MATH 3151 Analysis II, Spring 2017

Arthur J. Parzygnat

These are my personal notes. This is *not* a substitute for our textbooks. You will *not* be responsible for any **Remark**s in these notes. However, everything else, including what is in your textbooks (even if it's not here), is fair game for homework, quizzes, and exams. At the end of each lecture, I provide a list of homework problems that can be done after that lecture. I also provide additional exercises which I believe are good to know. You should also browse other books and do other problems as well to get better at writing proofs and understanding the material.

Contents

Acknowledgements.

The author has benefited from discussions with Keith Conrad, Lan-Hsuan Huang, Benjamin Russo, Ralf Schiffler,

1 January 19: Linear Algebra Review

I don't think there's a canonical method to teach Analysis II. There are several possibilities one can follow after learning the basics of sequences, convergence, functions, continuity, differentiation, and integration in $\mathbb{R}$. These possibilities are as follows.

1. Multivariable calculus: functions on Euclidean space, differentiation, implicit function theorem, integration, Fubini's theorem.

2. Differential topology: manifolds, Sard's theorem, differential forms, Stoke's theorem, de Rham cohomology, vector fields, cobordism theory, Hopf's theorem.

3. Metric spaces: notion of distance, contraction mapping theorem, fractals.

4. Functional analysis: Banach spaces, Hilbert spaces, Fourier analysis, convex analysis, operator theory, C^*-algebras, abstract probability theory.

5. Measure theory: measurable spaces, measure spaces, the Lebesgue integral.

Personally, my favorite are options 2 and 4 though 2 has significant overlap with a course on differential geometry. 5 is somewhat technical and potentially incredibly boring, but crucial for any serious analyst. 1 is pretty standard but not enough for a full semester. 3 seems interesting though I have less experience with it. I've decided to begin with 1 with a slight emphasis on some concepts from 2. Most of these topics are explored in Spivak's book [13] though a few in Milnor's book [8]. This will probably occupy us for about a little over one month. Then, we will proceed to parts of items 3 and 4, following Kolmogorov and Fomin's book [7] along with other sources. To avoid many technicalities, when we get to item 4, we might only focus on operators on finite-dimensional Hilbert spaces. This does a great disservice to the topic as much of the reason for its study is precisely due to the subtleties in infinite dimensions. However, there are already many wonderful applications of these ideas in probability theory, non-commutative geometry, and quantum mechanics, even at the level of finite-dimensional Hilbert spaces. I hope we will explore at least some of these applications rather than being held up by technical details, which you will eventually work through if you study further or specialize in any of these areas.

That being said, we will begin with multivariable calculus after having first reviewed some concepts from linear algebra. But even before that, we will recall the definition of a cartesian product from a different perspective that will become more and more useful as we progress in our studies. Warning: the following definitions may seem very foreign but will be used constantly throughout the course.

Definition 1.1. A *(finite) diagram of sets* consists of a finite collection of sets $\mathcal{X} := \{X_1, \ldots, X_n\}$ together with a collection of a finite set of functions $\mathcal{F} := \{f_{ij}^{\alpha} : X_j \to X_i\}_{\alpha \in A_{ij}}$, where A_{ij} is a finite set, for each $i, j \in \{1, \ldots, n\}$. A diagram may be denoted by $D := (\mathcal{X}, \mathcal{F})$.

Diagrams are often drawn explicitly such as

$$\begin{tikzcd}
& X_1 \arrow[dl, shift left] \arrow[dr] & & X_5 & \\
X_2 \arrow[ur, shift left] \arrow[ddr] & & X_4 \arrow[ddl] & & \\
& & & & X_6 \\
& X_3 & & &
\end{tikzcd}$$

Definition 1.2. Let $D = (\mathcal{X}, \mathcal{F})$ be a (finite) diagram of sets. A <u>*path*</u> in D consists of an ordered subset of functions $\gamma := \{f_1, \ldots, f_m\}$ of $\mathcal{F}$ such that the source of f_i equals the target of f_{i-1} for all $i \in \{2, \ldots, m\}$.

A path in the diagram drawn above may be depicted as the set of thick red arrows here.

$$\begin{tikzcd}
& X_1 \arrow[dl, shift left] \arrow[dr, red, thick] & & X_5 & \\
X_2 \arrow[ur, shift left] \arrow[ddr] & & X_4 \arrow[ddl, red, thick] & & \\
& & & & X_6 \\
& X_3 & & &
\end{tikzcd}$$

Definition 1.3. A diagram D <u>*commutes*</u> iff for every two paths $\gamma := \{f_1, \ldots, f_m\}$ and $\delta := \{g_1, \ldots, g_k\}$ in D with the source of f_1 and g_1 being the same and the targets of f_m and g_k being the same, the compositions are equal, namely

$$f_m \circ \cdots \circ f_1 = g_k \circ \cdots \circ g_1. \tag{1.4}$$

Two paths in the diagram drawn above are depicted as a set of thick red and blue arrows.

$$\begin{tikzcd}
& X_1 \arrow[dl, shift left, blue, thick] \arrow[dr, red, thick] & & X_5 & \\
X_2 \arrow[ur, shift left] \arrow[ddr, blue, thick] & & X_4 \arrow[ddl, red, thick] & & \\
& & & & X_6 \\
& X_3 & & &
\end{tikzcd}$$

The diagram commutes when both compositions are equal as functions for every such pair of paths. Another example consists of the two paths

$$\begin{tikzcd}
& X_1 \arrow[dl, shift left] \arrow[dr, red, thick] & & X_5 & \\
X_2 \arrow[ur, shift left, red, thick] \arrow[ddr, blue, thick] & & X_4 \arrow[ddl, red, thick] & & \\
& & & & X_6 \\
& X_3 & & &
\end{tikzcd}$$

Definition 1.5. Let X and Y be two sets. A <u>*Cartesian product*</u> of X and Y is a set U together with functions $\pi_X : U \to X$ and $\pi_Y : U \to Y$

$$\begin{tikzcd} & U \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_Y"] & \\ X & & Y \end{tikzcd} \tag{1.6}$$

satisfying the following ("universal") property. For any other set V equipped with functions $\nu_X : V \to X$ and $\nu_Y : V \to Y$

$$\begin{tikzcd} & V \arrow[dl, "\nu_X"'] \arrow[dr, "\nu_Y"] & \\ X & & Y \end{tikzcd} \tag{1.7}$$

there exists a unique function $h : V \to U$ (drawn as a dashed arrow) such that the diagram

$$\begin{tikzcd} & V \arrow[d, dashed, "h"] \arrow[ddl, bend right, "\nu_X"'] \arrow[ddr, bend left, "\nu_Y"] & \\ & U \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_Y"] & \\ X & & Y \end{tikzcd} \tag{1.8}$$

commutes, i.e.

$$\pi_X \circ h = \nu_X \qquad \text{and} \qquad \pi_Y \circ h = \nu_Y. \tag{1.9}$$

The functions $\pi_X : U \to X$ and $\pi_Y : U \to Y$ are called <u>*projections*</u> onto their respective factors.

Before we relate this definition to one that you might be familiar with, notice what the universal property is saying. First, for any set Z and a function $f : Z \to U$, one automatically obtains functions $\pi_X \circ f : Z \to X$ and $\pi_Y \circ f : Z \to Y$, i.e. functions onto the projections. Conversely, (and this is what the universal property says), given any two functions $f_X : Z \to X$ and $f_Y : Z \to Y$ onto the two factors, this uniquely determines a function $f : Z \to U$ such that $\pi_X \circ f = f_X$ and $\pi_Y \circ f = f_Y$. This is how we think of the product of two sets in terms of the characteristic properties that it satisfies rather than any particular construction, as in the following.

Theorem 1.10. *Given any two sets X and Y, a Cartesian product (U, π_X, π_Y) exists and is unique in the following sense. For any other Cartesian product (V, ν_X, ν_Y) of X and Y, there exist functions $f : U \to V$ and $g : V \to U$ such that*

(a) f and g are inverses of each other and

(b) $\nu_X \circ f = \pi_X$ and $\nu_Y \circ f = \pi_Y$.

Note that from this it follows that $\nu_X = \pi_X \circ g$ and $\nu_Y = \pi_Y \circ g$.

Proof. There are two things to prove: existence and uniqueness.
(*Proof of existence.*) Set U to be the set of ordered pairs

$$U := \{(x, y) \in X \times Y\} \tag{1.11}$$

and the projections π_X, π_Y to be $\pi_X(x, y) := x$ and $\pi_Y(x, y) = y$ for all $(x, y) \in X \times Y$. To see that U satisfies the universal property in Definition 1.5, let V be any other set equipped with functions $\nu_X : V \to X$ an $\nu_Y : V \to Y$. Then, define $h : V \to U$ by the assignment

$$V \ni v \mapsto h(v) := \big(\nu_X(v), \nu_Y(v)\big). \tag{1.12}$$

With this definition, $\pi_X \circ h = \nu_X$ and $\pi_Y \circ h = \nu_Y$ and in fact h is the only such function that satisfies this condition.
(*Proof of uniqueness.*) Let (V, ν_X, ν_Y) be another Cartesian product. Since (U, ν_X, ν_Y) satisfies the universal property, there exists a unique $g : V \to U$ such that $\nu_X = \pi_X \circ g$ and $\nu_Y = \pi_Y \circ g$. Since (V, ν_X, ν_Y) satisfies the universal property, there exists a unique $f : U \to V$ such that $\nu_X \circ f = \pi_X$ and $\nu_Y \circ f = \pi_Y$. Consider the compositions $f \circ g : V \to V$ and $g \circ f : U \to U$. Again by the universal property, there must be unique functions $h_V : V \to V$ and $h_U : U \to U$ such that $\nu_X \circ h_V = \nu_X$ and $\nu_Y \circ h_V = \nu_Y$ and similarly for h_U. The only such functions that achieve these conditions are the identity functions $\mathrm{id}_V : V \to V$ and $\mathrm{id}_U : U \to U$. Hence, $f \circ g = \mathrm{id}_V$ and $g \circ f = \mathrm{id}_U$. ■

Thus, it makes sense to say "the" Cartesian product of X and Y in a certain sense made precise by the previous theorem. Furthermore, it is common to denote this product by $X \times Y$. Part of the annoyance comes when defining the product of three sets X, Y, and Z. One may choose $(X \times Y) \times Z$ or $X \times (Y \times Z)$ and obtain different sets for example. The first has elements of the form $\big((x, y), z\big)$ and the second of the form $\big(x, (y, z)\big)$. Of course, this difference is irrelevant and we often denote "the" element by (x, y, z) in "the" set denoted by $X \times Y \times Z$.

Remark 1.13. Similarly, given any collection of sets $\{X_\alpha\}_{\alpha \in A}$ indexed by some set A, the Cartesian product of these sets is a set, denoted by

$$\prod_{\alpha \in A} X_\alpha \tag{1.14}$$

together with functions

$$\pi_\beta : \prod_{\alpha \in A} X_\alpha \to X_\beta \tag{1.15}$$

for each $\beta \in A$ satisfying a universal property analogous to ones mentioned above. However, we can no longer use the construction in terms of ordered tuples because A might be an infinite set, in fact uncountable. If we assume the Axiom of Choice, one possible construction of the Cartesian product is

$$\prod_{\alpha \in A} X_\alpha := \left\{ f : A \to \bigcup_{\alpha \in A} X_\alpha \ : \ f(\beta) \in X_\beta \text{ for all } \beta \in A \right\}. \tag{1.16}$$

Given an element $\beta \in A$ and a function $f : A \to \bigcup_{\alpha \in A} X_\alpha$, one can evaluate the function f on β providing an element $f(\beta) \in X_\beta$. This evaluation is precisely the projection, $\pi_\beta := \mathrm{ev}_\beta$, where

$\text{ev}_\beta(f) := f(\beta)$. One can obtain the special case of the product of two sets X_1 and X_2 by setting $A := \{1, 2\}$, the set consisting of the elements labeled "1" and "2." Both satisfy the universal property in the sense described in the following exercise.

Exercise 1.17. Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of sets indexed by some set A. Show that $\prod_{\alpha \in A} X_\alpha$ together with the functions $\text{ev}_\beta : \prod_{\alpha \in A} X_\alpha \to X_\beta$ for all $\beta \in A$ satisfies the universal property of Cartesian products, namely for any other set V together with functions $\{p_\beta : V \to X_\beta\}_{\beta \in A}$, there exists a unique function $h : \prod_{\alpha \in A} X_\alpha \to V$ such that $p_\beta = \text{ev}_\beta \circ h$ for all $\beta \in A$.

Exercise 1.18. Let A and X be two set and let $X_\alpha := X$ for all $\alpha \in A$. Show that $X^A := \{f : A \to X\}$, the set of all functions from A to X, is the Cartesian product $\prod_{\alpha \in A} X_\alpha$.

Definition 1.19. Let $\mathbb{R}^n$ denote n-dimensional Euclidean space, i.e.

$$\mathbb{R}^n := \big\{ (x_1, x_2, \ldots, x_n) \ : \ x_i \in \mathbb{R}, i \in \{1, 2, \ldots, n\} \big\} \equiv \overbrace{\mathbb{R} \times \cdots \times \mathbb{R}}^{n \text{ times}}. \tag{1.20}$$

Elements of $\mathbb{R}^n$ are called <u>*n-component vectors*</u> and occasionally the shorthand notation $\vec{x} := (x_1, x_2, \ldots, x_n)$ will be used to denote such vectors. The binary operation $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \tag{1.21}$$

is called the <u>*addition*</u> of n-component vectors. The function $\mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$c(x_1, x_2, \ldots, x_n) := (cx_1, cx_2, \ldots, cx_n) \tag{1.22}$$

is called the <u>*scalar multiplication*</u>. A <u>*subspace*</u> of $\mathbb{R}^n$ is a subset $V \subseteq \mathbb{R}^n$ such that

$$\vec{0} \in V, \tag{1.23}$$

$$c\vec{v} \in V \qquad \forall\, c \in \mathbb{R},\ \vec{v} \in V \tag{1.24}$$

and

$$\vec{v}_1 + \vec{v}_2 \in V \qquad \forall\, \vec{v}_1, \vec{v}_2 \in V. \tag{1.25}$$

Let $V \subseteq \mathbb{R}^n$ be a subspace. A finite set of vectors $\mathcal{S} := \{\vec{x}_1, \ldots, \vec{x}_m\}$ in V is said to be <u>*linearly dependent*</u> iff there exist real numbers $a_1, \ldots, a_m \in \mathbb{R}$, not all of which are zero, such that

$$\sum_{i=1}^{m} a_i \vec{x}_i = 0. \tag{1.26}$$

Otherwise $\mathcal{S}$ is said to be <u>*linearly independent*</u>. $\mathcal{S}$ <u>*spans*</u> V iff for all $\vec{v} \in V$, there exist numbers $a_1, \ldots, a_m \in \mathbb{R}$ such that

$$\sum_{i=1}^{m} a_i \vec{x}_i = \vec{v}. \tag{1.27}$$

$\mathcal{S}$ is a <u>*basis*</u> for V iff $\mathcal{S}$ spans V and $\mathcal{S}$ is linearly independent. Given any subset $X \subseteq \mathbb{R}^n$,

$$\text{span}(X) := \left\{ \sum_{i=1}^{m} a_i \vec{x}_i \ : \ a_i \in \mathbb{R},\ x_i \in X\ , i \in \{1, \ldots, m\},\ m \in \mathbb{N} \right\}, \tag{1.28}$$

i.e. the set of all linear combinations of vectors in X.

Together with this structure and the zero vector, $\mathbb{R}^n$ is a real vector space. In addition, $\mathbb{R}^n$ has a natural notion of distance and an inner product.

Definition 1.29. The *Euclidean norm/length* on $\mathbb{R}^n$ is the function $\mathbb{R}^n \to \mathbb{R}$ defined by

$$\big\|(x_1, x_2, \ldots, x_n)\big\| := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1.30}$$

The *Euclidean inner product* on $\mathbb{R}^n$ is the function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\big\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\big\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{1.31}$$

Often, the short-hand notation

$$\sum_{i=1}^{n} z_i := z_1 + z_2 + \cdots + z_n \tag{1.32}$$

will be used to denote the sum of n real numbers $z_i \in \mathbb{R}^n, i \in \{1, \ldots, n\}$.

With this structure, $\mathbb{R}^n$ is an inner product space.

Theorem 1.33. *$\mathbb{R}^n$ with these structures satisfies the following for all vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and for all numbers $c \in \mathbb{R}$,*

(a) $\langle \vec{x}, \vec{x}\rangle \geq 0$ and $\langle \vec{x}, \vec{x}\rangle = 0$ if and only if $\vec{x} = 0$,

(b) $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x}\rangle}$,

(c) $\langle \vec{x}, \vec{y}\rangle = \langle \vec{y}, \vec{x}\rangle$,

(d) $\langle c\vec{x}, \vec{y}\rangle = c\langle \vec{x}, \vec{y}\rangle = \langle \vec{x}, c\vec{y}\rangle$,

(e) $\langle \vec{x} + \vec{z}, \vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle + \langle \vec{z}, \vec{y}\rangle$ and $\langle \vec{x}, \vec{y} + \vec{z}\rangle = \langle \vec{x}, \vec{y}\rangle + \langle \vec{y}, \vec{z}\rangle$,

(f) $|\langle \vec{x}, \vec{y}\rangle| \leq \|\vec{x}\|\|\vec{y}\|$ and equality holds if and only if $\vec{x}$ and $\vec{y}$ are linearly dependent (Cauchy-Schwarz inequality),

(g) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (triangle inequality),

(h) $\langle \vec{x}, \vec{y}\rangle = \frac{\|\vec{x}+\vec{y}\|^2 - \|\vec{x}-\vec{y}\|^2}{4}$ (polarization identity).

Proof.

(a) Let $(x_1, \ldots, x_n) := \vec{x}$ denote the components of $\vec{x}$. Then,

$$\langle \vec{x}, \vec{x}\rangle = \sum_{i=1}^{n} x_i^2 \geq 0 \tag{1.34}$$

since the square of any real number is always at least 0. Furthermore, the sum of squares is zero if and only if each term is zero, but the square root of zero is zero so $x_i = 0$ for all i if and only if $\langle \vec{x}, \vec{x}\rangle = 0$.

(b) This follows immediately from the definitions of $\| \cdot \|$ in (1.30) and $\langle \cdot , \cdot \rangle$ in (1.31).

(c) This follows from commutativity of multiplication of real numbers and the formula (1.31).

(d) This follows from commutativity and associativity of multiplication of real numbers and the formula (1.31).

(e) This follows from the distributive law for real numbers and the formula (1.31).

(f) Suppose that $\vec{x}$ and $\vec{y}$ are linearly independent. Therefore, $\vec{x} \neq \lambda \vec{y}$ for all all $\lambda \in \mathbb{R}$. Hence,

$$0 < \|\lambda \vec{y} - \vec{x}\|^2 = \sum_{i=1}^{n} (\lambda y_i - x_i)^2 = \lambda^2 \sum_{i=1}^{n} y_i^2 - 2\lambda \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} x_i^2. \tag{1.35}$$

In particular, this is a quadratic equation in the variable λ that has no real solutions. Hence,

$$\left(-2 \sum_{i=1}^{n} x_i y_i \right)^2 - 4 \left(\sum_{i=1}^{n} y_i^2 \right) \left(\sum_{j=1}^{n} x_i^2 \right) < 0. \tag{1.36}$$

Rewriting this and canceling out the common factor of 4 gives

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_i x_j y_j < \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 y_j^2 \tag{1.37}$$

Applying the square root to both sides gives the desired result.

Now suppose $\vec{x} = \lambda \vec{y}$ for some $\lambda \in \mathbb{R}$. Then by parts (b) and (d), equality holds.

(g) Notice that by the Cauchy Schwarz inequality,

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \sum_{i=1}^{n} (x_i + y_i)^2 \\ &= \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\ &\leq \|\vec{x}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{y}\|^2 \\ &= \big(\|\vec{x}\| + \|\vec{y}\|\big)^2 \end{aligned} \tag{1.38}$$

Applying the square root and using parts (a) and (b) gives the desired result.

(h) This calculation is left to the reader.

■

Exercise 1.39. Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. The <u>*distance*</u> from $\vec{x}$ to $\vec{y}$ is

$$d(\vec{x}, \vec{y}) := \|\vec{y} - \vec{x}\|. \tag{1.40}$$

Prove that d satisfies the following conditions for all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$.

(a) $d(\vec{x},\vec{y}) = 0$ if and only if $\vec{x} = \vec{y}$,

(b) $d(\vec{x},\vec{y}) = d(\vec{y},\vec{x})$, and

(c) $d(\vec{x},\vec{z}) \le d(\vec{x},\vec{y}) + d(\vec{y},\vec{z})$.

Furthermore, using only these three conditions and not the explicit formula for d, prove that $d(\vec{x},\vec{y}) \ge 0$ for all $\vec{x},\vec{y} \in \mathbb{R}^n$.

Definition 1.41. Let $n,m \in \mathbb{N} \cup \{0\}$. A <u>*linear transformation*</u> from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ satisfying

$$T(\vec{x} + c\vec{y}) = T(\vec{x}) + cT(\vec{y}) \tag{1.42}$$

for all $c \in \mathbb{R}$ and $\vec{x},\vec{y} \in \mathbb{R}^n$.

Exercise 1.43. Prove that a function $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is a linear transformation if and only if

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) \tag{1.44}$$

and

$$T(c\vec{y}) = cT(\vec{y}) \tag{1.45}$$

for all $c \in \mathbb{R}$ and $\vec{x},\vec{y} \in \mathbb{R}^n$.

Definition 1.46. Let $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ denote the standard unit vectors of $\mathbb{R}^n$. Namely, $\vec{e}_i$ has 1 in the i-th component and 0 otherwise. Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. The <u>*ij-th matrix coordinate of T*</u> is

$$T_{ij} := \big\langle \vec{e}_i, T(\vec{e}_j) \big\rangle. \tag{1.47}$$

Exercise 1.48. Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. Prove that

$$T(\vec{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} x_j \vec{e}_i, \tag{1.49}$$

where

$$\vec{x} = \sum_{j=1}^{n} x_j \vec{e}_j. \tag{1.50}$$

In particular, show that

$$T(\vec{e}_k) = \sum_{i=1}^{m} T_{ik} \vec{e}_i. \tag{1.51}$$

Proposition 1.52. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ and $\mathbb{R}^n \xleftarrow{S} \mathbb{R}^p$ be linear transformations. The composition $\mathbb{R}^m \xleftarrow{T \circ S} \mathbb{R}^p$ is a linear transformation. Furthermore, its associated ij-th matrix coordinate is given by*

$$(T \circ S)_{ij} = \sum_{k=1}^{n} T_{ik} S_{kj}. \tag{1.53}$$

Proof. The verification of the linearity of $T \circ S$ is left to the reader. By definition of the composition of functions,

$$
\begin{aligned}
(T \circ S)_{ij} &= \big\langle \vec{e}_i, (T \circ S)(\vec{e}_j) \big\rangle \\
&= \big\langle \vec{e}_i, T\big(S(\vec{e}_j)\big) \big\rangle \quad \text{by Definition 1.46} \\
&= \left\langle \vec{e}_i, T\left(\sum_{k=1}^{n} S_{kj} \vec{e}_k \right) \right\rangle \quad \text{by Exercise 1.48} \\
&= \sum_{k=1}^{n} S_{kj} \big\langle \vec{e}_i, T(\vec{e}_k) \big\rangle \quad \text{by linearity of } T \text{ and by Theorem 1.33} \\
&= \sum_{k=1}^{n} \sum_{l=1}^{m} S_{kj} T_{lk} \big\langle \vec{e}_i, \vec{e}_l \big\rangle \quad \text{by similar arguments} \\
&= \sum_{k=1}^{n} S_{kj} T_{ik}.
\end{aligned}
\tag{1.54}
$$

■

This motivates the following notation for any linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$. The <u>*$m \times n$ matrix*</u> associated to T is the array

$$
\begin{bmatrix}
T_{11} & T_{12} & \cdots & T_{1n} \\
T_{21} & T_{22} & \cdots & T_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
T_{m1} & T_{m2} & \cdots & T_{mn}
\end{bmatrix}
\tag{1.55}
$$

This notation is convenient if we also view an n-component vector $(x_1, x_2, \ldots, x_n)$ as an $n \times 1$ matrix

$$
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
\tag{1.56}
$$

and view the linear transformation T acting on $\vec{x}$ in terms of <u>*matrix multiplication*</u>, analogous to the expression on the right-hand-side of Equation (1.53).

Definition 1.57. The <u>*determinant*</u> for $m \times m$ matrices is a function

$$
\det : \overbrace{\mathbb{R}^m \times \cdots \mathbb{R}^m}^{m \text{ times}} \to \mathbb{R}
\tag{1.58}
$$

satisfying the following conditions.

(a) For every m-tuple of vectors $(\vec{v}_1, \ldots, \vec{v}_m)$ in $\mathbb{R}^m$,

$$
\det\Big(\vec{v}_1, \ldots, \overset{\text{switch}}{\vec{v}_i, \ldots, \vec{v}_j}, \ldots, \vec{v}_m\Big) = -\det\Big(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_i, \ldots, \vec{v}_m\Big).
\tag{1.59}
$$

This is sometimes called the <u>*skew-symmetry*</u> of $\det$.

(b) det is <u>*multilinear*</u>, i.e.

$$\det\Big(\vec{v}_1,\ldots,a\vec{v}_i+b\vec{u}_i,\ldots,\vec{v}_m\Big)=a\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_m\Big)+b\det\Big(\vec{v}_1,\ldots,\vec{u}_i,\ldots,\vec{v}_m\Big) \quad (1.60)$$

for all $i=1,\ldots,m$ all scalars a,b and all vectors $\vec{v}_1,\ldots,\vec{v}_{i-1},\vec{v}_i,\vec{u}_i,\vec{v}_{i+1},\ldots,\vec{v}_m$.

(c) The determinant is <u>*normalized*</u> to 1, i.e.

$$\det\Big(\vec{e}_1,\ldots,\vec{e}_m\Big)=1. \quad (1.61)$$

For an $m\times m$ matrix A, the input for the determinant function consists of the columns of A written in order:

$$\det A:=\det\Big(A\vec{e}_1,\cdots,A\vec{e}_m\Big). \quad (1.62)$$

Remark 1.63. The determinant is a special example of a tensor. The above definition makes this definition manifest. We will probably not cover tensors, but this relation is described after Theorem 4.2. in [13].

This definition reproduces the usual formula for the determinant that you may have seen if you've studied some linear algebra. Let A be the $m\times m$ matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \quad (1.64)$$

Then

$$\begin{aligned} \det A &= \det\left(\sum_{i_1=1}^{m} a_{i_1 1}\vec{e}_{i_1},\ldots,\sum_{i_m=1}^{m} a_{i_m m}\vec{e}_{i_m}\right) \\ &= \sum_{i_1=1}^{m}\cdots\sum_{i_m=1}^{m} a_{i_1 1}\cdots a_{i_m m}\det\left(\vec{e}_{i_1},\ldots,\vec{e}_{i_m}\right) \\ &= \sum_{\substack{i_1=1,\ldots,i_m=1 \\ i_1\neq i_2\neq\cdots\neq i_m}}^{m} a_{i_1 1}\cdots a_{i_m m}\mathrm{sign}(\sigma_{i_1\ldots i_m}), \end{aligned} \quad (1.65)$$

where $\sigma_{i_1\ldots i_m}$ is the permutation defined by

$$\begin{pmatrix} 1 & 2 & \cdots & m \\ i_1 & i_2 & \cdots & i_m \end{pmatrix} \quad (1.66)$$

and sign of a permutation is -1 raised to the power of the number of single swaps needed to obtain the permutation (although the number of such swaps is not well-defined, whether the number is odd or even *is* well-defined). The notation used above was probably more familiar from earlier courses in linear algebra. However, from now on, we will drop the arrows over vectors and simply denote $\vec{x}$ by x. Furthermore, we will use the notation $|\cdot|$ instead of $\|\cdot\|$ for the norm of a vector.

Level I problems.
From Spivak [13]: 1.2, 1.3, 1.4, 1.9, 1.10, 1.11
From these notes: Exercise 1.17, 1.18, 1.43, 1.48

Level II problems.
From Spivak [13]: 1.1, 1.6, 1.7, 1.8, 1.12, 1.13
From these notes: Exercise 1.39

2 January 24: Topological Notions in Euclidean Spaces

Today we will review concepts of topology in $\mathbb{R}^n$. All of these concepts should be familiar to you in $\mathbb{R}$, and we will therefore be brief. We already know that some subsets of $\mathbb{R}$ can have interesting and surprising properties. For example, the Cantor set is an uncountable perfect set. The rational numbers are dense in $\mathbb{R}$ but are countable. Is there a difference in intuition between subsets of $\mathbb{R}$ and subsets of $\mathbb{R}^n$ for $n > 1$?

Definition 2.1. A *closed rectangle* in $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ of the form

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n], \tag{2.2}$$

where $a_i, b_i \in \mathbb{R}$ and $a_i \leq b_i$ for all $i \in \{1, 2, \ldots, n\}$. An *open rectangle* in $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ of the form

$$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n), \tag{2.3}$$

where $a_i, b_i \in \mathbb{R}$ and $a_i < b_i$ for all $i \in \{1, 2, \ldots, n\}$. A subset $U \subseteq \mathbb{R}^n$ is called *open* iff for any $x \in U$, there exists an open rectangle R such that $x \in R$ and $R \subseteq U$. A subset $C \subseteq \mathbb{R}^n$ is called *closed* iff $\mathbb{R}^n \setminus C$ is open.

Definition 2.4. Let $A \subseteq \mathbb{R}^n$. The *interior* of A is the set

$$A^\circ := \bigcup_{R \subseteq A} R, \tag{2.5}$$

where R is an open rectangle in A. The *exterior* of A is the set

$$\operatorname{ext}(A) := \big(\mathbb{R}^n \setminus A\big)^\circ. \tag{2.6}$$

The *closure* of A is the set

$$\overline{A} := \mathbb{R}^n \setminus \operatorname{ext}(A). \tag{2.7}$$

The *boundary* of A is the exterior of A minus the interior of A, i.e.

$$\partial A := \mathbb{R}^n \setminus \big(A^\circ \cup \operatorname{ext}(A)\big). \tag{2.8}$$

Exercise 2.9. Show that for each $c \in \mathbb{R}^n$ and $\delta > 0$, the set

$$V_\delta(c) := \big\{x \in \mathbb{R}^n \ : \ |x - c| < \delta\big\} \tag{2.10}$$

is open. $V_\delta(c)$ is called the *δ-neighborhood/open ball around c*.

Exercise 2.11. Let $U \subseteq \mathbb{R}^n$. Show that U is open if and only if for every $c \in U$, there exists a $\delta > 0$ such that $V_\delta(c) \subseteq U$.

Exercise 2.12. Let $A \subseteq \mathbb{R}^n$. Prove that

$$A^\circ = \mathbb{R}^n \setminus \big(\overline{\mathbb{R}^n \setminus A}\big). \tag{2.13}$$

Exercise 2.14. Let $A \subseteq \mathbb{R}^n$. Prove that $\overline{A}$ is the intersection of all closed sets containing A, i.e.

$$\overline{A} = \bigcap_{B \supseteq A} B, \tag{2.15}$$

where B is a closed subset of $\mathbb{R}^n$ containing A.

Exercise 2.16. Let $c \in \mathbb{R}^n$ and fix $\delta > 0$.

(a) Show that the closure of $V_\delta(c)$ is the set

$$\overline{V_\delta(c)} = \big\{ x \in \mathbb{R}^n \ : \ |x - c| \leq \delta \big\}. \tag{2.17}$$

(b) Show that the exterior of $V_\delta(c)$ is the set

$$\operatorname{ext}\big(V_\delta(c)\big) = \big\{ x \in \mathbb{R}^n \ : \ |x - c| > \delta \big\}. \tag{2.18}$$

(c) Show that the boundary of $V_\delta(c)$ is the set

$$\partial V_\delta(c) = \big\{ x \in \mathbb{R}^n \ : \ |x - c| = \delta \big\}. \tag{2.19}$$

Exercise 2.20. Let $A \subseteq \mathbb{R}^n$. Show that $\partial(\partial A) = \varnothing$.

Last semester, we learned about the Nested Interval Property. The analogous result for rectangles is true as well.

Theorem 2.21 (Nested Rectangle Property)**.** *Let*

$$R_1 \supseteq R_2 \supseteq R_3 \supseteq \cdots \tag{2.22}$$

be a nested sequence of closed rectangles in $\mathbb{R}^n$*. Then*

$$\bigcap_{k=1}^{\infty} R_k \neq \varnothing. \tag{2.23}$$

Proof. For each $i \in \{1, 2, \ldots, n\}$, the image $\pi_i(R_k)$ of each rectangle under the projection maps is an interval. Furthermore, these intervals satisfy

$$\pi_i(R_1) \supseteq \pi_i(R_2) \supseteq \pi_i(R_3) \supseteq \cdots . \tag{2.24}$$

Hence, by the Nested Interval Property, there exists a

$$c_i \in \bigcap_{k=1}^{\infty} \pi_i(R_k). \tag{2.25}$$

for each $i \in \{1, 2, \ldots, n\}$. Thus,

$$(c_1, c_2, \ldots, c_n) \in \bigcap_{k=1}^{\infty} R_k. \tag{2.26}$$

■

Last semester, we learned a slightly different definition of closed subsets. This definition is also valid here.

Exercise 2.27. A *sequence* in $\mathbb{R}^m$ is a function $a : \mathbb{N} \to \mathbb{R}^m$ whose value at $n \in \mathbb{N}$ is denoted by a_n. A sequence $a : \mathbb{N} \to \mathbb{R}^m$ *converges* to $\lim a \in \mathbb{R}^m$ iff for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$|a_n - \lim a| < \epsilon \qquad \forall\, n \geq N. \tag{2.28}$$

Let $A \subseteq \mathbb{R}^m$ be a subset. An element $x \in A$ is a *limit point* of A if and only if for any open set U containing x, $A \cap \big(U \setminus \{x\}\big) \neq \varnothing$. Otherwise, x is said to be an *isolated point*. Show that the following are equivalent for a subset $A \subseteq \mathbb{R}^m$.

(a) A is closed.

(b) For any sequence $a : \mathbb{N} \to A$ that converges to $\lim a$, then $\lim a \in A$.

(c) A contains all of its limit points.

Definition 2.29. A subset $A \subseteq \mathbb{R}^n$ is *bounded* iff there exists an $R > 0$ such that $A \subseteq V_R(0)$.

Definition 2.30. Let $A \subseteq \mathbb{R}^n$. An *open cover* of A is a set $\mathcal{O}$ of open subsets of $\mathbb{R}^n$ such that

$$A \subseteq \bigcup_{U \in \mathcal{O}} U. \tag{2.31}$$

A is *compact* iff for any open cover $\mathcal{O}$ of A, there exists a finite subset $\mathcal{V} = \{U_{\alpha_1}, \ldots, U_{\alpha_k}\}$ of $\mathcal{O}$ that covers A (a.k.a. a *finite subcover*).

The following three exercises were done in our first semester of analysis but for subsets of $\mathbb{R}$. Completely analogous proofs hold for $\mathbb{R}^n$ and are therefore left to the reader.

Exercise 2.32. Let $a : \mathbb{N} \to \mathbb{R}^m$ be a sequence. A *subsequence* of a consists of a one-to-one and monotone non-decreasing function $f : \mathbb{N} \to \mathbb{N}$ together with the sequence $a \circ f : \mathbb{N} \to \mathbb{R}^m$. Show that a subset $A \subseteq \mathbb{R}^m$ is compact if and only if for any sequence $a : \mathbb{N} \to A$, there exists a convergent subsequence whose limit is contained in A.

Exercise 2.33. Let

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq K_4 \supseteq \cdots \tag{2.34}$$

be a nested sequence of nonempty compact subsets of $\mathbb{R}^n$. Show that

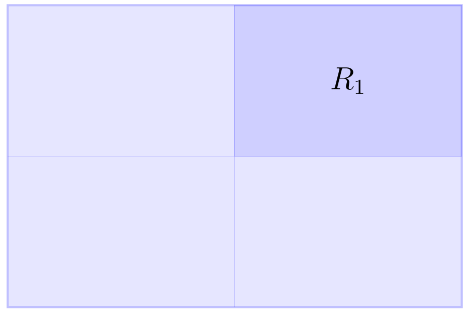
$$\bigcap_{m=1}^{\infty} K_m \neq \varnothing. \tag{2.35}$$

Exercise 2.36. Show that a subsequence of a convergent sequence in $\mathbb{R}^n$ is convergent and converges to the same value as the original sequence. Use this fact to show that compact subsets of $\mathbb{R}^n$ are closed.

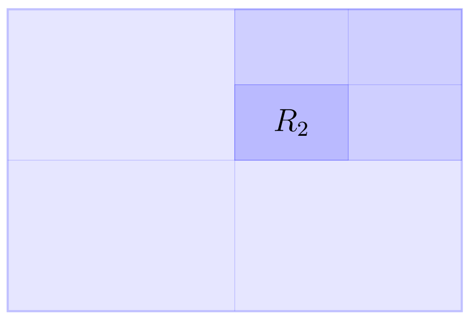
Exercise 2.37. Show that any closed subset of a compact set in $\mathbb{R}^n$ is compact. Use this to show that the intersection of an arbitrary collection of compact subsets of $\mathbb{R}^n$ is compact.

Theorem 2.38. *All closed rectangles in $\mathbb{R}^n$ are compact for all $n \in \mathbb{N}$.*

Proof. Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a closed rectangle. Suppose to the contrary that there exists an open cover $\mathcal{U} := \{U_\lambda\}_{\lambda \in \Lambda}$ of R with no finite subcover. Then, at least one of the 2^n halved n-cubes of R is not covered by finitely many of the U_λ. Let R_1 denote such a cube.

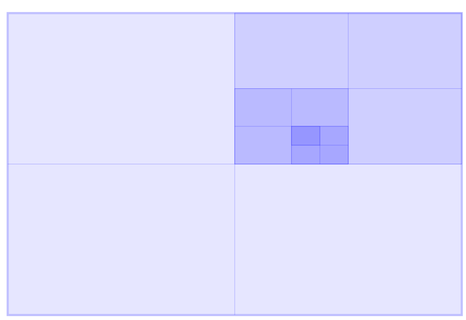


Similarly, at least one of the 2^n halved n-cubes of R_1 is not covered by finitely many of the U_λ. Let R_2 denote such a cube.



Continue this process so that you have a nested sequence of non-empty closed rectangles

$$R_1 \supseteq R_2 \supseteq R_3 \supseteq \cdots . \tag{2.39}$$



By the Nested Rectangle Property, there exists a $c \in \bigcap_{k=1}^{\infty} R_k$. Since $\mathcal{U}$ is an open cover of R, there exists an open set $U_{\lambda_0} \in \mathcal{U}$ such that $c \in U_{\lambda_0}$. Therefore, there exists an open rectangle Q such that $c \in Q \subseteq U_{\lambda_0}$. Q contains all but finitely many of the nested rectangles. Thus, there exists an $N \in \mathbb{N}$ such that $R_k \subseteq Q \subseteq U_{\lambda_0}$ for all $k \geq N$. This contradicts the assumption that there does not exist a finite subset of $\mathcal{U}$ that covers R_k. ■

Theorem 2.40 (Heine-Borel Theorem in $\mathbb{R}$)**.** *A subset $A \subseteq \mathbb{R}$ is compact if and only if A is closed and bounded.*

Proof. The proof of this theorem was given in class last semester. ■

The analogous theorem is also true in $\mathbb{R}^n$.

Theorem 2.41. *[Heine-Borel Theorem in $\mathbb{R}^n$] A subset $A \subseteq \mathbb{R}^n$ is compact if and only if A is closed and bounded.*

Proof.
($\Rightarrow$) Suppose A is compact. Then, by Exercise 2.36, A is closed. For each $k \in \mathbb{N}$, the sets $V_k(0)$ are open by Exercise 2.9. Furthermore, the collection $\big\{V_k(0)\big\}_{k\in\mathbb{N}}$ is an open cover of $\mathbb{R}^n$ and therefore covers A. Since A is compact, there exists a finite subcover $\big\{V_{k_1}(0), \dots, V_{k_m}(0)\big\}$. Set $M := \max\{k_1, \dots, k_m\}$. Then $A \subseteq V_M(0)$ so that A is bounded.
($\Leftarrow$) Suppose A is closed and bounded. Since A is bounded, there exists an $M > 0$ such that $A \subseteq \overline{V_M(0)}$. Thus, $A \subseteq [-M, M] \times \cdots \times [-M, M] \subseteq \mathbb{R}^n$. Since A is a closed subset of $[-M, M] \times \cdots \times [-M, M]$, which is compact by Theorem 2.38, A is compact by Exercise 2.37. ■

Definition 2.42. Two nonempty subsets $A, B \subseteq \mathbb{R}^n$ are said to be <u>*separated*</u> iff $\overline{A} \cap B$ and $A \cap \overline{B}$ are both empty. A subset $E \subseteq \mathbb{R}^n$ is <u>*disconnected*</u> iff there exist separated subsets $A, B \subseteq E$ with $E = A \cup B$. A subset $E \subseteq \mathbb{R}^n$ is <u>*connected*</u> iff it is not disconnected.

The following is also a useful characterization of connected subsets.

Theorem 2.43. *A subset $E \subseteq \mathbb{R}^n$ is disconnected if and only if there exist nonempty open sets $U \subseteq E$ and $V \subseteq E$ such that $U \cap V = \varnothing$ and $U \cup V = E$.*

Proof. This was proved last semester. ■

Connected subsets of $\mathbb{R}$ have a particularly simple characterization, which we recall.

Theorem 2.44. *A subset $E \subseteq \mathbb{R}$ is connected if and only if for any $a, b \in E$ with $a < b$ implies $(a, b) \subseteq E$. In particular, closed intervals are connected.*

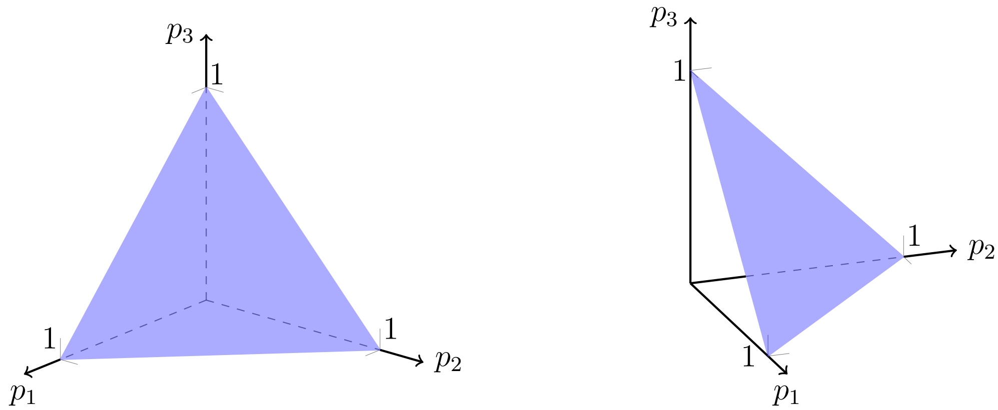
The analogous statement is *false* for subsets of $\mathbb{R}^n$ when $n > 1$. However, we do have the following generalization of this definition and result.

