Ordinary Differential Equations

Iosif Petrakis

Mathematisches Institut der Universität München
Summer term 2018

Acknowledgments

These lecture notes are based on [**6**]. I would like to thank Nils Köpp for constructing all figures included, and the students who followed this course for their comments and suggestions.

Contents

CHAPTER 1

Basic ideas of ODEs

1.1. Review of topology in $\mathbb{R}^n$

In this section we review the basic facts of the topology in $\mathbb{R}^n$ that we are going to use subsequently.

DEFINITION 1.1.1. Let X be a vector space over $\mathbb{R}$. A *real inner product* on X is a mapping $\langle\langle \cdot, \cdot \rangle\rangle : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ and $\lambda \in \mathbb{R}$ the following hold:
(i) $\langle\langle x, x\rangle\rangle \geq 0$ (positivity).
(ii) $\langle\langle x, x\rangle\rangle = 0 \Rightarrow x = 0$ (definiteness).
(iii) $\langle\langle x, y\rangle\rangle = \langle\langle y, x\rangle\rangle$ (symmetry).
(iv) $\langle\langle x + y, z\rangle\rangle = \langle\langle x, z\rangle\rangle + \langle\langle y, z\rangle\rangle$ (left additivity).
(v) $\langle\langle \lambda x, y\rangle\rangle = \lambda\langle\langle x, y\rangle\rangle$ (left homogeneous)
If $\langle\langle \cdot, \cdot \rangle\rangle$ is a real inner product on X, the pair $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ is called a *real inner product space*. A *real norm* on X is a mapping $||.|| : X \to \mathbb{R}$ such that for every $x, y \in X$ and $\lambda \in \mathbb{R}$ the following hold:
(i) $||x|| \geq 0$ (positivity).
(ii) $||x|| = 0 \Rightarrow x = 0$ (definiteness).
(iii) $||x + y|| \leq ||x|| + ||y||$ (triangle inequality).
(iv) $||\lambda x|| = |\lambda|||x||$.
If $||.||$ is a real norm on X, the pair $(X, ||.||)$ is called a *real normed space*. Unless stated otherwise, an inner product (space) means here a real inner product (space), and a norm(ed space) means here a real norm(ed space). We use the notation $X^* := X \setminus \{0\}$.

Because of symmetry an inner product is bilinear (i.e., it is also right additive and right homogeneous). Next we show that an inner product is determined by its diagonal entries.

PROPOSITION 1.1.2. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be an inner product space and $x, y \in X$.*
(i) (*Polarization identity*) $\langle\langle x, y\rangle\rangle = \frac{1}{4}\big(\langle\langle x + y, x + y\rangle\rangle - \langle\langle x - y, x - y\rangle\rangle\big)$.
(ii) $x = 0 \Leftrightarrow \forall_{z\in X}\big(\langle\langle x, z\rangle\rangle = 0\big)$.
(iii) $\forall_{z\in X}\big(\langle\langle x, z\rangle\rangle = \langle\langle y, z\rangle\rangle\big) \Rightarrow x = y$.

PROOF. (i) Clearly, $\langle\langle x+y, x+y\rangle\rangle - \langle\langle x-y, x-y\rangle\rangle = 4\langle\langle x, y\rangle\rangle$.
(ii) If $x = 0$, then $\langle\langle x, z\rangle\rangle = \langle\langle 0, z\rangle\rangle = \langle\langle 0+0, z\rangle\rangle = \langle\langle 0, z\rangle\rangle + \langle\langle 0, z\rangle\rangle$. Hence $\langle\langle 0, z\rangle\rangle = 0$. For the converse, if $\forall_{z\in X}\big(\langle\langle x, z\rangle\rangle = 0\big)$, then $\langle\langle x, x\rangle\rangle = 0$, hence $x = 0$.
(iii) By the hypothesis we get $\forall_{z\in X}\big(\langle\langle x-y, z\rangle\rangle = 0\big)$, hence by (ii) $x = y$. $\square$

If $x = 0$, then $||x|| = 0$. Moreover, if $x = 0$, or $y = 0$, or $y = \lambda x$, for some $\lambda > 0$, then equality holds in the triangle inequality.

DEFINITION 1.1.3. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are in $\mathbb{R}^n$, their *Euclidean inner product* is defined by

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

It is immediate to see that the Euclidean inner product is an inner product on $\mathbb{R}^n$. If we define the *Minkowski product* $(\cdot,\cdot)$ on $\mathbb{R}^4$ by

$$\big((x, s), (y, t)\big) := \sum_{i=1}^{3} x_i y_i - st,$$

for every $(x, s), (y, t) \in \mathbb{R}^4$, we get a function, which is symmetric, left additive and left homogeneous, but does not satisfy positivity and definiteness. Hence, positivity and definiteness are *independent* from the rest properties of an inner product. The pair $(\mathbb{R}^4, (\cdot,\cdot))$ is called the *Minkowski space*, and it is very important in the special theory of relativity. If we identify space with all pairs $(x, 0)$, then $\big((x, 0), (x, 0)\big) \geq 0$, and if we identify time with all pairs $(0, s)$, then $\big((0, s), (0, s)\big) \leq 0$. For this reason we say that an element (x, s) of the Minkowski space is *space-like*, if $\big((x, s), (x, s)\big) \geq 0$, and we say that it is *time-like*, if $\big((x, s), (x, s)\big) \leq 0$.

DEFINITION 1.1.4. If $x \in \mathbb{R}^n$, the *Euclidean norm* $|x|$ of x is defined by

$$|x| := \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

To show that the Euclidean norm is a norm we need the following.

PROPOSITION 1.1.5 (Inequality of Cauchy). *If $x, y \in \mathbb{R}^n$, then*

$$|\langle x, y\rangle| \leq |x||y|.$$

PROOF. (Bishop) By definition we need to show

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}},$$

which is equivalent to

$$A := \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) =: B.$$

This we get as follows:

$$\begin{aligned}
B - A &= \sum_{i=1}^n x_i^2 \sum_{j=1}^n y_j^2 - \sum_{i=1}^n x_i y_i \sum_{j=1}^n x_j y_j \\
&= \frac{1}{2} \sum_{i=1}^n x_i^2 \sum_{j=1}^n y_j^2 + \frac{1}{2} \sum_{j=1}^n x_j^2 \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i \sum_{j=1}^n x_j y_j \\
&= \sum_{i,j=1}^n \frac{1}{2} \big(x_i^2 y_j^2 + x_j^2 y_i^2 - 2 x_i y_i x_j y_j \big) \\
&= \sum_{i,j=1}^n \frac{1}{2} \big(x_i y_j - x_j y_i \big)^2 \\
&\geq 0.
\end{aligned}$$

□

An inner product on X induces a norm on X defined by

$$||x|| = \langle\langle x, x \rangle\rangle^{\frac{1}{2}}.$$

To show that $||.||$ is a norm on X we need the inequality

$$\big| \langle\langle x, y \rangle\rangle \big| \leq ||x|| \; ||y||,$$

which generalizes the inequality of Cauchy. Clearly, the Euclidean norm is the norm induced by the Euclidean inner product. Geometrically, if $x \in \mathbb{R}^n$, then $|x|$ is the *length* of the vector x and

$$\langle x, y \rangle = |x| \; |y| \cos \theta(x, y),$$

where θ is the *angle* between x and y, which for $x \neq 0$ and $y \neq 0$ is defined by

$$\theta(x, y) := \arccos \frac{\langle x, y \rangle}{|x||y|}.$$

If $\langle x, y \rangle = 0$, we say that x (y) is *orthogonal* to y (x).

If $(X, ||.||)$ is a normed space, the triangle inequality implies the *reverse* triangle inequality[1]

$$\big| ||x|| - ||y|| \big| \leq ||x - y||,$$

for every $x, y \in X$. If we replace y by $-y$, we get

$$||x|| - ||y|| \leq \big| ||x|| - ||y|| \big| \leq ||x + y||.$$

The next theorem is a sharp version of the triangle inequality. If $a, b \in \mathbb{R}$, we use the notations $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

[1]The reverse triangle inequality implies that $||.||$ is 1-Lipschitz on X with respect to $||.||$.

THEOREM 1.1.6 (Sharp triangle inequality). *If $(X, ||.||)$ is a normed space and $x, y \in X^*$, the following hold:*

$$||x+y|| \le ||x|| + ||y|| - \left(2 - \left|\left|\frac{x}{||x||} + \frac{y}{||y||}\right|\right|\right)\big(||x|| \wedge ||y||\big). \tag{1.1}$$

$$||x+y|| \ge ||x|| + ||y|| - \left(2 - \left|\left|\frac{x}{||x||} + \frac{y}{||y||}\right|\right|\right)\big(||x|| \vee ||y||\big). \tag{1.2}$$

Moreover, if either $||x|| = ||y||$ or $y = \lambda x$, for some $\lambda > 0$, then equality holds in both (1.1) and (1.2).

PROOF. (Maligranda) Without loss of generality we assume that $||x|| \le ||y||$, hence $||x|| \wedge ||y|| = ||x||$. Using the triangle inequality we have that

$$\begin{aligned}
||x+y|| &= \left|\left|\frac{||x||}{||x||}x + \frac{||x||}{||y||}y + \left(1 - \frac{||x||}{||y||}\right)y\right|\right| \\
&= \left|\left|\frac{||x||}{||x||}x + \frac{||x||}{||y||}y + \frac{||y|| - ||x||}{||y||}y\right|\right| \\
&\le \left|\left|\frac{||x||}{||x||}x + \frac{||x||}{||y||}y\right|\right| + \left|\left|\frac{||y|| - ||x||}{||y||}y\right|\right| \\
&= ||x||\left|\left|\frac{x}{||x||} + \frac{y}{||y||}\right|\right| + ||y|| - ||x|| \\
&= ||y|| + ||x||\left(\left|\left|\frac{x}{||x||} + \frac{y}{||y||}\right|\right| - 1\right) \\
&= ||x|| + ||y|| - \left(2 - \left|\left|\frac{x}{||x||} + \frac{y}{||y||}\right|\right|\right)||x||.
\end{aligned}$$

The rest of the proof is an exercise. □

THEOREM 1.1.7 (Jordan, von Neumann). *Let $(X, ||.||)$ be a normed space. The following are equivalent.*

(i) The norm $||.||$ is induced by some inner product $\langle\langle \cdot, \cdot \rangle\rangle$ on X.

(ii) The norm $||.||$ satisfies the parallelogram law i.e., for every $x, y \in X$

$$||x+y||^2 + ||x-y||^2 = 2\big(||x||^2 + ||y||^2\big).$$

PROOF. (i) $\Rightarrow$ (ii) It follows from a simple calculation.
(ii) $\Rightarrow$ (i) Due to the polarization identity it is natural to define

$$\langle\langle x, y \rangle\rangle := \frac{1}{4}\left(||x+y||^2 - ||x-y||^2\right). \tag{1.3}$$

Positivity, definiteness and symmetry of $\langle\langle x, y\rangle\rangle$ follow immediately. It is also straightforward to see that

$$\langle\langle -x, y\rangle\rangle = -\langle\langle x, y\rangle\rangle. \tag{1.4}$$

In order to show left additivity we have from the parallelogram law and the definition of $\langle\langle x, y\rangle\rangle$ that

$$\begin{aligned}
4\langle\langle x+z, y\rangle\rangle &= ||x+z+y||^2 - ||x+z-y||^2 \\
&= \left|\left|\left(x+\frac{y}{2}\right)+\left(z+\frac{y}{2}\right)\right|\right|^2 - \left|\left|\left(x-\frac{y}{2}\right)+\left(z-\frac{y}{2}\right)\right|\right|^2 \\
&= 2\left|\left|x+\frac{y}{2}\right|\right|^2 + 2\left|\left|z+\frac{y}{2}\right|\right|^2 - ||x-z||^2 - \\
&\qquad - \left(2\left|\left|x-\frac{y}{2}\right|\right|^2 + 2\left|\left|z-\frac{y}{2}\right|\right|^2 - ||x-z||^2\right) \\
&= 2\left(\left|\left|x+\frac{y}{2}\right|\right|^2 - \left|\left|x-\frac{y}{2}\right|\right|^2\right) + 2\left(\left|\left|z+\frac{y}{2}\right|\right|^2 - \left|\left|z-\frac{y}{2}\right|\right|^2\right) \\
&= 8\langle\langle x, \frac{y}{2}\rangle\rangle + 8\langle\langle z, \frac{y}{2}\rangle\rangle.
\end{aligned}$$

Hence we get

$$\langle\langle x+z, y\rangle\rangle = 2\left(\langle\langle x, \frac{y}{2}\rangle\rangle + \langle\langle z, \frac{y}{2}\rangle\rangle\right). \tag{1.5}$$

If in (1.5) we set $z = 0$, we get for every $x, y \in X$

$$\langle\langle x, y\rangle\rangle = 2\langle\langle x, \frac{y}{2}\rangle\rangle. \tag{1.6}$$

Consequently, (1.5) becomes

$$\langle\langle x+z, y\rangle\rangle = 2\left(\langle\langle x, \frac{y}{2}\rangle\rangle + \langle\langle z, \frac{y}{2}\rangle\rangle\right) = \langle\langle x, y\rangle\rangle + \langle\langle z, y\rangle\rangle.$$

The rest of the proof is an exercise. □

Note that in [**1**] one can find about 350 characterizations of a normed space induced by an inner product!

It is often convenient to work with norms on $\mathbb{R}^n$ other than the Euclidean norm. It is easy to show that the following mappings are norms on $\mathbb{R}^n$

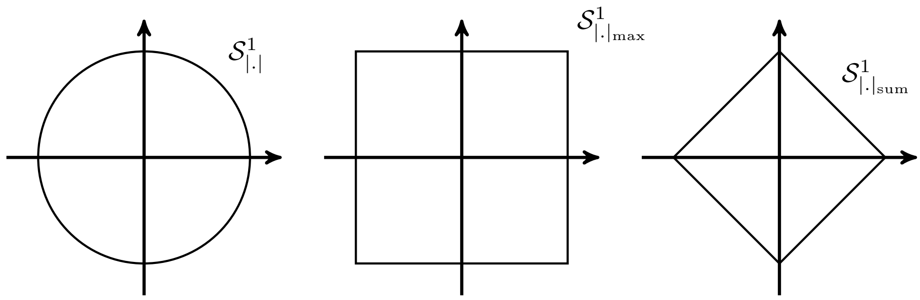
$$|x|_{\text{sum}} := \sum_{i=1}^{n} |x_i| =: \sum_i |x_i|,$$

$$|x|_{\max} := \max\{|x_i| \mid i \in \{1, \ldots, n\}\} =: \max_i |x_i|.$$

If $n = 1$ and $x \in \mathbb{R}$, then $|x|_{\text{sum}} = |x| = |x|_{\max}$. The *unit sphere* of a normed space $(X, ||.||)$ is the set

$$\mathcal{S}^1_{||.||} := \{x \in X \mid ||x|| = 1\}.$$

The unit spheres $\mathcal{S}^1_{|\cdot|}, \mathcal{S}^1_{|\cdot|_{\max}}$ and $\mathcal{S}^1_{|\cdot|_{\text{sum}}}$ of $\mathbb{R}^2$ are pictured as follows:



Especially for $\mathbb{R}^n$ we define the n-sphere $\mathbb{S}^n$, for $n \geq 1$, as follows:

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} \mid |x| = 1\}.$$

If $\mathcal{B} = \{f_1, \ldots, f_n\}$ is any basis for $\mathbb{R}^n$, there are $\mathcal{B}$-versions of the aforementioned norms on $\mathbb{R}^n$: if $x \in \mathbb{R}^n$ and

$$x = \sum_i t_i f_i,$$

then e.g., the $\mathcal{B}$-Euclidean norm and the $\mathcal{B}$-max norm are defined, respectively, as follows:

$$|x|_{\mathcal{B}} := \left(\sum_i t_i^2\right)^{\frac{1}{2}}$$

$$|x|_{\mathcal{B},\max} := \max_i |t_i|.$$

Definition 1.1.8. Let $(X, ||.||)$ be a normed space and $f : X \to \mathbb{R}$.
We say that f is *convex*, if

$$\forall_{x,y \in X} \forall_{t \in (0,1)} \big(f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)\big),$$

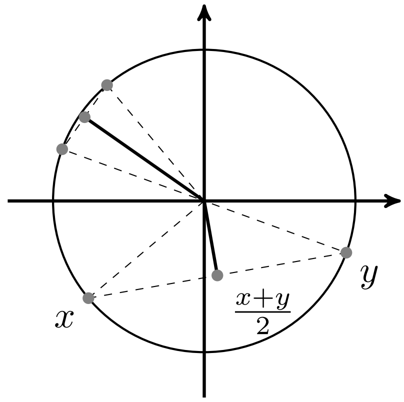
and we say that f is *strictly convex*, if

$$\forall_{x,y \in X} \big(x \neq y \Rightarrow \forall_{t \in (0,1)} (f(tx + (1-t)y) < tf(x) + (1-t)f(y))\big).$$

The normed space $(X, ||.||)$ is called *strictly convex*, if

$$\forall_{x,y \in X} \left(x \neq y \wedge ||x|| = 1 = ||y|| \Rightarrow \left|\left|\frac{x+y}{2}\right|\right| < 1\right).$$

The identity function $\mathrm{id}_{\mathbb{R}}$ on $\mathbb{R}$ is convex, but not strictly convex function. If a normed space is strictly convex, its unit sphere $\mathcal{S}^1_{||.||}$ includes no line segment, as the middle points are not in $\mathcal{S}^1_{||.||}$. The normed space $(\mathbb{R}^2, |.|)$ is strictly convex. A normed space generated by some inner product is always strictly convex.



PROPOSITION 1.1.9. *Let $(X, ||.||)$ be a normed space.*
(i) *The norm $||.||$ is a convex function, which is not strictly convex.*
(iii) *If the norm $||.||$ is induced by some inner product $\langle\langle\cdot,\cdot\rangle\rangle$ on X, then $(X, ||.||)$ is a strictly convex normed space.*

PROOF. Exercise. □

Using Proposition 1.1.9(iii) we can find norms that are not induced by some inner product (exercise).

PROPOSITION 1.1.10. *Let $(X, \langle\langle\cdot,\cdot\rangle\rangle)$ be an inner product space and let $||.||$ be the norm on X induced by $\langle\langle\cdot,\cdot\rangle\rangle$.*
(i) If $x, y \in X$, the following hold:

$$|\langle\langle x, y\rangle\rangle| = ||x||\ ||y|| \ \Leftrightarrow\ \langle\langle y, y\rangle\rangle x = \langle\langle x, y\rangle\rangle y,$$
$$||x + y|| = ||x|| + ||y|| \ \Leftrightarrow\ ||y||x = ||x||y.$$

(ii) *The function $||.||^2$ is a strictly convex function.*

PROOF. Exercise. □

DEFINITION 1.1.11. A *metric* on some set X is a mapping $d : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ the following hold:
(i) $d(x, y) \geq 0$.
(ii) $d(x, y) = 0 \Leftrightarrow x = y$.
(iii) $d(x, y) = d(y, x)$.
(iv) $d(x, y) \leq d(x, z) + d(z, y)$.
If d is a metric on X, the pair (X, d) is called a *metric space*.

A norm $||.||$ on the real vector space X induces a metric on X defined by

$$d(x, y) := ||x - y||.$$

DEFINITION 1.1.12. The *Euclidean metric* ε on $\mathbb{R}^n$ is the metric induced by the Euclidean norm on $\mathbb{R}^n$ i.e., $\varepsilon(x, y) := |x - y|$, for every $x, y \in \mathbb{R}^n$.

PROPOSITION 1.1.13. *Let X be a real vector space and let d be a metric on X. The following are equivalent:*

(i) There is a norm $||.||$ on X that induces d.
(ii) If $x, y, z \in X$ and $\lambda \in \mathbb{R}$, then d satisfies the following:
(a) $d(x, y) = d(x + z, y + z)$.
(b) $d(\lambda x, \lambda y) = |\lambda| d(x, y)$.

PROOF. Exercise. □

Using Proposition 1.1.13 we can find a metric on any real vector space X that is not induced by some norm on X.

DEFINITION 1.1.14. If $x \in \mathbb{R}^n$ and $\epsilon > 0$, the ϵ-*neighborhood* of x is the set

$$\mathcal{B}(x, \epsilon) := \{y \in \mathbb{R}^n \mid |y - x| < \epsilon\}.$$

The ϵ-*ball* around x is the set

$$\mathcal{B}(x, \epsilon] := \{y \in \mathbb{R}^n \mid |y - x| \leq \epsilon\}.$$

The 1-ball around 0 is called the *unit ball.* Let $X \subseteq \mathbb{R}^n$. X is *convex*, if for every $x, y \in X$ the *line segment* between x and y

$$\{tx + (1 - t)y \mid t \in (0, 1)\}$$

is included in X. We say that X is a *neighborhood* of x, if there is some $\epsilon > 0$ such that $\mathcal{B}(x, \epsilon) \subseteq X$, and we call X *open*, if X is a neighborhood of every $x \in X$. X is *bounded*, if there is $\epsilon > 0$ such that $X \subseteq \mathcal{B}(0, \epsilon)$. The convergence of a sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}^n$ to the limit $x \in \mathbb{R}^n$ is defined by

$$x_n \xrightarrow{n} x := \lim_{n \to \infty} x_n = x :\Leftrightarrow \forall_{\epsilon > 0} \exists_{n(\epsilon) \in \mathbb{N}} \forall_{n \geq n(\epsilon)} \big(|x_n - x| < \epsilon\big).$$

A sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}^n$ is a *Cauchy sequence*, if

$$\forall_{\epsilon > 0} \exists_{n(\epsilon) \in \mathbb{N}} \forall_{m, n \geq n(\epsilon)} \big(|x_m - x_n| < \epsilon\big).$$

X is *closed*, if every convergent sequence in X has its limit in X, and X is *compact*, if every sequence in X has a convergent subsequence in X.

All concepts found in Definition 1.1.14 and Definition 1.1.19 are generalized to arbitrary metric spaces. Note that the above notions of ϵ-neighborhood, neighborhood, open set, closed set, of convergence and the various continuity concepts are defined with respect to the Euclidean norm on $\mathbb{R}^n$. Usually we refer to them as a *Euclidean neighborhood*, a *Euclidean open set* and so on. Soon we will see that this is not a loss of generality. Convexity of sets is generalized to arbitrary normed spaces, and it is the necessary property of the domain of a convex function.

It is easy to see that the ϵ-neighborhoods and the ϵ-balls of a normed space are convex sets. This is not generally the case for the ϵ-neighborhoods and the ϵ-balls

$$\mathcal{B}_d(x, \epsilon) := \{y \in X \mid d(y, x) < \epsilon\},$$
$$\mathcal{B}_d(x, \epsilon] := \{y \in X \mid d(y, x) \leq \epsilon\}$$

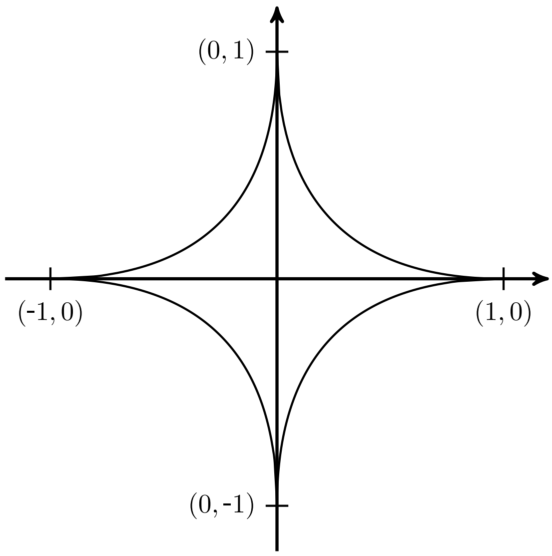
of a metric space (X, d). E.g., let the metric σ on $\mathbb{R}^2$ be defined by

$$\sigma(x, y) := \sqrt{|x_1 - y_1|} + \sqrt{|x_2 - y_2|},$$

for every $x, y \in \mathbb{R}^2$. If $x \in \mathbb{R}^2$ and $\epsilon > 0$, we show that $\mathcal{B}_\sigma(x, \epsilon)$ is not convex. If $\lambda \in (\frac{\epsilon^2}{2}, \epsilon)$, then $\sqrt{\lambda} < \epsilon$ and $\frac{\epsilon}{2} < \sqrt{\frac{\lambda}{2}}$. If $y = (x_1 + \lambda, x_2)$ and $z = (x_1, x_2 + \lambda)$, then $\sigma(x, y) = \sqrt{\lambda} = \sigma(x, z)$ i.e., $y, z \in \mathcal{B}_\sigma(x, \epsilon)$. Hence,

$$\begin{aligned}
\sigma\bigg(x, \frac{1}{2}x + \frac{1}{2}y\bigg) &= \sigma\bigg((x_1, x_2), (x_1 + \frac{\lambda}{2}, x_2 + \frac{\lambda}{2})\bigg) \\
&= \sqrt{\frac{\lambda}{2}} + \sqrt{\frac{\lambda}{2}} \\
&> \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

i.e., $\frac{1}{2}x + \frac{1}{2}y \notin \mathcal{B}_\sigma(x, \epsilon)$. The non-convex unit ball of σ looks as follows:



DEFINITION 1.1.15. If X is a vector space and d is a metric on X, then X has *convex ϵ-neighborhoods*, if for every $x \in X$ and $\epsilon > 0$ the set $\mathcal{B}_\sigma(x, \epsilon)$ is convex.

From now on we write "iff" instead of "if and only if".

PROPOSITION 1.1.16. *If X is a vector space and d is a metric on X, then X has convex ϵ-neighborhoods iff*

$$\forall_{x,y,z \in X} \forall_{t \in (0,1)} \big(d(x, ty + (1-t)z) \leq d(x, y) \vee d(x, z)\big).$$

PROOF. Suppose first that the condition holds and let $x, y, z \in X$ such that $y, z \in \mathcal{B}_d(x, \epsilon)$. If $t \in (0,1)$, then $d\big(x, ty + (1-t)z\big) \leq d(x,y) \vee d(x,z) < \epsilon$. For the converse suppose that there are $x, y, z \in X$ and $t \in (0,1)$ such that $d\big(x, ty + (1-t)z\big) > d(x,y) \vee d(x,z)$. If we take $\epsilon \in \mathbb{R}$ such that

$$d(x,y) \vee d(x,z) < \epsilon < d\big(x, ty + (1-t)z\big),$$

then $y, z \in \mathcal{B}_d(x, \epsilon)$ and $ty + (1-t)z \notin \mathcal{B}_d(x, \epsilon)$, which contradicts the convexity of $\mathcal{B}_d(x, \epsilon)$. □

It is easy to see that the set $\mathcal{T}$ of open sets of a normed (metric) space X are closed under arbitrary unions and finite intersections, and that X and $\emptyset$ are in $\mathcal{T}$, the so-called *topology* of X. We denote by $\mathcal{T}^c$ the set of the closed sets of X. If $A \subseteq X$, the *interior* $\mathring{A}$ of A and the *closure* $\bar{A}$ of A are defined by

$$\mathring{A} := \bigcup\{G \subseteq X \mid G \subseteq A \wedge G \in \mathcal{T}\},$$

$$\bar{A} := \bigcap\{F \subseteq X \mid F \supseteq A \wedge F \in \mathcal{T}^c\}.$$

Clearly, $\mathring{A}$ is the largest open set included in A and $\bar{A}$ is the smallest closed set that includes A. Moreover, A is open iff $\mathring{A} = A$, and A is closed iff $\bar{A} = A$. If $A, B \subseteq X$ and $\lambda \in \mathbb{R}$, we use the following notations"

$$A + B := \{a + b \mid a \in A, b \in B\},$$

$$\lambda A := \{\lambda a \mid a \in A\}.$$

PROPOSITION 1.1.17. *Let $(X, ||.||)$ be a normed space and $A, B \subseteq X$.*
(i) If A is open, then $A + B$ is open.
(ii) If A is open and $t > 0$, then tA is open.
(iii) If A is convex, then $\mathring{A}$ is convex and $\bar{A}$ is convex.
(iv) If A is a subspace of X, then $A \neq X \Leftrightarrow \mathring{A} = \emptyset$.
(v) If $f : X \to \mathbb{R}$ is linear and $f \neq 0$, then f is open i.e., it maps open sets of X onto open sets of $\mathbb{R}$.

PROOF. (i) If $a \in A$ and $\epsilon > 0$ such that $\mathcal{B}(x, \epsilon) \subseteq A$, and if $b \in B$, then

$$\mathcal{B}(x + y, \epsilon) = \mathcal{B}(x, \epsilon) + \{b\} \subseteq A + \{b\}$$

i.e., $A + \{b\}$ is open. Since,

$$A + B = \bigcup\{A + \{b\} \mid b \in B\},$$

we have that $A + B$ is open as a union of open sets.
(ii) Exercise.
(iii) Since $\mathring{A}$ is open, by (ii) we get $t\mathring{A}$ is open, hence by (i) we have that $t\mathring{A} + (1-t)\mathring{A}$ is open. Since A is convex, $tA + (1-t)A \subseteq A$, and since $\mathring{A} \subseteq A$, we conclude that $t\mathring{A} + (1-t)\mathring{A} \subseteq \mathring{A}$. Since $\mathring{A}$ is the largest open set included in A, we get

$t\mathring{A} + (1-t)\mathring{A} \subseteq \mathring{A}$ i.e., $\mathring{A}$ is convex. If $x, y \in \bar{A}$, there are sequences $(x_n)_{n=1}^{\infty} \subseteq A$ and $(y_n)_{n=1}^{\infty} \subseteq A$ such that $x_n \xrightarrow{n} x$ and $y_n \xrightarrow{n} y$. Since

$$tx_n + (1-t)y_n \xrightarrow{n} tx + (1-t)y,$$

and $tx_n + (1-t)y_n \in A$, by the convexity of A, we get $tx + (1-t)y \in \bar{A}$.
(iv) and (v) Exercises. □

PROPOSITION 1.1.18. *Let $X \subseteq \mathbb{R}^n$.*
(i) If $x_k = (x_{k1}, \ldots, x_{kn}) \in \mathbb{R}^n$, for every $k \in \mathbb{N}$, and $y \in \mathbb{R}^n$, then

$$\lim_{k\to\infty} x_k = y \Leftrightarrow \lim_{k\to\infty} x_{ki} = y_i, \text{ for every } i \in \{1, \ldots, n\}.$$

(ii) A sequence in $\mathbb{R}^n$ converges to a limit iff it is a Cauchy sequence.
(iii) X is closed iff its complement $\mathbb{R}^n \setminus X$ is open.
(iv)(Bolzano-Weierstrass) X is compact iff X is closed and bounded.
(v) If $n = 1$, and $X \neq \emptyset$ and compact, then X has a maximum and a minimum element.

PROOF. Left to the reader. See also [**4**] and [**12**]. □

DEFINITION 1.1.19. Let $X \subseteq \mathbb{R}^n$ and $f : X \to \mathbb{R}^m$. We say that f is *continuous at* $x_0 \in \mathbb{R}^n$, if

$$\forall_{\epsilon>0}\exists_{\delta>0}\forall_{x\in X}\big(|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon\big),$$

and f is *continuous on* X, if it is continuous at every element of X. We say that f is *sequentially continuous on* X, if for every $(x_n)_{n=1}^{\infty}$ in X and every $x \in X$

$$\lim_{n\to\infty} x_n = x \Rightarrow \lim_{n\to\infty} f(x_n) = f(x).$$

We say that f is *uniformly continuous on* X, if

$$\forall_{\epsilon>0}\exists_{\delta>0}\forall_{x,y\in X}\big(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon\big),$$

and f is *σ-Lipschitz on* X, where $\sigma \geq 0$, if

$$\forall_{x,y\in X}\big(|f(x) - f(y)| \leq \sigma|x - y|\big).$$

PROPOSITION 1.1.20. *Let $X \subseteq \mathbb{R}^n$ and $f : X \to \mathbb{R}^m$.*
(i) f is continuous on X iff f is sequentially continuous on X.
(ii) If f is σ-Lipschitz on X, it is uniformly continuous on X.
(iii) If f is uniformly continuous on X, it is continuous on X.
(iv) If X is compact and if f is continuous on X, then f is uniformly continuous and $f(X)$ is compact.
(v) If $m = 1$, and $X \neq \emptyset$ and compact, and if f is continuous on X, then f has a maximum and a minimum value.

PROOF. Left to the reader. See also [**4**] and [**12**]. □

One can show that $x \mapsto x^2$ is continuous on $\mathbb{R}$, but not uniformly continuous on $\mathbb{R}$, and $x \mapsto \sqrt{|x|}$ is uniformly continuous on $\mathbb{R}$, but not σ-Lipschitz, for every $\sigma > 0$, If $x \in \mathbb{R}^n$, then

$$\big(\max_i |x_i|\big)^2 \leq \sum_i x_i^2 \leq n\big(\max_i |x_i|\big)^2,$$

and taking square roots we get

$$|x|_{\max} \leq |x| \leq \sqrt{n}\ |x|_{\max},$$

or

$$\frac{1}{\sqrt{n}}\ |x| \leq |x|_{\max} \leq |x|.$$

Since $|x|_{\text{sum}} \leq n\ |x|_{\max} \leq n\ |x|$, we also have that

$$\frac{1}{n}\ |x|_{\text{sum}} \leq |x| \leq |x|_{\text{sum}}.$$

Such inequalities hold for every norm on $\mathbb{R}^n$.

LEMMA 1.1.21. *A norm $||.||$ on $\mathbb{R}^n$ is an (Mn)-Lipschitz function, where*

$$M := \max_i ||e_i||,$$

and $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{R}^n$.

PROOF. Let $x \in \mathbb{R}^n$ and let $x = \sum_i x_i e_i$. Then

$$||x|| = \left|\left|\sum_i x_i e_i\right|\right| \leq \sum_i ||x_i e_i|| = \sum_i |x_i| ||e_i|| \leq M \sum_i |x_i| = M|x|_{\text{sum}} \leq Mn|x|.$$

Hence, if $x, y \in \mathbb{R}^n$, we get

$$\big|||x|| - ||y||\big| \leq ||x - y|| \leq Mn|x - y|.$$

□

PROPOSITION 1.1.22 (Equivalence of norms). *Let $||.||$, $||.||_*$ be norms on $\mathbb{R}^n$.*

(i) There are $A > 0$ and $B > 0$ such that for every $x \in \mathbb{R}^n$ we have that

$$A|x| \leq ||x|| \leq B|x|.$$

(ii) There are $A' > 0$ and $B' > 0$ such that for every $x \in \mathbb{R}^n$ we have that

$$A'||x|| \leq ||x||_* \leq B'||x||.$$

PROOF. (i) Since the unit sphere $\mathbb{S}^1_{|.|}$ is non-empty, closed and bounded subset of $\mathbb{R}^n$, and since by Lemma 1.1.21 $||.||$ is continuous on $\mathbb{R}^n$, its restriction to $\mathbb{S}^1_{|.|}$ is continuous on $\mathbb{S}^1_{|.|}$. By Proposition 1.1.20(v) we have that $||.||$ has a maximum value B and a minimum value A on $\mathbb{S}^1_{|.|}$ i.e., for every $x \in \mathbb{R}^n$

$$|x| = 1 \Rightarrow A \leq ||x|| \leq B.$$

If $x = 0$, the inequalities $A|0| \leq ||0|| \leq B|0|$ hold trivially. If $x \neq 0$, then $|x| > 0$ and $\left|\frac{x}{|x|}\right| = 1$. Hence

$$A \leq \left|\left|\frac{x}{|x|}\right|\right| \leq B \Leftrightarrow A|x| \leq ||x|| \leq B|x|.$$

(ii) Exercise. □

Two norms satisfying the inequalities of Proposition 1.1.22(ii) are called *equivalent*. Hence, any two norms on $\mathbb{R}^n$ are equivalent. Two equivalent norms generate the same topology i.e., the same set of open sets, and "behave equivalently" in the sense of the next proposition. Of course, we have already seen that there are geometric properties, like the strict convexity of the resulting normed space, that are not shared by equivalent norms.

PROPOSITION 1.1.23. *Let $||.||$, $||.||_*$ be norms on $\mathbb{R}^n$, $X \subseteq \mathbb{R}^n$, $(x_n)_{n=1}^\infty \subset \mathbb{R}^n$, and $x \in \mathbb{R}^n$.*
(i) X is open with respect to $||.||$ iff X is open with respect to $||.||_*$.
(ii) X is closed with respect to $||.||$ iff X is closed with respect to $||.||_*$.
(iii) X is bounded with respect to $||.||$ iff X is bounded with respect to $||.||_*$.
(iv) $\lim_{n\to\infty}(x_n) = x$ with respect to $||.||$ iff $\lim_{n\to\infty}(x_n) = x$ with respect to $||.||_*$.
(v) $(x_n)_{n=1}^\infty$ is Cauchy with respect to $||.||$ iff $(x_n)_{n=1}^\infty$ is Cauchy with respect to $||.||_*$.
(vi) X is compact with respect to $||.||$ iff X is compact with respect to $||.||_*$.
(vii) The unit ball and the unit sphere with respect to $||.||$ are compact sets.

