbf{x}) = \mathrm{Id}_{\mathbb{R}^n}$

Consider the mapping:

$$\Phi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} - f(\mathbf{x}) + \mathbf{y} \qquad (\mathbf{x}, \mathbf{y} \in D)$$

Every fixed point of $\Phi_{\mathbf{y}}$ is the preimage of $\mathbf{y}$ under the mapping f. We next like to construct a closed ball $\overline{B}(0, 2r) = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}\| \leq 2r\}$ such that:

$$\Phi_{\mathbf{y}} : \overline{B}(0, 2r) \to \overline{B}(0, 2r)$$

and $\Phi_{\mathbf{y}}$ is contracting. Because $f'(\mathbf{0}) = \mathrm{Id}_{\mathbb{R}^n}$ and the continuity of f' there is a constant $r > 0$ such that for $\|\mathbf{x}\| \leq 2r$:

$$\|\mathrm{Id}_{\mathbb{R}^n} - f'(\mathbf{x})\| \leq \frac{1}{2}.$$

We have $\Phi_{\mathbf{0}}(\mathbf{x}) = \mathbf{x} - f(\mathbf{x})$ and $\Phi'_{\mathbf{0}}(\mathbf{x}) = \mathrm{Id}_{\mathbb{R}^n} - f'(\mathbf{x})$. From an application of the mean-value theorem we get the estimate:

$$\|\mathbf{x} - f(\mathbf{x})\| = \|\Phi_{\mathbf{0}}(\mathbf{x}) - \Phi_{\mathbf{0}}(\mathbf{0})\| = \|\Phi_{\mathbf{0}}(\nu\mathbf{x}, \mathbf{x})\| \leq \frac{1}{2}\|\mathbf{x}\|.$$

This estimate implies $\Phi_{\mathbf{y}}$ is a mapping for all $\mathbf{y} \in B(0, r)$ into $\overline{B}(0, 2r)$. We next show $\Phi_{\mathbf{y}}$ is contracting for every $\mathbf{y}$ fixed in $B(0, r)$. Take $\mathbf{x}_1, \mathbf{x}_2 \in \overline{B}(0, 2r)$ and apply the mean value theorem to get:

$$\Phi_{\mathbf{y}}(\mathbf{x}_2) - \Phi_{\mathbf{y}}(\mathbf{x}_1) = \Phi_{\mathbf{y}}(\mathbf{x}_1 + \nu(\mathbf{x}_2 - \mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1) \qquad (0 < \nu < 1)$$

Because $\Phi'_{\mathbf{y}} = \Phi'_{\mathbf{0}}$ and $\mathbf{x}_1 + \nu(\mathbf{x}_2 - \mathbf{x}_1) \in B(0, r)$ we get:

$$\|\Phi_{\mathbf{y}}(\mathbf{x}_2) - \Phi(\mathbf{x}_1)\| \leq \frac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|$$

We apply the Banach Fixed Point Theorem to $\Phi_{\mathbf{y}}$ and obtain a unique fixed point $\mathbf{x}$ in $\overline{B}(0, 2r)$, i.e. for every $\mathbf{y} \in B(0, r)$ there is *exactly* one $\mathbf{x} \in \overline{B}(0, 2r)$ such that:

$$f(\mathbf{x}) = \mathbf{y}.$$

The preimage of $B(0, r)$ is mapped bijectively to $B(0, r)$. Because f is continuous this preimage is an open set N. Therefore statements (i) and (ii) are proven.

Now we turn to (iii). We first look at the continuity of $f^{-1}|f(N)$, with $f(N) = B(0, r)$. Consider:

$$\begin{aligned} \Phi_{\mathbf{0}}(\mathbf{x}_2) &= \mathbf{x}_1 - f(\mathbf{x}_1), \\ \Phi_{\mathbf{0}}(\mathbf{x}_2) &= \mathbf{x}_2 - f(\mathbf{x}_2), \end{aligned} \qquad (\mathbf{x}_1, \mathbf{x}_2 \in \overline{B}(0, 2r))$$

