

ANALYSIS III/IV : MATH 3011/4201

DIRK SCHÜTZ

1. METRIC SPACES

The completeness axiom of $\mathbb{R}$. The real numbers are a totally ordered field, as is $\mathbb{Q}$, but there is one important difference, $\mathbb{R}$ satisfies the *completeness Axiom*

(C) Every nonempty subset $A \subset \mathbb{R}$ which is bounded above has a supremum.

Definition 1.1. A *supremum* of a set A is a real number s such that $a \leq s$ for all $a \in A$, and if m is another real number with $a \leq m$ for all $a \in A$, then $s \leq m$. We denote $\sup A = s$, if a supremum exists.

An *infimum* of a set A is a real number s such that $a \geq s$ for all $a \in A$, and if m is another real number with $a \geq m$ for all $a \in A$, then $s \geq m$. We denote $\inf A = s$, if an infimum exists.

If s_1 and s_2 are both suprema, it follows that $s_1 \leq s_2$ and $s_2 \leq s_1$, so a supremum is unique. The same holds for the infimum. A supremum is sometimes called a least upper bound, and one should interpret it as the smallest upper bound of a set A .

Example. Let $A = \{x \in \mathbb{Q} \mid x^2 < 2\}$. Then $\sup A = \sqrt{2}$, which is not in $\mathbb{Q}$. Therefore the completeness Axiom does not hold for $\mathbb{Q}$.

The reals $\mathbb{R}$ do not have a supremum, but this set is not bounded above.

Lemma 1.2. Let $I_n = [a_n, b_n]$ be a sequence of closed intervals with $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$ for all $n \geq 1$. Then

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset.$$

Proof. Let $a = \sup\{a_n \mid n \geq 1\}$, which exists by (C), as $a_n \leq b_1$ shows that the set is bounded above. Therefore $a_n \leq a$ for all $n \geq 1$. But we also have $a_n \leq b_k$ for every $n, k \geq 1$. By the supremum property we get $a \leq b_k$ for all $k \geq 1$. Therefore $a \in [a_n, b_n]$ for all $n \geq 1$ and a is in the intersection. $\square$

Metric spaces.

Definition 1.3. Let M be a set. A function $d : M \times M \rightarrow [0, \infty)$ is called a *metric* on M , if it satisfies

- (i) $d(x, y) = 0 \Leftrightarrow x = y$.
- (ii) $d(x, y) = d(y, x)$ for all $x, y \in M$ (*symmetry*).
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in M$ (*triangle inequality*).

Date: January 2, 2008.

The pair (M, d) is called a *metric space*.

We will often just call M a metric space and suppress the metric d , if this does not cause confusion.

Example. Let $M = \mathbb{R}$ and $d(x, y) = |x - y|$. The first two conditions are easy, the triangle inequality follows from

$$|a + b| \leq |a| + |b|$$

for all $a, b \in \mathbb{R}$, which can be seen by checking the various cases of the signs of a and b .

Now if (M, d) is a metric space and $N \subset M$, we get a metric on N by restricting d to $N \times N$. Therefore $\mathbb{Q}$, $\mathbb{Z}$, $\{0\}$, $(0, 1)$, $[a, b]$ are all metric spaces with the same metric as for $\mathbb{R}$.

Let $M = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, we can define a metric on the plane by setting

$$d_1((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| + |x_2 - y_2|.$$

Notice that $d_1((1, 1), (0, 0)) = d_1(1, 1), (1, 0)) + d_1((1, 0), (0, 0)) = 2$, which may not be what you expect as the distance of these two points. To get the standard metric on the plane, we set

$$d_2((x_1, x_2), (y_1, y_2)) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2}$$

The triangle inequality is geometrically obvious, but for an algebraic proof we want to introduce some linear algebra.

Definition 1.4. Let V be an $\mathbb{R}$ -vector space. An *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$, such that the following hold:

- (i) $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$ for all $x_1, x_2, y \in V$.
- (ii) $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$.
- (iii) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ for all $\lambda \in \mathbb{R}$ and $x, y \in V$.
- (iv) $\langle x, x \rangle \geq 0$ for all $x \in V$.
- (v) $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

Example. The standard example is $V = \mathbb{R}^n$ with

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$$

where $x = (x_1, \dots, x_n)$ with $x_i \in \mathbb{R}$ for all $i = 1, \dots, n$ and similarly for y . If we think of x and y as column vectors, we get $\langle x, y \rangle = x^t y$ by matrix-multiplication, where x^t is the transpose of x , a row vector.

Now if A is any symmetric $n \times n$ -matrix over $\mathbb{R}$, the formula

$$(x, y) = x^t A y$$

satisfies (i), (ii) and (iii) from Definition 1.4. In order to satisfy (iv) and (v), we need that the eigenvalues of A are all positive.

To get an example for an infinite dimensional vector space, we let

$$C[a, b] = \{f : [a, b] \rightarrow \mathbb{R} \mid f \text{ continuous}\}$$

Then

$$\langle f, g \rangle = \int_a^b f(x)g(x) dx$$

defines an inner product on $V = C[a, b]$. The most difficult point is (v). To see this, note that if f is not the zero-function, then $(f(x))^2 > 0$ for some $x \in [a, b]$, and by continuity we get $(f(y))^2 \geq c > 0$ for some small interval $[u, v]$ containing x and some small $c < (f(x))^2$. It follows that $\langle f, f \rangle \geq (v - u)c > 0$.

Cauchy-Schwarz Inequality. *Let V be an $\mathbb{R}$ -vector space and $\langle \cdot, \cdot \rangle$ be an inner product on V . Then*

$$|\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle} \cdot \sqrt{\langle y, y \rangle}$$

for all $x, y \in V$, with equality exactly when x and y are linearly dependent.

Proof. Note that $\langle x, 0 \rangle = \langle x, 0 \cdot x \rangle = 0 \langle x, x \rangle = 0$, so we assume $y \neq 0$. Now

$$\begin{aligned} (1) \quad 0 &\leq \langle x + \lambda y, x + \lambda y \rangle \\ &= \langle x, x \rangle + 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle. \end{aligned}$$

With $\lambda = -\frac{\langle x, y \rangle}{\langle y, y \rangle}$, this gives

$$\begin{aligned} 0 &\leq \langle x, x \rangle - 2 \frac{\langle x, y \rangle^2}{\langle y, y \rangle} + \frac{\langle x, y \rangle^2}{\langle y, y \rangle}, \quad \text{i.e.} \\ \langle x, y \rangle^2 &\leq \langle x, x \rangle \cdot \langle y, y \rangle. \end{aligned}$$

Taking the square-root gives the inequality.

If this is an equality, we also have equality in (1), which implies that $x + \lambda y = 0$, so that x and y are linearly dependent. Also, if x and y are linearly dependent, we can assume that $x = \lambda y$ for some $\lambda \in \mathbb{R}$ and it is easy to check that equality holds. $\square$

Lemma 1.5. *Let V be an $\mathbb{R}$ -vector space with inner product $\langle \cdot, \cdot \rangle$. Then $d : V \times V \rightarrow [0, \infty)$ defined by*

$$d(x, y) = \sqrt{\langle x - y, x - y \rangle}$$

gives a metric on V .

For simplicity, we write $\|x\| = \sqrt{\langle x, x \rangle}$ for $x \in V$, so that $d(x, y) = \|x - y\|$.

Proof. We only check the triangle inequality, let $a, b \in V$. Then

$$\begin{aligned} \|a + b\| &= \sqrt{\langle a + b, a + b \rangle} \\ &= \sqrt{\langle a, a \rangle + 2\langle a, b \rangle + \langle b, b \rangle} \\ &\leq \sqrt{\|a\|^2 + 2\|a\| \|b\| + \|b\|^2} \\ &= \|a\| + \|b\| \end{aligned}$$

where we used the Cauchy-Schwarz Inequality in the third line. The triangle inequality now follows from

$$d(x, z) = \|x - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z). \quad \square$$

Example. On $\mathbb{R}^n$ we have the metric

$$d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Definition 1.6. Let $f : M \rightarrow N$ be a function between metric spaces (M, d_M) and (N, d_N) . Then f is called *continuous at* $a \in M$, if for all $\varepsilon > 0$ there is a $\delta > 0$ with $d_N(f(x), f(a)) < \varepsilon$ for all $x \in M$ with $d_M(x, a) < \delta$.

If f is continuous at every $a \in M$, we call f *continuous*.

Remark. If $M, N \subset \mathbb{R}$, the statement for continuity at $a \in M$ reads as:

For all $\varepsilon > 0$ there is a $\delta > 0$ with $|f(x) - f(a)| < \varepsilon$ for all $x \in M$ with $|a - x| < \delta$.

Sequences and Cauchy sequences.

Definition 1.7. Let M be a metric space. A sequence $(a_n)_{n \in \mathbb{N}}$ in M consists of elements $a_n \in M$ for every non-negative integer n . Let $a \in M$. We say that the sequence $(a_n)_{n \in \mathbb{N}}$ *converges to* $a \in M$, if for every $\varepsilon > 0$ there is a $n_0 \geq 0$ such that $d(a_n, a) < \varepsilon$ for all $n \geq n_0$. We write $a = \lim_{n \rightarrow \infty} a_n$ or $a_n \rightarrow a$ for $n \rightarrow \infty$ and call a the *limit* of $(a_n)_{n \in \mathbb{N}}$, if $(a_n)_{n \in \mathbb{N}}$ converges to a .

Remark. The limit of a sequence is unique, if it exists. For if $b \in M$ is also a limit, then

$$d(a, b) \leq d(a, a_n) + d(a_n, b),$$

and the right side can be made arbitrary small by an appropriate choice of n , so $d(a, b) = 0$ and $a = b$.

Lemma 1.8. Let $f : M \rightarrow N$ be a function between metric spaces and $a \in M$. Then f is continuous at $a \in M$ if and only if $\lim_{n \rightarrow \infty} f(a_n) = f(a)$ for all sequences $(a_n)_{n \in \mathbb{N}}$ in M with $a_n \rightarrow a$ for $n \rightarrow \infty$.

Proof. Assume that f is continuous at a . Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in M with $a_n \rightarrow a$. Given $\varepsilon > 0$, there is a $\delta > 0$ such that $d(a, x) < \delta$ implies $d(f(a), f(x)) < \varepsilon$. Choose $n_0 \geq 0$ with $d(a, a_n) < \delta$ for $n \geq n_0$. Then $d(f(a), f(a_n)) < \varepsilon$ for $n \geq n_0$ and $f(a_n) \rightarrow f(a)$.

Now assume $\lim_{n \rightarrow \infty} f(a_n) = f(a)$ for all sequences $(a_n)_{n \in \mathbb{N}}$ in M with $a_n \rightarrow a$. Given $\varepsilon > 0$, assume that no $\delta > 0$ exists with $d(a, x) < \delta$ implying $d(f(a), f(x)) < \varepsilon$. Then we can find $a_n \in M$ with $d(a, a_n) < \frac{1}{n}$ and $d(f(a), f(a_n)) \geq \varepsilon$. But then $a_n \rightarrow a$, while $f(a_n) \not\rightarrow f(a)$, which is a contradiction. Therefore, given $\varepsilon > 0$ the appropriate $\delta > 0$ exists and f is continuous at a . $\square$

Definition 1.9. Let M be a metric space. A sequence $(a_n)_{n \in \mathbb{N}}$ in M is called a *Cauchy sequence*, if for every $\varepsilon > 0$ there is an $n_0 \geq 0$ such that for all $p, q \geq n_0$ we have $d(a_q, a_p) < \varepsilon$.

This is similar to a convergent sequence, but notice that we do not require an element to converge to.

Lemma 1.10. Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence in a metric space M . Then $(a_n)_{n \in \mathbb{N}}$ is a Cauchy sequence.

Proof. Let $a \in M$ such that $a_n \rightarrow a$. Given $\varepsilon > 0$, there exists an n_0 with $d(a, a_n) < \frac{\varepsilon}{2}$ for all $n \geq n_0$. So if $p, q \geq n_0$, we get

$$\begin{aligned} d(a_p, a_q) &\leq d(a_p, a) + d(a, a_q) \\ &< \varepsilon, \end{aligned}$$

so $(a_n)_{n \in \mathbb{N}}$ is a Cauchy sequence. $\square$

Remark. Cauchy sequences need not be convergent: take $M = (0, 1)$ and $a_n = \frac{1}{n}$. Then $|a_p - a_q| < \frac{1}{n_0}$ if $p, q \geq n_0$, so it is a Cauchy sequence, but it does not converge in M , of course only because $0 \notin (0, 1)$.

Another example is given by $M = \mathbb{Q}$ and we define a sequence in $\mathbb{Q}$ by letting $a_0 = 3$, $a_1 = 3.1$, $a_2 = 3.14$ and so on to achieve $a_n \rightarrow \pi$. Again, as $\pi \notin \mathbb{Q}$, the sequence is a Cauchy sequence which does not converge (in $\mathbb{Q}$).

Definition 1.11. A metric space is called *complete*, if every Cauchy sequence is convergent.

Theorem 1. $\mathbb{R}$ is complete.

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}$. Define a sequence of integers n_k as follows: $n_0 = 1$, n_{k+1} is the smallest integer $> n_k$ with

$$|a_p - a_q| < \frac{1}{2^{k+2}}$$

for all $p, q \geq n_{k+1}$. This sequence exists, as a_n form a Cauchy sequence. Let $I_k = [a_{n_k} - \frac{1}{2^k}, a_{n_k} + \frac{1}{2^k}]$.

Then $I_{k+1} \subset I_k$: If $x \in I_{k+1}$, then $|x - a_{n_{k+1}}| < \frac{1}{2^{k+1}}$. Also, $|a_{n_k} - a_{n_{k+1}}| < \frac{1}{2^{k+1}}$ by construction of n_k . Therefore

$$\begin{aligned} |x - a_{n_k}| &\leq |x - a_{n_{k+1}}| + |a_{n_{k+1}} - a_{n_k}| \\ &< \frac{1}{2^{k+1}} + \frac{1}{2^{k+1}} = \frac{1}{2^k}, \end{aligned}$$

so $x \in I_k$. By Lemma 1.2, there is $a \in \bigcap_{k=1}^{\infty} I_k$, and for $m \geq n_k$ we have

$$\begin{aligned} |a - a_m| &\leq |a - a_{n_k}| + |a_{n_k} - a_m| \\ &< \frac{1}{2^k} + \frac{1}{2^{k+1}} < \frac{1}{2^{k-1}}. \end{aligned}$$

This shows that $a_n \rightarrow a$. $\square$

Example. Let X be a set and

$$B(X) = \{f : X \rightarrow \mathbb{R} \mid f \text{ is bounded}\}$$

Bounded means that there is a $C_f \geq 0$ with $|f(x)| \leq C_f$ for all $x \in X$. For $f, g \in B(X)$, we define

$$d(f, g) = \sup\{|f(x) - g(x)| \in \mathbb{R} \mid x \in X\}$$

The supremum exists by the completeness axiom, as $|f(x) - g(x)| \leq C_f + C_g$ for all $x \in X$.