PROOF. (i) Let $A' > 0$ and $B' > 0$ such that $A'||x|| \leq ||x||_* \leq B'||x||$, for every $x \in \mathbb{R}^n$. Let X be open with respect to $||.||$ i.e., if $x \in X$, there is $\epsilon > 0$ such that

$$\mathcal{B}_{||.||}(x, \epsilon) = \{y \in \mathbb{R}^n \mid ||y - x|| < \epsilon\} \subseteq X.$$

If $y \in \mathbb{R}^n$ such that $||y - x||_* < \epsilon A'$, then, since $A'||y - x|| \leq ||y - x||_* < \epsilon A'$, we get $||y - x|| < \epsilon$. Consequently,

$$x \in \mathcal{B}_{||.||_*}(x, \epsilon A') \subseteq \mathcal{B}_{||.||}(x, \epsilon) \subseteq X.$$

If X is open with respect to $||.||_*$, then working similarly we get

$$x \in \mathcal{B}_{||.||}(x, \frac{\epsilon}{B'}) \subseteq \mathcal{B}_{||.||_*}(x, \epsilon) \subseteq X.$$

$(ii) - (vi)$ Immediately by (i) and the corresponding definitions.
(vii) The unit ball and the unit sphere with respect to the Euclidean norm are closed and bounded, hence compact. Then we use (vi) and Proposition 1.1.22(i). □

DEFINITION 1.1.24. Let $||.||$ be a norm on $\mathbb{R}^n$ and $||.||_*$ a norm on $\mathbb{R}^m$. If $X \subseteq \mathbb{R}^n$ and $f : X \to \mathbb{R}^m$, we call f *Lipschitz*, if there is $\sigma \geq 0$ such that f is σ-Lipschitz i.e.,

$$\forall_{x,y \in X}\big(||f(x) - f(y)||_* \leq \sigma||x - y||\big).$$

The Lipschitz-property does not depend on the choice of norms on $\mathbb{R}^n$ and $\mathbb{R}^m$

Corollary 1.1.25. *Let* $||.||$ *and* $||.||'$ *be norms on* $\mathbb{R}^n$, *and let* $||.||_*$ *and* $||.||'_*$ *be norms on* $\mathbb{R}^m$. *If* $X \subseteq \mathbb{R}^n$ *and* $f : X \to \mathbb{R}^m$, *then* f *is Lipschitz with respect to* $||.||$ *and* $||.||_*$ *iff* f *is Lipschitz with respect to* $||.||'$ *and* $||.||'_*$.

Proof. Exercise. □

Proposition 1.1.26. *Let* $(X, ||.||)$ *and* $(Y, ||.||_*)$ *be normed spaces and* $f : X \to Y$ *linear. The following are equivalent:*
(i) f *is continuous at* $x_0 \in X$.
(ii) f *is continuous at* 0.
(iii) There is $\sigma > 0$ *such that for all* $x \in X$ *we have that* $||f(x)||_* \leq \sigma ||x||$.

Proof. (i) $\Rightarrow$ (ii) If $x_n \xrightarrow{n} 0$, then $x_n + x_0 \xrightarrow{n} x_0$, hence $f(x_n) + f(x_0) \xrightarrow{n} f(x_0)$, which implies that $f(x_n) \xrightarrow{n} 0 = f(0)$.
(ii) Let $\delta(1) > 0$ such that $||x|| < \delta(1) \Rightarrow ||f(x)||_* < 1$, for every $x \in X$. If $x_0 \in X$ such that $x_0 \neq 0$, then

$$\left|\left|\frac{\delta(1)x_0}{2||x_0||}\right|\right| = \frac{\delta(1)}{2} < \delta(1),$$

hence

$$\left|\left|f\left(\frac{\delta(1)x_0}{2||x_0||}\right)\right|\right|_* < 1 \Leftrightarrow \frac{\delta(1)}{2||x_0||}||f(x_0)||_* < 1 \Leftrightarrow ||f(x_0)||_* < \sigma||x_0||,$$

where $\sigma := \frac{2}{\delta(1)}$. If $x_0 = 0$, then the inequality $||f(0)||_* \leq \sigma||0||$ holds trivially.
The implication (iii) $\Rightarrow$ (i) follows from Proposition 1.1.20(iii). □

If $X = \mathbb{R}^n$ we can show that a linear function on $\mathbb{R}^n$ is always continuous.

Proposition 1.1.27. *Let* E *be a normed space. If* $f : \mathbb{R}^n \to E$ *is linear, then* f *is Lipschitz.*

Proof. Exercise. □

The Lipschitz functions between metric spaces are defined as in Definition 1.1.24. A major difference between uniformly continuous functions and Lipschitz functions is that the latter send bounded subsets of their domain to bounded subsets of their codomain, as, for example,

$$\begin{aligned} ||f(x)|| &\leq ||f(x) - f(0)|| + ||f(0)|| \\ &\leq \sigma||x - 0|| + ||f(0)|| \\ &= \sigma M + ||f(0)||, \end{aligned}$$

while the former do not preserve, in general, boundedness; the identity $\mathrm{id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{R}$, where $\mathbb{N}$ is equipped with the discrete metric[2], is uniformly continuous, but $\mathrm{id}_{\mathbb{N}}(\mathbb{N}) = \mathbb{N}$ is not bounded in $\mathbb{R}$.

[2]The discrete metric on a set X is defined by $\rho(x, y) = 0 \Leftrightarrow x = y$ and $\rho(x, y) = 1$, otherwise.

PROPOSITION 1.1.28. *Let E be a subspace of $\mathbb{R}^n$.*
(i) If $||.||$ is a norm on $\mathbb{R}^n$, then its restriction $||.||_{|E}$ to E is a norm on E.
(ii) If $||.||_E$ is a norm on E, there is a norm $||.||$ on $\mathbb{R}^n$ such that $||.||_E = ||.||_{|E}$.
(iii) All norms on E are equivalent.
(iv) If $||.||_E$ is a norm on E, the unit ball and the unit sphere with respect to $||.||_E$ are compact sets.

PROOF. The proof of (i) is immediate. For (ii) we write $\mathbb{R}^n$ as the direct sum $\mathbb{R}^n = E + F$; take a basis $\{e_1, \ldots, e_n\}$ for $\mathbb{R}^n$ such that $\{e_1, \ldots, e_m\}$ is a basis for E, where $m < n$ (if $m = n$, what we want to show follows trivially). Then $F =< \{e_{m+1}, \ldots, e_n\} >$, the span of $\{e_{m+1}, \ldots, e_n\}$. Since an element x of $\mathbb{R}^n$ is written as

$$x = y + z, \quad y \in E, \; z \in F,$$

we define the function

$$||x|| := ||y||_E + |z|.$$

It is immediate to see that $||.||$ is a norm on $\mathbb{R}^n$. Also, $||y|| = ||y||_E$, for every $y \in E$.
(iii) If $|.||_E$, and $||.||_E^{'}$ are two norms on E, let $||.||$ and $||.||^{'}$ be their induced norms on $\mathbb{R}^n$. Since the latter are equivalent, the former are also equivalent.
(iv) The unit ball $\mathcal{B}(0,1] = \{x \in E \mid ||x||_E \leq 1\}$ is bounded with respect to the extension norm $||.||$ of $||.||_E$ to $\mathbb{R}^n$, hence by Proposition 1.1.23(iii) it is also bounded in $\mathbb{R}^n$ (with respect to the Euclidean norm). By the continuity of $||.||$ and the implied continuity of its restriction $||.||_E$, we have that $\mathcal{B}(0,1]$ is closed with respect to $||.||_E$. Hence, by Proposition 1.1.23(ii) it is also closed with respect to the Euclidean norm. $\square$

DEFINITION 1.1.29. If $(x_n)_{n=0}^{\infty} \subset \mathbb{R}^n$, the *sequence of partial sums* $(s_k)_{k=0}^{\infty}$ of $(x_n)_{n=1}^{\infty}$ is defined by

$$s_k := \sum_{i=0}^{k} x_i,$$

and it is often denoted by an infinite series

$$\sum_{k=0}^{\infty} x_k, \quad \text{or} \quad \sum_k x_k.$$

If $\lim_{k\to\infty} s_k = x$, for some $x \in \mathbb{R}^n$, we write

$$\sum_{k=0}^{\infty} x_k = x, \; \text{or} \quad \sum_k x_k = x.$$

If $||.||$ is a norm on $\mathbb{R}^n$, a series $\sum_k x_k$ is *absolutely convergent*, if the series

$$\sum_{k=0}^{\infty} ||x_k||$$

is convergent in $\mathbb{R}$.

If a series is absolutely convergent with respect to some norm $||.||$ on $\mathbb{R}^n$, then it is absolutely convergent with respect to any other norm $||.||_*$ on $\mathbb{R}^n$. For this let $\sigma_k := \sum_{i=0}^{k} ||x_i||$ and $\tau_k := \sum_{i=0}^{k} ||x_i||_*$. By the equivalence of norms there is some $C > 0$ such that if $n > m$,

$$|\tau_n - \tau_m| = \left| \sum_{i=m+1}^{n} ||x_i||_* \right| = \sum_{i=m+1}^{n} ||x_i||_* \leq C \sum_{i=m+1}^{n} ||x_i|| = C|\sigma_n - \sigma_m|.$$

Hence, absolute convergence of a series is independent of the norm on $\mathbb{R}^n$, and we speak of absolute convergence of a series in $\mathbb{R}^n$ without reference to some norm.

PROPOSITION 1.1.30 (Comparison test). *Let $||.||$ be a norm on $\mathbb{R}^n$ and let the series $\sum_k x_k$ in $\mathbb{R}^n$. If there is a series $\sum_k a_k$ in $\mathbb{R}$ such that:*
(i) $a_k \geq 0$, for every k;
(ii) $||x_k|| \leq a_k$, for every k;
(iii) $\sum_k a_k$ converges in $\mathbb{R}$,
then the series $\sum_k x_k$ converges absolutely.

PROOF. If $\tau_k := \sum_{i=0}^{k} ||x_i||$ and $\sigma_k = \sum_{i=0}^{k} a_i$, for every k, and since for $n > m$

$$|\tau_n - \tau_m| = \sum_{i=m+1}^{n} ||x_i|| \leq \sum_{i=m+1}^{n} a_i = \left| \sum_{i=m+1}^{n} a_i \right| = |\sigma_k - \sigma_m|,$$

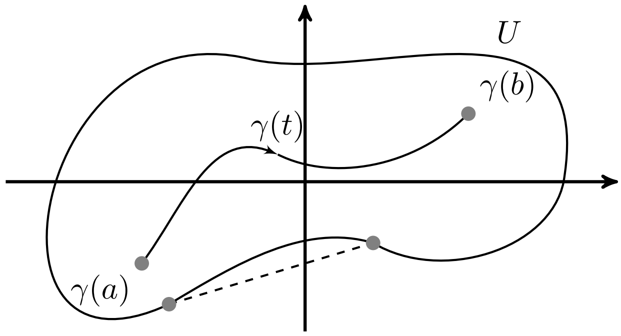
we use the Cauchy criterion for convergence. □

1.2. The Newtonian gravitational field and the method of integrals

The field of ordinary differential equations (ODEs) is closely related to physics. In this section we discuss Newton's second law that connects the physical concept of force field and the mathematical concept of differential equation, and lies at the root of classical mechanics. We shall be working with a particle moving in a field of force. We represent mathematically the notion of trajectory of a moving particle in $\mathbb{R}^n$ (usually $n \leq 3$) by a path in $\mathbb{R}^n$.

DEFINITION 1.2.1. Let $U \subseteq \mathbb{R}^n$. A *path* in U is a continuous function $\gamma : I \to U$, where I is an interval of $\mathbb{R}$. If γ is differentiable on I (i.e., each projection function γ_i is differentiable), the derivative of γ defines a function $\gamma' : I \to \mathbb{R}^n$. If γ' is continuous, we say that γ is C^1, or *continuously differentiable*. If γ' is C^1, we say that γ is C^2. Inductively one defines a function γ to be C^n, where $n > 0$. Moreover, γ is called C^∞, if it is C^n, for every $n > 0$. The set U is called *path-connected*, if for every $x, y \in U$ there is some path $\gamma : [a, b] \to U$ *from x to y* i.e., $\gamma(a) = x$ and $\gamma(b) = y$. Similarly, U is C^i *path-connected*, if there is a C^i path connecting any two points of U, where $i \in \mathbb{N}^+ \cup \{\infty\}$. A path from x to x in U is called a *closed* path, or a *loop* in U.

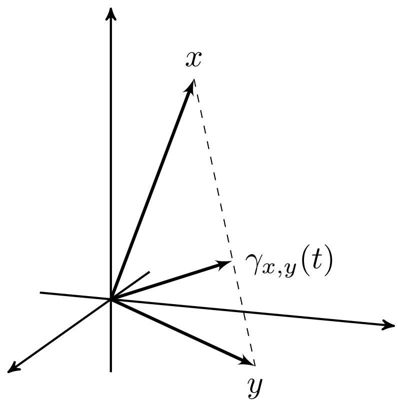
A convex subset of $\mathbb{R}^n$ is path-connected, but the converse is not generally true.



The space $\mathbb{R}^n$ is C^∞ path-connected in the following special way.

PROPOSITION 1.2.2. *Let $x, y \in \mathbb{R}^n$ such that $|y - x| > 0$.*
(i) The function $\gamma_{x,y} : [0, |y - x|] \to \mathbb{R}^n$, defined by

$$\gamma_{x,y}(t) := x + t\frac{y - x}{|y - x|},$$



for every $t \in [0, |y - x|]$ is a C^∞ path from x to y, which is an isometry i.e., for every $s, t \in [0, |y - x|]$

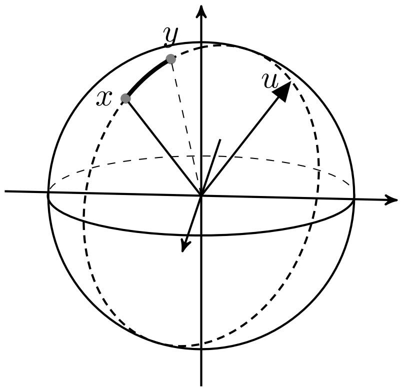
$$|\gamma_{x,y}(s) - \gamma_{x,y}(t)| = |s - t|.$$

(ii) If $\delta_{x,y} : [0, |y - x|] \to \mathbb{R}^n$ is a path from x to y that is an isometry, then $\delta_{x,y}$ is equal to $\gamma_{x,y}$.

PROOF. Exercise. □

Proposition 1.2.3. *Let $x, y \in \mathbb{S}^2$ such that $y \neq x$ and $y \neq -x$.*
(i) If $u \in \mathbb{S}^2$ is orthogonal to x, then the path $\sigma_{x,u} : \mathbb{R} \to \mathbb{S}^2$, defined by

$$\sigma_{x,u}(t) := x \cos t + u \sin t,$$



for every $t \in \mathbb{R}$, parametrizes the great circle $\langle \{x, u\} \rangle \cap \mathbb{S}^2$, where $\langle \{x, u\} \rangle$ is the linear span of x and u, which, since x, u are linearly independent, $\langle \{x, u\} \rangle$ is a plane.
(ii) There is a C^∞ path $\sigma_{x,y} : [0, |y - x|] \to \mathbb{S}^2$ that parametrizes the arc of the unique great circle from x to y.

Proof. Exercise. For (ii) use the vector

$$u := \frac{y - \langle y, x \rangle x}{|y - \langle y, x \rangle x|}.$$

□

Remark 1.2.4. (i) An inner product $\langle\langle \cdot, \cdot \rangle\rangle$ on $\mathbb{R}^n$ is a continuous function.
(ii) Let I be an interval of $\mathbb{R}$ and let $f, g : I \to \mathbb{R}^n$ be C^1.
(a) If $\langle\langle f, g \rangle\rangle : I \to \mathbb{R}$ is defined for every $t \in I$ by

$$\langle\langle f, g \rangle\rangle(t) := \langle\langle f(t), g(t) \rangle\rangle,$$

then, for every $t \in I$ we have that

$$\langle\langle f, g \rangle\rangle'(t) = \langle\langle f'(t), g(t) \rangle\rangle + \langle\langle f(t), g'(t) \rangle\rangle.$$

(b) For every $t \in I$ we have that

$$\langle\langle f'(t), f(t) \rangle\rangle = \frac{1}{2}\big(||f(t)||^2\big)'.$$

Proof. Exercise. □

Differentiability of a function f on an open subset of $\mathbb{R}^n$ means that locally f is well-approximated by some linear, therefore continuous, function on $\mathbb{R}^n$. First we consider a function f that takes values in $\mathbb{R}$.

DEFINITION 1.2.5. Let U be an open subset of $\mathbb{R}^n$, $x_0 \in U$ and $f : U \to \mathbb{R}$. We say that f is *differentiable at* x_0, if there are $A \in \mathbb{R}^n$ and a function ψ defined for all sufficiently small $h \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \psi(x) = 0,$$

and

$$f(x_0 + h) = f(x_0) + \langle A, h \rangle + |h| \psi(x_0).$$

Equivalently, we may write these two conditions in one as follows:

$$f(x_0 + h) = f(x_0) + \langle A, h \rangle + \mathrm{o}(x_0).$$

We say that f is *differentiable on* U, if it is differentiable at every point of U. We define the *gradient* of f at any point x at which all partial derivatives exist to be the vector

$$\mathrm{grad} f(x) := \big(D_1 f(x), \ldots, D_n f(x)\big) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right).$$

One should write $(\mathrm{grad} f)(x)$ but we omit the parentheses for simplicity.

Clearly, the differentiability of f at x_0 implies the continuity of f at x_0. If f, g are differentiable on U, and if $\lambda \in \mathbb{R}$, it is immediate to see that

$$\mathrm{grad}(f + g) = \mathrm{grad} f + \mathrm{grad} g, \quad \text{and } \mathrm{grad}(\lambda f) = \lambda \mathrm{grad} f.$$

PROPOSITION 1.2.6. *Let U be an open subset of $\mathbb{R}^n$, $x_0 \in U$ and $f : U \to \mathbb{R}$.*
(i) If f is differentiable at x_0, and if $A \in \mathbb{R}^n$ such that

$$f(x_0 + h) = f(x_0) + \langle A, h \rangle + \mathrm{o}(x_0),$$

then all partial derivatives of f at x_0 exist, and

$$A = \mathrm{grad} f(x_0).$$

(ii) If all partial derivatives of f exist in U and for each i the function

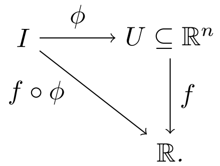
$$U \ni x \mapsto \frac{\partial f}{\partial x_i}(x)$$

is continuous[3]*, then f is differentiable at x_0.*

PROOF. See [**7**], p.322. □

PROPOSITION 1.2.7 (Chain rule). *Let I be an interval of $\mathbb{R}$, and $\phi : I \to \mathbb{R}^n$ differentiable on I such that $\phi(I) \subseteq U$, where U is an open subset of $\mathbb{R}^n$*

[3]In this case f is called C^1. As in Definition 1.2.1, one defines C^n functions for every $n > 0$.



If $f : U \to \mathbb{R}$ is differentiable, $f \circ \phi : I \to \mathbb{R}$ is differentiable and for every $t \in I$

$$(f \circ \phi)'(t) = \big\langle \mathrm{grad} f(\phi(t)), \phi'(t) \big\rangle.$$

Proof. See [**7**], pp.324-325. □

Unfolding the chain rule we get

$$\begin{aligned}(f \circ \phi)'(t) &= \left\langle \left(\frac{\partial f}{\partial x_1}(\phi(t)), \ldots, \frac{\partial f}{\partial x_n}(\phi(t)) \right), \phi'(t) \right\rangle \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_1}(\phi(t)) {\phi_i}'(t) \\ &=: \sum_{i} \frac{\partial f}{\partial x_i}(\phi(t)) \frac{d\phi_i}{dt}(t).\end{aligned}$$

An immediate consequence of the chain rule is that if $f : U \to \mathbb{R}$ is differentiable, and $U \subseteq \mathbb{R}^n$ is path connected, then

$$\mathrm{grad} f = 0 \Rightarrow f \text{ is constant.}$$

If $x \in U$, and $u \in \mathbb{R}^n$ is a fixed vector with $|u| = 1$, the *directional derivative* $D_u f(x)$ *of* $f : U \to \mathbb{R}$ *at* x *in the direction* u is defined by

$$D_u f(x) := \big(f(x + tu)\big)'(0) = g'(0),$$

where $g(t) := f(x + tu)$, for every $t \in J$, for some open interval J in $\mathbb{R}$. Since $g'(t) = \langle \mathrm{grad} f(x + tu), u\rangle$ and $g'(0) = \langle \mathrm{grad} f(x), u\rangle$, if $\mathrm{grad} f(x) \neq 0$, then $D_u f(x)$ becomes maximal precisely when u has the direction of $\mathrm{grad} f(x)$ i.e., $\mathrm{grad} f(x)$ points in the direction of the maximal increase of f at x. Moreover, from the implicit function theorem one can deduce that $\mathrm{grad} f(x)$ is perpendicular to the tangent plane of the *level hypersurface* $S_a(f)$ at x of level $a = f(x)$, where

$$S_a(f) := \{x \in U \mid f(x) = a\}.$$

Definition 1.2.8. Let U be an open subset of $\mathbb{R}^n$. A *vector field* on U is a function $F : U \to \mathbb{R}^n$. If F is represented by its coordinate functions i.e.,

$$F = (f_1, \ldots, f_n),$$

F is continuous (differentiable), if each $f_i : U \to \mathbb{R}$ is continuous (differentiable). F is called *conservative*, if there is a differentiable function $V : U \to \mathbb{R}$ such that[4]

$$F = -\mathrm{grad} V.$$

[4]The negative sign is only traditional, and it can be avoided.

In this case V is called a *potential energy* function for F.

If V is a potential energy function for F and $c \in \mathbb{R}$ some constant, then $V+c$ is also a potential energy function for F. If f is a differentiable function on U, then, because of Proposition 1.2.6(i), we get the vector field on U defined by

$$U \ni x \mapsto \operatorname{grad} f(x).$$

DEFINITION 1.2.9. Let $U \subseteq \mathbb{R}^n$ be open, $\gamma : [a,b] \to U$ a C^1 path and $F : U \to \mathbb{R}^n$ a continuous vector field. The *path integral* of F along γ is defined by

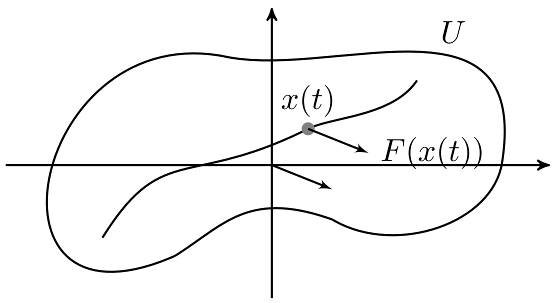
$$\int_\gamma F := \int_a^b \big\langle F(\gamma(t)), \gamma'(t) \big\rangle dt.$$

Note that by Remark 1.2.4(i) and our hypotheses on γ and F the function in the integral is continuous, hence Riemann-integrable.

PROPOSITION 1.2.10. *Let $U \subseteq \mathbb{R}^n$ be path-connected and open, and let $F : U \to \mathbb{R}^n$ be a continuous vector field on U. The following are equivalent.*
(i) F is conservative.
(ii) The path integral of F between any two points of U is independent of the path connecting them.
(iii) The path integral of F along any loop in U is equal to 0.

PROOF. Exercise. □

DEFINITION 1.2.11. Let U be an open subset of $\mathbb{R}^3$. A *force field* on U is a vector field $F : U \to \mathbb{R}^3$, where the vector $F(x)$ assigned to x is interpreted as a force acting on a particle placed at x. A *position function of a particle* in U is a function $x : J \to U$ that is C^2, where J is an open interval in $\mathbb{R}$. The vector $x(t)$ is interpreted as the position of the particle at time t.



If x is a position function of a particle in U and F is a force field on U, we may also say that the *particle is moving in F*. We use the term force field also for vector fields with values in $\mathbb{R}$ or in $\mathbb{R}^2$. If the mass of the particle is $m > 0$, the *kinetic energy* of the particle is the function $T : J \to \mathbb{R}$ defined by[5]

$$T(t) := \frac{1}{2}m\big|\dot{x}(t)\big|^2.$$

If F is conservative and V is a potential energy function for F, the *total energy* of the particle is the function $E : J \to \mathbb{R}$ defined by

$$E(t) := T(t) + V(x(t)).$$

If γ is a path in U from x_0 to x_1 in U and F is a force field on U, the path integral $\int_\gamma F$ of F along γ is the *work* done in moving a particle along this path.

If $x(t)$ is a position function of some particle with mass m, and F is a force field, *Newton's second law*

$$F = ma$$

asserts that a particle in a force field moves in such a way that the force vector at the location of the particle, at any instant, equals the acceleration vector of the particle times its mass. If we write the law as the equation

$$F(x(t)) = m\ddot{x}(t),$$

and rewrite it in the form

$$\ddot{x}(t) = \frac{1}{m}F(x(t)),$$

we get a *differential equation of second order* i.e., an equation the solution of which is a function and involves the derivatives of this function. From now on, *ode* means ordinary differential equation. The *order* of an ode is the order of the highest derivative that occurs explicitly in it. If we write Newton's second law as

$$F(x(t)) = m\dot{v}(t),$$

where $v(t) = \dot{x}(t)$, we get a first order ode in terms of $x(t)$ and $v(t)$. The term *ordinary* is used to distinguish these equations from differential equations involving partial derivatives of functions, which are called partial differential equations. In the next sections of this chapter we'll study *linear* odes i.e., equations of the form

$$a_0(x)f(x) + a_1(x)f'(x) + \ldots + a_n(x)f^{(n)}(x) + b(x) = 0,$$

where $a_0(x), \ldots, a_n(x)$ and $b(x)$ are differentiable functions. It is easy to see that if we consider the linear ode

$$\sum_{i=0}^{n} a_i(x)f^{(i)}(x) = 0, \tag{1.7}$$

where $f^{(i)}(x)$ denotes the i-th derivative of f at x, and g, h are solutions of equation (1.7), then $\lambda g + \mu h$ are also solutions. Note that Newton's second law, in its

[5]A standard way in physics texts to write the first and the second derivative of $x(t)$ with respect to time (only) is through the symbols $\dot{x}(t)$ and $\ddot{x}(t)$, respectively.

full generality, is a *non-linear* ode and its solutions do not form a vector space. In special cases though, it is reduced to a linear ode.

E.g., if we consider a particle of mass m attached to a wall by means of a spring, and $x : J \to \mathbb{R}$ is its position function, where $0 \in J$, such that $x(t)$ is the displacement of the particle from the equilibrium position $x(0)$, then according to Hooke's law $F(x(t)) = -Kx(t)$, where $K > 0$ is Hooke's constant. If we assume no friction, Newton's second law becomes the linear ode

$$\ddot{x}(t) + p^2 x(t) = 0, \tag{1.8}$$

where $p = \sqrt{\frac{K}{m}}$. The equation (1.8) is the *equation of the harmonic oscillator* in one dimension, that has as solutions the functions

$$x(t) = A\cos(pt) + B\sin(pt), \quad A, B \in \mathbb{R}. \tag{1.9}$$

One can show that (1.9) is the only solution of (1.8) satisfying the *initial conditions*

$$x(0) = A \quad \text{and} \quad \dot{x}(0) = pB.$$

Using the formula $\cos(\phi+\theta) = \cos\phi\cos\theta - \sin\phi\sin\theta$, solution (1.9) takes the form

$$x(t) = a\cos(pt + t_0), \tag{1.10}$$

where

$$a := \sqrt{A^2 + B^2}, \quad \text{and} \quad \cos t_0 = \frac{A}{\sqrt{A^2 + B^2}}.$$

In the proof of Theorem 1.2.20 we will consider the equation

$$\ddot{x}(t) + p^2 x(t) = C, \tag{1.11}$$

where K represents a constant disturbing force, and has solutions of the form

$$x(t) = A\cos(pt) + B\sin(pt) + \frac{C}{p^2}, \quad A, B \in \mathbb{R}, \tag{1.12}$$

which can take the form

$$x(t) = a\cos(pt + t_0) + \frac{C}{p^2}. \tag{1.13}$$

The two-dimensional version of the harmonic oscillator concerns a function $x : J \to \mathbb{R}^2$ and a force field F on $\mathbb{R}^2$ defined by $F(x(t)) = -Kx(t)$, for some $k > 0$. Newton's second law takes again the form

$$\ddot{x}(t) + p^2 x(t) = 0, \tag{1.14}$$

and has solutions of the form

$$x_1(t) = A\cos(pt) + B\sin(pt), \qquad x_2(t) = C\cos(pt) + D\sin(pt). \tag{1.15}$$

THEOREM 1.2.12 (Conservation of energy). *Let $U \subseteq \mathbb{R}^3$ be open. If $x(t)$ is the position function in U of a particle of mass m moving in a conservative force field F on U, then its total energy E is constant.*

Proof. Let $V : U \to \mathbb{R}$ a potential energy function for F. By the definition of a position function of a particle in U and the chain rule on $V \circ x : J \to \mathbb{R}$ we get

$$\begin{aligned}(V \circ x)'(t) &= \big\langle \mathrm{grad}V(x(t)), \dot{x}(t)\big\rangle \\ &= \big\langle -F(x(t)), \dot{x}(t)\big\rangle \\ &= -\big\langle F(x(t)), \dot{x}(t)\big\rangle.\end{aligned}$$

By Remark 1.2.4(ii)(b) and Newton's second law we have that

$$\begin{aligned}E'(t) &= T'(t) + (V \circ x)'(t) \\ &= m\big\langle \ddot{x}(t), \dot{x}(t)\big\rangle - \big\langle F(x(t)), \dot{x}(t)\big\rangle \\ &= \big\langle m\ddot{x}(t), \dot{x}(t)\big\rangle - \big\langle m\ddot{x}(t), \dot{x}(t)\big\rangle \\ &= \big\langle 0, \dot{x}(t)\big\rangle \\ &= 0.\end{aligned}$$

Hence the function E is constant on the interval J. □

The previous proof is independent from the choice of V, since any potential energy function V' for F has the property $\mathrm{grad}V'(x(t)) = -F(x(t))$, for every $t \in \mathbb{R}$.

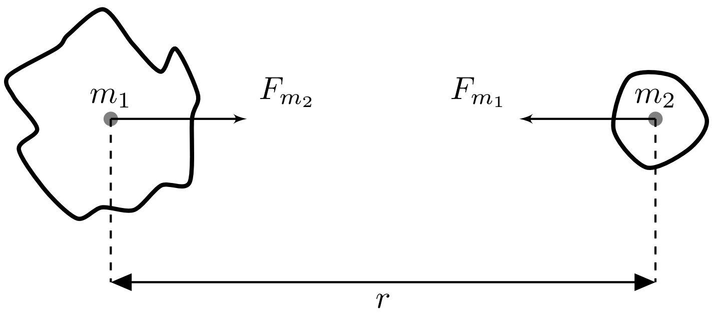
Definition 1.2.13. A force field F on an open subset U of $\mathbb{R}^3$ is called *central*, if there is $\mu : U \to \mathbb{R}$ such that for every $x \in U$

$$F(x) = \mu(x)x.$$

According to Newton's law of gravitation, a body of mass m_1 exerts a force F_{m_1} on a body of mass m_2 such that its magnitude is

$$\frac{gm_1m_2}{r^2},$$

where r is the distance of their centers of gravity and g is a constant, and the direction of F_{m_1} is from m_2 to m_1.



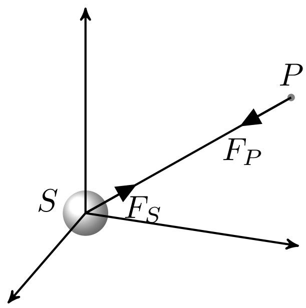
If m_1 is placed at the origin of $\mathbb{R}^3$ and m_2 at $x \in \mathbb{R}^3$, we have that

$$F_{m_1} := \Big(-\frac{gm_1m_2}{|x|^3}\Big)x.$$

The force F_{m_2} of m_2 on m_1 is $-F_{m_1}$. If m_1 is much larger than m_2, and since

$$a_1 = \frac{1}{m_1} F_{m_2} = \left(\frac{g m_2}{|x|^3} \right) x,$$

we may assume that m_1 does not move. In the case of planetary motion, where e.g., the sun has mass m_1 and a much smaller object of mass m_2 is considered, the assumption is natural. If we want to avoid this simplification, we may consider the center of mass of the sun at the origin.



DEFINITION 1.2.14. If we place the sun S at the origin of $\mathbb{R}^3$, the *Newtonian gravitational force field* to a much smaller planet P of mass m placed at

$$x \in \mathbb{R}^3 \setminus \{(0,0,0)\} =: U_0^{(3)}$$

is given by

$$F(x) = \left(-\frac{C}{|x|^3} \right) x.$$

If we use the notation $|U_0^{(3)}| := \{|x| \mid x \in U_0^{(3)}\}$, then

$$F(x) = f(|x|)x,$$

where, $f : |U_0^{(3)}| \to \mathbb{R}$ is defined by $f(t) := -\frac{C}{t^3}$, for every $t \in |U_0^{(3)}| = (0, +\infty)$, and C is the obviously defined constant. Clearly, F is a central force field on $U_0^{(3)}$, and it is conservative, since a simple calculation shows that

$$\left(\frac{C}{|x|^3} \right) x = \text{grad} V(x),$$

where

$$V(x) := -\frac{C}{|x|} = g(|x|),$$

where $g : |U_0^{(3)}| \to \mathbb{R}$ is defined by $g(t) := -\frac{C}{t}$, for every $t \in |U_0^{(3)}|$. As we show next, this situation is standard for conservative force fields that are central.

PROPOSITION 1.2.15. *If F is a conservative force field on $U_0^{(3)}$ and $V : U_0^{(3)} \to \mathbb{R}$ is a potential energy function for F, the following are equivalent:*
(i) F is central.
(ii) There is $f : |U_0^{(3)}| \to \mathbb{R}$ such that for every $x \in U_0^{(3)}$ we have $F(x) = f(|x|)x$.
(iii) There is a function $g : |U_0^{(3)}| \to \mathbb{R}$ such that $g \circ x$ is differentiable, for every position function $x(t)$ on $U_0^{(3)}$, and for every $x \in U_0^{(3)}$ we have $V(x) = g(|x|)$.

PROOF. *(iii)* $\Rightarrow$ *(ii)*: If we see $x \in U_0^{(3)}$ as $x(t)$ for some differentiable position function $x : J \to U_0^{(3)}$, then by the chain rule we have

$$(V \circ x)'(t) = \sum_{i=1}^{3} \frac{\partial V}{\partial x_i}(x(t)) \frac{dx_i}{dt}(t).$$

Moreover,

$$\begin{aligned}
(V \circ x)'(t) &= (g \circ |x|)'(t) \\
&= g'(|x(t)|)|x|'(t) \\
&= g'(|x(t)|)\frac{1}{2}\big(x_1^2(t) + x_2^2(t) + x_3^2(t)\big)^{-\frac{1}{2}}\big(x_1^2(t) + x_2^2(t) + x_3^2(t)\big)' \\
&= g'(|x(t)|)\frac{1}{2}\big(|x(t)|^2\big)^{-\frac{1}{2}}\left(2x_1(t)\frac{dx_1}{dt}(t) + 2x_2(t)\frac{dx_2}{dt}(t) + 2x_3(t)\frac{dx_3}{dt}(t)\right) \\
&= \sum_{i=1}^{3}\left(\frac{g'(|x(t)|)}{|x(t)|}x_i(t)\right)\frac{dx_i}{dt}(t).
\end{aligned}$$

Note that this is well-defined, since $0 \notin U_0^{(3)}$. Hence, for each $i \in \{1, 2, 3\}$ we have

$$\frac{\partial V}{\partial x_i}(x(t)) = \frac{g'(|x(t)|)}{|x(t)|}x_i(t).$$

Since $F(x) = -\mathrm{grad}V(x)$, we get

$$\begin{aligned}
F(x(t)) &= -\left(\frac{\partial V}{\partial x_1}(x(t)), \frac{\partial V}{\partial x_2}(x(t)), \frac{\partial V}{\partial x_3}(x(t))\right) \\
&= -\frac{g'(|x(t)|)}{|x(t)|}\big(x_1(t), x_2(t), x_3(t)\big) \\
&= -\frac{g'(|x(t)|)}{|x(t)|}x(t).
\end{aligned}$$

Hence we define $f : |U_0^{(3)}| \to \mathbb{R}$ by

$$f(|x|) := -\frac{g'(|x|)}{|x|}.$$

$(ii) \Rightarrow (i)$: We define $\mu : U_0^{(3)} \to \mathbb{R}$ by $\mu(x) := f(|x|)$, for every $x \in U_0^{(3)}$.
$(i) \Rightarrow (iii)$: It suffices to show that V is constant on each non-trivial sphere

$$\mathbb{S}_r = \{x \in \mathbb{R}^3 \mid |x| = r\} \subset U_0^{(3)},$$

where $r > 0$. Since then, for every $r > 0$

$$x, y \in \mathbb{S}_r \Rightarrow V(x) = V(y),$$

the function $g : |U_0^{(3)}| \to \mathbb{R}$, defined by $g(|x|) = V(x)$, is well-defined. The rest of the proof is an exercise. □

Next follows a remarkable consequence of the centrality of a force field.