We have:

$$\mathbf{x}_2 - \mathbf{x}_1 = \Phi_{\mathbf{0}}(\mathbf{x}_2) - \Phi_{\mathbf{0}}(\mathbf{x}_1) + f(\mathbf{x}_2) - f(\mathbf{x}_1)$$

and:
$$\|\mathbf{x}_2 - \mathbf{x}_1\| \leq \frac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_1\| + \|f(\mathbf{x}_2) - f(\mathbf{x}_1)\|.$$
It follows for $\mathbf{y}_1 = f(\mathbf{x}_1)$ and $\mathbf{y}_2 = f(\mathbf{x}_2)$ that:
$$\|f^{-1}(\mathbf{y}_2) - f^{-1}(\mathbf{y}_1)\| \leq 2\|\mathbf{y}_2 - \mathbf{y}_1\|.$$
This means f^{-1} is Lipschitz-continuous and therefore continuous on N.

We can now apply Theorem B2.4. It follows f^{-1} is differentiable on N. Consider the equality:
$$(f^{-1}(y))' = (f'(f^{-1}(\mathbf{y})))^{-1}$$
It follows $(f^{-1}(\mathbf{y}))'$ is continuous for every $\mathbf{y} \in N$.* This does imply $f^{-1}|N$ is a diffeomorphism and:
$$f^{-1}|N : N \to f(N).$$

□

*For $A \in M(n \times n) = \mathbb{R}^{n^2}$ consider a subset $U \subset \mathbb{R}^{n^2}$ and assume $\det(A) \neq 0$, U open, $A \in U$. Then:
$$A \mapsto A^{-1}$$
is a continuous mapping from U to U. Please check!

Example: Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be given by:
$$\begin{aligned} f_1(x_1, x_2, x_3) &= x_1 + x_2 + x_3 \\ f_2(x_1 + x_2 + x_3) &= x_2x_3 + x_3x_1 + x_1x_2 \\ f_3(x_1, x_2, x_3) &= x_1x_2x_3 \end{aligned}$$
These functions appear when calculating roots of a polynomial of degree 3:
$$(t - x_1)(t - x_2)(t - x_3) = t^3 - (x_1 + x_2 + x_3)t^2 + (x_2x_3 + x_3x + 1 + x_1x + 2_t - x_1x_2x_3.$$
f_1, f_2, f_3 are called "elementary symmetric functions" of x_1, x_2, x_3. We calculate $\det f'(x)$. We subtract the first line from the second and third to get:
$$\det f'(x) = \begin{vmatrix} 1 & 1 & 1 \\ x_2 + x_3 & x_3 + x_1 & x_1 + x_2 \\ x_2x_3 & x_3x_1 & x_1x_2 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ x_2 + x_3 & x_1 - x_2 & x_1 - x_3 \\ x_2x_3 & (x_1 - x_2)x_3 & (x_1 - x_3)x_2 \end{vmatrix} = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$$
We apply Theorem B6.1 to the equation:
$$\left.\begin{aligned} y_1 &= x_1 + x_2 + x_3, \\ y_2 &= x_2x_3 + x_3x_1 + x_1x_3, \\ y_3 &= x_1x_2x_3. \end{aligned}\right\} (*)$$
Assume above equation $(*)$ is solved by $x = \overline{x}$ and $y = \overline{y}$, and $(\overline{x_1} - \overline{x_2})(\overline{x_1}) - \overline{x_3})(\overline{x_2} - \overline{x_3}) \neq 0$. In this case the equation $(*)$ can also be solved for all $y \in f(N)$, with $f(N)$ a neighbourhood of $\overline{y}$ derived from Theorem B6.1:
$$\begin{aligned} x_1 &= f_1^{-1}(y_1, y_2, y_3) \\ x_2 &= f_2^{-1}(y_1, y_2, y_3) \\ x_3 &= f_3^{-1}(y_1, y_2, y_3) \end{aligned}$$
It also holds that:
$$(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \neq 0$$
for all x being the preimage of $y \in f(N)$. This implies the following for the cubic polynomial:
$$t^3 - y_1t^2 + y_2t - y_3 = 0 :$$
If this equation has three roots for $y = \overline{y}$ this remains valid for coefficients y in a neighbourhood of $\overline{y}$.