Definition 2.45. A subset $C \subseteq \mathbb{R}^n$ is <u>*convex*</u> iff for any two points $x, y \in C$, $ty + (1-t)x \in C$ for all $t \in [0, 1]$.

Exercise 2.46. The $(n-1)$-simplex Δ^{n-1} is defined to be

$$\Delta^{n-1} := \left\{ (p_1, \dots, p_n) \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} p_i = 1 \text{ and } p_i \geq 0 \ \forall \ i = 1, \dots, n \right\}. \tag{2.47}$$

For example, the 2-simplex looks like the following subset of $\mathbb{R}^3$ viewed from two different angles



Show that the $(n-1)$-simplex is convex for all $n \in \mathbb{N}$.

Theorem 2.48. *Every convex subset of $\mathbb{R}^n$ is connected.*

Proof. The proof of this will be postponed until next lecture. ■

The converse of Theorem 2.48 is false.

Example 2.49. Fix $n \in \mathbb{N}$. The set

$$S^{n-1} := \{x \in \mathbb{R}^n \ : \ |x| = 1\} \tag{2.50}$$

is called the *(standard) unit sphere* of dimension $n-1$. S^{n-1} is not convex for any $n \in \mathbb{N}$.

Exercise 2.51. Using only the definition of connectedness, show that S^{n-1} is connected for all $n \in \mathbb{N}$ with $n > 1$. Show that S^0 is disconnected.

Level I problems.
From Spivak [13]: 1.14, 1.15, 1.16, 1.18, 1.19
From these notes: Exercise 2.9, 2.11, 2.12, 2.14, 2.20, 2.36, 2.37, 2.46

Level II problems.
From Spivak [13]: 1.17, 1.21, 1.22
From these notes: Exercise 2.16, 2.27, 2.32, 2.33, 2.51

3 January 26: Continuous Functions

Let $A \subseteq \mathbb{R}^p$. By the universal property of the Cartesian product, a function $f : A \to \mathbb{R}^n$ determines and is uniquely determined by its n *component functions* $\pi_i \circ f$, where $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is the projection onto the i-th factor. Setting $f_i := \pi_i \circ f$, one may also write $f = (f_1, \ldots, f_n)$. A is called the *domain* of f and $\mathbb{R}^n$ is called its *codomain*. Let $B \subseteq \mathbb{R}^n$ with $f(A) \subseteq B$ and let $g : B \to \mathbb{R}^m$ be another function. The *composition* of f followed by g will be denoted by $g \circ f$ and we will often depict this via the diagram

$$\begin{tikzcd} & B \arrow[dl, "g"'] & \\ \mathbb{R}^m & & A \arrow[ul, "f"'] \arrow[ll, "g\circ f"] \end{tikzcd} \tag{3.1}$$

Definition 3.2. Let $A \subseteq \mathbb{R}^n$, let $f, g : A \to \mathbb{R}^m$ be two functions, and let $c \in \mathbb{R}$. The *sum* of f and g is the function $f + g : A \to \mathbb{R}^m$ defined by

$$A \ni a \mapsto (f+g)(a) := f(a) + g(a). \tag{3.3}$$

The *scalar multiple* of c with f is the function $cf : A \to \mathbb{R}^m$ defined by

$$A \ni a \mapsto (cf)(a) := cf(a). \tag{3.4}$$

Definition 3.5. Let $A \subseteq \mathbb{R}^n$ and let $f, g : A \to \mathbb{R}$ be two functions. The *product* of f with g is the function $fg : A \to \mathbb{R}$ defined by

$$A \ni a \mapsto (fg)(a) := f(a)g(a). \tag{3.6}$$

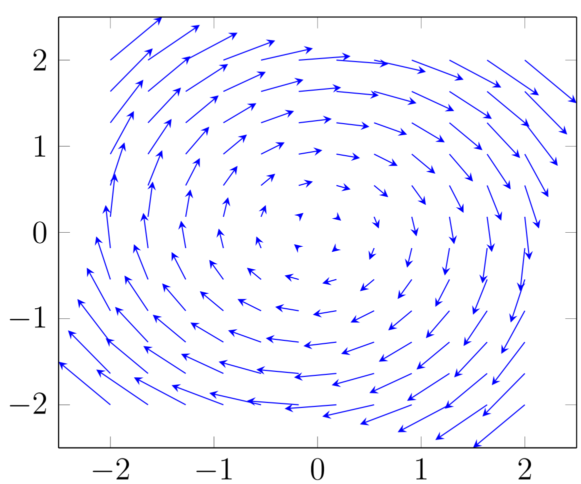
If $g(a) \neq 0$ for all $a \in A$, the quotient of f with g is the function $\frac{f}{g} : A \to \mathbb{R}$ defined by

$$A \ni a \mapsto \left(\frac{f}{g}\right)(a) := \frac{f(a)}{g(a)}. \tag{3.7}$$

There are several ways to visualize such functions.

Example 3.8. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) := (y, -x). \tag{3.9}$$

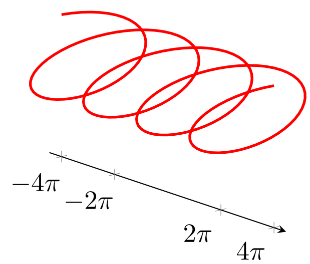


This function can be conveniently drawn with its values indicated directly on the domain (the arrows are scaled to 1/4 their actual lengths). This function is an example of a vector field, a concept that we will return to in more detail later.

Example 3.10. Let $\delta : \mathbb{R} \to \mathbb{R}^2$ be the function

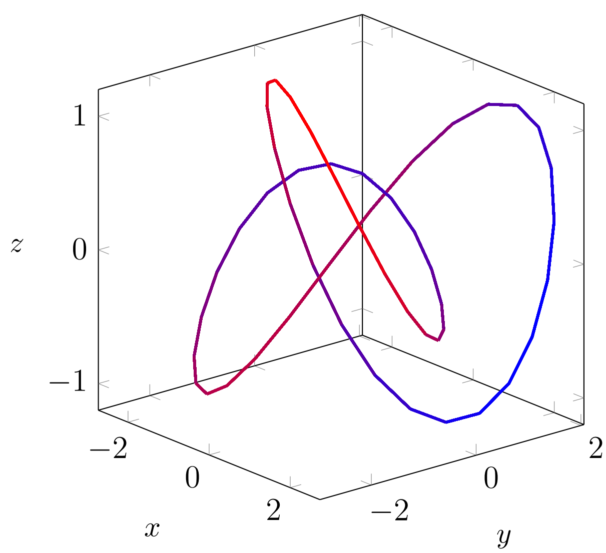
$$\mathbb{R} \ni t \mapsto \delta(t) := \big(\cos(t), \sin(t)\big). \tag{3.11}$$

This can be depicted as the graph of a function by drawing planes above each point t in $\mathbb{R}$ and drawing the value of δ at t on this plane. This traces out a helix as in the following plot.



Example 3.12. Let $\gamma : [0, 2\pi] \to \mathbb{R}^3$ be defined by

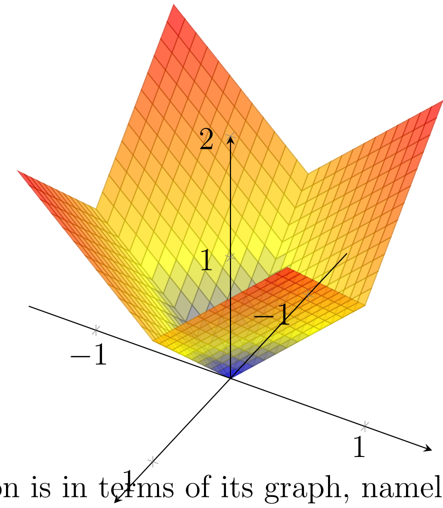
$$[0, 2\pi] \ni t \mapsto \gamma(t) := \big(\sin t + 2\sin 2t, \cos t - 2\cos 2t, -\sin 3t\big). \tag{3.13}$$



This is an example of a parametrized curve, more specifically a loop. This curve is known as the *trefoil knot*. The visualization of this function is in terms of its image in the codomain.

Example 3.14. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto h(x, y) := |x| + |y|. \tag{3.15}$$



The visualization of this function is in terms of its graph, namely, the set of points of the form

$$\Gamma := \Big\{ \big(x, y, h(x, y)\big) \in \mathbb{R}^3 \ : \ (x, y) \in \mathbb{R}^2 \Big\}. \tag{3.16}$$

The definitions of limits of functions and of continuity of functions is completely analogous to what it was for real-valued functions of a single variable.

Definition 3.17. Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Let c be a limit point of A. L is said to be a *limit of f as x approaches c*, written

$$\lim_{x \to c} f(x) = L, \tag{3.18}$$

if for any $\epsilon > 0$, there exists a $\delta > 0$ such that $f(x) \in V_\epsilon(L)$ for all $x \in V_\delta(c) \cap A \setminus \{c\}$. L is called a *functional limit* of f as x approaches c.

Exercise 3.19. Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}^m$, and let c be a limit point of A. If L and L' are both limits of f as x approaches c, show that $L = L'$, i.e. functional limits are unique.

Exercise 3.20. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function and let $c \in \mathbb{R}^n$ and $L \in \mathbb{R}^m$. Show that

$$\lim_{x \to c} f(x) = L \tag{3.21}$$

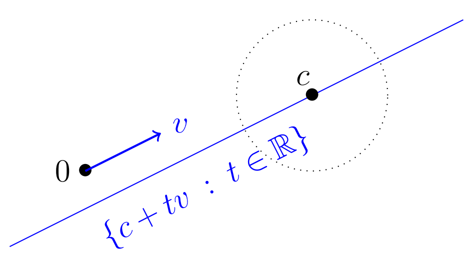
if and only if

$$\lim_{h \to 0} f(c + h) = L. \tag{3.22}$$

The following exercise is meant to illustrate that limits in more than one dimension are not strictly determined by limits of functions along straight paths.

Exercise 3.23. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function and let $c \in \mathbb{R}^n$. For every $v \in \mathbb{R}^n$ let $f_{c;v} : \mathbb{R} \to \mathbb{R}^m$ be the function defined by

$$\mathbb{R} \ni t \mapsto f_{c;v}(t) := f(c + tv). \tag{3.24}$$



Equivalently, if $\gamma_{c;v} : \mathbb{R} \to \mathbb{R}^n$ denotes the curve $\gamma_{c;v}(t) := c + tv$, then $f_{c;v} = f \circ \gamma_{c;v}$.

(a) Show that if $\lim_{x\to c} f(x) = L$ for some $L \in \mathbb{R}^m$, then

$$\lim_{t\to 0} f_{c;v}(t) = L \tag{3.25}$$

for every $v \in \mathbb{R}^n$.

(b) Give an example of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ (for some choices of m and n), a point $c \in \mathbb{R}^n$, and a vector $L \in \mathbb{R}^m$ for which

$$\lim_{t\to 0} f_{c;v}(t) = L \tag{3.26}$$

for every $v \in \mathbb{R}^n$ but for which f does not have a limit as x approaches c. [Hint: try $n = 2$ and $m = 1$ and do not use functions that are invariant under rotation about the origin.]

(c) For each $i \in \{1, \ldots, n\}$, let $\varphi_i : \mathbb{R} \to \mathbb{R}^n$ denote the i-th inclusion function, namely

$$\mathbb{R} \ni x \mapsto \varphi_i(x) := (0, \ldots, 0, x, 0, \ldots, 0) \tag{3.27}$$

where the non-zero entry is in the i-th coordinate. Find the appropriate $c \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ such that $f \circ \gamma_{c;v} = f \circ \varphi_i$.

This last part shows that fixing all the coordinates of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is a special case of precomposing f with a straight line curve.

Taking limits is a local operation in the following sense.

Exercise 3.28. Let $A \subseteq \mathbb{R}^n$, let $f, g : A \to \mathbb{R}^m$ be two functions, let $c \in A$ be a limit point of A, and suppose that $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ both exist. Suppose that there exists an open set U containing c such that $f = g$ on $U \cap A$. Show that

$$\lim_{x\to c} f(x) = \lim_{x\to c} g(x). \tag{3.29}$$

[Hint: Use Exercise 2.11.]

Definition 3.30. Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. f is <u>*continuous at* $c \in A$</u> if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $f(x) \in V_\epsilon\big(f(c)\big)$ for all $x \in V_\delta(c) \cap A$. If f is not continuous at $c \in A$, then f is said to be <u>*discontinuous at* $c \in A$</u>. f is <u>*continuous*</u> on A if f is continuous at c for all $c \in A$.

Exercise 3.31. Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. If $c \in A$ is an isolated point of A, show that f is continuous at a.

The following theorem gives some simple examples of continuous functions.

Theorem 3.32.

(a) The i-th projection $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous for all $i \in \{1, 2, \ldots, n\}$.

(b) Let $A \subseteq \mathbb{R}^n$. Then the identity function $\mathrm{id} : A \to \mathbb{R}^n$ *is continuous.*

(c) Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a constant function, meaning that $f(x) = f(y)$ for all $x, y \in \mathbb{R}^n$. Then f is continuous.

(d) The function $s : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto s(x, y) := x + y \tag{3.33}$$

is continuous.

Proof.

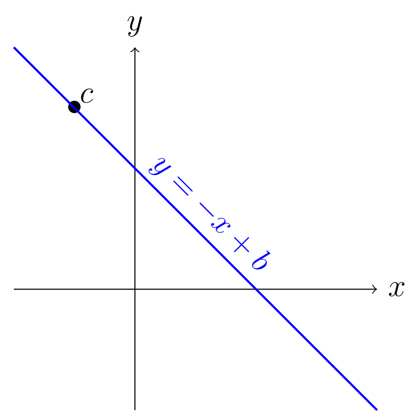
(a) Fix $\epsilon > 0$ and $c \in \mathbb{R}^n$. Set $\delta := \epsilon$. Then $\pi_i(x) \in V_\epsilon\big(\pi_i(c)\big)$ for all $x \in V_\delta(c)$.

(b) A similar proof shows that id is continuous.

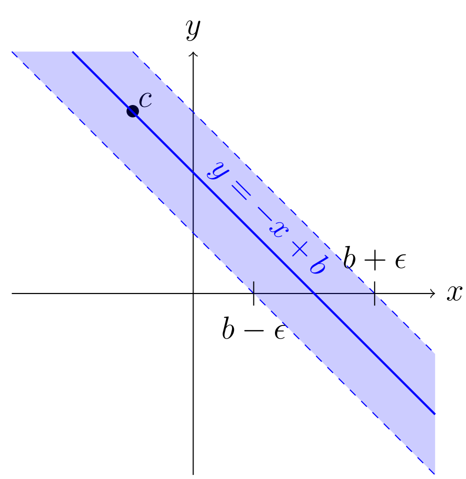
(c) Let $c \in A$ and fix $\epsilon > 0$. Set $\delta = 1$ (or anything for that matter). Let $d := f(c)$. Since $f(x) = d$ for all $x \in A$, $f(x) \in V_\epsilon(d)$ for all $x \in V_\delta(c)$.

(d) Fix $c \in \mathbb{R}^2$ and $\epsilon > 0$. Let $b := s(c)$. Then

$$s^{-1}(b) = \big\{(x, y) \in \mathbb{R}^2 \ : \ x + y = b\big\} = \big\{(x, y) \in \mathbb{R}^2 \ : \ y = -x + b\big\} \tag{3.34}$$



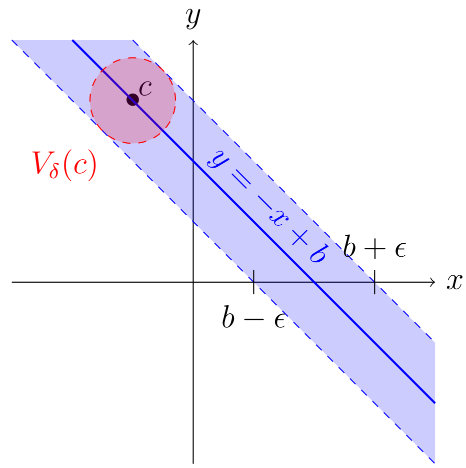
Hence, given $\epsilon > 0$, the set $s^{-1}\big(V_\epsilon(b)\big)$ is the following shaded subset of $\mathbb{R}^2$



Therefore, set

$$\delta := \frac{\epsilon}{\sqrt{2}}. \tag{3.35}$$

Then $s\big(V_\delta(c)\big) \subseteq V_\epsilon\big(s(c)\big)$ as can be seen from the following image.



Hence, s is continuous at $c \in \mathbb{R}^2$. Since c was arbitrary, s is continuous on all of $\mathbb{R}^2$.

■

Exercise 3.36. Show that the function $p : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto p(x, y) := xy \tag{3.37}$$

is continuous.

Exercise 3.38. Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}^m$ be a function, and let $c \in A$. Show that f is continuous at c if and only if for every sequence $a : \mathbb{N} \to A$ with $\lim a = c$, it follows that $\lim(f \circ a) = f(c)$. Furthermore, if c is a limit point of A, show that these conditions are equivalent to $\lim\limits_{x \to c} f(x) = f(c)$.

Exercise 3.39. Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}^m$ be a function. Show that f is continuous if and only if for any open set $U \subseteq \mathbb{R}^m$, there exists an open set $V \subseteq \mathbb{R}^n$ such that $V \cap A = f^{-1}(U)$.

Exercise 3.40. Let $A \subseteq \mathbb{R}^p$ and $B \subseteq \mathbb{R}^n$. Let $f : A \to B$ and $g : B \to \mathbb{R}^m$ be functions continuous at $c \in A$ and $f(c) \in B$, respectively. Show that the composition $g \circ f : A \to \mathbb{R}$ is continuous at c.

Theorem 3.41. *Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}^m$ be a function and fix $c \in A$. f is continuous at c if and only if $\pi_i \circ f : A \to \mathbb{R}^m \to \mathbb{R}$ are continuous at c for all $i \in \{1, 2, \ldots, n\}$.*

Proof.
($\Rightarrow$) Assume that f is continuous. Then by Theorem 3.32 and Exercise 3.40, $\pi_i \circ f$ is continuous.
($\Leftarrow$) Suppose that $\pi_i \circ f : A \to \mathbb{R}^n \to \mathbb{R}$ is continuous at c for all $i \in \{1, 2, \ldots, n\}$. Fix $\epsilon > 0$. For each $i \in \{1, 2, \ldots, n\}$, there exists a $\delta_i > 0$ such that $\pi_i\big(f(x)\big) \in V_\epsilon\Big(\pi_i\big(f(c)\big)\Big)$ for all $x \in A \cap V_{\delta_i}(c)$ since $\pi_i \circ f$ is continuous. Set

$$\delta := \min\{\delta_1, \ldots, \delta_n\}. \tag{3.42}$$

Then $f(x) \in V_\epsilon\big(f(c)\big)$ for all $x \in A \cap V_\delta(c)$. ■

Theorem 3.43 (Algebraic Continuity Theorem)**.** *Let $A \subseteq \mathbb{R}^n$, let $f, g : A \to \mathbb{R}^m$, and let $c \in A$. Furthermore, suppose that f and g are continuous at c. Then the following facts hold.*

(a) The function kf is continuous at c for all $k \in \mathbb{R}$.

(b) The function $f + g$ is continuous at c.

(c) When $n = 1$, the function fg is continuous at c.

(d) When $n = 1$, let $B \subseteq A$ be the domain over which g is nonzero and so that $c \in B$. Then the function $\frac{f}{g} : B \to \mathbb{R}$ is continuous at c.

Proof. The proof is similar to last semester. ■

Theorem 3.44. *Let $K \subseteq \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$ and let $f : K \to \mathbb{R}^n$ be continuous. Then $f(K)$ is a compact subset of $\mathbb{R}^m$.*

Proof. The same proof from last semester applies. ■

Theorem 3.45. *Let $A \subseteq \mathbb{R}^n$ be a connected subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be continuous. Then $f(A)$ is a connected subset of $\mathbb{R}^m$.*

Proof. The same proof from last semester applies. ■

Theorem 3.46. *The norm function $| \cdot | : \mathbb{R}^n \to \mathbb{R}$ defined by*

$$\mathbb{R}^n \ni x \equiv (x_1, \dots, x_n) \mapsto |x| := \sqrt{x_1^2 + \cdots + x_n^2} \tag{3.47}$$

is continuous on $\mathbb{R}^n$.

Proof. The norm function $| \cdot | : \mathbb{R}^n \to \mathbb{R}$ is the composition of the following functions

$$\mathbb{R} \xleftarrow{\sqrt{\cdot}} \mathbb{R} \xleftarrow{s} \mathbb{R}^n \xleftarrow{\mathrm{sq}} \mathbb{R}^n \tag{3.48}$$

where

$$\mathbb{R}^n \ni x \equiv (x_1, \dots, x_n) \mapsto \mathrm{sq}(x) := (x_1^2, \dots, x_n^2) \tag{3.49}$$

and

$$\mathbb{R}^n \ni y \equiv (y_1, \dots, y_n) \mapsto s(y) := y_1 + \cdots + y_n. \tag{3.50}$$

sq is continuous by Theorem 3.41 and the fact that the squaring operation is continuous on $\mathbb{R}$. s is continuous because it is the composition of several binary sum functions as in part (d) of Theorem 3.32, which is continuous. Finally, $\sqrt{\cdot}$ is continuous because the square-root function is continuous. Thus, by Theorem 3.40, the norm function is continuous. ■

Exercise 3.51. Show that the Euclidean inner product $\langle \cdot\,, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by sending $\big((x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n)\big) \in \mathbb{R}^n \times \mathbb{R}^n$ to

$$\big\langle (x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \big\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \tag{3.52}$$

is continuous.

Exercise 3.53. Show that the functions in Examples 3.8 and 3.12 are continuous.

Exercise 3.54. Show that the functions in Examples 3.10 and 3.14 are continuous.

Example 3.12 is particularly interesting as it is an example of a *path*.

Definition 3.55. Let $A \subseteq \mathbb{R}^n$. A *(Moore) path* in A is a continuous function of the form $\gamma : [a, b] \to A$ for some $a, b \in \mathbb{R}$ with $a \leq b$.

Definition 3.56. A subset $A \subseteq \mathbb{R}^n$ is *path connected* iff for any two points $x, y \in A$, there exist $r, s \in \mathbb{R}$ with $r \leq s$ and a path $\gamma : [r, s] \to A$ with $\gamma(r) = x$ and $\gamma(s) = y$.

Exercise 3.57. Show that a subset $A \subseteq \mathbb{R}^n$ is path connected if and only if for any two points $x, y \in A$, there exists a path $\gamma : [0, 1] \to A$ with $\gamma(0) = x$ and $\gamma(1) = y$.

Theorem 3.58. *Every convex subset of $\mathbb{R}^n$ is path connected and hence connected.*

Proof. Let $C \subseteq \mathbb{R}^n$ be a convex set. Let $x, y \in C$. Because C is convex, $(1-t)x + ty \in C$ for all $t \in [0, 1]$. Let $\gamma : [0, 1] \to \mathbb{R}^n$ be the path defined by this formula, namely $t \in [0, 1]$ gets sent to $\gamma(t) := (1-t)x + ty$. γ can be written as $\gamma(t) = t(y - x) - x$ and is therefore a linear function so is continuous. ■

Example 3.59. The converse is false, namely, there exist path connected sets that are not convex. An example is S^1, the unit circle in $\mathbb{R}^2$. To see this, let $(x_1, y_1), (x_2, y_2)$ be two points on the unit circle. There are three cases to consider (the following is not the most elegant proof but is meant to illustrate several methods of approaching the problem).

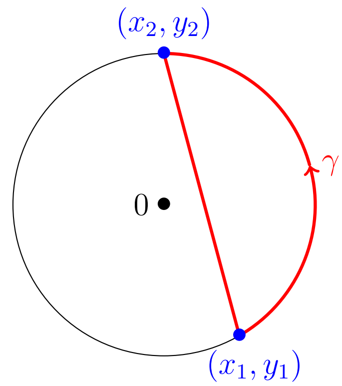
Case 1: If $(x_2, y_2) = (x_1, y_1)$, then the constant path, namely $\gamma(t) = (x_1, y_1)$ for all $t \in [0, 1]$, exhibits a path between the two points that lies in S^1. The constant path is continuous by Theorem 3.32.

Case 2: If $(x_2, y_2) = -(x_1, y_1)$, let θ be the unique angle in $[0, 2\pi)$ such that $x_1 = \cos\theta$ and $y_1 = \sin\theta$. Then the path $\gamma : [0, \pi] \to S^1$ defined by sending $t \in [0, \pi]$ to $\gamma(t) := \big(\cos(t + \theta), \sin(t + \theta)\big)$ satisfies $\gamma(0) = (x_1, y_1)$ and $\gamma(\pi) = (x_2, y_2)$. Let $T_\theta : \mathbb{R} \to \mathbb{R}$ denote the function that is translation by θ, namely $T_\theta(t) := t + \theta$ for all $t \in \mathbb{R}$. T_θ is linear and is therefore continuous by Exercise 1.25 in [13]. Hence, the component functions of γ, namely $\gamma_1 = \cos \circ T_\theta$ and $\gamma_2 = \sin \circ T_\theta$, are the compositions of continuous functions by Exercise 3.40. Since the component functions are continuous, γ is continuous by Theorem 3.41.

Case 3: In the final case, i.e. when $(x_2, y_2) \neq \pm(x_1, y_1)$, the path $\gamma : [0, 1] \to S^1$ defined by

$$[0, 1] \ni t \mapsto \gamma(t) := \frac{(1-t)(x_1, y_1) + t(x_2, y_2)}{|(1-t)(x_1, y_1) + t(x_2, y_2)|} \tag{3.60}$$

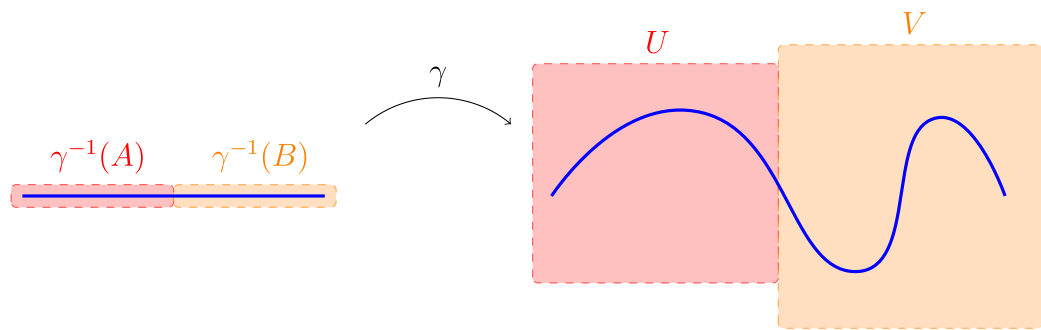
satisfies $\gamma(0) = (x_1, y_1)$ and $\gamma(1) = (x_2, y_2)$. Note that γ is well-defined because the denominator is never zero as can be seen in the following picture.



γ is continuous for the following reasons. First, the numerator is continuous because it is a straight path and the proof of Theorem 3.58 applies. The denominator is continuous because the norm function is continuous by Theorem 3.46. Finally, γ is the ratio of two continuous functions with the denominator never being zero and is therefore continuous by Theorem 3.43.

Theorem 3.61. *Every path connected subset of $\mathbb{R}^n$ is connected.*

Proof. Let $E \subseteq \mathbb{R}^n$ be path connected and suppose to the contrary that there exists two non-empty disjoint open subsets[1] $A, B \subseteq E$ such that $A \cup B = E$. Let $x \in A$ and $y \in B$. By assumption, there exists a path $\gamma : [0,1] \to E$ with $\gamma(0) = x$ and $\gamma(1) = y$. Since γ is continuous, $\gamma^{-1}(A)$ and $\gamma^{-1}(B)$ are open. Furthermore, they are nonempty since $0 \in \gamma^{-1}(A)$ and $1 \in \gamma^{-1}(B)$ and they are disjoint by assumption. This contradicts the fact that $[0,1]$ is connected by Theorem 2.44.



■

Notice that the proof of this theorem was quite simple. It relied heavily on the fact that $[0,1]$ is connected, a fact whose proof was significantly more involved and technical.

Exercise 3.62. Prove that S^{n-1}, the $(n-1)$-dimensional sphere (see Example 2.49) is path connected for all $n \in \mathbb{N}$ with $n > 1$. [Hint: 2 linearly independent vectors in $\mathbb{R}^n$ uniquely determine a plane.]

The converse of Theorem 3.61 is false.

[1] Though this is not necessary for the proof, recall that A is open in E iff there exists an open set $U \subseteq \mathbb{R}^n$ such that $U \cap E = A$.

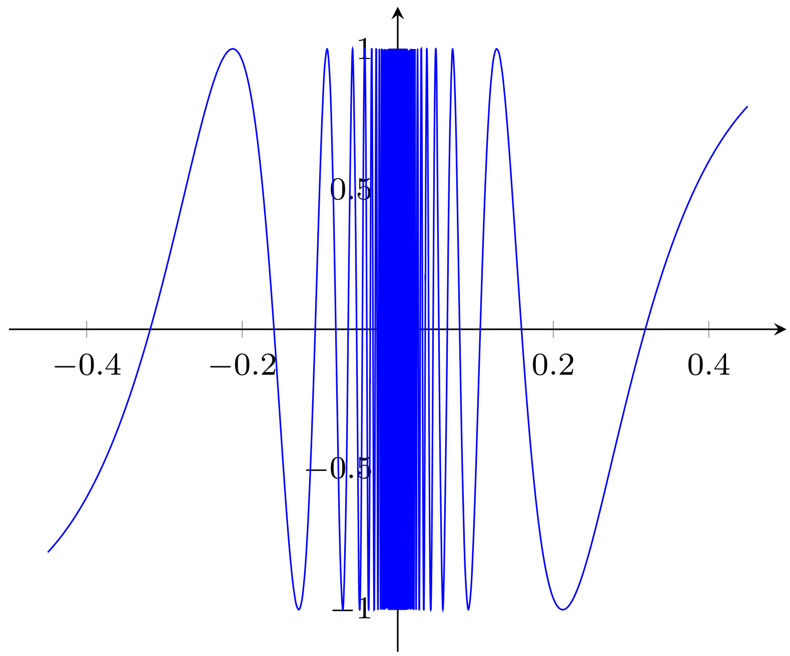
Example 3.63. Let $f : \mathbb{R} \to \mathbb{R}$ be the function

$$\mathbb{R} \ni x \mapsto f(x) := \begin{cases} 0 & \text{if } x = 0 \\ \sin\left(\frac{1}{x}\right) & \text{otherwise} \end{cases}. \tag{3.64}$$

Let

$$\Gamma := \Big\{ \big(x, f(x)\big) \in \mathbb{R}^2 \ : \ x \in \mathbb{R} \Big\} \tag{3.65}$$

be the graph of f (see Example 3.14). Then Γ as a subset of $\mathbb{R}^2$ is connected but not path connected. Γ is known as the <u>*topologist's sine curve*</u>.



Exercise 3.66. Prove that the topologist's sine curve is connected. Prove that it is not path connected.

Level I problems.
From Spivak [13]: 1.23, 1.25, 1.28
From these notes: Exercise 3.19, 3.20, 3.28, 3.31, 3.39, 3.40, 3.51, 3.53, 3.54, 3.57

Level II problems.
From Spivak [13]: 1.26, 1.29
From these notes: Exercise 3.23, 3.36, 3.38, 3.62, 3.66

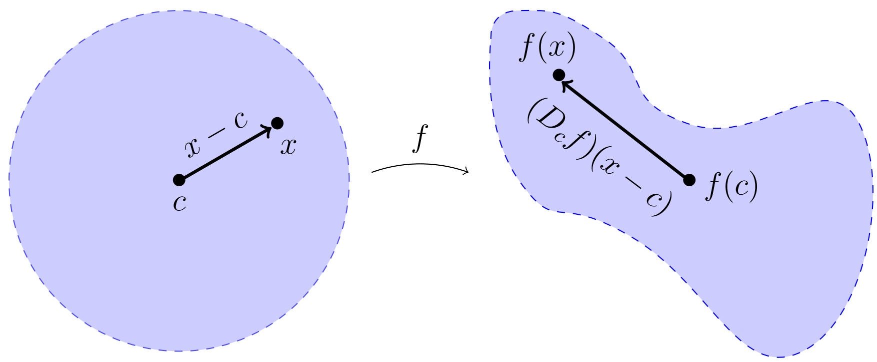
4 January 31: The Differential

Last semester, we studied two equivalent definitions for a function to be differentiable on $\mathbb{R}$. Here, we will present the one that is more useful in higher dimensions.

Definition 4.1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable* at $c \in \mathbb{R}^n$ if and only if there exists a linear transformation $D_c f : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to c} \frac{\big|f(x) - f(c) - (D_c f)(x - c)\big|}{|x - c|} = 0. \tag{4.2}$$

$D_c f$ is called the *differential* of f at c.



This cartoon picture provides some justification for the notation $D_c f$ for the differential of f at c. One "pushes" a vector v located at c (in the case of the picture, the vector v is $x - c$) along the function f to the vector $(D_c f)(v)$ at $f(c)$. This heuristic idea will be made more rigorous in a few lectures. Also notice that when $m = 1$ and $n = 1$, $D_c f : \mathbb{R} \to \mathbb{R}$ is completely determined by its slope. In fact, the slope is precisely $f'(c)$, the derivative of f at c, namely

$$(D_c f)(v) = f'(c)v \tag{4.3}$$

for all $v \in \mathbb{R}$. We proved this last semester. For more intuitive and less rigorous for why the differential should be a linear transformation, consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ of two variables and let $c = (a, b)$ for some $a, b \in \mathbb{R}$. Consider the two points $(a+h, b)$ and $(a, b+k)$ for some sufficiently small $h, k \in \mathbb{R}$. Then

$$f\big((a+h, b)\big) - f(c) = f(c+he_1) - f(c) \approx f(c) + \frac{df}{dx}\Big|_{(a,b)} h - f(c) + \text{h.o.t.} = \frac{df}{dx}\Big|_{(a,b)} h + \text{h.o.t.} \tag{4.4}$$

where we have taken the derivative of f viewing it as a function of just x (since the second coordinate is fixed) and h.o.t. stands for "higher order terms." Similarly,

$$f\big((a, b+k)\big) - f(c) = f(c+ke_2) - f(c) \approx f(c) + \frac{df}{dy}\Big|_{(a,b)} k - f(c) + \text{h.o.t.} = \frac{df}{dy}\Big|_{(a,b)} k + \text{h.o.t.} \tag{4.5}$$

Since a general point is a linear combination of these two unit vectors,

$$f\big((a+h, b+k)\big) - f(a, b) = f\big((a, b) + he_1 + ke_2\big) - f(a, b) \approx \frac{df}{dx}\Big|_{(a,b)} h + \frac{df}{dy}\Big|_{(a,b)} k + \text{h.o.t.} \tag{4.6}$$

In other words,

$$f\big((a+h,b+k)\big)-f(a,b)\approx\begin{bmatrix}\frac{df}{dx}\Big|_{(a,b)} & \frac{df}{dy}\Big|_{(a,b)}\end{bmatrix}\begin{bmatrix}h\\k\end{bmatrix}+\text{h.o.t.},\tag{4.7}$$

which shows how the derivative should be appropriately interpreted as a matrix acting on vectors in a neighborhood of the point (a,b).

Theorem 4.8. *Let $f:\mathbb{R}^n\to\mathbb{R}^m$ be differentiable at $c\in\mathbb{R}^n$ and suppose that $\mu:\mathbb{R}^n\to\mathbb{R}^m$ is a linear transformation that satisfies*

$$\lim_{x\to c}\frac{\big|f(x)-f(c)-\mu(x-c)\big|}{|x-c|}=0.\tag{4.9}$$

Then $\mu=D_cf$, i.e. if f is differentiable, then its differential is unique.

Proof. For ease of notation, let $\lambda:=D_cf$ and let $g:\mathbb{R}^n\to\mathbb{R}^m$ be defined by $g(x):=f(x)-f(c)$. Then,

$$\begin{aligned}\lim_{x\to c}\frac{\big|\mu(x-c)-\lambda(x-c)\big|}{|x-c|}&=\lim_{x\to c}\frac{\Big|\Big(g(x)-\lambda(x-c)\Big)+\Big(\mu(x-c)-g(x)\Big)\Big|}{|x-c|}\\&\le\lim_{x\to c}\frac{\big|g(x)-\lambda(x-c)\big|}{|x-c|}+\lim_{x\to c}\frac{\big|g(x)-\mu(x-c)\big|}{|x-c|}\\&=0.\end{aligned}\tag{4.10}$$

Since the quantity $\dfrac{\big|\mu(x-c)-\lambda(x-c)\big|}{|x-c|}$ is non-negative for all $x\in\mathbb{R}^n\setminus\{c\}$,

$$\lim_{x\to c}\frac{\big|\mu(x-c)-\lambda(x-c)\big|}{|x-c|}=0\tag{4.11}$$

by the squeeze theorem. By Exercise 3.20, this is equivalent to

$$\lim_{h\to0}\frac{\big|\mu(h)-\lambda(h)\big|}{|h|}=0\tag{4.12}$$

Let $v\in\mathbb{R}^n\setminus\{0\}$. Then

$$\begin{tikzcd}
 & \lim\limits_{t\to0}\dfrac{\big|\mu(tv)-\lambda(tv)\big|}{|tv|} \arrow[dl, equal, "\text{by linearity of } \lambda \text{ and } \mu"'] \arrow[dr, equal, "\text{by Exercise 3.23}"] & \\
\lim\limits_{t\to0}\dfrac{|t|\big|\mu(v)-\lambda(v)\big|}{|t||v|} \arrow[dr, equal] & & 0 \\
 & \dfrac{\big|\mu(v)-\lambda(v)\big|}{|v|} &
\end{tikzcd}\tag{4.13}$$

By this equality and since v is non-zero, it must be that $\mu(v)=\lambda(v)$. Since this is true for all $v\in\mathbb{R}^n\setminus\{0\}$ and because by linearity, $\mu(0)=0=\lambda(0)$, it follows that $\mu=\lambda$. ■

Exercise 4.14. Show that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable* at $c \in \mathbb{R}^n$ if and only if there exists a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\big|f(c+h) - f(c) - \lambda(h)\big|}{|h|} = 0. \tag{4.15}$$

How is λ related to $D_c f$?

The following fact should be compared to Exercise 3.28.

Exercise 4.16. Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ be two functions and let $c \in \mathbb{R}^n$. Suppose that f and g are both differentiable at c and suppose that there exists an open set $U \subseteq \mathbb{R}^n$ containing c such that $f = g$ on U. Show that $D_c f = D_c g$.

This shows that differentiability of a function at a point only depends on the values of the function in a neighborhood of that point, i.e. it is a local property. In fact, given any function $f : U \to \mathbb{R}^m$ on an open $U \subseteq \mathbb{R}^n$ that is differentiable at $c \in U$, the differential $D_c f$ is defined and is a linear function $D_c f : \mathbb{R}^n \to \mathbb{R}^m$. More precisely, we have the following definition.

Definition 4.17. Let $U \subseteq \mathbb{R}^n$ be an open set, let $c \in U$, and let $f : U \to \mathbb{R}^m$ be a function. f is *differentiable* at $c \in U$ iff there exists a linear transformation $D_c f : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\lim_{x \to c} \frac{\big|f(x) - f(c) - (D_c f)(x - c)\big|}{|x - c|} = 0. \tag{4.18}$$

$D_c f$ is called the *differential* of f at c.

One can define differentiability of functions on more general domains besides just Euclidean space or open sets [13], [8].

Definition 4.19. Let $A \subseteq \mathbb{R}^n$ and let $c \in A$. A function $f : A \to \mathbb{R}^m$ is *differentiable* at c if and only if there exists an open set U and a function $\tilde{f} : U \to \mathbb{R}^m$ such that

i) $c \in U$

ii) $\tilde{f} = f$ on $U \cap A$

iii) and $\tilde{f}$ is differentiable at c.

Similarly, $f : A \to \mathbb{R}^m$ is *differentiable* on A if and only if there exists an open set U and a function $\tilde{f} : U \to \mathbb{R}^m$ such that

i) $A \subseteq U$

ii) $\tilde{f} = f$ on A

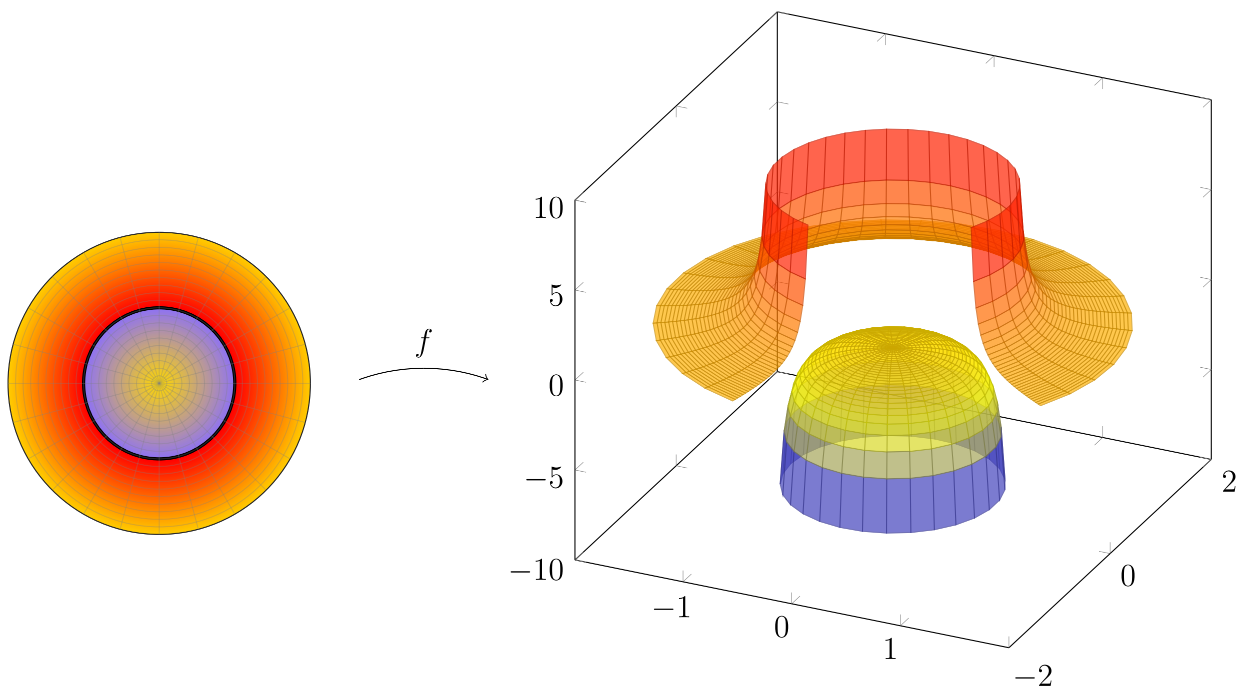
iii) and $\tilde{f}$ is differentiable on U.

Exercise 4.20. Let $A \subseteq \mathbb{R}^m$ and let $f : A \to \mathbb{R}^n$. Show that f is differentiable on A if and only if f is differentiable at c for every $c \in A$.