Proposition 1.12. $(B(X), d)$ is a complete metric space.

Proof. It is straightforward to check that d defines a metric on $B(X)$. Let $(f_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $B(X)$. Then for every $x \in X$, $(f_n(x))_{n \in \mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$:

$$|f_q(x) - f_p(x)| \leq d(f_q, f_p).$$

By Theorem 1, these sequences converge.

Define $f : X \rightarrow \mathbb{R}$ by $f(x) = \lim_{n \rightarrow \infty} f_n(x)$. We need that $f \in B(X)$ and $f_n \rightarrow f$. For this let $\varepsilon > 0$. There exists n_0 with $d(f_p, f_q) < \frac{\varepsilon}{2}$, and for every $x \in X$ there is a $n_1(x) \geq n_0$ with $|f_{n_1(x)}(x) - f(x)| < \frac{\varepsilon}{2}$.

So now let $x \in X$ and $n \geq n_0$. Then

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x) - f_{n_1(x)}(x)| + |f_{n_1(x)}(x) - f(x)| \\ (2) \quad &< d(f_n, f_{n_1(x)}) + \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

In particular,

$$|f(x)| \leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x)| \leq C_{f_{n_0}} + \varepsilon$$

for all $x \in X$, so that f is bounded. Furthermore, since (2) holds for all $x \in X$, we get that $d(f, f_n) \leq \varepsilon$ for all $n \geq n_0$, which gives that $f_n \rightarrow f$. $\square$

Topology of metric spaces.

Definition 1.13. Let (M, d) be a metric space, $x \in M$, $r > 0$. Define

$$B(x; r) = \{y \in M \mid d(x, y) < r\}$$

the *open ball around x of radius r* , and

$$D(x; r) = \{y \in M \mid d(x, y) \leq r\}$$

the *closed ball around x of radius r* .

A set $A \subset M$ is called *bounded*, if $A \subset D(x; r)$ for some $x \in M$, $r > 0$.

Example. Let $M = \mathbb{R}$, then $B(0; 1) = (-1, 1)$ and $D(0; 1) = [-1, 1]$. Note that $\mathbb{R}$ is not bounded.

In $\mathbb{R}^2$, these balls are nice and round, if we use the metric d_2 .

Definition 1.14. Let M be a metric space. A set $U \subset M$ is called *open*, if for every $x \in U$ there is an $r_x > 0$ such that $B(x; r_x) \subset U$.

Lemma 1.15. Let M be a metric space.

- (i) $M, \emptyset$ are open sets.
- (ii) Any union of open sets is open.
- (iii) A finite intersection of open sets is open.
- (iv) For every $x \in M$ and $r > 0$, $B(x; r)$ is open.

Proof. (i) For $x \in M$, we can choose any $r > 0$ and get $B(x; r) \subset M$, so M is open. As there are no $x \in \emptyset$, we do not have to check anything and get that $\emptyset$ is open.

(ii) Let U_α be open for $\alpha \in \mathcal{A}$. If $x \in \bigcup_{\alpha \in \mathcal{A}} U_\alpha$, then $x \in U_\beta$ for some $\beta \in \mathcal{A}$. There is $r_x > 0$ with $B(x; r_x) \subset U_\beta$, and therefore $B(x; r_x) \subset \bigcup_{\alpha \in \mathcal{A}} U_\alpha$.

(iii) Let U_i be open for $i = 1, \dots, n$. If $x \in \bigcap_{i=1}^n U_i$, then $x \in U_i$ for all $i = 1, \dots, n$ and so there is $r_i > 0$ with $B(x; r_i) \subset U_i$ for all $i = 1, \dots, n$. Choose $r = \min\{r_1, \dots, r_n\} > 0$, then $B(x; r) \subset \bigcap_{i=1}^n U_i$.

(iv) If $y \in B(x; r)$, then $d(x, y) < r$. Let $r_y = r - d(x, y) > 0$, then if $d(z, y) < r_y$, we get

$$d(z, x) \leq d(z, y) + d(y, x) < r_y + d(x, y) = r,$$

so $B(y; r_y) \subset B(x; r)$, and $B(x; r)$ is open. $\square$

Definition 1.16. Let M be a metric space. A subset $A \subset M$ is called *closed*, if its complement $M - A = \{x \in M \mid x \notin A\}$ is open.

Lemma 1.17. Let M be a metric space.

- (i) $M, \emptyset$ are closed sets.
- (ii) Any intersection of closed sets is closed.
- (iii) A finite union of closed sets is closed.
- (iv) For every $x \in M$ and $r > 0$, $D(x; r)$ is closed.

Proof. (i), (ii) and (iii) follow from Lemma 1.15 by standard properties of complements.

For (iv), let $y \notin D(x; r)$, so $d(x, y) > r$. Let $r_y = d(x, y) - r > 0$. If $d(z, y) < r_y$, then

$$\begin{aligned} d(x, z) &\geq d(x, y) - d(y, z) \\ &= r + r_y - d(y, z) > r. \end{aligned}$$

Therefore $B(y; r_y) \subset M - D(x; r)$ and $M - D(x; r)$ is open, hence $D(x; r)$ is closed by definition. $\square$

Example. Let $M = \mathbb{R}$, and $a < b$ be two real numbers. Then the open interval $(a, b) = B((a+b)/2; (b-a)/2)$ is open. Also, the intervals (a, ∞) and $(-\infty, b)$ are open as unions of bounded open intervals. The closed interval $[a, b] = D((a+b)/2; (b-a)/2)$ is closed, and $(-\infty, b] = \mathbb{R} - (b, \infty)$ is also closed.

The integers $\mathbb{Z}$ is closed as it is the complement of $\bigcup_{n=-\infty}^{\infty} (n, n+1)$, a union of open intervals.

The rational numbers $\mathbb{Q}$ are neither open nor closed, so are the intervals $(a, b]$ and $[a, b)$.

Warning. A closed set is not a set which is not open. There are sets which are both open and closed, and there are sets which are neither open nor closed.

Proposition 1.18. Let M be a metric space and $A \subset M$. Then A is closed, if and only if for every sequence $(a_n)_{n \in \mathbb{N}}$ in A which converges to $a \in M$, we have $a \in A$.

Proof. “ $\Leftarrow$ ”: Let A be closed and $(a_n)_{n \in \mathbb{N}}$ a sequence in A which converges to $a \in M$. Assume that $a \in M - A$, which is open. Then there exists $r > 0$ with $B(a; r) \subset M - A$. But $a_n \rightarrow a$ implies that $a_n \in B(a; r)$ for $n \geq n_0$ with some n_0 . This contradicts $a_n \in A$ for all $n \in \mathbb{N}$.

“ $\Rightarrow$ ”: Let $a \in M - A$, we need $r > 0$ such that $B(a; r) \subset M - A$. Assume otherwise. Then for every n we have $B(a; \frac{1}{n}) \cap A \neq \emptyset$, so there is $a_n \in A \cap B(a; \frac{1}{n})$. In particular, the a_n form a sequence in A , and $a_n \rightarrow a$ as $n \rightarrow \infty$. By assumption we get $a \in A$, a contradiction. $\square$

Theorem 2. *Let M be a complete metric space and $A \subset M$ closed. Then A with the induced metric is complete.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in A . This is also a Cauchy sequence in M , so we have $a_n \rightarrow a$ for some $a \in M$. By Proposition 1.18, we get $a \in A$, so A is complete. $\square$

Definition 1.19. Let M be a metric space. Then M is called *compact*, if every sequence $(a_n)_{n \in \mathbb{N}}$ has a convergent subsequence $(a_{n_k})_{k \in \mathbb{N}}$.

Example. The interval $(0, 1)$ is not compact. The sequence given by $a_n = \frac{1}{n+1} \in (0, 1)$ is not convergent in $(0, 1)$.

Example. The closed interval $[0, 1]$ is compact: Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1]$. Either $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$ contains infinitely many elements of the a_n (possibly both). Let I_1 be such an interval which contains infinitely many of the a_n . Listing these elements gives a subsequence $(a_{n_k}^1)_{k \in \mathbb{N}}$ in I_1 . Of this sequence we can find a subsequence $(a_{n_k}^2)_{k \in \mathbb{N}}$ with elements in I_2 , which is one half of I_1 .

Continuing, we get a sequence of intervals $I_m \supset I_{m+1}$ with

$$(3) \quad \text{diam } I_m = \frac{1}{2^m}$$

and sequences $(a_{n_k}^m)_{k \in \mathbb{N}}$ in I_m , each a subsequence of $(a_n)_{n \in \mathbb{N}}$. Define $b_m = a_{n_m}^m \in I_m$, a subsequence of $(a_n)_{n \in \mathbb{N}}$, which is a Cauchy sequence by (3). By Theorem 1 and 2, this converges in $[0, 1]$.

Proposition 1.20. *The cube $[a, b]^n \subset \mathbb{R}^n$ is compact.*

Proof. We can use the same argument by splitting $[a, b]^n$ into 2^n smaller cubes, or first find a subsequence where the first coordinate converges, then get a subsequence of this new sequence where the second coordinate converges. Continuing gives a sequence where all coordinates converge. Note that a subsequence of a convergent sequence is convergent. $\square$

Proposition 1.21. *Let $f : M \rightarrow N$ be continuous map between metric spaces with M compact. Then the image of f , $f(M) \subset N$, is compact.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in $f(M)$. Thus $a_n = f(b_n)$ for some $b_n \in M$. This has a convergent subsequence $(b_{n_k})_{k \in \mathbb{N}}$ in M , i.e. $b_{n_k} \rightarrow b \in M$ as $k \rightarrow \infty$. By Lemma 1.8 we get that $a_{n_k} = f(b_{n_k})$ satisfies $\lim_{k \rightarrow \infty} a_{n_k} = \lim_{k \rightarrow \infty} f(b_{n_k}) = f(b) \in f(M)$. $\square$

Proposition 1.22. *A closed subset A of a compact metric space M is compact.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in A , so it is also a sequence in M . Therefore it has a convergent subsequence $a_{n_k} \rightarrow a \in M$ for $k \rightarrow \infty$. As the subsequence is a sequence in A , we get $a = \lim_{k \rightarrow \infty} a_{n_k} \in A$ by Proposition 1.18. Therefore $(a_n)_{n \in \mathbb{N}}$ has a convergent subsequence in A . $\square$

Heine-Borel Theorem. *A subset $A \subset \mathbb{R}^n$ is compact, if and only if A is closed and bounded.*

Proof. “ $\implies$ ”: As A is bounded, we get $A \subset [a, b]^n$ for appropriate $a, b \in \mathbb{R}$. By Propositions 1.20 and 1.22 we get that A is compact.

“ $\impliedby$ ”: It follows from Proposition 1.18 that a compact A is closed. If A were not bounded, we could find a sequence $a_n \in A$ with $d(a_n, 0) \geq n$. Such a sequence cannot converge to any $x \in \mathbb{R}^n$ and every subsequence behaves the same way. $\square$

2. NORMED VECTOR SPACES

Banach and Hilbert spaces. The metric spaces we met so far were all subsets of vector spaces, and the metrics were usually obtained by measuring the difference of two elements in the vector space.

Definition 2.1. Let V be an $\mathbb{R}$ vector space. A *norm* on V is a function $\|\cdot\| : V \rightarrow [0, \infty)$ satisfying the following:

- (i) $\|x\| = 0 \Leftrightarrow x = 0$.
- (ii) $\|\lambda \cdot x\| = |\lambda| \cdot \|x\|$ for all $\lambda \in \mathbb{R}$ and $x \in V$.
- (iii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

The pair $(V, \|\cdot\|)$ is called a *normed vector space*.

Lemma 2.2. Let V be a normed vector space. Then $d(x, y) = \|x - y\|$ defines a metric on V . $\square$

Example. $\mathbb{R}$ with modulus $|\cdot|$, $\mathbb{C}$ with modulus $|\cdot|$.

If V has an inner product $\langle \cdot, \cdot \rangle$, define $\|x\| = \sqrt{\langle x, x \rangle}$.

On $\mathbb{R}^n$ we have the norms

$$\begin{aligned} \|x\|_1 &= \sum_{i=1}^n |x_i| \\ \|x\|_2 &= \sqrt{\sum_{i=1}^n |x_i|^2} \\ \|x\|_\infty &= \max\{|x_i| \mid i = 1, \dots, n\} \end{aligned}$$

Let X be a set, then $B(X)$ with

$$(4) \quad \|f\| = \sup\{|f(x)| \in \mathbb{R} \mid x \in X\},$$

is a normed vector space. This norm is called the *supremum-norm*.

Let X be a metric space and

$$C_b(X) = \{f : X \rightarrow \mathbb{R} \mid f \text{ is continuous and has bounded image}\}.$$

This is a subspace of $B(X)$ and we can use the supremum-norm to get a normed vector space.

Note that

$$C(X) = \{f : X \rightarrow \mathbb{R} \mid f \text{ is continuous}\}$$

cannot be given a norm by (4) as the supremum does not exist. If X is compact, the two spaces agree:

Proposition 2.3. Let X be a compact metric space. Then $C(X) = C_b(X)$.

Proof. By Proposition 1.21 $f(X) \subset \mathbb{R}$ is compact for every $f \in C(X)$. By the Heine-Borel Theorem we get that f is bounded, so $f \in C_b(X)$. $\square$

Remark. Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A function $f : V \rightarrow W$ is continuous at $x \in V$, if for all $\varepsilon > 0$ there is a $\delta > 0$ such that $\|x - y\|_V < \delta$ implies $\|f(x) - f(y)\|_W < \varepsilon$.

Definition 2.4. Let $(V, \|\cdot\|)$ be a normed vector space. We call V a *Banach space*, if V is complete with the metric induced by the norm.

Example. We have that $\mathbb{R}^n$ is a Banach space in all three metrics.

The space $B(X)$ with X a set, together with the supremum-norm is a Banach space by Proposition 1.12.