Nonlinear equations

Consider a system of m *non-linear* equations with n variables:

$$\begin{aligned} f_1(x_1, ..., x_n) &= 0 \\ &\vdots \\ f_m(x_1, ..., x_n) &= 0 \end{aligned} \qquad (m < n)$$

Assume $f_1, ..., f_m$ are continuously differentiable and one solution of the equation $\mathbf{s} = (s_1, ..., s_n)$ is known. We first look back at *linear* equations. If a linear system has rank m then we can express (after a possible relabelling) the first m variables through $x_{m+1}, ..., x_n$. This can best be done by the Gauss elimination procedure, which also answers the question about existence of a solution.

In the non-linear case it is not clear whether a solution exists at all. Here we assume that is the case. We will show that then there is a family of solutions depending on $m - n =: p$ parameters if the analogue to the rank condition is satisfied.

We label variables and parameters in the following way:

$$x_1, ..., x_m, \lambda_1, ..., \lambda_p$$

and look at the equation:

$$\begin{aligned} f_1(x_1, ..., x_m, \lambda_1, ..., \lambda_p) &= 0 \\ &\vdots \\ f_m(x_1, ..., x_m, \lambda_1, ..., \lambda_p) &= 0 \end{aligned}$$

and the condition:

$$\begin{vmatrix} D_1 f_1(\mathbf{s}) & \cdots & D_m f_1(\mathbf{s}) \\ \vdots & \ddots & \vdots \\ D_1 f_m(\mathbf{s}) & \cdots & D_m f_m(\mathbf{s}) \end{vmatrix} \neq 0$$

We know we can approximate $f_1, ..., f_m$ in a neighbourhood of $\mathbf{s}$ by linear functions which gives us hope we can express $x_1, ..., x_m$ as differentiable functions of $\lambda_1, ..., \lambda_p$ such that the nonlinear equation is satisfied for all points $(\lambda_1, ..., \lambda_p)$ in the neighbourhood of $(c_{m+1}, ..., c_n)$. We will use the following notation:

$$\begin{aligned} \mathbf{x} &:= (x_1, ..., x_m) & \boldsymbol{\lambda} &:= (\lambda_1, ..., \lambda_p) \\ \overline{\mathbf{x}} &:= (\overline{x_1}, ..., \overline{x_m}) & \overline{\boldsymbol{\lambda}} &:= (\overline{\lambda_1}, ..., \overline{\lambda_p}) \end{aligned}$$
$$(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) := \mathbf{s}.$$

The functions $f_1, ..., f_m$ are bundled into a vector function $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^m \times \mathbb{R}^p$.

The derivative $f'(\mathbf{c}, \boldsymbol{\lambda}; \mathbf{h}, \mathbf{k})$ can be splitted additively into:

$$f'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h}, \mathbf{k}) = f'_1(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h}) + f'_2(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{k}).$$

(This can easily be seen by using the Jacobi-matrix after fixing the standard basis). The rank condition becomes:

$$\det f'_1(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) \neq 0.$$

Theorem **B6.2** (Implicit Function Theorem)
Let $D \subset \mathbb{R}^m \times \mathbb{R}^p$ open and $f : D \to \mathbb{R}^m$ continuously differentiable. Further assume:

$$f(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) = 0 \text{ and } \det f'_1(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) \neq 0.$$

Then there is a neighbourhood N of $\overline{\boldsymbol{\lambda}}$ and a uniquely determined continuously differentiable mapping:

$$\varphi : N \to \mathbb{R}^m \qquad (N \subset \mathbb{R}^p)$$

such that for all $\boldsymbol{\lambda} \in N$:

$$f(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$$

Proof. Let $\mathbf{p} = (\mathbf{y}, \boldsymbol{\mu})$ and define:

$$(\mathbf{x}, \boldsymbol{\lambda}) \mapsto \mathbf{p} = F(\mathbf{x}, \boldsymbol{\lambda})$$

with:

$$\begin{aligned} \mathbf{y} &= f(\mathbf{x}, \boldsymbol{\lambda}), \\ \boldsymbol{\mu} &= \boldsymbol{\lambda}. \end{aligned}$$