The definition of differentiability of a function $f : A \to \mathbb{R}^m$ on some arbitrary subset $A \subseteq \mathbb{R}^n$ is a bit meaningless because there is no claim to the uniqueness of the differential. Indeed, consider the case where A is a single point $A = \{a\}$. Then $f : A \to \mathbb{R}^m$ can be extended to an open neighborhood of A in so many ways that the differential can be *any* linear transformation $\mathbb{R}^n \to \mathbb{R}^m$. As another interesting case, given a differentiable function on the Cantor set, what should its differential be? We will see that there *are* suitable subsets A for which the differential has a satisfactory meaning. Such subsets are called *manifolds* and we will study them in a few lectures.

Example 4.21. Let $f : \mathbb{R}^2 \setminus S^1 \to \mathbb{R}$ be the function

$$\mathbb{R}^2 \setminus S^1 \ni (x, y) \mapsto f(x, y) := \frac{1}{x^2 + y^2 - 1}. \tag{4.22}$$



Then f is differentiable on $\mathbb{R}^2 \setminus S^1$ and the differential of f at $(a, b) \in \mathbb{R}^2 \setminus S^1$ is the linear transformation $D_{(a,b)}f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\mathbb{R}^2 \ni (u, v) \mapsto \big(D_{(a,b)}f\big)(u, v) = \frac{-2(au + bv)}{(a^2 + b^2 - 1)^2}. \tag{4.23}$$

It is not only difficult but also un-enlightening to prove that this linear transformation is indeed the differential of f at (a, b). We will be able to differentiate functions in a much simpler manner when we prove some theorems of algebraic nature and then even more so when we discuss partial derivatives.

Because the differential $D_c f : \mathbb{R}^n \to \mathbb{R}^m$ of $f : \mathbb{R}^n \to \mathbb{R}^m$ at $c \in \mathbb{R}^n$ is a linear transformation, one can use the standard basis to express it as a matrix.

Definition 4.24. Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function that is differentiable at c. The *Jacobian matrix* of f at c is the $m \times n$ matrix $[D_c f]$ corresponding to $D_c f$ with respect to the standard basis in Euclidean space, namely

$$[D_cf] := \begin{bmatrix} \big\langle e_1, (D_cf)(e_1)\big\rangle & \big\langle e_1, (D_cf)(e_2)\big\rangle & \cdots & \big\langle e_1, (D_cf)(e_n)\big\rangle \\ \big\langle e_2, (D_cf)(e_1)\big\rangle & \big\langle e_2, (D_cf)(e_2)\big\rangle & \cdots & \big\langle e_2, (D_cf)(e_n)\big\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \big\langle e_m, (D_cf)(e_1)\big\rangle & \big\langle e_m, (D_cf)(e_2)\big\rangle & \cdots & \big\langle e_m, (D_cf)(e_n)\big\rangle \end{bmatrix}. \tag{4.25}$$

Theorem 4.26 (Chain Rule)**.** *Let $A \subseteq \mathbb{R}^p$ and $B \subseteq \mathbb{R}^n$ be open sets, let $f : A \to \mathbb{R}^n$ with $f(A) \subseteq B$, and let $g : B \to \mathbb{R}^m$. Let $c \in A$. Suppose that f is differentiable at c with differential D_cf and g is differentiable at $f(c)$ with differential $D_{f(c)}g$. Then $g \circ f$ is differentiable at c with differential*

$$D_c(g \circ f) = D_{f(c)}g \circ D_cf, \tag{4.27}$$

i.e. the diagram

$$\begin{tikzcd} & \mathbb{R}^n \arrow[dl, "D_{f(c)}g"'] & \\ \mathbb{R}^m & & \mathbb{R}^p \arrow[ul, "D_cf"'] \arrow[ll, "D_c(g\circ f)"] \end{tikzcd} \tag{4.28}$$

of linear transformations commutes.

The following proof comes from [9].[2]

Proof. The goal is to show that

$$\lim_{h\to 0} \frac{\Big|g\big(f(c+h)\big) - g\big(f(c)\big) - \big(D_{f(c)}g\big)\big((D_cf)(h)\big)\Big|}{|h|} = 0. \tag{4.29}$$

By assumption that g is differentiable at $f(c)$ and B is open, there exists an $\epsilon > 0$ such that $V_\epsilon\big(f(c)\big) \subseteq B$. Since f is continuous at a and A is open, there exists a $\delta > 0$ such that $V_\delta(a) \subseteq A$ and $f\big(V_\delta(a)\big) \subseteq V_\epsilon\big(f(c)\big)$. Let $\Delta_f : V_\delta(0) \to \mathbb{R}^n$ and $\Delta_g : V_\epsilon(0) \to \mathbb{R}^m$ be the functions defined by

$$V_\delta(0) \ni h \mapsto \Delta_f(h) := f(c+h) - f(c) \tag{4.30}$$

and

$$V_\epsilon(0) \ni k \mapsto \Delta_g(k) := g\big(f(c)+k\big) - g\big(f(c)\big). \tag{4.31}$$

In addition, let $F : V_\delta(0) \to \mathbb{R}^n$ and $G : V_\epsilon(0) \to \mathbb{R}^m$ be the functions defined by

$$V_\delta(0) \ni h \mapsto F(h) := \begin{cases} 0 & \text{for } h = 0 \\ \Delta_f(h) - (D_cf)(h) & \text{for } h \text{ satisfying } 0 < h < \delta \end{cases} \tag{4.32}$$

[2] Unfortunately, I can't seem to find a proof that is intuitive. The only intuition I can extract is by doing a "physics" calculation

$$g\big(f(c+h)\big) - g\big(f(c)\big) \approx g\big(f(c)+(D_cf)(h)\big) - g\big(f(c)\big) \approx g\big(f(c)\big) + \big(D_{f(c)}g\big)\big((D_cf)(h)\big) - g\big(f(c)\big) = \big(D_{f(c)}g\big)\big((D_cf)(h)\big),$$

which only gives us a good guess for what the formula for the derivative should be, but not any intuition for the proof that follows.

and

$$V_\epsilon(0) \ni k \mapsto G(k) := \begin{cases} 0 & \text{for } k = 0 \\ \Delta_g(k) - (D_{f(c)}g)(k) & \text{for } k \text{ satisfying } 0 < k < \epsilon \end{cases} \tag{4.33}$$

Finally, let $\varphi : V_\delta(0) \to \mathbb{R}^n$ be the function defined by

$$V_\delta(0) \ni h \mapsto \varphi(h) := g\big(f(c+h)\big) - g\big(f(c)\big) - \big(D_{f(c)}g\big)\big((D_c f)(h)\big). \tag{4.34}$$

Note that with these definitions,

$$\frac{\varphi(h)}{|h|} = \big(D_{f(c)}g\big)\big(F(h)\big) + \frac{\big|\Delta_f(h)\big|}{|h|} G\big(\Delta_f(h)\big) \qquad \forall\, h \in V_\delta(0) \setminus \{0\}. \tag{4.35}$$

The proof of the theorem will be complete if it is shown that the limit of the right-hand-side is 0 as $h \to 0$. By Exercise 1.10 in [13],there exists a real number $M \geq 0$ such that

$$\big|(D_c f)(h)\big| \leq M|h| \qquad \forall\, h \in \mathbb{R}^p. \tag{4.36}$$

Hence, by the triangle inequality and this result,

$$\begin{aligned} \frac{\big|\varphi(h)\big|}{|h|} &= \left| \big(D_{f(c)}g\big)\big(F(h)\big) + \frac{\big|D_c f)(h) + F(h)\big|}{|h|} G\big(\Delta_f(h)\big) \right| && \text{by Def'n of } \varphi \text{ and } F \\ &\leq \Big| \big(D_{f(c)}g\big)\big(F(h)\big) \Big| + \frac{\big|(D_c f)(h)\big| + \big|F(h)\big|}{|h|} \Big| G\big(\Delta_f(h)\big) \Big| && \text{by triangle inequality} \\ &\leq \Big| \big(D_{f(c)}g\big)\big(F(h)\big) \Big| + \left(M + \frac{\big|F(h)\big|}{|h|} \right) \Big| G\big(\Delta_f(h)\big) \Big| && \text{by Exercise 1.10 in [13].} \end{aligned} \tag{4.37}$$

Since F and G are continuous at 0 and equal 0 there, since $\lim\limits_{h\to 0} \Delta_f(h) = 0$, since $\lim\limits_{h\to 0} \frac{|F(h)|}{|h|}$ is bounded (since the derivative of f exists at c), since linear transformations are continuous, and since the absolute value is a continuous function,

$$\lim_{h \to 0} \left(\Big| \big(D_{f(c)}g\big)\big(F(h)\big) \Big| + \left(M + \frac{\big|F(h)\big|}{|h|} \right) \Big| G\big(\Delta_f(h)\big) \Big| \right) = 0 \tag{4.38}$$

by the Algebraic Limit Theorem. This concludes the proof of the Chain rule. ■

Exercise 4.39. In the proof of the Chain Rule, show that

$$\frac{\varphi(h)}{|h|} = \big(D_{f(c)}g\big)\big(F(h)\big) + \frac{\big|\Delta_f(h)\big|}{|h|} G\big(\Delta_f(h)\big) \qquad \forall\, h \in V_\delta(0) \setminus \{0\}. \tag{4.40}$$

Exercise 4.41. In the last step in the proof of the Chain Rule, carefully and explicitly describe all the steps needed to conclude that

$$\lim_{h \to 0} \left(\Big| \big(D_{f(c)}g\big)\big(F(h)\big) \Big| + \left(M + \frac{\big|F(h)\big|}{|h|} \right) \Big| G\big(\Delta_f(h)\big) \Big| \right) = 0. \tag{4.42}$$

Theorem 4.43 (Algebraic Differentiability Theorem)**.**

(a) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function, then $D_c f = 0$ for all $c \in \mathbb{R}^n$.

(b) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $D_c f = f$ for all $c \in \mathbb{R}^n$.

(c) A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $c \in \mathbb{R}^n$ if and only if the component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are differentiable at c for all $i \in \{1, \dots, m\}$. Furthermore,

$$(D_c f)(v) = \big((D_c f_1)(v), (D_c f_2)(v), \dots, (D_c f_m)(v)\big) \tag{4.44}$$

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

(d) If $s : \mathbb{R}^2 \to \mathbb{R}$ is the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto s(x, y) := x + y, \tag{4.45}$$

then $D_c s = s$ for all $c \in \mathbb{R}^2$.

(e) If $p : \mathbb{R}^2 \to \mathbb{R}$ is the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto p(x, y) := xy, \tag{4.46}$$

then

$$(D_{(a,b)} p)(u, v) = bu + av \tag{4.47}$$

for all $(a, b) \in \mathbb{R}^2$ and $(u, v) \in \mathbb{R}^2$.

(f) If $f, g : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $c \in \mathbb{R}^n$, then $f + g$ is differentiable at c and

$$D_c(f + g) = D_c f + D_c g. \tag{4.48}$$

(g) If $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $c \in \mathbb{R}^n$, then fg is differentiable at c and

$$D_c(fg) = g(c) D_c f + f(c) D_c g. \tag{4.49}$$

(h) If $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $c \in \mathbb{R}^n$ and $g(c) \neq 0$, then there exists an open set $U \subseteq \mathbb{R}^n$ with $c \in U$ such that $\frac{f}{g}$ is defined on U and differentiable at c and

$$D_c\left(\frac{f}{g}\right) = \frac{g(c) D_c f - f(c) D_c g}{\big(g(c)\big)^2}. \tag{4.50}$$

Part (g) is known as the <u>*product rule*</u> and part (h) is known as the <u>*quotient rule*</u>.

Proof. See [13]. ∎

Exercise 4.51. Let $f : \mathbb{R}^{n_1} \to \mathbb{R}^{m_1}$ and $g : \mathbb{R}^{n_2} \to \mathbb{R}^{m_2}$ be two functions. Let f be differentiable at $c_1 \in \mathbb{R}^{n_1}$ and g be differentiable at $c_2 \in \mathbb{R}^{n_2}$. Show that the function $f \times g : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{m_1} \times \mathbb{R}^{m_1}$ defined by

$$\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \ni (x, y) \mapsto (f \times g)(x, y) := \big(f(x), g(y)\big) \tag{4.52}$$

is differentiable at $(c_1, c_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$.

Exercise 4.53. Let $a \in \mathbb{R}^m$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that the function $f_a : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$\mathbb{R}^n \ni x \mapsto f_a(x) = a + f(x) \tag{4.54}$$

is differentiable and $D_c f_a = f$ for all $c \in \mathbb{R}^n$. Use this to conclude that $\gamma_{c;v} : \mathbb{R} \to \mathbb{R}^n$ (see Exercise 3.23 for notation) is differentiable for all $c \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ and its differential satisfies $(D_0\gamma_{c;v})(1) = v$.

Exercise 4.55. Prove that the j-th projection $\pi_j : \mathbb{R}^m \to \mathbb{R}$ is differentiable and show that

$$D_c\pi_j = \langle e_j, \,\cdot\, \rangle \tag{4.56}$$

for all $c \in \mathbb{R}^m$, i.e.

$$(D_c\pi_j)(v) = \langle e_j, v\rangle \tag{4.57}$$

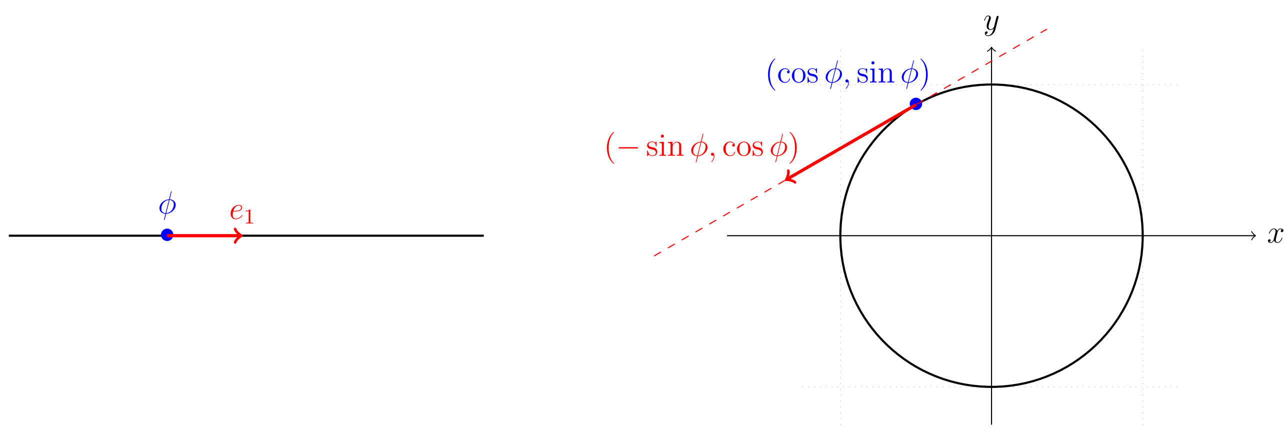
for all $v \in \mathbb{R}^m$ and for all $c \in \mathbb{R}^m$.

Example 4.58. Let $\gamma : \mathbb{R} \to \mathbb{R}^2$ be the function

$$\mathbb{R} \ni \theta \mapsto \gamma(\theta) := (\cos\theta, \sin\theta) \tag{4.59}$$

and let $\phi \in \mathbb{R}$. Then γ is differentiable at ϕ and $D_\phi\gamma : \mathbb{R} \to \mathbb{R}^2$ is the linear transformation

$$\mathbb{R} \ni v \mapsto (D_\phi\gamma)(v) = (-v\sin\phi, v\cos\phi). \tag{4.60}$$



The image on the right is magnified by a factor of 2 for better viewing purposes but the length of the unit vector e_1 at ϕ does not change in the image. The fact that $D_\phi\gamma$ as written is the correct differential follows from part (c) of Theorem 4.43 and the formulas for the derivatives of sine and cosine.

Example 4.61. Let $f : (0,\infty) \times \mathbb{R} \to \mathbb{R}^2$ be the function

$$(0,\infty) \times \mathbb{R} \ni (r,\theta) \mapsto f(r,\theta) := (r\cos\theta, r\sin\theta) \tag{4.62}$$

and let $(R,\phi) \in (0,\infty) \times \mathbb{R}$. Then

$$f_1(r,\theta) = r\cos\theta = p\big(\mathrm{id}(r), \cos(\theta)\big) \qquad \& \qquad f_2(r,\theta) = r\sin\theta = p\big(\mathrm{id}(r), \sin(\theta)\big), \tag{4.63}$$

where $p : \mathbb{R}^2 \to \mathbb{R}$ is the product function from part (g) of Theorem 4.43. Diagrammatically, these compositions can be written as

$$\begin{tikzcd} & (0,\infty)\times\mathbb{R} \arrow[dl, "p"'] & \\ \mathbb{R} & & (0,\infty)\times\mathbb{R} \arrow[ll, "f_1"] \arrow[ul, "\mathrm{id}\times\cos"'] \end{tikzcd} \quad \& \quad \begin{tikzcd} & (0,\infty)\times\mathbb{R} \arrow[dl, "p"'] & \\ \mathbb{R} & & (0,\infty)\times\mathbb{R} \arrow[ll, "f_2"] \arrow[ul, "\mathrm{id}\times\sin"'] \end{tikzcd} . \tag{4.64}$$

By the chain rule (Theorem 4.26), the diagrams

$$\begin{tikzcd} & \mathbb{R}^2 \arrow[dl, "D_{(R,\cos\phi)}p"'] & \\ \mathbb{R} & & \mathbb{R}^2 \arrow[ll, "D_{(R,\phi)}f_1"] \arrow[ul, "D_{(R,\phi)}(\mathrm{id}\times\cos)"'] \end{tikzcd} \quad \& \quad \begin{tikzcd} & \mathbb{R}^2 \arrow[dl, "D_{(R,\sin\phi)}p"'] & \\ \mathbb{R} & & \mathbb{R}^2 \arrow[ll, "D_{(R,\phi)}f_2"] \arrow[ul, "D_{(R,\phi)}(\mathrm{id}\times\sin)"'] \end{tikzcd} \tag{4.65}$$

commute. By Exercise 4.51, these diagrams become

$$\begin{tikzcd} & \mathbb{R}^2 \arrow[dl, "D_{(R,\cos\phi)}p"'] & \\ \mathbb{R} & & \mathbb{R}^2 \arrow[ll, "D_{(R,\phi)}f_1"] \arrow[ul, "(D_R\mathrm{id})\times(D_\phi\cos)"'] \end{tikzcd} \quad \& \quad \begin{tikzcd} & \mathbb{R}^2 \arrow[dl, "D_{(R,\sin\phi)}p"'] & \\ \mathbb{R} & & \mathbb{R}^2 \arrow[ll, "D_{(R,\phi)}f_2"] \arrow[ul, "(D_R\mathrm{id})\times(D_\phi\sin)"'] \end{tikzcd} . \tag{4.66}$$

By parts (b) and (e) of Theorem 4.43, problem 2-2 in [13], and together with knowledge of the derivatives of sines and cosines, applying these linear transformations to $(u, v) \in \mathbb{R}^2$ gives

$$\begin{tikzcd} & (u,-v\sin\phi) \arrow[dl, mapsto, "D_{(R,\cos\phi)}p"'] & \\ u\cos\phi - vR\sin\phi & & (u,v) \arrow[ll, mapsto, "D_{(R,\phi)}f_1"] \arrow[ul, mapsto, "\mathrm{id}\times(-\,\cdot\,\sin\theta)"'] \end{tikzcd} \quad \& \quad \begin{tikzcd} & (u,v\cos\phi) \arrow[dl, mapsto, "D_{(R,\sin\phi)}p"'] & \\ u\sin\phi + vR\cos\phi & & (u,v) \arrow[ll, mapsto, "D_{(R,\phi)}f_2"] \arrow[ul, mapsto, "\mathrm{id}\times(\,\cdot\,\cos\phi)"'] \end{tikzcd} . \tag{4.67}$$

Thus, f is differentiable at (R, ϕ) and $D_{(R,\phi)}f : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear transformation

$$\mathbb{R}^2 \ni (u,v) \mapsto (D_{(R,\phi)}f)(u,v) = \big(u\cos\phi - vR\sin\phi, u\sin\phi + vR\cos\phi\big). \tag{4.68}$$

Exercise 4.69. Using only the results known so far, calculate the differential of the inner product $\langle\,\cdot\,,\,\cdot\,\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ at the point 0. Prove that $\langle\,\cdot\,,\,\cdot\,\rangle$ is differentiable everywhere. Use this to prove that the norm squared $|\,\cdot\,| : \mathbb{R}^n \to \mathbb{R}$ is differentiable and calculate its differential.

As you might have guessed from parts (f), (g), and (h) of Theorem 4.43, for any $c \in \mathbb{R}^n$, the entity D_c seems to have meaning of its own independent of any differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ that is put in to produce $D_c f$. What type of mathematical object is D_c? It takes in functions that are differentiable at c and it provides linear transformations. We will explore this in further detail soon. D_c is an example of a *derivation*.

Level I problems.

From Spivak [13]: 2.1, 2.2, 2.3, 2.5, 2.6, 2.7, 2.10 parts (a) and (b), 2.10 parts (c) and (d), 2.10 part (e), 2.10 parts (f) and (g), 2.10 parts (h) and (i), 2.10 part (j), 2.11 part (a), 2.11 part (b), 2.11 part (c)
From these notes: Exercise 4.14, 4.16, 4.39, 4.51, 4.53, 4.55

Level II problems.
From Spivak [13]: 2.4, 2.9, 2.12, 2.13, 2.14, 2.15, 2.16
From these notes: Exercise 4.20, 4.41, 4.69

5 February 2: Partial Derivatives

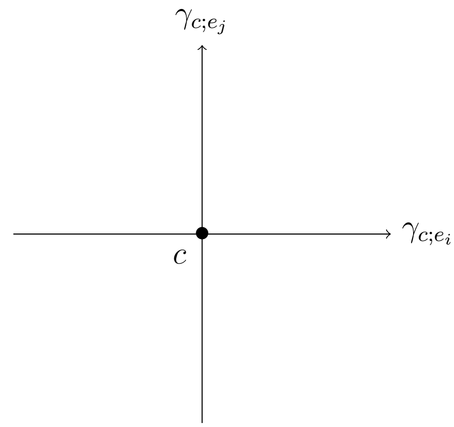
In Exercise 3.28, we introduced the inclusion functions $\varphi_i : \mathbb{R} \to \mathbb{R}^n$ onto the i-th coordinate. Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the resulting function $f \circ \varphi_i : \mathbb{R} \to \mathbb{R}^m$ is given by

$$\mathbb{R} \ni x_i \mapsto (f \circ \varphi_i)(x_i) := f(0, \ldots, 0, x_i, 0, \ldots, 0), \tag{5.1}$$

where the non-zero entry is in the i-th slot. In that same exercise, we saw that this was a special case of precomposing f with a straight curve through an arbitrary point. Recall, if $c \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ is a vector, let $\gamma_{c;v} : \mathbb{R} \to \mathbb{R}^n$ denote the curve

$$\mathbb{R} \ni t \mapsto \gamma_{c;v}(t) := c + tv. \tag{5.2}$$

As a special case, consider the vectors $\{e_i\}$ for all $i \in \{1, \ldots, n\}$ and the corresponding curves $\{\gamma_{c;e_i}\}$. These draw out a "coordinate axes" at the point c.



Precomposing f with any of these curves gives a function $f \circ \gamma_{c;e_i}$ of a single variable, namely, the i-th coordinate

$$\mathbb{R} \ni t \mapsto (f \circ \gamma_{c;e_i})(x_i) = f(c_1, \ldots, c_{i-1}, c_i + t, c_{i+1}, \ldots, c_n). \tag{5.3}$$

Here c_j denotes the j-th coordinate of c.

Definition 5.4. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $c \in \mathbb{R}^n$. The *i-th partial derivative of f at c* is the derivative of the function $f \circ \gamma_{c;e_i} : \mathbb{R} \to \mathbb{R}$ at 0, if it exists. It is denoted by[3]

$$(\partial_i f)(c) := \big(D_0(f \circ \gamma_{c;e_i})\big)(1) \equiv (f \circ \gamma_{c;e_i})'(0). \tag{5.5}$$

Explicitly, it is given by

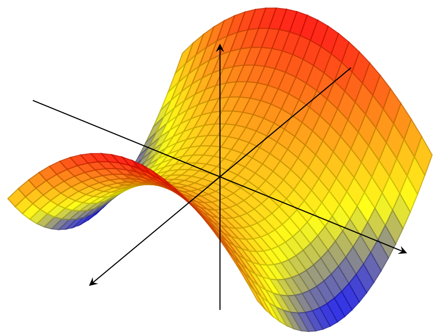
$$\lim_{t \to 0} \frac{f(c_1, \ldots, c_{i-1}, c_i + t, c_{i+1}, \ldots, c_n) - f(c_1, \ldots, c_{i-1}, c_i, c_{i+1}, \ldots, c_n)}{t}. \tag{5.6}$$

[3] All of this notation expresses the same mathematical entity. The first equality defines the left-hand-side, the middle expression is in terms of differentials, and the final expression is in terms of the notation for derivatives from our first semester of analysis.

Thus, the partial derivative of $f : \mathbb{R}^n \to \mathbb{R}$ at c is the slope of f when restricted to the i-th coordinate axis at the point c. Notice that it is *not* assumed that f is differentiable. The reason we are only focusing on functions whose codomain is $\mathbb{R}$ is simply for convenience (consult part (c) of Theorem 4.43) but also for the relationship to the slope from calculus on $\mathbb{R}$.

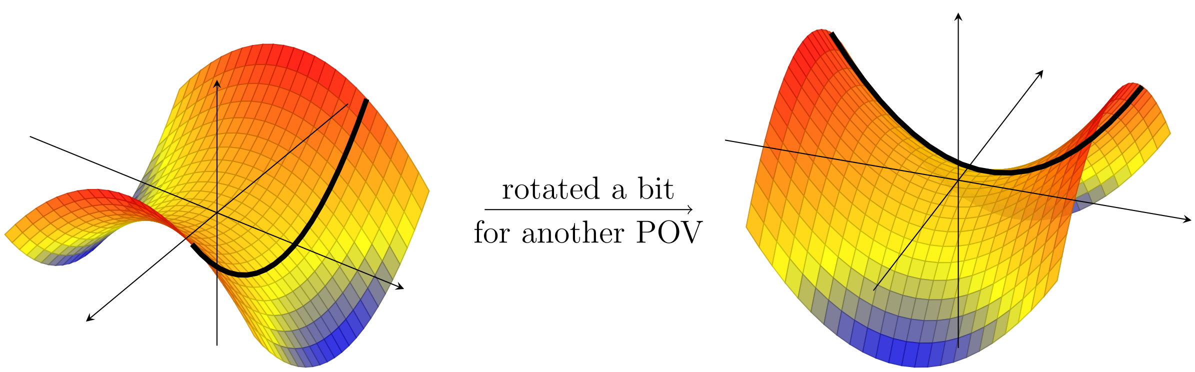
Example 5.7. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 - y^2. \tag{5.8}$$



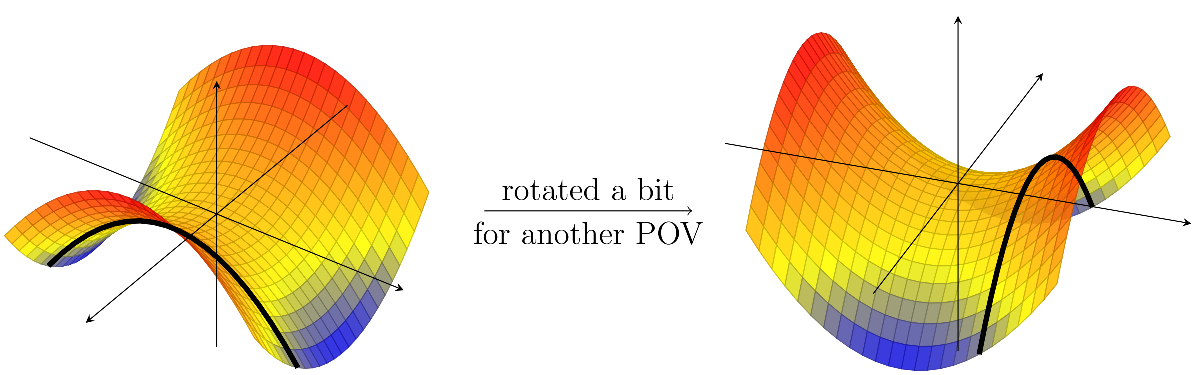
Let $c = (1, 1)$. Then, the functions $f \circ \gamma_{c;e_i}$ are given by

$$(f \circ \gamma_{c;e_1})(t) = f(1 + t, 1) = (1 + t)^2 - 1^2 = 2t + t^2 \tag{5.9}$$



and

$$(f \circ \gamma_{c;e_2})(t) = f(1, 1 + t) = 1^2 - (1 + t)^2 = -2t - t^2. \tag{5.10}$$



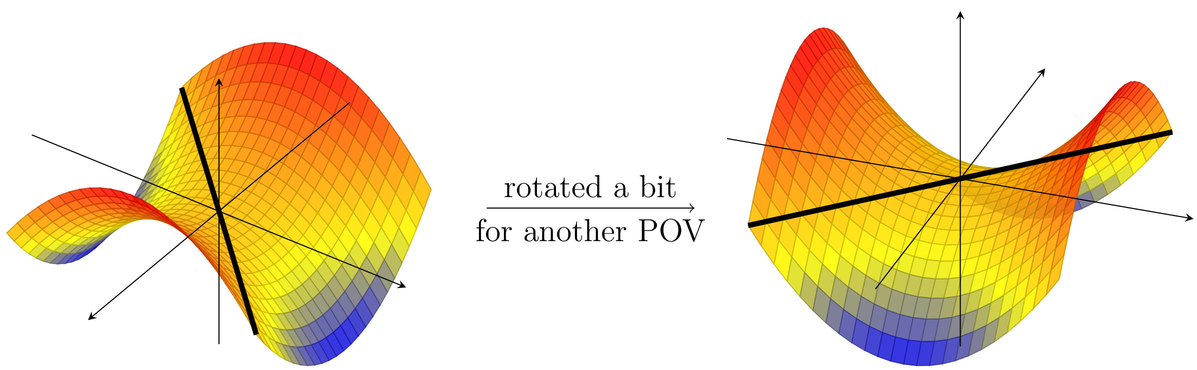
Be careful, these functions are *not* the same as $\mathbb{R} \ni x \mapsto f(x,1) = x^2 - 1$ and $\mathbb{R} \ni y \mapsto f(1,y) = 1-y^2$, respectively, though they are closely related (we will see exactly how in a moment). Anyway, the partial derivatives of f are therefore

$$(\partial_1 f)(c) = \lim_{t\to 0} \frac{f(1+t,1) - f(1,1)}{t} = \lim_{t\to 0} \frac{(1+t)^2 - 1^2 - 1^2 + 1^2}{t} = \lim_{t\to 0} \frac{2t}{t} = 2 \tag{5.11}$$

and

$$(\partial_2 f)(c) = \lim_{t\to 0} \frac{f(1,1+t) - f(1,1)}{t} = \lim_{t\to 0} \frac{1^2 - (1+t)^2 - 1^2 + 1^2}{t} = \lim_{t\to 0} \frac{-2t}{t} = -2. \tag{5.12}$$

Now consider the vector $v = (1,1)$ and the associated curve $\gamma_{c;v}$. This curve can be depicted on the graph of f as



as one can indeed show that $f \circ \gamma_{c;v} = 0$ is the zero function.

Our definition of the partial derivative is slightly different from Spivak's. Let us verify that the two definitions are the same.

Theorem 5.13. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $c \in \mathbb{R}^n$. Let $\varphi_{c;i} : \mathbb{R} \to \mathbb{R}^n$ be the function*

$$\mathbb{R} \ni h \mapsto \varphi_{c;i}(h) := (c_1, \ldots, c_{i-1}, h, c_{i+1}, \ldots, c_n). \tag{5.14}$$

The i-th partial derivative of f at c exists if and only if $f \circ \varphi_{c;i}$ is differentiable at c_i. Furthermore, when this happens,

$$(\partial_i f)(c) = (f \circ \varphi_{c;i})'(c_i). \tag{5.15}$$

Proof. By definition of the ordinary derivative

$$(f \circ \varphi_{c;i})'(c_i) = \lim_{h\to 0} \frac{f(c_1, \ldots, c_{i-1}, c_i + h, c_{i+1}, \ldots, c_n) - f(c_1, \ldots, c_{i-1}, c_i, c_{i+1}, \ldots, c_n)}{h}, \tag{5.16}$$

which is exactly the same as the expression in (5.6) in Definition 5.4. ■

The reason we use the definition of the partial derivative in terms of curves will become much more important when we consider derivatives of functions defined on subsets of $\mathbb{R}^n$ that are not necessarily open. In particular, the subsets we will look at will often be path-connected but they might not be "rectangular" and contain straight lines. Take for example the unit circle S^1 in $\mathbb{R}^2$. What should the derivative of a function $f : S^1 \to \mathbb{R}$ be? We will come back to this point again later.

Exercise 5.17. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $c \in \mathbb{R}^n$ and let $v \in \mathbb{R}^n$ be any vector. Show that $f \circ \gamma_{c;v} : \mathbb{R} \to \mathbb{R}^m$ is differentiable at 0 and

$$\big(D_0(f \circ \gamma_{c;v})\big)(1) = (D_c f)(v). \tag{5.18}$$

Another way to write this result is

$$(f \circ \gamma_{c;v})'(0) = (D_c f)(v). \tag{5.19}$$

Theorem 5.20. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $c \in \mathbb{R}^n$. Then the partial derivatives of the component function f_j at c exist for all $j \in \{1, \dots, m\}$. In fact,*

$$(\partial_i f_j)(c) = (D_c f_j)(e_i) \tag{5.21}$$

for all $i \in \{1, \dots, n\}$ and all $j \in \{1, \dots, m\}$. Furthermore,

$$(\partial_i f_j)(c) = \langle e_j, (D_c f)(e_i)\rangle \tag{5.22}$$

so that

$$[D_c f] = \begin{bmatrix} (\partial_1 f_1)(c) & (\partial_2 f_1)(c) & \cdots & (\partial_n f_1)(c) \\ (\partial_1 f_2)(c) & (\partial_2 f_2)(c) & \cdots & (\partial_n f_2)(c) \\ \vdots & \vdots & \ddots & \vdots \\ (\partial_1 f_m)(c) & (\partial_2 f_m)(c) & \cdots & (\partial_n f_m)(c) \end{bmatrix}. \tag{5.23}$$

Proof. Existence of the partial derivatives follows from Exercise 5.17. The explicit formula follows from the chain rule. To see this, first note that

$$(D_0 \gamma_{c;e_i})(1) = e_i \tag{5.24}$$

by Exercise 4.53. By definition of the partial derivative and this previous facts,

$$(\partial_i f_j)(c) = \big(D_0(f_j \circ \gamma_{c;e_i})\big)(1) = (D_c f_j)(e_i). \tag{5.25}$$

By the chain rule and Exercise 4.55,

$$(D_c f_j)(e_i) = (D_{f(c)}\pi_j)\big((D_c f)(e_i)\big) = \langle e_j, (D_c f)(e_i)\rangle. \tag{5.26}$$

■

To verify this with an example, recall Example 5.7.

Exercise 5.27. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by[4]

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ \frac{xy}{x^2+y^2} & \text{otherwise} \end{cases} \tag{5.28}$$

(a) For which vectors v does $(f \circ \gamma_{c;v})'$ exist? Evaluate it when it exists.

[4] The following few exercises are from [9].

(b) Do $\partial_1 f$ and $\partial_2 f$ exist at 0?

(c) Is f differentiable at 0?

(d) Is f continuous at 0?

Exercise 5.29. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ \frac{x^3}{x^2+y^2} & \text{otherwise} \end{cases} \tag{5.30}$$

(a) For which vectors v does $(f \circ \gamma_{c;v})'$ exist? Evaluate it when it exists.

(b) Do $\partial_1 f$ and $\partial_2 f$ exist at 0?

(c) Is f differentiable at 0?

(d) Is f continuous at 0?

Exercise 5.31. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := \sqrt{|xy|}. \tag{5.32}$$

(a) For which vectors v does $(f \circ \gamma_{c;v})'$ exist? Evaluate it when it exists.

(b) Do $\partial_1 f$ and $\partial_2 f$ exist at 0?

(c) Is f differentiable at 0?

(d) Is f continuous at 0?

Exercise 5.33. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ \frac{x|y|}{\sqrt{x^2+y^2}} & \text{otherwise} \end{cases} \tag{5.34}$$

(a) For which vectors v does $(f \circ \gamma_{c;v})'$ exist? Evaluate it when it exists.

(b) Do $\partial_1 f$ and $\partial_2 f$ exist at 0?

(c) Is f differentiable at 0?

(d) Is f continuous at 0?

The converse of Exercise 5.17 is false and is an analogue of the situation for limits (and hence continuity) from Exercise 3.28.

Exercise 5.35. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $c \in \mathbb{R}^n$. Construct an explicit example of a function f for which $f \circ \gamma_{c;v}$ is differentiable at $0 \in \mathbb{R}$ for all $v \in \mathbb{R}^n$ but for which f is not differentiable at c. [Hint: try to find an example with $n = 2$ and $m = 1$.]

Although the previous exercise makes us lose some hope, the following theorem brings our spirits back. But before we can state it, let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable on an open set $U \subseteq \mathbb{R}^n$ with $c \in U$. Consider the function $U \to \mathbb{R}$ defined by

$$U \ni x \mapsto \big\langle e_j, (D_x f)(e_i) \big\rangle = (\partial_i f_j)(x). \tag{5.36}$$

Even if f is differentiable on U, there is no reason for these partial derivatives to satisfy any particular properties such as continuity. Indeed, last semester, we saw that there were examples of functions whose derivatives are not continuous. However, a partial converse to Theorem 5.20 holds.

Theorem 5.37. *Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$. If there exists an open set $U \subseteq \mathbb{R}^n$ with $c \in U$ such that*

(a) $(\partial_i f_j)(x)$ exists for all $x \in U$ and

(b) $\partial_i f_j : U \to \mathbb{R}$ is continuous at c

for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$, then f is differentiable at c.

Proof. For simplicity, set $m = 1$. For f to be differentiable at c, there must exist a linear transformation $D_c f : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$\lim_{h \to 0} \frac{\big|f(c+h) - f(c) - (D_c f)(h)\big|}{|h|} = 0. \tag{5.38}$$

The proof of the theorem requires a lemma, which is just a special case of the Mean Value Theorem in one dimension.[5]

Lemma. Under the same assumptions in the Theorem, for each $h \in V_\epsilon(0) \subseteq \mathbb{R}^n$, let $\epsilon > 0$ be small enough so that $V_\epsilon(c) \subseteq U$. Set[6]

$$\vec{p}_i := \vec{p}_{i-1} + h_i e_i \qquad \forall\, i \in \{1, \dots, n\} \quad \text{and} \quad \vec{p}_0 := c. \tag{5.39}$$

Then, there exist vectors $\vec{q}_1, \dots, \vec{q}_n \in V_\epsilon(c)$ such that $\vec{q}_i$ lies on the straight line between $\vec{p}_{i-1}$ and $\vec{p}_i$ and

$$f(c+h) - f(c) = \sum_{i=1}^{n} (\partial_i f)(\vec{q}_i) h_i. \tag{5.40}$$

Proof of Lemma. Since $(\partial_i f)$ exists for all $x \in U$, the function $f \circ \varphi_{c;i} : \big[\pi_i(\vec{p}_{i-1}), \pi_i(\vec{p}_i)\big] \to \mathbb{R}^m$ is defined and differentiable on a closed interval. The Mean Value Theorem applies to each component so that there exists a $q_i \in \big(\pi_i(\vec{p}_{i-1}), \pi_i(\vec{p}_i)\big)$ such that

$$\big(f \circ \varphi_{c;i}\big)\big(\pi_i(\vec{p}_i)\big) - \big(f \circ \varphi_{c;i}\big)\big(\pi_i(\vec{p}_{i-1})\big) = \big(f \circ \varphi_{c;i}\big)'(q_i) h_i. \tag{5.41}$$

[5] The idea is to rewrite the difference $f(c+h) - f(c)$ for any h in a sufficiently small neighborhood of the origin in $\mathbb{R}^n$ in terms of "one-dimensional" differences in orthogonal directions. The one-dimensional Mean Value Theorem will then be used to express $f(c+h) - f(c)$ in terms of partial derivatives of f evaluated at nearby points. We will then use this fact to construct a linear transformation $D_c f : \mathbb{R}^n \to \mathbb{R}^m$. Finally, the second step will be to use continuity at c to take the limit and prove that it satisfies the requirements of a differential.

[6] The vectors have be drawn on the p_i's to emphasize that they are vectors and not components of a vector.

Set

$$\vec{q}_i := \big(c_1 + h_1, c_2 + h_2, \dots, c_{i-1} + h_{i-1}, c_i + q_i, c_{i+1}, \dots, c_n\big). \tag{5.42}$$

Then,

$$f(\vec{p}_i) - f(\vec{p}_{i-1}) = (\partial_i f)(\vec{q}_i) h_i. \tag{5.43}$$

Since this can be done for all $i \in \{1, \dots, n\}$, a telescoping sum argument gives

$$f(c+h) - f(c) = \sum_{i=1}^{n} f(\vec{p}_i) - f(\vec{p}_{i-1}) = \sum_{i=1}^{n} (\partial_i f)(\vec{q}_i) h_i. \tag{5.44}$$

■

(Back to proof of Theorem.) Since this can be done for each $h \in V_\epsilon(0) \subseteq \mathbb{R}^n$, define $D_c f : V_\epsilon(0) \to \mathbb{R}^m$ to be the function

$$V_\epsilon(0) \ni h \mapsto (D_c f)(h) := \sum_{i=1}^{n} (\partial_i f)(c) h_i. \tag{5.45}$$

For any vector $v \in \mathbb{R}^n$, there exists a vector $h \in V_\epsilon(0)$ and a scalar $\lambda \in \mathbb{R}$ such that $v = \lambda h$. Set $(D_c f)(v) := \lambda (D_c f)(h)$. This defines a linear transformation $D_c f : \mathbb{R}^n \to \mathbb{R}^m$. Using this definition and the Lemma,

$$\begin{aligned} \frac{f(c+h) - f(c) - (D_c f)(h)}{|h|} &= \frac{\sum_{i=1}^{n} (\partial_i f)(\vec{q}_i) h_i - \sum_{i=1}^{n} (\partial_i f)(c) h_i}{|h|} \\ &= \sum_{i=1}^{n} \Big((\partial_i f)(\vec{q}_i) - (\partial_i f)(c) \Big) \frac{h_i}{|h|} \end{aligned} \tag{5.46}$$

Since $(\partial_i f)$ is continuous at c for all $i \in \{1, \dots, n\}$, since $\lim\limits_{h \to 0} \vec{q}_i = c$ this implies

$$\lim_{h \to 0} (\partial_i f)(\vec{q}_i) = (\partial_i f)(c). \tag{5.47}$$

Furthermore, since $\left| \frac{h_i}{|h|} \right| \le 1$,

$$\lim_{h \to 0} \frac{\big| f(c+h) - f(c) - (D_c f)(h) \big|}{|h|} = 0. \tag{5.48}$$

by the Algebraic Limit Theorem. ■

Exercise 5.49. Explain at which point the proof of Theorem 5.37 needs to be modified for when $m \neq 1$ and precisely describe how to modify it.