Theorem 3. Let X be a metric space. Then $C_b(X)$ with the supremum-norm is a Banach space.

Proof. As $C_b(X) \subset B(X)$ and the norm on $C_b(X)$ is induced by the norm on $B(X)$, we only have to show that $C_b(X)$ is a closed subset of $B(X)$. So let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $C_b(X)$ converging to $f \in B(X)$. We need to show that f is continuous to get $f \in C_b(X)$.

Let $\varepsilon > 0$ and $x \in X$. There is $n_0 \geq 0$ such that $\|f_n - f\| < \frac{\varepsilon}{3}$ for all $n \geq n_0$. In particular $|f_{n_0}(y) - f(y)| < \frac{\varepsilon}{3}$ for all $y \in X$. As f_{n_0} is continuous, there is a $\delta > 0$ with $d(x, y) < \delta$ implying $|f_{n_0}(x) - f_{n_0}(y)| < \frac{\varepsilon}{3}$. Hence for $d(x, y) < \delta$

$$|f(x) - f(y)| \leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x) - f_{n_0}(y)| + |f_{n_0}(y) - f(y)| < \varepsilon$$

which shows that f is continuous. By Proposition 1.18, $C_b(X)$ is closed in $B(X)$ and by Theorem 2 we get the result. $\square$

Corollary 2.5. Let $a < b$ be two real numbers. Then $C[a, b]$ with the supremum norm is a Banach space. $\square$

Remark. Convergence of a sequence in $C_b(X)$ with respect to the supremum norm is called *uniform convergence*, a concept known from Complex Analysis II.

Remark. We can define a different norm on $C[a, b]$ using the inner product

$$(5) \quad \langle f, g \rangle = \int_a^b f(x)g(x) dx$$

It turns out that $C[a, b]$ with the norm induced from this inner product is not complete. To show this, start with the following example.

Example. Let $\|\cdot\|_2$ be the norm on $C[0, 1]$ coming from the inner product (5). Define a sequence $(f_n)_{n \in \mathbb{N}}$ in $C[0, 1]$ by $f_n(x) = x^n$. Then f_n converges to the 0-function:

$$\begin{aligned} \|f_n - 0\|_2 &= \int_0^1 (x^n)^2 dx \\ &= \frac{1}{2n+1} \end{aligned}$$

so indeed $\lim_{n \rightarrow \infty} f_n = 0$. Notice that $f_n(1) = 1$ for all n , but this does not hold for the limit. It is very important for the convergence here that we use norm from the inner product.

If we use the supremum-norm, the sequence does not converge to 0, it is not even a Cauchy sequence and does not converge at all. To see this let $n \in \mathbb{N}$. As x^n is continuous, we can find $0 < x_0 < 1$ with $x_0^n \geq \frac{3}{4}$. But $\lim_{k \rightarrow \infty} x_0^k = 0$, so we can also find k with $x_0^k < \frac{1}{4}$. Therefore $|x_0^k - x_0^n| \geq \frac{1}{2}$, so $\|f_n - f_k\| \geq \frac{1}{2}$. As n can be arbitrary, the f_n cannot be a Cauchy sequence with respect to the supremum norm.

Proposition 2.6. $C[a, b]$ with the norm $\|\cdot\|_2$ coming from the inner product (5) is not complete.

Proof. Let us start with $a = 0$ and $b = 2$. The example given in Exercise 3 from Problem set 1 gives a Cauchy sequence. Assume this sequence converges to $f \in C[0, 2]$. By restricting to $[0, 1]$ we get the Cauchy sequence $f_n|_{[0, 1]}$ converging to $f|_{[0, 1]}$. With an argument as in the example above, we get that $f|_{[0, 1]}$ is the zero-function. But by restricting to $[1, 2]$ we get that f is constant to 1 on that interval. This violates continuity of f . The case with arbitrary $[a, b]$ is no different. $\square$

Definition 2.7. Let $(V, \|\cdot\|)$ be a Banach space. If there is an inner product $\langle \cdot, \cdot \rangle$ on V such that $\|x\| = \sqrt{\langle x, x \rangle}$ for all $x \in V$, then V is called a *Hilbert space*.

Note that $C[a, b]$ with the norm induced by the inner product is not a Hilbert space, as it is not complete. This problem can be rectified in the following way, although we will not give a proof of this theorem.

Theorem 4. Let (M, d) be a metric space. There exists a complete metric space $(\bar{M}, \bar{d})$ and an embedding $i : M \rightarrow \bar{M}$ with $d(x, y) = \bar{d}(i(x), i(y))$ for all $x, y \in M$, such that for every point $\bar{x} \in \bar{M}$ there exists a sequence $(x_n)_{n \in \mathbb{N}}$ in M with $i(x_n) \rightarrow \bar{x}$ for $n \rightarrow \infty$. $\square$

Remark. The complete metric space $\bar{M}$ is called the *completion* of M . It is unique in the following sense: If (M', d') is another complete metric space with embedding $j : M \rightarrow M'$ satisfying $d(x, y) = d'(j(x), j(y))$ for all $x, y \in M$, then there is an embedding $\bar{i} : \bar{M} \rightarrow M'$ with $\bar{i} \circ i = j$ and $\bar{d}(x, y) = d'(\bar{i}(x), \bar{i}(y))$ for all $x, y \in \bar{M}$.

Example. The completion of $C[a, b]$ with the inner product metric is denoted $L^2([a, b])$. The inner product can be extended to $L^2([a, b])$ so that this space is a Hilbert space. The problem with this completion is that elements cannot be directly considered as functions. Nevertheless, $L^2([a, b])$ is still a very useful space.

Remark. Hilbert spaces are a special kind of Banach spaces. They have nicer properties, in particular, Linear Algebra is easier with Hilbert spaces than with general Banach spaces.

A norm for matrix spaces. Let $M_{n,k}(\mathbb{R})$ denote the $\mathbb{R}$ -vector space of $n \times k$ -matrices. As a vector space, this can be identified with $\mathbb{R}^{n \cdot k}$ so we have various norms at our disposal. But we want to introduce a norm which reflects the fact that such matrices correspond to linear maps.

Let $A = (A_{ij}) \in M_{n,k}(\mathbb{R})$ and define

$$(6) \quad \|A\| = \sup\{\|Ax\|_2 \mid x \in \mathbb{R}^k, \|x\|_2 \leq 1\}.$$

Here $Ax \in \mathbb{R}^n$, so we use the standard Euclidean norm. Note that we take the supremum of a compact set, as $x \mapsto \|Ax\|_2$ is a continuous map. By the Heine-Borel Theorem there is an $x \in \mathbb{R}^k$ with $\|x\|_2 \leq 1$ and $\|A\| = \|Ax\|_2$.

Remark. For $A \in M_{n,1}(\mathbb{R})$ or $a \in M_{1,k}(\mathbb{R})$ we get $\|A\| = \|A\|_2$, but in general these norms are different. Note that $\|I\| = 1$, where $I \in M_{n,n}(\mathbb{R})$ is the identity matrix.

Lemma 2.8. *The formula (6) defines a norm on $M_{n,k}(\mathbb{R})$ with*

$$(7) \quad \|Ax\|_2 \leq \|A\| \cdot \|x\|_2$$

for all $x \in \mathbb{R}^k$. Furthermore

$$(8) \quad \|AB\| \leq \|A\| \cdot \|B\|$$

with $A \in M_{n,k}(\mathbb{R})$ and $B \in M_{k,m}(\mathbb{R})$, and

$$(9) \quad \|A\|_\infty \leq \|A\| \leq k\sqrt{n} \|A\|_\infty$$

where $\|A\|_\infty = \max\{|A_{ij}|\}$ for $A \in M_{n,k}(\mathbb{R})$.

Proof. The norm properties are easily checked. By the Remark above, (7) is a special case of (8). Choose x with $\|x\|_2 \leq 1$ and $\|AB\| = \|ABx\|_2$. If $Bx = 0$, then $\|AB\| = 0$ and (8) holds. Otherwise we have $\|\frac{Bx}{\|Bx\|_2}\|_2 = 1$, and

$$\|A\| \geq \|A(\frac{Bx}{\|Bx\|_2})\|_2 = \frac{1}{\|Bx\|_2} \|A(Bx)\|_2.$$

Therefore

$$\|AB\| = \|ABx\|_2 \leq \|A\| \cdot \|Bx\|_2 \leq \|A\| \cdot \|B\|.$$

Denote the standard basis for $\mathbb{R}^k$ by e_i with $i = 1, \dots, k$ and denote $A_i = Ae_i$ for $i = 1, \dots, k$. Then

$$\begin{aligned} \|A\| &\geq \max\{\|A_i\|_2 \mid i = 1, \dots, k\} \\ &\geq \max\{\|A_i\|_\infty \mid i = 1, \dots, k\} = \|A\|_\infty. \end{aligned}$$

This gives one half of (9). For the other half let $x \in \mathbb{R}^k$ with $\|x\|_2 \leq 1$ satisfy $\|A\| = \|Ax\|_2$. Then $x = \sum x_i e_i$ with $x_i \in \mathbb{R}$. As $\|x\|_\infty \|x\|_2 \leq 1$, each $|x_i| \leq 1$. Then

$$\begin{aligned} \|A\| &= \|Ax\|_2 \leq \sum_{i=1}^k |x_i| \|A_i\|_2 \\ &\leq \sum_{i=1}^k \|A_i\|_2 \leq \sum_{i=1}^k \sqrt{n} \|A_i\|_\infty \\ &\leq k \cdot \sqrt{n} \|A\|_\infty \end{aligned}$$

as we claimed. □

The Contraction Mapping Theorem.

Contraction Mapping Theorem. Let (M, d) be a complete metric space, $0 \leq \lambda < 1$ and $f : M \rightarrow M$ a function with

$$(10) \quad d(f(x), f(y)) \leq \lambda \cdot d(x, y)$$

for all $x, y \in M$. Then f has a unique fixed point, that is, there is exactly one $x_0 \in M$ with $f(x_0) = x_0$.

An f satisfying (10) is called a *contraction*. Notice that a contraction is automatically continuous, given $\varepsilon > 0$ we simply choose $\delta = \varepsilon$.

It is important that M is complete:

Example. Let $M = (0, 1)$ and $f : M \rightarrow M$ defined by $f(x) = \frac{1}{2}x$. There is no $x \in (0, 1)$ with $f(x) = x$.

Proof of the Contraction Mapping Theorem. Let $x \in M$ be a point. We claim that $a_n = f^n(x) = f(f(\dots(f(x))\dots)) \in M$ is a Cauchy sequence. Note that

$$\begin{aligned} d(x, a_n) &\leq \sum_{i=0}^{n-1} d(f^i(x), f^{i+1}(x)) \\ &\leq \sum_{i=0}^{n-1} \lambda^i d(x, f(x)) \\ &= d(x, f(x)) \cdot \frac{1 - \lambda^n}{1 - \lambda} \\ &\leq \frac{d(x, f(x))}{1 - \lambda}. \end{aligned}$$

Let $n \geq m$, then

$$\begin{aligned} d(a_n, a_m) &= d(f^n(x), f^m(x)) \\ &\leq \lambda^m d(x, f^{m-n}(x)) \\ &\leq \frac{\lambda^m d(x, f(x))}{1 - \lambda} \end{aligned}$$

So for $n, m \geq n_0$ we get $d(a_n, a_m) \leq \frac{\lambda^{n_0}}{1 - \lambda}$ which can be made arbitrary small by choosing a large n_0 . As M is complete, the sequence $(a_n)_{n \in \mathbb{N}}$ converges to $a \in M$. Now

$$f(a) = f\left(\lim_{n \rightarrow \infty} a_n\right) = \lim_{n \rightarrow \infty} f(a_n) = \lim_{n \rightarrow \infty} a_{n+1} = a$$

where we use Lemma 1.8, so a is a fixed point.

It is the unique fixed point, for if $f(b) = b$, then

$$\begin{aligned} d(a, b) &= d(f(a), f(b)) \\ &\leq \lambda d(a, b) \end{aligned}$$

As $\lambda < 1$ this is only possible if $d(a, b) = 0$, so $a = b$. □

Example. Take a map of Durham and put it on the ground (in Durham). Then there is a unique point on the map which is lying directly over the point that it represents.

Think of Durham as a closed subset D of $\mathbb{R}^2$, e.g. the closed unit ball. Putting the map on the ground gives a function $f : D \rightarrow D$. Also $d(f(x), f(y)) \leq \lambda d(x, y)$, where λ is the scale of the map. Note that $\lambda < 1$ unless you have a 1-1 map of Durham, which seems unlikely.

In the next chapter we will see a more convincing application of the Contraction Mapping Theorem.

3. ORDINARY DIFFERENTIAL EQUATIONS

Example. Let $a \in \mathbb{R}$, then

$$\dot{x} = ax$$

is a differential equation, we look for a function $x(t)$ with $\dot{x}(t) = ax(t)$. The solutions are $x(t) = C \exp(at)$ with $C \in \mathbb{R}$. Note that by demanding a particular value $x(t_0)$ at a time t_0 , e.g. $x(0) = C$, there is exactly one solution.

To get a 2-dimensional example, say

$$\begin{aligned}\dot{x} &= -y \\ \dot{y} &= x\end{aligned}$$

The solutions are

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \alpha \cos t - \beta \sin t \\ \alpha \sin t + \beta \cos t \end{pmatrix}$$

With

$$\begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Again for a given starting point there is a unique solution.

Look at

$$\dot{x} = 3x^{\frac{2}{3}}.$$

Standard methods give

$$\begin{aligned}\frac{1}{3}\dot{x}x^{-\frac{2}{3}} &= 1 \\ x^{\frac{1}{3}} &= t + C\end{aligned}$$

so $x(t) = (t + C)^3$ is a solution. But notice that $x(-C) = 0$, so we should have been careful when we divided by x . In fact, $x(t) = 0$ also satisfies the differential equation and we could define for $C < D$

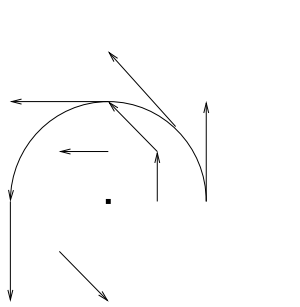
$$x_{D,C}(t) = \begin{cases} (t - D)^3 & t \leq D \\ 0 & t \in (D, C) \\ (t - C)^3 & t \geq C \end{cases}$$

So the solution is not determined by demanding some $x(t_0) = x_0 \in \mathbb{R}$. The problems with the uniqueness occur when $x = 0$, and notice that $f(x) = x^{\frac{2}{3}}$ is not differentiable at $x = 0$.