PROPOSITION 1.2.16. *If F is a central force field on an open $U \subseteq U_0^{(3)}$, a particle moving in F moves in a fixed plane.*

PROOF. Let $x : J \to U$ the position function of a particle moving in F. We fix some $t_0 \in J$ and let

$$P_{t_0} = P\big(x(t_0), v(t_0)\big)$$

the unique plane in $\mathbb{R}^3$ containing the position vector of the particle at t_0, the corresponding velocity vector and the origin. Since $F(x) = \mu(x)x$, for some $\mu : U \to \mathbb{R}$, the force vector $F(x(t_0))$ also lies in P_{t_0}. We show that the particle is moving in this plane i.e.,

$$\forall_{t \in \mathbb{R}}\big(x(t) \in P_{t_0}\big).$$

Using the Leibniz product rule for the cross product of $\mathbb{R}^3$-vector-valued differentiable functions u, w on $\mathbb{R}$

$$\frac{d(u \times w)}{dt}(t) = \big(\dot{u}(t) \times w(t)\big) + \big(u(t) \times \dot{w}(t)\big),$$

where $(u \times w)(t) := u(t) \times w(t)$, we have

$$\begin{aligned}\frac{d(x \times \dot{x})}{dt}(t) &= \big(\dot{x}(t) \times \dot{x}(t)\big) + \big(x(t) \times \ddot{x}(t)\big)\\ &= x(t) \times \ddot{x}(t)\\ &= x(t) \times \left[\frac{1}{m}\mu(x(t))\right]x(t)\\ &= 0.\end{aligned}$$

Hence the function $x \times \dot{x}$ is constant, and let $x(t) \times \dot{x}(t) = c \in \mathbb{R}^3$, for every $t \in J$.

If $c \neq 0$, then for every $t \in J$ the vectors $x(t)$ and $\dot{x}(t)$ lie in the plane orthogonal to the vector c, and this is the fixed plane in which the particle moves in. Since c is orthogonal to $x(0)$ and $\dot{x}(0)$, this plane is P_{t_0}

If $c = 0$, the equality $x(t) \times \dot{x}(t) = 0$ implies that there is some $g : J \to \mathbb{R}$ such that for every $t \in J$

$$\dot{x}(t) = g(t)x(t).$$

Hence, $F(x)$ and $v(t)$ are always directed along the line through the origin and the position $x(t)$ of the particle. Actually, in this case the particle *always* moves along the same line through the origin i.e.,

$$\forall_{t \in J}\Big(x(t) \in \big\langle \{x(t_0)\} \big\rangle\Big),$$

hence trivially it moves in P_{t_0}, which is just a line. To show this we work as follows. If $x(t) = (x_1(t), x_2(t), x_3(t))$, then

$$\frac{dx_i}{dt}(t) = g(t)x_i(t),$$

for each $i \in \{1, 2, 3\}$. Since U does not contain the origin, we have

$$\begin{aligned} h(t) &:= \int_{t_0}^{t} g(s)ds \\ &= \int_{t_0}^{t} \frac{1}{x_i(s)} \frac{dx_i}{ds}(s)ds \\ &= \int_{t_0}^{t} \big(\ln x_i(s)\big)' ds \\ &= \ln x_i(t) - \ln x_i(t_0). \end{aligned}$$

Since then $\ln x_i(t) = h(t) + \ln x_i(t_0)$, we get

$$x_i(t) = e^{h(t)} x_i(t_0).$$

Since this is the case for each i, the vector $x(t)$ is a scalar multiple of $x(t_0)$. □

Because of Proposition 1.2.16 we can assume without loss of generality that our central force field of study is defined on an open subset of $U_0^{(2)} := \mathbb{R}^2 \setminus \{(0,0)\}$.

DEFINITION 1.2.17. The *angular momentum* of a moving particle with position function $x : J \to \mathbb{R}^2$ is the function $h : J \to \mathbb{R}$ defined by

$$h(t) := mr^2(t)\frac{d\theta}{dt}(t) =: mr^2\dot{\theta},$$

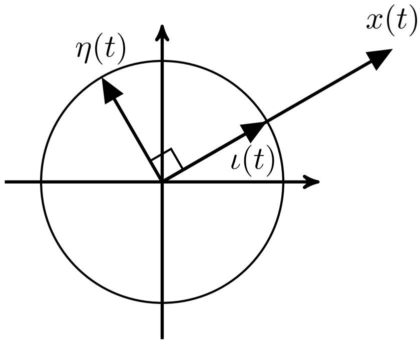
where $(r(t), \theta(t))$ are the polar coordinates of $x(t)$.

THEOREM 1.2.18 (Conservation of angular momentum). *The angular momentum of a particle moving in a central force field on an open $U \subseteq U_0^{(2)}$ is constant.*

PROOF. Let $x : J \to U$ the position function of the particle, and let $\iota(\theta(t))$ be the unit vector in the direction $x(t)$ i.e., for every $t \in J$

$$x(t) = r(t)\iota(\theta(t)).$$

Let $\eta(\theta(t))$ be the unit vector such the angle from $\iota(\theta(t))$ to $\eta(\theta(t))$ is $\frac{\pi}{2}$.



Since $\cos(\theta + \frac{\pi}{2}) = -\sin\theta$ and $\sin(\theta + \frac{\pi}{2}) = \cos\theta$ we have

$$\iota(\theta(t)) := \big(\cos\theta(t), \sin\theta(t)\big), \quad \text{and} \quad \eta(\theta(t)) := \big(-\sin\theta(t), \cos\theta(t)\big),$$

hence taking the derivatives with respect to time we get

$$\dot{\imath} = \eta\dot{\theta}, \quad \text{and} \quad \dot{\eta} = -\iota\dot{\theta}.$$

E.g., for the first equality we have $(\iota \circ \theta)'(t) = \iota'(\theta(t))\theta'(t) = \eta(\theta(t))\theta'(t)$. Hence,

$$\dot{x} = \dot{r}\iota(\theta(t)) + r\eta(\theta(t))\dot{\theta}, \tag{1.16}$$

and since U does not contain the origin we have

$$\begin{aligned}
\ddot{x} &= \ddot{r}\iota(\theta(t)) + \dot{r}\eta(\theta(t))\dot{\theta} + \dot{r}\eta(\theta(t))\dot{\theta} + r(-\iota(\theta(t))\dot{\theta}^2 + r\eta(\theta(t))\ddot{\theta} \\
&= (\ddot{r} - r\dot{\theta}^2)\iota + \big(2\dot{r}\dot{\theta} + r\ddot{\theta}\big)\eta \\
&= (\ddot{r} - r\dot{\theta}^2)\iota + \frac{1}{r}\big((2r\dot{r}\dot{\theta} + r^2\ddot{\theta}\big)\eta \\
&= (\ddot{r} - r\dot{\theta}^2)\iota + \left[\frac{1}{r}\frac{d}{dt}\big(r^2\dot{\theta}\big)\right]\eta.
\end{aligned}$$

Since $\ddot{x} = m^{-1}F(x) = m^{-1}\mu(x)x$, for some $\mu : U \to \mathbb{R}$, the vector $\ddot{x}(t)$ has zero component perpendicular to $x(t)$. Hence

$$\frac{d}{dt}\big(r^2\dot{\theta}\big) = 0,$$

and this implies $\dot{h} = 0$, hence h is constant on J. □

Because of Proposition 1.2.16 we study the motion of a planet in the Newtonian gravitational field (of the sun placed at the origin) on $U_0^{(2)}$, which is defined by

$$F(x) := -\frac{x}{|x|^3},$$

where the constant C in Definition 1.2.14 is avoided with appropriate change of the units. Let $s(t)$ be a *solution curve* of $\ddot{x}(t) = m^{-1}F(x(t))$. By Theorems 1.2.12 and 1.2.18 the total energy E and the angular momentum h are constant at all points of the curve $s(t)$. If $h = 0$, then $\dot{\theta} = 0$, hence θ is constant i.e., the planet moves along a straight line toward or away from the sun. Therefore, we assume that $h \neq 0$. If $s(t) = \big(r(t), \theta(t)\big)$, and since $r^2\dot{\theta}$ is a non-zero constant function of time, the sign of $\dot{\theta}$ is constant along $s(t)$, hence $\theta(t)$ is either an increasing or a decreasing function of time. In order to have constant h along $s(t)$ we need to have r as a function of θ along the curve $s(t)$ i.e., $r = r(\theta)$. We define

$$u(t) := \frac{1}{r(t)}$$

i.e.,

$$u(t) = -V(s(t)),$$

where $V(x) = -\frac{1}{|x|}$. Note that since $r = r(\theta)$, we also get $u = u(\theta)$.

Lemma 1.2.19. *Let $s(t)$ be a solution curve of $\ddot{x}(t) = m^{-1}F(x(t))$, where $F(x)$ is the Newtonian gravitational field on $U_0^{(2)}$, and h is non-zero along $s(t)$.*
(i) The kinetic energy T along $s(t)$ satisfies the following formula:

$$T = \frac{1}{2}\frac{h^2}{m}\left[\left(\frac{du}{d\theta}\right)^2 + u^2\right]. \tag{1.17}$$

(ii) Along $s(t)$ the functions u, θ and E satisfy the following ode:

$$\left(\frac{du}{d\theta}\right)^2 + u^2 = \frac{2m}{h^2}\big(E + u\big). \tag{1.18}$$

(iii) Along $s(t)$ the functions u and θ satisfy the following ode:

$$\frac{d^2u}{d\theta^2} + u = \frac{m}{h^2}. \tag{1.19}$$

Proof. Exercise. □

Theorem 1.2.20. *Let P be a planet moving in the Newtonian gravitational field (of the sun placed at the origin) on $U_0^{(2)}$. If the angular momentum h along a solution curve $s(t)$ of $\ddot{x}(t) = m^{-1}F(x(t))$ is non-zero, then P moves along a conic of eccentricity*

$$\epsilon = \left(1 + \frac{2Eh^2}{m}\right)^{\frac{1}{2}}.$$

Proof. Equation (1.19) has the form of equation (1.11), where $p = 1$ and $C = \frac{m}{h^2}$, hence it has a solution of the form

$$u(\theta) = a\cos(\theta + \theta_0) + \frac{m}{h^2}, \tag{1.20}$$

where $a, \theta_0 \in \mathbb{R}$. Hence

$$\frac{du}{d\theta} = -a\sin(\theta + \theta_0). \tag{1.21}$$

Substituting equations (1.20) and (1.21) in (1.18) we get

$$a = \pm\frac{1}{h^2}\left(2mh^2E + m^2\right)^{\frac{1}{2}}.$$

Hence (1.20) becomes

$$\begin{aligned} u(\theta) &= \pm\frac{1}{h^2}\sqrt{2mh^2E + m^2}\cos(\theta+\theta_0) + \frac{m}{h^2} \\ &= \frac{m}{h^2} \pm \frac{1}{h^2}\sqrt{\frac{2Em^2h^2}{m} + m^2}\ \cos(\theta+\theta_0) \\ &= \frac{m}{h^2} \pm \frac{m}{h^2}\sqrt{\frac{2Eh^2}{m} + 1}\ \cos(\theta+\theta_0) \\ &= \frac{m}{h^2}\left[1 \pm \sqrt{\left(1+\frac{2Eh^2}{m}\right)}\ \cos(\theta+\theta_0)\right]. \end{aligned}$$

Since $\cos(\theta+\theta_0+\pi) = -\cos(\theta+\theta_0)$, and θ_0 is arbitrary, hence it can be written as $\phi+\pi$, we can use only one sign in the last equation. Hence we get

$$u(\theta) = \frac{m}{h^2}\left[1 + \sqrt{\left(1+\frac{2Eh^2}{m}\right)}\ \cos(\theta+\theta_0)\right]. \tag{1.22}$$

If we change the variable θ to $\theta' = \theta - \theta_0$, then

$$u(\theta') = u(\theta-\theta_0) = \frac{m}{h^2}\left[1 + \sqrt{\left(1+\frac{2Eh^2}{m}\right)}\ \cos\theta\right]. \tag{1.23}$$

Since the equation of a conic in polar coordinates is

$$u = \frac{1}{r}, \qquad u = \frac{1}{l}(1+\epsilon\cos\theta),$$

where l is the *latus rectum* and $\epsilon \geq 0$ is the *eccentricity* of the conic, we get

$$l = \frac{h^2}{m}, \qquad \epsilon = \sqrt{\left(1+\frac{2Eh^2}{m}\right)}.$$

□

In the equation of a conic in polar coordinates, if $\epsilon > 1$, then conic is a hyperbola, if $\epsilon = 1$, then conic is a parabola, and if $\epsilon < 1$, then conic is an ellipse. The special case $\epsilon = 0$ corresponds to a circle. Hence, if $E > 0$, the orbit of the planet around the sun is a hyperbola, if $E = 0$, the orbit of the planet around the sun is a parabola, and if $E < 0$, the orbit of the planet is an ellipse.

COROLLARY 1.2.21 (Kepler's first law). *Let P be a planet moving in the Newtonian gravitational field (of the sun placed at the origin) on $U_0^{(2)}$. If the angular momentum h along a solution curve $s(t)$ of $\ddot{x}(t) = m^{-1}F(x(t))$ is non-zero, then P moves along an ellipse.*

PROOF. The quantity $u = \frac{1}{r}$ is always positive. Hence by equation (1.23)

$$\sqrt{\left(1 + \frac{2Eh^2}{m}\right)} \cos\theta > -1.$$

Since for planets like the earth $\cos\theta = -1$ has been observed at least once a year, and since E is constant, we get

$$\sqrt{\left(1 + \frac{2Eh^2}{m}\right)} < 1,$$

which implies $E < 0$. □

While in the planetary model of Copernicus the speed of the planet in orbit around the sun is constant, for Kepler neither the velocity nor the angular velocity is constant, but the *areal velocity* is.

COROLLARY 1.2.22. *If a particle moves in a central force field on some open* $U \subseteq U_0^{(2)}$, *it sweeps out equal areas in equal intervals of time.*

PROOF. Let $A(t)$ be the area swept out by the moving particle $x(t)$ in the time from t_0 to t. In polar coordinates we get $dA = \frac{1}{2}r^2 d\theta$ and we define

$$\dot{A} := \frac{1}{2}r^2\dot{\theta}.$$

By Theorem 1.2.18 we have that $\dot{A}$ is constant. □

In the case of the Newtonian gravitational field Corollary 1.2.22 becomes Kepler's second law.

COROLLARY 1.2.23 (Kepler's second law). *A line segment joining a planet and the sun sweeps out equal areas in equal intervals of time.*

Intuitively, a *state* of a physical system is information characterizing it at a given time. E.g., a state for the harmonic oscillator in one dimension is a pair of vectors $(x(t), v(t))$ and in this case the *space of states* of the harmonic oscillator is the open set $\mathbb{R}^3 \times \mathbb{R}^3$. Since Newton's second law can be written as the ode

$$m\dot{v}(t) = F(x(t)),$$

a solution to it is a curve $s(t) = (x(t), v(t))$ in the state space $\mathbb{R}^3 \times \mathbb{R}^3$ such that

$$\dot{x}(t) = v(t), \quad \text{and} \quad \dot{v}(t) = \frac{1}{m}F(x(t)),$$

for every $t \in J$. Trivially, if $x(t)$ is a solution to the 2nd-order ode of Newton's second law, we get a solution of the 1st-order version of it by setting $v(t) = \dot{x}(t)$. The other direction is also trivial. Moreover, the function

$$A : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3,$$

$$A\big(x(t), v(t)\big) := \big(v(t), \frac{1}{m}F(x(t))\big)$$

is a vector field on the space of states

$$\boldsymbol{S} := \mathbb{R}^3 \times \mathbb{R}^3$$

that *defines* the 1st-order ode of Newton's second law. A solution curve $s(t) = (x(t), v(t))$ describes the evolution of the state of the system in time. We can view the total energy of a particle as the function

$$E : \boldsymbol{S} \to \mathbb{R},$$

$$E(x(t), v(t)) := \frac{1}{2}m|v(t)|^2 + V(x(t)),$$

and when we say that *the total energy is an integral* we mean that the composition

$$\begin{array}{ccc} J & \xrightarrow{\;s\;} & \boldsymbol{S} \\ & {\scriptstyle E \circ s} \searrow & \downarrow {\scriptstyle E} \\ & & \mathbb{R} \end{array}$$

is constant, or E is constant on the solution curve in the state space. According to Theorem 1.2.18, the angular momentum is also an integral for $m\dot{v}(t) = F(x(t))$. In the nineteenth century the solution of an ode was related to the construction of appropriate integrals. This method of integrals, which uses results from basic calculus, does not suffice though, for the solution of more general odes, for the solution of which we need to employ tools and results from more abstract theories.

1.3. The simplest ode, but one of the most important

If $a \in \mathbb{R}$ and $x : J \to \mathbb{R}$ is differentiable, one can show (exercise) that the ode

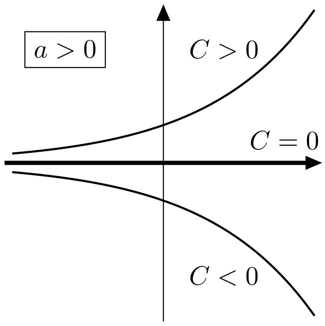
$$\dot{x}(t) = ax(t) \tag{1.24}$$

has as set of solutions the set

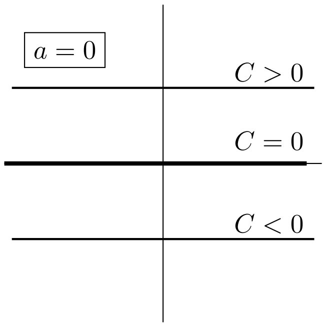
$$\texttt{Solutions}(1.24) = \big\{s : J \to \mathbb{R} \mid \exists_{C\in\mathbb{R}}\forall_{t\in J}\big(s(t) = Ce^{at}\big)\big\}.$$

Equation (1.24) is the simplest ode. If $s \in \texttt{Solutions}(1.24)$, then $s(0) = C$. Conversely, there is a unique function $s \in \texttt{Solutions}(1.24)$ such that $s(0) = C$. This is a special case of the existence of a unique $s \in \texttt{Solutions}(1.24)$ satisfying the initial condition $s(t_0) = s_0$, where $t_0 \in J$.

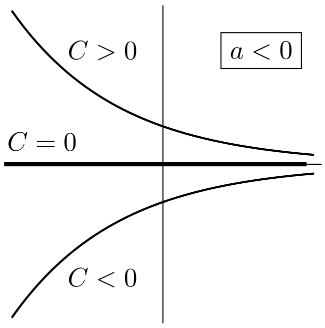
The *parameter* a in (1.24) influences dramatically the way the solution curve s looks like. If $a > 0$, then we have the following three cases:



If $C > 0$, then $\lim_{t \longrightarrow +\infty} Ce^{at} = +\infty$, and if $C < 0$, then $\lim_{t \longrightarrow +\infty} Ce^{at} = -\infty$. If $a = 0$, the solution curves are constant functions



If $a < 0$, we have the following three cases:



In this case, if $C \neq 0$, then

$$\lim_{t \longrightarrow +\infty} Ce^{at} = C \lim_{t \longrightarrow +\infty} e^{-|a|t} = C \lim_{t \longrightarrow +\infty} \frac{1}{e^{|a|t}} = 0.$$

The above graphs reflect the *qualitative behavior* of the solution curves. If $a \neq 0$, equation (1.24) is *stable* in the following sense: If a is replaced by some a' sufficiently close to a, the qualitative behavior of the solution curves does not change. E.g., we have that

$$|a' - a| < |a| \Rightarrow \mathtt{sign}(a') = \mathtt{sign}(a),$$

since, if $a > 0$, then $|a' - a| < a \Leftrightarrow -a < a' - a < a \Rightarrow 0 < a' < 2a$, while, if $a < 0$, then $|a' - a| < -a \Leftrightarrow a < a' - a < -a \Rightarrow 2a < a' < 0$. If $a = 0$, equation (1.24) is *unstable*, since the slightest change in the value of a implies a big change in the qualitative behavior of the solution curves. For this reason we say that $a = 0$ is a *bifurcation point* in the one-parameter family of equations

$$\Big(\dot{x}(t) = ax(t)\Big)_{a\in\mathbb{R}}.$$

Let the following *system* of two odes in two unknown functions:

$$\begin{aligned} \dot{x}_1(t) &= a_1 x_1(t), \\ \dot{x}_2(t) &= a_2 x_2(t). \end{aligned} \tag{1.25}$$

Since there is no relation between $x_1(t)$ and $x_2(t)$, we have that

$$\mathtt{Solutions}(1.25) = \Big\{ s : J \to \mathbb{R}^2 \mid \exists_{C_1, C_2 \in \mathbb{R}} \forall_{t\in J} \Big(s(t) = \big(C_1 e^{a_1 t}, C_2 e^{a_2 t}\big) \Big) \Big\}.$$

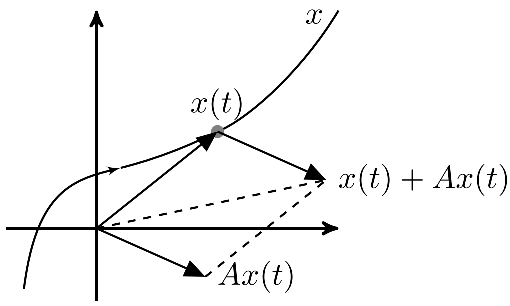
If $s_1(t) = C_1 e^{a_1 t}$ and $s_2(t) = C_2 e^{a_2 t}$, we get $C_1 = s_1(0)$ and $C_2 = s_2(0)$. Equation (1.25) can be written as

$$\dot{x}(t) = Ax(t), \tag{1.26}$$

where

$$A : \mathbb{R}^2 \to \mathbb{R}^2,$$
$$A(x_1, x_2) := (a_1 x_1, a_2, x_2)$$

is a vector field on $\mathbb{R}^2$, which geometrically we understand that it assigns to each vector $x \in \mathbb{R}^2$ the directed line segment from x to $x + Ax$.



We can write equation (1.25) using matrices as follows

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \tag{1.27}$$

A *dynamical system* is a way of describing the passage in time of all states s in the space of states $\boldsymbol{S}$ of a physical system. Here $\boldsymbol{S}$ will be an open subset of $\mathbb{R}^n$, and a dynamical system on $\boldsymbol{S}$ tells us for every $s \in \boldsymbol{S}$ the *history* of s i.e., its future and past positions in time. A dynamical system on $\boldsymbol{S}$ is an appropriately defined[6] function of type

$$\phi : \mathbb{R} \times \boldsymbol{S} \to \boldsymbol{S},$$

such that for every $s \in \boldsymbol{S}$, the function

$$\phi_s : \mathbb{R} \to \boldsymbol{S},$$
$$\phi_s(t) := \phi(t, s)$$

represents the history of the state s.

The ode (1.25) generates a dynamical system. If we consider $\boldsymbol{S} := \mathbb{R}^2$, the dynamical system on $\mathbb{R}^2$ generated by (1.25) is the function

$$\begin{aligned} &\phi : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2 \\ &\phi(t, u) := \big(u_1 e^{a_1 t}, u_2 e^{a_2 t}\big). \end{aligned} \tag{1.28}$$

We can visualize a dynamical system on $\mathbb{R}^2$ as particles placed at each point of $\mathbb{R}^2$ and moving simultaneously, like dust particles under a steady wind. In order to prove some properties of the aforementioned dynamical system on $\mathbb{R}^2$, it is useful to recall the following definitions and facts.

Definition 1.3.1. Let $L(\mathbb{R}^n, \mathbb{R}^m)$ denote the space of (continuous) linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ we define the norm

$$||T|| := \inf\big\{\sigma > 0 \mid \forall_{x \in \mathbb{R}^n}\big(|T(x)| \leq \sigma|x|\big)\big\}.$$

Proposition 1.3.2. *If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, then*

$$\begin{aligned} ||T|| &= \sup\left\{\frac{|T(x)|}{|x|} \mid x \in \mathbb{R}^n \text{ and } |x| > 0\right\} \\ &= \sup\big\{|T(x)| \mid x \in \mathbb{R}^n \text{ and } |x| \leq 1\big\} \\ &= \sup\big\{|T(x)| \mid x \in \mathbb{R}^n \text{ and } |x| = 1\big\}. \end{aligned}$$

Proof. Exercise. □

By Proposition 1.1.18(ii) the Euclidean normed space $(\mathbb{R}^n, |.|)$ is a *Banach space* i.e., a normed space where every Cauchy sequence in it is convergent.

Theorem 1.3.3. *The normed space $(L(\mathbb{R}^n, \mathbb{R}^m), ||.||)$ is a Banach space.*

Proof. Exercise. □

Definition 1.3.4. Let U be an open subset of $\mathbb{R}^n$, $x_0 \in U$ and $f : U \to \mathbb{R}^m$. We say that f is *differentiable at* x_0, if there is a linear map $\lambda_{x_0} : \mathbb{R}^n \to \mathbb{R}^m$ and a function ψ defined for all sufficiently small $h \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \psi(x) = 0,$$

[6]We will define and study dynamical systems later in this course.

and

$$f(x_0 + h) = f(x_0) + \lambda_{x_0}(h) + |h|\psi(h).$$

We say that f is *differentiable on* U, if it is differentiable at every point of U. In that case, the *derivative* f' is a map

$$f' : U \to L(\mathbb{R}^n, \mathbb{R}^m),$$
$$x_0 \mapsto \lambda_{x_0} =: f'(x_0).$$

We say that f is C^1, if f is differentiable on U and the derivative f' is continuous, where the space $L(\mathbb{R}^n, \mathbb{R}^m)$ is equipped with the norm in Definition 1.3.1.

In many cases, to show that some $f : U \to \mathbb{R}^m$ is C^1 we use the following.

PROPOSITION 1.3.5. *Let* U *be an open subset of* $\mathbb{R}^n$*, and* $f : U \to \mathbb{R}^m$*. The following are equivalent.*
(i) f *is* C^1*.*
(ii) The partial derivatives $\frac{\partial f_i}{\partial x_j} : U \to \mathbb{R}$*, where* $i \in \{1, \ldots, m\}$ *and* $j \in \{1, \ldots, n\}$*, exist and are continuous functions.*

PROOF. See [**7**], p.371. □

PROPOSITION 1.3.6. *Let* ϕ *be the dynamical system defined by equation (1.28).*
(i) ϕ *is* C^1*.*
(ii) If $t \in \mathbb{R}$*, the function* $\phi_t : \mathbb{R}^2 \to \mathbb{R}^2$*, defined by*

$$\phi_t(u) := \phi(t, u),$$

for every $u \in \mathbb{R}^2$*, is linear.*
(iii) If $t = 0$*, the function* $\phi_0 : \mathbb{R}^2 \to \mathbb{R}^2$ *is the identity function on* $\mathbb{R}^2$*.*
(iv) If $s, t \in \mathbb{R}$*, then* $\phi_s \circ \phi_t = \phi_{s+t}$*.*

PROOF. Exercise. □

The above result is a special case of a general fact that we will prove later, namely that an arbitrary ode generates a dynamical system ϕ. As we will see later, the converse also holds i.e., a dynamical system ϕ on a state space $\mathcal{S}$ generates an ode by differentiating ϕ_t with respect to time t.

The equations in the system (1.25) are in *uncoupled*, or *diagonal form*, as the matrix corresponding to it is diagonal. Usually, in a system of odes the equations are *coupled*, as e.g., in the system

$$\dot{x}_1(t) = 5x_1(t) + 3x_2(t),$$
$$\dot{x}_2(t) = -6x_1(t) - 4x_2(t). \tag{1.29}$$

In the next section we will explain why we can choose to define

$$y_1(t) = 2x_1(t) + x_2(t),$$
$$y_2(t) = x_1(t) + x_2(t),$$

and hence we get

$$x_1(t) = y_1(t) - y_2(t),\\ x_2(t) = -y_1(t) + 2y_2(t). \tag{1.30}$$

Since

$$\dot{y}_1(t) = 2\dot{x}_1(t) + \dot{x}_2(t),\\ \dot{y}_2(t) = \dot{x}_1(t) + \dot{x}_2(t),$$

we get by 1.30 the system

$$\dot{y}_1(t) = 2y_1(t),\\ \dot{y}_2(t) = -y_2(t),$$

where its equations are in a diagonal form. Hence, if $y(t) = (y_1(t), y_2(t))$ is its solution with initial value $(y_1(0), y_2(0)) = (u_1, u_2)$ i.e.,

$$y_1(t) = u_1 e^{2t},\\ y_2(t) = u_2 e^{-t},$$

we can solve the original system (1.29) by substituting these solutions to the system (1.30). Finally we get

$$x_1(t) = (2u_1 + u_2)e^{2t} - (u_1 + u_2)e^{-t},\\ x_2(t) = -(2u_1 + u_2)e^{2t} + 2(u_1 + u_2)e^{-t}.$$

1.4. Linear systems with constant coefficients & real eigenvalues

If $x_1, \ldots, x_n : J \to \mathbb{R}$ are differentiable functions, and $a_{ij} \in \mathbb{R}$, for every $i, j \in \{1, \ldots, n\}$, the following generalization of the system (1.29) is formed

$$\begin{aligned} \dot{x}_1(t) &= a_{11}x_1(t) + \ldots + a_{1n}x_n(t),\\ &\vdots\\ \dot{x}_i(t) &= a_{i1}x_1(t) + \ldots + a_{in}x_n(t),\\ &\vdots\\ \dot{x}_n(t) &= a_{n1}x_1(t) + \ldots + a_{nn}x_n(t). \end{aligned} \tag{1.31}$$

We can write equation (1.31) using matrices as follows

$$\begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_i(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_i(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \tag{1.32}$$

or, generalizing the simplest ode, we can write it in the form

$$\dot{x}(t) = Ax(t), \tag{1.33}$$

where

$$A := \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} =: [a_{ij}]. \tag{1.34}$$

The right-hand side of equation (1.33) is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$. Next we investigate the use of matrices and linear maps in the study of the system of odes given by equation (1.33). The aim of this section is to prove the fundamental theorem of linear odes with constant coefficients and real eigenvalues.

DEFINITION 1.4.1. The set $L(\mathbb{R}^n, \mathbb{R}^n)$ is denoted by $L(\mathbb{R}^n)$ and an element T of $L(\mathbb{R}^n)$ is called an *operator*. Usually[7], we write Tx instead of $T(x)$. The constant *zero operator* in $L(\mathbb{R}^n)$ is denoted by O_n, and the *identity operator* in $L(\mathbb{R}^n)$ is denoted by I_n. The norm $||.||$ on $L(\mathbb{R}^n)$ defined in Definition 1.3.1 is called the *operator norm*. If $T \in L(\mathbb{R}^n)$ and $m \in \mathbb{N}$, we define

$$T^m := \begin{cases} I_n & , \text{if } m = 0 \\ T \circ T^{m-1} & , \text{if } m > 0. \end{cases}$$

The set of $n \times m$ matrices with entries in $\mathbb{R}$ is denoted by $M_{n,m}(\mathbb{R})$, and the set $M_{n,n}(\mathbb{R})$ is denoted by $M_n(\mathbb{R})$. A *diagonal* matrix in $M_n(\mathbb{R})$ is written as follows

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} =: \mathtt{Diag}(\lambda_1, \dots, \lambda_n).$$

We also denote by I_n the *unit* matrix in $M_n(\mathbb{R})$ i.e.,

$$I_n := \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} =: [\delta_{ij}],$$

where

$$\delta_{ij} := \begin{cases} 1 & , \text{if } i = j \\ 0 & , \text{otherwise} \end{cases}$$

The *zero* matrix in $M_n(\mathbb{R})$ is also denoted by O_n. If $A, B \in M_n(\mathbb{R})$ such that $A = [a_{ij}]$ and $B = [b_{ij}]$, then the algebra operations on $M_n(\mathbb{R})$ are defined as

[7]This is due to Proposition 1.4.3.

follows: $A + B := [c_{ij}]$, $\lambda A := [e_{ij}]$, and $AB := [d_{ij}]$, where $c_{ij} := a_{ij} + b_{ij}$, $e_{ij} := \lambda a_{ij}$ and $d_{ij} := \sum_{k=1}^{n} a_{ik}b_{kj}$. An $A \in M_n(\mathbb{R})$ is called *invertible*, if there is $B \in M_n(\mathbb{R})$ such that $AB = BA = I_n$. A $T \in L(\mathbb{R}^n)$ is called *invertible*, if there is $S \in L(\mathbb{R}^n)$ such that $S \circ T = T \circ S = I_n$. Since B and S are unique, we write $B = A^{-1}$ and $S = T^{-1}$, respectively.

PROPOSITION 1.4.2. *Let* $S, T \in L(\mathbb{R}^n)$, $x \in \mathbb{R}^n$ *and* $m \in \mathbb{N}$.

(i) $|Tx| \leq ||T|||x|$.
(ii) $||S \circ T|| \leq ||S|| \cdot ||T||$.
(iii) $||I_n|| = 1$.
(iv) $||T^m|| \leq ||T||^m$.
(v) If T *is invertible, then* $||T|| \cdot ||T^{-1}|| \geq 1$.

PROOF. Exercise. □

By Proposition 1.4.2(ii)-(iii) and Theorem 1.3.3, $L(\mathbb{R}^n)$ is a *Banach algebra*.

PROPOSITION 1.4.3. *There is a mapping* $\mathcal{T} : M_n(\mathbb{R}) \to L(\mathbb{R}^n)$

$$A \mapsto \mathcal{T}_A := \mathcal{T}(A),$$

where, if $A = [a_{ij}]$, *the mapping* $\mathcal{T}_A : \mathbb{R}^n \to \mathbb{R}^n$ *is defined by*

$$\mathcal{T}_A(x) := \bigg(\sum_{j=1}^{n} a_{1j}x_j, \ldots, \sum_{j=1}^{n} a_{ij}x_j, \ldots, \sum_{j=1}^{n} a_{nj}x_j\bigg),$$

or in matrix form

$$\begin{bmatrix} \mathcal{T}_A(x)_1 \\ \vdots \\ \mathcal{T}_A(x)_i \\ \vdots \\ \mathcal{T}_A(x)_n \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}, \tag{1.35}$$

for every $x \in \mathbb{R}^n$. *There is a function* $\mathcal{A} : L(\mathbb{R}^n) \to M_n(\mathbb{R})$

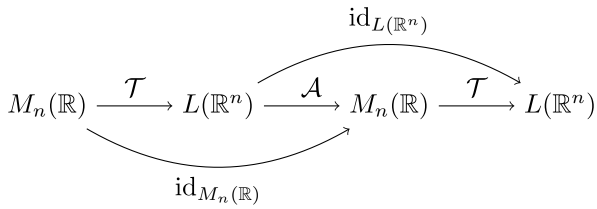
$$T \mapsto \mathcal{A}_T := \mathcal{A}(T),$$

where, if $T \in L(\mathbb{R}^n)$ $\{e_1, \ldots, e_n\}$ *is the standard basis for* $\mathbb{R}^n$, *the matrix* $\mathcal{A}_T \in M_n(\mathbb{R})$ *is defined by*

$$\mathcal{A}_T := \begin{bmatrix} T(e_1)_1 & \ldots & T(e_n)_1 \\ \vdots & \vdots & \vdots \\ T(e_1)_i & \ldots & T(e_n)_i \\ \vdots & \vdots & \vdots \\ T(e_1)_n & \ldots & T(e_n)_n \end{bmatrix} =: \big[T(e_j)_i\big]. \tag{1.36}$$

The mappings $\mathcal{T}$ *and* $\mathcal{A}$ *satisfy the following conditions:*

(i) $\mathcal{A} \circ \mathcal{T} = \mathrm{id}_{M_n(\mathbb{R})}$ *and* $\mathcal{T} \circ \mathcal{A} = \mathrm{id}_{L(\mathbb{R}^n)}$



(ii) $\mathcal{T}_{AB} = \mathcal{T}_A \circ \mathcal{T}_B$.
(iii) $\mathcal{T}_{I_n} = I_n$ *and* $\mathcal{T}_{O_n} = O_n$.
(iv) $\mathcal{T}_{A+B} = \mathcal{T}_A + \mathcal{T}_B$.
(v) $\mathcal{T}_{\lambda A} = \lambda \mathcal{T}_A$.
(vi) If A *is invertible, then* $\mathcal{T}_A$ *is invertible and* $\mathcal{T}_A^{-1} = \mathcal{T}_{A^{-1}}$.
(vii) $\mathcal{A}_{S \circ T} = \mathcal{A}_S \mathcal{A}_T$.
(viii) $\mathcal{A}_{I_n} = I_n$ *and* $\mathcal{A}_{O_n} = O_n$.
(ix) $\mathcal{A}_{S+T} = \mathcal{A}_S + \mathcal{A}_T$.
(x) $\mathcal{T}_{\lambda T} = \lambda \mathcal{T}_A$.
(xi) If T *is invertible, then* $\mathcal{A}_T$ *is invertible and* $\mathcal{A}_T^{-1} = \mathcal{A}_{T^{-1}}$.