Definition 5.50. Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$. f is <u>*continuously differentiable at c*</u> iff there exists an open set $U \subseteq \mathbb{R}^n$ with $c \in U$ such that

(a) $(\partial_i f_j)(x)$ exists for all $x \in U$ and

(b) $\partial_i f_j : U \to \mathbb{R}$ is continuous at c

for all i and j. f is *continuously differentiable on* $\mathbb{R}^n$ (or some open subset $U \subseteq \mathbb{R}^n$) iff f is continuously differentiable at c for all $c \in \mathbb{R}^n$ (or for all $c \in U$). The set of all continuously differentiable functions on $\mathbb{R}^n$ is denoted by $C^1(\mathbb{R}^m, \mathbb{R}^n)$ (or $C^1(\mathbb{R}^m, U)$) and such a function is said to be of *class* C^1.

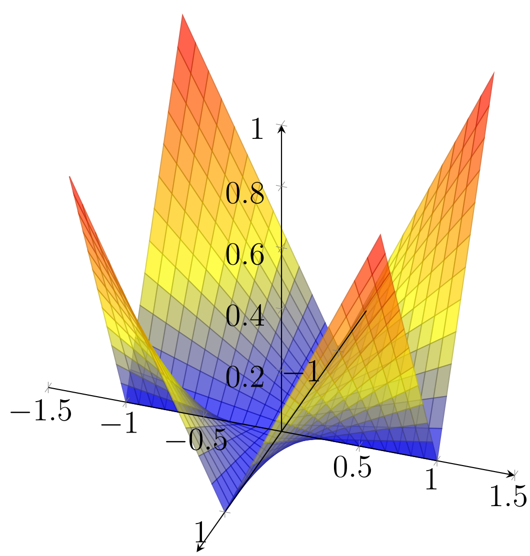
We have already seen in Exercise 5.35 that we cannot drop the continuity of the partial derivatives in Theorem 5.37. However, notice that if the partial derivatives exist, nothing is stopping us from *defining* the $m \times n$ matrix of partial derivatives

$$[D_c f] = \begin{bmatrix} (\partial_1 f_1)(c) & (\partial_2 f_1)(c) & \cdots & (\partial_n f_1)(c) \\ (\partial_1 f_2)(c) & (\partial_2 f_2)(c) & \cdots & (\partial_n f_2)(c) \\ \vdots & \vdots & \ddots & \vdots \\ (\partial_1 f_m)(c) & (\partial_2 f_m)(c) & \cdots & (\partial_n f_m)(c) \end{bmatrix}. \tag{5.51}$$

Hence, a linear transformation $D_c f$ *exists* but the *problem* is that it does not satisfy the limiting condition required for $D_c f$ to be the *differential* of f at c. Indeed, in the previous proof, we explicitly used the assumption that the partial derivatives are continuous to prove that the limit condition is satisfied. We know that the topologist's sine curve is differentiable but not of class C^1. However, it is also a rather complicated function. The following example illustrates that more "reasonable" functions of several variables can be differentiable but not of class C^1.

Exercise 5.52. Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := |xy| \tag{5.53}$$



is differentiable at 0 but is not continuously differentiable.

Theorem 5.37 can be used to prove that many kinds of functions are differentiable.

Exercise 5.54. Use Theorem 5.37 to prove that the function

$$\mathbb{R}^3 \ni (x, y, z) \mapsto \cos(e^{xy} z) \in \mathbb{R} \tag{5.55}$$

is differentiable.

Theorem 5.56. *Let $A \subseteq \mathbb{R}^n$ be an open subset and let $f : A \to \mathbb{R}$. Suppose that f attains its maximum at $a \in A$ and $(\partial_i f)(a)$ exists for all $i \in \{1, \ldots, n\}$. Then $(\partial_i f)(a) = 0$.*

Proof. This follows from the analogous theorem in one dimension by restricting to each variable. ■

Theorem 5.57 (Chain Rule v.2)**.** *Let $a \in \mathbb{R}^n$, let $g : \mathbb{R}^n \to \mathbb{R}^m$, and let $f : \mathbb{R}^m \to \mathbb{R}$. Suppose that $g_j := \pi_j \circ g$ is continuously differentiable at a and f is differentiable at $g(a)$. Then*

$$\big(\partial_i(f \circ g)\big)(a) = \sum_{j=1}^{m} (\partial_j f)\big(g(a)\big)(\partial_i g_j)(a). \tag{5.58}$$

Proof. See [13]. ■

Exercise 5.59. Let $A \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, let $c \in A$, and let $f : A \to \mathbb{R}$ be a function for which the partial derivatives all exist and are bounded in an open neighborhood of c. Show that f is continuous at c.

Level I problems.
From Spivak [13]: 2.17 parts (a) and (b), 2.17 parts (c) and (d), 2.17 part (e), 2.17 parts (f) and (g), 2.17 parts (h) and (i), 2.20 part (a), 2.20 part (b), 2.20 parts (c) and (d), 2.20 part (e), 2.22, 2.28 part (a), 2.28 part (b), 2.28 part (c), 2.28 part (d)
From these notes: Exercises 5.17, 5.35, 5.49, 5.52, 5.54

Level II problems.
From Spivak [13]: 2.19, 2.23, 2.25, 2.27, 2.29 (warning: my notation is *very* different from Spivak), 2.30 (refer to 2.29 for the notation, but if possible, please be consistent with my notation in the notes), 2.31, 2.32 part (b), 2.33, 2.34, 2.35
From these notes: Exercise 5.59

Level III problems.
From Spivak [13]: 2.26

6 February 7: Vector Fields

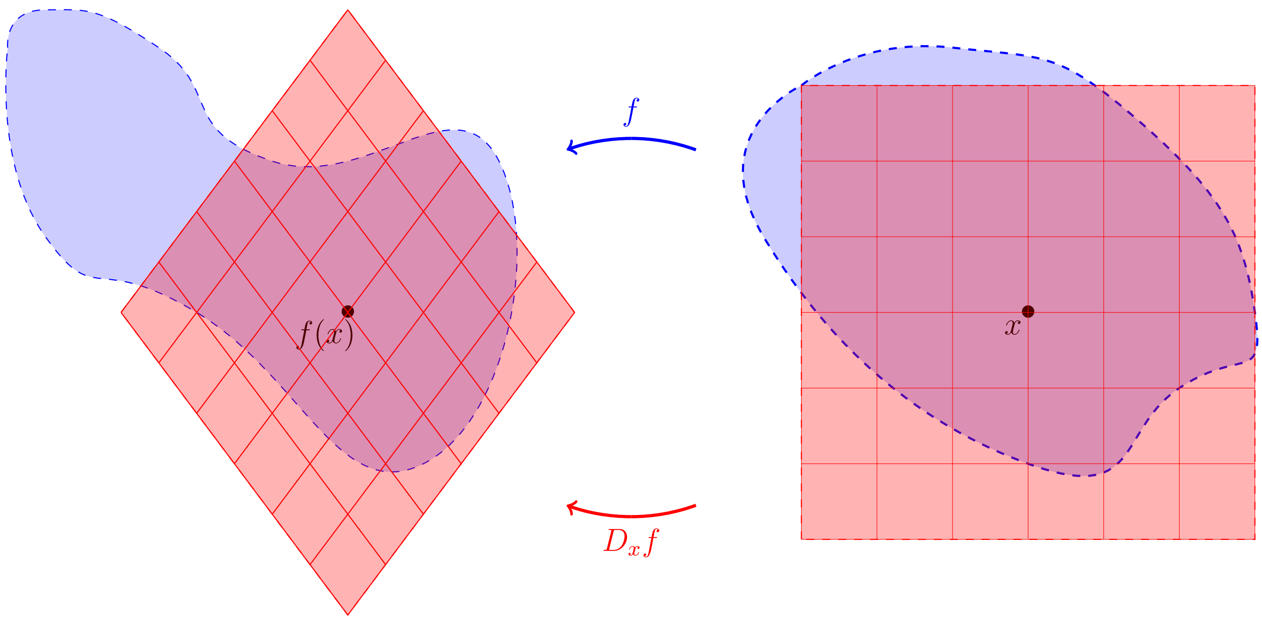
Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ that is differentiable on an open set $U \subseteq \mathbb{R}^n$, what happens to the differential $D_x f$ as $x \in U$ varies? What sort of mathematical object is the assignment $U \ni x \mapsto D_x f$? This assignment defines a function

$$U \xrightarrow{D_\square f} \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n), \tag{6.1}$$

where

$$\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) := \left\{ \mathbb{R}^m \xleftarrow{T} \mathbb{R}^n \ : \ T \text{ is linear} \right\} \tag{6.2}$$

is the set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. The "Hom" stands for "(linear) homomorphism." The box $\square$ is meant to denote that the input $x \in U$ gets placed there. From this perspective, one can interpret $D_x f$ as the first order (linear) approximation to the differentiable function f at x.[7]



Since a linear transformation, such as $D_x f : \mathbb{R}^n \to \mathbb{R}^m$ is determined by what it does on the standard basis, the amount of skewness provides some local information as to the behavior of the function f near x.

From the case in one dimension, this reduces to

$$U \xrightarrow{D_\square f} \mathrm{Hom}(\mathbb{R}, \mathbb{R}) \tag{6.3}$$

and since the function $D_x f$ is linear for every $x \in U$, this means it is completely determined by its value at 1 (this is the slope $f'(x)$). Recall, that for any two sets X and Y, the set of functions from Y to X, denoted by X^Y, comes equipped with a family of functions $\mathrm{ev}_y : X^Y \to X$, for every element $y \in Y$, given by

$$X^Y \ni g \mapsto \mathrm{ev}_y(g) := g(y) \tag{6.4}$$

[7]Higher order derivatives might be discussed later in the course, but the idea behind this might be familiar. The first derivative is just the first order term in the Taylor expansion of a function.

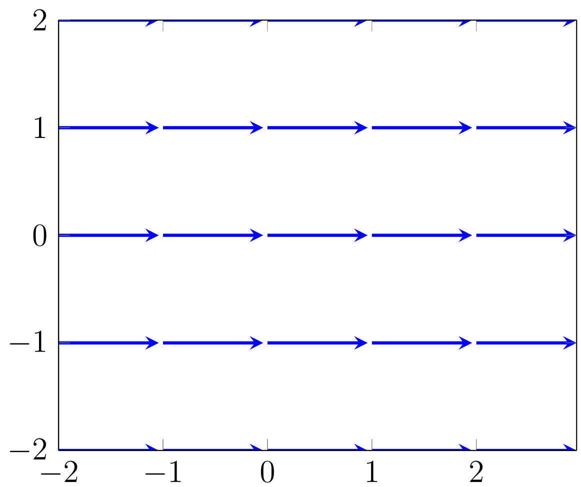
and known as *evaluations*. Since $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) \subseteq (\mathbb{R}^m)^{\mathbb{R}^n}$, it, too, comes equipped with such evaluation functions for any vector in $\mathbb{R}^n$. Consider, then, the following composition of functions

$$\mathbb{R} \xleftarrow{\pi_j} \mathbb{R}^m \xleftarrow{\mathrm{ev}_{e_i}} \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) \xleftarrow{D_\square f} U \tag{6.5}$$

i.e.

$$U \ni x \mapsto \langle e_j, (D_x f)(e_i)\rangle = (\partial_i f_j)(x). \tag{6.6}$$

Because of the relationship between partial derivatives in terms of straight curves and differentiation and the differential, the function $U \to \mathbb{R}^m$ on U defined by sending $x \in U$ to $(D_x f)(e_i)$ can therefore be visualized by evaluating the differential of f on the vectors e_i scattered at every point of the neighborhood U.



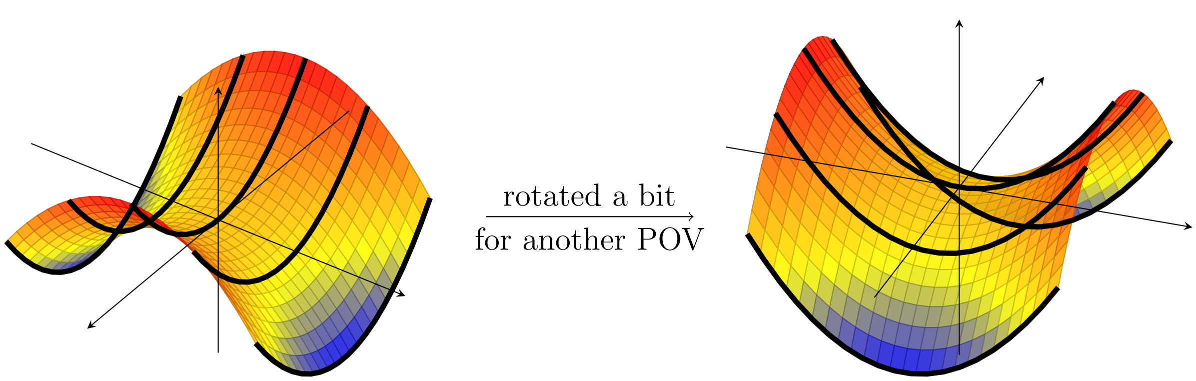
In particular, for a differentiable function of the form $f : U \to \mathbb{R}$ with $U \subseteq \mathbb{R}^n$ open, the composition reduces to

$$U \ni x \mapsto (D_x f)(e_i) = (\partial_i f)(x). \tag{6.7}$$

Example 6.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function from Example 5.7, namely

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 - y^2. \tag{6.9}$$

Consider the set of curves $\{\gamma_{(x,y);e_1}\}_{(x,y)\in\mathbb{R}^2}$ and the associated functions $f \circ \gamma_{(x,y);e_1} : \mathbb{R} \to \mathbb{R}$. For example, in the following plots, the graphs of some of these functions are depicted.



Explicitly,

$$\mathbb{R} \ni t \mapsto (f \circ \gamma_{(x,y);e_1})(t) = f(x+t, y) \tag{6.10}$$

Set $U := \mathbb{R}^2$. Thus, the function $(D_\Box f)(e_1) : U \to \mathbb{R}$ is the function

$$\mathbb{R}^2 \ni (x,y) \mapsto \lim_{t \to 0} \frac{f(x+t,y) - f(x,y)}{t} = \lim_{t \to 0} \frac{2xt + t^2}{t} = 2x. \tag{6.11}$$

This is indeed nothing more than the function

$$U \ni (x,y) \mapsto (\partial_1 f)(x,y), \tag{6.12}$$

the partial derivative of f with respect to the first coordinate as we expected.

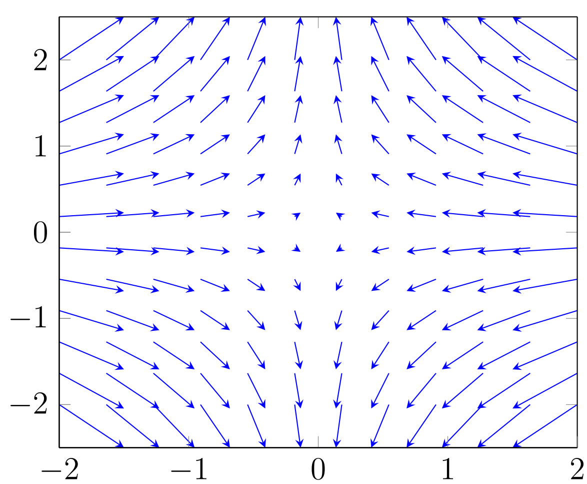
Why on earth would we complicate the partial derivative in such a manner? The point is to view it as a very special case of evaluating the differential $D_\Box f$ on vectors that might *also* depend on the input! For example, given the set of varying vectors in Example 3.8, we could evaluate the differential on this set of "moving vectors" and obtain a completely different function on U.

Definition 6.13. A *vector field* on $\mathbb{R}^n$ is a function $V : \mathbb{R}^n \to \mathbb{R}^n$. A vector field is *continuous* (*differentiable*) iff it is continuous (differentiable) as a function.

Example 6.14. Let $W : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^2 \ni (x,y) \mapsto W(x,y) := (-x, y). \tag{6.15}$$

Then W is a vector field, which can be depicted as in the following plot.

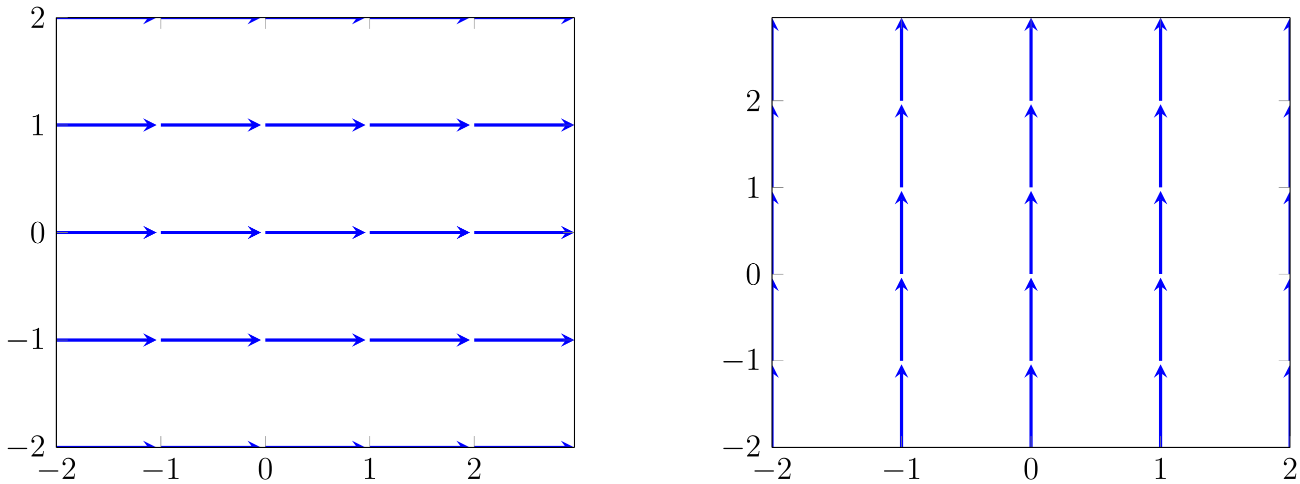


The arrows are scaled to 1/4 their actual lengths for easier view.

Example 6.16. For each $i \in \{1, \ldots, n\}$, let $E_i : \mathbb{R}^n \to \mathbb{R}^n$ be the constant function given by

$$\mathbb{R}^n \ni x \mapsto E_i(x) := e_i. \tag{6.17}$$

Then E_i is a vector field for all $i \in \{1, \ldots, n\}$. These vector fields are the *standard Euclidean vector fields* in $\mathbb{R}^n$. For example, E_1 and E_2 in $\mathbb{R}^2$ are plotted as



respectively.

Definition 6.18. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Set $\mathcal{V}f : \mathbb{R}^n \to \mathbb{R}$ to be the function defined by[8]

$$\mathbb{R}^n \ni x \mapsto (\mathcal{V}f)(x) := (D_x f)\big(V(x)\big). \tag{6.19}$$

This definition also makes sense for functions on open subsets. Diagrammatically, given such a function $f : U \to \mathbb{R}^m$ on an open subset $U \subseteq \mathbb{R}^n$, the new function $\mathcal{V}f : U \to \mathbb{R}^m$ is the composition

$$\mathbb{R}^m \xleftarrow{\mathrm{Ev}} \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) \times \mathbb{R}^n \xleftarrow{D_\square f \times V} U \times U \xleftarrow{\Delta} U, \tag{6.20}$$

where $\Delta : U \to U \times U$ is the diagonal map defined by sending $x \in U$ to $\Delta(x) := (x, x)$ and where $\mathrm{Ev} : \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}^m$ is the evaluation function defined by sending a linear transformation $T \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ and a vector $u \in \mathbb{R}^n$ to $\mathrm{Ev}(T, u) := T(u)$. Therefore, vector fields $V : U \to \mathbb{R}^n$ can also be viewed as operations $\mathcal{V}$ that take in differentiable functions $f : U \to \mathbb{R}^m$ and spit out functions $\mathcal{V}f : U \to \mathbb{R}^m$.

$$\boxed{\text{functions}} \longleftarrow \boxed{\begin{array}{c}\text{vector}\\ \text{fields}\end{array}} \longleftarrow \boxed{\begin{array}{c}\text{differentiable}\\ \text{functions}\end{array}}$$

Example 6.21. As a special case, consider the vector fields E_i on any open subset $U \subseteq \mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a real-valued differentiable function on U. Then

$$\mathcal{E}_i f = \partial_i f, \tag{6.22}$$

the partial derivative of f in the i-th variable.

[8] We will use the notation $\mathcal{V}$ to represent the *operator* associated to the vector field V instead of calling it V again (most texts use the same notation). The reason we should use new notation is to avoid the ambiguous situation of a vector field defined on $\mathbb{R}$ and a function $f : U \to \mathbb{R}$ defined on an open subset $U \subseteq \mathbb{R}$. In this case, Vf could have two possible meanings. On the one hand, it could mean the function $U \ni x \mapsto (D_x f)\big(V(x)\big)$ but on the other hand, it could simply mean the function $U \ni x \mapsto V(x)f(x)$. Of course, we could simply clarify depending on the context, but let us just be safe and use slightly different notation all-together to avoid the ambiguity.

Example 6.23. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) := (y, -x). \tag{6.24}$$

and let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 - y^2. \tag{6.25}$$

Then

$$(\mathcal{V}f)(x, y) = (D_{(x,y)}f)\big(V(x, y)\big) = 4xy. \tag{6.26}$$

Exercise 6.27. Let $W : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto W(x, y) := (-x, y) \tag{6.28}$$

and let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 - y^2. \tag{6.29}$$

Calculate $(\mathcal{W}f)(x, y) \equiv (D_{(x,y)}f)\big(W(x, y)\big)$ for all $(x, y) \in \mathbb{R}^2$.

Vector fields have numerous properties reminiscent of the differential.

Theorem 6.30. *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions, let $\lambda \in \mathbb{R}$, and let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. Then*

(a) $\mathcal{V}(f + g) = \mathcal{V}(f) + \mathcal{V}(g)$,

(b) $\mathcal{V}(\lambda f) = \lambda\mathcal{V}(f)$, *and*

(c) $\mathcal{V}(fg) = \mathcal{V}(f)g + f\mathcal{V}(g)$.

Proof. (a) and (b) follow from linearity of $D_x f$ and the definition of $\mathcal{V}f$. (c) follows from the product rule. ■

This theorem shows that vector fields can be thought of as linear transformations from the vector space of differentiable functions to the vector space of all functions that also satisfy a Leibniz property (the product rule). We will come back to this crucial observation when discussing arbitrary vector fields on manifolds. Also note that all of the above definitions make sense for functions $f : \mathbb{R}^n \to \mathbb{R}^m$ except where one multiplies functions in which case m must be 1.

Exercise 6.31. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field.

(a) Show that there exist a unique set of functions $\{\psi_i : \mathbb{R}^n \to \mathbb{R}\}_{i\in\{1,\ldots,n\}}$ such that

$$V = \sum_{i=1}^{n} \psi_i E_i. \tag{6.32}$$

In fact, show that these functions are given by

$$\mathbb{R}^n \ni x \mapsto \psi_i(x) = \big\langle e_i, V(x)\big\rangle \tag{6.33}$$

for all $i \in \{1, \ldots, n\}$.

(b) Use this to conclude that

$$\mathcal{V}f = \sum_{i=1}^{n} \psi_i(\partial_i f). \tag{6.34}$$

(c) Use this to show that V is differentiable if and only if ψ_i is differentiable for all $i \in \{1, \dots, n\}$.

Given an open set $U \subseteq \mathbb{R}^n$, a vector field $V : U \to \mathbb{R}^n$, and a *differentiable* function $f : U \to \mathbb{R}^m$, we get a function $\mathcal{V}f : U \to \mathbb{R}^m$. *Suppose* this function is *also* differentiable. Then, for another vector field $W : U \to \mathbb{R}^n$, we can obtain the function $\mathcal{W}(\mathcal{V}f) : U \to \mathbb{R}^m$. Suppose that $\mathcal{W}f : U \to \mathbb{R}^m$ is also differentiable. Then we can compare the two functions $\mathcal{W}(\mathcal{V}f)$ and $\mathcal{V}(\mathcal{W}f)$ by looking at their difference

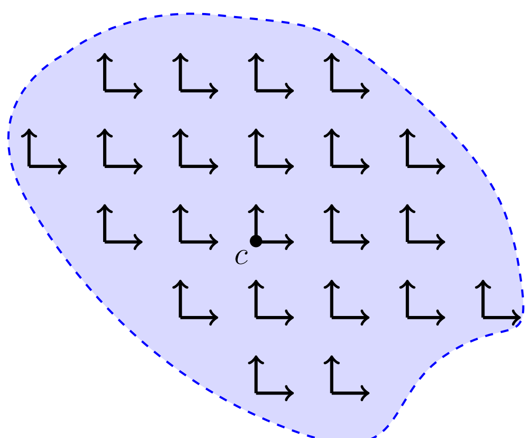
$$\mathcal{V}(\mathcal{W}f) - \mathcal{W}(\mathcal{V}f). \tag{6.35}$$

When restricted to the set of all functions $f : U \to \mathbb{R}^n$ for which $\mathcal{V}f$ and $\mathcal{W}f$ are differentiable, we can define the vector field operator

$$[\mathcal{V}, \mathcal{W}] := \mathcal{V} \circ \mathcal{W} - \mathcal{W} \circ \mathcal{V}, \tag{6.36}$$

but we must be careful about its domain of definition due to the fact that our functions must be twice differentiable in a sense that will be made more precise later in this lecture. Ignoring this technicality for the moment, in general, this expression is non-zero. In fact, it need not be zero even in the special case where $\mathcal{V} = \mathcal{E}_i$ and $\mathcal{W} = \mathcal{E}_j$. Indeed, there exist functions $f : U \to \mathbb{R}^m$ for which

$$\mathcal{E}_i(\mathcal{E}_j f) - \mathcal{E}_j(\mathcal{E}_i f) \equiv \partial_i(\partial_j f) - \partial_j(\partial_i f) \neq 0. \tag{6.37}$$



Fortunately, with an additional assumption on these functions, the result is zero and the partial derivatives are said to commute.

Theorem 6.38. *Let $A \subseteq \mathbb{R}^n$ be open, fix $c \in A$, let $f : A \to \mathbb{R}$, and fix $i, j \in \{1, \dots, n\}$. Suppose there exists an open set $U \subseteq \mathbb{R}^n$ with $c \in U \subseteq A$ such that $\partial_i f$, $\partial_j f$ $\partial_i(\partial_j f)$, and $\partial_j(\partial_i f)$ are all well-defined on U and suppose that both $\partial_i(\partial_j f)$ and $\partial_j(\partial_i f)$ are continuous at c. Then*

$$\big(\partial_i(\partial_j f)\big)(c) = \big(\partial_j(\partial_i f)\big)(c). \tag{6.39}$$

The following proof comes directly out of [9].

Proof. It suffices to prove the theorem for a function f of two variables. Hence, write c as $c = (a, b)$. Therefore, the goal is to prove

$$(\partial_2\partial_1 f)(c) = (\partial_1\partial_2 f)(c). \tag{6.40}$$

By the definition of the ordinary derivative of a single variable (the second variable in this case),

$$\big(\partial_2(\partial_1 f)\big)(c) = \lim_{k\to 0} \frac{(\partial_1 f)(a, b+k) - (\partial_1 f)(a, b)}{k}. \tag{6.41}$$

Applying this definition one more time (to the first variable now) to $\partial_1 f$ gives

$$\big(\partial_2(\partial_1 f)\big)(c) = \lim_{k\to 0} \frac{1}{k}\left(\lim_{h\to 0} \frac{f(a+h, b+k) - f(a, b+k)}{h} - \lim_{h\to 0} \frac{f(a+h, b) - f(a, b)}{h}\right). \tag{6.42}$$

By the Algebraic Continuity Theorem, this becomes (please note the order of limits must be preserved)

$$\big(\partial_2(\partial_1 f)\big)(c) = \lim_{k\to 0}\lim_{h\to 0} \frac{f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b)}{hk}. \tag{6.43}$$

The expression for $(\partial_1\partial_2 f)(c)$ is the same but with the limits performed in the opposite order. This calculation motivates the following definitions and notation. Because U is open, there is a closed rectangle R of the form $[a, a+h] \times [b, b+k]$ for some $h, k > 0$ contained in U. Let $\lambda(h, k)$ be the expression

$$\lambda(h, k) := f(a, b) - f(a+h, b) - f(a, b+k) + f(a+h, b+k). \tag{6.44}$$

Our goal, therefore, is to prove

$$\lim_{k\to 0}\lim_{h\to 0} \frac{\lambda(h, k)}{hk} = \lim_{h\to 0}\lim_{k\to 0} \frac{\lambda(h, k)}{hk}. \tag{6.45}$$

To do this, we will need a lemma.
Lemma. There exist points $p, q \in R$ such that

$$\lambda(h, k) = \big((\partial_2\partial_1 f)(p)\big)hk \qquad \& \qquad \lambda(h, k) = \big((\partial_1\partial_2 f)(q)\big)hk. \tag{6.46}$$

Proof of Lemma. Let $\phi : [a, a+h] \to \mathbb{R}$ be the function defined by

$$[a, a+h] \ni r \mapsto \phi(r) := f(r, b+k) - f(r, b). \tag{6.47}$$

Note that

$$\phi(a+h) - \phi(a) = f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b) = \lambda(h, k). \tag{6.48}$$

Because $\partial_1 f$ exists in U, ϕ is differentiable in an open interval containing $[a, a+h]$. By the Mean Value Theorem (MVT), there exists an $r_0 \in [a, a+h]$ such that

$$\begin{tikzcd}
\phi(a+h)-\phi(a) \arrow[r, equal, bend left, "\text{by MVT}"] & \big(\phi'(r_0)\big)h \arrow[d, equal, bend left, "\text{by (6.47)}"] \\
\lambda(h,k) \arrow[u, equal, bend left, "\text{by (6.48)}"] & \big((\partial_1 f)(r_0,b+k)-(\partial_1 f)(r_0,b)\big)h
\end{tikzcd} \tag{6.49}$$

The MVT can be used once again applied to the function

$$[b, b+k] \ni s \mapsto (\partial_1 f)(r_0, s) \tag{6.50}$$

so that there exists an $s_0 \in [b, b+k]$ such that

$$(\partial_1 f)(r_0, b+k) - (\partial_1 f)(r_0, b) = (\partial_2 \partial_1 f)(r_0, s_0)k. \tag{6.51}$$

Setting $p := (r_0, s_0)$ gives

$$\lambda(h,k) = (\partial_2 \partial_1 f)(p)hk. \tag{6.52}$$

A similar proof works considering the function $\psi : [b, b+k] \to \mathbb{R}$ defined by

$$[b, b+k] \ni t \mapsto \psi(t) := f(a+h, t) - f(a, t). \tag{6.53}$$

from which q can be constructed satisfying the right-hand-side of (6.46). ■

(Back to proof of Theorem.) Since this lemma is true for arbitrary sufficiently small h and k, for every sufficiently small t (such as for all t satisfying $0 \le t \le \min\{h, k\}$), the assignment $t \mapsto \lambda(t,t)$ defined on the rectangle

$$R_t := [a, a+t] \times [b, b+t] \tag{6.54}$$

is well-defined. By the previous Lemma, for each such t, there exist $p_t, q_t \in R_t$ such that

$$\lambda(t,t) = \big((\partial_2 \partial_1 f)(p_t)\big)t^2 \qquad \& \qquad \lambda(t,t) = \big((\partial_1 \partial_2 f)(q_t)\big)t^2. \tag{6.55}$$

By continuity of both $\partial_1 \partial_2 f$ and $\partial_2 \partial_1 f$ at c and because $\lim\limits_{t \to 0} p_t = c = \lim\limits_{t \to 0} q_t$,

$$\lim_{t \to 0} \frac{\lambda(t,t)}{t^2} = (\partial_2 \partial_1 f)(c) \qquad \& \qquad \lim_{t \to 0} \frac{\lambda(t,t)}{t^2} = (\partial_1 \partial_2 f)(c) \tag{6.56}$$

respectively. Since the left-hand-sides are the same for each of these equalities, this proves that

$$(\partial_2 \partial_1 f)(c) = (\partial_1 \partial_2 f)(c). \tag{6.57}$$

■

Exercise 6.58. Indicate how the proof of Theorem changes if the function f in the statement is assumed to have codomain $\mathbb{R}^m$ instead of $\mathbb{R}$.

Definition 6.59. Fix $k \in \mathbb{N}$. Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$. f is *k times continuously differentiable at c* iff there exists an open set $U \subseteq \mathbb{R}^n$ with $c \in U$ such that

(a) $(\partial_{i_1} \cdots \partial_{i_l} f_j)(x)$ exists for all $x \in U$ and

(b) $\partial_{i_1} \cdots \partial_{i_l} f_j : U \to \mathbb{R}$ is continuous at c

for all $i_1, \dots, i_l \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ for all $l \in \{1, \dots, k\}$. f is *k times continuously differentiable on* $\mathbb{R}^n$ iff f is k times continuously differentiable at c for all $c \in \mathbb{R}^n$. The set of all k times continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ is denoted by $C^k(\mathbb{R}^m, \mathbb{R}^n)$ and such a function is said to be of *class* C^k on $\mathbb{R}^n$. Sometimes, the shorthand notation $C^k(\mathbb{R}^n)$ is used if it is clear from context that one is considering real-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ and not functions with more than one component.

Exercise 6.60. Let $g : \mathbb{R} \to \mathbb{R}$ be a function of class C^2 on $\mathbb{R}$. Show that

$$g''(c) = \lim_{h \to 0} \frac{g(c+h) - 2g(c) + g(c-h)}{h^2}. \tag{6.61}$$

[Hint: Use ideas from the proof of Theorem 6.38.]

Definition 6.62. Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$. f is *infinitely continuously differentiable at c* iff f is k-times continuously differentiable at c for all $k \in \mathbb{N}$. $f : \mathbb{R}^n \to \mathbb{R}^m$ is *infinitely continuously differentiable on* $\mathbb{R}^n$ iff f is infinitely continuously differentiable at c for all $c \in \mathbb{R}^n$. The set of all infinitely times continuously differentiable functions on $\mathbb{R}^n$ is denoted by $C^\infty(\mathbb{R}^m, \mathbb{R}^n)$ and such a function is said to be of *class* C^∞ on $\mathbb{R}^n$.

Analogous definitions can be made for functions $f : A \to \mathbb{R}^m$ defined on an open subset $A \subseteq \mathbb{R}^n$.

Exercise 6.63. Prove or disprove the following:[9] "Let $c \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$. f is infinitely continuously differentiable at c if and only if there exists an open set $A \subseteq \mathbb{R}^n$ such that $c \in A$ and $f \in C^\infty(\mathbb{R}^m, A)$." Prove any direction that is true and provide a counter-example to any direction that is false. [Warning: The subtlety that I am noticing is that the definition we gave states there exists a $U_k \subseteq \mathbb{R}^n$ for each $k \in \mathbb{N}$, but what happens if these open sets get smaller and smaller as k increases?]

Theorem 6.38 implies the following result.

Corollary 6.64. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be a function of class C^2 on U. Then*

$$\partial_i \partial_j f = \partial_j \partial_i f \tag{6.65}$$

for all $i, j \in \{1, \dots, n\}$.

There is a subtle but important distinction between continuously differentiable functions and just differentiable functions. The same is true for higher derivatives, but we need to define what it means for a function to be twice differentiable.

[9] I don't know the answer!

Definition 6.66. Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^m$ be a function. f is *twice differentiable* at c iff there exists an open set $U \subseteq \mathbb{R}^n$ such that

(a) $c \in U \subseteq A$,

(b) f is differentiable on U, and

(c) $\partial_i f_j$ is differentiable at c for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$.

The above definition uses partial derivatives in a crucial way. However, as we will be studying manifolds in the coming lectures, it will be important to provide a definition that is independent of coordinates (for arbitrary domains, it might not be possible to define $\partial_i f$).

Theorem 6.67. *Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^m$ be a function. Then f is twice differentiable at c iff there exists an open set $U \subseteq \mathbb{R}^n$ such that*

(a) $c \in U \subseteq A$,

(b) f is differentiable on U, and

(c) for every differentiable vector field $V : U \to \mathbb{R}^n$, the functions $\mathcal{V} f_j : U \to \mathbb{R}$ are differentiable at c for all $j \in \{1, \dots, m\}$.

Proof.
($\Leftarrow$) Fix $i \in \{1, \dots, n\}$. Set $V := E_i$. Since E_i is differentiable (it is constant), $\mathcal{E}_i f_j = \partial_i f_j$ is differentiable at c. Since i and j were arbitrary, this proves that f is twice differentiable.
($\Rightarrow$) Let $V : U \to \mathbb{R}^n$ be a differentiable vector field. By Exercise 6.31, there exist a unique set of functions $\{\psi_i : \mathbb{R}^n \to \mathbb{R}\}_{i \in \{1,\dots,n\}}$ such that

$$V = \sum_{i=1}^{n} \psi_i E_i. \tag{6.68}$$

Therefore,

$$\mathcal{V} f_j = \sum_{i=1}^{n} \psi_i \mathcal{E}_i f_j = \sum_{i=1}^{n} \psi_i \partial_i f_j, \tag{6.69}$$

which is a finite sum of a product of differentiable functions. Hence, $\mathcal{V} f_j$ is differentiable for all $j \in \{1, \dots, m\}$. ■

Notice that f is required to be differentiable not just at c but on some open neighborhood of c in A (the definition would not make sense if $D_{\square} f$ was not defined in a neighborhood of c since to take the next derivative, we need to know the values of $D_{\square} f$ in a neighborhood of c). This is so that we can iterate and use our earlier definition. In particular, this definition can be used to define what it means for a vector field to be twice differentiable. Therefore, this can be used to inductively define what it means for a function to be k times differentiable for any $k \in \mathbb{N}$.

Definition 6.70. Fix $k \in \mathbb{N}$. Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^m$ be a function. f is *k times differentiable* at c iff for every $(k-1)$ times differentiable vector field $V : U \to \mathbb{R}^n$, the functions $\mathcal{V} f_j : U \to \mathbb{R}$ are $(k-1)$ times differentiable at c for all $j \in \{1, \dots, m\}$.

Exercise 6.71. Make a statement analogous to Theorem 6.67 so as to characterize what it means for a function $f : A \to \mathbb{R}^m$ to be k times differentiable without using vector fields and only using partial derivatives as in Definition 6.66. Prove your claim.

Finally, we can then use this to define infinitely differentiable functions.

Definition 6.72. Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^m$ be a function. f is *infinitely times differentiable at c* iff f is k times differentiable at c for all $k \in \mathbb{N}$.

Exercise 6.73. Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}^m$ be a function. Prove or disprove: "f is infinitely times differentiable on A if and only if f is of class C^∞ on A."

Exercise 6.74. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be the function defined by

$$\mathbb{R}^3 \ni (\rho, \phi, \theta) \mapsto f(\rho, \phi, \theta) := \big(\rho \cos\phi \sin\theta, \rho \sin\phi \sin\theta, \rho \cos\theta\big). \tag{6.75}$$

(a) Calculate the matrix associated to the linear transformation $D_{(\rho,\phi,\theta)} f : \mathbb{R}^3 \to \mathbb{R}^3$ with respect to the standard Euclidean basis.

(b) Calculate the determinant $\det\big(D_{(\rho,\phi,\theta)} f\big)$.

(c) Let

$$S := [1, 2] \times \left[0, \frac{\pi}{2}\right] \times \left[0, \frac{\pi}{2}\right]. \tag{6.76}$$

Sketch the image of S under f. On the same drawing, sketch the images of the unit vectors $e_1, e_2, e_3 \in \mathbb{R}^3$ at any point of your choice in S.

Level I problems.
From Spivak [13]:
From these notes: Exercises 6.27, 6.58, 6.71

Level II problems.
From Spivak [13]: 2.24
From these notes: Exercises 6.31, 6.60, 6.63, 6.73, 6.74

7 February 14: Differential Equations and Dynamical Systems

This lecture is now the only content for Week #04.

As we saw from the last few lectures, there are many technical and subtle differences between different definitions of differentiability. This can get annoying when trying to formulate theorems since their validity may depend on the precise nature of the assumptions. As a result, we will often assume that our functions are of class C^k for some $k \in \mathbb{N}$. Let us now get back to vector fields and their applications. But before we do that, we will refresh our memory by recalling some points from ordinary differential equations of a single variable. A more through explanation of this material will be differed to when we learn about integration. In many exercises, you will be asked to plot vector fields and plot some integral curves. You may use whatever program you deem appropriate, but I highly recommend using Mathematica's VectorPlot and StreamPlot features.

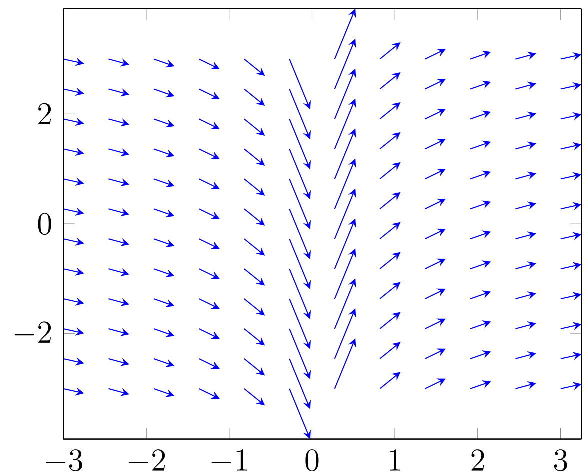
Example 7.1. The present example is motivated by the discussion in [12]. Without considerations to domain, suppose you had a function x of a single variable t that satisfies

$$x'(t) = \frac{1}{t} \tag{7.2}$$

for all t for which the right-hand-side makes sense. What should the function x be? Upon some thought, perhaps $x(t) = \log(t)$ would work. Or more generally, you might guess $x(t) = b + \log(t)$ for some constant b. However, the derivative function $x'(t)$ makes sense for all values of t away from 0. Meanwhile, the domain of log consists only of positive real numbers. Upon more contemplation, another solution that works is $x(t) = a + \log(-t)$ for some constant a but this time the domain is different. Only after putting these two solutions together do we obtain a general solution to the differential equation (7.2)

$$x(t) = \begin{cases} a + \log(-t) & \text{for } t \in (-\infty, 0) \\ b + \log(t) & \text{for } t \in (0, \infty) \end{cases} \tag{7.3}$$

for some constants $a, b \in \mathbb{R}$. Notice the important point of not only the domain issue mentioned above, but also the existence of *two* integration constants! In calculus, you were probably shown that the anti-derivative of a function usually involves *one* integration constant. This is true for integrable functions defined on a *connected* domain (we will discuss integration later and for now I am relying on your experience with having taken calculus). In general, there is an integration constant for each connected subset of the domain. One way to visualize these functions without even attempting to solve the differential equation is to plot the graphs of the slopes of the function x using the differential equation. Namely, for every point $t \in \mathbb{R} \setminus \{0\}$, draw the vector $(1, x'(t))$ at the points (t, s) for every $s \in \mathbb{R}$. This defines a vector field on $\mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big)$. Visually, we can see that the solutions of (7.2) "hug" these vectors.