Definition 3.1. Let $U \subset \mathbb{R}^n$ be open. A *vector field* or a *autonomous differential equation* on U is a continuous map $v : U \rightarrow \mathbb{R}^n$. Here U is called the *phase space* of v . For $x \in U$ and $\tau \in \mathbb{R}$, a continuously differentiable curve $\alpha : (a, b) \rightarrow U$ is called an *integral curve of v at (x, τ)* , if $\tau \in (a, b)$ and $\alpha'(t) = v(\alpha(t))$ with $\alpha(\tau) = x$.

Integral curves are the solutions of the differential equation $\dot{\alpha} = v(\alpha)$.

Example. If $U = \mathbb{R}^2$, let $v : U \rightarrow \mathbb{R}^2$ be given by $v(x, y) = (-y, x)$. We interpret $v(x, y)$ as a ‘direction’, starting at (x, y) . Integral curves have their tangent vector always equal to that direction.



We can look at more general differential equations by letting the vector field change with time.

Definition 3.2. Let $U \subset \mathbb{R}^n$ be open and $I \subset \mathbb{R}$ an open interval. A *differential equation* is a continuous map $v : U \times I \rightarrow \mathbb{R}^n$. A *solution of v at $x \in U$ and $\tau \in I$* is a continuously differentiable curve $\alpha : I \rightarrow U$ with $\alpha'(t) = v(\alpha(t), t)$ and $\alpha(\tau) = x$.

In the following theorem $\|\cdot\|_2$ always refers to the Euclidean norm.

Picard-Lindelöf Theorem. Let $U \subset \mathbb{R}^n$ be open, $I \subset \mathbb{R}$ an open interval and $v : U \times I \rightarrow \mathbb{R}^n$ a differential equation which satisfies the following: There is $L \geq 0$ with

$$\|v(x_1, t) - v(x_2, t)\|_2 \leq L \cdot \|x_1 - x_2\|_2$$

for all $x_1, x_2 \in U$ and $t \in I$.

Given $(u, \tau) \in U \times I$, there exist $a, b > 0$ such that, with

$$U_1 = \{x \in U \mid \|x - u\|_2 < a\}$$

$$I_1 = \{t \in I \mid |t - \tau| < b\}$$

the differential equation $v|_{U_1 \times I_1}$ has for every $x \in U_1$ and $\tau \in I_1$ a unique solution.

Furthermore, the resulting $\alpha : U_1 \times I_1 \rightarrow U$ given by $\alpha(x, t) = \alpha_x(t)$, where α_x is the solution of v at x and τ (i.e. $\frac{\partial}{\partial t} \alpha(x, t) = v(\alpha(x, t), t)$ and $\alpha(x, \tau) = x$) is continuous.

Before we start the proof, let us interpret this result. If the differential equation satisfies a condition, we do get for given (u, τ) a unique solution, at least for a short period of time (represented by the interval I_1). We even get unique solutions for starting points near u (points in U_1). Finally, these solutions can be put together to form the continuous map α . This can be interpreted as saying that solutions which are “close” at time τ , will stay “close”.

The fact that we get solutions only for a short time should not be surprising, if U is just a small ball in $\mathbb{R}^n$ and the vector field simply points constantly and non-zero in one direction, the solution curves will move out of the ball, and if they start near the boundary, this can happen quite fast.

Now let us take a closer look at the condition on v , and let us assume that v does not depend on $t \in I$, that is, let us look at the case that $v : U \rightarrow \mathbb{R}^n$ is an autonomous differential equation. Then

$$\|v(x_1) - v(x_2)\|_2 \leq L \cdot \|x_1 - x_2\|_2$$

is called a *Lipschitz condition*. A function satisfying this is continuous (choose $\delta = \frac{\varepsilon}{L}$). It can be violated easily, if we allow U to be very big.

Example. Let $U = \mathbb{R}$ and $v(x) = x^2$. Then

$$|v(x) - v(y)| = |x - y| \cdot |x + y|$$

so no $L > 0$ can work for all $x, y \in \mathbb{R}$, as we can always find x, y with $x + y > L$.

On the other hand, given $x \in \mathbb{R}$, we can use $L = 2|x| + 1$ for $U = (x - 1, x + 1)$ as the Lipschitz constant.

If the Lipschitz condition is satisfied locally near $(u, \tau) \in U \times I$, we still get a solution. Note that since we only get the solution on $U_1 \times I_1$, we might as well reduce U and I after choosing (u, τ) .

Example. Let $v(x) = x^{\frac{2}{3}}$. Then v does not satisfy a Lipschitz condition on any closed interval containing 0. As we do not get a unique solution with $\alpha(0) = 0$, we see that the Lipschitz condition in the theorem cannot be dropped.

If v is continuously differentiable, it satisfies a Lipschitz condition locally, as we will see in the next chapter.

Proof of the Picard-Lindelöf Theorem. We look for α with

$$\frac{\partial}{\partial t} \alpha(x, t) = v(\alpha(x, t), t)$$

and $\alpha(x, t) = x$. By integrating, this means we want

$$(11) \quad \alpha(x, t) = x + \int_{\tau}^t v(\alpha(x, s), s) ds.$$

Given any β , we can define a new function $A\beta$ by

$$A\beta(x, t) = x + \int_{\tau}^t v(\beta(x, s), s) ds$$

and (11) then simply means $A\alpha = \alpha$. In view of the Contraction Mapping Theorem, we need a complete metric space M such that A defines a contraction from M to M .

Start by choosing $a_1 > 0$ with

$$D = D(u; 2a_1) \subset U$$

and $b_1 > 0$ with $D(\tau; b_1) \subset I$. As $D \times D(\tau; b_1)$ is compact by the Heine-Borel Theorem, we get a $K > 0$ such that

$$\|v(x, t)\|_2 \leq K$$

for all $(x, t) \in D \times D(\tau; b_1)$.

Let $0 < a < a_1$, and let $b > 0$ satisfy

$$(12) \quad b < \min\{b_1, \frac{a}{K}, \frac{1}{L}\}.$$

We set $U_1 = B(u; a) \subset U \subset \mathbb{R}^n$ and $I_1 = B(\tau; b) \subset I \subset \mathbb{R}$. Now let

$$\begin{aligned} M &= \{\beta : U_1 \times I_1 \rightarrow D \text{ continuous} \mid \beta(x, \tau) = x \text{ for all } x \in U_1\} \\ &\subset (C_b(U_1 \times I_1))^n \end{aligned}$$

where the latter is a Banach space with the supremum norm

$$\|\beta\| = \sup\{\|\beta(x, t)\|_2 \mid (x, t) \in U_1 \times I_1\},$$

see Theorem 3 for the case $n = 1$, the case $n > 1$ being basically the same.

Then M is closed: Let $(\beta_n)_{n \in \mathbb{N}}$ be a sequence in M with $\beta_n \rightarrow \beta \in (C_b(U_1 \times I_1))^n$.

For every $(x, t) \in U_1 \times I_1$, we have

$$\|\beta(x, t) - \beta_n(x, t)\|_2 \leq \|\beta - \beta_n\|,$$

so $\beta(x, t) = \lim_{n \rightarrow \infty} \beta_n(x, t) \in D$, as D is closed. Also $\beta(x, \tau) = \lim_{n \rightarrow \infty} \beta_n(x, \tau) = x$, as $\beta_n(x, \tau) = x$ for all n . This shows that $\beta \in M$ and M is closed by Proposition 1.18.

By Theorem 2 we also get M is complete.

Define for $\beta \in M$

$$A\beta(x, t) = x + \int_{\tau}^t v(\beta(x, s), s) ds.$$

We want $A\beta \in M$ and $A : M \rightarrow M$ a contraction. We certainly have $A\beta(x, \tau) = x$. Let us check that $A\beta(x, t) \in D$. Note

$$\begin{aligned} \|A\beta(x, t) - u\|_2 &\leq \|A\beta(x, t) - x\|_2 + \|x - u\|_2 \\ &\leq \int_{\tau}^t \|v(\beta(x, s), s)\|_2 ds + a \\ &\leq K \cdot |t - \tau| + a \leq 2a_1 \end{aligned}$$

by the choice of a and b , see (12).

Now let us check that $A\beta$ is continuous. We have

$$\begin{aligned} \|A\beta(x, t) - A\beta(y, t')\|_2 &\leq \|y - x\|_2 + \left\| \int_{t'}^t v(\beta(y, s), s) ds \right\|_2 \\ &\quad + \left\| \int_{\tau}^t v(\beta(x, s), s) - v(\beta(y, s), s) ds \right\|_2 \\ &\leq \|y - x\|_2 + K \cdot |t - t'| \\ &\quad + L \cdot |t - \tau| \cdot \sup\{\|\beta(x, s) - \beta(y, s)\|_2 \mid s \in [\tau, t]\} \end{aligned}$$

which can be made arbitrary small by choosing (x, t) and (y, t') close enough, note that the supremum will get small by the continuity of β and the compactness of $[\tau, t]$. To see this, assume there is no $\delta > 0$ with $\|x - y\|_2 < \delta$, implying $\|\beta(x, s) - \beta(y, s)\| < \varepsilon$ for all $s \in [\tau, t]$ and some $\varepsilon > 0$. Then there is a sequence (y_n, s_n) with $\|x - y_n\|_2 < \frac{1}{n}$ and $\|\beta(x, s_n) - \beta(y_n, s_n)\| \geq \varepsilon$. Clearly $y_n \rightarrow x$, and since $[\tau, t]$ is compact, we can assume that $s_n \rightarrow s \in [\tau, t]$. Then

$$\|\beta(x, s_n) - \beta(y_n, s_n)\|_2 \leq \|\beta(x, s_n) - \beta(x, s)\|_2 + \|\beta(x, s) - \beta(y_n, s_n)\|$$

where the left-hand side can be made arbitrarily small by continuity of β at (x, s) . This is a contradiction.

Therefore $A : M \rightarrow M$. Now for $\beta_1, \beta_2 \in M$ we get

$$\begin{aligned} \|A\beta_1(x, t) - A\beta_2(x, t)\|_2 &\leq \int_{\tau}^t \|v(\beta_1(x, s), s) - v(\beta_2(x, s), s)\|_2 ds \\ &\leq L \int_{\tau}^t \|\beta_1(x, s) - \beta_2(x, s)\|_2 ds \\ &\leq L \cdot |t - \tau| \cdot \|\beta_1 - \beta_2\| \\ &\leq (L \cdot b) \|\beta_1 - \beta_2\| \end{aligned}$$

As this works for all $(x, t) \in U_1 \times I_1$, we get

$$\|A\beta_1 - A\beta_2\| \leq (L \cdot b) \|\beta_1 - \beta_2\|$$

and since $Lb < 1$ by (12), we have that A is a contraction. By the Contraction Mapping Theorem we get a unique $\alpha : U_1 \times I_1 \rightarrow U$ with

$$\frac{\partial}{\partial t} \alpha(x, t) = v(\alpha(x, t), t)$$

and $\alpha(x, t) = x$. Note that there could be $\bar{\alpha} : I_1 \rightarrow U$ with $\bar{\alpha}'(t) = v(\bar{\alpha}(t), t)$ and $\bar{\alpha}(\tau) = x$ not fitting into a continuous $\bar{\alpha} : U_1 \times I_1 \rightarrow U$. To rule this out, we can repeat the contraction argument by defining A and M for a fixed $x \in U_1$, that is, using

$$M_x = \{\beta : I_1 \rightarrow D \mid \beta(\tau) = x\}$$

and $A_x : M_x \rightarrow M_x$. □

The proof does even suggest a way to find the solution: start with $\alpha_0(x, t) = x$ and for $n \geq 0$ form $\alpha_{n+1} = A\alpha_n$. Then $\alpha = \lim_{n \rightarrow \infty} \alpha_n$ is the solution. Let us do this in an easy example.

Example. Let B be a $n \times n$ matrix and define $v(x, t) = Bx$, so v does not depend on t . Clearly v satisfies the Lipschitz condition with $L = \|B\|$. Starting with $\alpha_0(x, t) = x$, we get

$$\begin{aligned} \alpha_1(x, t) &= x + \int_0^t Bx ds = x + tBx \\ \alpha_2(x, t) &= x + \int_0^t B(x + tBx) ds = x + tBx + \frac{t^2}{2} B^2 x \end{aligned}$$

and continuing we get

$$\alpha_n(x, t) = x + tBx + \frac{t^2}{2} B^2 x + \dots + \frac{t^n}{n!} B^n x$$

The limit is a series which resembles the power series of the exponential function:

$$\alpha(x, t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} B^n x = e^{tB} \cdot x$$

Notice that we think of e^{tB} as an $n \times n$ -matrix, which is then multiplied with $x \in \mathbb{R}^n$ to give an element of $\mathbb{R}^n$.

4. DIFFERENTIATION IN $\mathbb{R}^n$

Let $U \subset \mathbb{R}^n$ be open. Recall that a function $f : U \rightarrow \mathbb{R}^p$ is called *differentiable* at $x \in U$ with derivative $Df(x) \in M_{p,n}(\mathbb{R})$, if near x we can write

$$f(x+h) = f(x) + Df(x) \cdot h + R(h)$$

where R is defined near $0 \in \mathbb{R}^n$ with $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$.

The matrix $Df(x)$ is the matrix of partial derivatives $Df(x) = \left(\frac{\partial f_j}{\partial x_i}(x) \right)$, $i = 1, \dots, n$, $j = 1, \dots, p$. If $f : U \rightarrow \mathbb{R}^p$ is differentiable for every $x \in U$, we get a new function, the derivative $Df : U \rightarrow M_{p,n}(\mathbb{R}) \cong \mathbb{R}^{p \cdot n}$, and if Df is continuous, we call f *continuously differentiable*, or of class C^1 . As $Df : U \rightarrow \mathbb{R}^{pn}$, it can also be differentiable, with second derivative $D^2f : U \rightarrow \mathbb{R}^{pn^2}$, and so on. If f is k -times differentiable and $D^k f$ is continuous, we say that f is of class C^k . A function which is of class C^k for all $k \geq 1$ is called *smooth* or of class C^∞ .

Remark. We will often speak of a function being C^k with $k \geq 1$. This will also include the possibility that $k = \infty$.