PROOF. Left to the reader. □

COROLLARY 1.4.4. *If* $A \in M_n(\mathbb{R})$*, we define*

$$||A|| := ||\mathcal{T}_A||.$$

(i) $||.||$ *is a norm on* $M_n(\mathbb{R})$.
(ii) The mappings $\mathcal{T}$ *and* $\mathcal{A}$ *are norm-preserving.*

PROOF. (i) $||A|| = 0 \Leftrightarrow ||\mathcal{T}_A|| = 0 \Leftrightarrow \mathcal{T}_A = 0 \Leftrightarrow A = 0$, where the implication $\mathcal{T}_A = 0 \Rightarrow A = 0$ is shown as follows: By Proposition 1.4.3(i) we have that $\mathcal{A}_{\mathcal{T}_A} = A$, and by definition $\mathcal{A}_{\mathcal{T}_A} = [\mathcal{T}_A(e_j)_i] = O_n$. The rest properties of the norm follow easily from Proposition 1.4.3(iv)-(v).
(ii) $||\mathcal{A}(T)|| = ||\mathcal{A}_T|| = ||\mathcal{T}_{\mathcal{A}_T}|| = ||T||$. □

According to the equality

$$\big[\mathcal{T}_A(x)\big]_i = \sum_{j=1}^{n} a_{ij} x_j \tag{1.37}$$

the i-row of A expresses the i-coordinate of $\mathcal{T}_A(x)$. Since

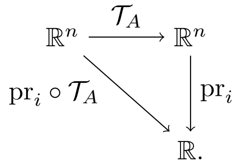
$$\mathcal{T}_A(e_j) = Ae_j = \sum_{i=1}^{n} a_{ij} e_i \tag{1.38}$$

the j-column of A gives the j-coordinate of $T_A(e_j)$. If $x = \sum_{i=1}^{n} x_i e_i$, with respect to the standard basis, the *coordinate functions* $\mathrm{pr}_i : \mathbb{R}^n \to \mathbb{R}$, are defined by

$$x \mapsto x_i := \mathrm{pr}_i(x),$$

and equation (1.37) is written as

$$\mathrm{pr}_i \circ \mathcal{T}_A = \sum_{j=1}^{n} a_{ij} \mathrm{pr}_j.$$



We also have that $Tx = \mathcal{A}_T x$, since

$$\begin{aligned}
\mathcal{A}_T x &= \begin{bmatrix} T(e_1)_1 & \dots & T(e_n)_1 \\ \vdots & \vdots & \vdots \\ T(e_1)_i & \dots & T(e_n)_i \\ \vdots & \vdots & \vdots \\ T(e_1)_n & \dots & T(e_n)_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \\
&= \left(\sum_{i=1}^{n} T(e_i)_1 x_i, \dots, \sum_{i=1}^{n} T(e_i)_n x_i \right) \\
&= \left(T\left(\sum_{i=1}^{n} x_i e_i \right)_1, \dots, T\left(\sum_{i=1}^{n} x_i e_i \right)_n \right) \\
&= \big(T(x)_1, \dots, T(x)_n\big). \\
&= Tx.
\end{aligned}$$

PROPOSITION 1.4.5. *Let $T \in L(\mathbb{R}^n)$ and $\mathcal{B} = \{f_1, \dots, f_n\}$ a basis for $\mathbb{R}^n$. If B is the matrix of $\mathcal{T}$ with respect to $\mathcal{B}$, there is an invertible matrix $Q \in M_n(\mathbb{R})$ such that $B = Q\mathcal{A}_T Q^{-1}$.*

PROOF. If

$$f_i = \sum_{j=1}^{n} p_{ij} e_j,$$

and $P := [p_{ij}]$, it is easy to see that $P^t := [p_{ji}]$, the *transpose* of P, is invertible, and if we define

$$Q = \left[P^t\right]^{-1},$$

the coordinates x_i and y_i of some $z \in \mathbb{R}^n$ with respect to the standard basis and $\mathcal{B}$, respectively, satisfy

$$y = Qx, \quad \text{and} \quad x = Q^{-1}y.$$

The corresponding coordinates $\mathcal{A}_T x$ and By of the image $\mathcal{T}(z)$ satisfy

$$By = Q\mathcal{A}_T x = Q\mathcal{A}_T Q^{-1} y,$$

for every $y \in \mathbb{R}^n$, hence $B = QA_TQ^{-1}$. □

Matrices that are related as B and $\mathcal{A}_T$ are called *similar*, and it is easy to see the converse of Proposition 1.4.5 is also the case. Namely, if two matrices in $M_n(\mathbb{R})$ are similar, they *represent* the same operator with respect to different bases of $\mathbb{R}^n$.

DEFINITION 1.4.6. We call a property P on $M_n(\mathbb{R})$ an *operator property*, if P is preserved under similarity i.e.,

$$P(A) \Rightarrow P(QAQ^{-1}),$$

for every $A \in M_n(\mathbb{R})$ and every invertible $Q \in M_n(\mathbb{R})$.

Note that if P is an operator property on $M_n(\mathbb{R})$, the converse implication $P(QAQ^{-1}) \Rightarrow P(A)$ also holds. Clearly, an operator property on $M_n(\mathbb{R})$ defines a property P on $L(\mathbb{R}^n)$, since its validity is independent from the choice of the matrix representing an operator. Recall that there is a unique mapping

$$\mathtt{Det} : M_n(\mathbb{R}) \to \mathbb{R}$$

satisfying the following conditions:

(D_1) $\mathtt{Det}(AB) = \mathtt{Det}(A)\mathtt{Det}(B)$,
(D_2) $\mathtt{Det}(I_n) = 1$,
(D_3) $\mathtt{Det}(A) \neq 0$ iff A is invertible.

If B is invertible, then it is immediate to see that

(D_4) $\mathtt{Det}(B^{-1}) = \mathtt{Det}(B)^{-1}$,
(D_5) $\mathtt{Det}(BAB^{-1}) = \mathtt{Det}(A)$.

Because of (D_5), the property $P_\lambda(A) := (\mathtt{Det}(A) = \lambda)$ is an operator property, and we can define the *determinant* $\mathtt{Det}(T)$ of an operator T to be the determinant of any matrix representing T.

PROPOSITION 1.4.7. *If $T \in L(\mathbb{R}^n)$, the following are equivalent:*

(i) $\mathtt{Det}(T) \neq 0$.
(ii) $\mathtt{Ker}(T) := \{x \in \mathbb{R}^n \mid Tx = 0\} = \{0\}$.
(iii) T is an injection.
(iv) T is a surjection.
(v) T is invertible.

PROOF. For (i) ⇒ (ii) we use (D_3). The rest is left to the reader. □

Consequently, $\mathtt{Det}(T) = 0$ iff $Tx = 0$, for some $x \neq 0$. The trace $\mathtt{Tr}(A)$ of a matrix $A = [a_{ij}] \in M_n(\mathbb{R})$ is defined by

$$\mathtt{Tr}(A) := \sum_i^n a_{ii},$$

and since

$$\mathtt{Tr}(AB) = \mathtt{Tr}(BA),$$

if $B \in M_n(\mathbb{R})$ is invertible, we get

$$\mathtt{Tr}(BAB^{-1}) = \mathtt{Tr}(B^{-1}BA) = \mathtt{Tr}(A)$$

i.e., $\mathtt{Tr}(A) = \lambda$ is an operator property on $M_n(\mathbb{R})$. Hence we define the *trace* $\mathtt{Tr}(T)$ *of an operator* $T \in L(\mathbb{R}^n)$ to be the trace of any matrix representing T. The correspondence between matrices and operators facilitates also the transfer of concepts from operators to matrices, other than the norm. If $A \in M_n(\mathbb{R})$ the *rank* $\mathtt{Rank}(A)$ of A is defined as the $\mathtt{Rank}(\mathcal{T}_A)$, which is $\dim(\mathtt{Im}(\mathcal{T}_A))$. If $S, T \in L(\mathbb{R}^n)$, we say that they are *similar*, if there is invertible $R \in L(\mathbb{R}^n)$ such that

$$S = R \circ T \circ R^{-1}.$$

By Proposition 1.4.3 we get that if S, T are similar operators, then $\mathcal{A}_S, \mathcal{A}_T$ are similar matrices, and if A, B are similar matrices, then $\mathcal{T}_A, \mathcal{T}_B$ are similar operators. Note that the concept of an operator property on $M_n(\mathbb{R})$ does not have its counterpart for properties on $L(\mathbb{R}^n)$, since the definition of $\mathcal{T}_A$ does not depend on a basis for $\mathbb{R}^n$.

Definition 1.4.8. Let $E_1, \ldots, E_k$ be subspaces of $\mathbb{R}^n$. We say that $\mathbb{R}^n$ is the *direct sum* of $E_1, \ldots, E_k$, if

$$\forall_{x \in \mathbb{R}^n} \exists!_{x_1 \in E_1, \ldots x_k \in E_k} \left(x = \sum_{i=1}^{k} x_i \right).$$

In this case we write

$$\mathbb{R}^n = E_1 \oplus \ldots \oplus E_k =: \bigoplus_{i=1}^{k} E_i.$$

If $T \in L(\mathbb{R}^n)$ and $T_i : E_i \to E_i$ are operators, we say that T is the *direct sum* of $T_1, \ldots, T_k$, if $\mathbb{R}^n = \bigoplus_{i=1}^{k} E_i$ and $Ty = T_i y$, for every $y \in E_i$ and every $i \in \{1, \ldots, k\}$ In this case we write

$$T = T_1 \oplus \ldots \oplus T_k =: \bigoplus_{i=1}^{k} T_i.$$

If A_i is the matrix of T_i with respect to some basis $\mathcal{B}_i$ for E_i, then

$$A := \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix} =: \mathtt{Diag}(A_1, \ldots, A_k)$$

is a matrix of T with respect to the basis

$$\mathcal{B} = \bigcup_{i=1}^{k} \mathcal{B}_i$$

for $\mathbb{R}^n$. We also have that

$$\mathtt{Det}\left(\bigoplus_{i=1}^{k} T_i\right) = \prod_{i=1}^{k} \mathtt{Det}(T_i),$$

since

$$\mathtt{Det}\big(\mathtt{Diag}(A_1, \ldots A_k)\big) = \prod_{i=1}^{k} \mathtt{Det}(A_j),$$

and we have that

$$\mathtt{Tr}\left(\bigoplus_{i=1}^{k} T_i\right) = \sum_{i=1}^{k} \mathtt{Tr}(T_i),$$

since

$$\mathtt{Tr}\big(\mathtt{Diag}(A_1, \ldots A_k)\big) = \sum_{i=1}^{k} \mathtt{Tr}(A_i).$$

Definition 1.4.9. If $T \in L(\mathbb{R}^n)$, a vector $x \in \mathbb{R}^n \setminus \{0\}$ is a (real) *eigenvector* of T, if there is $\lambda \in \mathbb{R}$ such that $Tx = \lambda x$. In this case λ is a *real eigenvalue* of T, and we also say that *x belongs to λ*. The subspace $\mathtt{Ker}(T - \lambda I_n)$ of $\mathbb{R}^n$ is called the *λ-eigenspace* of T. Similar notions are defined for an operator T on a subspace X of $\mathbb{R}^n$, where in this case the λ-eigenspace of $T : X \to X$ is $\mathtt{Ker}(T - \lambda I_X)$, and I_X is the identity on X.

Clearly, λ is an eigenvalue of T iff $\mathtt{Ker}(T - \lambda I_n) \neq \{0\}$, and $\mathtt{Ker}(T - \lambda I_n)$ is the set of all eigenvectors belonging to λ, together with 0. By Proposition 1.4.7

$$\mathtt{Ker}(T - \lambda I_n) \neq \{0\} \Leftrightarrow \mathtt{Det}(T - \lambda I_n) = 0,$$

hence to find the eigenvalues of T we solve the polynomial $p(\lambda)$ generated by the equation

$$\mathtt{Det}(A - \lambda I_n) = 0,$$

where A is any matrix that represents T with respect to some basis for $\mathbb{R}^n$. If B is some other matrix of T, then by Proposition 1.4.5 there is some invertible $Q \in M_n(\mathbb{R})$ such that $B = QAQ^{-1}$, hence by the properties of $\mathtt{Det}$ we get

$$\begin{aligned} \mathtt{Det}(B - \lambda I_n) &= \mathtt{Det}(QAQ^{-1} - \lambda I_n) \\ &= \mathtt{Det}\big(Q(A - \lambda I_n)Q^{-1}\big) \\ &= \mathtt{Det}(Q)\mathtt{Det}(A - \lambda I_n)\mathtt{Det}(Q)^{-1} \\ &= \mathtt{Det}(A - \lambda I_n). \end{aligned}$$

Since $P_\lambda(A) := \mathtt{Det}(A - \lambda I_n) = 0$ is an operator property on $M_n(\mathbb{R})$, we can call $p(\lambda)$ the *characteristic polynomial of T*. A complex root of $p(\lambda)$ is called a *complex eigenvalue* of T. If λ is a real eigenvalue of T and A is a matrix of T, we determine the λ-eigenspace of T by solving the equation

$$(A - \lambda I_n)x = 0.$$

Now we can explain why we chose the new coordinates

$$y_1(t) = 2x_1(t) + x_2(t),$$
$$y_2(t) = x_1(t) + x_2(t),$$

for the solution of the coupled system of odes (1.29). The matrix of this system is

$$A = \begin{bmatrix} 5 & 3 \\ -6 & -4 \end{bmatrix},$$

and the characteristic polynomial of $\mathcal{T}_A$ is $p(\lambda) = (\lambda - 2)(\lambda + 1)$ i.e., its eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = -1$. If we solve the equation $(A - 2I_2)x = 0$, we find that 2-eigenspace of $\mathcal{T}_A$ is the one-dimensional space $\{(t, -t) \mid t \in \mathbb{R}\}$ and let $f_1 = (1, -1)$ form a basis for it. Working similarly, we find that the (-1)-eigenspace of $\mathcal{T}_A$ is the one-dimensional space $\{(t, -2t) \mid t \in \mathbb{R}\}$ and let $f_2 = (-1, 2)$ form a basis for it. From the proof of Proposition 1.4.5 we find that the matrix of $\mathcal{T}_A$ with respect to the basis $\{f_1, f_2\}$ for $\mathbb{R}^2$ is the diagonal matrix

$$B = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix},$$

hence

$$(x_1, x_2) = (y_1 - y_2, -y_1 + 2y_2).$$

Definition 1.4.10. An operator $T \in L(\mathbb{R}^n)$ is called *diagonalizable*, if its matrix with respect to some basis $\mathcal{B} = \{f_1, \dots, f_n\}$ for $\mathbb{R}^n$ is diagonal.

Remark 1.4.11. Let $T \in L(\mathbb{R}^n)$. If $\mathcal{B} = \{f_1, \dots, f_n\}$ is a basis for $\mathbb{R}^n$ such that $f_1, \dots, f_n$ are eigenvectors of T, and if $\lambda_1, \dots, \lambda_n$ are the corresponding eigenvalues of T, then the matrix of T with respect to $\mathcal{B}$ is $\texttt{Diag}(\lambda_1, \dots, \lambda_n)$, hence T is diagonalizable.

Proof. Just note that if $x = (x_1, \dots, x_n)$ with respect to $\mathcal{B}$, then $Tx = (\lambda_1 x_1, \dots, \lambda_n x_n)$ with respect to $\mathcal{B}$. □

Theorem 1.4.12 (Criterion of diagonalizability). *Let $T \in L(\mathbb{R}^n)$. If the characteristic polynomial $p(\lambda)$ of T has n distinct real roots $\lambda_1, \dots, \lambda_n$ and $f_1, \dots, f_n$ are corresponding eigenvectors, then their set $\mathcal{B} = \{f_1, \dots, f_n\}$ is a basis for $\mathbb{R}^n$, and T is diagonalizable.*

Proof. We show that $\mathcal{B}$ is a basis for $\mathbb{R}^n$, hence by Remark 1.4.11 we have that T is diagonalizable. Suppose that $\mathcal{B}$ is not a basis for $\mathbb{R}^n$, and let the elements of $\mathcal{B}$ be ordered such that there is $m < n$ with the property $\{f_1, \dots, f_m\}$ is a maximal independent subset of $\{f_1, \dots, f_n\}$. Clearly, $m \geq 1$, and

$$e_n = \sum_{j=1}^{m} a_j f_j$$

for some $a_1, \ldots, a_m \in \mathbb{R}$. Since f_n belongs to λ_n, we have that

$$\begin{aligned} 0 &= (T - \lambda_n I_n) f_n \\ &= T f_n - \lambda_n f_n \\ &= T\bigg(\sum_{j=1}^{m} a_j f_j\bigg) - \lambda_n \sum_{j=1}^{m} a_j f_j \\ &= \sum_{j=1}^{m} a_j T f_j - \sum_{j=1}^{m} a_j \lambda_n f_j \\ &= \sum_{j=1}^{m} a_j (T f_j - \lambda_n f_j) \\ &= \sum_{j=1}^{m} a_j (\lambda_j f_j - \lambda_n f_j) \\ &= \sum_{j=1}^{m} a_j (\lambda_j - \lambda_n) f_j. \end{aligned}$$

Since the vectors $f_1, \ldots, f_m$ are linearly independent, we get

$$a_j(\lambda_j - \lambda_n) = 0, \quad j \in \{1, \ldots, m\}.$$

Since $\lambda_1, \ldots, \lambda_n$ are distinct, we get

$$a_j = 0, \quad j \in \{1, \ldots, m\},$$

hence $f_n = 0$, which contradicts the hypothesis that f_n is an eigenvector of T. □

Corollary 1.4.13. *If $A \in M_n(\mathbb{R})$ such that $\mathtt{Det}(A - \lambda I_n)$ has n distinct real roots $\lambda_1, \ldots, \lambda_n$, then there exists an invertible $Q \in M_n(\mathbb{R})$ such that*

$$QAQ^{-1} = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n).$$

Proof. By Theorem 1.4.12 there is a basis $\mathcal{B} = \{f_1, \ldots, f_n\}$ for $\mathbb{R}^n$ with $f_1, \ldots, f_n$ eigenvectors that correspond to the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the operator $\mathcal{T}_A$. Since A is the matrix of $\mathcal{T}_A$ with respect to the standard basis for $\mathbb{R}^n$, the matrix B of $\mathcal{T}_A$ with respect to $\mathcal{B}$ is by Proposition 1.4.5 equal to QAQ^{-1}, for some invertible $Q \in M_n(\mathbb{R})$. Moreover $B = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n)$, by Remark 1.4.11. □

Remark 1.4.14. Let $T \in L(\mathbb{R}^2)$ with matrix $A \in M_2(\mathbb{R})$, and let

$$\Delta(A) := \mathtt{Tr}(A)^2 - 4\mathtt{Det}(A).$$

(i) If $\Delta(A) > 0$, then T has two distinct real eigenvalues and it is diagonalizable.
(ii) If $\Delta(A) < 0$, then T has two non-real complex eigenvalues.
(iii) If $\Delta(A) = 0$, then T has two equal real eigenvalues. In this case, every matrix of T is diagonal, or no matrix of T is diagonal.

PROOF. If $A = \mathcal{A}_T$ and

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then

$$\begin{aligned} p_T(\lambda) &= \mathtt{Det}(A - \lambda I_2) \\ &= (a-\lambda)(d-\lambda) - bd \\ &= \lambda^2 - (a+d)\lambda + ad - bc \\ &= \lambda^2 - \mathtt{Tr}(A)\lambda + \mathtt{Det}(A). \end{aligned}$$

Hence,

$$\lambda_{1,2} = \frac{\mathtt{Tr}(A) \pm \sqrt{\Delta(A)}}{2},$$

and (i)-(ii) follow immediately. For case (ii) we work as follows. If T is diagonalizable, then it has a matrix of the form

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix},$$

hence every matrix representing T is diagonal (why?). If T is not diagonalizable, then by definition no matrix of T is diagonal. □

REMARK 1.4.15. If $x_1, \ldots, x_n, y_1, \ldots, y_n : J \to \mathbb{R}$ are differentiable functions and $A \in M_n(\mathbb{R})$ such that $y(t) = Ax(t)$, then $\dot{y}(t) = A\dot{x}(t)$.

PROOF. By hypothesis $y_i(t) = \sum_{j=1}^n a_{ij}x_j(t)$, hence $\dot{y}_i(t) = \sum_{j=1}^n a_{ij}\dot{x}_j(t)$. □

REMARK 1.4.16. If $A = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n)$, for some $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, and $u \in \mathbb{R}^n$, then the system of linear odes

$$\dot{x}(t) = Ax(t) \ ; \quad x(0) = u$$

has a unique solution $x(t) = \big(x_1(t), \ldots, x_n(t)\big)$, where for each $i \in \{1, \ldots, n\}$

$$x_i(t) = u_i e^{\lambda_i t}.$$

PROOF. The system

$$\begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

with initial condition $x(0) = u$ is equivalent to the system of odes $\dot{x}_i(t) = \lambda_i x_i(t)$ with initial condition $x_i(0) = u_i$, for each i, hence we get the above solutions. □

In the previous remark $\lambda_1, \ldots, \lambda_n$ need not be distinct.

THEOREM 1.4.17 (Fundamental theorem of linear odes with constant coefficients and real, distinct eigenvalues). *If $A \in M_n(\mathbb{R})$ with n distinct, real eigenvalues $\lambda_1, \ldots, \lambda_n$, and $u \in \mathbb{R}^n$, then the system of linear odes*

$$\dot{x}(t) = Ax(t) \ ; \quad x(0) = u$$

has a unique solution $x(t) = \big(x_1(t), \ldots, x_n(t)\big)$, where for each $i \in \{1, \ldots, n\}$

$$x_i(t) = \sum_{j=1}^{n} d_{ij} e^{\lambda_j t},$$

for unique constants d_{ij} that depend on u.

PROOF. By Corollary 1.4.13 there exists an invertible $Q \in M_n(\mathbb{R})$ such that

$$QAQ^{-1} = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n).$$

With the following matrix equation we introduce the new coordinates

$$y = Qx, \quad \text{hence} \quad x = Q^{-1}y.$$

By Remark 1.4.15 we have

$$\dot{y}(t) = Q\dot{x}(t) = QAx(t) = QAQ^{-1}y(t),$$

hence

$$\dot{y}(t) = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n)y(t).$$

By Remark 1.4.16 this system together with the initial condition

$$y(0) = Qu$$

has as unique solutions the curve

$$y(t) = \big(y_1(t), \ldots, y_n(t)\big) = \Big((Qu)_1 e^{\lambda_1 t}, \ldots, (Qu)_n e^{\lambda_n t}\Big).$$

We show that the function $x(t) = \big(x_1(t), \ldots, x_n(t)\big)$ defined by

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_i(t) \\ \vdots \\ x_n(t) \end{bmatrix} = Q^{-1} \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}, \tag{1.39}$$

is the unique solution of the initial system. By Remark 1.4.15 we have that

$$\begin{aligned} \dot{x}(t) &= Q^{-1}\dot{y}(t) \\ &= Q^{-1}QAQ^{-1}y(t) \\ &= AQ^{-1}y(t) \\ &= Ax(t). \end{aligned}$$

Moreover,

$$x(0) = Q^{-1}y(0) = Q^{-1}Qu = u.$$

The uniqueness of this solution follows from the uniqueness of the solution of the system $\dot{y} = \texttt{Diag}(\lambda_1, \ldots, \lambda_n)y$ with initial condition $y(0) = Qu$. If $x(t)$ is a solution of $\dot{x}(t) = Ax(t)$ with initial condition $x(0) = u$, then $y(t) = Qx(t)$ is a solution for $\dot{y} = \texttt{Diag}(\lambda_1, \ldots, \lambda_n)y$, since

$$\begin{aligned}\dot{y}(t) &= Q\dot{x}(t)\\ &= QAx(t)\\ &= QAQ^{-1}y(t)\\ &= \texttt{Diag}(\lambda_1, \ldots, \lambda_n)y(t),\end{aligned}$$

and

$$y(0) = Qx(0) = Qu.$$

If $Q^{-1} = [{q_{ij}}']$, equation 1.39 gives us

$$\begin{aligned}x_i(t) &= \sum_{j=1}^{n} {q_{ij}}'(Qu)_j e^{\lambda_j t}\\ &= \sum_{j=1}^{n} d_{ij} e^{\lambda_j t},\end{aligned}$$

where each term

$$d_{ij} := {q_{ij}}'(Qu)_j$$

depends on u. The uniqueness of the terms d_{ij} follows from the fact that the functions $e^{\lambda_1 t}, \ldots, e^{\lambda_n t}$ are linearly independent, since $\lambda_1, \ldots, \lambda_n$ are distinct[8]. □

One can show (exercise) that if $\lambda_1, \ldots, \lambda_n$ are distinct, the solution of the system in Remark 1.4.16 is a special case of the solution of the system in Theorem 1.4.17.

The *direct algorithm* of finding the solution of the system

$$\dot{x}(t) = Ax(t) \; ; \quad x(0) = u$$

that is extracted from the proof of Theorem 1.4.17 is the following:

Step 1: Find the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A i.e., the roots of $\texttt{Det}(A - \lambda I_n)$. This can be difficult.

Step 2: For each eigenvalue λ_i find an eigenvector f_i that belongs to λ_i i.e., solve the system $(A - \lambda_i I_n)f_i = 0$. This is mechanical.

Step 3: Find $P = [p_{ij}]$, by $f_i = \sum_{j=1}^{n} p_{ij}e_j$ and $x = P^t y$, or equivalently

$$x_j = \sum_{i=1}^{n} p_{ij} y_i,$$

---

[8]The proof of this standard fact makes use of the Vandermonde determinant.

for every $j \in \{1, \ldots, n\}$.

Step 4: The system in the new coordinates is $\dot{y}(t) = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n)y(t)$ and

$$y_i(t) = a_i e^{\lambda_i t}, \quad a_i = y_i(0).$$

The general solution to the original system, where $j \in \{1, \ldots, n\}$, is given by

$$x_j(t) = \sum_{i=1}^{n} p_{ij} a_i e^{\lambda_i t}.$$

If one is interested in a specific u, it is easier to solve the equations

$$u_j = \sum_{i=1}^{n} p_{ij} a_j$$

than to invert P^t and solve

$$a = (P^t)^{-1} u.$$

A second algorithm is extracted from the form of solutions and not from the proof of Theorem 1.4.17. We rewrite the equation $\dot{x}(t) = Ax(t)$ as

$$\begin{bmatrix} \sum_{j=1}^{n} \lambda_j d_{1j} e^{\lambda_j t} \\ \vdots \\ \sum_{j=1}^{n} \lambda_j d_{ij} e^{\lambda_j t} \\ \vdots \\ \sum_{j=1}^{n} \lambda_j d_{nj} e^{\lambda_j t} \end{bmatrix} = A \begin{bmatrix} \sum_{j=1}^{n} d_{1j} e^{\lambda_j t} \\ \vdots \\ \sum_{j=1}^{n} d_{ij} e^{\lambda_j t} \\ \vdots \\ \sum_{j=1}^{n} d_{nj} e^{\lambda_j t} \end{bmatrix}, \tag{1.40}$$

and we solve (1.40) with respect to d_{ij}. E.g., using this algorithm the system

$$\begin{aligned} \dot{x}_1(t) &= x_1(t), \\ \dot{x}_2(t) &= x_1(t) + 2x_2(t), \\ \dot{x}_3(t) &= x_1(t) - x_3(t). \end{aligned}$$

with initial condition $x(0) = (1, 0, 0)$ has as solution the curve

$$x(t) = \big(e^t, -e^t + e^{2t}, \frac{1}{2}e^t - \frac{1}{2}e^{-t}\big).$$

Remark 1.4.18. Theorem 1.4.17 doesn't hold if some of the real eigenvalues of A are equal.

Proof. Let

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

with eigenvalues $\lambda_1 = \lambda_2 = 1$, and let the system

$$\begin{aligned} \dot{x}_1(t) &= x_1(t), \\ \dot{x}_2(t) &= x_1(t) + x_2(t) \end{aligned}$$

with $x_1(0) = a$ and $x_2(0) = b$. If $a \neq 0$, then this system cannot be solved according to Theorem 1.4.17. If it could be, then

$$x_1(t) = d_{11}e^t + d_{12}e^t = (d_{11} + d_{12})e^t = ae^t$$
$$x_2(t) = d_{21}e^t + d_{22}e^t = (d_{21} + d_{22})e^t = be^t.$$

But then the second equation of the original system becomes

$$be^t = ae^t + be^t \Leftrightarrow ae^t = 0 \Leftrightarrow a = 0.$$

□

One can show (exercise) that the unique solution to the above system is

$$x(t) = \big(ae^t, e^t(at + b)\big).$$

THEOREM 1.4.19 (Lipschitz continuity of solutions in initial conditions). *Let $A \in M_n(\mathbb{R})$ with n distinct, real eigenvalues. We define the function*

$$\phi_A : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$$

$$\phi_A(t, u) = x(t),$$

where $x(t)$ is the unique solution of the system

$$\dot{x}(t) = Ax(t)\ ; \quad x(0) = u.$$

Let $t \in \mathbb{R}$ be fixed. We define

$$\phi_{A,t} : \mathbb{R}^n \to \mathbb{R}^n$$

$$\phi_{A,t}(u) = \phi_A(t, u).$$

Then there are constants $C \geq 0$ and $k \in \mathbb{R}$ such that for every $u, w \in \mathbb{R}^n$

$$|\phi_{A,t}(u) - \phi_{A,t}(w)| \leq \sigma|u - w|,$$

where

$$\sigma := Ce^{kt}.$$

PROOF. Using the form of solutions in Theorem 1.4.17 (exercise). □

Note that Theorem 2.1.15 implies trivially the continuity of solutions in initial conditions i.e., the property

$$\lim_{u \to u_0} \phi_{A,t}(u) = \phi_{A,t}(u_0),$$

which can be shown (exercise) without using the form of solutions in Theorem 1.4.17.

1.5. Linear systems with constant coefficients & complex eigenvalues

DEFINITION 1.5.1. If $a, b \in \mathbb{R}$ and $b \neq 0$, we define the matrix

$$A_{a,b} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

and $T_{a,b}$ is the operator in $L(\mathbb{R}^2)$ that is represented by $A_{a,b}$.

The eigenvalues of $A_{a,b}$ are $\lambda_1 = a + bi$ and $\lambda_2 = a - bi$ in $\mathbb{C} \setminus \mathbb{R}$.

PROPOSITION 1.5.2. *If $b \neq 0$, the operator $T_{a,b}$ is the composition of a stretching or shrinking and a rotation.*

PROOF. Let $a = r\cos\theta$ and $b = r\sin\theta$, where $r = \sqrt{a^2 + b^2}$. We have that

$$A_{a,b} = \begin{bmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

hence

$$T_{a,b} = r \circ R_\theta,$$

where $R_\theta(x)$ is the θ-counterclockwise rotation of the vector x, and we use for simplicity the symbol r for the mapping $x \mapsto rx$, which is the stretching or shrinking of x by the factor r. □

If we identify $\mathbb{R}^2$ with $\mathbb{C}$, and if $z = x + iy$, then

$$T_{a,b}z = A_{a,b} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax - by \\ bx + ay \end{bmatrix},$$

hence

$$T_{a,b}z = z(a + bi) \tag{1.41}$$

i.e., algebraically speaking, $T_{a,b}$ is multiplication by $a + bi$. The identification between $\mathbb{R}^2$ and $\mathbb{C}$ can be used to solve the system of odes

$$\begin{aligned} \dot{x}(t) &= ax(t) - by(t), \\ \dot{y}(t) &= bx(t) + ay(t), \end{aligned} \tag{1.42}$$

which is also written

$$\dot{z}(t) = A_{a,b}z = T_{a,b}z = (a + bi)z. \tag{1.43}$$

Therefore, for some $C = u + iv$ the solution of (1.43) is

$$z(t) = Ce^{(a+bi)t} = (u + iv)e^{at}e^{ibt}.$$

Since

$$e^{ibt} = \cos(bt) + i\sin(bt),$$

we get

$$\begin{aligned} x(t) &= ue^{at}\cos(bt) - ve^{at}\sin(bt) \\ y(t) &= ue^{at}\sin(bt) + ve^{at}\cos(bt). \end{aligned}$$

In this section we explain how one can reduce different linear systems with constant coefficients and non-real, complex eigenvalues to a system like the above.

DEFINITION 1.5.3. A vector space over $\mathbb{C}$ is called a *complex* vector space. The *complex Cartesian space* $\mathbb{C}^n$ is a complex vector space where

$$(z_1, \dots, z_n) + (w_1, \dots, w_n) := (z_1 + w_1, \dots, z_n + w_n),$$
$$\lambda(z_1, \dots, z_n) := (\lambda z_1, \dots, \lambda z_n) \ ; \quad \lambda \in \mathbb{C}.$$

A subset F of $\mathbb{C}^n$ is a *complex subspace*, if it is closed in $\mathbb{C}^n$ under addition and complex multiplication. We denote the set of operators $T : F \to F$ by $L(F)$. An eigenvalue of $T \in L(F)$ is some $\lambda \in \mathbb{C}$ such that $Tv = \lambda v$, for some $v \in F \setminus \{0\}$. In this case v is an eigenvector of T that "belongs to" λ. If $M_n(\mathbb{C})$ is the set of $n \times n$ matrices with entries in $\mathbb{C}$, an isomorphism between the complex algebras $L(\mathbb{C}^n)$ and $M_n(\mathbb{C})$ can be established, as in the real case. The polynomial with complex coefficients $p_T(\lambda) = \mathtt{Det}(T - \lambda I_F)$ is the characteristic polynomial of T. An operator $T \in L(F)$ is called diagonalizable, if it has a matrix in diagonal form.

Note that an element $g \in \mathbb{C}^n$ can be written as

$$\begin{aligned}\mathbb{C}^n \ni g &= (z_1, \dots, z_n) \\ &= (a_1 + b_1, \dots, a_n + ib_n) \\ &= (a_1, \dots, a_n) + i(b_1, \dots, b_n) \\ &= u + iv, \qquad\qquad u, v \in \mathbb{R}^n.\end{aligned}$$

If $g, g' \in \mathbb{C}^n$ such that $g = u + iv$ and $g' = u' + iv'$, where $u, v, u', v' \in \mathbb{R}^n$, then

$$g = g' \Leftrightarrow u = u' \text{ and } v = v'.$$

THEOREM 1.5.4 (Criterion of diagonalizability). *Let F be a complex subspace of $\mathbb{C}^n$ and $T \in L(F)$. If the characteristic polynomial $p_T(\lambda)$ of T has distinct roots $\lambda_1, \dots, \lambda_m$, where $m = \dim(F)$, and $f_1, \dots, f_m$ are corresponding eigenvectors, then their set $\mathcal{B} = \{f_1, \dots, f_m\}$ is a basis for F, and T is diagonalizable.*

PROOF. Similar to the proof of Theorem 1.4.12. □

DEFINITION 1.5.5. If F is a complex subspace of $\mathbb{C}^n$, the *space of real vectors $F_{\mathbb{R}}$ in F* is defined by

$$F_{\mathbb{R}} := F \cap \mathbb{R}^n.$$

If E is a real subspace of $\mathbb{R}^n$, the *complexification $E_{\mathbb{C}}$ of E* is defined by

$$E_{\mathbb{C}} := \left\{ w \in \mathbb{C}^n \mid w = \sum_{i=1}^{k} \lambda_i w_i, \ k \in \mathbb{N}^+, \ w_1, \dots, w_k \in E, \lambda_1, \dots, \lambda_k \in \mathbb{C} \right\}.$$

We say that F is *decomplexifiable*, if there is E such that $F = E_{\mathbb{C}}$.