In the figure above, the vector field

$$\mathbb{R}^2 \setminus \big(\{0\} \times \mathbb{R}\big) \ni (t, x) \mapsto V(t, x) := \left(1, \frac{1}{t}\right) \tag{7.4}$$

has been drawn with is vectors scaled by a factor of $\frac{1}{4}$ for easier viewing purposes.

Exercise 7.5. Consider the differential equation

$$t(2x^3 - t^3)x' + x(2t^3 - x^3) = 0, \tag{7.6}$$

where x is assumed to be a function of t so that x' means the derivative of x with respect to t. Calculate the associated vector field (being careful to explain the possibility of dividing by 0), plot it on the domain $[-1, 1]$, and describe its integral curves without finding their explicit expressions.

It is not necessarily true that if a solution exists to a differential equation then its solution through a point is unique. We will later study assumptions which guarantee the existence and solutions of differential equations later when we learn about metric spaces.

Example 7.7. Consider the differential equation

$$\big(x'(t)\big)^2 = t \tag{7.8}$$

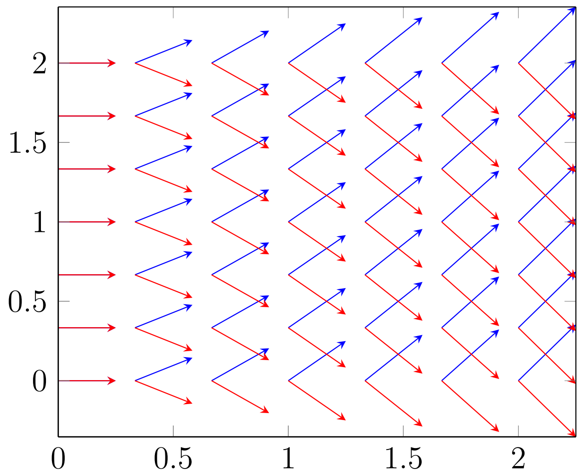
which is defined for all $t \geq 0$. For every $t \geq 0$, there are two possible vector fields associated with this differential equation. These vector fields, denoted by $V_+, V_- : [0, \infty) \times \mathbb{R} \to \mathbb{R}^2$, are given by

$$[0, \infty) \times \mathbb{R} \ni (t, x) \mapsto V_+(t, x) := \big(1, \sqrt{t}\big) \tag{7.9}$$

and

$$[0, \infty) \times \mathbb{R} \ni (t, x) \mapsto V_-(t, x) := \big(1, -\sqrt{t}\big), \tag{7.10}$$

respectively.



The vectors were scaled by a factor of $\frac{1}{4}$ for easier viewing purposes.

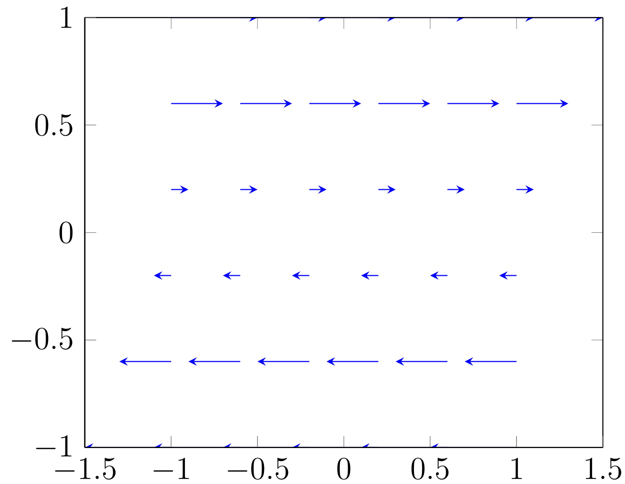
Example 7.11. The present example is also from [12]. Consider now the *second* order differential equation

$$x''(t) = 0, \tag{7.12}$$

where x is a function of $t \in \mathbb{R}$. Rather than working through the solution of this differential equation (which you should verify is of the form $x(t) = at + b$ with $a, b \in \mathbb{R}$), write this as two first order differential equations in the form

$$\begin{aligned} x'(t) &= y(t) \\ y'(t) &= 0 \end{aligned} \tag{7.13}$$

by introducing the function y whose domain is $\mathbb{R}$. By substituting back, you can check that the original ODE is obtained. In this way, a solution to the original ODE (7.12) can be expressed as a solution to (7.13) which describes a parametric curve $t \mapsto \big(x(t), y(t)\big)$. There is a vector field associated to the ODE (7.13)



whose solutions are curves "hugging" along the vector field lines (the vectors have been scaled by a factor of $\frac{1}{2}$). Indeed, the solutions are $y = a$ for some constant $a \in \mathbb{R}$. The behavior of these solutions depends on whether $a > 0, a = 0$, or $a < 0$. Thus x as a function of time either increases, remains in place, or decreases in these respective possibilities.

As the previous example illustrates, for several functions of a single real variable, it can be useful to use vector fields to provide a visualization of differential equations. Besides the reason of providing visualizations of second order differential equations, another reason is that the derivative of a function $f : \mathbb{R} \to \mathbb{R}^n$ can no longer be interpreted as just a slope. Instead, we visualize solutions of differential equations as integral curves.

Definition 7.14. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field and let $c \in \mathbb{R}^n$. An <u>*integral curve*</u> through c consists of an open interval $U \subseteq \mathbb{R}$ and a differentiable curve $\gamma : U \to \mathbb{R}^n$ such that

(a) $0 \in U$,

(b) $\gamma(0) = c$, and

(c) $(D_t\gamma)(1) = V\big(\gamma(t)\big)$ for all $t \in \mathbb{R}$.

Example 7.15. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) := (y, -x) \tag{7.16}$$

and consider the case $c = (a, b) \in \mathbb{R}^2$ as an arbitrary initial condition with not both a and b equal to zero. Suppose that an integral curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ through c exists. Let $\gamma_1, \gamma_2 : \mathbb{R} \to \mathbb{R}$ be the components of such an integral curve. Then such curves must satisfy the conditions

$$\begin{aligned} \gamma_1'(t) &= \gamma_2(t) \\ \gamma_2'(t) &= -\gamma_1(t) \end{aligned} \tag{7.17}$$

and the initial conditions

$$\gamma_1(0) = a \qquad \& \qquad \gamma_2(0) = b. \tag{7.18}$$

This is an example of an initial value problem (an ordinary differential equation). We are postulating that solutions exist and we will confirm whether or not the set of solutions is non-empty by attempting to find an explicit formula for a solution. We assume that γ is twice-differentiable, in which case the initial value problem becomes

$$\begin{aligned} \gamma_1''(t) &= \gamma_2'(t) = -\gamma_1(t) \\ \gamma_2''(t) &= -\gamma_1'(t) = -\gamma_2''(t) \end{aligned} \tag{7.19}$$

whose solutions might be more transparent: the sine and cosine functions satisfy the condition that if you differentiate them twice, they are negatives of themselves. Thus, the solutions should be of the form

$$\begin{aligned} \gamma_1(t) &= A\cos t + B\sin t \\ \gamma_2(t) &= C\cos t + D\sin t \end{aligned} \tag{7.20}$$

for some as of yet undetermined coefficients $A, B, C, D \in \mathbb{R}$. To determine these constants, we use the initial conditions and the first order differential equation itself. The initial conditions provide for us

$$a = \gamma_1(0) = A \qquad \& \qquad b = \gamma_2(0) = C. \tag{7.21}$$

Differentiating (7.20) gives

$$\begin{aligned}\gamma_1'(t) &= -a\sin t + B\cos t\\ \gamma_2'(t) &= -b\sin t + D\cos t.\end{aligned} \tag{7.22}$$

Using this and (7.17) gives the constraints

$$-a\sin t + B\cos t = b\cos t + D\sin t \qquad \& \qquad -b\sin t + D\cos t = -a\cos t - B\sin t \tag{7.23}$$

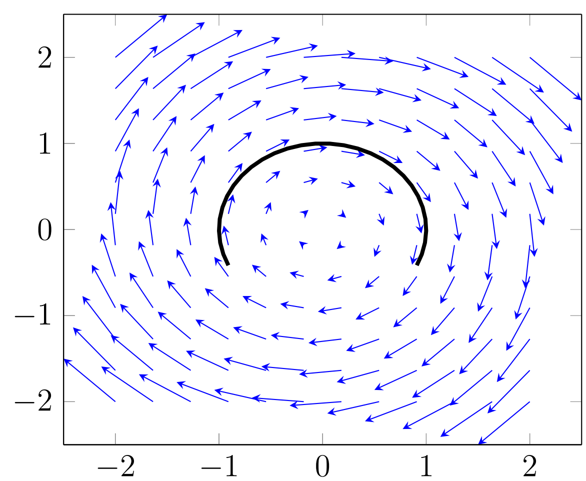
for all t in some domain which we will soon discover. Whatever this domain is, it must contain $t = 0$ so plugging this in gives

$$B = b \qquad \& \qquad D = -a. \tag{7.24}$$

Hence, the supposed solution to the initial value problem is

$$\mathbb{R} \ni t \mapsto \gamma(t) = \big(a\cos t + b\sin t, b\cos t - a\sin t\big) \tag{7.25}$$

which is indeed defined for all $t \in \mathbb{R}$ and solves the initial value problem.



The previous example has integral curves that are periodic.

Definition 7.26. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field. An integral curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ is <u>*periodic*</u> iff there exists a $T \in \mathbb{R}$ such that

$$\gamma(t + T) = \gamma(t) \qquad \forall\, t \in \mathbb{R}. \tag{7.27}$$

When γ is periodic, the smallest such non-negative value of T is called the <u>*period*</u> of the integral curve γ.

Exercise 7.28. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$V(x, y) = (-x, y). \tag{7.29}$$

Plot V on a graph and calculate all of the integral curves of V specifying their precise domains.

Exercise 7.30. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$V(x, y) = (x^2, y). \tag{7.31}$$

Plot V on a graph. Calculate the integral curves of V and see which ones are defined on all of $\mathbb{R}$. In particular, given an arbitrary initial condition, define the domain on which an integral curve exists satisfying the initial conditions.

One of the reasons for studying vector fields is that they describe systems that can be described by differential equations. To see this, consider a simple n-th order differential equation (in a single variable) of the form

$$f^{(n)} - a_{n-1}f^{(n-1)} - \cdots - a_1 f^{(1)} - a_0 f = 0 \tag{7.32}$$

where f is a real-valued function of a single variable and the a_i's are real numbers.[10] Then by setting

$$x_1 := f, \qquad x_2 := f^{(1)}, \quad \ldots, \quad x_n := f^{(n-1)}, \tag{7.33}$$

we get n coupled first order differential equations

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= x_3, \\ &\vdots \\ x_{n-1}' &= x_n \\ x_n' &= a_0 x_1 + a_1 x_2 + \cdots + a_{n-1} x_n \end{aligned} \tag{7.34}$$

This last set of equations is a differential equation obtained from a vector field on $\mathbb{R}^n$. The vector field $\mathbb{R}^n \to \mathbb{R}^n$ in question is defined by

$$\mathbb{R}^n \ni (x_1, \ldots, x_n) \mapsto \big(x_2, x_3, \ldots, x_n, a_0 x_1 + a_1 x_2 + \cdots + a_{n-1} x_n\big). \tag{7.35}$$

Therefore, from an n-th order ordinary differential equation, we obtained a vector field. The integral curves of these vector fields are precisely the set of solutions to the n-th order differential equation once the variables are changed. In general, the system of (coupled) first order differential equations associated to an arbitrary vector field $V : \mathbb{R}^n \to \mathbb{R}^n$ are given by

$$\begin{aligned} x_1' &= V_1(x_1, x_2, \ldots, x_n) \\ x_2' &= V_2(x_1, x_2, \ldots, x_n) \\ &\vdots \\ x_n' &= V_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{7.36}$$

where $V_i := \pi_i \circ V$ is the i-th component of V. Equivalently, if γ is assumed to be an integral curve of V, this can be rewritten more concisely as

$$\gamma'(t) = V\big(\gamma(t)\big). \tag{7.37}$$

[10] In fact, the a_i's do not have to be constant coefficients and can more generally be functions of t. This generalization is explored in Exercise 7.38.

It turns out that it is *not* always possible to obtain a single n-th order differential equation from such a system. Therefore, vector fields describe a far more general set of differential equations than higher order differential equations (a more precise statement will be provided in Exercise 7.38). Since we only described a very simple type of n-th order differential equation, the following exercise should convince you that far more general equations are allowed.

Exercise 7.38. Let $f : \mathbb{R} \to \mathbb{R}$ be n times differentiable and let $G : \mathbb{R}^n \to \mathbb{R}$ be a function. Show that the n-th order differential equation

$$f^{(n)} - G\big(f, f^{(1)}, \dots, f^{(n-1)}\big) = 0 \tag{7.39}$$

can be reduced to a system of n coupled first order differential equations.

Definition 7.40. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field on $\mathbb{R}^n$. The ***critical points*** of the dynamical system associated to V are the points $x \in \mathbb{R}^n$ such that $V(x) = 0$.

The critical points of a vector field often provide a lot of useful information about the integral curves and their behavior in the vicinity of critical points.

Exercise 7.41. Consider the differential equation in $\mathbb{R}^2$ given by

$$\begin{aligned} \dot{x} &= ax + by \\ \dot{y} &= cx + dy \end{aligned} \tag{7.42}$$

where $a, b, c, d \in \mathbb{R}$. The dot is another notation for the derivative, typically used when we interpret the quantity to change in time. Let A be the 2×2 matrix given by

$$A := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{7.43}$$

and

$$\gamma(t) := \big(x(t), y(t)\big) \qquad \forall\, t \in \mathbb{R}. \tag{7.44}$$

Then, the differential equation (7.42) can be expressed as

$$\dot{\gamma}(t) = A\gamma(t) \tag{7.45}$$

for all $t \in \mathbb{R}$.

(a) Show that this first order linear differential equation can be expressed as a second order differential equation of the form

$$\ddot{x} - \operatorname{tr}(A)\dot{x} + \det(A)x = 0 \tag{7.46}$$

(and similarly for y), where $\ddot{x}$ is the second derivative of x, $\operatorname{tr}(A)$ is the trace of A (the sum of its diagonal entries), and $\det(A)$ is the determinant of A. Use this to argue that if P is another matrix, then PAP^{-1} gives exactly the same set of second order differential equations.

(b) Solve the system assuming the initial conditions are $x(0) = x_0$ and $y(0) = y_0$. Note: you will need to split the solutions into the following four cases

i) $\det(A) = 0$ and $\operatorname{tr}(A) = 0$,

ii) $\det(A) = 0$ and $\operatorname{tr}(A) \neq 0$,

iii) $\det(A) \neq 0$ and $\operatorname{tr}(A)^2 \neq 4\det(A)$, and

iv) $\det(A) \neq 0$ and $\operatorname{tr}(A)^2 = 4\det(A)$.

(c) Further subdivide the previous solutions based on the behavior of the integral curves in the vicinity of the critical point. Note: be aware that complex eigenvalues are possible.

Exercise 7.47. Let $V : [0,\infty) \times [0,\infty) \to \mathbb{R}^2$ be the vector field given by

$$[0,\infty) \times [0,\infty) \ni (x,y) \mapsto V(x,y) := \big(0.4x - 0.4xy, -0.1y + 0.2xy\big) \tag{7.48}$$

so that the associated differential equation describing dynamical system is

$$\begin{aligned} \dot{x} &= 0.4x - 0.4xy \\ \dot{y} &= -0.1y + 0.2xy \end{aligned} \tag{7.49}$$

This dynamical system describes a simple model for a predator-prey community. Here x is the prey and y is the predator.

(a) Plot the vector field V on the domain $[0,2] \times [0,2]$.

(b) Calculate the critical points of V.

(c) In physical terms, describe (do not calculate) the integral curves of V in the vicinity of the critical points. What happens as you move slightly away from the critical points but still remain within the domain $(0,\infty) \times (0,\infty)$?

(d) Describe the integral curves for the initial conditions where $x(0) = 0$.

(e) Describe the integral curves for the initial conditions where $y(0) = 0$.

Exercise 7.50. Let $V : [0,\infty) \times [0,\infty) \to \mathbb{R}^2$ be the vector field given by

$$[0,\infty) \times [0,\infty) \ni (x,y) \mapsto V(x,y) = \left(2x\left(1 - \frac{x}{2}\right) - xy, 3y\left(1 - \frac{y}{3}\right) - 2xy\right). \tag{7.51}$$

The dynamical system associated to is a simple model describing two species x and y competing for the same source of food.

(a) Plot the vector field V on the domain $[0,2] \times [0,2]$.

(b) Calculate the critical points of V.

(c) In physical terms, describe (do not calculate) the integral curves of V in the vicinity of the critical points. What happens as you move slightly away from the critical points but still remain within the domain $(0,\infty) \times (0,\infty)$?

(d) Describe the integral curves for the initial conditions where $x(0) = 0$.

(e) Describe the integral curves for the initial conditions where $y(0) = 0$.

Exercise 7.52. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field with associated system of differential equations given by

$$\begin{aligned} \dot{x} &= \big(1 - (x^2 + y^2)\big)x - y \\ \dot{y} &= \big(1 - (x^2 + y^2)\big)y + x \end{aligned} \tag{7.53}$$

(a) Plot the vector field V on the domain $[-2, 2] \times [-2, 2]$.

(b) Find the critical points of this dynamical system.

(c) Without explicitly solving the system, describe the integral curves.

(d) Are there any periodic orbits?

(e) What do all the trajectories look like they are being "attracted" to?

(f) Can you convert this set of coupled first order differential equations into a single second order differential equation? If so, try to solve it and find explicit solutions.

Exercise 7.54. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) = \big(y, -x - (x^2 - 1)y\big). \tag{7.55}$$

(a) Plot the vector field V on the domain $[-3, 3] \times [-4, 4]$.

(b) Find the critical points of this dynamical system.

(c) Without explicitly solving the system, describe the integral curves.

(d) Are there any periodic orbits?

(e) What do all the trajectories look like they are being "attracted" to?

(f) Can you convert this set of coupled first order differential equations into a single second order differential equation? If so, try to solve it and find explicit solutions.

Exercise 7.56. In 1963, Edward Lorenz, a mathematician and meteorologist born in West Hartford, Connecticut, discovered a mathematical model which he used to describe atmospheric convection. The differential equation he found to describe the motion of particles in air is given by

$$\begin{aligned} \dot{x} &= 10(y - x) \\ \dot{y} &= x(28 - z) - y \\ \dot{z} &= xy - \frac{8}{3}z \end{aligned} \tag{7.57}$$

Incidentally, this model can be used to describe a wide variety of physical phenomena.

(a) Plot the vector field associated to this differential equation on any domain that you believe illustrates the overall features of the vector field (you may want to rescale the vectors in some way so that it is easier to visualize). Warning: do not attempt to draw this by hand.

(b) Calculate the critical points of V.

(c) Many initial conditions tend to be "attracted" to a subset of $\mathbb{R}^3$ and the trajectories asymptotically approach this subset. Plot one such trajectory, enough to see a qualitative description of this attraction. The set is known as a strange attractor. For a precise definition (which you are capable of understanding!), see https://en.wikipedia.org/wiki/Attractor though you do not need to know the definition to analyze this problem.

A large class of examples of dynamical systems comes from classical physics. The motion of N particles in $\mathbb{R}^m$ subject to a force is described adequately by a function $H : \mathbb{R}^{2mN} \to \mathbb{R}$ known as the *Hamiltonian*. Let us write $\mathbb{R}^{2mN} = \mathbb{R}^{mN}_{\text{pos}} \times \mathbb{R}^{mN}_{\text{mom}}$ to distinguish between the two variables—the first stands for the position of the particles and the second for their momentum (momentum is closely related to velocity). Therefore, we will often denote the variable for position by x and the variable for momentum by p and use indices for the different particles. Many Hamiltonians are of the form

$$H(\vec{x}_1, \ldots, \vec{x}_N, \vec{p}_1, \ldots, \vec{p}_N) = \sum_{i=1}^{N} \frac{|\vec{p}_i|^2}{2m_i} + V(\vec{x}_1, \ldots, \vec{x}_N), \tag{7.58}$$

where m_i is a non-negative real number interpreted as the mass of the i-th particle and $V : \mathbb{R}^{mN}_{\text{pos}} \to \mathbb{R}$ is the potential, a function closely related to the force F by the formula $F = -\nabla V$, where ∇ is the gradient (see exercises in [13]). For convenience, we will consider a single particle so that $N = 1$ and we only have a single index for our position and momentum variables. When H is differentiable, it gives rise to a vector field $V_H : \mathbb{R}^m_{\text{pos}} \times \mathbb{R}^m_{\text{mom}} \to \mathbb{R}^m_{\text{pos}} \times \mathbb{R}^m_{\text{mom}}$ by the formula

$$\mathbb{R}^m_{\text{pos}} \times \mathbb{R}^m_{\text{mom}} \ni (x, p) \mapsto V_H(x, p) := \big(\partial_p H, -\partial_x H\big), \tag{7.59}$$

where $\partial_p H$ and $\partial_x H$ are shorthand notation for

$$\partial_p H := \big(\partial_{m+1} H, \partial_{m+2} H, \ldots, \partial_{2m} H\big) \qquad \& \qquad \partial_x H := \big(\partial_1 H, \partial_2 H, \ldots, \partial_m H\big) \tag{7.60}$$

though sometimes the notation

$$\partial_p H := \big(\partial_{p_1} H, \partial_{p_2} H, \ldots, \partial_{p_m} H\big) \qquad \& \qquad \partial_x H := \big(\partial_{x_1} H, \partial_{x_2} H, \ldots, \partial_{x_m} H\big) \tag{7.61}$$

is used instead in case it is not clear that the position variables are always placed first. In other words, the differential equation associated to this vector field is

$$\begin{aligned} \dot{x}_1 &= \partial_{p_1} H \\ &\vdots \\ \dot{x}_m &= \partial_{p_m} H \\ \dot{p}_1 &= -\partial_{x_1} H \\ &\vdots \\ \dot{p}_m &= -\partial_{x_m} H \end{aligned} \qquad \begin{matrix} \text{or remaining consistent} \\ \text{with our earlier notation} \end{matrix} \qquad \begin{aligned} \dot{x}_1 &= \partial_1 H \\ &\vdots \\ \dot{x}_m &= \partial_m H \\ \dot{p}_1 &= -\partial_{m+1} H \\ &\vdots \\ \dot{p}_m &= -\partial_{2m} H \end{aligned} \tag{7.62}$$

These are known as *Hamilton's equations of motion.*

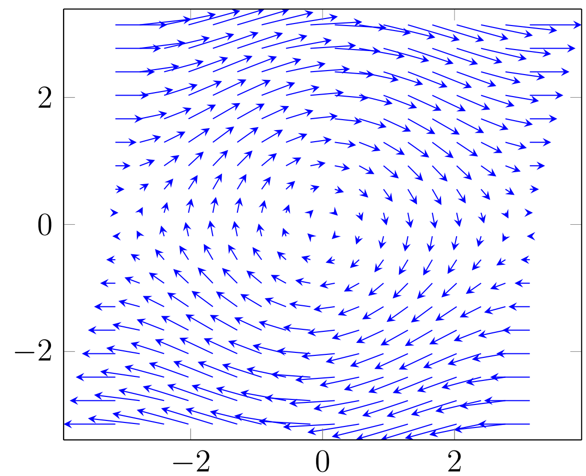
Example 7.63. Let $H : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (\theta, p) \mapsto H(\theta, \omega) := \frac{p^2}{2} + 1 - \cos\theta. \tag{7.64}$$

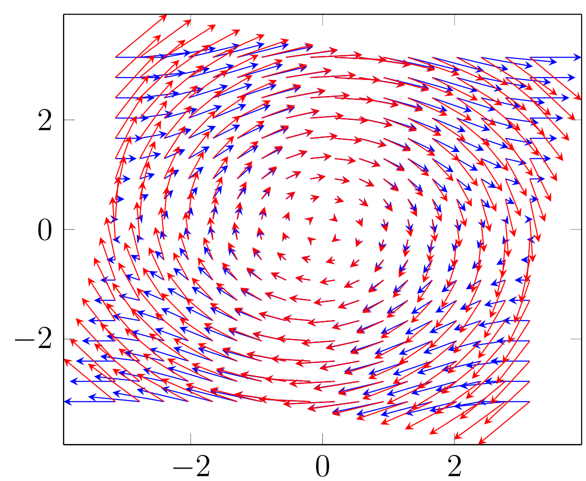
This describes the motion of a pendulum (of mass 1, length 1, and gravitational constant set to 1), where θ is the angle between the rod that holds the pendulum and the direction of gravity and p is the angular momentum. The associated system of differential equations (Hamilton's equations) is

$$\begin{aligned} \dot{\theta} &= p \\ \dot{p} &= -\sin\theta \end{aligned} \tag{7.65}$$

and the vector field looks like



Notice that for small θ, this looks like Example 7.15, which makes sense because $\sin\theta \approx \theta + \text{h.o.t.}$ (remember, h.o.t. stands for higher order terms). In fact, we can compare these two vector fields by overlaying them.



You can see that near the critical point, the two vector fields behave very similarly.

Later in this course, we will specify conditions under which integral curves exist for vector fields and their uniqueness. This will follow from the existence and uniqueness of solutions to ordinary differential equations, which is a theorem we will prove using contraction mappings on metric spaces along with existence and uniqueness of fixed points. However, to get to this point, we will need to discuss integration and metric spaces, and will therefore take a few more weeks.

Definition 7.66. A vector field $V : \mathbb{R}^n \to \mathbb{R}^n$ is *integrable* iff for every $c \in \mathbb{R}^n$, there exists an integral curve through c whose domain is all of $\mathbb{R}$.

One of the important things about the Hamiltonian is that it is a constant of the motion.

Definition 7.67. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be an integrable vector field. A *constant of the motion* (also known as an *integral of motion*) is a function $I : \mathbb{R}^n \to \mathbb{R}$ such that the composition $I \circ \gamma : \mathbb{R} \to \mathbb{R}^n \to \mathbb{R}$ is constant for all integral curves $\gamma : \mathbb{R} \to \mathbb{R}^n$ for the vector field V.

Theorem 7.68. *Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a differentiable function, a Hamiltonian, and suppose that the associated Hamiltonian vector field $V_H : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is integrable. Then H is a constant of the motion.*

Proof. This follows from the chain rule and Hamilton's equations of motion. ■

Exercise 7.69. Prove Theorem 7.68.

Constants of the motion help us solve complicated differential equations. Given an integrable vector field $V : \mathbb{R}^n \to \mathbb{R}^n$ and a constant of the motion $I : \mathbb{R}^n \to \mathbb{R}$, we know that if $\gamma : \mathbb{R} \to \mathbb{R}^n$ is an integral curve, then since $I \circ \gamma : \mathbb{R} \to \mathbb{R}$ is constant, let $c \in \mathbb{R}$ denote this constant. The curve γ must lie on the set in $\mathbb{R}^n$ described by

$$I^{-1}(c) := \big\{(x_1, \dots, x_n) \in \mathbb{R}^n \ : \ I(x_1, \dots, x_n) = c\big\}. \tag{7.70}$$

Often, this set is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$ (the definition of which will be given in a few lectures). For example, a torus in $\mathbb{R}^3$ can be described by a function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x, y, z) = \left(x^2 + y^2 + z^2 + R^2 - r^2\right)^2 - 4R^2\left(x^2 + y^2\right) \tag{7.71}$$

with $R > r > 0$ constants.

There is an important theorem that guarantees when a function $I : \mathbb{R}^n \to \mathbb{R}$ and a constant $c \in I(\mathbb{R}^n)$ in the image of I describes a manifold via the set of points $x \in \mathbb{R}^n$ satisfying $I(x) = c$. We will also state and prove this theorem later in the course. This provides two important observations that will motivate the study of manifolds and vector fields on manifolds.

1. When a constant of the motion exists, generically, a system of n coupled first order differential equations is reduced to a system of $(n-1)$ coupled first order differential equations. Therefore, the system is simpler to analyze from this perspective.

2. However, a complication arises in the following sense. Even though we began with a vector field in $\mathbb{R}^n$, constants of motion restrict attention to subsets of $\mathbb{R}^n$ that need not be vector subspaces. More often, they are submanifolds, and the vector fields on $\mathbb{R}^n$ restrict to these

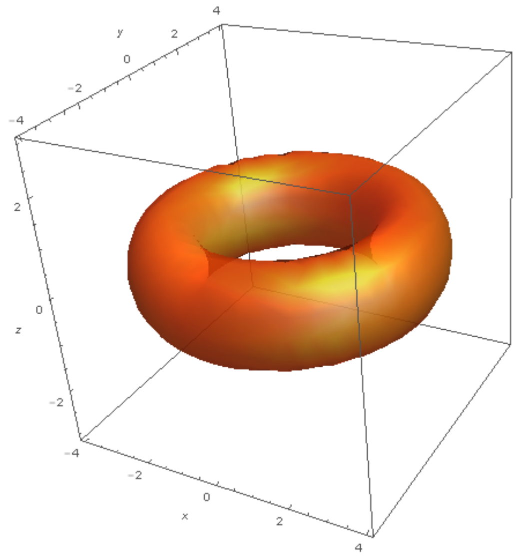


Figure 1: A plot of the manifold $f^{-1}(0)$ with $R = 3$ and $r = 1$. This was drawn in Mathematica.

submanifolds in a precise sense (that will be described later) and provide new dynamical systems on the corresponding manifolds. Interestingly, the topology of a manifold can very often give you a lot of information about the behavior of such vector fields and the existence of critical points. For example, it is guaranteed that every continuous vector field on the sphere S^2 has at least one critical point (this statement is false in $\mathbb{R}^2$—just take the vector field E_1 for example).

Not every vector field is integrable. Notice that the vector field from Exercise 7.30 is not integrable because of the first coordinate. Indeed, the vector field on $\mathbb{R}$ given by

$$\mathbb{R} \ni x \mapsto x^2 \tag{7.72}$$

is not integrable. An integrable vector field gives rise to a very special type of function.

Definition 7.73. Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be an integrable vector field. The *flow generated by V* is the collection of functions $\{\Phi_t : \mathbb{R}^n \to \mathbb{R}^n\}_{t\in\mathbb{R}}$ defined by

$$\mathbb{R}^n \ni x \mapsto \Phi_t(x) := \gamma_x(t), \tag{7.74}$$

where γ_x is an integral curve of V through x.

Later in the course, we will actually prove a theorem that guarantees the existence of solutions to initial value problems. In particular, we will prove a theorem about the existence of flows locally for vector fields that are not necessarily integrable. Furthermore, depending on the type of vector field (continuous, differentiable, continuously differentiable, etc.), the flow will have different properties. With such conditions, one of the important properties of a flow is that it is invertible. to explore this in more detail, we will first need to discuss inverse functions and their differentiability.

Level II problems.

From these notes: Exercises 7.5, 7.28, 7.30, 7.38, 7.69

Level III problems.
From these notes: Exercises 7.41, 7.47, 7.50, 7.52, 7.54, 7.56

8 February 23: Inverse Functions

This is now covered on the second day of Week #05 and is the only content for the week.

We now come back to more rigorous aspects of analysis. Today, we will study inverse functions and their properties. First, we will *suppose* that an inverse exists and is differentiable and use this to calculate the differential of the inverse. Then, we will supply sufficient conditions for when the inverse exists with a differentiable inverse. The implicit function theorem will be stated and proved in the following lectures. This theorem is crucial, for example, in understanding the local structure of manifolds.

Theorem 8.1. *Let $A \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, let $c \in A$, and let $f : A \to \mathbb{R}^n$ be a function. Furthermore, suppose there exist open neighborhoods U of c and V of $f(c)$ together with a function $g : V \to \mathbb{R}^n$ such that*

(a) $c \in U \subseteq A$ and $f(c) \subseteq V$,

(b) f is differentiable at $c \in A$,

(c) g is differentiable at $f(c) \in V$,

(d) $f(U) = V$ and $g(V) = U$, and

(e) $g \circ f = \mathrm{id}_U$ and $f \circ g = \mathrm{id}_V$.

Then,

$$D_{f(c)}g = (D_c f)^{-1}, \tag{8.2}$$

where $(D_c f)^{-1}$ denotes the inverse of the linear transformation $D_c f$.

Proof. By assumption, the diagrams

$$\begin{tikzcd} & V \arrow[dl, "g"'] & \\ U & & U \arrow[ul, "f"'] \arrow[ll, "\mathrm{id}_U"] \end{tikzcd} \quad \& \quad \begin{tikzcd} & U \arrow[dl, "f"'] & \\ V & & V \arrow[ul, "g"'] \arrow[ll, "\mathrm{id}_V"] \end{tikzcd} \tag{8.3}$$

both commute. Applying the chain rule to these diagrams says the diagrams

$$\begin{tikzcd} & \mathbb{R}^n \arrow[dl, "D_{f(c)}g"'] & \\ \mathbb{R}^n & & \mathbb{R}^n \arrow[ul, "D_c f"'] \arrow[ll, "D_c(g\circ f)=D_c(\mathrm{id}_U)=\mathrm{id}_{\mathbb{R}^n}"] \end{tikzcd} \quad \& \quad \begin{tikzcd} & \mathbb{R}^n \arrow[dl, "D_c f"'] & \\ \mathbb{R}^n & & \mathbb{R}^n \arrow[ul, "D_{f(c)}g"'] \arrow[ll, "D_{f(c)}(f\circ g)=D_{f(c)}(\mathrm{id}_V)=\mathrm{id}_{\mathbb{R}^n}"] \end{tikzcd} \tag{8.4}$$

both commute. The equalities on the bottom arrows follows from Theorem 4.43 and the fact that the derivative of a linear map equals that same linear map (the derivative of the identity is the identity). ∎

Recall that the inverse of a finite-dimensional matrix exists if and only if its determinant is non-zero. Therefore, in order for the inverse function to be differentiable, it must be that the differential of f has non-zero determinant. Consider, for example, the function $f : \mathbb{R} \to \mathbb{R}$ defined by $\mathbb{R} \ni x \mapsto f(x) := x^3$. Then $D_0 f = 0$. Now, the inverse *function* $g : \mathbb{R} \to \mathbb{R}$ of f *exists* everywhere since it is given by $\mathbb{R} \ni y \mapsto g(y) = \sqrt[3]{y}$ but its derivative is not defined at the origin. Note that g can be written as $f^{-1} := g$ so that Equation (8.2) can be written more suggestively as

$$D_{f(c)} f^{-1} = (D_c f)^{-1}. \tag{8.5}$$

Written this way, the formula for the inverse might be much easier to remember than it would be if expressed in terms of partial derivatives.

Moving on, we wish to drop several of the assumptions in Theorem 8.1 and find sufficient conditions for when an inverse exists, at least locally. For example, one of the assumptions is that f is one-to-one in some open neighborhood of U containing c (this is the condition that a *function* g exists satisfying $g \circ f = \mathrm{id}_U$). This actually follows from the assumption that the differential of f is one-to-one. Before stating and proving the inverse function theorem, we need a few more facts from linear algebra. Recall Exercise 1.10 in [13], which states that for any linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$, there exists a real number $M \geq 0$ such that

$$\big|T(v)\big| \leq M|v| \qquad \forall\, v \in \mathbb{R}^n. \tag{8.6}$$

The following exercise expands on this concept.

Exercise 8.7. For any linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$, define

$$\|T\| := \inf \Big\{ M \geq 0 \ : \ \big|T(v)\big| \leq M|v| \quad \forall\, v \in \mathbb{R}^n \Big\}. \tag{8.8}$$

(a) Prove that $\|T\|$ exists and is well-defined for any linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$.

(b) Prove that $\|T\| \geq 0$ for all linear transformations $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ and $\|T\| = 0$ if and only if $T = 0$.

(c) Prove that $\|\lambda T\| = |\lambda|\|T\|$ for all $\lambda \in \mathbb{R}$ and linear transformations $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$.

(d) Prove that $\|T + S\| \leq \|T\| + \|S\|$ for all linear transformations $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ and $\mathbb{R}^m \xleftarrow{S} \mathbb{R}^n$.

(e) Prove that $\|T \circ S\| \leq \|T\|\|S\|$ for all composable pairs of linear transformations $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n \xleftarrow{S} \mathbb{R}^p$.

(f) Prove that

$$\|\mathrm{id}_{\mathbb{R}^n}\| = 1 \qquad \& \qquad \|\pi_i\| = 1, \tag{8.9}$$

where $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is the i-th projection.

(g) Viewing a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ as an $m \times n$ matrix (using the standard basis), prove that the function $\|\cdot\| : \mathbb{R}^{mn} \to \mathbb{R}$ given by

$$\mathbb{R}^{mn} \ni T \mapsto \|T\| \tag{8.10}$$

is continuous.

Given a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$, the quantity $\|T\|$ is called the <u>*operator norm*</u> of T.

Theorem 8.11 (Inverse Function Theorem)**.** *Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^n$ be of class C^1 on A. If $\det(D_c f) \neq 0$, then there exists open sets $U, V \subseteq \mathbb{R}^n$ and a function $g : V \to \mathbb{R}^n$ such that*

(a) $c \in U$ and $f(c) \in V$,

(b) $f : U \to V$ and $g : V \to U$ are inverses of each other,

(c) g is continuous on V and differentiable at $f(c)$ with $D_{f(c)}g = (D_c f)^{-1}$.

The proof of this theorem is lengthy and will be broken into several steps.

Proof. The following proof will be broken up into four steps. First, it will be shown that f is one-to-one in some open neighborhood of c. Second, a construction of g on some open neighborhood of $f(c)$ will be provided. Third, it will be shown that g is continuous in some open neighborhood of $f(c)$. Finally, it will be shown that g is differentiable at $f(c)$.
Step 1. To prove that f is one-to-one, a stronger result will be shown, which will be useful for a later part of the proof of the theorem.
Lemma. Under the same hypothesis as in the the statement of the theorem, there exists an $\epsilon > 0$ and an $\alpha > 0$ such that

$$\big|f(x) - f(y)\big| \geq \alpha|x - y| \tag{8.12}$$

for all $x, y \in V_\epsilon(c)$.

Proof of Lemma. Set

$$\alpha := \frac{1}{2\|(D_c f)^{-1}\|}, \tag{8.13}$$

which is well-defined by parts (a) and (b) of Exercise 8.7. Then,

$$2\alpha|x - y| = 2\alpha\Big|(D_c f)^{-1}\big((D_c f)(x - y)\big)\Big| \leq 2\alpha\big\|(D_c f)^{-1}\big\|\big|(D_c f)(x - y)\big| = \big|(D_c f)(x - y)\big|. \tag{8.14}$$

This inequality will be used shortly. Define the function $H : A \to \mathbb{R}^n$ by

$$H := f - D_c f. \tag{8.15}$$

Then $D_x H = D_x f - D_c f$ for all $x \in A$ and in particular $D_c H = 0$. Because f is C^1 (and because all linear transformations are C^1), H is C^1. Since $\|\cdot\|$ is continuous by part (g) of Exercise 8.7, $A \ni x \mapsto \|D_x H\|$ is a continuous function on A since it is the composition of two continuous functions, namely $A \xrightarrow{D_\square H} \mathbb{R}^{mn} \xrightarrow{\|\cdot\|} \mathbb{R}$.[11] Because $\|D_c H\| = 0$, there exists an $\epsilon > 0$ such that

$$\|D_x H\| < \frac{\alpha}{\sqrt{n}} \qquad \forall\, x \in V_\epsilon(c). \tag{8.16}$$

[11] $A \xrightarrow{D_\square H} \mathbb{R}^{mn}$ is continuous because each of the projections $\pi_{ij} \circ D_\square H$ are continuous since these are precisely the partial derivatives $\partial_j H_i$ and are continuous since H is of class C^1 on A.

By the Mean Value Theorem (restricting to the i-th component function of H), for any $x, y \in V_\epsilon(c)$, there exists an $a \in V_\epsilon(c)$ such that

$$\begin{aligned}
\big|H_i(x) - H_i(y)\big| &= \big|(D_a H_i)(x-y)\big| && \text{by MVT} \\
&\le \|D_a H_i\||x-y| && \text{by Def'n of } \|\cdot\| \\
&= \|\pi_i \circ D_a H\||x-y| && \text{by the Chain Rule} \\
&\le \|\pi_i\|\|D_a H\||x-y| && \text{by part (e) of Exercise 8.7} \\
&= \|D_a H\||x-y| && \text{by part (f) of Exercise 8.7} \\
&< \frac{\alpha}{\sqrt{n}}|x-y| && \text{by Equation (8.16).}
\end{aligned} \tag{8.17}$$

Since this result is true for all $i \in \{1, \dots, n\}$, it follows that

$$\big|H(x) - H(y)\big| = \sqrt{\sum_{i=1}^{n} \big|H_i(x) - H_i(y)\big|^2} \le \sqrt{\sum_{i=1}^{n} \frac{\alpha^2}{n}|x-y|^2} = \sqrt{\frac{\alpha^2}{n} n|x-y|^2} = \alpha|x-y|. \tag{8.18}$$

Therefore,

$$\begin{aligned}
2\alpha|x-y| - \big|f(x) - f(y)\big| &\le \big|(D_c f)(x) - (D_c f)(y)\big| - \big|f(x) - f(y)\big| \quad \text{by (8.14)} \\
&= \big|f(x) - (D_c f)(x) - f(y) + (D_c f)(y) + f(y) - f(x)\big| - \big|f(x) - f(y)\big| \\
&\le \big|f(x) - (D_c f)(x) - f(y) + (D_c f)(y)\big| + \big|f(y) - f(x)\big| - \big|f(x) - f(y)\big| \\
&\qquad \text{by the triangle inequality} \\
&= \big|H(x) - H(y)\big| \quad \text{by Def'n of } H \\
&\le \alpha|x-y| \quad \text{by (8.18)}
\end{aligned} \tag{8.19}$$

Canceling terms, this gives $\big|f(x) - f(y)\big| \ge \alpha|x-y|$ as needed. ■

(Back to Proof of Theorem.) From this Lemma, it follows that f is one-to-one on $V_\epsilon(c)$ because

$$\big|f(x) - f(y)\big| \ge \alpha|x-y| > 0 \tag{8.20}$$

for all distinct x and y in $V_\epsilon(c)$, which says $f(x) \neq f(y)$ for all distinct x and y in $V_\epsilon(c)$. By choosing ϵ to be smaller if necessary, it also follows that f is one-to-one and satisfies this inequality on $\overline{V_\epsilon(c)}$.
Step 2. The result from Step 1 implies, in particular, that the restriction of f to $V_\epsilon(c)$ is one-to-one. In this step, it will be shown that $f\big(V_\epsilon(c)\big)$ is an open subset of $\mathbb{R}^n$ containing $f(c)$. This can then be used to define an inverse function and the fact that $f\big(V_\epsilon(c)\big)$ is open will be used in proving that the inverse is continuous in Step 3.
Lemma. Under the same hypothesis as in the the statement of the theorem, let $\epsilon > 0$ be small enough so that $f : \overline{V_\epsilon(c)} \to \mathbb{R}^n$ is one-to-one. Then $f\big(V_\epsilon(c)\big)$ is an open subset of $\mathbb{R}^n$ that contains $f(c)$.