The Mean Value Theorem for a differentiable function $f : [a, b] \rightarrow \mathbb{R}$ states that there is a $c \in (a, b)$ with $f(b) - f(a) = f'(c)(b - a)$. This can be generalized to functions $f : U \rightarrow \mathbb{R}$ with $U \subset \mathbb{R}^n$, but it does not quite work for functions to $\mathbb{R}^p$, as we would get a different c for every coordinate of f . But we get the following version, which is a generalization of the Fundamental Theorem of Calculus.

Mean Value Theorem. Let $U \subset \mathbb{R}^n$ be open, $x \in U$, $h \in \mathbb{R}^n$ such that $x + th \in U$ for all $t \in [0, 1]$, and $f : U \rightarrow \mathbb{R}^p$ be continuously differentiable. Then

$$(13) \quad f(x+h) - f(x) = \int_0^1 Df(x+th) \cdot h \, dt.$$

Note that $Df(x+th) \cdot h \in \mathbb{R}^p$ and the integral is a coordinatewise integral.

Proof. For $i = 1, \dots, p$ denote $f_i : U \rightarrow \mathbb{R}$ the i -th coordinate. Let $g_i(t) = f_i(x+th)$. Then $g'_i(t) = Df_i(x+th) \cdot h$ by the chain rule. Hence

$$\begin{aligned} \int_0^1 Df_i(x+th) \cdot h \, dt &= g_i(1) - g_i(0) \\ &= f_i(x+h) - f_i(x). \end{aligned}$$

Putting the coordinates together gives (13). □

Corollary 4.1. Let $U \subset \mathbb{R}^n$ be open and convex (that is, for all $x, y \in U$ and $t \in [0, 1]$ we have $tx + (1-t)y \in U$), and let $f : U \rightarrow \mathbb{R}^p$ be continuously differentiable. Assume that $C = \sup\{\|Df(x)\| \in \mathbb{R} \mid x \in U\}$ exists. Then

$$\|f(y) - f(x)\|_2 \leq C \cdot \|y - x\|_2$$

for all $x, y \in U$.

Proof. By the Mean Value Theorem, we have

$$\begin{aligned}
\|f(x+h) - f(x)\|_2 &= \left\| \int_0^1 Df(x+th)h \, dt \right\|_2 \\
&\leq \int_0^1 \|Df(x+th)h\|_2 \, dt \\
&\leq \int_0^1 \|Df(x+th)\| \cdot \|h\|_2 \, dt \\
&\leq C \cdot \|h\|_2,
\end{aligned}$$

where we use that $|\int_a^b g(t) \, dt| \leq \int_a^b |g(t)| \, dt$ and Lemma 2.8. Substituting $h = y - x$ gives the result. $\square$

Remark. Locally, the supremum of $\|Df(x)\|$ will always exist, for if $x \in U$, there is a $r > 0$ with $D(x; r) \subset U$ and on $D(x; r)$ the supremum exists, as $D(x; r)$ is compact.

In particular, if $v : U \times I \rightarrow \mathbb{R}^n$ is continuously differentiable, v will satisfy the condition of the Picard-Lindelöf Theorem near every (u, τ) .

Definition 4.2. Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^p$ be open. A C^1 function $f : U \rightarrow V$ is called a *diffeomorphism*, if there is an inverse function $f^{-1} : V \rightarrow U$, that is, $f^{-1} \circ f = \text{id}_U$ and $f \circ f^{-1} = \text{id}_V$, which is also differentiable.

Remark. By the chain rule

$$\begin{aligned}
D(f^{-1} \circ f)(x) &= Df^{-1}(f(x))Df(x) = I_n \\
D(f \circ f^{-1})(y) &= Df(f^{-1}(y))Df^{-1}(y) = I_p
\end{aligned}$$

where $I_m \in M_{m,m}(\mathbb{R})$ is the identity matrix. So with $y = f(x)$, $Df^{-1}(y) = (Df(x))^{-1}$, the inverse matrix. It follows from basic Linear Algebra that $n = p$.

In particular, if $f : U \rightarrow V$ is a diffeomorphism, then $Df(x)$ is an invertible matrix for every $x \in U$.

Example. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = x^3$ is not a diffeomorphism, even though the inverse $f^{-1}(x) = \sqrt[3]{x}$ exists. Notice that $f^{-1}(x)$ is not differentiable at $x = 0$, and $f'(0) = 0$ is not invertible.

Let $GL_n(\mathbb{R}) \subset M_{n,n}(\mathbb{R})$ be the set of invertible matrices.

Lemma 4.3. Let n be a positive integer and I the identity matrix.

- (1) $GL_n(\mathbb{R})$ is an open subset of $M_{n,n}(\mathbb{R})$.
- (2) If $A \in M_{n,n}(\mathbb{R})$ satisfies $\|A\| < 1$, then $I - A \in GL_n(\mathbb{R})$.
- (3) The map $\text{inv} : GL_n(\mathbb{R}) \rightarrow GL_n(\mathbb{R})$ given by $\text{inv}(A) = A^{-1}$ is a smooth diffeomorphism.

Proof. The map $\det : M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ is differentiable. To see this, we use induction on n (the case $n = 1$ gives $\det = \text{id}$) with the Laplace formula

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A'_{ij}$$

where $i \in \{1, \dots, n\}$ and A'_{ij} the matrix obtained from A by removing the i -th row and j -th column. In particular, the determinant is just a polynomial from $\mathbb{R}^{n^2}$ to $\mathbb{R}$ so it is even smooth. Now $GL_n(\mathbb{R}) = \det^{-1}(\mathbb{R} - \{0\})$ and if $\det A \neq 0$, then $\det B \neq 0$ for all B near A . This gives (1).

If $\|A\| < 1$, define $B_n = I + A + \dots + A^n$. Then $(B_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $M_{n,n}(\mathbb{R})$:

$$\|B_n - B_m\| \leq \sum_{k=\min\{n,m\}+1}^{\max\{n,m\}} \|A\|^k \leq \frac{\|A\|^{\min\{n,m\}+1}}{1 - \|A\|}$$

Hence $B = \lim_{n \rightarrow \infty} B_n \in M_{n,n}(\mathbb{R})$ exists, and

$$\begin{aligned} (I - A) \cdot B &= (I - A) \cdot \lim_{n \rightarrow \infty} B_n \\ &= \lim_{n \rightarrow \infty} (I - A) \cdot B_n \\ &= \lim_{n \rightarrow \infty} I - A^{n+1} = I \end{aligned}$$

where in the second line we used that matrix multiplication is continuous. Thus B is the inverse of $I - A$ which gives (2).

For (3), look at Cramer's rule: if $A = (a_{ij})$ is invertible, $A^{-1} = (b_{ij})$ is given by

$$b_{ij} = \frac{\det A_{ji}}{\det A}$$

where A_{ji} is the matrix obtained from A by replacing the i -th column with the j -th standard basis vector. As the determinant is smooth, the b_{ij} depend smoothly on the a_{ij} , and so inv is smooth. Note that $\text{inv} \circ \text{inv} = \text{id}$, so inv is its own inverse. Therefore it is a diffeomorphism. $\square$

The Inverse Function Theorem. We would like to say that a differentiable function $f : U \rightarrow V$ is a diffeomorphism, if $Df(x)$ is invertible for every $x \in U$. This is not true, as the next example shows.

Example. Let $f : (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^2 - \{0\}$ be defined by $f(r, \varphi) = r(\cos \varphi, \sin \varphi)$. Then

$$Df(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix}$$

Then $\det Df(r, \varphi) = r \neq 0$, but f is not a diffeomorphism, as f is not injective. If we restrict f to $(0, \infty) \times I$, where I is an open interval of length 2π , it is injective and a diffeomorphism onto its image. For example, $f : (0, \infty) \times (0, 2\pi) \rightarrow \mathbb{R}^2 - [0, \infty) \times \{0\}$. For $(x, y) \in \mathbb{R}^2 - [0, \infty) \times \{0\}$ we get $r = \sqrt{x^2 + y^2}$ and φ is determined by $\cos \varphi = \frac{x}{r}$ and $\sin \varphi = \frac{y}{r}$. As the inverse functions of the trigonometric functions are smooth, we get f^{-1} to be smooth as well.

Definition 4.4. Let $U \subset \mathbb{R}^n$ be open. A C^k function $f : U \rightarrow \mathbb{R}^n$ is called *locally invertible at $x \in U$* , if there is an open set $U_1 \subset U$ containing x and V_1 containing $f(x)$, such that $f|_{U_1} : U_1 \rightarrow V_1$ is a diffeomorphism.

Inverse Function Theorem. Let $U \subset \mathbb{R}^n$ be open, $f : U \rightarrow \mathbb{R}^n$ a C^k function for $k \geq 1$ and $u \in U$. Then f is locally invertible at u , if and only if $Df(u)$ is invertible. In this case the local inverse is also C^k .

Proof. If f is locally invertible at u , it restricts to a diffeomorphism near u and so $Df(u)$ is invertible by the remark above.

So assume that $Df(u)$ is invertible, we need to construct a local inverse and show that it is differentiable. We can assume that $u = 0 = f(u)$ and $Df(0) = I$, the identity matrix. After all, we can look at $g(x) = A^{-1} \circ (f(x + u) - f(u))$, where $A = Df(u)$. Then $g(0) = 0$ and $Dg(0) = I$ by the chain rule. If we can find an inverse of g , we also get an inverse of $f(x) = A(g(x - u)) + f(u)$.

Define for $y \in \mathbb{R}^n$

$$g_y(x) = y + (x - f(x)).$$

Then the equation $x = g_y(x)$ is equivalent to $y = f(x)$, and, after all, we try to solve $y = f(x)$ for x . As we have turned this into a fixed point problem, the Contraction Mapping Theorem will be helpful. Note that $Dg_y(x) = Dg_0(x) = I - Df(x)$ is independent of y and $Dg_y(0) = 0$. By continuity there is a $r > 0$ with $\|Dg_y(x)\| \leq \frac{1}{2}$ for all $x \in D(0; r)$.

By Corollary 4.1,

$$(14) \quad \|g_y(x_1) - g_y(x_2)\|_2 \leq \frac{1}{2} \|x_1 - x_2\|_2.$$

With $x_2 = 0$, we furthermore get

$$\|g_y(x) - y\|_2 \leq \frac{1}{2} \|x\|_2$$

So if $\|y\|_2 \leq \frac{r}{2}$ and $x \in D(0; r)$, we get

$$\|g_y(x)\|_2 \leq \|y\|_2 + \frac{1}{2} \|x\|_2 \leq r,$$

which means, with (14), that $g_y : D(0; r) \rightarrow D(0; r)$ is a contraction of a complete metric space. By the Contraction Mapping Theorem there is a unique $x \in D(0; r)$ with $y = f(x)$ for every $y \in D(0; \frac{r}{2})$.

Define $U_1 = \{x \in U \mid \|x\|_2 < r \text{ and } \|f(x)\|_2 < \frac{r}{2}\}$, which is an open set by continuity of f . Let $V_1 = f(U_1)$, then $V_1 \subset D(0; \frac{r}{2})$, so $f : U_1 \rightarrow V_1$ is bijective, hence the inverse $\varphi = f^{-1} : V_1 \rightarrow U_1$ exists.

We claim that $V_1 = B(0; \frac{r}{2})$. If $y \in B(0; \frac{r}{2})$, there is $x \in D(0; r)$ with $f(x) = y$, so we only need $x \in B(0; r)$. Given $x_1, x_2 \in D(0; r)$, we have

$$\begin{aligned} \|x_1 - x_2\|_2 &= \|g_0(x_1) + f(x_1) - (g_0(x_2) + f(x_2))\|_2 \\ &\leq \|g_0(x_1) - g_0(x_2)\|_2 + \|f(x_1) - f(x_2)\|_2 \\ &\leq \frac{1}{2} \|x_1 - x_2\|_2 + \|f(x_1) - f(x_2)\|_2. \end{aligned}$$

Hence

$$(15) \quad \|x_1 - x_2\|_2 \leq 2 \|f(x_1) - f(x_2)\|_2.$$

Use $x_2 = 0$, then $\|f(x_1)\|_2 < \frac{r}{2}$ implies $\|x_1\|_2 < r$, and so $V_1 = B(0; \frac{r}{2})$ is open. Also, (15) implies that

$$(16) \quad \|\varphi(y_1) - \varphi(y_2)\|_2 \leq 2 \|y_1 - y_2\|_2,$$

so φ is continuous. Note that $Df(x)$ is invertible for all $x \in D(0; r)$. This follows from Lemma 4.3, because $Df(x) = I - Dg_0(x)$ and $\|Dg_0(x)\| < \frac{1}{2}$ for $x \in D(0; r)$.

That f is differentiable means that

$$f(x_1) - f(x_2) = Df(x_2)(x_1 - x_2) + R(x_1 - x_2)$$

and $R(h)/\|h\|_2 \rightarrow 0$ for $h \rightarrow 0$. So

$$y_1 - y_2 = Df(x_2)(\varphi(y_1) - \varphi(y_2)) + R(\varphi(y_1) - \varphi(y_2))$$

and therefore

$$(17) \quad \varphi(y_1) - \varphi(y_2) = (Df(\varphi(y_2)))^{-1}(y_1 - y_2) + \bar{R}(y_1 - y_2)$$

with $\bar{R}(y_1 - y_2) = (Df(\varphi(y_2)))^{-1}(R(\varphi(y_1) - \varphi(y_2)))$. Then by (16) and Lemma 2.8

$$\frac{\|(Df(\varphi(y_2)))^{-1}(R(\varphi(y_1) - \varphi(y_2)))\|_2}{\|y_1 - y_2\|_2} \leq \|(Df(x_2))^{-1}\| \cdot 2 \frac{\|R(\varphi(y_1) - \varphi(y_2))\|_2}{\|\varphi(y_1) - \varphi(y_2)\|_2}$$

and as $\|(Df(x_2))^{-1}\|$ is bounded for $x_2 \in D(0; r)$, and since $\|y_1 - y_2\|_2 \rightarrow 0$ implies $\|\varphi(y_1) - \varphi(y_2)\|_2 \rightarrow 0$, we get that the right hand side tends to zero by the property of R . Therefore (17) gives that φ is differentiable, and the derivative is given by

$$D\varphi(y) = (Df(\varphi(y)))^{-1} = \text{inv} \circ Df \circ \varphi(y).$$

This shows, that if f is C^k , then so is φ , by the chain rule and Lemma 4.3. $\square$

Given an equation of the form

$$f(x, y) = 0$$

one often would like to solve this for y , that is, write $y = g(x)$ such that $f(x, g(x)) = 0$. The next theorem gives a criterion when this is possible.