REMARK 1.5.6. If F is a complex subspace of $\mathbb{C}^n$ and E is a real subspace of $\mathbb{R}^n$, the following hold:
(i) $F_{\mathbb{R}}$ is a real vector space such that $F_{\mathbb{R}} \subseteq F$.
(ii) $E_{\mathbb{C}}$ is a complex vector space such that $E \subseteq E_{\mathbb{C}}$.
(iii) $(E_{\mathbb{C}})_{\mathbb{R}} = E$.
(iv) $(F_{\mathbb{R}})_{\mathbb{C}} \subseteq F$.

PROOF. Left to the reader. □

DEFINITION 1.5.7. If A is a complex algebra, an *involution* on A is a function ${}^* : A \to A$ that satisfies the following conditions:
(I_1) $(x+y)^* = x^* + y^*$.
(I_2) $(\lambda x)^* = \bar{\lambda} x^*$, where $\bar{\lambda}$ is the conjugate of λ.
(I_3) $(xy)^* = y^* x^*$
(I_4) $(x^*)^* = x$.
The pair $(A, {}^*)$ is called a **-algebra*. The *fixed points* of * is the set $\{a \in A \mid a^* = a\}$. A subspace B of A is called **-invariant*, if $B^* := \{b^* \mid b \in B\} \subseteq B$. If $(A, {}^*)$ and $(B, {}^{\circledast})$ are *-algebras, a function $\varphi : A \to B$ is called **-preserving*, if for every $x \in A$
$$\varphi(x^*) = \varphi(x)^{\circledast}.$$

The conjugate function $z \mapsto \bar{z}$ is an involution on $\mathbb{C}$ with $\mathbb{R}$ as the set of its fixed points. We can also define the function ${}^* : \mathbb{C}^n \to \mathbb{C}^n$ defined by
$$(z_1, \dots, z_n)^* := (\bar{z}_1, \dots, \bar{z}_n)$$
on the vector space $\mathbb{C}^n$, which has $\mathbb{R}^n$ as the set of its fixed points.

PROPOSITION 1.5.8. *A complex subspace F of $\mathbb{C}^n$ is decomplexifiable iff F is *-invariant.*

PROOF. Exercise. □

DEFINITION 1.5.9. Let E be a real subspace of $\mathbb{R}^n$ and $T \in L(E)$. The *complexification* $T_{\mathbb{C}}$ of T is the linear operator
$$T_{\mathbb{C}} : E_{\mathbb{C}} \to E_{\mathbb{C}}$$
defined by
$$T_{\mathbb{C}}(w) = T_{\mathbb{C}}\left(\sum_{i=1}^{k} \lambda_i w_i\right) := \sum_{i=1}^{k} \lambda_i T(w_i).$$
An $S \in L(E_{\mathbb{C}})$ is called *decomplexifiable*, if there is $T \in L(E)$ such that $S = T_{\mathbb{C}}$.

Note that if $u \in E$, then by definition we have that $T_{\mathbb{C}}(u) = T(u)$.

REMARK 1.5.10. Let E be a real subspace of $\mathbb{R}^n$, $\mathcal{B} = \{e_1, \dots, e_m\}$ a basis for E, $T \in L(E)$, and $\lambda \in \mathbb{C}$. The following hold:
(i) $\mathcal{B}$ is a basis for $E_{\mathbb{C}}$.

(ii) The definition of the complexification $T_{\mathbb{C}}$ of T is independent from the choice of representation of $w \in E_{\mathbb{C}}$.
(iii) If $B \in M_m(\mathbb{R})$ is the matrix of T with respect to $\mathcal{B}$ (as a basis for E), then B is the matrix of $T_{\mathbb{C}}$ with respect to $\mathcal{B}$ (as a basis for $E_{\mathbb{C}}$).
(iv) $p_T(\lambda) = p_{T_{\mathbb{C}}}(\lambda)$.
(v) λ is an eigenvalue of T iff λ is an eigenvalue of $T_{\mathbb{C}}$.

PROOF. Exercise. □

PROPOSITION 1.5.11. *2 If E is a real subspace of $\mathbb{R}^n$ and $S \in L(E_{\mathbb{C}})$, then S is decomplexifiable iff S is *-preserving.*

PROOF. Exercise. □

COROLLARY 1.5.12. *Let E be a real vector subspace of $\mathbb{R}^n$, $T \in L(E)$, and $\lambda \in \mathbb{C}$. If λ is an eigenvalue of T, then $\bar{\lambda}$ is an eigenvalue of T.*

PROOF. By Remark 1.5.10(v) λ is an eigenvalue of $T_{\mathbb{C}}$ i.e., there is some non-zero $w \in E_{\mathbb{C}}$ such that $T_{\mathbb{C}}(w) = \lambda w$. Since $T_{\mathbb{C}}$ is trivially decomplexifiable, by Proposition 1.5.8(ii) $T_{\mathbb{C}}$ is *-preserving, hence

$$T_{\mathbb{C}}(w^*) = \big(T_{\mathbb{C}}(w)\big)^* = (\lambda w)^* = \bar{\lambda} w^*,$$

and $\bar{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$ with w^* as a vector in $\mathbb{C}^n$ belonging to $\bar{\lambda}$. By Remark 1.5.10(v) we conclude that $\bar{\lambda}$ is an eigenvalue of T. □

By Corollary 1.5.12 the eigenvalues of some $T \in L(E)$ can be listed as

$$\lambda_1, \ldots, \lambda_k \in \mathbb{R},$$

$$\mu_1, \bar{\mu}_1, \ldots, \mu_l, \bar{\mu}_l \in \mathbb{C} \setminus \mathbb{R}.$$

DEFINITION 1.5.13. Let X be a vector space, $Y, Y_1, \ldots, Y_l$ subspaces of X and $T \in L(X)$. We say that Y is *T-invariant*, if $TY := \{Ty \mid y \in Y\} \subseteq Y$. If X is the direct sum of $Y_1, \ldots, Y_l$, we say that $Y_1, \ldots, Y_l$ form a *T-invariant direct sum decomposition for X*, if Y_j is T-invariant, for every $j \in \{1, \ldots, l\}$.

If $T \in L(X)$, the subspaces X and $\{0\}$ are T-invariant, and every subspace is id_X-invariant. Since $R_0 = \mathrm{id}_{\mathbb{R}^2}$, every subspace of $\mathbb{R}^2$ is R_0-invariant, and since an one-dimensional subspace of $\mathbb{R}^2$ is a line through the origin, every subspace of $\mathbb{R}^2$ is also R_π-invariant. Note also that $R_\pi = -\mathrm{id}_{\mathbb{R}^2}$.

THEOREM 1.5.14 (Direct sum decomposition for an operator with distinct eigenvalues). *Let E be a real vector subspace of $\mathbb{R}^n$ and $T \in L(E)$. If all eigenvalues of T are distinct, then there are subspaces E_r, E_c of E and operators $T_r \in L(E_r), T_c \in L(E_c)$ such that:*

(i) $E = E_r \oplus E_c$;

(ii) $T = T_r \oplus T_c$;

(iii) *T_r has real eigenvalues and T_c has non-real, complex eigenvalues.*

Proof. Let $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$, and $\mu_1, \bar{\mu}_1, \ldots, \mu_l, \bar{\mu}_l \in \mathbb{C} \setminus \mathbb{R}$ be the distinct eigenvalues of T, and let $e_1, \ldots, e_k, d_1, {d_1}', \ldots, d_l, {d_l}'$ the corresponding eigenvectors of T. By Remark 1.5.10(v) these are exactly the eigenvalues of its complexification $T_{\mathbb{C}} \in L(E_{\mathbb{C}})$. By Theorem 1.5.4 the set

$$\mathcal{B} := \{f_1, \ldots, f_k, g_1, g_1^*, \ldots, g_l, g_l^*\}$$

is a basis for $E_{\mathbb{C}}$, where its elements are eigenvectors of $T_{\mathbb{C}}$ that belong to the corresponding eigenvalues of $T_{\mathbb{C}}$. Let

$$F_r := <\{f_1, \ldots, f_k\}>_{\mathbb{C}},$$
$$F_c := <\{g_1, g_1^*, \ldots, g_l, g_l^*\}>_{\mathbb{C}}$$

be the complex linear span of $f_1, \ldots, f_k$ and $g_1, g_1^*, \ldots, g_l, g_l^*$, respectively. The sets F_r, F_c are complex subspaces of $E_{\mathbb{C}}$ that are $T_{\mathbb{C}}$-invariant, since they are generated by eigenvectors of $T_{\mathbb{C}}$. By the definition of $\mathcal{B}$ we get

$$E_{\mathbb{C}} = F_r \oplus F_c.$$

We define the following subspaces of E:

$$E_r := E \cap F_r, \quad \text{and} \quad E_c := E \cap F_c.$$

By Proposition 1.5.8 we have that F_r and F_c are *-invariant, since

$$F_r = (E_r)_{\mathbb{C}}, \quad \text{and} \quad F_c = (E_c)_{\mathbb{C}}$$

i.e., they are decomplexifiable. We show only the first equality, and for the second we work similarly. By the corresponding definitions we get

$$E_r = \left\{ u \in E \mid \exists_{k \in \mathbb{N}^+, \sigma_1, \ldots, \sigma_k \in \mathbb{C}} \left(u = \sum_{i=1}^{k} \sigma_i f_i \right) \right\},$$

$$(E_r)_{\mathbb{C}} = \left\{ w \in \mathbb{C}^n \mid w = \sum_{j=1}^{m} \tau_j u_j, \ m \in \mathbb{N}^+, \ u_1, \ldots, u_m \in E_r, \tau_1, \ldots, \tau_m \in \mathbb{C} \right\}.$$

Clearly, $(E_r)_{\mathbb{C}} \subseteq F_r$. For the converse inclusion it suffices to show that $f_1, \ldots, f_k \in (E_r)_{\mathbb{C}}$. Since $e_1, \ldots, e_k$ belong to $\lambda_1, \ldots, \lambda_k$, and since $E \subseteq E_{\mathbb{C}}$, for every $\nu \in \{1, \ldots, k\}$, we have that

$$e_\nu = \sum_{i=1}^{k} \rho_i f_i + \sum_{j=1}^{l} \sigma_j g_j + \sum_{j=1}^{l} \tau_j g_j^*,$$

for some $\rho_1, \ldots, \rho_k, \sigma_1, \ldots, \sigma_l, \tau_1, \ldots, \tau_l \in \mathbb{C}$. Since $T_{\mathbb{C}}(e_\nu) = T(e_\nu) = \lambda_\nu e_\nu$, we get

$$\begin{aligned} T_{\mathbb{C}}(e_\nu) &= \sum_{i=1}^{k} \rho_i T_{\mathbb{C}} f_i + \sum_{j=1}^{l} \sigma_j T_{\mathbb{C}} g_j + \sum_{j=1}^{l} \tau_j T_{\mathbb{C}} g_j^* \\ &= \sum_{i=1}^{k} \rho_i \lambda_i f_i + \sum_{j=1}^{l} \sigma_j \mu_j g_j + \sum_{j=1}^{l} \tau_j \bar{\mu}_j g_j^* \end{aligned}$$

$$= \lambda_\nu e_\nu$$

$$= \sum_{i=1}^{k} \rho_i \lambda_\nu f_i + \sum_{j=1}^{l} \sigma_j \lambda_\nu g_j + \sum_{j=1}^{l} \tau_j \lambda_\nu g_j^*.$$

Hence, for each $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, l\}$ we have that

$$\rho_i \lambda_i = \rho_i \lambda_\nu \Leftrightarrow \rho_i(\lambda_i - \lambda_\nu) = 0,$$
$$\sigma_j \mu_j = \sigma_j \lambda_\nu \Leftrightarrow \sigma_j(\mu_j - \lambda_\nu) = 0,$$
$$\tau_j \bar{\mu}_j = \tau_j \lambda_\nu \Leftrightarrow \tau_j(\bar{\mu}_j - \lambda_\nu) = 0.$$

Since all eigenvalues are distinct, we get $\rho_i = 0$, if $i \neq \nu$, and $\sigma_j = 0 = \tau_j$, for every $j \in \{1, \ldots, l\}$. Consequently,

$$e_\nu = \rho_\nu f_\nu,$$

for some $\rho_\nu \neq 0$, since e_ν is an eigenvector, and $e_\nu \in E_r$. Hence

$$f_\nu = \frac{1}{\rho_\nu} e_\nu, \quad \frac{1}{\rho_\nu} \in \mathbb{C}, \quad e_\nu \in E_r$$

i.e., $f_\nu \in (E_r)_{\mathbb{C}}$. Since $e_\nu = e_\nu + 0 \in E_r \oplus E_c$, and since similarly we have that $d_j, d_j{}' \in E_r \oplus E_c$, for every $j \in \{1, \ldots, l\}$, we get $E \subseteq E_r \oplus E_c$. The converse inclusion $E_r \oplus E_c \subseteq E$ holds trivially. Hence

$$E = E_r \oplus E_c.$$

We define $T_r \in L(E_r)$ and $T_c \in L(E_c)$ by $T_r := T_{|E_r}$ and $T_c := T_{|E_c}$, respectively. These are well-defined mappings, since if e.g., $u = \sum_{i=1}^{k} \mu_i f_i \in E_r$, then

$$Tu = T_{\mathbb{C}} u$$
$$= T_{\mathbb{C}} \sum_{i=1}^{k} \mu_i f_i$$
$$= \sum_{i=1}^{k} \mu_i T_{\mathbb{C}} f_i$$
$$= \sum_{i=1}^{k} \mu_i \lambda_i f_i \in E_r.$$

Clearly, T_r has real eigenvalues and T_c has non-real, complex eigenvalues. □

Corollary 1.5.15. *Let E be a real vector subspace of $\mathbb{R}^n$ and $T \in L(E)$. If all eigenvalues of T are distinct, then then the system of linear odes*

$$\dot{x}(t) = Tx(t),$$

is rewritten as

$$\dot{x}_r(t) = T_r x_r(t), \quad \dot{x}_c(t) = T_c x_c(t),$$

where $x(t) = x_r(t) + x_c(t) \in E = E_r \oplus E_c$ and $T = T_r \oplus T_c$.

PROOF. By Theorem 1.5.14, we have that let $B_r := \{e_1, \ldots, e_k\}$ and $B_c := \{d_1, d_1', \ldots, d_l, d_l'\}$ are the bases for E_r and E_c, respectively. If A_r is the matrix of T_r with respect to B_r, and if A_c is the matrix of T_c with respect to B_c, then, by the comment following Definition 1.4.8, the matrix of T with respect to $\mathcal{B} = B_r \cup B_c$ is

$$A = \begin{bmatrix} A_r & 0 \\ 0 & A_c \end{bmatrix} = \mathtt{Diag}(A_r, A_c),$$

and the original system is written

$$\begin{bmatrix} \dot{x}_r(t) \\ \dot{x}_c(t) \end{bmatrix} = \begin{bmatrix} A_r & 0 \\ 0 & A_c \end{bmatrix} \begin{bmatrix} x_r(t) \\ x_c(t) \end{bmatrix}.$$

□

Next follows the direct sum decomposition for the operator T_c.

THEOREM 1.5.16 (Direct sum decomposition for an operator with distinct, non-real eigenvalues). *Let E be a real vector subspace of $\mathbb{R}^n$ and $T \in L(E)$. If all eigenvalues of T are the distinct, non-real complex numbers $\mu_1, \bar{\mu}_1, \ldots, \mu_l, \bar{\mu}_l$, there are subspaces $E_1, \ldots E_l$ of E and operators $T_1 \in L(E_1), \ldots, T_l \in L(E_l)$ such that:*

(i) $E_1, \ldots, E_l$ are two-dimensional;

(ii) T_1 has eigenvalues $\mu_1, \bar{\mu}_1$, ..., T_l has eigenvalues $\mu_l, \bar{\mu}_l$;

(iii) $E = E_1 \oplus \ldots \oplus E_l$ is a T-invariant direct sum decomposition for E;

(iv) $T = T_1 \oplus \ldots \oplus T_l$.

PROOF. Let $g_1, g_1^*, \ldots, g_l, g_l^*$ be the corresponding eigenvectors of $T_{\mathbb{C}}$. For every $j \in \{1, \ldots, l\}$ we define the complex subspace

$$F_j :=< \{g_j, g_j^*\} >_{\mathbb{C}},$$

of $E_{\mathbb{C}}$. If $E_j := F_j \cap E$, we work as in the proof of Theorem 1.5.14. □

Because of Theorem 1.5.14 the study of an operator $T \in L(E)$ with distinct eigenvalues is reduced to the study of T_r and T_c. For the operator T_r we use Theorem 1.4.17, while the study of T_c is reduced by Theorem 1.5.16 to the study of an operator $T' \in L(E')$, where E' is a two-dimensional real subspace of $\mathbb{R}^n$ and T' has non-real, complex eigenvalues.

THEOREM 1.5.17. *Let E be a two-dimensional real vector subspace of $\mathbb{R}^n$ and $T \in L(E)$ with eigenvalues $\mu = a+ib$ and $\bar{\mu} = a-ib$, where $b \neq 0$. Then the matrix*

$$A_{ab} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

is the matrix of T with respect to some basis for E.

Proof. The complexification $T_{\mathbb{C}} \in L(E_{\mathbb{C}})$ of T has eigenvectors g, g^* that belong to μ and $\bar{\mu}$, respectively. By the remark following Definition 1.5.3 there are $u, v \in \mathbb{R}^n$ such that $g = u + iv$. Hence $g^* = u - iv$, and

$$u = \frac{1}{2}(g + g^*), \quad v = \frac{1}{2i}(g - g^*) = \frac{i}{2}(g^* - g).$$

The linear independence of g, g^* implies the linear independence of u, v. If $x, y \in \mathbb{R}$,

$$\begin{aligned} xu + yv = 0 &\Rightarrow x\frac{1}{2}(g + g^*) + y\frac{i}{2}(g^* - g) = 0 \\ &\Rightarrow \left(\frac{x}{2} - \frac{yi}{2}\right)g + \left(\frac{x}{2} + \frac{yi}{2}\right)g^* = 0 \\ &\Rightarrow (x - yi) = 0 = (x + yi) \\ &\Leftrightarrow x = 0 = y. \end{aligned}$$

Hence $\mathcal{B} := \{v, u\}$ is a basis for E. If $e = (x, y)$ with respect to $\mathcal{B}$ i.e., $e = xv + yu$, then from the equalities

$$\begin{aligned} T_{\mathbb{C}}g &= \mu g \\ &= (a + ib)(u + iv) \\ &= (au - bv) + i(bu + av), \end{aligned}$$

and

$$T_{\mathbb{C}}g = T_{\mathbb{C}}(u + iv) = T_{\mathbb{C}}u + T_{\mathbb{C}}(iv) = Tu + iTv$$

we get $Tu = au - bv$ and $Tv = bu + av$. Hence

$$\begin{aligned} Te &= T_{\mathbb{C}}e \\ &= T_{\mathbb{C}}(xv + yu) \\ &= xT_{\mathbb{C}}v + yT_{\mathbb{C}}u \\ &= xTv + yTu \\ &= x(bu + av) + y(au - bv) \\ &= (xa - yb)v + (xb + ya)u, \end{aligned}$$

or in matrix form

$$\begin{bmatrix} xa - yb \\ xb + ya \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

□

Corollary 1.5.18. *Let E be a two-dimensional real vector subspace of $\mathbb{R}^n$ and $T \in L(E)$ with eigenvalues $\mu = a + ib$ and $\bar{\mu} = a - ib$, where $b \neq 0$. If g is an eigenvector of the complexification $T_{\mathbb{C}} \in L(E_{\mathbb{C}})$ of T that belongs to μ, such that*

$$g = u + iv, \quad u, v \in \mathbb{R}^n,$$

then $\mathcal{B} := \{v, u\}$ is a basis for E, and the matrix of T with respect to $\mathcal{B}$ is A_{ab}.

Proof. By inspection of the proof of Theorem 1.5.17. □

Note that if we had used as basis for E the set $\mathcal{B}' := \{u, v\}$, then working as above we see that the matrix of T with respect to $\mathcal{B}'$ is $A_{a(-b)}$.

Let for example the following system of odes

$$\dot{x}_1(t) = -2x_2(t),$$
$$\dot{x}_2(t) = x_1(t) + 2x_2(t),$$

with matrix

$$A = \begin{bmatrix} 0 & -2 \\ 1 & 2 \end{bmatrix}.$$

The eigenvalues of A are $\lambda = 1 + i$ and $\bar{\lambda} = 1 - i$. We find a non-real, complex eigenvector $w \in \mathbb{C}^2$ that belongs to λ by solving the equation

$$\big(A - (i + i)I_2\big)w = 0 \Leftrightarrow \begin{bmatrix} -1 - i & -2 \\ 1 & 1 - i \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 0$$
$$\Leftrightarrow (-1 - i)w_1 - 2w_2 = 0 \text{ and } w_1 + (1 - i)w_2 = 0.$$

Since by multiplying the equation $w_1 + (1-i)w_2 = 0$ by $(-1-i)$ we get the equation $(-1 - i)w_1 - 2w_2 = 0$, the two equations are equivalent. Since

$$w_1 = (-1 + i)w_2,$$

we can choose $w_2 = -i$ and $w_1 = 1 + i$. Hence

$$w = (1 + i, -i) = \big(1 + i1, 0 + i(-1)\big) = (1, 0) + i(1, -1) = u + iv,$$
$$u := (1, 0), \quad v := (1, -1).$$

Let $\mathcal{B} := \{v, u\}$ the new basis for $\mathbb{R}^2$. By Corollary 1.5.18 the matrix of A with respect to the new basis is

$$A_{11} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

If $x(t)$ is a solution curve to the system, and if $x(t) = Py(t)$, where $y(t)$ are the coordinates of the solution curve with respect to $\mathcal{B}$, we get $y(t) = P^{-1}x(t)$, therefore

$$\begin{aligned} \dot{y}(t) &= P^{-1}\dot{x}(t) \\ &= P^{-1}Ax(t) \\ &= P^{-1}APy(t) \\ &= A_{11}y(t). \end{aligned}$$

Since, as we already know, the system (1.42) has solutions the curves

$$x(t) = K_1e^{at}\cos(bt) - K_2e^{at}\sin(bt)$$
$$y(t) = K_1e^{at}\sin(bt) + K_1e^{at}\cos(bt),$$

we get

$$y_1(t) = K_1e^t\cos t - K_2e^t\sin t$$
$$y_2(t) = K_1e^t\sin t + K_2e^t\cos t.$$

Since

$$\begin{aligned} x &= (x_1, x_2) \\ &= y_1 v + y_2 u \\ &= y_1(1, -1) + y_2(1, 0) \\ &= (y_1 + y_2, -y_1), \end{aligned}$$

we get

$$\begin{aligned} x_1 &= y_1 + y_2 \\ x_2 &= -y_1. \end{aligned}$$

Hence the solution curve of the original system is

$$\begin{aligned} x_1(t) &= (K_1 + K_2)e^t \cos t + (K_1 - K_2)e^t \sin t, \\ x_2(t) &= -K_1 e^t \cos t + K_2 e^t \sin t. \end{aligned}$$

1.6. Exponentials of operators and homogeneous linear systems

The aim of this section is to solve the system of linear odes

$$\dot{x}(t) = Ax(t),$$

where $A \in M_n(\mathbb{R})$, without supposing that the eigenvalues of A are distinct. In order to do this we use the concept of the exponential of an operators, a generalization of the exponential function on reals. Recall that $\exp : \mathbb{R} \to \mathbb{R}^+$ can be defined through the power series

$$\exp(x) =: e^x := \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

DEFINITION 1.6.1. If $T \in L(\mathbb{R}^n)$, its *exponential operator* $\mathtt{exp}(T)$, or $\mathtt{e}^T$, is defined through its *exponential series* in $L(\mathbb{R}^n)$:

$$\mathtt{exp}(T) =: \mathtt{e}^T := \sum_{k=0}^{\infty} \frac{T^k}{k!}.$$

Recall that the operator T^k, where $k \in \mathbb{N}$, is defined in Definition 1.4.1, and all concepts defined in Definition 1.1.29 extend to a general normed space.

PROPOSITION 1.6.2. *The exponential series of* $\mathtt{e}^T$ *is absolutely convergent.*

PROOF. We show that the series

$$\sum_{k=0}^{\infty} \left|\left| \frac{T^k}{k!} \right|\right|$$

is convergent. By Proposition 1.4.2(iv) we get for every $k \in \mathbb{N}$

$$\left|\left| \frac{T^k}{k!} \right|\right| \leq \frac{||T||^k}{k!},$$

and since

$$\sum_{k=0}^{\infty} \frac{||T||^k}{k!} = e^{||T||},$$

by the comparison test we get the required convergence. □

REMARK 1.6.3. If $T \in L(\mathbb{R}^n)$, then $\mathsf{e}^T \in L(\mathbb{R}^n)$ and

$$||\mathsf{e}^T|| \leq e^{||T||}.$$

PROOF. If $x \in \mathbb{R}^n$, then

$$\mathsf{e}^T(x) = \sum_{k=0}^{\infty} \frac{T^k(x)}{k!},$$

and the linearity of e^T follows immediately from the linearity of each T^k and the properties of infinite series. If $||x|| = 1$, then by Proposition 1.4.2(i) and (iii) $|T^k(x)| \leq ||T^k||\ |x| \leq ||T||^k$, hence

$$|\mathsf{e}^T(x)| = \left| \sum_{k=0}^{\infty} \frac{T^k(x)}{k!} \right| \leq \sum_{k=0}^{\infty} \left| \frac{T^k(x)}{k!} \right| \leq \sum_{k=0}^{\infty} \frac{||T||^k}{k!} = e^{||T||},$$

hence by Proposition 1.3.2 we get $||\mathsf{e}^T|| \leq e^{||T||}$. □

Note that if $(T_n)_{n=0}^{\infty}$ is an absolutely convergent sequence in $L(\mathbb{R}^n)$, then it is also convergent in $L(\mathbb{R}^n)$ i.e.,

$$\sum_{n=0}^{\infty} ||T_n|| < \infty \Longrightarrow \sum_{n=0}^{\infty} T_n \text{ converges in } L(\mathbb{R}^n).$$

If τ_n is the n-th partial sum of the series $\sum_{n=0}^{\infty} T_n$, σ_n is the n-th partial sum of the series $\sum_{n=0}^{\infty} ||T_n||$, and $n > m$, then

$$||\tau_n - \tau_m|| = \left|\left| \sum_{i=m+1}^{n} T_i \right|\right| \leq \sum_{i=m+1}^{n} ||T_i|| = |\sigma_n - \sigma_m|,$$

and we use the fact that $L(\mathbb{R}^n)$ is a Banach space (Theorem 1.3.3). Note that when absolutely convergence of a series in a normed space X implies its convergence in X, then X is a Banach space (left to the reader).

LEMMA 1.6.4. *If $R = \sum_{j=0}^{\infty} R_j$ and $S = \sum_{k=0}^{\infty} S_k$ are absolutely convergent series in $L(\mathbb{R}^n)$, then*

$$R \circ S =: T = \sum_{l=0}^{\infty} T_l,$$

$$T_l := \sum_{j+k=l} R_j \circ S_k.$$

Proof. Let the sequences of the partial sums

$$\rho_n := \sum_{j=0}^{n} R_j, \quad \sigma_n := \sum_{k=0}^{n} S_k, \quad \tau_n := \sum_{l=0}^{n} T_l.$$

We have that

$$R \circ S = \lim_{n\to\infty} (\rho_n \circ \sigma_n) \Leftrightarrow ||R \circ S - (\rho_n \circ \sigma_n)|| \overset{n}{\longrightarrow} 0,$$

since

$$\begin{aligned} ||R \circ S - (\rho_n \circ \sigma_n)|| &= ||R \circ S - \rho_n \circ S + \rho_n \circ S - \rho_n \circ \sigma_n|| \\ &\leq ||R \circ S - \rho_n \circ S|| + ||\rho_n \circ S - \rho_n \circ \sigma_n|| \\ &= ||(R - \rho_n) \circ S|| + ||\rho_n \circ (S - \sigma_n)|| \\ &\leq ||R|| \cdot ||R - \rho_n|| + ||\rho_n|| \cdot ||S - \sigma_n||. \end{aligned}$$

Since $||R - \rho_n|| \overset{n}{\longrightarrow} 0$ and $||S - \sigma_n|| \overset{n}{\longrightarrow} 0$, and since the sequence $(||\rho_n||)_{n=1}^{\infty}$ is bounded (for each $n \in \mathbb{N}$ we have that $||\rho_n|| \leq \sum_{j=0}^{n} ||R_j|| \leq \sum_{j=0}^{\infty} ||R_j|| < \infty$), we conclude that

$$||R \circ S - (\rho_n \circ \sigma_n)|| \overset{n}{\longrightarrow} 0.$$

By hypothesis we have that

$$T = \lim_{n\to\infty} \tau_{2n} \Leftrightarrow ||T - \tau_{2n}|| \overset{n}{\longrightarrow} 0.$$

We also have that

$$\begin{aligned} \rho_n \circ \sigma_n &= \left(\sum_{j=0}^{n} R_j\right) \circ \left(\sum_{k=0}^{n} S_k\right) \\ &= R_0 \circ S_0 + (R_0 \circ S_1 + R_1 \circ S_0) + \ldots + \\ &+ (R_{n-1} \circ S_n + R_n \circ S_{n-1}) + R_n \circ S_n \\ &= \sum_{j+k\leq 2n,\ 0\leq j\leq n, 0\leq k\leq n} R_j \circ S_k. \end{aligned}$$

Since

$$\tau_{2n} = \sum_{l=0}^{2n} \sum_{j+k=l} R_j \circ S_k,$$

we have that

$$\begin{aligned} \tau_{2n} &= \rho_n \circ \sigma_n + \\ &+ \sum_{j+k\leq 2n,\ 0\leq j\leq n, n+1\leq k\leq 2n} R_j \circ S_k \\ &+ \sum_{j+k\leq 2n,\ n+1\leq j\leq 2n, 0\leq k\leq n} R_j \circ S_k. \end{aligned}$$

By the hypothesis of the absolute convergence of the series we get

$$\begin{aligned}||\tau_{2n} - \rho_n \circ \sigma_n|| &= \Bigg|\Bigg| \sum_{j+k\leq 2n,\ 0\leq j\leq n, n+1\leq k\leq 2n} R_j \circ S_k \ + \\ &+ \sum_{j+k\leq 2n,\ n+1\leq j\leq 2n, 0\leq k\leq n} R_j \circ S_k \Bigg|\Bigg| \\ &\leq \sum_{j+k\leq 2n,\ 0\leq j\leq n, n+1\leq k\leq 2n} ||R_j|| \cdot ||S_k|| \ + \\ &+ \sum_{j+k\leq 2n,\ n+1\leq j\leq 2n, 0\leq k\leq n} ||R_j|| \cdot ||S_k|| \\ &\leq \left(\sum_{j=0}^{\infty} ||R_j||\right)\left(\sum_{k=n+1}^{2n} ||S_k||\right) + \\ &+ \ \left(\sum_{k=0}^{\infty} ||S_k||\right)\left(\sum_{j=n+1}^{2n} ||R_j||\right).\end{aligned}$$

Since $\sum_{k=n+1}^{2n} ||S_k|| \xrightarrow{n} 0$ and $\sum_{j=n+1}^{2n} ||R_j|| \xrightarrow{n} 0$, we get that

$$||\tau_{2n} - \rho_n \circ \sigma_n|| \xrightarrow{n} 0.$$

Since

$$\begin{aligned}||R \circ S - \tau_{2n}|| &= ||R \circ S - \rho_n \circ \sigma_n + \rho_n \circ \sigma_n - \tau_{2n}|| \\ &\leq ||R \circ S - \rho_n \circ \sigma_n|| + ||\rho_n \circ \sigma_n - \tau_{2n}||,\end{aligned}$$

we conclude that $||R \circ S - \tau_{2n}|| \xrightarrow{n} 0$. □

REMARK 1.6.5. Let $S, T \in L(\mathbb{R}^n)$, and $(T_n)_{n=1}^{\infty} \subseteq L(\mathbb{R}^n)$, such that $T_n \xrightarrow{n} T$.
(i) $S \circ T_n \xrightarrow{n} S \circ T$.
(ii) $T_n \circ S \xrightarrow{n} T \circ S$.

PROOF. Left to the reader. □

PROPOSITION 1.6.6. *Let $R, S, T \in L(\mathbb{R}^n)$, and $a, b \in \mathbb{R}$.*
(i) If R is invertible, then $\mathsf{e}^{R\circ S\circ R^{-1}} = R \circ \mathsf{e}^S \circ R^{-1}$.
(ii) If $S \circ T = T \circ S$, then $\mathsf{e}^{S+T} = \mathsf{e}^S \circ \mathsf{e}^T$.
(iii) $\mathsf{e}^{-S} = (\mathsf{e}^S)^{-1}$.
(iv) If $n = 2$ and the matrix of T is

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

then the matrix of e^T *is*

$$e^a \begin{bmatrix} \cos b & -\sin b \\ \sin b & \cos b \end{bmatrix}.$$

PROOF. (i) It is easy to show by induction on $\mathbb{N}$ that for every $k \in \mathbb{N}$

$$(R \circ S \circ R^{-1})^k = R \circ S^k \circ R^{-1}.$$

Since

$$R \circ \left(\sum_{k=0}^{n} \frac{S^k}{k!} \right) \circ R^{-1} = \sum_{k=0}^{n} \frac{R \circ S^k \circ R^{-1}}{k!} = \sum_{k=0}^{n} \frac{(R \circ S \circ R^{-1})^k}{k!},$$

by Remark 1.6.5 we have that

$$\begin{aligned} \mathsf{e}^{R \circ S \circ R^{-1}} &= \sum_{k=0}^{\infty} \frac{(R \circ S \circ R^{-1})^k}{k!} \\ &= \lim_{n \to \infty} \sum_{k=0}^{n} \frac{(R \circ S \circ R^{-1})^k}{k!} \\ &= \lim_{n \to \infty} R \circ \left(\sum_{k=0}^{n} \frac{S^k}{k!} \right) \circ R^{-1} \\ &= R \circ \left(\lim_{n \to \infty} \sum_{k=0}^{n} \frac{S^k}{k!} \right) \circ R^{-1} \\ &= R \circ \mathsf{e}^S \circ R^{-1}. \end{aligned}$$

(ii) Using the binomial expansion we get

$$\begin{aligned} (S+T)^n &= \sum_{k=0}^{n} \binom{n}{k} S^{n-k} \circ T^k \\ &= \sum_{k=0}^{n} \frac{n!}{(n-k)!\ k!} S^{n-k} \circ T^k \\ &= n! \sum_{j+k=n} \left(\frac{S^j}{j!} \right) \circ \left(\frac{T^k}{k!} \right). \end{aligned}$$

Hence by Lemma 1.6.4 we get

$$\begin{aligned} \mathsf{e}^{S+T} &= \sum_{n=0}^{\infty} \frac{(S+T)^n}{n!} \\ &= \sum_{n=0}^{\infty} \left(\sum_{j+k=n} \left(\frac{S^j}{j!} \right) \circ \left(\frac{T^k}{k!} \right) \right) \end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{j=0}^{\infty} \frac{S^j}{j!}\right) \circ \left(\sum_{k=0}^{\infty} \frac{T^k}{k!}\right) \\
&= \mathsf{e}^S \circ \mathsf{e}^T.
\end{aligned}
$$

(iii) First we observe that

$$
\mathsf{e}^0 = \sum_{k=0}^{\infty} \frac{0^k}{k!} = I_n + 0^1 + \frac{0^2}{2!} + \ldots = I_n.
$$

Since $S \circ (-S) = (-S) \circ S$, by case (ii) we get that $\mathsf{e}^{S+(-S)} = \mathsf{e}^0 = \mathsf{e}^S \circ \mathsf{e}^{-S}$, and similarly $\mathsf{e}^{-S+S} = \mathsf{e}^0 = \mathsf{e}^{-S} \circ \mathsf{e}^S$.
(iv) If $x_1, x_2 \in \mathbb{R}$, then

$$
\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 - bx_2 \\ bx_1 + ax_2 \end{bmatrix},
$$

hence, identifying $\mathbb{R}^2$ with $\mathbb{C}$ and viewing (x_1, x_2) as $x_1 + ix_2 = z$, we get

$$
Tz = (a + ib)z.
$$

Since for every $k \in \mathbb{N}$ we get then $T^k z = (a + ib)^k z$, we have that

$$
\begin{aligned}
\mathsf{e}^T(z) &= \sum_{k=0}^{\infty} \frac{(a+ib)^k z}{k!} \\
&= z \sum_{k=0}^{\infty} \frac{(a+ib)^k}{k!} \\
&= z e^{a+ib} \\
&= z e^a e^{ib} \\
&= (x_1 + ix_2) e^a (\cos b + i \sin b) \\
&= e^a \big(x_1 \cos b - x_2 \sin b + i(x_2 \cos b + x_1 \sin b)\big),
\end{aligned}
$$

hence using matrices we get

$$
\begin{bmatrix} e^a (x_1 \cos b - x_2 \sin b) \\ e^a (x_2 \cos b + x_1 \sin b) \end{bmatrix} = e^a \begin{bmatrix} \cos b & -\sin b \\ \sin b & \cos b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
$$

□

Proposition 1.6.7. *If $L(\mathbb{R}^n)^{-1}$ is the set of all invertible operators in $L(\mathbb{R}^n)$, the following hold:*
(i) The function `exp` *: $L(\mathbb{R}^n) \to L(\mathbb{R}^n)$, defined by $T \mapsto \mathsf{e}^T$, is a function from $L(\mathbb{R}^n)$ to $\mathrm{L}(\mathbb{R}^n)^{-1}$.*
(ii) The function `exp` *is continuous.*
(iii) If $T \in L(\mathbb{R}^n)$ such that $||T|| < 1$, then
(a) the series $\sum_{k=0}^{\infty} T^k$ converges,

(b) $I_n - T \in \mathrm{L}(\mathbb{R}^n)^{-1}$, *and*

$$\sum_{k=0}^{\infty} T^k = \frac{1}{I_n - T}.$$

(iv) The set $\mathrm{L}(\mathbb{R}^n)^{-1}$ *is an open subset of* $L(\mathbb{R}^n)$.