Proof of Lemma. The goal is to show that for any $y \in f\big(V_\epsilon(c)\big)$, there exists an open neighborhood $U \subseteq \mathbb{R}^n$ such that $y \in U \subseteq f\big(V_\epsilon(c)\big)$. Without loss of generality, it suffices to focus on the point $y := f(c)$. Therefore, the goal is to show that there exists a $\delta > 0$ such that $V_\delta\big(f(c)\big) \subseteq f\big(V_\epsilon(c)\big)$.

Since $\partial\overline{V_\epsilon(c)}$ is compact, $f\left(\partial\overline{V_\epsilon(c)}\right)$ is a compact subset of $\mathbb{R}^n$. Since f is one-to-one, $f(c) \notin f\left(\partial\overline{V_\epsilon(c)}\right)$. Since $\{f(c)\}$ and $f\left(\partial\overline{V_\epsilon(c)}\right)$ are disjoint compact subsets of $\mathbb{R}^n$, Problem 1.21 in [13] shows that there exists a $\delta > 0$ such that

$$d\Big(\big\{f(c)\big\}, f\left(\partial\overline{V_\epsilon(c)}\right)\Big) > 2\delta. \tag{8.21}$$

Hence, $\overline{V_{2\delta}\big(f(c)\big)}$ is disjoint from $f\left(\partial\overline{V_\epsilon(c)}\right)$. Let $z \in V_\delta\big(f(c)\big)$. The goal is to show that $z = f(x)$ for some $x \in V_\epsilon(c)$ because it must be shown that $f\big(V_\epsilon(c)\big)$ contains each point of some open neighborhood around $f(c)$, i.e. the goal is to show that $V_\delta\big(f(c)\big) \subseteq f\big(V_\epsilon(c)\big)$. To do this, first define $\phi : A \to \mathbb{R}$ to be the function

$$A \ni x \mapsto \phi(x) := \big|f(x) - z\big|^2. \tag{8.22}$$

By Exercise 4.69, ϕ is C^1. Since $\overline{V_\epsilon(c)}$ is compact, f restricted to $\overline{V_\epsilon(c)}$ attains a minimum value (by Problem 1.29 in [13]), say at $x \in \overline{V_\epsilon(c)}$. Because $z \in \overline{V_\delta\big(f(c)\big)}$,

$$\phi(c) = \big|f(c) - z\big|^2 < \delta^2. \tag{8.23}$$

Therefore, the minimum value of ϕ must not only be less than δ^2, but also cannot occur on $\left(\partial\overline{V_\epsilon(c)}\right)$ by (8.21). Therefore, $x \in V_\epsilon(c)$. By Theorem 5.56,

$$D_x\phi = 0. \tag{8.24}$$

By the Chain Rule,

$$D_x\phi = \big(D_{f(x)-z}|\ \cdot\ |^2\big) \circ (D_x f). \tag{8.25}$$

By Exercise 4.69,

$$D_{f(x)-z}|\ \cdot\ |^2 = 2\langle f(x) - z,\ \cdot\ \rangle. \tag{8.26}$$

Since $\det(D_x f) \neq 0$, one can multiply both sides of (8.25) by $(D_x f)^{-1}$ turning equation (8.25) into

$$0 = 2\langle f(x) - z,\ \cdot\ \rangle. \tag{8.27}$$

By non-degeneracy of the inner product (part (a) of Theorem 1.33), it follows that $f(x) - z = 0$, i.e. $f(x) = z$. ■

(Back to Proof of Theorem.)
Step 3. In this step, it will be shown that the inverse $g : f\big(V_\epsilon(c)\big) \to \mathbb{R}^n$ of f is continuous. Let $U \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$ and set $V := U \cap V_\epsilon(c)$. Then, $g^{-1}(V) = f(V)$ since f is one-to-one on $V_\epsilon(c)$. Since V is an open subset of $V_\epsilon(c)$, Step 2 shows that $f(V)$ is open in $\mathbb{R}^n$. Hence $g^{-1}(V)$ is open in $\mathbb{R}^n$ and therefore also open in $f\big(V_\epsilon(c)\big)$.
Step 4. In the final step, it will be shown that the inverse $g : f\big(V_\epsilon(c)\big) \to \mathbb{R}^n$ of f is differentiable at $f(c)$ with differential $D_{f(c)}g = (D_c f)^{-1}$. By Step 2, there exist open neighborhoods $U \subseteq \mathbb{R}^n$ around c contained in A and $V := f(U)$ around $f(c)$ on which the inverse $g : V \to U$ is defined and continuous. Let

$$W := \{h \in \mathbb{R}^n\ :\ f(c) + h \in V\} \setminus \{0\}. \tag{8.28}$$

Define the function $G : W \to \mathbb{R}^n$ by

$$W \ni k \mapsto G(k) := \frac{g\big(f(c)+k\big) - g\big(f(c)\big) - \big((D_c f)^{-1}\big)(k)}{|k|} \tag{8.29}$$

It is also convenient to define $\Delta : W \to \mathbb{R}^n$ by

$$W \ni k \mapsto \Delta(k) := g\big(f(c)+k\big) - g\big(f(c)\big). \tag{8.30}$$

The goal is to show $\lim\limits_{k\to 0} \big|G(k)\big| = 0$. To do this, it helps to rewrite the function G as

$$G(k) = -\big((D_c f)^{-1}\big) \underbrace{\left(\frac{k - (D_c f)\big(\Delta(k)\big)}{\big|\Delta(k)\big|} \right)}_{(*)} \frac{\big|\Delta(k)\big|}{|k|}, \tag{8.31}$$

which is valid for all $k \in W$. Each of the three terms will be analyzed separately in the limit as k tends to 0. The first term is a constant and is unaffected by the limit. For the second term, note that by definition of Δ,

$$g\big(f(c)\big) + \Delta(k) = g\big(f(c)+k\big). \tag{8.32}$$

Applying f to both sides gives

$$\begin{tikzcd} f\Big(g\big(f(c)\big) + \Delta(k)\Big) \arrow[r, equal, bend left] \arrow[d, equal, bend right, "\text{since } g\circ f=\mathrm{id}"'] & f\Big(g\big(f(c)+k\big)\Big) \arrow[d, equal, bend left, "\text{since } f\circ g=\mathrm{id}"] \\ f\big(c+\Delta(k)\big) & f(c)+k \end{tikzcd} \tag{8.33}$$

Solving for k and plugging it into $(*)$ gives

$$(*) = \frac{k - (D_c f)\big(\Delta(k)\big)}{\big|\Delta(k)\big|} = \frac{f\big(c+\Delta(k)\big) - f(c) - (D_c f)\big(\Delta(k)\big)}{\big|\Delta(k)\big|}. \tag{8.34}$$

Since g is continuous, $\lim\limits_{k\to 0} \Delta(k) = 0$ (see Exercise 4.3.4. in [1] and generalize the claim to higher dimensions). Using this and the fact that f is continuous and differentiable at c,

$$\lim_{k\to 0} \frac{\big|k - (D_c f)\big(\Delta(k)\big)\big|}{\big|\Delta(k)\big|} = 0. \tag{8.35}$$

Finally, for the third term in the expression for $G(k)$, using the Lemma from Step 1 and setting

$$x := g\big(f(c)+k\big) \qquad \& \qquad y := g\big(f(c)\big) \tag{8.36}$$

gives the inequality

$$\big|f(c)+k-f(c)\big| \ge \alpha \Big| g\big(f(c)+k\big) - g\big(f(c)\big)\Big|, \tag{8.37}$$

which says

$$\frac{1}{\alpha} \geq \frac{\big|\Delta(k)\big|}{|k|}. \tag{8.38}$$

Hence, $W \ni k \mapsto \frac{|\Delta(k)|}{|k|}$ is a bounded function on W. Therefore, $\lim_{k\to 0} G(k) = 0$ (see Exercise 4.2.7. in [1] and generalize the claim to higher dimensions) proving that g is differentiable at $f(c)$. With this, the proof is complete. ■

The following two exercises were used in the proof of the inverse function theorem above and are generalizations of exercises 4.3.4 and 4.2.7 in [1], respectively.

Exercise 8.39. Let $f : \mathbb{R}^p \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. Suppose that $\lim_{x\to a} f(x) = b \lim_{y\to b} g(y) = c$. Show that if g is continuous at b, then

$$\lim_{x\to a} g\big(f(x)\big) = c. \tag{8.40}$$

Exercise 8.41. Let $A \subseteq \mathbb{R}^n$ and let $f, g : A \to \mathbb{R}$ be two functions on A such that f is bounded, i.e. there exists an $M > 0$ such that $\big|f(x)\big| \leq M$ for all $x \in A$. Let c be a limit point of A. Show that if $\lim_{x\to c} g(x) = 0$ then $\lim_{x\to c} \big(g(x)f(x)\big) = 0$.

Exercise 8.42. Prove the following closely related theorem: "Let $A \subseteq \mathbb{R}^n$ be open, let $c \in A$, and let $f : A \to \mathbb{R}^n$ be of class C^k on A for some $k \in \mathbb{N}$. If $\det(D_c f) \neq 0$, then there exists open sets $U, V \subseteq \mathbb{R}^n$ and a function $g : V \to \mathbb{R}^n$ such that

(a) $c \in U$ and $f(c) \in V$,

(b) $f : U \to V$ and $g : V \to U$ are inverses of each other,

(c) g is of class C^k on V."

Exercise 8.43. Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}^n$ be continuously differentiable, and suppose that $\det(D_x f) \neq 0$ for all $x \in A$. Show that $f(A)$ is an open subset of $\mathbb{R}^n$.

Remark 8.44. Recall the following result from Analysis I: "Let $a, b \in \mathbb{R}$ with $a \leq b$ and let $f : [a, b] \to \mathbb{R}$ be a continuous function that is one-to-one. Then f^{-1} exists and is also continuous on its domain." The following exercise asks if a similar theorem is true in higher dimensions.

Exercise 8.45. Let $R \subseteq \mathbb{R}^n$ be a closed rectangle and let $f : R \to \mathbb{R}^n$ be a continuous function that is one-to-one. Prove or disprove: f^{-1} exists and is also continuous on its domain. [Warning: I think this is very difficult to prove and might be beyond what we learn in analysis—according to [9], this result (or rather a closely related one) is due to Brouwer.]

The next three exercises are from [9].

Exercise 8.46. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := (x^2 - y^2, 2xy). \tag{8.47}$$

Let

$$A := \big\{(x, y) \in \mathbb{R}^2 \ : \ x > 0\big\}. \tag{8.48}$$

(a) Show that f is one-to-one on A.

(b) Describe mathematically and draw the image $B := f(A)$ of A under f explicitly.

(c) Let $g : B \to \mathbb{R}^2$ denote the inverse function. Using the standard Euclidean basis, express $D_{(0,1)}g$ as a 2×2 matrix.

Exercise 8.49. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := (e^x \cos y, e^x \sin y). \tag{8.50}$$

Let

$$A := \big\{(x, y) \in \mathbb{R}^2 \ : \ 0 < y < 2\pi\big\}. \tag{8.51}$$

(a) Show that f is one-to-one on A.

(b) Describe mathematically and draw the image $B := f(A)$ of A under f explicitly.

(c) Let $g : B \to \mathbb{R}^2$ denote the inverse function. Using the standard Euclidean basis, express $D_{(0,1)}g$ as a 2×2 matrix.

Exercise 8.52. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}^3$ be the functions defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto g(x, y) := (2ye^{2x}, xe^y) \tag{8.53}$$

and

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := (3x - y^2, 2x + y, xy + y^3), \tag{8.54}$$

respectively.

(a) Show that there exists an open neighborhood U of $(0, 1) \in \mathbb{R}^2$ and an open neighborhood V of $(2, 0) \in \mathbb{R}^2$ such that $g : U \to V$ is a bijection.

(b) Using the standard Euclidean basis, express $D_{(2,0)}(f \circ g^{-1})$ as a matrix.

The following exercise provides an intuition for the norm of an operator.

Exercise 8.55. Let $\mathbb{R}^n \xleftarrow{T} \mathbb{R}^n$ be a *symmetric* linear transformation, i.e.

$$\big\langle T(u), v\big\rangle = \big\langle u, T(v)\big\rangle \qquad \forall\, u, v \in \mathbb{R}^n,$$

or equivalently, if $[T]$ is the matrix associated to T, then $[T]^T = [T]$ (the superscript here means the transpose of $[T]$). Furthermore, suppose there exists a basis $\{x_1, \ldots, x_n\}$ of vectors in $\mathbb{R}^n$ and numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that $Tx_i = \lambda_i x_i$ for all $i \in \{1, \ldots, n\}$. Prove that

$$\|T\| = \max_{i \in \{1,\ldots,n\}} \big\{|\lambda_i|\big\}. \tag{8.56}$$

Level I problems.
From these notes: Exercise 8.7 part (a), 8.39, 8.41

Level II problems.
From Spivak [13]: 2.37, 2.38, 2.39
From these notes: Exercises 8.7 parts (b), (c) and (d), 8.7 part (e), 8.7 part (f), 8.7 part (g), 8.42, 8.43, 8.46, 8.49, 8.52, 8.55

Level III problems.
From these notes: Exercise 8.45 (warning: I do not know how easy this is)

9 February 16: Brief recap (buffer for exam review)

There is a quiz today and afterwards, we will review today in class for the exam. You may present solutions for any week that passed, especially if there was a particular week which you wish to obtain more points for. I will distribute the points to your benefit. If you have sufficient points for earlier weeks, I will distribute them towards Week #05 even though you don't present Week #05 problems today.

Do not forget that there is a midterm on February 21 covering Lectures 1 through 8!

10 February 28: Implicit Functions

This content is now covered on the first day of Week #06.

Consider a small patch of the surface of the earth such that the curvature of the earth can be neglected.[12] Setting the origin of $\mathbb{R}$ to be sea level and setting $\mathbb{R}^2$ to represent a point at sea level on earth, the height above sea level at any such point defines a function $h : \mathbb{R}^2 \to \mathbb{R}$. This function can be depicted by a topographical map.[13] For each value h_0 of the height, the curves depict the set of points $(x, y) \in \mathbb{R}^2$ for which $h(x, y) = h_0$.

As you can see from this figure, given a fixed $h_0 \in h(\mathbb{R}^2)$, it is not trivial to explicitly solve for the variable x in terms of the variable y or vice versa. The best we could hope for is to solve for the variable x in terms of y *near* some particular points (x_0, y_0). Let's say you sprained your ankle hiking a mountain and it hurts to walk up and down steep trails on a mountain. On your map, you see that if you walk from your current position (x_0, y_0) at elevation h_0 to another position (x_0', y_0') at that same elevation, the trail becomes significantly less steep and you can make it somewhat safely to help. Initially, if you move a certain number of steps in the x direction, you need to move a certain number of steps in the y direction. At every moment, the distance you travel in the two directions changes. If you wanted, you could define a function that tells you at all times how many steps to move in the y direction given that you move a certain number in the x direction. This defines a function y as a function of x. Unfortunately, this method could fail after some steps though you can always reverse it and define a function x of y when this happens.

In general, you might have a function $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ of a *vector* variable $x \in \mathbb{R}^n$ and another variable $y \in \mathbb{R}$. In this case, you might want to solve for y in terms of x given some fixed constant $h_0 \in f(\mathbb{R}^n \times \mathbb{R})$. One example of such a thing is a more general differential equation than the one we've studied in the previous lecture. For instance, if $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a function and $f : \mathbb{R} \to \mathbb{R}$

[12] The present discussion is motivated by one of the responses given at http://matheducators.stackexchange.com/questions/22/what-are-some-good-examples-to-motivate-the-implicit-function-theorem.

[13] Image obtained from Al Walsh's post on http://www.justtrails.com/nav-skills/navigation-skills-topographic-maps/ (2013).

is an n-times differentiable function, then a general differential equation involving f is

$$F\big(f, f^{(1)}, \ldots, f^{(n-1)}, f^{(n)}\big) = 0. \tag{10.1}$$

It would be convenient occasionally to be able to solve for $f^{(n)}$ and express the differential equation in the form we had from last lecture. At least, we can hope to obtain an expression for $f^{(n)}$ in terms of the lower derivatives locally and solve the differential equation in some suitably small domain.

The previous discussion was the case $m = 1$, i.e. wanting to solve for a single variable. Even more generally, you might have a system of functions $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ with associated m component functions.

Exercise 10.2. Come up with a reasonable physical situation of such a system where solving for some set of variables would provide significant insight into the physical phenomenon (for $n > 1$ and $m > 1$).

The Implicit Function Theorem provides a sufficient condition for when one can do this. But before stating this theorem, a converse of it is helpful in understanding it from a more mathematical perspective [9]. And even before stating the converse, some notation should be mentioned.

Definition 10.3. For every $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$, let $\varphi_y : \mathbb{R}^k \to \mathbb{R}^{k+n}$ and $\psi_x : \mathbb{R}^n \to \mathbb{R}^{k+n}$ be the functions defined by[14]

$$\mathbb{R}^k \ni x' \mapsto \varphi_y(x') := (x', y) \qquad \& \qquad \mathbb{R}^n \ni y' \mapsto \psi_x(y') := (x, y'). \tag{10.4}$$

Note that the differentials of these functions at $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$, respectively, are given by

$$\mathbb{R}^k \ni u \mapsto (D_x\varphi_y)(u) := (u, 0) \qquad \& \qquad \mathbb{R}^n \ni v \mapsto (D_y\psi_x)(v) := (0, v), \tag{10.5}$$

respectively. For brevity in the following theorem, these inclusions at x and y will be denoted by $i : \mathbb{R}^k \to \mathbb{R}^k \times \mathbb{R}^n$ and $j : \mathbb{R}^n \to \times\mathbb{R}^k \times \mathbb{R}^n$, respectively:

$$i := D_x\varphi_y \qquad \& \qquad j := D_y\psi_x. \tag{10.6}$$

Theorem 10.7. *Let $A \subseteq \mathbb{R}^{k+n}$ be an open subset of the form $A = U \times V$ with $U \subseteq \mathbb{R}^k$ and $V \subseteq \mathbb{R}^n$ open subsets and let $f : A \to \mathbb{R}^m$ be differentiable on A. Let $B \subseteq \mathbb{R}^k$ be an open subset and let $g : B \to \mathbb{R}^n$ be a differentiable function satisfying*

$$B \times g(B) \subseteq A \tag{10.8}$$

and

$$f\big(x, g(x)\big) = 0 \qquad \forall\, x \in B. \tag{10.9}$$

Then[15]

$$D_x(f \circ \varphi_{g(x)}) + \big(D_{g(x)}(f \circ \psi_x)\big) \circ D_x g = 0, \tag{10.10}$$

[14] These functions are closely related to the ones from Theorem 5.13. Those were the special case of inclusions of single axes at a point. The functions here are inclusions of higher-dimensional planes at a point.

[15] For this equation to make sense, it must be understood that $\varphi_{g(x)}$ is restricted to $U \subseteq \mathbb{R}^k$ and ψ_x is restricted to $V \subseteq \mathbb{R}^n$.

i.e. using the notation from (10.6) after setting $y = g(x)$,[16]

$$\big(D_{(x,g(x))}f\big) \circ \big(i + j \circ D_x g\big) = 0. \tag{10.11}$$

Another way of writing (10.10) is to avoid the introduction of the functions φ_y and ψ_x and instead write it as

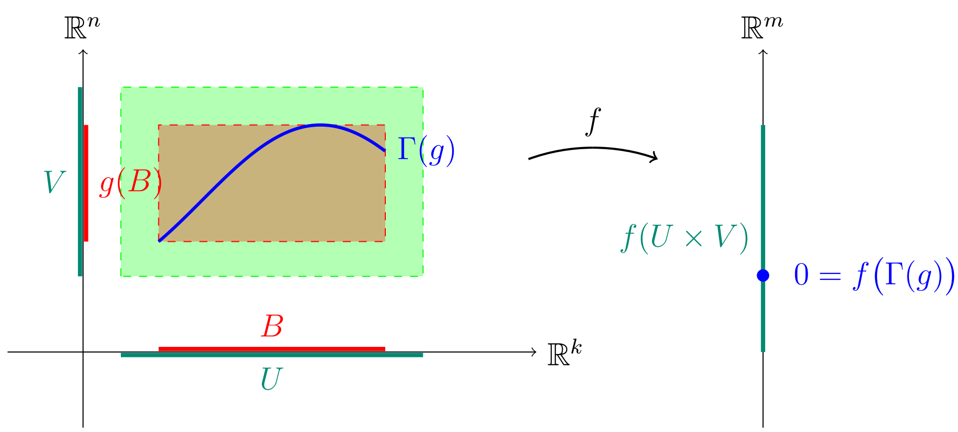
$$D_x f\big(\,\cdot\,, g(x)\big) + \big(D_{g(x)} f(x,\,\cdot\,)\big) \circ D_x g = 0,$$

where it is understood that the functions that are being placed inside the differential D are those for which the variable is left blank. Munkres also writes this theorem in a slightly different way than I have, which is probably easier to understand if you've taken calculus. His version of equation (10.10) is written as

$$\frac{\partial f}{\partial x}\big(x, g(x)\big) + \frac{\partial f}{\partial y}\big(x, g(x)\big) D_x g = 0, \tag{10.12}$$

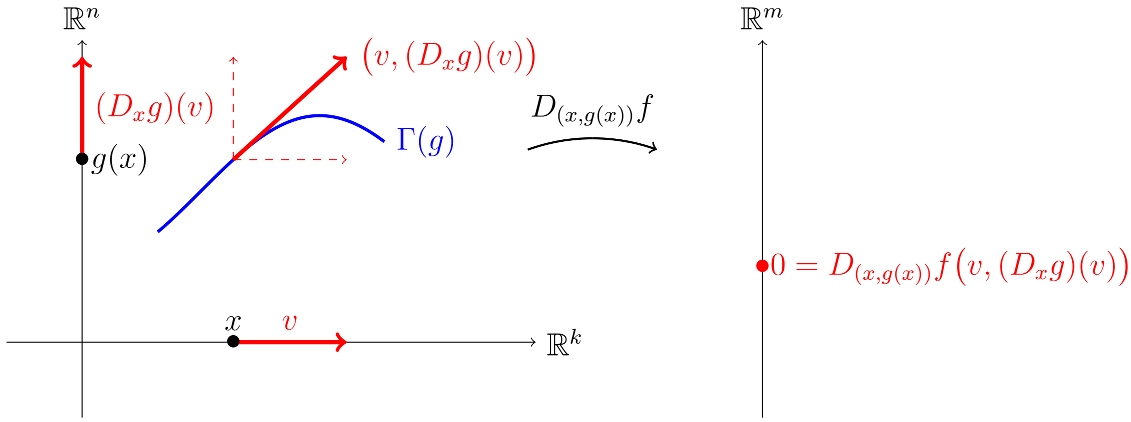
where $\frac{\partial f}{\partial x}\big(x, g(x)\big)$ is the $m \times k$ matrix of partial derivatives of f with respect to the first k coordinates evaluated at the point $\big(x, g(x)\big)$ and $\frac{\partial f}{\partial y}\big(x, g(x)\big)$ is the $m \times n$ matrix of partial derivatives of f with respect to the last n coordinates evaluated at the point $\big(x, g(x)\big)$. For a proof of the theorem using this kind of notation, please refer to [9].

Unfortunately, all of these different ways of expressing the result might make it difficult to understand. However, equation (10.11) can actually be expressed quite intuitively via the following *geometric* interpretation of the theorem. The *assumptions* in the theorem can be summarized in the following figure

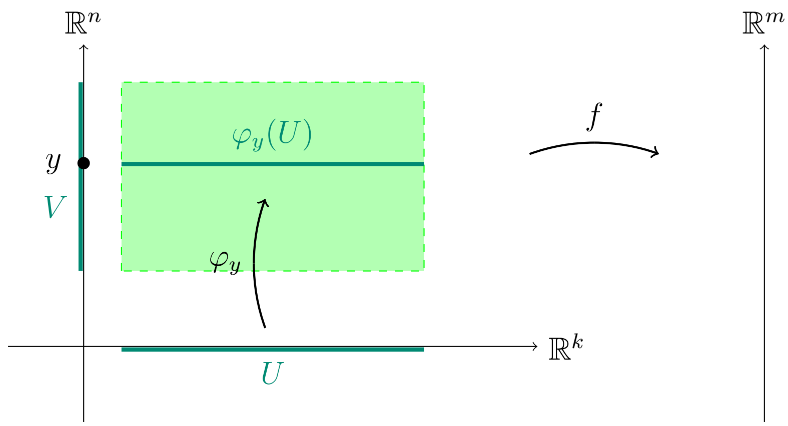


where $\Gamma(g)$ denotes the graph of g. The *conclusions* of the theorem can be summarized in the following version of the previous figure together with images of tangent vectors being pushed forward by the differentials.

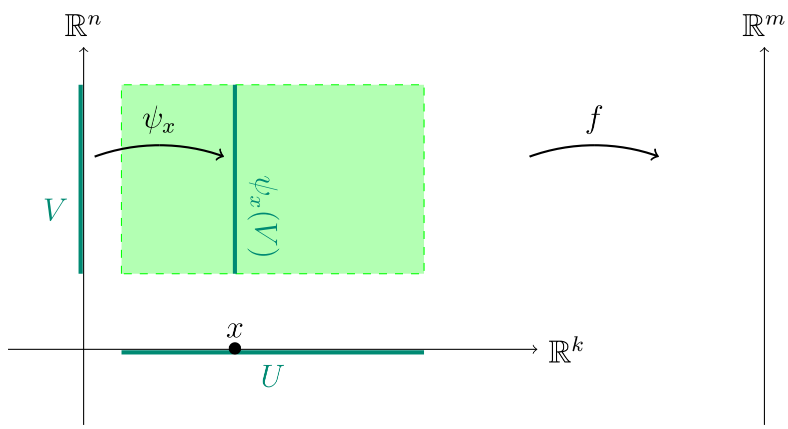
[16]The Chain Rule was applied here.



For completeness of visualizations, we also draw the functions $f \circ \varphi_y$ and $f \circ \psi_x$ as the following cartoon.



and



respectively.

In what follows, we will give a more diagrammatic proof of the theorem.

Proof of Theorem 10.7. Conditions (10.8) and (10.9) can be expressed by saying that the diagram

$$
\begin{tikzcd}
A \arrow[d, "f"'] & B \times g(B) \arrow[l, hook'] & B \times B \arrow[l, "\mathrm{id}\times g"'] \\
\mathbb{R}^m & & B \arrow[ll, "0"] \arrow[u, "\Delta"']
\end{tikzcd}
\tag{10.13}
$$

commutes. Here $\Delta : B \to B \times B$ is the diagonal map that sends x to (x, x) for all $x \in B$ and $\hookrightarrow$ is just the inclusion of sets. All of these functions are differentiable, and hence the Chain Rule applies producing the commutative diagram

$$
\begin{tikzcd}
\mathbb{R}^k \times \mathbb{R}^n \arrow[d, "D_{(x,g(x))}f"'] & & \mathbb{R}^k \times \mathbb{R}^k \arrow[ll, "\mathrm{id}_{\mathbb{R}^k}\times D_x g"'] \\
\mathbb{R}^m & & \mathbb{R}^k \arrow[ll, "0"] \arrow[u, "\Delta"']
\end{tikzcd}
\tag{10.14}
$$

because $D_x\Delta = \Delta$ since Δ is the restriction of a linear map. Commutativity of this diagram says that

$$
(D_{(x,g(x))}f) \circ (\mathrm{id}_{\mathbb{R}^k} \times D_x g) \circ \Delta = 0, \tag{10.15}
$$

which is an equality of linear transformations from $\mathbb{R}^k$ to $\mathbb{R}^m$. It suffices to show that the left-hand-side of this is equal to (10.10). For every vector $v \in \mathbb{R}^k$, this linear transformation acts by

$$
\begin{aligned}
\big((D_{(x,g(x))}f) \circ (\mathrm{id}_{\mathbb{R}^k} \times D_x g) \circ \Delta\big)(v) &= (D_{(x,g(x))}f)\big(v, (D_x g)(v)\big) \\
&= (D_{(x,g(x))}f)\Big(\big(v, 0\big) + \big(0, (D_x g)(v)\big)\Big) \\
&= (D_{(x,g(x))}f)\Big(i(v) + j\big((D_x g)(v)\big)\Big) \\
&= (D_{(x,g(x))}f)\big(i + j \circ (D_x g)\big)(v),
\end{aligned}
\tag{10.16}
$$

which proves the claim. ■

In terms of matrix notation, the matrix associated to $D_{(x,y)}f$ for any $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$ can be expressed as

$$
[D_{(x,y)}f] = \begin{bmatrix} \big[D_x(f \circ \varphi_y)\big] & \big[D_y(f \circ \psi_x)\big] \end{bmatrix} \tag{10.17}
$$

and the vector $v \in \mathbb{R}^k$ is applied via matrix multiplication

$$
\begin{bmatrix} \big[D_x(f \circ \varphi_{g(x)})\big] & \big[D_{g(x)}(f \circ \psi_x)\big] \end{bmatrix} \begin{bmatrix} v \\ (D_x g)(v) \end{bmatrix} \tag{10.18}
$$

to produce the last line in (10.16).

Exercise 10.19. In the statement and proof of Theorem 10.7, it seems that a rather stronger assumption, namely that $B \times g(B) \subseteq A$ (this assumption is not explicitly stated in [9]), was used.

This assumption can be weakened since the differential of a function is well-defined in some open neighborhood of a point. Hence, one can replace B in the proof by an open neighborhood when calculating the differential at any point $x \in B$. In addition, $A \subseteq \mathbb{R}^{k+n}$ need not be in the form of a product of two open sets, one in $\mathbb{R}^k$ and one in $\mathbb{R}^n$.

(a) Prove that the statement and proof of the theorem still remain valid if A is replaced with an arbitrary open subset of $\mathbb{R}^{k+n}$ and condition (10.8) is replaced with the following: "there exists an open cover $\{U_\alpha\}_{\alpha \in \mathcal{I}}$ of B such that

$$U_\alpha \times g(U_\alpha \cap B) \subseteq A \qquad \forall\, \alpha \in \mathcal{I}." \tag{10.20}$$

[Hint: I think you will also need to cover A by sufficiently small open rectangles, you may need to refine the cover of B, and you may need to use continuity of g to prove that g maps the cover of B inside the cover of A.]

(b) In the statement of this modified version of Theorem 10.7, do we need to assume that $g(U_\alpha \cap B)$ is an open subset of $\mathbb{R}^n$? Explain your answer.

(c) Is the theorem still true if we do not assume the existence of such an open cover $\{U_\alpha\}_{\alpha \in \mathcal{I}}$?

Corollary 10.21. *Under the same hypotheses as in Theorem 10.7 but with the additional assumptions that $m = n$ and $\det\big(D_{g(x)}(f \circ \psi_x)\big) \neq 0$, then*

$$D_x g = -\big(D_{g(x)}(f \circ \psi_x)\big)^{-1} \circ \big(D_x(f \circ \varphi_{g(x)})\big). \tag{10.22}$$

Exercise 10.23. Let B be an $n \times k$ matrix and let C be an $n \times n$ matrix.

(a) Prove that

$$\det \begin{bmatrix} \mathbb{1}_k & 0 \\ B & C \end{bmatrix} = \det(C). \tag{10.24}$$

(b) If A is a $k \times k$ matrix, is there a simple expression for

$$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} ? \tag{10.25}$$

If so, find it.

A partial converse of Theorem 10.7 is the following (the presentation mostly comes from [9] though the statement is a blend of what is in [9] and [13]).

Theorem 10.26 (Implicit Function Theorem)**.** *Let $A \subseteq \mathbb{R}^k \times \mathbb{R}^n$ be an open subset of the form $A = A_1 \times A_2$ with $A_1 \subseteq \mathbb{R}^k$ and $A_2 \subseteq \mathbb{R}^n$ open subsets, let $f : A \to \mathbb{R}^n$ be of class C^r on A, and let $c := (a, b) \in A$ with $a \in A_1$ and $b \in A_2$ be a point satisfying*

$$f(a, b) = 0 \tag{10.27}$$

and

$$\det\big(D_b(f \circ \psi_a)\big) \neq 0. \tag{10.28}$$

Then, there exists an open set $B \subseteq \mathbb{R}^k$ with $a \in B$ and a unique function $g : B \to \mathbb{R}^n$ on B with the following properties:

(a) g *is of class* C^r *on* B,

(b) $g(a) = b$,

(c) $B \times g(B) \subseteq A$, *and*

(d) $f\big(x, g(x)\big) = 0 \qquad \forall\, x \in B$.

Using calculus notation, the assumption (10.28) in the theorem would be written as

$$\det\left(\frac{\partial f}{\partial y}(a,b)\right) \neq 0. \tag{10.29}$$

Proof. Define the function $F : A \to \mathbb{R}^{k+n}$ to be the composition in the diagram

$$\begin{tikzcd} & A_1 \times A_1 \times A_2 \arrow[dl, "\mathrm{id}_1 \times f"'] & \\ A_1 \times \mathbb{R}^n & & A_1 \times A_2 \arrow[ll, "F"] \arrow[ul, "\Delta_1 \times \mathrm{id}_2"'] \end{tikzcd} \tag{10.30}$$

i.e.

$$A_1 \times A_2 \ni (x,y) \mapsto F(x,y) := \big(x, f(x,y)\big). \tag{10.31}$$

The Chain Rule implies

$$D_{(x,y)}F = \big(D_{(x,x,y)}(\mathrm{id}_1 \times f)\big) \circ \big(D_{(x,y)}(\Delta_1 \times \mathrm{id}_2)\big) = \big(\mathrm{id}_{\mathbb{R}^k} \times D_{(x,y)}f\big) \circ \big(\Delta_{\mathbb{R}^k} \times \mathrm{id}_{\mathbb{R}^n}\big). \tag{10.32}$$

In other words, for vectors $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^n$,

$$(D_{(x,y)}F)(u,v) = \big(\mathrm{id}_{\mathbb{R}^k} \times D_{(x,y)}f\big)(u,u,v) = \big(u, (D_{(x,y)}f)(u,v)\big), \tag{10.33}$$

which means that in the standard basis, the matrix associated to $D_{(x,y)}F$ is

$$[D_{(x,y)}F] = \begin{bmatrix} \mathbb{1}_k & 0 \\ \big[D_x(f \circ \varphi_y)\big] & \big[D_y(f \circ \psi_x)\big] \end{bmatrix}. \tag{10.34}$$

Thus, by Exercise 10.23,

$$\det\big(D_{(x,y)}F\big) = \det\big(D_y(f \circ \psi_x)\big) \tag{10.35}$$

for all $x \in A_1$ and $y \in A_2$. In particular,

$$\det\big(D_{(a,b)}F\big) = \det\big(D_b(f \circ \psi_a)\big) \neq 0 \tag{10.36}$$

by assumption. In addition

$$F(a,b) = (a,0). \tag{10.37}$$

by the other assumptions. Hence, by the Inverse Function Theorem and Exercise 8.42, there exist open sets $U \subseteq \mathbb{R}^k$, $V \subseteq \mathbb{R}^n$, and $W \subseteq \mathbb{R}^{k+n}$ such that

(a) $(a,b) \in U \times V \subseteq A_1 \times A_2$ and $(a,0) \in W$,

(b) $F : U \times V \to W$ is a one-to-one C^r map with $F(U \times V) = W$, and

(c) the inverse function $G : W \to U \times V$ of F is of class C^r.

Define the function $h : W \to V \subseteq \mathbb{R}^n$ by the composition

$$\begin{tikzcd} & U \times V \arrow[dl, "\pi_V"'] & \\ V & & W \arrow[ul, "G"'] \arrow[ll, "h"] \end{tikzcd} \qquad W \ni w \mapsto h(w) := \pi_V\big(G(w)\big), \tag{10.38}$$

where π_V stands for the projection onto V. Since the projection is C^∞, h is C^r. In addition, G preserves the first coordinate in the sense that the diagram

$$\begin{tikzcd} U \times V \arrow[dr, "\pi_U"'] & & W \arrow[ll, "G"'] \arrow[dl, "\pi_{\mathbb{R}^k}\big|_W"] \\ & \mathbb{R}^k & \end{tikzcd} \tag{10.39}$$

commutes. This shows that G is a function of the form

$$W \ni (x,z) \mapsto G(x,z) = \big(x, h(x,z)\big). \tag{10.40}$$

Because $(a,0) \in W$ and W is open, there exists an open rectangle $R \subseteq W$ containing $(a,0)$. Let $B := \pi_{\mathbb{R}^k}\big|_W(R)$ be the projection of this open rectangle onto $\mathbb{R}^k$. Let $g : B \to \mathbb{R}^n$ be the composition

$$\begin{tikzcd} & W \arrow[dl, "h"'] & \\ \mathbb{R}^n & & B \arrow[ul, "\varphi_0"'] \arrow[ll, "g"] \end{tikzcd} \qquad B \ni x \mapsto g(x) := h(x,0), \tag{10.41}$$

or equivalently, by the definition of h,

$$\begin{tikzcd} U \times V \arrow[d, "\pi_V"'] & W \arrow[l, "G"'] \\ V & B \arrow[l, "g"] \arrow[u, "\varphi_0"'] \end{tikzcd} \qquad B \ni x \mapsto g(x) := \pi_V\big(G(x,0)\big). \tag{10.42}$$

Finally, since $U \times V$ contains $G(R)$, there exists an open set Q of the form $Q = Q_1 \times Q_2$ with $Q_1 \subseteq U$ and $Q_2 \subseteq V$ such that $G(R) \subseteq Q \subseteq U \times V$ (in fact, Q_1 can be taken to be B itself since

G fixes the first coordinates). With all of these facts combined, we have the following diagram.

$$
\begin{tikzcd}
A \arrow[dddd, "f"'] \arrow[dr, phantom, "\begin{array}{c}\text{By def'n of}\\ \text{inclusion}\end{array}"] & & B \times g(B) \arrow[ll, hook'] \arrow[dl, hook] \arrow[dr, phantom, "\text{By def'n of } g \text{ (10.41)}"] & & B \times B \arrow[ll, "\mathrm{id}\times g"'] \arrow[dl, "\mathrm{id}_B\times\varphi_0"] \\
& B \times Q_2 \arrow[ul, hook] & & B \times R \arrow[ll, "\mathrm{id}_B\times h"] \arrow[dl, "\mathrm{id}_B\times G"] \arrow[dd, hook] \arrow[ddr, phantom, "\text{By def'n of } \varphi_0"] & \\
& & B \times Q_1 \times Q_2 \arrow[ul, "\mathrm{id}_B\times\pi_{Q_2}"'] \arrow[u, phantom, "\begin{array}{c}\text{By def'n of } h\\ \text{(10.38)}\end{array}"] \arrow[ll, phantom, "\text{By diagram chase}"] \arrow[r, phantom, "\begin{array}{c}\text{Since}\\ G := F^{-1}\end{array}"] \arrow[dl, "\mathrm{id}_B\times\Delta_{Q_1}\times\mathrm{id}_{Q_2}"'] \arrow[dr, "\mathrm{id}_B\times F"] \arrow[d, phantom, "\begin{array}{c}\text{By def'n of } F\\ \text{(10.30)}\end{array}"] & & \\
& B \times Q_1 \times Q_1 \times Q_2 \arrow[rr, "\mathrm{id}_{B\times Q_1}\times f"'] & & B \times Q_1 \times \mathbb{R}^n \arrow[dlll, two heads, "\pi_{\mathbb{R}^n}"'] & \\
\mathbb{R}^n & & & & B \arrow[llll, two heads, "0"] \arrow[uuuu, "\Delta"']
\end{tikzcd}
\tag{10.43}
$$

The outer most diagram is just (10.13). The outer most diagram commutes because every sub-diagram commutes by the previous results, which are referenced in the centers of the sub-diagrams. The phrase "by diagram chase" means that it follows from a short calculation to verify that the diagram indeed commutes. Note that $\twoheadrightarrow$ stands for surjection (due to the projection maps). In other words, commutativity of this diagram shows that g satisfies the condition that

$$f\big(x, g(x)\big) = 0 \qquad \forall\, x \in B. \tag{10.44}$$

Explicitly following the diagram on a given element $x \in B$ gives

$$
\begin{aligned}
f\big(x, g(x)\big) &= f\big(x, h(x,0)\big) \\
&= \pi_2\Big(x, f\big(x, h(x,0)\big)\Big) \\
&= \pi_2\Big(F\big(x, h(x,0)\big)\Big) \\
&= \pi_2\Big(F\big(G(x,0)\big)\Big) \\
&= \pi_2(x,0) \\
&= 0.
\end{aligned}
\tag{10.45}
$$

Furthermore,

$$g(a) = h(a,0) = \pi_V\big(G(a,0)\big) = \pi_V(a,b) = b. \tag{10.46}$$

This concludes the construction of $g : B \to \mathbb{R}^n$. Note that by Exercise 8.42, g is of class C^r on B. ■

Exercise 10.47. Finish the proof of the Implicit Function Theorem by proving that for any other open set $C \subseteq \mathbb{R}^k$ with $a \in C$ and for any other function $\alpha : C \to \mathbb{R}^n$ satisfying

(a) α is of class C^r on C,

(b) $\alpha(a) = b$,

(c) $C \times g(C) \subseteq A$, and

(d) $f\big(x, \alpha(x)\big) = 0 \qquad \forall\, x \in B$,

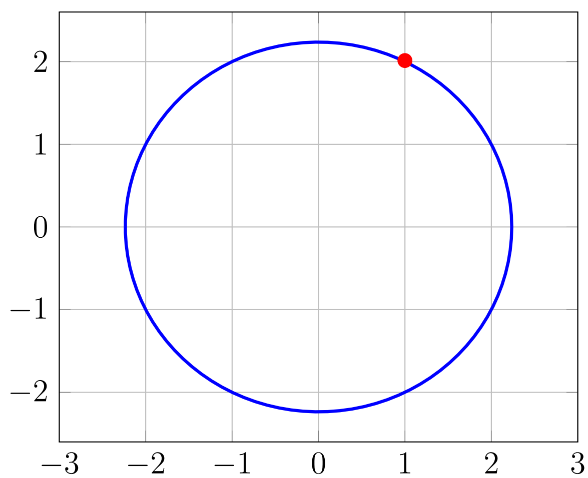
then the restrictions of α and g to $B \cap C$ are equal.

Actually applying the Implicit Function Theorem is far simpler in practice than the statement and proof of the theorem. To illustrate this point, we will take a simple example from [9] and go through each step of the proof to illustrate how it works.

Example 10.48. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 + y^2 - 5. \tag{10.49}$$

Here $k = 1$ and $n = 1$. $A_1 = \mathbb{R}$ and $A_2 = \mathbb{R}$. Thus, $A = \mathbb{R} \times \mathbb{R}$ and f is of class C^∞ on A. Consider the point $(a, b) = (1, 2)$. Then $f(1, 2) = 0$.