Implicit Function Theorem. Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open, and let $f : U \times V \rightarrow \mathbb{R}^m$ be a C^k -function with $k \geq 1$. Let $(u, v) \in U \times V$ and assume that the $m \times m$ -matrix $\left(\frac{\partial f_i}{\partial y_j}(u, v)\right)$ is invertible. Let $c = f(u, v)$. Then there is a C^k -function $\eta : U_1 \rightarrow V_1 \subset V$ with $U_1 \subset U$ an open set containing u and $V_1 \subset V$ an open set containing v , such that $\eta(u) = v$ and $f(x, \eta(x)) = c$ for all $x \in U_1$. Furthermore, if $f(x, y) = c$ with $(x, y) \in U_1 \times V_1$, then $y = \eta(x)$.

Remark. The theorem states that the equation $f(x, y) = c$ can be solved uniquely for y near a point (u, v) , if the linear equation

$$Df(u, v) \cdot (x, y) = 0$$

can be solved uniquely for y . Note that

$$Df(u, v) \cdot (x, y) = \left(\frac{\partial f_i}{\partial x_s}(u, v)\right) \cdot x + \left(\frac{\partial f_i}{\partial y_j}(u, v)\right) \cdot y = 0$$

and this can be solved uniquely for y exactly if $\left(\frac{\partial f_i}{\partial y_j}(u, v)\right)$ is an invertible matrix.

Proof. We want to use the Inverse Function Theorem, so define $\varphi : U \times V \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ by $\varphi(x, y) = (x, f(x, y))$. Then

$$D\varphi = \left(\begin{array}{c|c} I & 0 \\ \hline \partial f_i / \partial x_s & \partial f_i / \partial y_j \end{array} \right)$$

and so $\det D\varphi(u, v) = \det \left(\frac{\partial f_i}{\partial y_j}(u, v)\right) \neq 0$. By the Inverse Function Theorem, there is an open set W containing (u, v) such that φ restricts to a diffeomorphism

$\varphi : W \rightarrow W'$. Since $\varphi(x, y) = (x, f(x, y))$, we get that $\varphi^{-1}(a, b) = (a, g(a, b))$. Set $\eta(x) = g(x, c)$. Then

$$\begin{aligned} f(x, \eta(x)) &= f(x, g(x, c)) = f(\varphi^{-1}(x, c)) \\ &= p_2(\varphi(\varphi^{-1}(x, c))) = c. \end{aligned}$$

So $f(x, y) = c$ if and only if $y = \eta(x)$ for $(x, y) \in W$. Now choose $U_1 \times V_1 \subset W$ with $(u, v) \in U_1 \times V_1$ and $\eta(U_1) \subset V_1$. $\square$

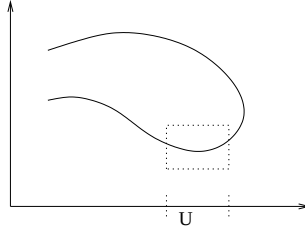
Example. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be smooth. To apply the Implicit Function Theorem, we only need one coordinate with $\frac{\partial f}{\partial x_i}(x) \neq 0$. Let $f(x, y) = x^2 + y^2$. Then $f^{-1}(\{1\})$ is a circle. Also, $f(1, 0) = 1$. Note that $\frac{\partial f}{\partial y}(1, 0) = 0$, but $\frac{\partial f}{\partial x}(1, 0) = 2 \neq 0$. So we get $\eta : I \rightarrow \mathbb{R}$ where I is a small open interval containing 0 with $f(\eta(y), y) = 1$.

If $f(x, y, z) = x^2 + y^2 + z^2$, then $f(0, 0, 1) = 1$ and $\frac{\partial f}{\partial z}(0, 0, 1) \neq 0$. By the Implicit Function Theorem we get a small disc U around $(0, 0)$ and a parametrization of the unit sphere near $(0, 0, 1)$ by $(x, y, \eta(x, y))$, with $\eta : U \rightarrow \mathbb{R}$.

More generally, if the conditions of the Implicit Function Theorem apply, then the set

$$\{(x, y) \in \mathbb{R}^{n+m} \mid f(x, y) = c\} = M \subset \mathbb{R}^{n+m}$$

can locally around $(u, v) \in M$ be parametrized by an open set $U \subset \mathbb{R}^n$; any point in $M \cap (U \times V)$ is of the form $(x, \eta(x))$, which we can think of as the graph of η .



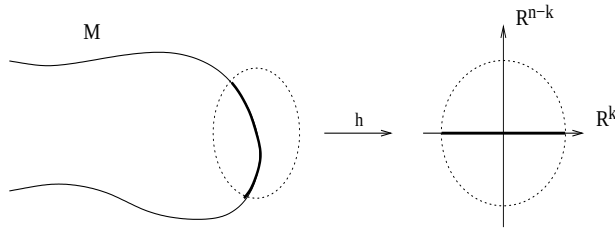
Also, as can be seen in the example of the circle at $(1, 0)$, if we cannot parametrize in the first coordinate, we might still be able to parametrize in the second coordinate.

This leads to the following definition of a manifold.

Definition 4.5. Let $k \geq 0$ and $l \geq 1$. A subset $M \subset \mathbb{R}^n$ is called a *k-dimensional C^l -manifold*, if the following holds: every point $p \in M$ is contained in an open set $U \subset \mathbb{R}^n$ on which there is a C^l -diffeomorphism $h : U \rightarrow U' \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ with $h(U \cap M) = U' \cap (\mathbb{R}^k \times \{0\})$.

Such an h is called a *chart* of M , and a family $\{h_\lambda : U_\lambda \rightarrow U'_\lambda \mid \lambda \in \Lambda\}$ of charts is called an *atlas*, if $M \subset \bigcup_{\lambda \in \Lambda} U_\lambda$.

So in local coordinates, a manifold looks like some $\mathbb{R}^k$ in $\mathbb{R}^n$.



Remark. In the situation of the Implicit Function Theorem, if $f : U \times V \rightarrow \mathbb{R}^{n-k}$, with $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^{n-k}$ open, and $(u, v) \in U \times V$ satisfies $\left(\frac{\partial f_i}{\partial y_j}(u, v)\right)$ is invertible, we get the chart from the map $\varphi : W \rightarrow W'$ given by $\varphi(x, y) = (x, f(x, y) - c)$, where $c = f(u, v)$ and W is an open set containing (u, v) so that φ is a diffeomorphism, compare the proof of the Implicit Function Theorem. In particular, $W \cap \{(x, y) \in \mathbb{R}^n \mid f(x, y) = c\}$ is a k -dimensional manifold, an atlas just consisting of the one chart φ .

Example. Let $U \subset \mathbb{R}^n$ be open. Then U is a n -dimensional C^∞ -manifold, the identity map $\text{id} : U \rightarrow U$ being a chart.

The set $S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_2 = 1\}$, called the n -sphere, is an n -dimensional manifold. Look at $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ given by $f(x) = \|x\|_2^2$. Note that we have to use different coordinates i to get $\frac{\partial f}{\partial x_i}(x) \neq 0$ for all $x \in S^n$, so we need more than one chart for an atlas.

Definition 4.6. Let $f : U \rightarrow \mathbb{R}^k$ be a C^1 -function, where $U \subset \mathbb{R}^n$ is open. A point $x \in U$ is called a *critical point* of f , if $Df(x)$ has rank less than k . Otherwise it is called a *regular point*. If x is a critical point, $f(x) \in \mathbb{R}^k$ is called a *critical value* of f . A point $u \in \mathbb{R}^k$ is called a *regular value* of f , if it is not a critical value.

Example. Define $f : \mathbb{R}^n \rightarrow \mathbb{R}$ by $f(x) = \|x\|_2^2$, then $0 \in \mathbb{R}$ is the only critical point, so $0 \in \mathbb{R}$ is a critical value, and $\mathbb{R} - \{0\}$ are all regular values. Note that negative numbers are not in the image, and hence automatically are regular values.

If $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is C^1 , and $k > n$, the rank of $Df(x)$ can never be k , so all points of $\mathbb{R}^n$ are critical. So the regular values are exactly those points not in the image.

Define $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by $f(x, y, z) = (e^z x, (y - 1) \sin z)$. Then

$$Df(x, y, z) = \begin{pmatrix} e^z & 0 & e^z x \\ 0 & \sin z & (y - 1) \cos z \end{pmatrix}$$

From the first two columns we see that $\sin z = 0$ is necessary for a critical point. Then $\cos z = \pm 1$, and from the first and the third row we see that $y = 1$ is necessary for a critical point. This is also sufficient for being a critical point, so

$$\{(x, 1, k\pi) \in \mathbb{R}^3 \mid k \in \mathbb{Z}\}$$

is the set of critical points.

Theorem 5. Let $U \subset \mathbb{R}^n$ be open and $f : U \rightarrow \mathbb{R}^k$ a C^l -map with $l \geq 1$. If $y \in \mathbb{R}^k$ is a regular value, then $f^{-1}(\{y\})$ is a $(n - k)$ -dimensional C^l -manifold.

Proof. Let $x \in f^{-1}(\{y\})$, then x is not a critical point of f . After rearranging the coordinates, we can assume that $\left(\frac{\partial f_i}{\partial x_j}(x)\right)$, with $i = 1, \dots, k$ and $j = n - k + 1, \dots, n$, is invertible and so we get a chart around x as in the remark above. $\square$

Remark. Note that points not in the image of f are regular values, but the empty-set is a manifold of any dimension. After all, as there are no points in the emptyset, the condition for being a manifold is vacuously satisfied.

Example. We have already seen that S^n is a manifold, but as $S^n = f^{-1}(\{1\})$ with $f(x) = \|x\|_2^2$, we get this also from Theorem 5.

Let $M = \{(x, y, z) \in \mathbb{R}^3 \mid x^3 + y^3 + z^3 = 2\}$, this is a manifold by Theorem 5, using $g(x, y, z) = x^3 + y^3 + z^3$, which has only a critical point at $(0, 0, 0) \notin M$.

Define $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^2$ by $F(x, y) = (\|x\|_2^2, \|y\|_2^2)$. Again (a, b) with $a \neq 0$ and $b \neq 0$ are regular values, and $F^{-1}(\{(1, 1)\}) = S^{n-1} \times S^{m-1}$.

Let $f(x, y, z) = (e^z x, (y - 1) \sin z)$. We have seen that the critical points are of the form $(x, 1, k\pi)$ with $k \in \mathbb{Z}$. Hence critical values are of the form $(e^{k\pi} x, 0)$, and so all (a, b) with $b \neq 0$ are regular. Then $f^{-1}(\{(a, b)\})$ is a 1-dimensional manifold.

5. TANGENT SPACES AND VECTOR FIELDS

Definition 5.1. Let $M \subset \mathbb{R}^n$ be a k -dimensional C^l -manifold with $l \geq 1$. A *tangent vector* v at $p \in M$ is an element $v \in \mathbb{R}^n$ of the form $v = \gamma'(0)$, where $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ is a C^1 -curve with $\varepsilon > 0$ and $\gamma(0) = p$. The set of tangent vectors, denoted $T_p M$, is called the *tangent space* of M at p .

Proposition 5.2. Let $M \subset \mathbb{R}^n$ be a C^l -manifold of dimension k and $p \in M$. Then $T_p M$ is a k -dimensional sub-vector space of $\mathbb{R}^n$. In fact, if $h : U \rightarrow U' \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ is a chart around p with $h(p) = 0$, then

$$T_p M = Dh^{-1}(0)(\mathbb{R}^k \times \{0\})$$

Proof. Let $h : U \rightarrow U'$ be a chart with $p \in U$ and $h(p) = 0$. If $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ is a C^1 -curve with $\gamma(0) = p$, we can assume that $\gamma(\varepsilon, \varepsilon) \subset U$ after possibly decreasing $\varepsilon > 0$. So $h \circ \gamma$ is a C^1 -curve in $\mathbb{R}^k \times \{0\}$ with $h \circ \gamma(0) = 0$. Note that $w = (h \circ \gamma)'(0) \in \mathbb{R}^k \times \{0\}$. Now $\gamma = h^{-1} \circ (h \circ \gamma)$, and so

$$\begin{aligned} \gamma'(0) &= Dh^{-1}(h \circ \gamma(0)) \cdot (h \circ \gamma)'(0) \\ &= Dh^{-1}(0) \cdot w. \end{aligned}$$

So every element of $T_p M$ is in the image of the linear map $Dh^{-1}(0)$ restricted to $\mathbb{R}^k \times \{0\}$. Also, for $w \in \mathbb{R}^k$, the curve $\gamma(t) = h^{-1}(tw)$ satisfies $\gamma'(0) = Dh^{-1}(0)w$, so $T_p M = Dh^{-1}(0)(\mathbb{R}^k \times \{0\})$. As $Dh^{-1}(0)$ is injective, $\dim T_p M = \dim \mathbb{R}^k = k$. $\square$

Our way to obtain manifolds is to look at C^l -functions $g : U \times V \rightarrow \mathbb{R}^m$ with $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ open and $\det \left(\frac{\partial g_i}{\partial y_j}(u, v) \right) \neq 0$ and take points with near (u, v) with $g(x, y) = g(u, v) = c$.

The tangent space of a manifold resulting this way can be described by g :

Theorem 6. Let $U \subset \mathbb{R}^n$ be open and $g : U \rightarrow \mathbb{R}^{n-k}$ a C^l -function such that $c \in \mathbb{R}^{n-k}$ is a regular value of g . Let $M = g^{-1}(\{c\})$, a k -dimensional manifold. Then

$$T_p M = \ker Dg(p)$$

for every $p \in M$.