F inherits the property that it is continuously differentiable. We have that:

$$\begin{aligned} F'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h}, \mathbf{k}) &= F_1'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h}) + F_2'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{k}) \\ &= (f_1'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{h}) + f_2'(\mathbf{x}, \boldsymbol{\lambda}; \mathbf{k}), \mathbf{k}). \end{aligned}$$

The derivative $F'(\mathbf{x}, \boldsymbol{\lambda})$ has therefore the following matrix form:

$$F'(\mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} f_1'(\mathbf{x}, \boldsymbol{\lambda}) & f_2'(\mathbf{x}, \boldsymbol{\lambda}) \\ 0 & \mathrm{Id}_{\mathbb{R}^n} \end{pmatrix}$$

Therefore:

$$\det F'(\mathbf{x}, \boldsymbol{\lambda}) = \det f_1' \mathbf{x}, \boldsymbol{\lambda}.$$

Because $\det f_1'(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) \neq 0$ by assumption we have:

$$\det F'(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) \neq 0$$

F does therefore satisfy the local invertibility theorem B6.1: there is a neighbourhood M of $(\overline{\mathbf{x}}, \overline{\boldsymbol{\lambda}}) \in D$ such that $F|M$ possesses a unique continuously differentiable inverse mapping:

$$F^{-1} : F(M) \to M.$$

Because $F(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{y}, \boldsymbol{\mu})$ one obtains for F^{-1}:

$$F^{-1}(\mathbf{y}, \boldsymbol{\mu}) = (\mathbf{x}, \boldsymbol{\lambda}),\ x = f^{-1}(\mathbf{y}, \boldsymbol{\mu}),\ \boldsymbol{\lambda} = \boldsymbol{\mu}.$$

Here f^{-1} is a continuously differentiable function with:

$$f^{-1} : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^m$$

The mapping f^{-1} determined the implicitly defined function φ by setting $\mathbf{y} = \mathbf{0}$. The domain of definition of $\mathbf{y}$ is the open set $F(M)$, with $(\mathbf{0}, \overline{\boldsymbol{\lambda}}) \in F(M)$. Therefore there is a neighbourhood $N \subset \mathbb{R}^p$ of $\overline{\boldsymbol{\lambda}}$ such that for all $\boldsymbol{\mu} \in N$ the family of points $(\mathbf{0}, \boldsymbol{\mu})$ belong to $F(M)$. For such $\boldsymbol{\mu}$ we have:

$$\varphi(\boldsymbol{\mu}) = f^{-1}(\mathbf{0}, \boldsymbol{\mu}).$$

Because $F \circ F^{-1} = \mathrm{Id}$ we also have:

$$f(f^{-1}(\mathbf{y}, \boldsymbol{\mu}), \boldsymbol{\mu}) = \mathbf{y} \text{ for } (\mathbf{y}, \boldsymbol{\mu}) \in F(M).$$

With $\mathbf{y} = \mathbf{0}$ and $\boldsymbol{\mu} = \boldsymbol{\lambda}$ one obtains:

$$f(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0 \text{ for all } \boldsymbol{\lambda} \in N$$

This shows that φ given by:

$$\varphi(\boldsymbol{\lambda}) := f^{-1}(\mathbf{0}, \boldsymbol{\lambda})$$

solves the equation:

$$f(\mathbf{x}, \boldsymbol{\lambda}) = 0$$

by parameterising $\mathbf{x}$ with $\boldsymbol{\lambda}$.