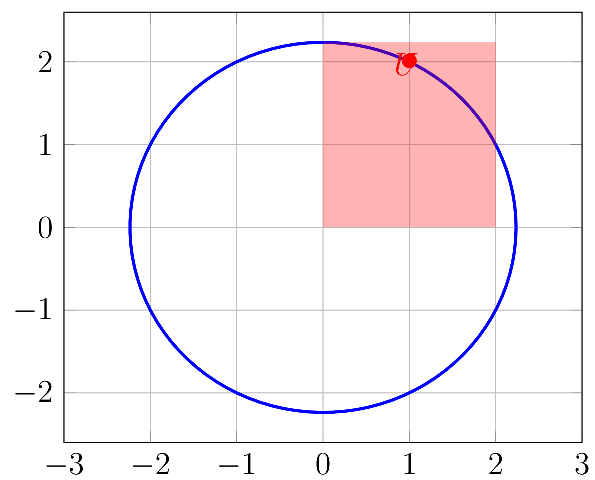


Furthermore, $[D_2(f \circ \psi_1)]$ is the one-by-one matrix $\begin{bmatrix}4\end{bmatrix}$ (this is a matrix whose sole entry is the number 4—this is not referring to reference [4]) since $[D_b(f \circ \psi_a)] = \begin{bmatrix}2b\end{bmatrix}$ for arbitrary a, b in the appropriate domains. The function $F : A \to \mathbb{R}^{1+1}$ is given by

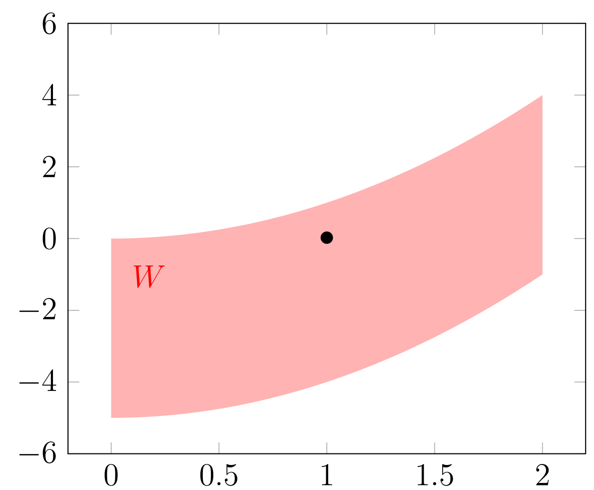
$$A \ni (x, y) \mapsto F(x, y) := (x, x^2 + y^2 - 5). \tag{10.50}$$

By the inverse function theorem, we know the open sets $U \subseteq \mathbb{R}^1$, $V \subseteq \mathbb{R}^1$, and $W \subseteq \mathbb{R}^{1+1}$ exist. We can set them to be (for instance)

$$U = (0, 2), \qquad V = \left(0, \sqrt{5}\right), \qquad \& \qquad W = \left\{(x, x^2 + y^2 - 5) \in \mathbb{R}^2 \ : \ x \in U, y \in V\right\}. \tag{10.51}$$



W is just the image of $U \times V$ under F and looks like the following subset of $\mathbb{R}^2$



with the $\bullet$ depicting the point $F\big((a,b)\big) = (a,0) = (1,0)$. Then $F : U \times V \to W$ is a bijection. The inverse $G : W \to U \times W$ is given by

$$W \ni (z,w) \mapsto G(z,w) = (z, \sqrt{w + 5 - z^2}). \tag{10.52}$$

Note that because of the way the domain W is defined, G is indeed of class C^∞ on W (the inside of the square root is positive for all $(w,z) \in W$). Furthermore, one can check by explicit computation that G and F are inverses:

$$F\big(G(z,w)\big) = F(z, \sqrt{w+5-z^2}) = \big(z, z^2 + (w+5-z^2) - 5\big) = (z,w) \tag{10.53}$$

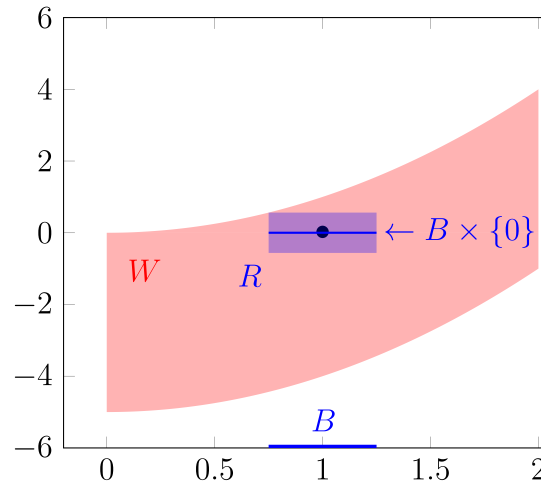
and

$$G\big(F(x,y)\big) = G(x, x^2+y^2-5) = (x, \sqrt{x^2+y^2-5+5-x^2}) = (x, \sqrt{y^2}) = (x,y) \tag{10.54}$$

since y is positive. Therefore, the function $h : W \to V$ is just the projection of G onto the second factor, i.e.

$$W \ni (z,w) \mapsto h(z,w) = \sqrt{w+5-z^2} \tag{10.55}$$

as we expected from the proof of the theorem. Now, one can choose R to be a rectangle containing $(1,0) \in W$ such as the one depicted in the following figure



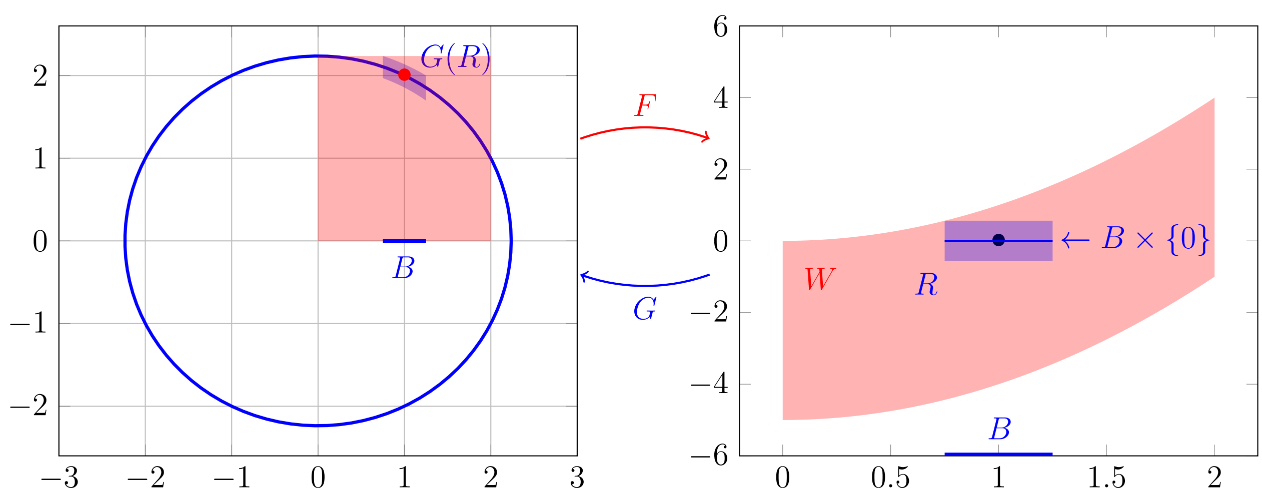
where the central line is the set $B \times \{0\}$, where B is the projection of R onto the first component. In the figure, the rectangle R is given by

$$R = \left(\frac{3}{4}, \frac{5}{4}\right) \times \left(-\frac{9}{16}, \frac{9}{16}\right) \tag{10.56}$$

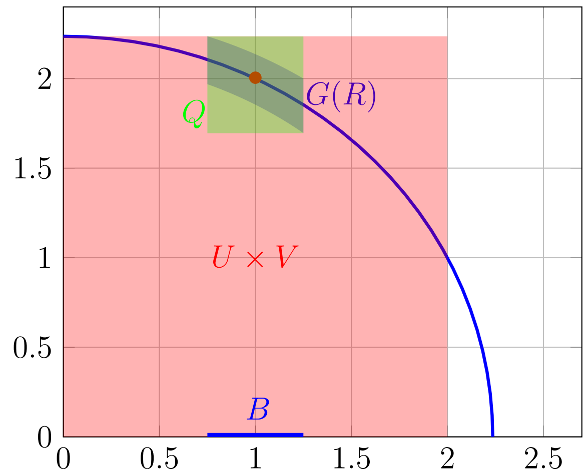
so that

$$B = \left(\frac{3}{4}, \frac{5}{4}\right). \tag{10.57}$$

We then look at $G(R)$ as a subset of $U \times V$ and the relationship between all these different subsets.



The rectangle $Q = Q_1 \times Q_2$ such that $G(R) \subseteq Q \subseteq U \times V$ can be taken to be the following subset

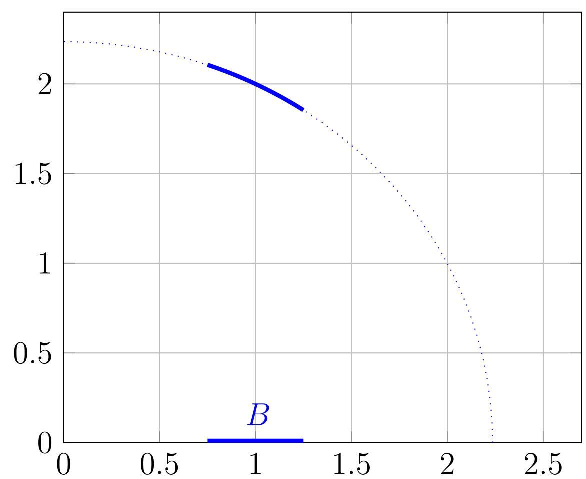


which is

$$Q = \left(\frac{3}{4}, \frac{5}{4}\right) \times \left(\sqrt{\frac{23}{8}}, \sqrt{5}\right). \tag{10.58}$$

Finally, the function $g : B \to V$ is given by the composition $g = h \circ \varphi_0$, which in this case is just

$$B \ni x \mapsto g(x) = h(x, 0) = \sqrt{5 - x^2}. \tag{10.59}$$



This is the long sought after function (the graph of which is drawn over the subset $B \subseteq U$). It would have been much simpler to simply solve the equation $x^2 + y^2 - 5 = 0$ for the variable y near the point $(1, 2)$ in this particular example. However, in more complicated examples, such manipulations might not be so easy. It is therefore helpful to illustrate the proof with this simple example as a stepping stone. Furthermore, the construction in the proof of the Implicit Function Theorem is completely general and works for *any* function f satisfying the assumptions of the theorem.

Exercise 10.60. Does the proof of the Implicit Function Theorem in the case of Example 10.48 still hold if U is chosen to be the open interval

$$U = \left(0, \sqrt{5}\right) \tag{10.61}$$

instead of $U = (0, 2)$? Explain.

Exercise 10.62. Can the rectangle R in Example 10.48 be chosen to contain the subset $(0,2) \times \{0\} \subseteq W$? Explain.

Exercise 10.63. Let $f : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ be of class C^1 and suppose that $f(3,-1,2) = 0$ and

$$[D_{(3,-1,2)}f] = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}. \tag{10.64}$$

(a) Show that there exists an open neighborhood B in $\mathbb{R}$ around 0 together with a C^1 function $g : B \to \mathbb{R}^2$ satisfying

$$f\big(x, g_1(x), g_2(x)\big) = 0 \qquad \forall\, x \in B \tag{10.65}$$

and

$$g(3) = (-1, 2). \tag{10.66}$$

(b) Calculate the matrix $[D_3 g]$.

(c) Discuss the problem of solving the equation $f(x, y_1, y_2) = 0$ for an arbitrary pair of variables in terms of the third variable near the point $(3,-1,2)$. For example, is it possible to solve for x and y_2 in terms of y_1 near this point? Ask and answer similar questions for the other possibilities as well.

Exercise 10.67. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be of class C^1 with $f(2,-1) = -1$. Define the functions $G, H : \mathbb{R}^3 \to \mathbb{R}$ by

$$\mathbb{R}^3 \ni (x,y,u) \mapsto G(x,y,u) := f(x,y) + u^2 \tag{10.68}$$

and

$$\mathbb{R}^3 \ni (x,y,u) \mapsto H(x,y,u) := ux + 3y^2 + u^3. \tag{10.69}$$

Note that

$$G(2,-1,1) = 0 = H(2,-1,1). \tag{10.70}$$

(a) What conditions on Df ensure that there exists an open neighborhood $U \subseteq \mathbb{R}$ of $-1 \in \mathbb{R}$ and C^1 functions $g, h : U \to \mathbb{R}$ such that

$$g(-1) = 2 \qquad \& \qquad h(-1) = 1 \tag{10.71}$$

and

$$G\big(g(y), y, h(y)\big) = 0 \qquad \& \qquad H\big(g(y), y, h(y)\big) = 0 \qquad \forall\, y \in U? \tag{10.72}$$

(b) Under the conditions you found in part (a) and in addition assuming that the differential has matrix form given by

$$[D_{(2,-1)}f] = \begin{bmatrix} 1 & -3 \end{bmatrix}, \tag{10.73}$$

calculate $g'(-1)$ and $h'(-1)$.

The following exercises are from [9] and provide a glimpse towards the theory of manifolds, which will be studied in the next few lectures.

Exercise 10.74. Let $f, g : \mathbb{R}^3 \to \mathbb{R}$ be functions of class C^1. Let $(a, b, c) \in \mathbb{R}^3$ be a point that satisfies

$$f(a, b, c) = 0 \qquad \& \qquad g(a, b, c) = 0. \tag{10.75}$$

Let $h : \mathbb{R}^3 \to \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto h(x, y, z) := \big(f(x, y, z), g(x, y, z)\big). \tag{10.76}$$

Show that if the differential $D_{(a,b,c)}h : \mathbb{R}^3 \to \mathbb{R}^2$ is surjective, then there exists an open neighborhood U of (a, b, c) for which two of the variables x, y, z can be solved in terms of the third. Note that this solution describes two surfaces intersecting at a common curve.

Another exercise in this direction but slightly more general is the following [9].

Exercise 10.77. Let $f : \mathbb{R}^{k+n} \to \mathbb{R}^n$ be of class C^1, let $a \in \mathbb{R}^{k+n}$ be a point for which $f(a) = 0$, and suppose that $D_a f : \mathbb{R}^{k+n} \to \mathbb{R}^n$ is onto. Show that there exists an open neighborhood $U \subseteq \mathbb{R}^n$ around 0 for satisfying the condition that for any $c \in U$, there exists an $x \in \mathbb{R}^{n+k}$ such that $f(x) = c$.

Exercise 10.78. Following the outline of Example 10.48, go through the proof of the implicit function theorem applied to the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto f(x, y, z) := x^2 + y^2 + z^2 - 5 \tag{10.79}$$

about the point $c := (0, 1, 2)$.

Exercise 10.80. Following the outline of Example 10.48, go through the proof of the implicit function theorem applied to the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto f(x, y, z) := x^3 + y^2 - z^2 \tag{10.81}$$

about the point $c := (0, 1, 1)$.

Exercise 10.82. Following the outline of Example 10.48, go through the proof of the implicit function theorem applied to the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto f(x, y, z) := (x^2 + y^2 + z^2 + 8)^2 - 36(x^2 + y^2) \tag{10.83}$$

about the point $c := (0, 3, 1)$.

Level I problems.
From these notes: Exercises 10.60, 10.62

Level II problems.
From Spivak [13]: 2.40, 2.41
From these notes: Exercises 10.23, 10.47, 10.63, 10.67, 10.74, 10.77, 10.78, 10.80

Level III problems.
From these notes: Exercises 10.19, 10.82

11 March 2: Manifolds

Intuitively, manifolds are subsets of $\mathbb{R}^k$ that *locally* look Euclidean. In particular, they should have no corners, and they should have a well-defined (non negative and integral) dimension. Making this intuition precise relies on many of the definitions we've made and the results we have proven up until this point. We have already seen many examples of manifolds. For instance, graphs of differentiable functions are manifolds (once the definition below is provided, this observation will be a theorem and is listed as Problem 5.6 in Spivak [13]). However, not all manifolds can be described as the graph of a function. Indeed, we have seen several examples of manifolds described by level sets of functions. A simple example is the unit sphere in any Euclidean space. Another example is the torus. However, as Exercises 10.78, 10.80, and 10.82 suggest, we can express these subsets *locally* as graphs of functions on suitable small domains, and then "patch" these *local* "charts" to describe the entire subset. Furthermore, if the subset is compact, we can guarantee that a finite number of such graphs suffices to cover the entire subset (these claims will be made precise soon). Before providing the definition, we first describe what we mean by saying that "something looks like something else."

Definition 11.1. Fix $r \in \{0\} \cup \mathbb{N} \cup \{\infty\}$. Let $U, V \subseteq \mathbb{R}^k$. A <u>*$C^r$-diffeomorphism*</u> between U and V consists of functions

$$\begin{tikzcd} U \arrow[r, shift left, "\varphi"] & V \arrow[l, shift left, "\psi"] \end{tikzcd} \tag{11.2}$$

of class C^r, where φ and ψ are inverses of each other.[17]

By the uniqueness of inverses, it is common to simply say that $\varphi : U \to V$ is a C^r-diffeomorphism without referring to the inverse $\psi = \varphi^{-1}$.

The observations made at the beginning of this lecture motivate the following definition of a manifold. There are several equivalent definitions of a manifold.

Definition 11.3. Fix $r \in \{0\} \cup \mathbb{N} \cup \{\infty\}$. A subset $M \subseteq \mathbb{R}^k$ is a <u>*C^r-manifold (without boundary) of dimension m*</u> iff for every point $c \in M$, there exists

(a) an open neighborhood $U \subseteq \mathbb{R}^k$ of c,

(b) an open set $V \subseteq \mathbb{R}^k$, and

(c) C^r-diffeomorphism

$$\begin{tikzcd} U \arrow[r, shift left, "\varphi"] & V \arrow[l, shift left, "\psi"] \end{tikzcd} \tag{11.4}$$

[17] Recall from Definition 4.19 that a function defined on a *subset* of Euclidean space is differentiable/C^r at a point iff there exists an open neighborhood of that point and an extension of that function that is differentiable/C^r at that point. This concept is used to make sense of differentiability/C^r-ness of both φ and ψ in this definition. A function such as φ or ψ is then by definition differentiable/C^r on its entire domain if and only if there exists an open set and a differentiable/C^r function on this open set that agrees with the original function. It was (essentially) stated in Exercise 4.20 that a function is differentiable/C^r on a subset of $\mathbb{R}^k$ if and only if it is differentiable/C^r at every point of its domain. Although this is true, proving it uses techniques that we have not yet covered. We need to know about partitions of unity, which we may or may not discuss. Nevertheless, we will assume this equivalence is true without proving it.

such that

$$\varphi(U \cap M) = V \cap \big(\mathbb{R}^m \times \{0\}\big), \tag{11.5}$$

where the latter set $\mathbb{R}^m \times \{0\} \subseteq \mathbb{R}^k$ is explicitly described as

$$\mathbb{R}^m \times \{0\} := \{y \in \mathbb{R}^k \ : \ y_{m+1}, y_{m+2}, \dots, y_k = 0\}. \tag{11.6}$$

The terminology *smooth manifold* will occasionally be used instead of C^∞-manifold. The term *manifold* will be used whenever it is a C^r manifold for some r. The map $\varphi|_{U\cap M}$, which is φ restricted to $U \cap M$, is called a *coordinate patch* of M about c and the map ψ is called a *parametrization* of M about c.

The phrase "C^r" in Definition 11.3 may also be replaced with "r-times differentiable" though we will rarely discuss such manifolds. A few exercises are listed at the end of this lecture that discuss this in some detail.

Please see Chapter 5 of Spivak [13] for some nice illustrations of this definition.

Example 11.7. In Example 10.48 we constructed such a diffeomorphism pair about the point $c = (1, 2) \in M$, where M was given by $f^{-1}(0)$ with $f : \mathbb{R}^2 \to \mathbb{R}$ the function defined by sending $(x, y) \in \mathbb{R}^2$ to $f(x, y) := x^2 + y^2 - 5$. Unfortunately, some of the notation conflicts. The following chart summarizes the corresponding open sets and diffeomorphisms.

Notation from Definition 11.3	Notation from Example 10.48
U	$U \times V$
V	W
φ	F
ψ	G

Exercise 11.8. Example 11.7 only provides a diffeomorphism pair about the point $c = (1, 2)$. To prove that S^1 is a manifold, one must be able to construct a diffeomorphism pair about every point. This exercises asks you to do this.

(a) Using this definition directly prove that the unit circle $S^1 \subseteq \mathbb{R}^2$ is a C^∞-manifold of dimension 1 in $\mathbb{R}^2$. [Hint: refer to Example 10.48 for an idea for a parametrization.]

(b) Show that the function $\alpha : [0, 1) \to S^1$ given by

$$[0, 1) \ni t \mapsto \alpha(t) := \big(\cos(2\pi t), \sin(2\pi t)\big) \tag{11.9}$$

is not a coordinate patch on S^1 (besides the fact that $[0, 1)$ is not an open subset of $\mathbb{R}$).

Note that a given coordinate patch might cover several points on S^1 so that a finite number of patches will suffice to cover all of S^1.

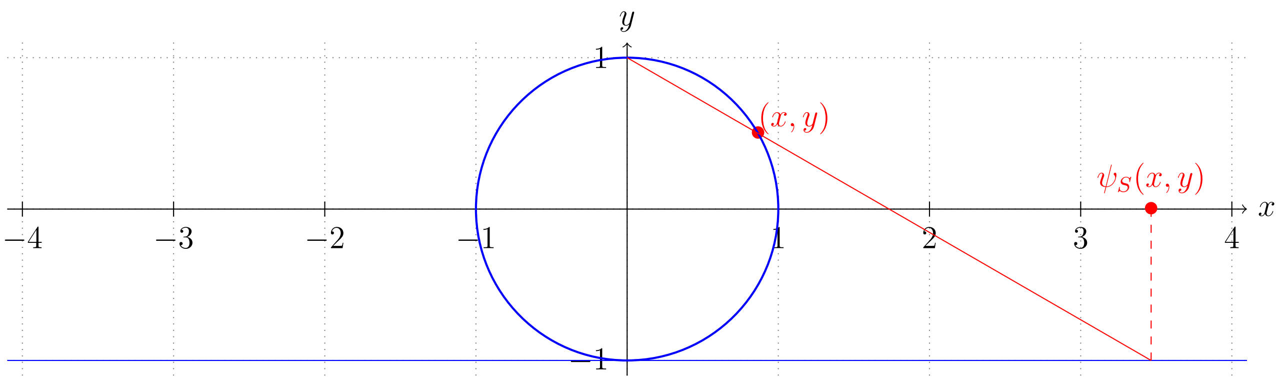
Exercise 11.10. Prove that the unit circle,

$$S^1 := \big\{(x, y) \in \mathbb{R}^2 \ : \ x^2 + y^2 = 1\big\}, \tag{11.11}$$

is a one-dimensional manifold in $\mathbb{R}^2$ by using the following coordinate charts, known as a *stereographic projections*, and completing the following steps.

(a) Show that $U_N := S^{-1} \setminus \{(0,-1)\}$ and $U_S := S^{-1} \setminus \{(0,1)\}$ are open subsets of S^1. Here $(0,1) \in \mathbb{R}^2$ is the "north pole" of S^1 and $(0,-1)$ is the "south pole."

(b) Find an explicit formula for the function $\psi_S : U_S \to \mathbb{R}$ that sends a point $(x,y) \in U_S$ to the horizontal coordinate obtained by the straight line that passes from the north pole through (x,y) and intersecting the line $y = -1$. For a visualization, see the following figure:



Note that $(0,-1)$ gets sent to 0 under this map and $\psi_S(0,1)$ is not defined because one point does not uniquely define a line. Find a similar formula for U_N this time emanating from the south pole passing through the point (x,y) on S^1 and intersecting the line $y = 1$.

(c) Prove that your functions $U_S \to \mathbb{R}$ and $U_N \to \mathbb{R}$ are C^∞ and are indeed coordinate charts of S^1 about the points $(0,-1)$ and $(0,1)$ respectively. Since every point (x,y) on S^1 is either in U_N or U_S, this provides another proof that S^1 is a one-dimensional manifold.

Exercise 11.12. Using the definition of a manifold directly prove that the 2-sphere $S^2 \subseteq \mathbb{R}^3$ is a C^∞-manifold of dimension 2. [Hint: one way (though certainly not the only way) to do this is to try stereographic projections (draw a straight line from the north pole to the $z = -1$ plane and assign the point that intersects the sphere the corresponding point that intersects the $z = 1$ plane, then do something analogous for the south pole and the $z = 1$ plane).]

Theorem 11.13. *Let $M \subseteq \mathbb{R}^k$ be a compact m-dimensional manifold. Then there exists a finite collection*

$$\left\{ U_i \underset{\psi_i}{\overset{\varphi_i}{\rightleftarrows}} V_i \right\}_{i \in \{1,\dots,N\}} \tag{11.14}$$

for some $N \in \mathbb{N}$ such that for every point $x \in M$, there exists an $i \in \{1,\dots,N\}$ such that $U_i \underset{\psi_i}{\overset{\varphi_i}{\rightleftarrows}} V_i$ is a coordinate system about the point x.

Proof. Since M is a manifold, for every x, there exists a coordinate system $U_x \underset{\psi_x}{\overset{\varphi_x}{\rightleftarrows}} V_x$ of M about x. Since M is compact and $\{U_x\}_{x \in M}$ is an open cover of M, there exists a finite subcover

$\{U_{x_i}\}_{x_i \in M,\, i \in \{1,\ldots,N\}}$ of M so that

$$\left\{ \begin{tikzcd} U_{x_i} \arrow[r, shift left, "\varphi_{x_i}"] & V_{x_i} \arrow[l, shift left, "\psi_{x_i}"] \end{tikzcd} \right\}_{i \in \{1,\ldots,N\}} \tag{11.15}$$

is a system of coordinates for M satisfying the required conditions. ■

Before we give more examples of manifolds, we provide another equivalent definition. The present definition can be found in [9] (Spivak has the special case of this Theorem for C^∞ manifolds [13]).

Theorem 11.16. *Fix $r \in \{0\} \cup \mathbb{N} \cup \{\infty\}$. A subset $M \subseteq \mathbb{R}^k$ is a C^r-manifold (without boundary) of dimension m if and only if for every point $c \in M$, there exists*

(a) an open neighborhood $U \subseteq \mathbb{R}^k$ of c,

(b) an open set $W \subseteq \mathbb{R}^m$, and

(c) a 1-1 function $\alpha : W \to \mathbb{R}^k$ of class C^r

such that

i) $\alpha(W) = M \cap U$,

ii) $D_y\alpha$ has rank m for each $y \in W$, and

iii) $\alpha^{-1} : \alpha(W) \to W$ is continuous.

The map α satisfying these conditions is called a parametrization *of M about the point c. The map α^{-1} is called a* coordinate patch *of M about the point c. Together, these two maps form a* coordinate system.

I personally initially found this theorem surprising because α^{-1} is only assumed to be continuous and not C^r. A slight modification of the proof in [13] shows that this theorem is also true for C^r manifolds.

Proof. See Spivak [13] for the C^∞ case and Munkres [9] for the more general C^r case. ■

Remark 11.17. My terminology is not the same as what is in the references. What I call a coordinate patch is an assignment that sends points on a manifold to points in Euclidean space of the dimension of that manifold. In other words, you *assign coordinates* to the points on the manifold. That's why it's called a coordinate patch. Spivak and Munkres don't seem to give this map a name. The inverse map associates points on your manifold from some coordinates in Euclidean space. This coincides with what you'd normally call a *parametrization* of your manifold. That's why I call it a parametrization. Spivak and Munkres call this a coordinate chart/patch/system. Instead, I reserve the terminology "coordinate system" to the combination of a coordinate patch and parametrization. As yet another example, think about a flat map of the Earth. This describes (as close as one can get to) a *coordinate chart* of Earth. We see the coordinates being written out on a paper map when we read a map (ignoring the distortions, which have to do with distance).

Yet another definition can be found in other sources. The following one is a slight modification of the one in Milnor's book [8] obtained by replacing C^∞ (aka "smooth") with C^r.

Theorem 11.18. *Fix $r \in \{0\} \cup \mathbb{N} \cup \{\infty\}$. A subset $M \subseteq \mathbb{R}^k$ is a C^r-manifold (without boundary) of dimension m if and only if for every point $c \in M$, there exists*

(a) an open neighborhood $V \subseteq \mathbb{R}^k$ of c,

(b) an open set $U \subseteq \mathbb{R}^m$, and

(c) C^r-diffeomorphism

$$U \underset{\psi}{\overset{\varphi}{\rightleftarrows}} V \cap M \tag{11.19}$$

The map $\varphi : U \to V \cap M$ is called a parametrization *of $V \cap M$. $\psi : V \cap M \to U$ is also called a* coordinate patch *on $V \cap M$. Both diffeomorphisms are called a* system of coordinates.

Exercise 11.20. Prove Theorem 11.18. [Hint: use Definition 11.1 and assume a C^r version of Exercise 4.20 is true.]

Exercise 11.21. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold of class C^r. Let $c \in M$ and let $\alpha : W \to \mathbb{R}^k$ and $\beta : V \to \mathbb{R}^k$ be two parametrizations of M about the point c as in Theorem 11.16 (so that $W, V \subseteq \mathbb{R}^m$ are both open). Let $Z := \alpha(W) \cap \beta(V) \subseteq \mathbb{R}^k$. Prove that the composition

$$\mathbb{R}^m \xleftarrow{\beta^{-1}} Z \xleftarrow{\alpha} \alpha^{-1}(Z) \tag{11.22}$$

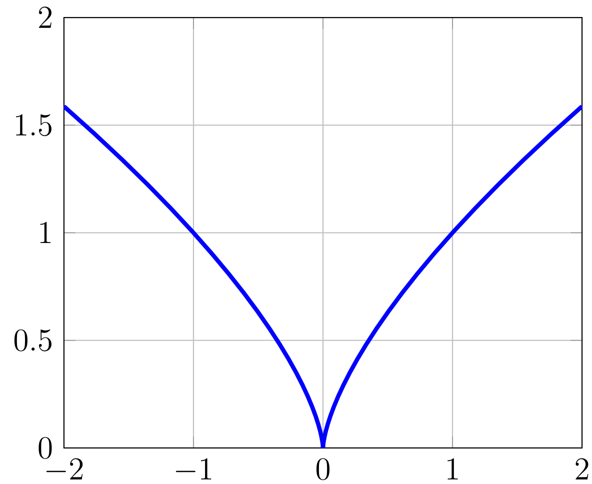
is of class C^r with associated differential having non-vanishing determinant. [Careful: Z is not necessarily an open subset of $\mathbb{R}^k$.]

Before providing examples of manifolds (which we have already actually seen), we provide several *non*-examples following Section 23 of [9] very closely. Note that by Exercise 11.20, showing any one of the conditions in Theorem 11.16 fail implies that the example cannot be a manifold.

In the notation of Theorem 11.16, what happens if the differential $D_y\alpha : \mathbb{R}^m \to \mathbb{R}^k$ associated to the map $\alpha : W \to M \cap U$ does not have rank m?

Example 11.23. Let M be the image of the curve $\alpha : \mathbb{R} \to \mathbb{R}^2$ defined by

$$\mathbb{R} \ni t \mapsto \alpha(t) := (t^3, t^2). \tag{11.24}$$



Intuitively, we see that there is a problem at the "cusp." What is the precise problem? α is C^∞ on $\mathbb{R}$ so there is no problem there. Furthermore, α^{-1} is continuous on M. To see this, first notice that

$$M := \big\{(x,y) \in \mathbb{R}^2 \;:\; x^2 = y^3\big\} = \big\{(x, x^{2/3}) \in \mathbb{R}^2 \;:\; x \in \mathbb{R}\big\} \tag{11.25}$$

Then, one can check that

$$M \ni (x,y) \mapsto \alpha^{-1}(x,y) := x^{1/3} \tag{11.26}$$

is the inverse. We know this is continuous (but not differentiable at 0). Finally, note that

$$[D_0\alpha] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{11.27}$$

and therefore has zero rank, which is less than 1.

Warning: We did not actually prove that M is not a C^∞ manifold here. All we did was show that for a *particular* candidate for a coordinate system near $(0,0)$ in M, this candidate does not satisfy the required conditions of a coordinate system. What we really have to show is that there is *no* coordinate patch around $(0,0)$ in M. However, the argument is similar to the one above. Suppose there exists a parametrization $\alpha : W \to U \cap M$ around $(0,0)$, but one for which no assumption on $D_{\alpha^{-1}(0,0)}\alpha$ is made, with $W \subseteq \mathbb{R}$ open and $U \subseteq \mathbb{R}^2$ an open set containing $(0,0)$. Without loss of generality, we may assume that $0 \in W$ and $\alpha(0) = (0,0)$ (by translating the domain U and the argument of the function α). To prove the claim, we must therefore show that $D_0\alpha$ must vanish. This is left as an exercise.

Exercise 11.28. Referring to the last paragraph in Example 11.23, prove that $D_0\alpha$ must vanish.

Exercise 11.29. Following similar reasoning to Example 11.23, complete the following tasks.

(a) Explain why the image of the curve $\beta : \mathbb{R} \to \mathbb{R}^2$ described by the equation

$$\mathbb{R} \ni t \mapsto \beta(t) := \big(t^3, |t|^3\big) \tag{11.30}$$

is not a C^1 manifold (plot it!).

(b) What about the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ described by the equation

$$\mathbb{R} \ni t \mapsto \beta(t) := \big(t, |t|\big)? \tag{11.31}$$

(c) Prove that the subset

$$S := \big(\{0\} \times [0,1]\big) \cup \big([0,1] \times \{1\}\big) \cup \big(\{1\} \times [0,1]\big) \cup \big([0,1] \times \{0\}\big) \tag{11.32}$$

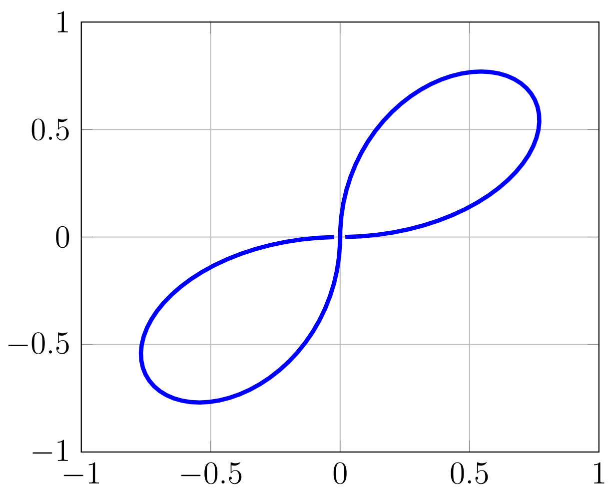
of $\mathbb{R}^2$ is not a C^1 manifold.

In the notation of Theorem 11.16, what happens if $\alpha^{-1} : M \cap U \to W$ is not continuous?

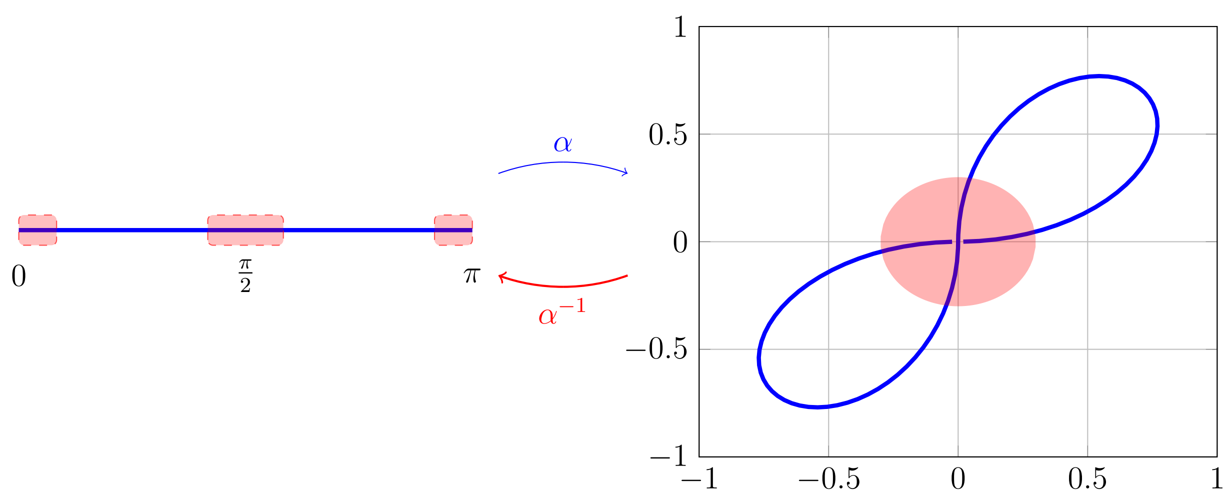
Example 11.33. Let $\alpha : (0, \pi) \to \mathbb{R}^2$ be the function defined by

$$(0,\pi) \mapsto \alpha(t) := \big(\sin(2t)|\cos(t)|, \sin(2t)\sin(t)\big) \tag{11.34}$$

and let M be the image of α (drawn below).



Let $c := (0,0) \in M$ and set $\epsilon := 0.3$. Note that $\alpha^{-1}(c) = \frac{\pi}{2}$. Then for every $\delta > 0$, $\alpha^{-1}\big(V_\delta(c)\big)$ is not contained in $V_\epsilon\left(\frac{\pi}{2}\right) \subseteq (0, \pi)$ as the following picture illustrates



On the right, the shaded red region is $V_\delta(c)$. On the left, the image of $V_\delta(c)$ under the map α^{-1} splits: part of it is concentrated near $\frac{\pi}{2}$ but part of it also goes near 0 and π. Hence, if ϵ is small, $V_\epsilon\left(\frac{\pi}{2}\right)$ cannot contain these left and right parts. Therefore, α^{-1} is not continuous.

In the notation of Theorem 11.16, what happens if $\alpha : W \to M \cap U$ is not C^r? Although many examples can be given to show why M would not be a manifold, Exercise 5.6 from Spivak [13] addresses this point in great generality. Here is one example.

Example 11.35. Let $\alpha : \mathbb{R} \to \mathbb{R}^2$ be the function

$$\mathbb{R} \ni x \mapsto \alpha(x) := (x, |x|), \tag{11.36}$$

i.e. the graph of the function $f : \mathbb{R} \to \mathbb{R}$ given by $x \in \mathbb{R}$ being sent to $f(x) := |x|$. The function f is not differentiable, nor is it 1-1 so that it does not have an inverses. However, α satisfies all of the required conditions except that it is not differentiable at 0. α does have an inverse, which is just the projection function $\mathbb{R}^2 \to \mathbb{R}$ onto the first factor. This function is certainly continuous, even on the image $\alpha(\mathbb{R})$, which is the graph of f. Note that α satisfies the rank condition vacuously since the derivative is not defined at 0 and it has rank 1 at all other points.

The previous non-examples illustrate how every condition in Theorem 11.16 is necessary to be rightfully called a manifold. What are some other examples of manifolds? Many examples of smooth manifolds are often obtained from implicit functions [8].

Definition 11.37. Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \to \mathbb{R}^m$ be a differentiable function. A *critical point* of f is a point $a \in U$ such that $D_a f : \mathbb{R}^n \to \mathbb{R}^m$ is not onto (has rank strictly less than m). Let C_f denote the critical points of f. A *regular value* of $f : U \to \mathbb{R}^m$ is a point in $\mathbb{R}^m \setminus f(C_f)$, the complement of the image of all the critical points of f. This set is denoted by $R_f := \mathbb{R}^m \setminus f(C_f)$. A *regular point* of $f : U \to \mathbb{R}^m$ is an element of $f^{-1}(R_f)$.

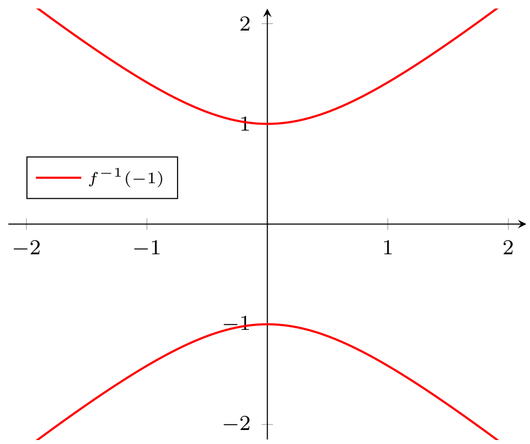
It is unfortunate that the phrase "critical point" is used also in the context of vector fields in a different way than this phrase. We will try to be clear to distinguish depending on the context by saying "critical point of the function..." or "critical point of the vector field..." to be clear.

Example 11.38. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

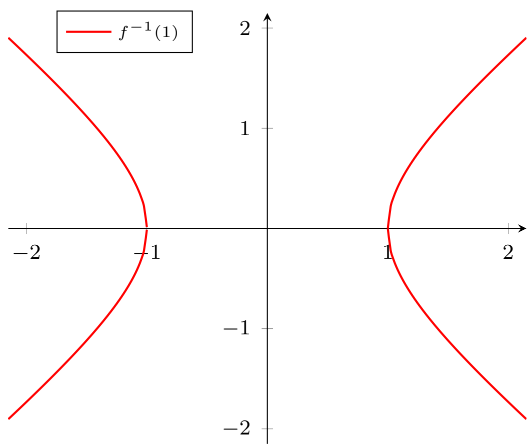
$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2 - y^2 \tag{11.39}$$

and let $\ell \in \mathbb{R}$. Before studying the critical points, regular values, and regular points, we consider the level sets for various values of ℓ.

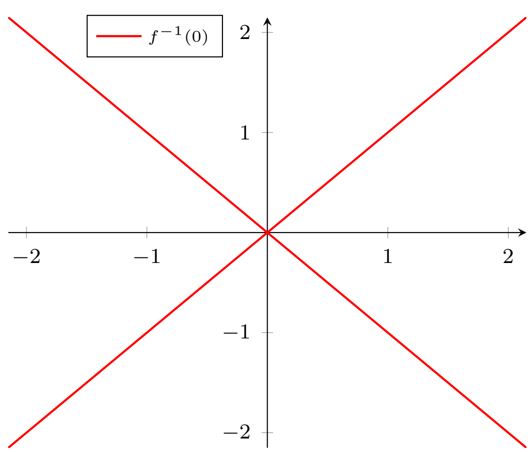
(a) If $\ell < 0$, then $f^{-1}(\ell)$ looks something like the subset of $\mathbb{R}^2$ shown in the following figure



(b) If $\ell > 0$, then $f^{-1}(\ell)$ looks something like the subset of $\mathbb{R}^2$ shown in the following figure



(c) If $\ell = 0$, then $f^{-1}(\ell)$ looks something like the subset of $\mathbb{R}^2$ shown in the following figure



The matrix associated to the differential of f at $(a, b) \in \mathbb{R}^2$ is given by

$$[D_{(a,b)}f] = \begin{bmatrix} 2a & -2b \end{bmatrix}. \tag{11.40}$$

This has rank 1 for all (a, b) except at $(0, 0)$. Hence

$$C_f = \{0\}. \tag{11.41}$$

The set of regular values is therefore

$$R_f = \mathbb{R} \setminus \{0\}. \tag{11.42}$$

Finally, since

$$f^{-1}(0) = \Gamma(\mathrm{id}_{\mathbb{R}}) \cup \Gamma(-\mathrm{id}_{\mathbb{R}}) \tag{11.43}$$

(recall, Γ of a function is the graph of that function), the set of regular points is

$$\mathbb{R}^2 \setminus \big(\Gamma(\mathrm{id}_{\mathbb{R}}) \cup \Gamma(-\mathrm{id}_{\mathbb{R}})\big). \tag{11.44}$$

One of the important observations in the previous theorem is that the inverse image of a regular value is a manifold!