Proof. Let $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ be a C^1 -curve with $\gamma(0) = p$. Then $g \circ \gamma$ is a constant map. So, for $v = \gamma'(0)$ we get

$$Dg(p)v = Dg(p)\gamma'(0) = D(g \circ \gamma)(0) = 0$$

which shows that $T_p M \subset \ker Dg(p)$. Now $\dim \ker Dg(p) + \text{rank}(Dg(p)) = n$. But $\text{rank}(Dg(p)) = n - k$, so $\dim \ker Dg(p) = k$ and therefore $T_p M = \ker Dg(p)$. $\square$

Example. Let $M = \{(x, y, z) \in \mathbb{R}^3 \mid x^3 + y^3 + z^3 = 1\}$. Let $p = (1, -1, 1) \in M$. Then $T_p M = \ker Dg(p)$ with $g(x, y, z) = x^3 + y^3 + z^3 - 1$. Then $Dg(p) = \begin{pmatrix} 3 & 3 & 3 \end{pmatrix}$, so

$$T_p M = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}.$$

With $q = (1, 0, 0)$ we get $Dg(q) = \begin{pmatrix} 3 & 0 & 0 \end{pmatrix}$, hence

$$T_q M = \{(x, y, z) \in \mathbb{R}^3 \mid x = 0\}.$$

We want to have differentiable functions $f : M \rightarrow \mathbb{R}^k$ on manifolds M . As $M \subset \mathbb{R}^n$, we can demand that f is defined on an open set U containing M and that this function is differentiable. For the derivative we are only interested in the parts that come from M . We are led to the following definition.

Definition 5.3. Let $M \subset \mathbb{R}^n$ be a C^1 -manifold, let $U \subset \mathbb{R}^n$ be an open set containing M and $f : U \rightarrow \mathbb{R}$ a C^1 -function. Then $p \in M$ is called a *critical point* of $f|_M$, if for every C^1 -curve $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ with $\gamma(0) = p$ we get $(f \circ \gamma)'(0) = 0$.

For example, if f has a local extremum on M , this gives a critical point, as $f \circ \gamma$ will have an extremum at such a point.

By the chain rule, $Df(p) \cdot \gamma'(0) = 0$, and $\gamma'(0) \in T_p M$, so the condition of f having a critical point at $p \in M$ is the same as saying $Df(p)|_{T_p M} = 0$.

Method of Lagrange Multipliers. Let $U \subset \mathbb{R}^{n+m}$ be open, $g : U \rightarrow \mathbb{R}^n$ a C^1 -function and $0 \in \mathbb{R}^n$ a regular value of g . Let $f : U \rightarrow \mathbb{R}$ be a C^1 -function. Then $p \in M = g^{-1}(\{0\})$ is a critical point of $f|_M$, if and only if there exist $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ with $D(f + \lambda_1 g_1 + \dots + \lambda_n g_n)(p) = 0$.

Proof. Note that

$$D(f + \lambda_1 g_1 + \dots + \lambda_n g_n)(p) = 0$$

is the same as

$$Df(p) = \sum_{i=1}^n -\lambda_i Dg_i(p)$$

that is, the row vector $Df(p)$ is a linear combination of the row vectors $Dg_i(p)$.

Now $f|_M$ is critical at $p \in M$ if and only if $Df(p)|_{T_p M} = 0$, and by Theorem 6 $T_p M = \ker Dg(p)$. We can think of $Df(p)$ and the $Dg_i(p)$ as linear maps from $\mathbb{R}^{n+m}$ to $\mathbb{R}$, that is, as elements of the dual space $(\mathbb{R}^{n+m})^*$. Now the $Dg_i(p)$ vanish on $T_p M$, so if $Df(p)$ is a linear combination of them, it also vanishes on $T_p M$, so p is a critical point.

On the other hand, the $Dg_i(p)$ are all linearly independent, as $Dg(p)$ has rank n , and the space of linear maps vanishing on $T_p M$ has dimension $n + m - \dim M = n$, so the $Dg_i(p)$ form a basis of that space. If p is a critical point of $f|_M$, $Df(p)$ vanishes on $T_p M$ and is therefore a linear combination of the $Dg_i(p)$. This finishes the proof. $\square$

If M is compact, then f will take the maximum and minimum on M and they are critical points. The Method of Lagrange gives us a way to find the critical points: If we define $F : U \times \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$F(x, \lambda_1, \dots, \lambda_n) = f(x) + \lambda_1 g_1(x) + \dots + \lambda_n g_n(x)$$

we get $2n + m$ equations

$$\frac{\partial F}{\partial x_i} = 0$$

for $i = 1, \dots, n + m$ and

$$\frac{\partial F}{\partial \lambda_j} = 0$$

for $j = 1, \dots, n$, and we can try to solve them for x_i and λ_j . We will often get finitely many points in M , and by evaluating f at them we can find the maximum and minimum. It should always be checked though that M is compact, which by the Heine-Borel Theorem is equivalent to closed and bounded. To get a manifold to be closed, the following purely topological criterion is useful.

Lemma 5.4. *Let $g : \mathbb{R}^n \rightarrow \mathbb{R}^k$ be continuous and $x \in \mathbb{R}^k$. Then $g^{-1}(\{x\})$ is closed.*

Proof. Let $(y_n)_{n \in \mathbb{N}}$ be a sequence in $g^{-1}(\{x\})$ with $y_n \rightarrow y \in \mathbb{R}^n$. By continuity, $g(y) = \lim_{n \rightarrow \infty} g(y_n) = \lim_{n \rightarrow \infty} x = x$, so $y \in g^{-1}(\{x\})$. So by Proposition 1.18, $g^{-1}(\{x\})$ is closed. $\square$

One should keep in mind though, that often g is only defined on an open subset U of $\mathbb{R}^n$, and then $g^{-1}(\{x\})$ is only closed in U , but not necessarily in $\mathbb{R}^n$. If g can be extended continuously to a closed set containing U , or if restricting g to a closed set does not change $g^{-1}(\{x\})$, we still get $g^{-1}(\{x\})$ to be closed in $\mathbb{R}^n$.

Example. Find $\max\{x + y \mid x^4 + y^4 = 1\}$.

We set $f(x, y) = x + y$ and $g(x, y) = x^4 + y^4 - 1$. Then 0 is a regular value of g , and we set $F(x, y, \lambda) = x + y + \lambda(x^4 + y^4 - 1)$. This gives rise to three equations

$$\begin{aligned} 1 + 4\lambda x^3 &= 0 \\ 1 + 4\lambda y^3 &= 0 \\ x^4 + y^4 - 1 &= 0. \end{aligned}$$

Thus $x = y = \left(-\frac{1}{4\lambda}\right)^{\frac{1}{3}}$, and

$$\left(-\frac{1}{4\lambda}\right)^{\frac{4}{3}} + \left(-\frac{1}{4\lambda}\right)^{\frac{4}{3}} - 1 = 0$$

This can be solved for λ , and we get $\lambda = \pm \frac{8^{\frac{1}{4}}}{4}$. So we get two critical points $\pm \left(\frac{1}{\sqrt[4]{2}}, \frac{1}{\sqrt[4]{2}}\right)$, and the maximum is $\frac{2}{\sqrt[4]{2}}$. Note that M is compact, so this has to be the maximum.

Example. Find the extrema of $f(x, y, z) = 5x + y - 3z$ on the intersection of the plane $x + y + z = 0$ with S^2 .

Define $g(x, y, z) = (x + y + z, x^2 + y^2 + z^2 - 1)$, then

$$Dg(x, y, z) = \begin{pmatrix} 1 & 1 & 1 \\ 2x & 2y & 2z \end{pmatrix}$$

Critical points satisfy $x = y = z$, but $g(x, x, x) \neq 0$ for all x , so 0 is a regular value of g . Define $F(x, y, z, \lambda, \mu) = 5x + y - 3z + \lambda(x + y + z) + \mu(x^2 + y^2 + z^2 - 1)$, and

we get five equations

$$\begin{aligned} 5 + \lambda + 2\mu x &= 0 \\ 1 + \lambda + 2\mu y &= 0 \\ -3 + \lambda + 2\mu z &= 0 \\ x + y + z &= 0 \\ x^2 + y^2 + z^2 - 1 &= 0. \end{aligned}$$

Adding the first three equations, together with the fourth equation gives $3 + 3\lambda = 0$, so $\lambda = -1$. Thus we get two equations

$$\begin{aligned} 4 + 2\mu x &= 0 \\ 2\mu y &= 0 \end{aligned}$$

which gives $\mu \neq 0$ and $y = 0$. The fourth equation thus gives $x = -z$, and the fifth $2x^2 = 1$, that is, $x = \pm \frac{1}{\sqrt{2}}$. We get the points

$$\begin{aligned} a &= \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right) \\ b &= \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right), \end{aligned}$$

and with $\mu = \pm 2\sqrt{2}$, these points are indeed critical for $f|_M$. Also, M is compact as a closed subset of S^2 , so $f(a) = 4\sqrt{2}$ is the maximum, and $f(b) = -4\sqrt{2}$ is the minimum.

Even if the manifold is not compact, we can sometimes find extremal points, for example if we are looking for a point $p \in M$ which has minimum distance to a given point $x \in \mathbb{R}^n$.

Proposition 5.5. *Let $A \subset \mathbb{R}^n$ be compact and $B \subset \mathbb{R}^n$ closed, and assume that both sets are non-empty. Then there is $a \in A$ and $b \in B$ with*

$$\|a - b\| \leq \|x - y\|$$

for all $x \in A$ and $y \in B$. Here $\|\cdot\|$ can be any norm on $\mathbb{R}^n$.

Proof. Let $d = \inf\{\|x - y\| \mid x \in A, y \in B\}$, we have to find $a \in A$ and $b \in B$ with $\|a - b\| = d$. Because of the infimum property we can find $a_n \in A$ and $b_n \in B$ with $\|a_n - b_n\| < d + \frac{1}{n}$. As A is compact, we can, by passing to a subsequence, assume that $a_n \rightarrow a \in A$. Then, for large n , we get

$$\begin{aligned} \|b_n\| &\leq \|b_n - a_n\| + \|a_n - a\| + \|a\| \\ &\leq d + 1 + \|a\|. \end{aligned}$$

As $B \cap D(0; d + 1 + \|a\|)$ is compact, and $b_n \in B \cap D(0; d + 1 + \|a\|)$ for large n , we can assume that $b_n \rightarrow b \in B \cap D(0; d + 1 + \|a\|)$. Now

$$\begin{aligned} \|b - a\| &\leq \|b - b_n\| + \|b_n - a_n\| + \|a_n - a\| \\ &\leq d + \frac{1}{n} + \|b - b_n\| + \|a_n - a\|. \end{aligned}$$

As $\|b - b_n\|$ and $\|a_n - a\|$ can be arbitrarily small for large n , we get $\|b - a\| < d + \varepsilon$ for every $\varepsilon > 0$, and therefore $\|b - a\| = d$. $\square$

Example. Find a point in $M = \{(x, y, z) \in \mathbb{R}^3 \mid 2x^2 + y^2 + z = 1\}$ which has minimal distance to $p = (0, 0, -5)$.

We see that M is a closed manifold, but it is not bounded, as z can take arbitrary negative values. So M is not compact. By Proposition 5.5, with $A = \{p\}$ and $B = M$, there is $b \in M$ minimizing the distance. Such b is a critical point of the squared distance function $f(x, y, z) = x^2 + y^2 + (z + 5)^2 = \|(x, y, z) - p\|_2^2$. Define $F(x, y, z, \lambda) = x^2 + y^2 + (z + 5)^2 + \lambda(2x^2 + y^2 + z - 1)$, then we get the equations

$$\begin{aligned} 2x + 4\lambda x &= 0 \\ 2y + 2\lambda y &= 0 \\ 2z + 10 + \lambda &= 0 \\ 2x^2 + y^2 + z - 1 &= 0 \end{aligned}$$

From the first equation we either get $\lambda = -1/2$, or $x = 0$. If $\lambda = -1/2$, then $y = 0$ and $z = -19/4$, so $x = \pm\sqrt{23/8}$. Then $f(\pm\sqrt{23/8}, 0, -19/4) = 23/8 + 1/16 = 47/16$.

If $x = 0$, the second equation gives $y = 0$ or $\lambda = -1$. If $y = 0$, we get $z = 1$ and $f(0, 0, 1) = 36$. If $\lambda = -1$, then $z = -9/2$ and $y = \pm\sqrt{11/2}$. Then $f(0, \pm\sqrt{11/2}, -9/2) = 23/4$.

The smallest values are taken at $(\pm\sqrt{23/8}, 0, -19/4)$.

We will now generalize the notion of critical points/values to maps between manifolds. Let $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ be C^l -manifolds with $l \geq 1$. Assume we have a continuous map $f : M \rightarrow N$ which extends to a C^1 -map $f : U \rightarrow \mathbb{R}^n$, where $U \subset \mathbb{R}^m$ is an open set containing M . Then we define

$$T_p f : T_p M \rightarrow T_{f(p)} N$$

in the following way: If $v \in T_p M$, then $v = \gamma'(0)$ for some $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ with $\gamma(0) = p$. Then $f \circ \gamma : (-\varepsilon, \varepsilon) \rightarrow N$, so $(f \circ \gamma)'(0) \in T_{f(p)} N$. Note that by the chain rule

$$\begin{aligned} (f \circ \gamma)'(0) &= Df(\gamma(0))\gamma'(0) \\ &= Df(p)v, \end{aligned}$$

so we can simply define $T_p f(v) = Df(p)v$, and therefore $T_p f : T_p M \rightarrow T_{f(p)} N$ is linear.

Definition 5.6. We call $f : M \rightarrow N$ a C^l -map, if f extends to a C^l -map $f : U \rightarrow \mathbb{R}^n$ as above. In this situation, if $T_p f$ is not surjective for a $p \in M$, we call p a *critical point* of f , and $f(p) \in N$ a *critical value*. A *regular point* of f is a $p \in M$ with $T_p f$ surjective, and a *regular value* $x \in N$ is a point which is not a critical value.

The following is a slight generalization of Theorem 5.

Theorem 7. Let $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$ be C^l -manifolds, and $f : M \rightarrow N$ a C^l -map. If $x \in N$ is a regular value, then $f^{-1}(\{x\}) \subset M$ is a C^l -manifold. The dimension is $\dim M - \dim N$.