To show $\varphi(\lambda)$ is the only solution we look at:

$$f^{-1}(f(\mathbf{x}, \boldsymbol{\lambda}), \boldsymbol{\lambda}) = \mathbf{x}.$$

If $f(\mathbf{x}, \boldsymbol{\lambda}) = 0$ we obtain:

$$f^{-1}(\mathbf{0}, \boldsymbol{\lambda}) = \mathbf{x},$$
$$\mathbf{x} = \varphi(\boldsymbol{\lambda}).$$

After showing φ is differentiable we like to determine φ. We make use of the chain rule and note that:

$$f(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0 \text{ for all } \boldsymbol{\lambda} \in N.$$

The result is:

$$f_1'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}; \varphi'(\boldsymbol{\lambda}, \mathbf{h})) + f_2'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}; \mathbf{h}) = 0$$

Because $\det f_1'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \neq 0$ for $\boldsymbol{\lambda} \in N$ one can solve for $\varphi'(\boldsymbol{\lambda}, \mathbf{h})$:

$$\varphi'(\boldsymbol{\lambda}, \mathbf{h}) = -(f_1'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}))^{-1} f_2'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}; \mathbf{h}).$$

One can also leave away the argument $\boldsymbol{\lambda}$:

$$\varphi'(\boldsymbol{\lambda}) = -(f_1'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda})^{-1} \circ f_2'(\varphi(\boldsymbol{\lambda}), \boldsymbol{\lambda}).$$

The right side of this equation can be determined before $\mathbf{x} = \varphi(\boldsymbol{\lambda})$ is computed/from:

$$-(f_1'(\mathbf{x}, \boldsymbol{\lambda}))^{-1} \circ (f_2'(\mathbf{x}, \boldsymbol{\lambda})).$$

□

Examples:

1. If there is only one equation:

$$f(x_1, ..., x_n) = 0$$

one computes $\nabla f(\mathbf{x})$ and checks if a component is different from 0. If for example $D_n f(s_1, ..., s_n) \neq 0$ then the equation can be solved in a neighbourhood of $(s_1, ..., s_{n-1})$ for x_n:

$$x_n = \varphi(x_1, ..., x_{n-1})$$

Next one computes $\nabla\varphi$ from:

$$f(x_1, ..., x_{n-1}, \varphi(x_1, ..., x_{n-1})) = 0$$

noting that:

$$D_1 f + D_n f \cdot D_1 \varphi = 0$$
$$\vdots$$
$$D_{n-1} f + D_n f \cdot D_{n-1} \varphi = 0.$$

It follows:

$$\nabla\varphi = \left(-\frac{D_1 f}{D_n f}, ..., -\frac{D_{n-1} f}{D_n f}\right)$$

2. Consider two equations:

$$f_1(x_1, x_2, x_3) = 0,$$
$$f_2(x_1, x_2, x_3) = 0.$$

In general the intersection of two surfaces in $\mathbb{R}^n$ determines a curve:
This can be made more precise. One computes the Jacobi-matrix:

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{pmatrix}$$

If $\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1} \neq 0$ at a point for which $f_1 = f_2 = 0$ then there is a neighbourhood of this point such that x_1 and x_2 can be parameterised by x_3:

$$x_1 = \varphi_1(x_3)$$
$$x_2 = \varphi_2(x_3)$$

The derivative $\begin{pmatrix} \varphi_1' \\ \varphi_2' \end{pmatrix}$ can be computed from the equation:

$$\begin{pmatrix} \varphi_1' \\ \varphi_2' \end{pmatrix} = -\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} \end{pmatrix} = \begin{pmatrix} \dfrac{\frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_3} - \frac{\partial f_1}{\partial x_3}\frac{\partial f_2}{\partial x_2}}{\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_3}\frac{\partial f_2}{\partial x_1}} \\ \dfrac{\frac{\partial f_1}{\partial x_3}\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_3}}{\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1}} \end{pmatrix}$$

This result can also be obtained from the equations:

$$\frac{\partial f_1}{\partial x_1}\varphi_1' + \frac{\partial f_1}{\partial x_2}\varphi_2' + \frac{\partial f_1}{\partial x_3} = 0$$

and:

$$\frac{\partial f_2}{\partial x_1}\varphi_1' + \frac{\partial f_2}{\partial x_2}\varphi_2' + \frac{\partial f_2}{\partial x_3} = 0$$

This follows from applying the chain rule to:

$$f_1(\varphi_1(x_3), \varphi_2(x_3), x_3) = 0,$$
$$f_2(\varphi_1(x_3), \varphi_2(x_3), x_3) = 0.$$