PROOF. Exercise. □

PROPOSITION 1.6.8. *Let* $T \in L(\mathbb{R}^n)$, $\lambda \in \mathbb{R}$, $x \in \mathbb{R}^n$, *and* E *a subspace of* $\mathbb{R}^n$.
(i) If λ *is an eigenvalue of* T *and* x *is an eigenvector of* T *that belongs to* λ, *then* x *is an eigenvector of* e^T *that belongs to* e^λ.
(ii) If E *is* T*-invariant, then* E *is* e^T*-invariant.*

PROOF. Exercise. □

Note that if $\lambda \in \mathbb{R}$, then

$$\mathsf{e}^{\lambda I_n} = \sum_{k=0}^{\infty} \frac{(\lambda I_n)^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k I_n^k}{k!} = \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) I_n = e^\lambda I_n.$$

PROPOSITION 1.6.9. *If*

$$A = \begin{bmatrix} a & 0 \\ b & a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ b & 0 \end{bmatrix} =: aI_2 + B,$$

then the matrix of $\mathsf{e}^{\mathcal{T}_A}$ *is*

$$e^a \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}.$$

PROOF. Since $aI_2 \cdot B = B \cdot aI_2$, we get $aI_2 \circ \mathcal{T}_B = \mathcal{T}_B \circ aI_2$, hence by Proposition 1.6.6(ii) and the previous remark we have that

$$\mathsf{e}^{\mathcal{T}_A} = \mathsf{e}^{aI_2 + \mathcal{T}_B} = \mathsf{e}^{aI_2} \circ \mathsf{e}^{\mathcal{T}_B} = (e^a I_2) \circ \mathsf{e}^{\mathcal{T}_B} = e^a (I_2 \circ \mathsf{e}^{\mathcal{T}_B}) = e^a \mathsf{e}^{\mathcal{T}_B}.$$

Since

$$\mathcal{T}_B x = \begin{bmatrix} 0 & 0 \\ b & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ bx_1 \end{bmatrix},$$

we have that

$$\mathcal{T}_B(\mathcal{T}_B x) = \begin{bmatrix} 0 & 0 \\ b & 0 \end{bmatrix} \begin{bmatrix} 0 \\ bx_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and similarly $(\mathcal{T}_B)^k = 0$, for every $k > 1$. Hence

$$\mathsf{e}^{\mathcal{T}_B} = \sum_{k=0}^{\infty} \frac{\mathcal{T}_B^k}{k!} = I_2 + \mathcal{T}_B + 0 + +0 \ldots = I_2 + \mathcal{T}_B,$$

and $\mathsf{e}^{\mathcal{T}_A} = e^a(I_2 + \mathcal{T}_B)$. Therefore the matrix of $\mathsf{e}^{\mathcal{T}_A}$ is $e^a(I_2 + B)$. □

PROPOSITION 1.6.10. *Let $S, T \in L(\mathbb{R}^n)$ such that $S \circ T = T \circ S$.*
(i) $\mathsf{e}^S \circ \mathsf{e}^T = \mathsf{e}^T \circ \mathsf{e}^S$.
(ii) $\mathsf{e}^S \circ T = T \circ \mathsf{e}^S$.

PROOF. Exercise. □

DEFINITION 1.6.11. If $T \in L(\mathbb{R}^n)$, the map $\mathtt{exp}_T : \mathbb{R} \to \mathrm{L}(\mathbb{R}^n)^{-1}$ is defined by

$$t \mapsto \mathtt{exp}_T := \mathsf{e}^{tT}.$$

If $A \in M_n(\mathbb{R})$, we write for simplicity $\mathtt{exp}_A(t) = \mathsf{e}^{tA}$ instead of $\mathtt{exp}_{\mathcal{T}_A}(t) = \mathsf{e}^{t\mathcal{T}_A}$.

Since $L(\mathbb{R}^n)$ can be identified with $M_n(\mathbb{R})$, and hence with $\mathbb{R}^{n^2}$, it is meaningful to study the differentiability of $\mathtt{exp}_A$. In the rest we identify $\mathcal{T}_A$ with A.

PROPOSITION 1.6.12. *If $A \in M_n(\mathbb{R})$, the function $\mathtt{exp}_A$ is differentiable and*

$$\dot{\mathtt{exp}}_A(t) = A \circ \mathtt{exp}_A(t) = \mathtt{exp}_A(t) \circ A.$$

PROOF. If $h, t \in \mathbb{R}$, then $tA \circ hA = hA \circ tA$, and Proposition 1.6.6(ii) gives

$$\begin{aligned}
\dot{\mathtt{exp}}_A(t) &= \lim_{h \to 0} \frac{\mathtt{exp}_A(t+h) - \mathtt{exp}_A(t)}{h} \\
&= \lim_{h \to 0} \frac{\mathsf{e}^{(t+h)A} - \mathsf{e}^{tA}}{h} \\
&= \lim_{h \to 0} \frac{\mathsf{e}^{tA} \circ \mathsf{e}^{hA} - \mathsf{e}^{tA}}{h} \\
&= \lim_{h \to 0} \frac{\mathsf{e}^{tA} \circ (\mathsf{e}^{hA} - I_n)}{h} \\
&= \lim_{h \to 0} \left(\mathsf{e}^{tA} \circ \left(\frac{\mathsf{e}^{hA} - I_n}{h} \right) \right) \\
&= \mathsf{e}^{tA} \circ \lim_{h \to 0} \left(\frac{\mathsf{e}^{hA} - I_n}{h} \right) \\
&= \mathsf{e}^{tA} \circ A,
\end{aligned}$$

where the last equality is justified as follows. By definition of e^{hA} we get

$$\begin{aligned}
\frac{\mathsf{e}^{hA} - I_n}{h} &= \frac{\left(I_n + hA + h^2 \frac{A^2}{2} + \dots\right) - I_n}{h} \\
&= \frac{hA + h^2 \frac{A^2}{2} + h^3 \frac{A^3}{3!} + \dots}{h} \\
&= A + h\left(\frac{A^2}{2} + h \frac{A^3}{3!} + \dots \right) \\
&=: A + hB,
\end{aligned}$$

hence by Proposition 1.6.2, and since $|h| \to 0$, we get

$$\begin{aligned}\left|\left|\frac{\mathsf{e}^{hA}-I_n}{h}-A\right|\right| &= ||hB|| \\ &= |h|\ ||B|| \\ &\leq |h|\left(\left|\left|\frac{A^2}{2}\right|\right| + \left|\left|h\frac{A^3}{3!}\right|\right| + \dots\right) \\ &\leq |h|\left(\left|\left|\frac{A^2}{2}\right|\right| + \left|\left|\frac{A^3}{3!}\right|\right| + \dots\right) \\ &\leq |h| \sum_{k=0}^{\infty}\left|\left|\frac{A^k}{k!}\right|\right|.\end{aligned}$$

Since $A \circ (tA) = (tA) \circ A$, by Proposition 1.6.10(ii) $A \circ \mathtt{exp}_A(t) = \mathtt{exp}_A(t) \circ A$. □

THEOREM 1.6.13 (Fundamental theorem of linear odes with constant coefficients). *If $A \in M_n(\mathbb{R})$, the system of linear odes*

$$\dot{x}(t) = Ax(t); \quad x(0) = K \in \mathbb{R}^n$$

has as unique solution the function

$$x(t) = (\mathtt{exp}_A(t))(K) = \mathsf{e}^{tA}K.$$

PROOF. First we show that $x(t)$ is a solution. By Proposition 1.6.12 we get[9]

$$\begin{aligned}\dot{x}(t) &= \lim_{h\to 0}\frac{\mathsf{e}^{(t+h)A}K - \mathsf{e}^{tA}K}{h} \\ &= \lim_{h\to 0}\frac{\mathsf{e}^{(t+h)A} - \mathsf{e}^{tA}}{h}K \\ &= \left[\lim_{h\to 0}\frac{\mathsf{e}^{(t+h)A} - \mathsf{e}^{tA}}{h}\right]K \\ &= \dot{\mathtt{exp}}_A(t)K \\ &= (A \circ \mathtt{exp}_A(t))K \\ &= A\mathsf{e}^{tA}K \\ &= Ax(t).\end{aligned}$$

Moreover, $x(t)$ satisfies the given initial condition, since

$$x(0) = \mathsf{e}^{0A}K = \mathsf{e}^0K = I_nK = K.$$

For the uniqueness of the solution of the system we work as in the case of the proof of uniqueness of solution to the simplest ode. If $x(t)$ is a solution of the system and

[9]We freely pass from an expression like $(A \circ \mathtt{exp}_A(t))K$, which is understood as a formula between operators, to an expression like $A\mathsf{e}^{tA}K$, which is understood as a formula between matrices.

$y(t) = (\mathtt{exp}_A(-t))(x(t)) = \mathtt{e}^{-tA}x(t)$, then by Proposition 1.6.12 we have that

$$\begin{aligned}\dot{y}(t) &= \left(\frac{d}{dt}\mathtt{e}^{-tA}\right)x(t) + \mathtt{e}^{-tA}\dot{x}(t)\\ &= -A\mathtt{e}^{-tA}x(t) + \mathtt{e}^{-tA}Ax(t)\\ &= \mathtt{e}^{-tA}(-A + A)x(t)\\ &= 0,\end{aligned}$$

hence $y(t)$ is constant with value $y(0) = \mathtt{e}^0 x(0) = I_n x(0) = x(0) = K$. Hence $K = \mathtt{e}^{-tA}x(t)$, therefore $\mathtt{e}^{tA}K = \mathtt{e}^{tA}\mathtt{e}^{-tA}x(t) = \mathtt{e}^0 x(t) = I_n x(t) = x(t)$. □

Note that if $n = 1$, the general solution to the system

$$\dot{x}(t) = Ax(t); \quad x(0) = K \in \mathbb{R}$$

is $x(t) = \mathtt{e}^{ta}K$, and since

$$\mathtt{e}^{ta} = e^{ta},$$

we get the known unique solution of the simplest ode. If we consider the system

$$\begin{aligned}\dot{x}_1(t) &= ax_1(t); & x_1(0) &= K_1 \in \mathbb{R},\\ \dot{x}_2(t) &= bx_1(t) + ax_2(t); & x_2(0) &= K_2 \in \mathbb{R},\end{aligned} \tag{1.44}$$

with matrix

$$A = \begin{bmatrix} a & 0 \\ b & a \end{bmatrix},$$

then by Proposition 1.6.9 we know that

$$tA = \begin{bmatrix} ta & 0 \\ tb & ta \end{bmatrix} \Longrightarrow \mathtt{e}^{tA} = e^{ta}\begin{bmatrix} 1 & 0 \\ tb & 1 \end{bmatrix}.$$

By Theorem 1.6.13 the unique solution of the system is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = e^{ta}\begin{bmatrix} 1 & 0 \\ tb & 1 \end{bmatrix}\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} e^{ta}K_1 \\ e^{ta}(tbK_1 + K_2) \end{bmatrix}.$$

If $A \in M_n(\mathbb{R})$, the dynamical system that is generated by the system of odes

$$\dot{x}(t) = Ax(t)$$

is the function $\phi_A : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$

$$\phi_A(t, u) := x(t),$$

where $x(t)$ is the unique solution of the system

$$\dot{x}(t) = Ax(t)\ ; \quad x(0) = u \in \mathbb{R}^n.$$

By Theorem 1.6.13 we get

$$\phi_A(t, u) = \mathtt{e}^{tA}u.$$

Let $t \in \mathbb{R}$ be fixed. The function

$$\phi_{A,t} : \mathbb{R}^n \to \mathbb{R}^n$$

$$\phi_{A,t}(u) := \phi_A(t,u) = \mathsf{e}^{tA}u$$

is linear. The family of maps

$$(\phi_t)_{t\in\mathbb{R}}$$

is called the *flow* that corresponds to the above system of odes. This flow is *linear*, as the maps ϕ_t are linear, for every $t \in \mathbb{R}$. If $s,t \in \mathbb{R}$, the flow satisfies the fundamental property

$$\phi_{A,s} \circ \phi_{A,t} = \phi_{A,s+t},$$

since for every $u \in \mathbb{R}^n$ we have that

$$\begin{aligned}(\phi_{A,s} \circ \phi_{A,t})(u) &= \phi_{A,s}(\phi_{A,t}u)\\ &= \phi_{A,s}\big(\mathsf{e}^{tA}u\big)\\ &= \mathsf{e}^{sA}\mathsf{e}^{tA}u\\ &= \mathsf{e}^{(s+t)A}u\\ &= \phi_{A,s+t}(u).\end{aligned}$$

The Lipschitz continuity of solutions in initial conditions (see Theorem 2.1.15) follows in this case easily, since

$$\begin{aligned}|\phi_{A,t}(u) - \phi_{A,t}(w)| &= |\mathsf{e}^{tA}u - \mathsf{e}^{tA}w|\\ &= |\mathsf{e}^{tA}(u-w)|\\ &\leq ||\mathsf{e}^{tA}|| \cdot |u-w|\\ &\leq e^{||tA||} \cdot |u-w|\\ &= e^{|t|\cdot||A||} \cdot |u-w|.\end{aligned}$$

If $A \in M_2(\mathbb{R})$, one can show that there is invertible $P \in M_2(\mathbb{R})$ such that $B = PAP^{-1}$ has one of the following forms:

$$\begin{bmatrix}\lambda & 0\\ 0 & \mu\end{bmatrix}; \qquad \begin{bmatrix}a & -b\\ b & a\end{bmatrix}; \qquad \begin{bmatrix}\lambda & 0\\ 1 & \lambda\end{bmatrix}.$$

Correspondingly, the exponential e^B has one of the following forms:

$$\begin{bmatrix}e^{\lambda} & 0\\ 0 & e^{\mu}\end{bmatrix}; \qquad e^a\begin{bmatrix}\cos b & -\sin b\\ \sin b & \cos b\end{bmatrix}; \qquad e^{\lambda}\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}.$$

The firs case is an exercise, while the third follows from the solution of the system (1.44) for $t = 1 = b$. By Proposition 1.6.6(i) we get

$$\mathsf{e}^A = \mathsf{e}^{P^{-1}BP} = P^{-1}\mathsf{e}^B P$$

i.e., we can can compute e^A, for every $A \in M_2(\mathbb{R})$. Consequently, we can *explicitly* solve the system $\dot{x}(t) = Ax(t)$, for every $A \in M_2(\mathbb{R})$. We consider the following cases:

(I) A has eigenvalues $\lambda, \mu \in \mathbb{R}$ such that $\lambda \cdot \mu < 0$ (*saddle*): By Corollary 1.4.13

$$B = \mathtt{Diag}(\lambda, \mu) = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}.$$

(II) All eigenvalues have negative real part (*sink*): one can show that every solution $x(t)$ of the corresponding system satisfies

$$\lim_{t \to \infty} x(t) = 0.$$

E.g., if $\lambda = a + ib$ and $\mu = a - ib$ and $a < 0$, by Corollary 1.5.18 after changing the system of coordinates we get the equivalent system $\dot{y}(t) = By(t)$, where $B = A_{ab}$. Since

$$\mathtt{e}^{tB} = e^{ta} \begin{bmatrix} \cos(tb) & -\sin(tb) \\ \sin(tb) & \cos(tb) \end{bmatrix},$$

the solutions are

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = e^{ta} \begin{bmatrix} \cos(tb) & -\sin(tb) \\ \sin(tb) & \cos(tb) \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} e^{ta}\big(K_1 \cos(tb) - K_2 \sin(tb)\big) \\ e^{ta}\big(K_1 \sin(tb) + K_2 \cos(tb)\big) \end{bmatrix}.$$

Since $|\cos(tb)| \leq 1$ and $|\sin(tb)| \leq 1$, and since $a < 0$, we get $\lim_{t \to \infty} y(t) = 0$, and since $x(t) = Py(t)$, we conclude that $\lim_{t \to \infty} x(t) = 0$.

(III) All eigenvalues have positive real part (*source*): one can show as in case (II) that every solution $x(t)$ of the corresponding system satisfies

$$\lim_{t \to \infty} |x(t)| = \infty, \qquad \lim_{t \to -\infty} |x(t)| = 0.$$

(IV) All eigenvalues are pure imaginary (*center*): one can show (exercise) that all solutions are *periodic* with the same period i.e., there is some $p > 0$ such that

$$\forall_{t \in \mathbb{R}} \big(x(t + p) = x(t)\big).$$

1.7. Variation of constants

Definition 1.7.1. If $A \in M_n(\mathbb{R})$ and $B : \mathbb{R} \to \mathbb{R}^n$ is continuous, the system of odes

$$\dot{x}(t) = Ax(t) + B(t) \tag{1.45}$$

is called a *non-homogeneous, non-autonomous* system of odes.

Equation (1.45) is called non-homogeneous because the term $B(t)$ prevents it from being linear, and it is called non-autonomous, since $\dot{x}(t)$ depends explicitly on the time parameter t.

Theorem 1.7.2. (*i*) *Equation (1.45) has as a solution the function*

$$x(t) = \mathtt{e}^{tA} \left[\int_0^t \mathtt{e}^{-sA} B(s) ds + K \right], \qquad K \in \mathbb{R}^n, \tag{1.46}$$

and every solution of equation (1.45) is of this form.

(ii) A solution of equation (1.45) has the form

$$x(t) = u(t) + v(t),$$

where $u(t)$ is a solution of equation (1.45) and $v(t)$ is a solution of the homogeneous equation $\dot{x}(t) = Ax(t)$.

(iii) The sum of a solution of equation (1.45) and of the homogeneous equation $\dot{x}(t) = Ax(t)$ is a solution to equation (1.45).

Proof. (i) We suppose that the solution of equation (1.45) has the form

$$x(t) = \mathsf{e}^{tA} f(t), \tag{1.47}$$

for some differentiable function $f : \mathbb{R} \to \mathbb{R}^n$, and we determine the exact form of $f(t)$. Note that if $B(t) = 0$, for every $t \in \mathbb{R}$, then by Theorem 1.6.13 $f(t) = K$, for every $T \in \mathbb{R}$ and for some $K \in \mathbb{R}^n$ (that is why this method of solution of equation (1.45) is called variation of constants). By Proposition 1.6.12 we get

$$\begin{aligned} Ax(t) + B(t) &= \dot{x}(t) \\ &= (\mathsf{e}^{tA})' f(t) + \mathsf{e}^{tA} f'(t) \\ &= \big(A\mathsf{e}^{tA}\big) f(t) + \mathsf{e}^{tA} f'(t) \\ &= A\big(\mathsf{e}^{tA} f(t)\big) + \mathsf{e}^{tA} f'(t) \\ &= Ax(t) + \mathsf{e}^{tA} f'(t), \end{aligned}$$

hence

$$f'(t) = \mathsf{e}^{-tA} B(t).$$

By integration we get

$$f(t) = \int_0^t \mathsf{e}^{-sA} B(s) ds + K,$$

for some $K \in \mathbb{R}^n$. Note that the function $g : \mathbb{R} \to \mathbb{R}^n$, defined by

$$g(s) := \mathsf{e}^{-sA} B(s)$$

is continuous, hence it is integrable, and

$$\int_0^t g(s) ds = \bigg(\int_0^t g_1(s) ds, \ldots, \int_0^t g_n(s) ds \bigg) \in \mathbb{R}^n.$$

First we show that equation (1.47) is indeed a solution to equation (1.45):

$$\begin{aligned} \dot{x}(t) &= \big(\mathsf{e}^{tA}\big)' \bigg[\int_0^t \mathsf{e}^{-sA} B(s) ds + K \bigg] + \mathsf{e}^{tA} \bigg[\int_0^t \mathsf{e}^{-sA} B(s) ds + K \bigg]' \\ &= A\mathsf{e}^{tA} \bigg[\int_0^t \mathsf{e}^{-sA} B(s) ds + K \bigg] + \mathsf{e}^{tA} \mathsf{e}^{-tA} B(t) \\ &= Ax(t) + B(t). \end{aligned}$$

Next we show that a solution $y : \mathbb{R} \to \mathbb{R}^n$ of equation (1.45) is of this form. Since $\dot{y}(t) = Ay(t) + B(t)$, we get

$$(x \dot{-} y)(t) = \dot{x}(t) - \dot{y}(t) = A\big(x(t) - y(t)\big),$$

hence by Theorem 1.6.13 there is some $\Lambda \in \mathbb{R}^n$ such that $x(t) - y(t) = \mathsf{e}^{tA}\Lambda$, hence

$$\begin{aligned} y(t) &= x(t) - \mathsf{e}^{tA}\Lambda \\ &= \mathsf{e}^{tA}\left[\int_0^t \mathsf{e}^{-sA}B(s)ds + K\right] - \mathsf{e}^{tA}\Lambda \\ &= \mathsf{e}^{tA}\left[\int_0^t \mathsf{e}^{-sA}B(s)ds + (K - \Lambda)\right] \\ &= \mathsf{e}^{tA}\left[\int_0^t \mathsf{e}^{-sA}B(s)ds + K'\right], \end{aligned}$$

where $K' := K - \Lambda \in \mathbb{R}^n$.
(ii) The general solution of equation (1.45) is written as

$$x(t) = u(t) + \mathsf{e}^{tA}K,$$

where

$$u(t) := \mathsf{e}^{tA}\int_0^t \mathsf{e}^{-sA}B(s)ds$$

is also a solution of equation (1.45).
(iii) Let $u(t)$ be a solution of equation (1.45) and $v(t)$ be a solution of $\dot{x}(t) = Ax(t)$. Then $x(t) = u(t) + v(t)$ is a solution of equation (1.45), since

$$\begin{aligned} \dot{x}(t) &= \dot{u}(t) + \dot{v}(t) \\ &= Au(t) + B(t) + Av(t) \\ &= A\big(u(t) + v(t)\big) + B(t) \\ &= Ax(t) + B(t). \end{aligned}$$

□

If $B(t)$ is of non-trivial complexity, it is hard to compute the integral in (1.47). If $B(t)$ is simple, we calculate $x(t)$ following the obvious steps:

(i) We determine the matrices A and $B(t)$.

(ii) We calculate the matrices e^{-sA} and e^{tA}.

(iii) We calculate the $(n\times 1)$-matrix that corresponds to the integral $\int_0^t \mathsf{e}^{-sA}B(s)ds$.

(iv) We find the product of the matrix e^{tA} and the $(n\times 1)$-matrix $\int_0^t \mathsf{e}^{-sA}B(s)ds+K$.

1.8. Higher order linear odes

An ode of higher order is a a linear ode with constant coefficients that involves derivatives higher than the first.

DEFINITION 1.8.1. If $n \geq 2$, $s : \mathbb{R} \to \mathbb{R}$ is an n-differentiable function and $a_1, \ldots, a_n \in \mathbb{R}$, the ode

$$s^{(n)}(t) + a_1 s^{(n-1)}(t) + \ldots + a_{n-1}\dot{s}(t) + a_n s(t) = 0 \tag{1.48}$$

is an ode of *higher order* n. If $n = 2$, equation (1.48) becomes

$$\ddot{s}(t) + a_1\dot{s}(t) + a_2 s(t) = 0. \tag{1.49}$$

If we introduce new variables, equation (1.49) is reduced to a linear system of odes with constant coefficients. Namely, if $x_1 = s$ and $x_2 = \dot{x}_1 = \dot{s}$, equation (1.49) becomes $\dot{x}_2 + a_1 x_2 + a_2 x_1 = 0$, hence we get the following system of odes:

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -a_2 x_1 - a_1 x_2. \end{aligned} \tag{1.50}$$

If (x_1, x_2) is a solution of the system (1.50), then $s = x_1$ is a solution of equation (1.49), and if s is a solution of equation (1.49), then $(s, \dot{s})$ is a solution of the system (1.50). The matrix of the system (1.50) is

$$A_2 = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix},$$

with characteristic polynomial

$$p_{A_2}(\lambda) = \mathtt{Det}(A_2 - \lambda I_2) = \lambda^2 + a_1\lambda + a_2.$$

Similarly, equation (1.48) is reduced to a linear system of odes with constant coefficients. If we define $x_1 = s$, $x_2 = \dot{x}_1 = \dot{s}$, $\ldots x_n = \dot{x}_{n-1}$, we get the following system of odes:

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\vdots \\ \dot{x}_{n-1} &= x_n, \\ \dot{x}_n &= -a_n x_1 - a_{n-1} x_2 - \ldots - a_1 x_n. \end{aligned} \tag{1.51}$$

The matrix of the system (1.51) is

$$A_n = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ -a_n & -a_{n-1} & \ldots & -a_2 & -a_1 \end{bmatrix}.$$

PROPOSITION 1.8.2. *If $n \geq 2$, the characteristic polynomial of A_n is given by*

$$p_{A_n}(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n.$$

PROOF. By induction on $n \geq 2$. Case $n = 2$ is shown above, and the inductive case is straightforward (the details are left to the reader). □

THEOREM 1.8.3. *Let λ_1, λ_2 be the roots of the characteristic polynomial p_{A_2} of A_2. For the solution $s(t)$ of the ode (1.49) the following hold:*
(i) If $\lambda_1, \lambda_2 \in \mathbb{R}$ are distinct, there are $C_1, C_2 \in \mathbb{R}$ such that

$$s(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}.$$

(ii) If $\lambda_1 = \lambda_2 = \lambda \in \mathbb{R}$, there are $C_1, C_2 \in \mathbb{R}$ such that

$$s(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}.$$

(iii) If $\lambda_1, \lambda_2 \in \mathbb{C} \setminus \mathbb{R}$, and $\lambda_1 = u + iv$, there are $C_1, C_2 \in \mathbb{R}$ such that

$$s(t) = e^{ut}\big(C_1 \cos(vt) + C_2 \sin(vt)\big).$$

PROOF. (i) By Theorem 1.4.17 there are $K_1, K_2 \in \mathbb{R}$ such that for the diagonalizing system of coordinates $\big(y_1(t), y_2(t)\big)$ we have that $y_1(t) = K_1 e^{\lambda_1 t}$ and $y_2(t) = K_2 e^{\lambda_2 t}$. For the original system $\big(x_1(t), x_2(t)\big)$ we have that

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix},$$

hence $s(t) = x_1(t) = p_{11}K_1 e^{\lambda_1 t} + p_{12}K_2 e^{\lambda_2 t}$.
(ii) One can show that in this case A_2 is similar to a matrix B of the form

$$B = \begin{bmatrix} \lambda & 0 \\ \beta & \lambda \end{bmatrix}; \qquad \beta \neq 0.$$

As we have seen already in the solution of system (1.44), the solutions of the system $\dot{y}(t) = By(t)$ are

$$\begin{aligned} y_1(t) &= K_1 e^{\lambda t}, \\ y_2(t) &= K_2 e^{\lambda t} + K_1 \beta t e^{\lambda t}, \end{aligned}$$

where $K_1, K_2 \in \mathbb{R}$. As in the previous case, the solutions $x_1(t)$ and $x_2(t)$ of the original system are linear combinations of $y_1(t)$ and $y_2(t)$.
(iii) By Corollary 1.5.18 and since we know the solutions of system (1.42), the solutions of the system

$$\dot{y}(t) = A_{uv} y(t)$$

are

$$\begin{aligned} y_1(t) &= K_1 e^{ut}\cos(vt) - K_2 e^{ut}\sin(vt) = e^{ut}\big(K_1\cos(vt) - K_2\sin(vt)\big), \\ y_2(t) &= K_1 e^{ut}\sin(vt) + K_2 e^{ut}\cos(vt) = e^{ut}\big(K_1\sin(vt) + K_2\cos(vt)\big). \end{aligned}$$

Since the solutions $x_1(t)$ and $x_2(t)$ of the original system are linear combinations of $y_1(t)$ and $y_2(t)$, the result follows. □

Proposition 1.8.4. *Let $S(a_1, \ldots, a_n)$ be the set of solutions of the higher ode*

$$s^{(n)}(t) + a_1 s^{(n-1)}(t) + \ldots + a_{n-1}\dot{s}(t) + a_n s(t) = 0.$$

(i) Equipped with pointwise addition and multiplication by reals the set $S(a_1, \ldots, a_n)$ is a real vector space.
(ii) If $f \in S(a_1, \ldots, a_n)$, then f is $(n+1)$-differentiable and $\dot{f} \in S(a_1, \ldots, a_n)$.

Proof. (i) Straightforward and left to the reader.
(ii) If $f \in S(a_1, \ldots, a_n)$, then $x = (f, \dot{f}, \ldots, f^{(n-1)})$ is a solution to the system (1.51). By Theorem 1.6.13 $x(t)$ has derivatives of all orders (i.e., it is infinitely differentiable), hence f is $(n+1)$-differentiable. To get $\dot{f} \in S(a_1, \ldots, a_n)$ we take the derivatives on both sides of the higher ode. □

Proposition 1.8.5. *If $\mathfrak{C}^\infty(\mathbb{R})$ is the set of infinitely differentiable functions of type $\mathbb{R} \to \mathbb{R}$, the following hold:*
(i) The constant functions $\texttt{Const}(\mathbb{R})$ is a subset of $\mathfrak{C}^\infty(\mathbb{R})$.
(ii) Equipped with pointwise addition and multiplication by reals the set $\mathfrak{C}^\infty(\mathbb{R})$ is a real vector space.
(iii) The solutions $S(a_1, \ldots, a_n)$ of the higher ode (1.48) is a subspace of $\mathfrak{C}^\infty(\mathbb{R})$.
(iv) The differentiation operator $D : \mathfrak{C}^\infty(\mathbb{R}) \to \mathfrak{C}^\infty(\mathbb{R})$ defined by

$$Df := \dot{f},$$

for every $f \in \mathfrak{C}^\infty(\mathbb{R})$, is in $L\big(\mathfrak{C}^\infty(\mathbb{R})\big)$.
(v) For every $\lambda \in \mathbb{R}$, the mapping $M_\lambda : \mathfrak{C}^\infty(\mathbb{R}) \to \mathfrak{C}^\infty(\mathbb{R})$ defined by

$$M_\lambda f := \lambda f,$$

for every $f \in \mathfrak{C}^\infty(\mathbb{R})$, is in $L\big(\mathfrak{C}^\infty(\mathbb{R})\big)$. Moreover, $M_1 = \mathrm{id}_{\mathfrak{C}^\infty(\mathbb{R})}$ and $M_0 = \bar{0}$, the zero operator in $L\big(\mathfrak{C}^\infty(\mathbb{R})\big)$.
(vi) The mapping $M_{\mathrm{id}_\mathbb{R}} : \mathfrak{C}^\infty(\mathbb{R}) \to \mathfrak{C}^\infty(\mathbb{R})$ defined by

$$M_{\mathrm{id}_\mathbb{R}} f := \mathrm{id}_\mathbb{R} \cdot f,$$

for every $f \in \mathfrak{C}^\infty(\mathbb{R})$, is in $L\big(\mathfrak{C}^\infty(\mathbb{R})\big)$.
(vii) If D^n is the n-th application of D to itself, and if

$$p(t) = t^n + a_1 t^{n-1} + \ldots + a_{n-1} t + a_n \in \mathbb{R}[t],$$

the mapping $p(D) : \mathfrak{C}^\infty(\mathbb{R}) \to \mathfrak{C}^\infty(\mathbb{R})$ defined by

$$p(D) := D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n I_{\mathfrak{C}^\infty(\mathbb{R})},$$

i.e.,

$$p(D)f = f^{(n)} + a_1 f^{(n-1)} + \ldots + a_{n-1}\dot{f} + a_n f,$$

for every $f \in \mathfrak{C}^\infty(\mathbb{R})$, is in $L\big(\mathfrak{C}^\infty(\mathbb{R})\big)$.

Proof. Straightforward and left to the reader. □

Clearly, if $p(t)$ is the polynomial corresponding to equation (1.48), and $f \in \mathtt{Ker}(p(D))$, then f is a solution to equation (1.48). Hence the problem of solving (1.48) is reduced to the problem of finding elements of $\mathtt{Ker}(p(D))$.

PROPOSITION 1.8.6. *If $p(t), q(t), r(t) \in \mathbb{R}[t]$ such that $p(t) = q(t) \cdot r(t)$, the following hold:*
(i) $\mathtt{Ker}(r(D)) \subseteq \mathtt{Ker}(p(D))$ and $\mathtt{Ker}(q(D)) \subseteq \mathtt{Ker}(p(D))$.
(ii) If $f \in \mathtt{Ker}(q(D))$ and $g \in \mathtt{Ker}(r(D))$, then $f + g \in \mathtt{Ker}(p(D))$.

PROOF. The proof of (i) is straightforward, while if $q(D)f = 0 = r(D)g$, then by case (i) $p(D)f = 0 = p(D)g$, hence $p(D)(f+g) = p(D)f + p(D)g = 0$. □

From now on we denote $M_{\mathrm{id}_{\mathbb{R}}}$ by M_t and $M_{\mathrm{id}_{\mathbb{R}}^k}$ by M_{t^k} i.e.,

$$M_t f := tf, \quad \text{and} \quad M_{t^k} f := t^k f.$$

LEMMA 1.8.7. *If $\lambda \in \mathbb{R}$, then for every $k \geq 1$ we have that*

$$(D - M_\lambda) \circ M_{t^k} - M_{t^k} \circ (D - M_\lambda) = kM_{t^{k-1}}.$$

PROOF. By induction on $k \geq 1$. If $k = 1$, we show that

$$(D - M_\lambda) \circ M_t - M_t \circ (D - M_\lambda) = M_1 = \mathrm{id}_{\mathfrak{C}^\infty(\mathbb{R})}.$$

First we observe that

$$D \circ M_t - M_t \circ D = M_1,$$

since by the Leibniz rule we get

$$\big[D \circ M_t - M_t \circ D\big]f = D(tf) - tDf = \dot{t}f + tDf - tDf = f.$$

Since $(M_\lambda \circ M_t)f = \lambda M_t f = \lambda t f = t(\lambda f) = (M_t \circ M_\lambda)f$, we get

$$\begin{aligned} (D - M_\lambda) \circ M_t - M_t \circ (D - M_\lambda) &= D \circ M_t - M_\lambda \circ M_t - M_t \circ D + M_t \circ M_\lambda \\ &= D \circ M_t - M_t \circ D \\ &= M_1. \end{aligned}$$

For the inductive step we observe first that

$$\begin{aligned} (D - M_\lambda) \circ M_{t^{k+1}} - M_{t^{k+1}} \circ (D - M_\lambda) &= D \circ M_{t^{k+1}} - M_\lambda \circ M_{t^{k+1}} - \\ &\quad - M_{t^{k+1}} \circ D + M_{t^{k+1}} \circ M_\lambda \\ &= D \circ M_{t^{k+1}} - M_{t^{k+1}} \circ D, \end{aligned}$$

and we reach our conclusion by the following equalities:

$$\begin{aligned} \big[D \circ M_{t^{k+1}} - M_{t^{k+1}} \circ D\big]f &= D(t^{k+1}f) - t^{k+1}Df \\ &= (k+1)t^k f + t^{k+1}Df - t^{k+1}Df \\ &= (k+1)t^k f \\ &= (k+1)M_{t^k} f. \end{aligned}$$

□

PROPOSITION 1.8.8. *If $m \in \mathbb{N}^+, \lambda \in \mathbb{R}$, and $p(t) \in \mathbb{R}[t]$, then*

$$(t-\lambda)^m \mid p(t) \Longrightarrow \forall_{k \in \{0,\ldots,m-1\}}\bigg(t^k e^{\lambda t} \in \mathtt{Ker}(p(D))\bigg).$$

PROOF. It suffices to show that

$$\forall_{k\in\mathbb{N}}\bigg((D-M_\lambda)^{k+1}t^k e^{\lambda t}=0\bigg),$$

since then we get all required cases:

$$(D-M_\lambda)e^{\lambda t}=0,$$
$$(D-M_\lambda)^2 te^{\lambda t}=0,$$
$$\vdots$$
$$(D-M_\lambda)^m t^{m-1}e^{\lambda t}=0,$$

since, if $\sigma(t) := t-\lambda$, and by hypothesis $\sigma^j(t) \mid p(t)$, for every $j \in \{1,\ldots,m\}$, we get that $t^j e^{\lambda t} \in \mathtt{Ker}(\sigma^j(D))$, hence by Proposition 1.8.6(i) $t^j e^{\lambda t} \in \mathtt{Ker}(p(D))$.