Proof. Let $y \in f^{-1}(\{x\}) \subset M$. Assume that $\dim M = r$ and $\dim N = s$. We need to find a chart around y , that is, a diffeomorphism $\psi : U \rightarrow U' \subset \mathbb{R}^{r-s} \times \mathbb{R}^{m+s-r}$

such that $\psi(U \cap f^{-1}(\{x\})) = U' \cap (\mathbb{R}^{r-s} \times \{0\})$, and U is an open subset of $\mathbb{R}^m$ containing y . Let $g : V \rightarrow V' \subset \mathbb{R}^s \times \mathbb{R}^{n-s}$ be a chart around x with $g(x) = 0$, and $h : W \rightarrow W' \subset \mathbb{R}^r \times \mathbb{R}^{m-r}$ a chart around y with $h(y) = 0$. By possibly shrinking W we can assume that $f(W) \subset V$. Let $B \subset \mathbb{R}^r$ be a small open disc such that $B \times \{0\} \subset W'$. Define $\phi : B \rightarrow \mathbb{R}^s$ be defined by

$$B \longrightarrow W' \cap (\mathbb{R}^s \times \{0\}) \xrightarrow{h^{-1}} W \cap M \xrightarrow{f} V \cap M \xrightarrow{g} \mathbb{R}^s \times \{0\} = \mathbb{R}^s$$

Note that by Proposition 5.2, the tangent space $T_y M$ is spanned by $Dh^{-1}(0)(\mathbb{R}^s)$, and since $T_z f$ is surjective for every $z \in M$ with $f(z) = x$, and g is a chart, we get that $0 \in \mathbb{R}^s$ is a regular value of ϕ . By Theorem 5, $\phi^{-1}(\{0\})$ is a manifold, so there is a chart $\chi : O \rightarrow O' \subset \mathbb{R}^{r-s} \times \mathbb{R}^s$ around $0 \in \mathbb{R}^r$, with $O \subset B \subset \mathbb{R}^r$ a small open subset around 0. Choose an open subset $U \subset \mathbb{R}^m$ containing y and contained in W , with $p \circ h(U) \subset O$, where $p : \mathbb{R}^s \times \mathbb{R}^{m-s} \rightarrow \mathbb{R}^s$ is projection. Then $\psi = (\chi \times \text{id}) \circ h : U \rightarrow U' \subset \mathbb{R}^{r-s} \times \mathbb{R}^s \times \mathbb{R}^{m+s-r}$ is the required chart. $\square$

Remark. The basic idea is to work locally, and working locally on a manifold is the same as working in some $\mathbb{R}^r$, via the charts. We will use this idea to discuss differential equations defined on manifolds. First we need the definition of a vector field on a manifold.

Definition 5.7. Let $M \subset \mathbb{R}^n$ be a C^l -manifold, $l \geq 1$. A continuous map $v : M \rightarrow \mathbb{R}^n$ is called a *vector field on M* , if $v(p) \in T_p M$ for all $p \in M$. It is called a C^l -vector field, if there is an open set $U \subset \mathbb{R}^n$ containing M , such that v extends to a C^l -function $v : U \rightarrow \mathbb{R}^n$.

Example. Let $S^{n-1} = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \|x\|_2 = 1\}$ and $p \in S^{n-1}$. Then

$$T_p S^{n-1} = \{x \in \mathbb{R}^n \mid \langle x, p \rangle = 0\}$$

where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product on $\mathbb{R}^n$. To see this, note that $S^{n-1} = g^{-1}(\{1\})$ with $g(x) = \|x\|_2^2$, and by Theorem 6, $T_p S^{n-1} = \ker Dg(p)$. But $Dg(p)x = 2 \sum_{i=1}^n p_i x_i = 2 \langle p, x \rangle$.

If $n = 2m$, define $v : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$v(x_1, \dots, x_{2m}) = (-x_2, x_1, -x_4, x_3, \dots, -x_{2m}, x_{2m-1}).$$

Clearly $\langle x, v(x) \rangle = 0$ for all $x \in \mathbb{R}^n$, so v defines a C^∞ -vector field on S^{n-1} . Furthermore, $v(x) \neq 0$ for all $x \in S^{n-1}$, so v is a *non-vanishing* vector field.

Remark. There are no non-vanishing vector fields on S^{2m} , that is, for every vector field v on an even-dimensional sphere, there is a point x with $v(x) = 0$. This may be proven in the second half of the course.

Example. We want to describe a torus in cylindrical coordinates. Let $0 < \varepsilon < 1$ and define $\Phi : (1 - \varepsilon, 1 + \varepsilon) \times \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by

$$\Phi(r, \phi, \theta) = \begin{pmatrix} (2 + r \cos \phi) \cos \theta \\ (2 + r \cos \phi) \sin \theta \\ r \sin \phi \end{pmatrix}$$

This function puts a ring $((1 - \varepsilon, 1 + \varepsilon) \times S^1)$ into the xz -plane, and rotates it around the z -axis. The image, with fixed $r = 1$, is the 2-torus. Note that by restricting Φ to small angle values we get charts. So

$$T^2 = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z) = \Phi(1, \phi, \theta), (\phi, \theta) \in \mathbb{R}^2\}$$

is a compact manifold (it is in fact the image of Φ restricted to $\{1\} \times [0, 2\pi] \times [0, 2\pi]$). We can get a non-vanishing vector field on T^2 by choosing a constant vector field on $\mathbb{R}^2$, and then push it onto T^2 by Φ . Let us make this more precise.

Let $U, V \subset \mathbb{R}^n$ be open and $h : U \rightarrow V$ a C^∞ -diffeomorphism. Let $v : U \rightarrow \mathbb{R}^n$ be a C^l -vector field, $l \geq 1$. Then we can define a new vector field h_*v on V by

$$(h_*v)(x) = Dh(h^{-1}(x))v(h^{-1}(x))$$

for $x \in V$. First we use h^{-1} to get back to U , then we can apply the vector field v , and then we use the derivative Dh to transform the vector. Since h is C^∞ , we get that h_*v is a C^l -vector field.

Example. Define a vector field on $\mathbb{R}^3$ by $v(x) = (0, c)$ with $c \in \mathbb{R}^2$. Now define a vector field on T^2 by Φ_*v , where Φ is from the example above. Notice that Φ is only locally invertible, but if x, y satisfy $\Phi(x) = \Phi(y)$, we also have $D\Phi(x) = D\Phi(y)$, so the vector field on T^2 is well defined. If $c \neq 0$, the vector field is non-vanishing.

We now want to solve differential equations on manifolds. Let us start by observing how integral curves transform under diffeomorphisms.

Lemma 5.8. *Let $U, V \subset \mathbb{R}^n$ be open, $h : U \rightarrow V$ a C^∞ -diffeomorphism and v a C^l -vector field on U . If $\gamma : I \rightarrow U$ is an integral curve of v , then $h \circ \gamma$ is an integral curve of h_*v .*

Proof. We have $\gamma'(t) = v(\gamma(t))$. Now

$$\begin{aligned} (h \circ \gamma)'(t) &= Dh(\gamma(t))\gamma'(t) \\ &= Dh(\gamma(t))v(\gamma(t)) \\ &= Dh(h^{-1}(h(\gamma(t))))v(h^{-1}(h(\gamma(t)))) \\ &= h_*v(h \circ \gamma(t)), \end{aligned}$$

which is the condition for being an integral curve. $\square$

Now if v is a C^l -vector field on the C^∞ -manifold $M \subset \mathbb{R}^n$, let $h : U \rightarrow U' \subset \mathbb{R}^m \times \mathbb{R}^{n-m}$ be a chart. We can assume that v is defined on U and a C^l -vector field on U . Then h_* is a vector field on U' , and when restricted to $U' \cap (\mathbb{R}^m \times \{0\})$, the values are in $\mathbb{R}^m \times \{0\}$. So we can think of h_*v as a vector field defined on an open set $V \subset \mathbb{R}^m$. By the Picard-Lindelöf Theorem, there exist integral curves for this vector field. The diffeomorphism h^{-1} then brings these integral curves back into M . Notice that

$$\begin{aligned} h_*^{-1}h_*v(x) &= Dh^{-1}(h(x))h_*v(h(x)) \\ &= Dh^{-1}(h(x))Dh(h^{-1}(h(x)))v(h^{-1}(h(x))) \\ &= Dh^{-1}(h(x))Dh(x)v(x) \\ &= v(x) \end{aligned}$$

so we recover the original vector field. We collect this in the following Lemma.

Lemma 5.9. *Let M be a C^∞ -manifold, and v a C^l -vector field on M , $l \geq 1$. Then for any $p \in M$ there exists an open interval $I \subset \mathbb{R}$ containing 0 and a curve $\gamma : I \rightarrow M$ with $\gamma(0) = p$, and $\gamma'(t) = v(\gamma(t))$ for all $t \in I$. $\square$*

Recall that the Picard-Lindelöf Theorem also gives us uniqueness. We can use this to look at maximal integral curves through a given point.

Lemma 5.10. *Let M be a C^∞ -manifold, and v a C^l -vector field on M , $l \geq 1$. For $i = 1, 2$ let $\gamma_i : I_i \rightarrow M$ be integral curves of v with $\gamma_1(0) = p = \gamma_2(0)$, where I_1 and I_2 are open intervals containing 0. Then $\gamma_1(t) = \gamma_2(t)$ for all $t \in I_1 \cap I_2$.*

Proof. The two curves have to agree in a small open set around 0 by the uniqueness statement of the Picard-Lindelöf Theorem. Let $B = \{t \in I_1 \cap I_2 \cap [0, \infty) \mid \gamma_1(s) = \gamma_2(s) \text{ for all } 0 \leq s \leq t\}$. If B is unbounded, then $\gamma_1(t) = \gamma_2(t)$ for all $t \geq 0$. Otherwise let $T = \sup B$. Then $T \notin I_1 \cap I_2$, for otherwise let $t_n < T$ be a sequence in B converging to T . By continuity and the fact that the $t_n \in B$, we get $\gamma_1(T) = \gamma_2(T) = x \in M$. As γ_1 and γ_2 are integral curves around x , they agree in a small open set around T (recall that I_1 and I_2 are open), but then T cannot be the supremum of B .

Therefore $T \notin I_1 \cap I_2$, and γ_1 and γ_2 agree on $I_1 \cap I_2 \cap [0, \infty)$. To see that they agree on negative numbers, we use a symmetric argument. $\square$

This means that γ_1 and γ_2 can be combined to an integral curve on $I_1 \cup I_2$. Given a point $p \in M$, we can now look at the maximal interval I , on which an integral curve $\gamma : I \rightarrow M$ with $\gamma(0) = p$ is defined. If M is compact, this interval is always $\mathbb{R}$.

Proposition 5.11. *Let M be a compact C^∞ -manifold, and v a C^l -vector field on M with $l \geq 1$. Then for every $p \in M$ there is an integral curve $\gamma : \mathbb{R} \rightarrow M$ with $\gamma(0) = p$.*

Proof. Let $\gamma : I \rightarrow M$ be the maximal integral curve through $p \in M$, and assume $T = \sup I < \infty$. Choose a sequence $t_n \in I$ converging to T . Then $\gamma(t_n)$ is a sequence in M , and by compactness, there is a subsequence that converges to a point $x \in M$. By the Picard-Lindelöf Theorem we can find a continuous $\alpha : U_1 \times I_1 \rightarrow U$, where $U_1 \subset U$ are open sets containing x and I_1 is a small open interval containing 0, with $\alpha(y, 0) = y$ for all $y \in U_1$ and $\frac{\partial}{\partial t} \alpha(y, t) = v(\alpha(y, t))$. Choose n_0 with $T - t_{n_0} \in I_1$ and $\gamma(t_{n_0}) \in U_1$. Define $\beta(t) = \alpha(\gamma(t_{n_0}), t - t_{n_0})$. Then β is an integral curve with $\beta(t_{n_0}) = \gamma(t_{n_0})$, so by Lemma 5.10 they agree on the intersection of domains. Note that β is defined for T and some values bigger than T . Therefore the integral curve γ can be extended using β , which contradicts maximality. It follows that γ is defined for all $t \geq 0$. The argument for negative numbers is completely analogous. $\square$

Recall that the Picard-Lindelöf Theorem gave us even more than just existence and uniqueness of integral curves, it also states that integral curves can be pieced together along their starting points to yield a continuous map. We use this to define a continuous action of $\mathbb{R}$ on a compact manifold M .

Let v be a C^l -vector field on a compact C^∞ -manifold M . Define

$$\begin{aligned} \Phi : M \times \mathbb{R} &\rightarrow M \\ \Phi(x, t) &= \gamma_x(t) \end{aligned}$$

where $\gamma_x : \mathbb{R} \rightarrow M$ is the unique maximal integral curve with $\gamma_x(0) = x$ for $x \in M$. This map is called the *flow* of v .

Theorem 8. *Let M be a compact C^∞ -manifold and v a C^l -vector field with $l \geq 1$. Then the flow of v is a continuous function satisfying*

- (1) $\Phi(x, 0) = x$ for all $x \in M$.
- (2) $\Phi(\Phi(x, t), s) = \Phi(x, t + s)$ for all $x \in M$ and $t, s \in \mathbb{R}$.

Proof. To see that Φ is continuous, let $(x, t) \in M \times \mathbb{R}$. Near (x, t) , the function Φ is just the collection of integral curves, by the uniqueness statement of the Picard-Lindelöf Theorem it agrees with the function $\alpha : U_1 \times I_1 \rightarrow U$ provided by that theorem. Here U is a chart neighborhood of x . As α is continuous, so is Φ .

Statement (1) is clear by construction. To see (2), let $y = \Phi(x, t)$, so $\gamma_x(t) = y$. If we define $\gamma(u) = \gamma_x(u + t)$, then γ is an integral curve, and $\gamma(0) = y$, so we have $\gamma = \gamma_y$. In particular

$$\begin{aligned} \Phi(\Phi(x, t), s) &= \Phi(y, s) \\ &= \gamma_y(s) \\ &= \gamma_x(s + t) \\ &= \Phi(x, s + t). \end{aligned}$$

This finishes the proof. □

Remark. If we write $x \cdot t = \Phi(x, t)$ for $x \in M$ and $t \in \mathbb{R}$, then the properties of the theorem are

- (1) $x \cdot 0 = x$ for all $x \in M$.
- (2) $(x \cdot t) \cdot s = x \cdot (t + s)$ for all $x \in M$ and $t, s \in \mathbb{R}$.

These are exactly the conditions needed for an action of the abelian group $\mathbb{R}$ on the set M . The fact that Φ is continuous can be expressed by saying that the topological group $\mathbb{R}$ acts on the topological space M .

In particular, every C^1 -vector field on a compact manifold gives rise to an action of the real numbers on that manifold.

Remark. It is possible to define a flow on non-compact manifolds, but we do not get integral curves to be defined for all times in general, so the domain will usually be smaller than $M \times \mathbb{R}$. The conditions (1) and (2) still hold, if $s, t, s + t$ fall into the domain.

Remark. The flow Φ is also as differentiable as the vector field v , but in order to show this, we would have to put more effort into differential equations.

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF DURHAM, SCIENCE LABORATORIES,
SOUTH RD, DURHAM DH1 3LE, UNITED KINGDOM

E-mail address: `dirk.schuetz@durham.ac.uk`