If $k=0$, the equality $De^{\lambda t}=\lambda e^{\lambda t}$ is written as $(D-M_\lambda)e^{\lambda t}=0$ By Lemma 1.8.7, and since $(D-M_\lambda)e^{\lambda t}=0$, we get

$$\begin{aligned}(D-M_\lambda)^{k+1}t^k e^{\lambda t} &= (D-M_\lambda)^{k+1}\big(M_{t^k}e^{\lambda t}\big)\\ &= (D-M_\lambda)^k\bigg[\bigg((D-M_\lambda)\circ M_{t^k}\bigg)e^{\lambda t}\bigg]\\ &= (D-M_\lambda)^k\bigg[\bigg(M_{t^k}\circ(D-M_\lambda)+kM_{t^{k-1}}\bigg)e^{\lambda t}\bigg]\\ &= (D-M_\lambda)^k\bigg[t^k(D-M_\lambda)e^{\lambda t}+kt^{k-1}e^{\lambda t}\bigg]\\ &= (D-M_\lambda)^k kt^{k-1}e^{\lambda t}\\ &= k(D-M_\lambda)^k t^{k-1}e^{\lambda t}\\ &= 0,\end{aligned}$$

since by the inductive hypothesis $(D-M_\lambda)^k t^{k-1}e^{\lambda t}=0$. □

Everything we said in this section so far works also for $\mathbb{C}$ instead of $\mathbb{R}$. Recall[10] that a polynomial $p(t) \in \mathbb{C}[t]$ of degree ≥ 1 has a factorization

$$p(t)=cp_1(t)\ldots p_s(t),$$

where $p_1(t)\ldots p_s(t) \in \mathbb{C}[t]$ are irreducible, with leading coefficient 1, $c \in \mathbb{C}$, and this factorization is unique up to permutation. This factorization holds for the polynomials in $\mathbb{R}[t]$ too, since if $\mathbb{K}$ is a field, the integral domain $\mathbb{K}[t]$ is a principal ideal domain, hence a unique factorization domain. The use of $\mathbb{C}[t]$ though, is

[10]See [**8**], Chapter XI, section 3.

crucial in the form of the irreducible polynomials occurring in the factorization of $p(t)$. Namely, a *monic* polynomial $p(t) \in \mathbb{C}[t]$, i.e., a polynomial with leading coefficient 1, is written as

$$p(t) = (t - \lambda_1)^{m_1} \cdot \ldots \cdot (t - \lambda_r)^{m_r},$$

where m_j is the *multiplicity* of $(t - \lambda_j)$ in $p(t)$, or the *multiplicity* of λ_j in $p(t)$ for every $j \in \{1, \ldots, r\}$.

COROLLARY 1.8.9. *The following n functions belong to the set $S(a_1, \ldots, a_n)$ of solutions of the higher ode (1.48):*
(i) The functions

$$f(t) = t^k e^{\lambda t},$$

where λ is any of the distinct real roots of the polynomial[11] *of $p(t)$ that corresponds to (1.48), and $k \in \mathbb{N}$ is between 0 and the multiplicity of λ in $p(t)$.*
(ii) The functions

$$g(t) = t^k e^{at} \cos(bt), \qquad h(t) = t^k e^{at} \sin(bt),$$

where $\lambda = a + ib$ is any of the non-real, complex roots of $p(t)$ with $b > 0$ and $k \in \mathbb{N}$ is between 0 and the multiplicity of λ in $p(t)$.

PROOF. (i) It follows from the above factorization of $p(t)$ and Proposition 1.8.8. If m is the multiplicity of λ, then the following m functions are in $S(a_1, \ldots, a_n)$:

$$e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1} e^{\lambda t}.$$

(ii) Since $p(D)$ has real coefficients, we have that

$$p(D) i f = i p(D) f,$$

for every n-differentiable function $f : \mathbb{R} \to \mathbb{R}$, hence by the generalization of Proposition 1.8.8 to $\lambda \in \mathbb{C}$ and $p(t) \in \mathbb{C}[t]$ we get

$$\begin{aligned} 0 &= p(D) t^k e^{\lambda t} \\ &= p(D) t^k e^{(a+ib)t} \\ &= p(D) t^k \big(\cos(bt) + i \sin(bt) \big) \\ &= p(D) t^k e^{at} \cos(bt) + i p(D) t^k e^{at} \sin(bt), \end{aligned}$$

therefore $p(D) t^k e^{at} \cos(bt) = 0 = p(D) t^k e^{at} \sin(bt)$. □

Note that a non-real, complex root of $p(t)$ of the form $a - ib$ generates the functions $g(t)$ and $-h(t)$, hence it adds no new functions to the linear span of the functions mentioned in Corollary 1.8.9. Next we show that these functions not only belong to $S(a_1, \ldots, a_n)$, but also form a basis for it. For the proof of this fact we need some preparation.

[11] Note that by Proposition 1.8.2 the polynomial $p(t)$ is the characteristic polynomial of the matrix of the system (1.51) that corresponds to equation (1.48).

First we note that the definition in Proposition 1.8.5(vii) is generalized to any complex vector space X and $T \in L(X)$ i.e, if $p(t)t^n + a_1 t + \dots a_{n-1}t + a_n \in \mathbb{C}[t]$, then $p(T) \in L(X)$ is defined by

$$p(T) := T^n + a_1 T^{n-1} + \dots + a_{n-1}T + a_n I_X.$$

Hence every $p(t) \in \mathbb{C}[t]$ determines the function

$$\begin{aligned} p(\cdot) : L(X) &\to L(X) \\ T &\mapsto p(T), \end{aligned}$$

and consequently we get the mapping

$$\begin{aligned} (\cdot) : \mathbb{C}[t] &\to \big(L(X) \to L(X)\big) \\ p &\mapsto p(\cdot). \end{aligned}$$

It is immediate to see that if 1 is the constant polynomial 1 in $\mathbb{C}[t]$, then

$$1(T) = I_X$$

i.e., $1(\cdot)$ is the constant mapping I_X on $L(X)$. Moreover, if $p(t), q(t) \in \mathbb{C}[t]$, then

$$p(T) \circ q(T) = (p \cdot q)(T) = (q \cdot p)(T) = q(T) \circ p(T).$$

For simplicity we may write $p(T)q(T)$ instead of $p(T) \circ q(T)$.

Proposition 1.8.10. *If X be a complex vector space, $T \in L(X)$, and $p(t) \in \mathbb{C}[t]$ such that $p(t) = q(t)r(t)$, for some $q(t), r(t) \in \mathbb{C}[t]$ with degree ≥ 1 and greatest common divisor equal to 1, and $p(T) = 0$, then*

$$X = Y_1 \oplus Y_2,$$

where $Y_1 = \mathtt{Ker}\big(q(T)\big)$ and $Y_2 = \mathtt{Ker}\big(r(T)\big)$.

Proof. Let $\sigma(t), \tau(t) \in \mathbb{C}[t]$ such that $\sigma(t)q(t) + \tau(t)r(t) = 1$. Hence

$$\sigma(T)q(T) + \tau(T)r(T) = I_X,$$

and

$$x = I_X x = \big[\sigma(T)q(T) + \tau(T)r(T)\big]x = \sigma(T)q(T)x + \tau(T)r(T)x.$$

Since

$$\begin{aligned} r(T)\sigma(T)q(T)x &= \sigma(T)r(T)q(T)x \\ &= \sigma(T)q(T)r(T)x \\ &= \sigma(T)p(T)x \\ &= \sigma(T)0x \\ &= 0, \end{aligned}$$

we get $\sigma(T)q(T)x \in Y_2$. Similarly we get $\tau(T)r(T)x \in Y_1$, hence $X = Y_1 + Y_2$. If $x = y_1 + y_2$, where $y_1 \in Y_1$ and $y_2 \in Y_2$, then

$$\sigma(T)q(T)x = \sigma(T)q(T)(y_1 + y_2)$$

$$\begin{aligned} &= \sigma(T)q(T)y_1 + \sigma(T)q(T)y_2 \\ &= 0 + \sigma(T)q(T)y_2 \\ &= \sigma(T)q(T)y_2. \end{aligned}$$

Hence we get

$$\begin{aligned} y_2 &= I_X y_2 \\ &= \big[\sigma(T)q(T) + \tau(T)r(T)\big] y_2 \\ &= \sigma(T)q(T)y_2 + \tau(T)r(T)y_2 \\ &= \sigma(T)q(T)y_2 + 0 \\ &= \sigma(T)q(T)y_2 \\ &= \sigma(T)q(T)x. \end{aligned}$$

Similarly we get $y_1 = \tau(X)r(T)x$ i.e., y_1, y_2 are uniquely determined. □

LEMMA 1.8.11. *Let X be a complex vector space and $T \in L(X)$. If $p(t) \in \mathbb{C}[t]$, then $\mathtt{Ker}\big(p(T)\big)$ is T-invariant.*

PROOF. If $x \in \mathtt{Ker}\big(p(T)\big)$, we show that $Tx \in \mathtt{Ker}\big(p(T)\big)$. Since $p(t)\cdot t = t\cdot p(t)$,

$$p(T) \circ T = T \circ p(T),$$

hence $p(T)Tx = Tp(T)x = T0 = 0$. □

THEOREM 1.8.12. *Let $r \geq 2$. If X is a complex vector space, $T \in L(X)$, and $p(t) \in \mathbb{C}[t]$ such that $p(t) = (t - \lambda_1)^{m_1} \cdot \ldots \cdot (t - \lambda_r)^{m_r}$, for distinct $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$, and $p(T) = 0$, then*

$$X = Y_1 \oplus \ldots \oplus Y_r,$$

where $Y_1 = \mathtt{Ker}\big((T - \lambda_1 I_X)^{m_1}\big), \ldots, Y_r = \mathtt{Ker}\big((T - \lambda_r I_X)^{m_r}\big)$.

PROOF. If $T = \lambda_i I_X$, for some $i \in \{1, \ldots, r\}$, then $\mathtt{Ker}\big((T - \lambda_i I_X)^{m_i}\big) = X$ and $\mathtt{Ker}\big((T - \lambda_j I_X)^{m_j}\big) = \{0\}$, for every $j \in \{1, \ldots, r\} \setminus \{i\}$. Hence we get immediately what we want to show. Suppose next that $T \neq \lambda_i I_X$, for every $i \in \{1, \ldots, r\}$. We prove what we want by induction on $r \geq 2$. The case $r = 2$ follows immediately from Proposition 1.8.10. If $r > 2$, let

$$Z := \mathtt{Ker}\Bigg((T - \lambda_2 I_X)^{m_2} \circ \ldots \circ (T - \lambda_r I_X)^{m_r}\Bigg).$$

Since $\lambda_1, \ldots, \lambda_r$ are distinct, in the factorization

$$p(t) = (t - \lambda_1)^{m_1}\Bigg[(t - \lambda_2)^{m_2} \cdot \ldots \cdot (t - \lambda_r)^{m_r}\Bigg]$$

of $p(t)$ the polynomials $q(t) := (t - \lambda_1)^{m_1}$ and $s(t) := (t - \lambda_2)^{m_2} \cdot \ldots \cdot (t - \lambda_r)^{m_r}$ have greatest common divisor equal to 1. Hence by Proposition 1.8.10 we get

$$X = Y_1 \oplus Z.$$

If $T' = T_{|Z}$, then T' is linear, and by Lemma 1.8.11 we have that if $z \in \mathtt{Ker}\big(s(T)\big)$, then $T'z = Tz \in \mathtt{Ker}\big(s(T)\big)$, therefore $T' \in L(Z)$. By definition of Z we have that, if $z \in Z$, then $s(T')z = 0$ i.e., $s(T')$ is the zero element of $L(Z)$. Hence by the inductive hypothesis on $r-1$ for Z, T', and $s(t)$ we get

$$Z = Z_2 \oplus \ldots \oplus Z_r,$$

$$Z_2 := \mathtt{Ker}\big((T - \lambda_2 I_Z)^{m_2})\big), \ldots, Z_r := \mathtt{Ker}\big((T - \lambda_r I_Z)^{m_r})\big).$$

It suffices to show that for every $j \in \{2, \ldots, r\}$

$$Z_j = \mathtt{Ker}\big((T - \lambda_j I_X)^{m_j})\big).$$

For this it suffices to show that

$$Z_j \supseteq \mathtt{Ker}\big((T - \lambda_j I_X)^{m_j})\big).$$

Since

$$X = Y_1 \oplus Z_2 \oplus \ldots \oplus Z_r,$$

if $x \in \mathtt{Ker}\big((T - \lambda_j I_X)^{m_j})\big)$, there are unique $y_1 \in Y_1, z_2 \in Z_2, \ldots, z_r \in Z_r$ such that

$$x = y_1 + z_2 + \ldots + z_r.$$

Since the polynomials in the factorization of $s(t)$ commute, the compositions of the corresponding operators also commute, and since $(T - \lambda_j I_X)^{m_j} x = 0$, we also get

$$\left[(T - \lambda_2 I_X)^{m_2} \circ \ldots \circ (T - \lambda_r I_X)^{m_r}\right] x = 0$$

i.e., $x \in Z$. Consequently, $y_1 = 0$ and $x = z_j \in Z_j$. $\square$

Corollary 1.8.13. *Let $p(t) \in \mathbb{C}[t]$ and $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ are distinct such that*

$$p(t) = t^n + a_1 t^{n-1} + \ldots + a_{n-1} t + a_n$$
$$= (t - \lambda_1)^{m_1} \cdot \ldots \cdot (t - \lambda_r)^{m_r}.$$

If $S(a_1, \ldots, a_n)$ is the complex vector space of the solutions of the ode

$$s^{(n)} + a_1 s^{(n-1)} + \ldots + a_{n-1} \dot{s} + a_n s = 0,$$

then

$$S(a_1, \ldots, a_n) = Y_1 \oplus \ldots \oplus Y_r,$$

where Y_j is the space of solutions of the ode

$$\big(D - \lambda_j I_{S(a_1, \ldots, a_n)}\big)^{m_j} s = 0,$$

for every $j \in \{1, \ldots, r\}$.

Proof. Immediately by Theorem 1.8.12 for $X = S(a_1, \ldots, a_n)$ and $T = D$. $\square$

Lemma 1.8.14. *Let the space $S(a_1, \ldots, a_n)$ be as in Corollary 1.8.13, and let $s \in S(a_1, \ldots, a_n)$. If $m \geq 1$, then for every $\lambda \in \mathbb{C}$*

$$(D - \lambda I_{S(a_1, \ldots, a_n)})^m s = e^{\lambda t} D^m \big(e^{-\lambda t} s\big).$$

PROOF. Exercise. □

THEOREM 1.8.15. *Let $\lambda \in \mathbb{C}$ and $m \geq 1$. If S_λ is the set of solutions of the ode*

$$(t-\lambda)^m(D)s = 0,$$

then the m functions

$$e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1}e^{\lambda t}$$

form a basis for S_λ.

PROOF. By Lemma 1.8.14 we have that

$$s \in \mathtt{Ker}\big((D-\lambda I_{S_\lambda})^m\big) \Leftrightarrow D^m\big(e^{-\lambda t}s\big) = 0.$$

The only functions the m-derivative of which is constant 0 are the polynomials of degree $\leq m-1$. Hence there are $b_0, \ldots, b_{m-1} \in \mathbb{C}$ such that

$$e^{-\lambda t}s = b_0 \ \vee\ \ldots\ \vee\ e^{-\lambda t}s = b_{m-1}t^{m-1}.$$

Hence

$$s = b_0 e^{\lambda t} \ \vee\ \ldots\ \vee\ s = b_{m-1}t^{m-1}e^{\lambda t}$$

i.e., the functions $e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1}e^{\lambda t}$ generate the space S_λ. The fact that these functions are linearly independent is an exercise. □

CHAPTER 2

Fundamental theorems of ODEs

2.1. The fundamental local theorem of odes

A dynamical system is a mathematical description of the passage in time of the points in some space $\boldsymbol{S}$, which is usually understood as the space of states of some physical system. From now on $\boldsymbol{S}$ denotes an open subset of $\mathbb{R}^n$.

DEFINITION 2.1.1. A *dynamical system* on $\boldsymbol{S}$ is a C^1 function $\phi : \mathbb{R} \times \boldsymbol{S} \to \boldsymbol{S}$

$$(t, u) \mapsto \phi(t, u),$$

such that if for every $t \in \mathbb{R}$ we define the function $\phi_t : \boldsymbol{S} \to \boldsymbol{S}$

$$u \mapsto \phi_t(u) := \phi(t, u),$$

the following properties are satisfied:
(i) $\phi_0 = \mathrm{id}_{\boldsymbol{S}}$.
(ii) $\forall_{s,t\in\mathbb{R}}(\phi_s \circ \phi_t = \phi_{s+t})$.

REMARK 2.1.2. If ϕ is a dynamical system on $\boldsymbol{S}$, the following hold:
(i) $\forall_{t\in\mathbb{R}}(\phi_t \text{ is } C^1)$.
(ii) $\forall_{t\in\mathbb{R}}(\phi_t \text{ has a } C^1 \text{ inverse})$.

PROOF. Left to the reader. □

DEFINITION 2.1.3. The vector field f on $\boldsymbol{S}$ generated by a dynamical system ϕ on $\boldsymbol{S}$ is given by

$$f(x) := \frac{d}{dt}\phi_t(x)\bigg|_{t=0} \tag{2.1}$$

i.e., $f(x)$ is a vector in $\mathbb{R}^n$, which is tangent to the curve $x : J \to \boldsymbol{S}$, defined by $t \mapsto x(t) := \phi_t(x)$, at $t = 0$. We rewrite equation (2.1) as the initial value problem

$$\dot{x} = f(x) \tag{2.2}$$

and $x(0) = \phi_0(x) = x$

As we have already seen, the linear ode $\dot{x}(t) = Ax(t)$ generates the dynamical system $\phi_A : \mathbb{R} \to \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\phi_A(t, u) = \mathsf{e}^{tA}u.$$

Next we generalize this fact. Given an ode of the form (2.2) we associate to it an object that would be a dynamical system if it were definable in $\mathbb{R}$.

Definition 2.1.4. Let $f : \boldsymbol{S} \to \mathbb{R}^n$ be continuous. A *solution* of the ode

$$\dot{x} = f(x) \tag{2.3}$$

is a differentiable function

$$u : J \to \boldsymbol{S},$$

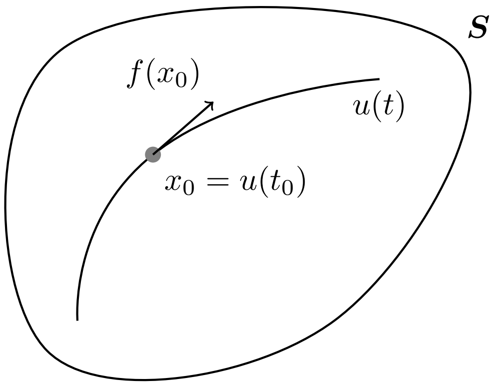
where, if $a, b \in \mathbb{R}$ and $a < b$, J is an interval that has one of the following forms[1]

$$(a, b),\ (a, b],\ [a, b),\ [a, b],\ (-\infty, b),\ (-\infty, b],\ (a, \infty),\ [a, \infty),\ (-\infty, \infty),$$

such that for all $t \in J$

$$\dot{u}(t) = f(u(t)).$$

From the geometric point of view a solution u to equation (2.3) is a curve in $\boldsymbol{S}$ whose tangent vector $\dot{u}(t)$ is the vector $f(u(t))$.



Generally, there are more than one solutions of the ode (2.3). E.g., the ode

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

where $\boldsymbol{S} = \mathbb{R}$, has both $u_0(t) = 0$, for every $t \in \mathbb{R}$, and $u_1(t) = t^3$, for every $t \in \mathbb{R}$, as solutions. As we will show, we get uniqueness, if f is C^1, while for existence continuity of f suffices. As we saw in the previous example, continuity of f does not imply uniqueness of solutions to (2.3).

Definition 2.1.5. If $(X, ||.||), (Y, ||.||')$ are normed spaces, a function $f : X \to Y$ is called *locally Lipschitz*, if for every $x \in X$ there is a neighborhood V_x of x such that the restriction $f_{|V_x}$ of f to V_x is Lipschitz i.e., there is some $K > 0$, which depends on x and V_x, such that for all $y, z \in V_x$

$$||f(y) - f(z)||' \leq K||y - z||.$$

[1] I.e., J does not have the form $[-\infty, b], [-\infty, b)$, or $[a, \infty], (a, \infty]$.

LEMMA 2.1.6. *If $f : \boldsymbol{S} \to \mathbb{R}^n$ is C^1, then f is locally Lipschitz.*

PROOF. Let $x_0 \in \boldsymbol{S}$. Since $\boldsymbol{S}$ is open, there is some $\epsilon_0 > 0$ such that

$$V_{x_0} := \mathcal{B}(x_0, \epsilon_0] = \{y \in \mathbb{R}^n \mid |y - x_0| \leq \epsilon_0\} \subseteq \boldsymbol{S}.$$

By Definition 1.3.4 the function $Df : \boldsymbol{S} \to L(\mathbb{R}^n)$, where $Df(x) \in L(\mathbb{R}^n)$ satisfies

$$Df(x)u = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h},$$

is continuous. Since V_{x_0} is closed and bounded, hence by Proposition 1.1.18(iv) it is compact, the composition $||.|| \circ Df$ has a maximum on V_{x_0}. Let

$$K := \max\{||Df(y)|| \mid y \in V_{x_0}\}.$$

We could have also taken K to be any bound of $||.|| \circ Df$ on V_{x_0}. Since V_{x_0} is a closed ball, it is also a convex set. Let $y, z \in V_{x_0}$, and let

$$u := z - y.$$

If $s \in [0, 1]$, then $y + su \in V_{x_0}$, since $y + su = y + s(z - y) = (1 - s)y + sz$. Let $\theta : [0, 1] \to V_{x_0}$ defined by $\theta(s) := y + su$, and let

$$\phi := f_{|V_{x_0}} \circ \theta : [0, 1] \to \mathbb{R}^n$$
$$\phi(s) = f(y + su).$$

Applying the chain rule to the coordinate functions

$$\begin{array}{ccc} [0,1] & \xrightarrow{\ \theta\ } & V_{x_0} \subseteq \boldsymbol{S} \subseteq \mathbb{R}^n \\ & {}_{\phi_i}\searrow & \downarrow f_i \\ & & \mathbb{R} \end{array}$$

we get

$$\begin{aligned} \dot{\phi}_i(s) &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\theta(s)) \frac{d\theta_j}{ds}(s) \\ &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\theta(s)) \frac{d(y_j + su_j)}{ds}(s) \\ &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(y + su)u_j. \end{aligned}$$

Since

$$Df(y + su)u = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(y + su) & \dots & \frac{\partial f_1}{\partial x_n}(y + su) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1}(y + su) & \dots & \frac{\partial f_n}{\partial x_n}(y + su) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix},$$

we conclude that

$$\dot{\phi}(s) = Df(y+su)u.$$

Since $\theta(0) = y$ and $\theta(1) = y + u = y + (z - y) = z$, we get

$$\begin{aligned} f(z) - f(y) &= f(\theta(1)) - f(\theta(0)) \\ &= \phi(1) - \phi(0) \\ &= \int_0^1 \dot{\phi}(s)ds \\ &= \int_0^1 Df(y+su)uds. \end{aligned}$$

Hence we get

$$\begin{aligned} |f(z) - f(y)| &= \left| \int_0^1 Df(y+su)uds \right| \\ &\leq \int_0^1 \big|Df(y+su)u\big|ds \\ &\leq \int_0^1 ||Df(y+su)|| \cdot |u|ds \\ &\leq \int_0^1 K|u|ds \\ &= K|u| \int_0^1 ds \\ &= K|u| \\ &= K|z-y|. \end{aligned}$$

□

One can use Lemma 2.1.6 to find locally Lipschitz functions that are not Lipschitz.

COROLLARY 2.1.7. *If $f : \boldsymbol{S} \to \mathbb{R}^n$ is C^1, and $V \subseteq \boldsymbol{S}$ is convex such that $||Df(x)|| \leq K$, for some $K > 0$ and for every $x \in V$, then K is a Lipschitz constant for $f_{|V}$.*

PROOF. It follows immediately by inspection of the proof of Lemma 2.1.6. □

LEMMA 2.1.8. *Let J be an open interval of $\mathbb{R}$ such that $0 \in J$, $x_0 \in \boldsymbol{S}$, and $x : J \to \boldsymbol{S}$ is differentiable. The following are equivalent:*

(i) $\dot{x}(t) = f(x(t))$ *and* $x(0) = x_0$.

(ii) $x(t) = x_0 + \int_0^t f(x(s))ds$.

PROOF. Exercise. □

LEMMA 2.1.9 (Cauchy criterion of uniform convergence). *Let* $(f_n)_{n=0}^{\infty}$ *a sequence of continuous functions from a closed interval* $[a,b]$ *to* $\mathbb{R}^n$. *If*

$$\forall_{\epsilon>0}\exists_{N>0}\forall_{m,n>N}\forall_{t\in[a,b]}\big(|f_m(t)-f_n(t)|<\epsilon\big),$$

then there is a continuous $f:[a,b]\to\mathbb{R}^n$ *such that*

$$\forall_{\epsilon>0}\exists_{N>0}\forall_{n>N}\forall_{t\in[a,b]}\big(|f_n(t)-f(t)|<\epsilon\big).$$

PROOF. This is a standard result, and the proof is left to the reader. □

The conclusion of the previous lemma is usually formulated by the expression "f is the uniform limit of $(f_n)_{n=0}^{\infty}$".

LEMMA 2.1.10. *Let* $(f_n)_{n=0}^{\infty}$ *be a sequence of continuous functions from* $[a,b]$ *to* $K\subseteq\mathbb{R}^n$ *compact,* $f:[a,b]\to\mathbb{R}^n$, *and* $g:K\to\mathbb{R}^m$ *uniformly continuous. If* f *is the uniform limit of* $(f_n)_{n=0}^{\infty}$, *then the following hold:*
(*i*) f *is integrable on* $[a,b]$, *and*

$$\lim_{n\to\infty}\int_a^b f_n=\int_a^b\lim_{n\to\infty}f_n=\int_a^b f.$$

(*ii*) $g\circ f$ *is the uniform limit of* $(g\circ f_n)_{n=0}^{\infty}$.

PROOF. This is a standard result, and the proof is left to the reader. □

THEOREM 2.1.11 (Fundamental local theorem of odes). *If* $f:\boldsymbol{S}\to\mathbb{R}^n$ *is* C^1, *and* $x_0\in\boldsymbol{S}$, *then there is* $a>0$ *and a unique solution* $x:(-a,a)\to\boldsymbol{S}$ *of the ode* $\dot{x}=f(x)$ *that satisfies the initial condition* $x(0)=x_0$.

PROOF. By Lemma 2.1.6 f is locally Lipschitz on $V_{x_0}=\mathcal{B}(x_0,\epsilon_0]$, for some $\epsilon_0>0$, and has Lipschitz constant on V_{x_0}. Since V_{x_0} is compact, the continuous function $|f|$ is bounded on V_{x_0}. Let $M>0$ such that

$$\forall_{y\in V_{x_0}}(|f(y)|\leq M).$$

Moreover, let

$$0<a<\min\left\{\frac{\epsilon_0}{M},\frac{1}{K}\right\}, \tag{2.4}$$

$$J:=[-a,a].$$

Next we define a sequence $(u_n)_{n=0}^{\infty}$ of continuous functions from J to V_{x_0} as follows:

$$u_0(t):=x_0,\qquad t\in J,$$

and if we suppose that $u_n(t)$, where has been defined such that it satisfies

$$|u_n(t)-x_0|\leq\epsilon_0,\qquad t\in J, \tag{2.5}$$

a condition that holds trivially for $n=0$, we define

$$u_{n+1}(t):=x_0+\int_0^t f(u_n(s))ds,\qquad t\in J. \tag{2.6}$$

If we suppose that u_n is continuous, then the composition $f \circ u_n$ is also continuous, hence integrable. Clearly, u_n is C^1, for every n. We show that if $u_n : J \to V_{x_0}$, then $u_{n+1} : J \to V_{x_0}$ i.e.,

$$\forall_{t \in J} (|u_{n+1}(t) - x_0| \leq \epsilon_0).$$

If $t \in J$, then

$$\begin{aligned} |u_{n+1}(t) - x_0| &= \left| \int_0^t f(u_n(s)) ds \right| \\ &\leq \int_0^t \big| f(u_n(s)) \big| ds \\ &\leq \int_0^t M ds \\ &= Mt \\ &\leq Ma \\ &< M \frac{\epsilon_0}{M} \\ &= \epsilon_0. \end{aligned}$$

Next we show that the sequence $(u_n)_{n=0}^{\infty}$ satisfies the hypothesis of Lemma 2.1.9. First we need to show a useful inequality. Let

$$L := \max\{|u_1(t) - u_0(t)| \mid t \in J\}.$$

We show that for all $n \in \mathbb{N}$ and $t \in J$ we have that

$$|u_{n+1}(t) - u_n(t)| \leq (Ka)^n L. \tag{2.7}$$

The case $n = 0$ holds by definition. For the inductive step we have that

$$\begin{aligned} |u_{n+1}(t) - u_n(t)| &= \left| \int_0^t \big(f(u_n(s)) - f(u_{n-1}(s)) \big) ds \right| \\ &\leq \int_0^t \big| f(u_n(s)) - f(u_{n-1}(s)) \big| ds \\ &\leq \int_0^t K |u_n(s) - u_{n-1}(s)| ds \\ &\leq K (Ka)^{n-1} Lt \\ &\leq (Ka)^n L. \end{aligned}$$

If we fix some $\epsilon > 0$, we can find $N > 0$ such that for all $m, n > N$, and without loss of generality let $m > n$, and for all $t \in J$ we have that

$$|u_m(t) - u_n(t)| \leq \sum_{k=N}^{\infty} |u_{k+1}(t) - u_k(t)|$$

$$\begin{aligned}
&\leq \sum_{k=N}^{\infty}(Ka)^k L \\
&\leq L\sum_{k=N}^{\infty}(Ka)^k \\
&\leq \epsilon,
\end{aligned}$$

since by (2.4) we have that $Ka < 1$. Hence there is continuous $x : J \to \mathbb{R}^n$, which is the uniform limit of $(u_n)_{n=0}^{\infty}$. One can show (exercise) that actually $x : J \to V_{x_0}$.

Taking limits in the equality (2.6) and using Lemma 2.1.10 we get

$$\begin{aligned}
x(t) &= x_0 + \lim_{n\to\infty}\int_0^t f(u_n(s))ds \\
&= x_0 + \int_0^t \big[\lim_{n\to\infty} f(u_n(s))\big]ds \\
&= x_0 + \int_0^t f(x(s))ds,
\end{aligned}$$

hence by Lemma 2.1.8 $x(t)$ is a solution of the ode $\dot{x} = f(x)$ and satisfies $x(0) = x_0$ To show the uniqueness of the solution we suppose that there are $x : J \to V_{x_0}$ and $y : J \to V_{x_0}$ such that $\dot{x} = f(x)$ and $\dot{y} = f(y)$ and $x(0) = x_0 = y(0)$. Note that we can take without loss of generality J to be the same closed interval around 0. We show that $x(t) = y(t)$, for every $t \in J$. We define

$$\Lambda := \max\{|x(t) - y(t)| \mid t \in J\},$$

and let $t_\Lambda \in J$ such that $|x(t) - y(t)|$ attains Λ at t_Λ. If $\Lambda > 0$, we have that

$$\begin{aligned}
\Lambda &= |x(t_\Lambda) - y(t_\Lambda)| \\
&= \left|\int_0^{t_\Lambda}\dot{x}(s)ds - \int_0^{t_\Lambda}\dot{y}(s)\right| \\
&= \left|\int_0^{t_\Lambda}\big(\dot{x}(s) - \dot{y}(s)\big)ds\right| \\
&= \left|\int_0^{t_\Lambda}\big(f(x(s)) - f(y(s))\big)ds\right| \\
&\leq \int_0^{t_\Lambda}\big|f(x(s)) - f(y(s))\big|ds \\
&\leq \int_0^{t_\Lambda} K|x(s) - y(s)|ds \\
&\leq K\Lambda t_\Lambda \\
&\leq aK\Lambda \\
&< \Lambda,
\end{aligned}$$

since $Ka < 1$ and the hypothesis $\Lambda > 0$ implies $Ka\Lambda < \Lambda$. Since we reached a contradiction, we conclude that $\Lambda = 0$, and consequently $x = y$. $\square$

COROLLARY 2.1.12. *Let* $V_{x_0} := \mathcal{B}(x_0, \epsilon_0] \subseteq \boldsymbol{S}$*, for some* $\epsilon_0 > 0$*,* $M, K > 0$*, and*

$$0 < a < \min\left\{\frac{\epsilon_0}{M}, \frac{1}{K}\right\}.$$

If $f : \boldsymbol{S} \to \mathbb{R}^n$ *satisfies the conditions:*

(i) $\max\{|f(x)| \mid x \in V_{x_0}\} \leq M$*, and*

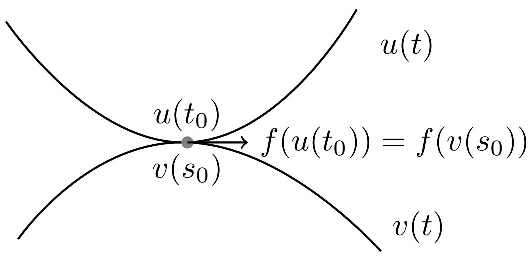
(ii) $f_{|V_{x_0}}$ *is* K*-Lipschitz,*

there is a unique solution $x : (-a, a) \to \boldsymbol{S}$ *of the ode* $\dot{x} = f(x)$ *that satisfies the initial condition* $x(0) = x_0$*.*

PROOF. It follows by inspection of the proof of Theorem 2.1.11. $\square$

COROLLARY 2.1.13. *Let* $f : \boldsymbol{S} \to \mathbb{R}^n$ *be* C^1*, and* $x_0 \in \boldsymbol{S}$*. Suppose that* $u : I \to \boldsymbol{S}$ *and* $v : J \to \boldsymbol{S}$ *are solutions of the ode* $\dot{x} = f(x)$ *that satisfy* $u(t_0) = v(s_0)$*, for some* $t_0 \in I$ *and* $s_0 \in J$*. Then there is some subinterval* I' *of* I *around* t_0 *and a subinterval* J' *of* J *around* s_0 *such that* $u(I') = v(J')$*.*

PROOF. Suppose that $u(t_0) = v(s_0)$ is a *crossing point*, as it is indicated in the following figure:



We define $x : I \to \boldsymbol{S}$ by

$$x(t) := v(s_0 - t_0 + t), \qquad t \in I.$$

Since $\dot{x}(t) = \dot{v}(s_0 - t_0 + t) = f(v(s_0 - t_0 + t)) = f(x(t))$, and since $x(t_0) = v(s_0) = u(t_0)$, by the uniqueness of the solution around t_0, there is an interval I_0 around t_0 such that $u_{|I_0} = x_{|I_0}$. If t is close to t_0, then $s_0 - t_0 + t$ is close to s_0, hence u and v meet again. $\square$

PROPOSITION 2.1.14. *Let* $a > 0$ *and let* $u : [0, a] \to [0, +\infty)$ *be continuous. If* $C \geq 0$ *and* $L \geq 0$ *such that for every* $t \in [0, a]$

$$u(t) \leq C + \int_0^t Lu(s)ds,$$

then for every $t \in [0, a]$ *we have that*

$$u(t) \leq Ce^{Lt}.$$

PROOF. Suppose first that $C > 0$. We define $U : [0, a] \to [0, +\infty)$ by

$$U(t) := C + \int_0^t Lu(s)ds.$$

Since $C > 0$ and $Lu(s) \geq 0$, we get that $U(t) > 0$, for every $t \in [0, a]$. By our hypothesis we have that for every $t \in [0, a]$

$$u(t) \leq U(t).$$

Since $\dot{U}(t) = Lu(t)$, we get

$$\frac{\dot{U}(t)}{U(t)} = \frac{Lu(t)}{U(t)} \leq L,$$

or equivalently

$$\frac{d}{dt}\big[\log(U(t))\big] \leq L.$$

hence

$$\begin{aligned}
\int_0^t \frac{d}{ds}\big[\log(U(s))\big]ds \leq \int_0^t Lds &\Leftrightarrow \log(U(t)) - \log(U(0)) \leq Lt \\
&\Leftrightarrow \log(U(t)) \leq \log(U(0)) + Lt \\
&\Leftrightarrow \log(U(t)) \leq \log(C) + Lt \\
&\Rightarrow e^{\log(U(t))} \leq e^{\log(C)+Lt} \\
&\Leftrightarrow U(t) \leq e^{\log(C)}e^{Lt} \\
&\Leftrightarrow U(t) \leq Ce^{Lt}.
\end{aligned}$$

The proof for case $C = 0$ is an exercise. □

THEOREM 2.1.15 (Continuity of solutions in initial conditions for Lipschitz function f). *Suppose that the* C^1 *function* $f : \boldsymbol{S} \to \mathbb{R}^n$ *has Lipschitz constant* $\sigma > 0$. *If* $y : [t_0, t_1] \to \boldsymbol{S}$ *and* $z : [t_0, t_1] \to \boldsymbol{S}$ *are solutions of the ode* $\dot{x} = f(x)$ *on* $[t_0, t_1]$, *then for every* $t \in [t_0, t_1]$

$$|y(t) - z(t)| \leq |y(t_0) - z(t_0)|e^{\sigma(t-t_0)}.$$

PROOF. For every $t \in [t_0, t_1]$ we define

$$w(t) := |y(t) - z(t)|.$$

Since

$$\begin{aligned}
y(t) - z(t) &= y(t_0) + \int_{t_0}^t f(y(s))ds - z(t_0) - \int_{t_0}^t f(z(s))ds \\
&= \big(y(t_0) - z(t_0)\big) + \int_{t_0}^t \big[f(y(s)) - f(z(s))\big]ds,
\end{aligned}$$

we get

$$\begin{aligned} w(t) &\leq \big|y(t_0) - z(t_0)\big| + \left| \int_{t_0}^{t} \big[f(y(s)) - f(z(s))\big] ds \right| \\ &\leq w(t_0) + \int_{t_0}^{t} \big|f(y(s)) - f(z(s))\big| ds \\ &\leq w(t_0) + \int_{t_0}^{t} \sigma \big|y(s) - z(s)\big| ds \\ &\leq w(t_0) + \int_{t_0}^{t} \sigma w(s) ds. \end{aligned}$$

If $a := t_1 - t_0 > 0$, then for the continuous function $u : [0, a] \to [0, +\infty)$, defined by

$$u(r) := w(r + t_0),$$

then $w(t_0) = u(0)$, and $w(t) = u(t - t_0)$. Moreover, if $g(r) := r + t_0$, we have that

$$\int_{t_0}^{t} \sigma w(s) ds = \int_{g(0)}^{g(t-t_0)} \sigma w(s) ds = \int_{0}^{t-t_0} \sigma w(g(r)) \dot{g}(r) dr = \int_{0}^{t-t_0} \sigma w(g(r)) dr.$$

Hence the inequality

$$w(t) \leq w(t_0) + \int_{t_0}^{t} \sigma w(s) ds$$

is written as

$$u(t - t_0) \leq u(0) + \int_{0}^{t-t_0} \sigma w(g(r)) dr.$$

By Proposition 2.1.14 we get

$$\begin{aligned} u(t - t_0) \leq u(0) e^{\sigma(t-t_0)} &\Leftrightarrow w(t) \leq w(t_0) e^{\sigma(t-t_0)} \\ &\Leftrightarrow |y(t) - z(t)| \leq |y(t_0) - z(t_0)| e^{\sigma(t-t_0)}. \end{aligned}$$

□

Definition 2.1.16. A sequence of continuous functions $(f_n)_{n=0}^{\infty}$ from $[a, b] \to \mathbb{R}^n$ is called *uniformly bounded*, if there is $M > 0$ such that

$$\forall_{n\in\mathbb{N}} \forall_{t\in[a,b]} \big(|f_n(x)| \leq M\big),$$

and it is called *equicontinuous*, if

$$\forall_{\epsilon>0} \exists_{\delta>0} \forall_{s,t\in[a,b]} \forall_{n\in\mathbb{N}} \big(|s - t| < \delta \Rightarrow |f_n(s) - f_n(t)| < \epsilon\big).$$

Theorem 2.1.17 (Arzela-Ascoli). *If $(f_n)_{n=0}^{\infty}$ is a sequence of continuous functions from $[a, b]$ to $\mathbb{R}^n$, which is uniformly bounded and equicontinuous, then $(f_n)_{n=0}^{\infty}$ has a subsequence $(f_{k_n})_{n=0}^{\infty}$ that converges uniformly on $[a, b]$.*

PROPOSITION 2.1.18. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous such that $\forall_{x\in\mathbb{R}^n}\big(|f(x)| \leq M\big)$, and let $x_0 \in \mathbb{R}^n$. Moreover, let $(x_n)_{n=0}^{\infty}$ be a sequence of functions from $[0,1]$ to $\mathbb{R}^n$ such that*

(i) x_n is a solution of the ode $\dot{x} = f(x)$, for every $n \in \mathbb{N}$, and
(ii) $\lim_{n\to\infty} x_n(0) = x_0$.

Then there is a subsequence of $(x_n)_{n=0}^{\infty}$ that converges uniformly on $[0,1]$ to a solution of $\dot{x} = f(x)$.

PROOF. Exercise. □

LEMMA 2.1.19. *Let $f : \boldsymbol{S} \to \mathbb{R}^n$ be C^1, and $u : I \to \boldsymbol{S}$, $v : I \to \boldsymbol{S}$ solutions of the ode $\dot{x} = f(x)$ such that $u(t_0) = v(t_0)$, where $t_0 \in I$. Then $u(t) = v(t)$, for every $t \in I$.*

PROOF. By Theorem 2.1.11 there is an open subinterval J_0 of I such that $t_0 \in J_0$ and $u_{|_{J_0}} = v_{|_{J_0}}$. Hence

$$\mathfrak{I} := \big\{J \subseteq I \mid t_0 \in J \wedge u_{|_J} = v_{|_J} \wedge J \text{ open interval}\big\} \neq \emptyset.$$

Since the union of intervals with a common point is an interval, the set

$$I^* := \bigcup \mathfrak{I} = \{t \mid \exists_{J\in\mathfrak{I}}(t \in J)\}$$

is an open interval. By its definition I^* is the largest open subinterval of I that contains t_0 and the restrictions of u and v to it are equal. If t_l, t_r are the endpoints of I^*, we show that

$$I \subseteq I^* = (t_l, t_r).$$

Suppose that this is not the case. Then at least one of the endpoints of I^* has to be in I. Let $t_r \in I$, and if $t_l \in I$, we work similarly. Since $u_{|_{I^*}} = v_{|_{I^*}}$, and since (t_l, t_r) is dense in $(t_l, t_r]$, by the continuity of u and v on $(t_l, t_r]$ we get $u(t_r) = v(t_r)$. By Theorem 2.1.11 there is an open subinterval J_r of I such that $t_r \in J_r$ and $u_{|_{J_r}} = v_{|_{J_r}}$. Hence $u_{|_{I^*\cup J_r}} = v_{|_{I^*\cup J_r}}$, and

$$I^* \subsetneq I^* \cup J_r \in \mathfrak{I},$$

which is a contradiction. The equality $I^* = I$ implies what we want to show. □

A solution $x(t)$ to an ode $\dot{x} = f(x)$ is not always extendable to $\mathbb{R}$. E.g., the ode $\dot{x} = 1 + x^2$ has a s solution the curve $x(t) = \tan(t - c)$, with $\big(c - \frac{\pi}{2}, c + \frac{\pi}{2}\big)$ as the largest interval of definition.

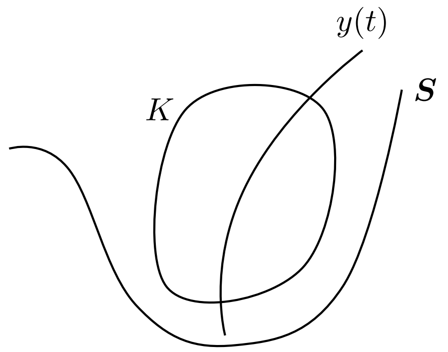
PROPOSITION 2.1.20. *Let $f : \boldsymbol{S} \to \mathbb{R}^n$ be C^1. For every $x_0 \in \boldsymbol{S}$, there is a maximum open interval (α, β), where $\alpha, \beta \in \mathbb{R}\cup\{-\infty, +\infty\}$, such that the following hold:*

(i) $0 \in (\alpha, \beta)$, and
(ii) there is a solution $x : (\alpha, \beta) \to \boldsymbol{S}$ of the ode $\dot{x} = f(x)$ such that $x(0) = x_0$.

PROOF. Exercise. □

Next we see how a solution curve $y(t)$ behaves close to the limits of its domain. We include only the result for the right endpoint of the interval of definition of $y(t)$. As we will show, if the domain of $y(t)$ cannot be extended, then $y(t)$ "leaves" any compact set in $\boldsymbol{S}$.

THEOREM 2.1.21. *Let $f : \boldsymbol{S} \to \mathbb{R}^n$ be C^1 and $\beta \in \mathbb{R}$. If $y : (\alpha, \beta) \to \boldsymbol{S}$ is a solution of $\dot{x} = f(x)$ on the maximal open interval (α, β), then for every compact $K \subseteq \boldsymbol{S}$, there is $t \in (\alpha, \beta)$ such that $y(t) \notin K$.*



PROOF. We fix some compact subset K of $\boldsymbol{S}$, and we suppose that

$$\forall_{t\in(\alpha,\beta)}\big(y(t) \in K\big).$$

Since $f_{|K}$ is continuous, there is some $M > 0$ such that $\forall_{x\in K}\big(|f(x)| \leq M\big)$. Next we show that y is Lipschitz with Lipschitz constant M. If $s, t \in (\alpha, \beta)$ such that $s < t$, then

$$\begin{aligned} |y(s) - y(t)| &= \left| \int_s^t \dot{y}(z)dz \right| \\ &\leq \int_s^t \big|\dot{y}(z)\big| dz \\ &= \int_s^t \big|f(y(z))\big| dz \\ &\leq M(t - s) \\ &= M|t - s|. \end{aligned}$$

Since y is uniformly continuous, and (α, β) is dense in $(\alpha, \beta]$, y can be extended[2] to a uniformly continuous function $y^* : (\alpha, \beta] \to \mathbb{R}^n$. Actually, we have that[3]

[2]Here we use the following standard fact: If D is a dense subset of a metric space X, and $f : D \to Y$ is a uniformly continuous function from D to a complete metric space Y, then f is extended to a uniformly continuous function $f^* : X \to Y$.

[3]This follows from the above result, if we take $y : (\alpha, \beta) \to K$, where K is complete, as a closed subset of the complete space $\mathbb{R}^n$.

$y^* : (\alpha, \beta] \to K$. Next we show that y^* is differentiable at β. If $\gamma \in (\alpha, \beta)$, then taking the limit $t \to \beta$ on both sides of the equation

$$y(t) = y(\gamma) + \int_\gamma^t \dot{y}(z)dz,$$

and since $y(\alpha, \beta) \subseteq K \subseteq \boldsymbol{S}$, we get

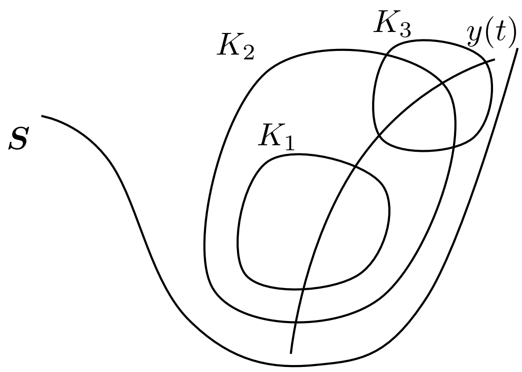
$$\begin{aligned} y^*(\beta) &= y(\gamma) + \lim_{t\to\beta} \int_\gamma^t \dot{y}(z)dz \\ &= y(\gamma) + \lim_{t\to\beta} \int_\gamma^t f(y(z))dz \\ &= y(\gamma) + \int_\gamma^\beta f(y(z))dz. \end{aligned}$$

hence for every $t \in [\gamma, \beta]$ we have that

$$y^*(t) = y(\gamma) + \int_\gamma^t f(y(z))dz.$$

Hence y^* is differentiable also at β and $(y^*)'(\beta) = f(y(\beta))$. Therefore, y^* is a solution on $[\gamma, \beta]$. Since by Theorem 2.1.11 there is a solution on an interval around β, there is a solution on some interval $[\beta, \delta)$, where $\delta > \beta$. But then we can extend y to the interval (α, δ), which contradicts the maximality of (α, β). $\square$

By Theorem 2.1.21 we have that when t approaches β, then $y(t)$ approaches the boundary of $\boldsymbol{S}$, or $|y(t)|$ tends to $+\infty$.



Corollary 2.1.22. *Let $K \subseteq \boldsymbol{S}$ be compact, $x_0 \in K$, and let $f : \boldsymbol{S} \to \mathbb{R}^n$ be C^1. Suppose that every solution $x : [0, \beta] \to \boldsymbol{S}$ with $x(0) = x_0$ satisfies the property*

$$\forall_{t\in[0,\beta]}\big(x(t) \in K\big).$$

Then there is a solution $x^* : [0, +\infty) \to \boldsymbol{S}$ with $x^*(0) = x_0$ and

$$\forall_{t\geq 0}\big(x^*(t) \in K\big).$$

PROOF. Exercise. □

2.2. The fundamental global theorem of odes

LEMMA 2.2.1. *If $f : \boldsymbol{S} \to \mathbb{R}^n$ is locally Lipschitz and $K \subseteq \boldsymbol{S}$ is compact, then $f_{|K}$ is Lipschitz.*

PROOF. Since f is locally Lipschitz, f is continuous. Let $M > 0$ such that

$$\forall_{x\in K}\big(|f(x)| \leq M\big).$$

Suppose that $f_{|K}$ is not Lipschitz i.e., for every $\sigma > 0$ there are $x, y \in K$ such that

$$|f(x) - f(y)| > \sigma|x - y|.$$

Consequently, for every $n > 0$ there are $x_n, y_n \in K$ such that

$$|f(x_n) - f(y_n)| > n|x_n - y_n|.$$

By compactness of K there is a subsequence $(x_{k_n})_{n=1}^{\infty}$ of $(x_n)_{n=1}^{\infty}$ and some $x' \in K$ such that $x_{k_n} \xrightarrow{n} x'$. If we consider the sequence $(y_{k_n})_{n=1}^{\infty}$ in K, there is a subsequence $(y_{\lambda_{k_n}})_{n=1}^{\infty}$ of $(y_{k_n})_{n=1}^{\infty}$ and some $y' \in K$ such that $y_{\lambda_{k_n}} \xrightarrow{n} y'$. Clearly, $x_{\lambda_{k_n}} \xrightarrow{n} x'$ too. We define

$$x_n{}' := x_{\lambda_{k_n}}, \quad y_n{}' := y_{\lambda_{k_n}}, \quad n > 0.$$

Hence

$$\begin{aligned}|x_n{}' - y_n{}'| &= \big|x_{\lambda_{k_n}} - y_{\lambda_{k_n}}\big| \\ &< \frac{1}{\lambda(k(n))}\big|f(x_{\lambda_{k_n}}) - f(y_{\lambda_{k_n}})\big| \\ &< \frac{1}{n}\big|f(x_n{}') - f(y_n{}')\big|.\end{aligned}$$

Taking limits we have that

$$\begin{aligned}0 &\leq \lim_{n\to\infty} |x_n{}' - y_n{}'| \\ &\leq \lim_{n\to\infty}\left(\frac{1}{n}\big|f(x_n{}') - f(y_n{}')\big|\right) \\ &\leq \lim_{n\to\infty}\left(\frac{1}{n}\big(|f(x_n{}')| + |f(y_n{}')|\big)\right) \\ &\leq \lim_{n\to\infty}\frac{2M}{n} \\ &= 0.\end{aligned}$$

Since

$$\begin{aligned}0 &\leq \big||x_n{}' - y_n{}'| - |x' - y'|\big| \\ &\leq |(x_n{}' - y_n{}') - (x' - y')| \\ &\leq |x_n{}' - x'| + |y_n{}' - y'|,\end{aligned}$$

we get $|x' - y'| = \lim_{n\to\infty} |x_n{}' - y_n{}'| = 0$ i.e., $x' = y'$. Since f is locally Lipschitz, there is some neighborhood $V_{x'}$ of x' in $\boldsymbol{S}$ such that $f_{|V_{x'}}$ has Lipschitz constant σ. Since $x_n{}' \xrightarrow{n} x'$ and $y_n{}' \xrightarrow{n} x'$, there is some $n_0 \in \mathbb{N}^+$ such that for every $n > n_0$ $|f(x_n{}') - f(y_n{}')| \leq \sigma |x_n{}' - y_n{}'|$. If $n > \sigma$, hence $\lambda(k(n)) > \sigma$, we get

$$\begin{aligned} \sigma \big| x_{\lambda_{k_n}} - y_{\lambda_{k_n}} \big| &< \lambda(k(n)) \big| x_{\lambda_{k_n}} - y_{\lambda_{k_n}} \big| \\ &< \big| f(x_{\lambda_{k_n}}) - f(y_{\lambda_{k_n}}) \big| \\ &\leq \sigma \big| x_{\lambda_{k_n}} - y_{\lambda_{k_n}} \big|, \end{aligned}$$

which is a contradiction.

□

LEMMA 2.2.2. *Let $y : [t_0, t_1] \to \boldsymbol{S}$ be continuous.*
(i) There exists $\epsilon_0 > 0$ such that for every $x \in \mathbb{R}^n$ the following implication holds

$$\exists_{t\in[t_0,t_1]} \big(|x - y(t)| \leq \epsilon_0\big) \Rightarrow x \in \boldsymbol{S}.$$

(ii) If for this ϵ_0 we define the set

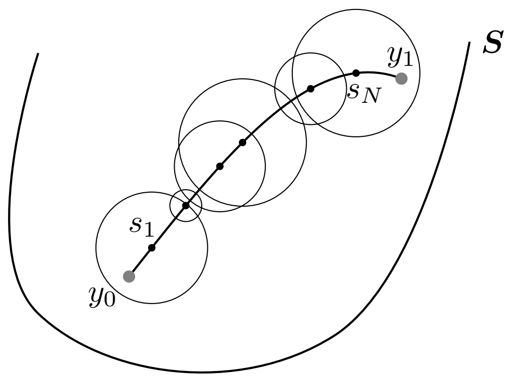
$$K_{\epsilon_0} := \big\{x \in \mathbb{R}^n \mid \exists_{t\in[t_0,t_1]} \big(|x - y(t)| \leq \epsilon_0\big)\big\},$$

then K_{ϵ_0} is a compact subset of $\boldsymbol{S}$.

PROOF. (i) If $t \in [t_0, t_1]$, $y(t) \in \boldsymbol{S}$, and since $\boldsymbol{S}$ is open, there is $\epsilon_t > 0$ with $\mathcal{B}\big(y(t), \epsilon_t\big) \subseteq \boldsymbol{S}$. Since y is continuous, $y^{-1}\big[\mathcal{B}\big(y(t), \frac{\epsilon_t}{2}\big)\big]$ is open in $[t_0, t_1]$. Clearly,

$$[t_0, t_1] = \bigcup_{t\in[t_0,t_1]} y^{-1}\left[\mathcal{B}\left(y(t), \frac{\epsilon_t}{2}\right)\right].$$

By the compactness[4] of the closed interval $[t_0, t_1]$ there are $s_1, \ldots, s_N \in [t_0, t_1]$, for some $N \in \mathbb{N}^+$, such that



[4]Here we use the theorem of Heine-Borel, according to which, a subset K of $\mathbb{R}^n$ is compact if and only if every open covering of K has a finite subcover. In the figure $y_0 = y(t_0)$ and $y_1 = y(t_1)$.

$$[t_0, t_1] = \bigcup_{j=1}^{n} y^{-1}\left[\mathcal{B}\left(y(s_j), \frac{\epsilon_{s_j}}{2}\right)\right].$$

We define

$$\epsilon_0 := \min\left\{\frac{\epsilon_{s_1}}{2}, \ldots, \frac{\epsilon_{s_N}}{2}\right\}.$$

Let $x \in \mathbb{R}^n$ and $t \in [t_0, t_1]$ such that $|x - y(t)| \leq \epsilon_0$. For this t there is some $j \in \{1, \ldots, N\}$ such that

$$\begin{aligned} t \in y^{-1}\left[\mathcal{B}\left(y(s_j), \frac{\epsilon_{s_j}}{2}\right)\right] &\Leftrightarrow y(t) \in \mathcal{B}\left(y(s_j), \frac{\epsilon_{s_j}}{2}\right) \\ &\Leftrightarrow |y(t) - y(s_j)| < \frac{\epsilon_{s_j}}{2}. \end{aligned}$$

Hence

$$\begin{aligned} |x - y(s_j)| &\leq |x - y(t)| + |y(t) - y(s_j)| \\ &< \epsilon_0 + \frac{\epsilon_{s_j}}{2} \\ &\leq \frac{\epsilon_{s_j}}{2} + \frac{\epsilon_{s_j}}{2} \\ &= \epsilon_{s_j}, \end{aligned}$$

i.e., $x \in \mathcal{B}\big(y(s_j), \epsilon_{s_j}\big) \subseteq \boldsymbol{S}$.

(ii) By case (i) we get $K_{\epsilon_0} \subseteq \boldsymbol{S}$. Next we show that K_{ϵ_0} is bounded. Let $M > 0$ such that $|y(t)| \leq M$, for every $t \in [t_0, t_1]$. If $x, x' \in K_{\epsilon_0}$, there are $t, t' \in [t_0, t_1]$ such that $|x - y(t)| \leq \epsilon_0$ and $|x' - y(t')| \leq \epsilon_0$. Hence

$$\begin{aligned} |x - x'| &\leq |x - y(t)| + |y(t) - y(t')| + |y(t') - x'| \\ &\leq \epsilon_0 + |y(t) - y(t')| + \epsilon_0 \\ &\leq \epsilon_0 + 2M + \epsilon_0. \end{aligned}$$

Next we show that K_{ϵ_0} is closed. If $x_0 \in \overline{K_{\epsilon_0}}$, where $\overline{K_{\epsilon_0}}$ is the closure of K_{ϵ_0}, we show that $x_0 \in K_{\epsilon_0}$. If $\epsilon > 0$, there is some $x \in K_{\epsilon_0}$ such that $|x - x_0| < \epsilon$. If $t \in [t_0, t_1]$ such that $|x - y(t)| \leq \epsilon_0$, we get

$$|x_0 - y(t)| \leq |x_0 - x| + |x - y(t)| \leq \epsilon + \epsilon_0$$

i.e., we showed that

$$\forall_{\epsilon > 0} \exists_{t \in [t_0, t_1]} \big(|x_0 - y(t)| \leq \epsilon + \epsilon_0\big).$$

Suppose that $x_0 \notin K_{\epsilon_0}$ i.e.,

$$\forall_{t \in [t_0, t_1]} \big(|x_0 - y(t)| > \epsilon_0\big).$$

We define the function $\rho : [t_0, t_1] \to (0, +\infty)$ by

$$\rho(t) := |x_0 - y(t)| - \epsilon_0.$$

Since ρ is continuous, it attains its minimum value μ at some point $s \in [t_0, t_1]$ i.e.,

$$\rho(t) \geq \rho(s) = |x_0 - y(s)| - \epsilon_0 = \mu > 0.$$

Since $\mu > 0$, there is some $t' \in [t_0, t_1]$ such that $|x_0 - y(t')| \leq \frac{\mu}{2} + \epsilon_0$, hence

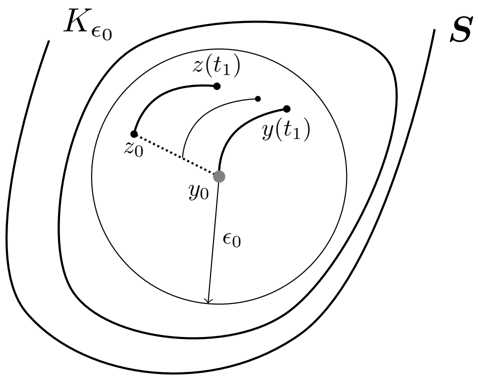
$$\rho(t') = |x_0 - y(t')| - \epsilon_0 \leq \frac{\mu}{2} < \mu = \rho(s),$$

which is a contradiction. Hence $x_0 \in K_{\epsilon_0}$. □

Next follows a stronger form of the continuity of solutions in initial conditions. While in Theorem 2.1.15 both solutions were *assumed* to be defined on the same interval $[t_0, t_1]$, in the theorem that follows the solutions starting at nearby points will be *shown* that they are defined on the same interval $[t_0, t_1]$ and remain close to each other in $[t_0, t_1]$. Moreover, f is not assumed to be Lipschitz.

THEOREM 2.2.3 (Fundamental global theorem of odes). *Let $f : \boldsymbol{S} \to \mathbb{R}^n$ be C^1 and $y : [t_0, t_1] \to \boldsymbol{S}$ a solution of $\dot{x} = f(x)$ with $y(t_0) = y_0$. There is a neighborhood $V_{y_0} \subseteq \boldsymbol{S}$ of y_0 and there is a constant $\sigma > 0$ such that for every $z_0 \in V_{y_0}$ there is a unique solution $z : [t_0, t_1] \to \boldsymbol{S}$ of $\dot{x} = f(x)$ with $z(t_0) = z_0$ and*

$$\forall_{t \in [t_0, t_1]} \Big(|y(t) - z(t) \leq |y_0 - z_0| e^{\sigma(t - t_0)} \Big).$$



PROOF. Since $y : [t_0, t_1] \to \boldsymbol{S}$ is continuous, let ϵ_0 and K_{ϵ_0} as in Lemma 2.2.2. By the definition of K_{ϵ_0} we get immediately that $\forall_{t \in [t_0, t_1]} \big(y(t) \in K_{\epsilon_0} \big)$. Since K_{ϵ_0} is a compact subset of $\boldsymbol{S}$, by Lemma 2.2.1 the function $f_{|K_{\epsilon_0}}$ has Lipschitz constant σ, for some $\sigma > 0$. There is $\delta > 0$ such that

$$\delta \leq \epsilon_0 \quad \text{and} \quad \delta e^{\sigma(t_1 - t_0)} \leq \epsilon_0.$$

We define

$$V_{y_0} := \mathcal{B}(y_0, \delta),$$

and we show that if $z_0 \in V_{y_0}$, there is a unique solution $z : [t_0, t_1] \to \boldsymbol{S}$ of $\dot{x} = f(x)$ with $z(t_0) = z_0$. Since

$$|z_0 - y_0| = |z_0 - y(t_0)| < \delta \leq \epsilon_0,$$

we get $z_0 \in K_{\epsilon_0}$, hence

$$V_{y_0} \subseteq K_{\epsilon_0} \subseteq \boldsymbol{S}.$$

By Theorem 2.1.11 there is a solution $z(t)$ through z_0 defined on a maximal interval $[t_0, \beta)$, for some $\beta \in \mathbb{R} \cup \{+\infty\}$. We show that $\beta > t_1$, where $>$ is here the ordering of the extended reals. Suppose that $\beta \leq t_1$, and let $t \in [t_0, \beta)$. Then by Theorem 2.1.15 on $f_{V_{y_0}}$, which is σ-Lipschitz, and on the solutions y and z defined on the common interval $[t_0, s]$, where $t < s < \beta$, we get

$$\begin{aligned} |z(t) - y(t)| &\leq |z_0 - y_0| e^{\sigma(t - t_0)} \\ &\leq \delta e^{\sigma(t-t_0)} \\ &\leq \epsilon_0. \end{aligned}$$

Hence $z(t)$ lies in K_{ϵ_0}. By Theorem 2.1.21 the interval $[t_0, \beta)$ cannot be a maximal solution domain, which contradicts our hypothesis. Therefore, $\beta > t_1$. Since now $z : [t_0, \beta) \to \boldsymbol{S}$ and $[t_0, t_1] \subset [t_0, \beta)$, we conclude that z is defined on $[t_0, t_1]$. The inequality

$$|y(t) - z(t) \leq |y_0 - z_0| e^{\sigma(t-t_0)}$$

for every $t \in [t_0, t_1]$ follows from Theorem 2.1.15, and the uniqueness of the solution z on $[t_0, t_1]$ follows from Lemma 2.1.19. □

Hence, if $f : \boldsymbol{S} \to \mathbb{R}^n$ is C^1 and $y : [t_0, t_1] \to \boldsymbol{S}$ is a solution of $\dot{x} = f(x)$, then for all z_0 sufficiently close to $y_0 = y(t_0)$ there is a unique solution on $[t_0, t_1]$ starting at z_0. If we write

$$z(t) = \phi(t, z_0), \quad y(t) = \phi(t, y_0),$$

then $z_0 = \phi(0, z_0)$ and $y_0 = \phi(0, y_0)$, and by Theorem 2.2.3

$$\lim_{z_0 \to y_0} \phi(t, z_0) = \phi(t, y_0)$$

uniformly on $[t_0, t_1]$ i.e., the solution through z_0 "depends continuously" on z_0.

2.3. The flow of an ode

Definition 2.3.1. If $f : \boldsymbol{S} \to \mathbb{R}^n$ is C^1, and since for every $u \in \boldsymbol{S}$ there is a unique solution $x_u : J(u) \to \boldsymbol{S}$ of the ode $\dot{x} = f(x)$ such that $x_u(0) = u$ and $J(u)$ is the maximal open interval of u, we define the set

$$\Omega := \big\{(t, u) \in \mathbb{R} \times \boldsymbol{S} \mid t \in J(u), u \in \boldsymbol{S}\big\},$$

and the function $\phi : \Omega \to \boldsymbol{S}$,

$$\phi(t, u) := x_u(t) =: \phi_t(u),$$

for every $(t, u) \in \Omega$, which is called the *flow* of the ode $\dot{x} = f(x)$.

Note that since $0 \in J(u)$, for every $u \in \boldsymbol{S}$, we have that $\{0\} \times \boldsymbol{S} \subset \Omega$, and

$$\phi(0, u) = x_u(0) = u.$$

PROPOSITION 2.3.2. *Let $s, t \in \mathbb{R}$, $u \in \boldsymbol{S}$ and ϕ the flow of $\dot{x} = f(x)$, for some C^1 function $f : \boldsymbol{S} \to \mathbb{R}^n$. The following hold:*

(i) If $t \in J(u)$ and $s \in J(\phi_t(u))$, then $s + t \in J(u)$ and $\phi_{s+t}(u) = \phi_s(\phi_t(u))$.

(ii) If $s + t \in J(u)$, then $t \in J(u)$, $s \in J(\phi_t(u))$ and $\phi_{s+t}(u) = \phi_s(\phi_t(u))$.

PROOF. We show only (i), and we consider the case $s, t > 0$. The other cases are shown similarly. Let $J(u) = (\alpha, \beta)$ and let $t \in J(u)$ i.e. $\alpha < t < \beta$, where $<$ is the ordering of the extended reals. We show that $s + t \in J(u) \Leftrightarrow \alpha < s + t < \beta$. Since $t > 0$, $\alpha < s + t$, hence it remains to show that $s + t < \beta$. We define the function $y : (\alpha, s + t] \to \boldsymbol{S}$ by

$$y(r) := \begin{cases} \phi(r, u) & , \text{ if } \alpha < r \leq t \\ \phi(r - t, \phi_t(u)) & , \text{ if } t \leq r \leq s + t. \end{cases}$$

Note that y is continuous at t, since $\phi(t - t, \phi_t(u)) = \phi(0, \phi_t(u)) = \phi_t(u)$, hence y is continuous on $(\alpha, s + t]$. Moreover, u is a solution curve on $(\alpha, s + t]$. If $\alpha < r \leq t$, then

$$\dot{y}(r) = \dot{\phi}(r, u) = \dot{x}_u(r) = f(x_0(r)) = f(\phi(r, u)) = f(y(r)).$$

If $t \leq r \leq s + t$, and $s(r) := r - t$, then

$$\begin{aligned}
\dot{y}(r) &= \frac{d}{dr}\big[\phi(r - t, \phi_t(u))\big] \\
&= \frac{d}{dr}\big[x_{\phi_t(u)}(r - t)\big] \\
&= \frac{d}{ds}\big[x_{\phi_t(u)}(s)\big]\frac{ds}{dr} \\
&= \frac{d}{ds}\big[x_{\phi_t(u)}(s)\big] \\
&= f\big(x_{\phi_t(u)}(s)\big) \\
&= f\big(x_{\phi_t(u)}(r - t)\big) \\
&= f(y(r)).
\end{aligned}$$

Since $0 \in (\alpha, \beta)$ and $\alpha < 0 < t$, by our hypothesis on t, then by the definition of y, we get

$$y(0) = \phi(0, u) = x_u(0) = u.$$

Hence the maximal open interval $J(u)$ must include $(\alpha, s + t]$, i.e., $s + t < \beta$. By the uniqueness of solutions on $(\alpha, s + t]$ that agree on $0 \in (\alpha, s + t]$ (Lemma 2.1.19) and the definition of y we get

$$\begin{aligned}
\phi_{s+t}(u) &= \phi(s + t, u) \\
&= x_u(s + t)
\end{aligned}$$

$$\begin{aligned}
&= y(s+t) \\
&= \phi((s-t)-t, \phi_t(u)) \\
&= \phi(s, \phi_t(u)) \\
&= \phi_s(\phi_t(u)).
\end{aligned}$$

□

THEOREM 2.3.3. *If* Ω *and* ϕ *are as in Definition 2.3.1, then*
(i) Ω *is an open subset of* $\mathbb{R} \times \boldsymbol{S}$*, and*
(ii) ϕ *is continuous.*

PROOF. Left to the reader. □

Bibliography

[1] D. Amir: *Characterizations of inner product spaces*, Springer Basel, 1986.

[2] E. Bishop: *Foundations of Constructive Analysis*, McGraw-Hill, 1967.

[3] O. Forster: *Analysis I, Differential- und Integralrechnung einer Veränderlichen*, Vieweg+Teubner, 2011.

[4] O. Forster: *Analysis II, Differentialrechnung im* $\mathbb{R}^n$*, Gewöhnliche Diferentialgleichungen*, Vieweg+Teubner, 2008.

[5] P. Hartmann: *Ordinary Differential Equations*, Society for Industrial and Applied Mathematics, 2002.

[6] M. Hirsch, S. Smale: *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, 1974.

[7] S. Lang: *Undergraduate Analysis*, Springer, 1983.

[8] S. Lang: *Linear Algebra*, Springer, 1987.

[9] B. V. Limaye: *Linear Functional Analysis for Scientists and Engineers*, Springer, 2016

[10] L. Maligranda: Some remarks on the triangle inequality for norms, Banach J. Math. Anal. 2, 2008, no. 2, 31-41.

[11] N. G. Markley: *Principles of Differential Equations*, Wiley-Interscience, 2004.

[12] M. Simmons: *Foundations of Topology and Modern Analysis*, McGraw-Hill .