Theorem 11.45. *Fix $n, m \in \mathbb{N} \cup \{0\}$ with $n \geq m$. Let $U \subseteq \mathbb{R}^n$ be an open set, let $f : U \to \mathbb{R}^m$ be a smooth function, and let $c \in \mathbb{R}^m$ be a regular value. Then $f^{-1}(c) \subseteq \mathbb{R}^n$ is a smooth manifold of dimension $n - m$.*

We will prove a much more general version of this theorem soon! But for now, we point out that by redefining the function f by shifting by ℓ, namely setting $g := f - \ell$, then this describes manifolds as subsets whose points z get mapped to 0 under g, i.e. $g(z) = 0$. By splitting up the input coordinates $z = (x, y)$, we can then use the Implicit Function Theorem, when it holds, to describe the manifold locally in terms of charts. According to Spivak, Theorem 11.45 can be proved "immediately" using the following fact [13].

Theorem 11.46. *Let $a \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be of class C^1 in some open set $U \subseteq \mathbb{R}^n$ containing a with $m \leq n$. Suppose that $f(a) = 0$ and $D_a f$ has rank m. Then there exists an open set $A \subseteq \mathbb{R}^n$ containing a and a diffeomorphism $h : A \to \mathbb{R}^n$ such that*

$$(f \circ h)(x_1, \ldots, x_n) = (x_{n-m+1}, \ldots, x_n) \qquad \forall\, (x_1, \ldots, x_n) \in A, \tag{11.47}$$

i.e. the diagram

$$\begin{tikzcd} \mathbb{R}^n \arrow[d, "f"'] & A \arrow[l, "h"'] \arrow[d, hook] \\ \mathbb{R}^m & \mathbb{R}^{n-m} \times \mathbb{R}^m \arrow[l, "\pi_{\mathbb{R}^m}"] \end{tikzcd} \tag{11.48}$$

commutes.

Proof. See Theorem 2-13 in [13]. ■

Exercise 11.49. Use Theorem 11.46 to prove Theorem 11.45.

Exercise 11.50. Prove that the n-sphere

$$S^n = \big\{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \ :\ x_1^2 + \cdots + x_{n+1}^2 = 1\big\} \tag{11.51}$$

is a manifold by showing that 0 is a regular value for the function $f : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$\mathbb{R}^{n+1} \ni (x_1, \ldots, x_{n+1}) \mapsto x_1^2 + \cdots + x_{n+1}^2 - 1. \tag{11.52}$$

Exercise 11.53. For the present example, let $\ell \in \mathbb{R}$ and set $f_\ell : \mathbb{R}^3 \to \mathbb{R}$ to be the function

$$\mathbb{R}^3 \ni (x, y, z) \mapsto f_\ell(x, y, z) := x^2 + y^2 - z^2 - \ell. \tag{11.54}$$

(a) For $\ell > 0$, prove that 0 is a regular value so that $f_\ell^{-1}(0)$ is a two-dimensional manifold in $\mathbb{R}^3$. This manifold is known as a *one-sheeted hyperboloid*.

(b) For $\ell < 0$, prove that 0 is a regular value so that $f_\ell^{-1}(0)$ is a two-dimensional manifold in $\mathbb{R}^3$. This manifold is known as a *two-sheeted hyperboloid*.

(c) If $\ell = 0$, is $f_\ell^{-1}(0)$ a manifold? Prove your claim.

Exercise 11.55. Let $R > r > 0$. Prove that for the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x, y, z) = \left(x^2 + y^2 + z^2 + R^2 - r^2\right)^2 - 4R^2\left(x^2 + y^2\right), \tag{11.56}$$

0 is a regular value. If necessary (for ease of calculation), set $R = 3$ and $r = 1$. This shows that the set $f^{-1}(0)$ in Figure 1 on page 72 of Lecture 7 is a two-dimensional manifold in $\mathbb{R}^3$. Since f is C^∞ in a neighborhood of 0, the manifold is of class C^∞. This manifold is known as a two-dimensional *torus*.

Exercise 11.57. Referring to the function $f : \mathbb{R}^3 \to \mathbb{R}$ from Exercise 11.55 setting $R = 3$ and $r = 1$, find a set of neighborhoods $\{U_1, U_2, \dots, U_n\}$ in $\mathbb{R}^2$ (thought of as open subsets of different two-dimensional planes in $\mathbb{R}^3$) and a set of one-to-one functions $g_1 : U_1 \to \mathbb{R}, \dots, g_n : U_n \to \mathbb{R}$ of class C^∞ whose graphs describe the torus $f^{-1}(0)$. In other words, prove that $f^{-1}(0)$ is a two-dimensional manifold directly from the Definition of a manifold. [Hint: find a suitable "parametrization" of $f^{-1}(0)$ that "covers" $f^{-1}(0)$ several times.]

Remark 11.58. Although our definitions of a smooth manifold demand that they be subsets of Euclidean space, a more abstract definition exists. However, one requires some notions from topology to describe it which would lead us on a somewhat technical detour (such a definition may be presented in a course on differential geometry). It is a theorem of Whitney that every such abstract smooth manifold (let's assume it is compact for now—I'm not sure how general the actual result is) can be embedded in Euclidean space. As a side note, being a subset of Euclidean space, every manifold inherits a notion of distance so that it becomes a metric space. This is obtained by using the shortest path between any two points on a manifold. An infinitesimal version of this notion is called a metric. One can also abstract this definition to define what is known as a Riemannian manifold, not as a subset of Euclidean space, but as an abstract space. It is a theorem of Nash that every such abstract Riemannian manifold (well, as long as it is compact—again, I'm not sure exactly how general the result is) can be embedded in Euclidean space and its metric is induced by the metric in that Euclidean space.

Exercise 11.59. Let $M \subseteq \mathbb{R}^k$ be a smooth m-dimensional manifold, let $c \in M$, and let $f : M \to \mathbb{R}^n$ be a function. Show that f is differentiable at $c \in M$ if and only if for any parametrization $\alpha : W \to M \cap U$ of M about the point c, the composition $f \circ \alpha : W \to \mathbb{R}^n$ is differentiable at $\alpha^{-1}(c)$.

As promised, we provide some example of manifolds that are differentiable but not C^1. One such example is the graph of the function

$$\mathbb{R} \ni x \mapsto \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases} \tag{11.60}$$

Similarly, there are manifolds that are n times differentiable but not of class C^n.

Exercise 11.61. Provide an example of a manifold that is n times differentiable but not of class C^n.

Here are some additional problems, some of which come from, or are motivated by, [9].

Exercise 11.62. Let

$$\mathrm{GL}_m := \big\{ A : \mathbb{R}^m \to \mathbb{R}^m \ : \ A \text{ linear and invertible} \big\} \tag{11.63}$$

be the set of invertible $m \times m$ (real) matrices.

(a) Prove GL_m is an m^2-dimensional manifold in $\mathbb{R}^{m^2}$. [Hint: consider the function $\det : \mathbb{R}^{m^2} \to \mathbb{R}$.]

(b) Prove that the product (composition) of matrices

$$\begin{aligned} \mathrm{GL}_m \times \mathrm{GL}_m &\to \mathrm{GL}_m \\ (A, B) &\mapsto AB \end{aligned} \tag{11.64}$$

is smooth.

(c) Prove that the inversion for invertible matrices

$$\begin{aligned} \mathrm{GL}_m &\to \mathrm{GL}_m \\ A &\mapsto A^{-1} \end{aligned} \tag{11.65}$$

is smooth.

The previous exercise shows that GL_m is a Lie group.

Definition 11.66. A *Lie Group* is a smooth manifold G (that is a subset of some Euclidean space) together with

(a) a function $e : \{*\} \to G$ called the *identity element*,[18]

(b) a smooth function $\mu : G \times G \to G$ known as the *product*, and

(c) a smooth function $\mathrm{inv} : G \to G$ known as the *inversion*

[18] A function from a single element set to another set picks a single element in the latter set.

such that the following diagrams commute[19]

$$\begin{tikzcd}
G \times G \times G \arrow[r, "\mu\times\mathrm{id}_G"] \arrow[d, "\mathrm{id}_G\times\mu"'] & G \times G \arrow[d, "\mu"] \\
G \times G \arrow[r, "\mu"'] & G
\end{tikzcd}, \tag{11.67}$$

$$\begin{tikzcd}
\{*\} \times G \arrow[r, "e\times\mathrm{id}_G"] \arrow[dr, "\pi_G"'] & G \times G \arrow[d, "\mu"'] & G \times \{*\} \arrow[l, "\mathrm{id}_G\times e"'] \arrow[dl, "\pi_G"] \\
& G &
\end{tikzcd} \tag{11.68}$$

$$\begin{tikzcd}
& G \times G \arrow[rr, "\mathrm{id}_G\times\mathrm{inv}"] & & G \times G \arrow[dr, "\mu"] & \\
G \arrow[ur, "\Delta"] \arrow[drr, "!"'] & & & & G \\
& & \{*\} \arrow[urr, "e"'] & &
\end{tikzcd}, \tag{11.69}$$

and another diagram just like (11.69) but with the top arrow $\mathrm{id}_G \times \mathrm{inv}$ replaced by $\mathrm{inv} \times \mathrm{id}_G$. In these diagrams, $! : G \to \{*\}$ is the unique function that sends every element of G to the single element $*$ in $\{*\}$ and $\Delta : G \to G \times G$ is the duplication function that sends an element $g \in G$ to (g, g). Since $\{*\}$ is the single element set, it is common to denote $e(*) \in G$ as simply e.

Exercise 11.70. Explain concretely what the three diagrams in the definition of a Lie group mean in terms of group elements.

The next two exercises come directly out of [9].

Exercise 11.71. Let $O(3)$ denote the set of all orthogonal 3×3 matrices considered as a subspace of $\mathbb{R}^9$.

(a) Define a function $f : \mathbb{R}^9 \to \mathbb{R}^6$ such that $O(3)$ is the solution set of the equation $f(x) = 0$.

(b) Show that $O(3)$ is a compact three-dimensional manifold in $\mathbb{R}^9$. [Hint: Show that the rows of $D_x f$ are independent if $x \in O(3)$.]

(c) Prove that $O(3)$ is a Lie Group with the induced structure from GL_3. [Hint: Use the results of Exercise 11.62.]

Exercise 11.72. Fix $n \in \mathbb{N}$. Let $O(n)$ denote the set of all orthogonal $n \times n$ matrices considered as a subspace of $\mathbb{R}^{n^2}$. Prove that $O(n)$ is a compact manifold and is in fact a Lie group with the induced structure from GL_n.

Here are some additional exercises.

[19] You should convince yourself that these are the axioms of a group! See Exercise 11.70.

Exercise 11.73. The present exercise gives some background for Theorem 11.45. Notice how the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x,y) := (x,0)$ has Jacobian whose associated rank is 1. Furthermore, $f(\mathbb{R}^2)$ is a 1-dimensional manifold. The relationship between these the rank of a Jacobian is made precise by Theorem 11.45 as well as the Inverse Function Theorem. Occasionally, the rank of the Jacobian of a differentiable function indicates something about the dimension of the image of that function, and sometimes it doesn't. One should be cautious about making generalizations.

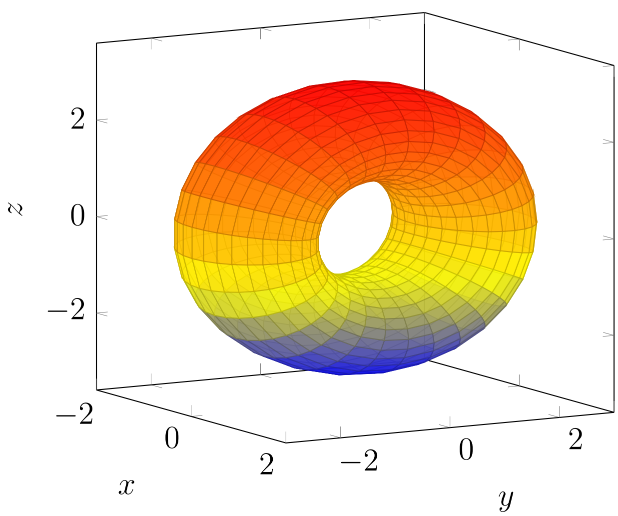
(a) Draw the image of the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $g(x,y) := (x, x^2)$. Calculate the rank of $D_{(x,y)}f$ for all $(x,y) \in \mathbb{R}^2$. Finally, find $f^{-1}(x,x^2)$ as a subset of $\mathbb{R}^2$ for some fixed $x \in \mathbb{R}$.

(b) Give an example of a function $f : \mathbb{R} \to \mathbb{R}^2$ such that $D_x f$ has rank 1 for all $x \in \mathbb{R}$ but for which $f(\mathbb{R})$ is not a manifold. [Hint: a small modification of a function that exists in this lecture can be used.]

Exercise 11.74. Let $X : [0,2\pi) \times [0,2\pi) \to \mathbb{R}^3$ be defined by the following component functions sending $(s,t) \in [0,2\pi) \times [0,2\pi)$ to

$$\begin{aligned} X_1(s,t) &:= \sin(t) \\ X_2(s,t) &:= \big(2+\cos(t)\big)\cos(s) \\ X_3(s,t) &:= \big(2+\cos(t)\big)\sin(s). \end{aligned} \tag{11.75}$$

Let M be the image of X.



(a) Let $r \in \mathbb{Q}$ be a rational number and let $\gamma_r : \mathbb{R} \to \mathbb{R}^3$ be the curve defined by

$$\mathbb{R} \ni s \mapsto \gamma_r(s) := \Big(\sin(rs), \big(2+\cos(rs)\big)\cos(s), \big(2+\cos(rs)\big), \sin(s)\Big). \tag{11.76}$$

Let $L_r := \gamma_r(\mathbb{R})$ be the image of this curve. Prove that L_r is a one-dimensional manifold. [Hint: Show that $L_r \subseteq M$ and find out what $X^{-1}(L_r)$ looks like as a subset of $[0,2\pi) \times [0,2\pi)$ for a few values of r such as $r = \frac{1}{3}, \frac{1}{2}, 1, 2, 3, \ldots$ perhaps using Mathematica.]

(b) Let $\lambda \in \mathbb{I}$ be an irrational number and let $\gamma_\lambda : \mathbb{R} \to \mathbb{R}^3$ be the curve defined by

$$\mathbb{R} \ni s \mapsto \gamma_\lambda(s) := \Big(\sin(rs), \big(2 + \cos(rs)\big)\cos(s), \big(2 + \cos(rs)\big), \sin(s) \Big). \tag{11.77}$$

Let $L_\lambda := \gamma_\lambda(\mathbb{R})$ be the image of this curve. Prove that L_λ is *not* a manifold. [Hint: Show that $L_r \subseteq M$ and find out what $X^{-1}(L_\lambda)$ looks like as a subset of $[0, 2\pi) \times [0, 2\pi)$ for $\lambda = e$ (an irrational number) perhaps using Mathematica.]

Level I problems.
From Spivak [13]: Exercises 5.5, 5.6,
From these notes: Exercises 11.12, 11.20, 11.21, 11.28, 11.50, 11.70, 11.73

Level II problems.
From Spivak [13]: Exercises 5.4, 5.7
From these notes: Exercises 11.8, 11.29, 11.49, 11.53, 11.55, 11.57, 11.59, 11.61, 11.62, 11.71, 11.72, 11.74

12 March 21: Tangent Spaces

This material has been moved to the beginning of Week #08 so that we may spend more time on manifolds during Week #07.

We introduce the concept of a tangent space to a manifold at a point. The tangent space to an m-dimensional manifold $M \subseteq \mathbb{R}^k$ at a point $c \in M$ is the hyperplane tangent to M at the point c viewed as a vector space whose origin coincides with the point c. Furthermore, the dimension of this hyperplane agrees with the dimension of the manifold. This notion would not make sense for an arbitrary subset of $\mathbb{R}^k$. For instance, referring back to Example 11.23, what should the tangent space be at the point $(0,0)$? At best, it could be a vertical line, namely the y axis, but that doesn't quite work. As another example, consider the point $(0,0)$ in Example 11.33. It looks like there are two possibilities for the tangent line. In either case, a unique tangent line doesn't make sense. To make sense of the tangent space for a *manifold*, we first define it for an open subset $V \subseteq \mathbb{R}^k$.

Definition 12.1. Let $V \subseteq \mathbb{R}^k$ be open and let $c \in V$. The *tangent space to V at c* is the set

$$T_cV := \{c\} \times \mathbb{R}^k, \tag{12.2}$$

whose elements are written as[20] $(c;v)$ or more frequently as v_c, equipped with the following vector space structure.

(a) The zero vector is 0_c (i.e. $(c,0)$).

(b) The sum of two vectors $u_c, v_c \in T_cV$ is defined to be

$$u_c + v_c := (u+v)_c \tag{12.3}$$

(i.e. $(c;u) + (c;v) := (c;u+v)$).

(c) The scalar multiplication of a real number $\lambda \in \mathbb{R}$ with a vector $v_c \in T_cV$ is defined to be

$$\lambda v_c := (\lambda v)_c, \tag{12.4}$$

(i.e. $\lambda(c;v) := (c;\lambda v)$).

This is just formalizing something we have already been very familiar with since we first learned about the differential. Note that the point c just labels the vector and does not alter the algebraic structure in any way. In other words, the only thing we have done is that we are now keeping track of the point c in our notation explicitly. As usual, we will think of T_cV as $\mathbb{R}^k$ itself but whose origin is at the point c and whose vectors emanate out of c. For example, if $f: V \to \mathbb{R}^m$ is a differentiable function, then the differential $D_cf : \mathbb{R}^k \to \mathbb{R}^m$ will now be written more precisely as $D_cf : T_cV \to T_{f(c)}\mathbb{R}^m$ so that our notation keeps track of specifically where our vector space origin is. However, the way we compute this differential does not change at all. Also, if $V \subseteq \mathbb{R}^k$ is open and $c \in V$, then $T_cV = T_c\mathbb{R}^k$ (check this!). But to avoid clutter, we may sometimes still write $\mathbb{R}^k$ instead of T_cV or $T_c\mathbb{R}^k$ whenever convenient. Furthermore, the notation now stresses the following picture

[20] A semi-colon has been used instead of a comma to avoid confusion with usual ordered pairs since the algebraic structure will essentially ignore c completely.

$$\boxed{\text{finite-dimensional vector spaces}} \longleftarrow \boxed{\text{open subsets of Euclidean space with a basepoint}}$$

$$T_cV \longmapsfrom (V,c)$$

for what happens to domains and

$$\boxed{\text{linear transformations}} \longleftarrow \boxed{\text{basepoint-preserving differentiable functions}}$$

$$\Big(T_dU \xleftarrow{D_cf} T_cV\Big) \longmapsfrom \Big((U,d) \xleftarrow{f} (V,c)\Big)$$

for what happens to differentiable functions. The fact that the differentiable functions are basepoint-preserving just means that $f(c) = d$ in the notation above. This, in particular, reminds us of why the domain changes when we take the differential of a function. We're no longer looking at functions on open subsets of Euclidean space—we are looking at linear transformations of vector spaces! We started off with a function $f : V \to U$ between open sets that is differentiable at c. When we take the differential $D_cf : T_cV \to T_{f(c)}U$, this linear transformation is defined for *all* vectors in T_cV, regardless of the domain on which f was defined (as long as it is open). Phrased another way, the direction derivative of a function makes sense in any direction for any magnitude of a vector that is plugged in. Keeping this subtle distinction in mind will be crucial for what follows.

We now provide a definition of the tangent space for a manifold M at an arbitrary point $c \in M$.

Definition 12.5. Let $M \subseteq \mathbb{R}^k$ be a smooth m-dimensional manifold, let $c \in M$, and let $U \subseteq \mathbb{R}^m$ be an open set, let $V \subseteq \mathbb{R}^k$ be an open set with $c \in V$, and let $\varphi : U \to M \cap V$ and $\psi : M \cap V \to U$ be a diffeomorphism pair with φ a parametrization and ψ a coordinate patch about c. The *tangent space* of M at c is the image of $T_{\psi(c)}U$ (which is basically $\mathbb{R}^m$) in $T_c\mathbb{R}^k$ (which is basically $\mathbb{R}^k$) under the linear transformation $D_{\psi(c)}\varphi : T_cV \to T_{f(c)}\mathbb{R}^k$. The tangent space of M at c is denoted by T_cM,

$$T_cM := \big(D_{\psi(c)}\varphi\big)(T_{\psi(c)}U) \equiv \big\{(D_{\psi(c)}\varphi)(v_{\psi(c)}) \ : \ v_{\psi(c)} \in T_{\psi(c)}U\big\}. \tag{12.6}$$

A *tangent vector* at c in M is an element of T_cM.

We can use Definition 12.1 to provide T_cM with a vector space structure by using the vector space structure of $T_c\mathbb{R}^k$ by viewing T_cM as a subspace of $T_c\mathbb{R}^k$ (the image of a linear transformation is a subspace of the codomain of that linear transformation).

Before proving that T_cM is independent of the coordinate patch chosen around c, we go through an explicit example to calculate the tangent space.

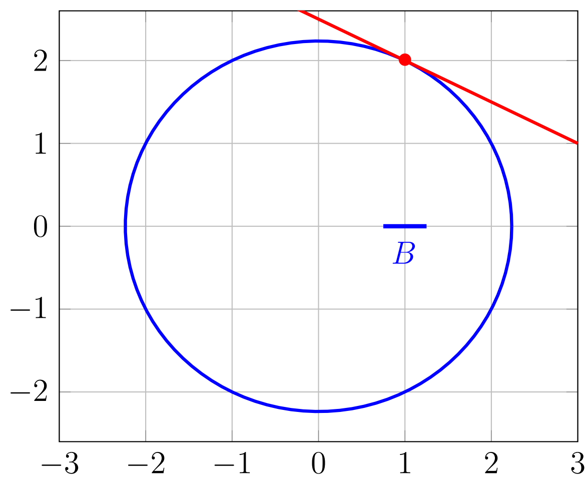
Example 12.7. Let M be the circle of radius $\sqrt{5}$ centered at the origin and let $c = (1,2)$ as in Example 10.48. In this example, we found a function $g : B \to \mathbb{R}$ whose graph parametrizes a part of S^1 containing c. Namely, let $\alpha : B \to \mathbb{R}^2$ be given by $\alpha(x) := \big(x, g(x)\big)$. One of the homework exercises from the previous lecture asks to prove that the graph of a differentiable function is a manifold so we know that $\alpha(B)$ is a 1-dimensional manifold. Let us calculate the differential of α and use it to calculate the tangent space $T_{(1,2)}S^1$. The formula for g was given by $g(x) = \sqrt{5-x^2}$ for $x \in B$. Hence, the matrix associated to the differential of α is the 2×1 matrix given by

$$[D_x\alpha] = \begin{bmatrix} 1 \\ \frac{-1}{\sqrt{5-x^2}} \end{bmatrix} \tag{12.8}$$

Hence, the unit vector $e_1 \in \mathbb{R}$ (this is just the number 1) gets mapped to

$$(D_x\alpha)(e_1) = e_1 - \frac{1}{\sqrt{5-x^2}} e_2. \tag{12.9}$$

Therefore, the tangent space T_cS^1 is depicted (in red) in the following figure and is the span of this vector in $T_c\mathbb{R}^2$.



Be careful: keep in mind that the origin of $T_c\mathbb{R}^2$, the codomain of $D_x\alpha$, here is viewed as a vector space whose origin is at the point c. We have mentioned in Lecture 6 and above that the differential should be viewed as a linear transformation between vector spaces whose origins correspond to the point at which we calculate the differentials.

Exercise 12.10. Let $c = (0,1,2) \in \mathbb{R}^3$ and let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function

$$\mathbb{R}^3 \ni (x,y,z) \mapsto f(x,y,z) := x^2 + y^2 + z^2 - 5. \tag{12.11}$$

Let $M := f^{-1}(0)$. Find T_cM and provide a basis for it. [I especially recommend this problem for those who did Exercise 10.78.]

Exercise 12.12. Let $c = (0,1,1) \in \mathbb{R}^3$ and let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function

$$\mathbb{R}^3 \ni (x,y,z) \mapsto f(x,y,z) := x^3 + y^2 - z^2. \tag{12.13}$$

Let $M := f^{-1}(0)$. Find T_cM and provide a basis for it. [I especially recommend this problem for those who did Exercise 10.80.]

Exercise 12.14. Let $c = (0,3,1) \in \mathbb{R}^3$ and let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function

$$\mathbb{R}^3 \ni (x,y,z) \mapsto f(x,y,z) := (x^2+y^2+z^2+8)^2 - 36(x^2+y^2). \tag{12.15}$$

Let $M := f^{-1}(0)$. Find T_cM and provide a basis for it. [I especially recommend this problem for those who did Exercise 10.82.]

A few more alternative and increasingly abstract definitions of T_cM will be given in the exercises throughout this lecture. One important point to make is that the definition of T_cM is a-priori ambiguous because one could have chosen a different diffeomorphism pair of a neighborhood of c in M [8]. In fact, T_cM is independent of such a diffeomorphism pair as the following result shows.

Theorem 12.16. *Let $M \subseteq \mathbb{R}^k$ be a smooth m-dimensional manifold, let $c \in M$, let $U_1, U_2 \subseteq \mathbb{R}^m$ be open sets, let $V_1, V_2 \subseteq \mathbb{R}^k$ be open sets with $c \in V_1 \cap V_2$, and let*

$$U_1 \underset{\psi_1}{\overset{\varphi_1}{\rightleftarrows}} M \cap V_1 \qquad \& \qquad U_2 \underset{\psi_2}{\overset{\varphi_2}{\rightleftarrows}} M \cap V_2 \tag{12.17}$$

be two diffeomorphism pairs. Then

$$\mathrm{Image}\big(D_{\psi_1(c)}\varphi_1\big) = \mathrm{Image}\big(D_{\psi_2(c)}\varphi_2\big). \tag{12.18}$$

Proof. Set $V := V_1 \cap V_2$, $U_1' := \psi_1(M \cap V)$, and $U_2' := \psi_2(M \cap V)$. Note that U_1' and U_2' are open subsets of $\mathbb{R}^m$ due to the proof of the Inverse Function Theorem. Then, the diagrams

$$\begin{tikzcd} & \mathbb{R}^k & \\ U_1' \arrow[ur, "\varphi_1"] \arrow[rr, "\psi_2\circ\varphi_1"'] & & U_2' \arrow[ul, "\varphi_2"'] \end{tikzcd} \qquad \& \qquad \begin{tikzcd} & \mathbb{R}^k & \\ U_1' \arrow[ur, "\varphi_1"] & & U_2' \arrow[ul, "\varphi_2"'] \arrow[ll, "\psi_1\circ\varphi_2"] \end{tikzcd} \tag{12.19}$$

of differentiable functions (on open domains) both commute. By the chain rule, the diagrams

$$\begin{tikzcd} & \mathbb{R}^k & \\ \mathbb{R}^m \arrow[ur, "D_{\psi_1(c)}\varphi_1"] \arrow[rr, "D_{\psi_1(c)}(\psi_2\circ\varphi_1)"'] & & \mathbb{R}^m \arrow[ul, "D_{\psi_2(c)}\varphi_2"'] \end{tikzcd} \qquad \& \qquad \begin{tikzcd} & \mathbb{R}^k & \\ \mathbb{R}^m \arrow[ur, "D_{\psi_1(c)}\varphi_1"] & & \mathbb{R}^m \arrow[ul, "D_{\psi_2(c)}\varphi_2"'] \arrow[ll, "D_{\psi_2(c)}(\psi_1\circ\varphi_2)"] \end{tikzcd} \tag{12.20}$$

therefore both commute. The diagrams show that

$$\mathrm{Image}\big(D_{\psi_1(c)}\varphi_1\big) \subseteq \mathrm{Image}\big(D_{\psi_2(c)}\varphi_2\big) \qquad \& \qquad \mathrm{Image}\big(D_{\psi_2(c)}\varphi_2\big) \subseteq \mathrm{Image}\big(D_{\psi_1(c)}\varphi_1\big), \tag{12.21}$$

respectively. For example, to see the left containment, let $v \in \mathrm{Image}\big(D_{\psi_1(c)}\varphi_1\big)$. Then, there exists a $u \in \mathbb{R}^m$ such that $(D_{\psi_1(c)}\varphi_1)(u) = v$. Then, by commutativity of the first diagram,

$$\Big(D_{\psi_2(c)}\varphi_2\Big)\Big(\big(D_{\psi_1(c)}(\psi_2 \circ \varphi_1)\big)(u)\Big) = (D_{\psi_1(c)}\varphi_1)(u) = v \tag{12.22}$$

showing that $v \in \mathrm{Image}\big(D_{\psi_2(c)}\varphi_2\big)$. A similar argument using the other commutative diagram shows the other containment. ∎

Theorem 12.23. *Let $M \subseteq \mathbb{R}^k$ be a smooth m-dimensional manifold and let $c \in M$. Then T_cM is an m-dimensional vector space.*

Proof. The fact that T_cM is a vector space follows from its definition (Definition 12.5) since the image of any linear transformation is always a vector space. To see that it is m-dimensional, let $\varphi : U \to M \cap V$ be a parametrization of M about c. Then $D_{\varphi^{-1}(c)}\varphi : \mathbb{R}^m \to \mathbb{R}^k$ has rank m by assumption. This implies (by the Invertible Matrix Theorem from linear algebra) that $\dim T_cM = m$. ■

There is an alternative definition of the tangent space of a manifold at a point using equivalence classes of curves. This definition is useful for abstract manifolds that are not assumed to be embedded in some Euclidean space.

Exercise 12.24. Let $M \subseteq \mathbb{R}^k$ be a smooth m-dimensional manifold and let $c \in M$. Let $C^\infty\big((M,c),(\mathbb{R},0)\big)$ be the set of smooth functions from $\gamma : \mathbb{R} \to \mathbb{R}^k$ such that $\gamma(t) \in M$ for all $t \in \mathbb{R}$ and $\gamma(0) = c$, i.e.

$$C^\infty\big((M,c),(\mathbb{R},0)\big) := \big\{\gamma : \mathbb{R} \to M \ : \ \gamma \in C^\infty(\mathbb{R},M) \text{ and } \gamma(0) = c\big\} \tag{12.25}$$

(the reason for the complicated notation in $C^\infty\big((M,c),(\mathbb{R},0)\big)$ is to keep track of all the data—the functions considered are basepoint preserving). The claim of this exercise is that

$$\operatorname{span}\Big\{\gamma'(0) \in T_cM \ : \ \gamma \in C^\infty\big((M,c),(\mathbb{R},0)\big)\Big\} = T_cM. \tag{12.26}$$

However, the set $C^\infty\big((M,c),(\mathbb{R},0)\big)$ has infinitely many elements so this span is difficult to calculate. In fact, all we care about is $\gamma'(0)$ and not the whole curve γ. This motivates defining an equivalence relation on $C^\infty\big((M,c),(\mathbb{R},0)\big)$ that identifies such curves whose derivatives at 0 agree. Define this equivalence relation on $C^\infty\big((M,c),(\mathbb{R},0)\big)$ by demanding that $\gamma \sim \delta$ if and only if

$$\gamma'(0) = \delta'(0). \tag{12.27}$$

The goal of this exercise then is to show that the sets $C^\infty\big((M,c),(\mathbb{R},0)\big)/_\sim$ and T_cM are isomorphic. Do this by constructing functions in both directions in the following manner.

(a) First show that for any curve $\gamma \in C^\infty\big((M,c),(\mathbb{R},0)\big)$, the vector $(D_0\gamma)(1)$ (the image of the vector $1 \in \mathbb{R}$ under the differential $D_0\gamma : \mathbb{R} \to \mathbb{R}^k$), which is a-priori just a vector in $\mathbb{R}^k$, is actually a vector in T_cM. Also explain why this vector is independent of the representative curve γ chosen. Note that this defines a function $C^\infty\big((M,c),(\mathbb{R},0)\big)/_\sim \to T_cM$.

(b) Show that for any vector $v \in T_cM$, there exists a curve $\gamma : \mathbb{R} \to M$ such that $\gamma(0) = c$ and $\gamma'(0) = v$ (if you can only construct this curve on some domain $(-\epsilon,\epsilon) \subseteq \mathbb{R}$ instead of all of $\mathbb{R}$, that suffices because by redefining your ϵ, you can make it so that this curve is constant outside of the domain $(-2\epsilon, 2\epsilon)$). Then take the equivalence class associated to this curve. This defines a function $T_cM \to C^\infty\big((M,c),(\mathbb{R},0)\big)/_\sim$.

(c) Show that the functions from part (a) and part (b) are indeed inverses of each other.

[Hint: This is a fun problem! Draw pictures and use the definition of a smooth manifold!]

There is yet another definition of the tangent space in terms of derivations (operators acting on differentiable functions satisfying a type of "product rule") and is close to another perspective we explored earlier in Lecture 6. However, there was something from that lecture that we should have discussed, which was how to obtain a vector field in $\mathbb{R}^n$ from a derivation. All we did was describe how to construct a derivation from a vector field. Before describing it, we review some important concepts from linear algebra, abstract algebra, and differential algebra (Exercises 12.28 and 12.44 are borderline "remarks" and their claims are not entirely required to understand everything else that follows).

Exercise 12.28. Let V be a vector space and let X be a set. The set

$$V^X := \left\{ V \xleftarrow{f} X \right\} \tag{12.29}$$

is a vector space by setting the following structure.

(a) Let $0 : X \to V$ be the constant function $0(x) := 0$ for all $x \in V$.

(b) Set the sum $f + g$ of two functions $f, g :\in V^X$ to be

$$X \ni x \mapsto (f+g)(x) := f(x) + g(x). \tag{12.30}$$

(c) Set the scalar multiplication λf of a function $f \in V^X$ by a number $\lambda \in \mathbb{R}$ to be

$$X \ni x \mapsto (\lambda f)(x) := \lambda f(x). \tag{12.31}$$

Prove that this is the *unique* vector space structure on V^X that satisfies the condition that the evaluation function $\mathrm{ev}_x : V^X \to \mathbb{R}$ is linear for every $x \in X$. Recall, the definition of the evaluation map is

$$V^X \ni f \mapsto \mathrm{ev}_x(f) := f(x). \tag{12.32}$$

By the way, this exercise explains in a precise sense what is meant by the "reasonable" or "natural" vector space structure on V^X, which you may have heard people say but not explain.[21] [Hint: This problem is not difficult and is just an unwinding of the definitions. First show that the vector space structure above satisfies the required property. Then, suppose that $0'$ is another zero vector, $+'$ is another sum, and use some notation to distinguish another possible scalar multiplication on the set V^X. Use the definitions and assumptions to prove that $0' = 0$, $+' = +$, and the scalar multiplications agree.]

For the exercise that will follow, we introduce two definitions: an associative algebra and an algebra homomorphism (only the former is important for the concept of derivations while the latter is only needed for Exercise 12.44 and not so important for the rest of the lecture). These are used to describe the algebraic structure of vector fields viewed as operations on smooth functions.

[21] It makes sense to me *why* this works as a vector space structure on V^X, but nobody every told me what *characterized* it! Think about it: V^X is a *huge* set, even if X is finite! So there could be *lots* of vector spaces structures on it (and there are). But by demanding the restrictions I've outlined, there's a unique one that stands out among all the rest. Where are these restrictions coming from? Well, the evaluation maps are natural mathematical objects associated with V^X, and here the word "natural" means something concrete: the evaluation maps are part of the data.

Definition 12.33. An *associative algebra* (or just *algebra* for short) consists of a vector space $\mathcal{A}$ together with a binary operation $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ called the *product*, whose value on (a, b) is written as ab, such that

(a) $(ab)c = a(bc)$

(b) $a(b + c) = ab + ac$

(c) $(a + b)c = ac + bc$

(d) $\lambda(ab) = (\lambda a)b = a(\lambda b)$

for all $a, b, c \in \mathcal{A}$ and all $\lambda \in \mathbb{R}$. An algebra is *commutative* iff $ab = ba$ for all $a, b \in \mathcal{A}$. If $\mathcal{A}$ is a set, an *algebra structure* on $\mathcal{A}$ is a vector space structure on $\mathcal{A}$ together with a binary operation $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ satisfying the above postulates.

Example 12.34. The set $\mathbb{R}$ of real numbers under the usual addition and multiplication operations is a commutative algebra.

Example 12.35. The set of $m \times m$ matrices is an algebra using the component-wise addition and matrix multilplication for the product. It is commutative if and only if $m = 1$.

Example 12.36. The set of upper-triangular $m \times m$ matrices is an algebra using the component-wise addition and matrix multilplication for the product. It is commutative if and only if $m = 1$.

Example 12.37. The set of diagonal $m \times m$ matrices is a commutative algebra using the component-wise addition and matrix multiplication for the product.

Example 12.38. The set of smooth functions $C^\infty(\mathbb{R}, \mathbb{R}^n)$ on $\mathbb{R}^n$ is an associative algebra with the natural algebra structure obtained from $\mathbb{R}$ (see Exercise 12.44). Namely, given two functions $f, g : \mathbb{R}^n \to \mathbb{R}$, the product fg is defined to be

$$\mathbb{R} \ni x \mapsto f(x)g(x). \tag{12.39}$$

Example 12.40. The set of invertible $m \times m$ matrices using the matrix sum and product is *not* an algebra.

Example 12.41. The set of vectors in $\mathbb{R}^3$ with usual addition but the cross product as the product is *not* an associative algebra.

Definition 12.42. Let $\mathcal{A}$ and $\mathcal{B}$ be two algebras. A function $f : \mathcal{A} \to \mathcal{B}$ is an *algebra homomorphism* iff f is a linear transformation and

$$f(ab) = f(a)f(b) \qquad \forall\, a, b \in \mathcal{A}. \tag{12.43}$$

Exercise 12.44. Let X be a set and $\mathcal{A}$ an algebra. Let $\mathcal{A}^X$ denote the set of functions from X to $\mathcal{A}$. Prove that there is a unique algebra structure on $\mathcal{A}^X$ such that the evaluation map $\mathrm{ev}_x : \mathcal{A}^X \to \mathcal{A}$ is an algebra homomorphism for every $x \in X$. Describe the product explicitly.

Definition 12.45. Let $c \in \mathbb{R}^n$. A *derivation on $\mathbb{R}^n$ at c* is a linear function

$$\mathcal{V}_c : C^\infty(\mathbb{R}, \mathbb{R}^n) \to \mathbb{R} \tag{12.46}$$

satisfying

$$\mathcal{V}_c(fg) = (\mathcal{V}_c f)g(c) + f(c)(\mathcal{V}_c g) \tag{12.47}$$

for all $f, g \in C^\infty(\mathbb{R}, \mathbb{R}^n)$.

We already know several examples of derivations at points in Euclidean space. In fact, we have studied many properties of derivations in general in Lecture 6 (particularly Theorem 6.30) though abstracting from that and defining general derivations on associative algebras would take us on too far of a tangent (no pun intended). Instead, we focus on derivations at points and hope to understand the relationship between them and vectors, namely the tangent space at a point.

Exercise 12.48. Let $c \in \mathbb{R}^n$ and let $V_c \in T_c\mathbb{R}^n$ be a vector at c. Show that the function $\mathcal{V}_c : C^\infty(\mathbb{R}, \mathbb{R}^n) \to \mathbb{R}$ defined by sending a smooth (technically, all you need is differentiable here) function f to

$$\mathcal{V}_c f := (D_c f)(v_c) \tag{12.49}$$

is a derivation at c.

Is every derivation of $\mathbb{R}^n$ at c of this form? To answer this, we will first explore some facts.

Exercise 12.50. Let $c \in \mathbb{R}^n$ and let $\mathcal{V}_c : C^\infty(\mathbb{R}, \mathbb{R}^n) \to \mathbb{R}$ be a derivation at c.

(a) Show that

$$\mathcal{V}_c f = 0 \tag{12.51}$$

for any constant function f. [Hint: consider the constant 1 and then use linearity.]

(b) Show that if $f, g \in C^\infty(\mathbb{R}, \mathbb{R}^n)$ satisfy $f(c) = g(c) = 0$, then

$$\mathcal{V}_c(fg) = 0. \tag{12.52}$$

A close version of the following fact was stated as Problem 2.35 in Spivak [13]. Since it is important, we provide the proof.

Theorem 12.53 (Hadamard's Lemma)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then there exists a function $g : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$f(x) = f(0) + \sum_{i=1}^{n} x_i g_i(x) \qquad \forall\, x \in \mathbb{R}^n, \tag{12.54}$$

i.e. such that the diagram

$$\begin{tikzcd}
\mathbb{R} & \mathbb{R}^n \arrow[l, "f"'] \arrow[dd, "\Delta"] \\
\mathbb{R} \arrow[u, "+f(0)"] & \\
\mathbb{R}^n \times \mathbb{R}^n \arrow[u, "\langle \cdot , \cdot \rangle"] & \mathbb{R}^n \times \mathbb{R}^n \arrow[l, "\mathrm{id}\times g"]
\end{tikzcd} \tag{12.55}$$

commutes, and such that

$$g_i(0) = (\partial_i f)(0) \qquad \forall\, i \in \{1, \dots, n\}. \tag{12.56}$$

Furthermore, if f is of class C^r, then g can be chosen to be of class C^{r-1}.

Proof. For each $x \in \mathbb{R}^n$, define the function $h_x : \mathbb{R} \to \mathbb{R}$ by

$$\mathbb{R} \ni t \mapsto h_x(t) := f(tx). \tag{12.57}$$

In other words,

$$h_x = f \circ \gamma_{0;x} \tag{12.58}$$

in terms of our earlier notation. Recall, $\gamma_{0;x}(t) := tx$ for all $t \in \mathbb{R}$. By the Chain Rule

$$D_t h_x = (D_{tx} f) \circ (D_t \gamma_{0;x}) \tag{12.59}$$

so that

$$\begin{aligned} h'_x(t) &= (D_t h_x)(1) \\ &= (D_{tx} f)\big((D_t \gamma_{0;x})(1)\big) \\ &= (D_{tx} f)(x) \\ &= \big[(\partial_1 f)(tx) \quad \cdots \quad (\partial_n f)(tx)\big] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= \sum_{i=1}^{n} (\partial_i f)(tx) x_i. \end{aligned} \tag{12.60}$$

By this result and the Fundamental Theorem of Calculus,

$$f(x) - f(0) = h_x(1) - h_x(0) = \int_0^1 h'_x(t) dt = \sum_{i=1}^{n} \left(\int_0^1 (\partial_i f)(tx) dt \right) x_i \tag{12.61}$$

for all $x \in \mathbb{R}$. Therefore, for each $i \in \{1, \dots, n\}$, set $g_i : \mathbb{R}^n \to \mathbb{R}$ to be the function

$$\mathbb{R}^n \ni x \mapsto g_i(x) := \int_0^1 (\partial_i f)(tx) dt \tag{12.62}$$

and set $g : \mathbb{R}^n \to \mathbb{R}^n$ to be the function sending $x \in \mathbb{R}^n$ to $\big(g_1(x), g_2(x), \dots, g_n(x)\big)$. Then g satisfies the required conditions. ■

Theorem 12.63. *Let $\mathcal{V}_0$ be a derivation on $\mathbb{R}^n$ at 0. Then there exists a vector $v_0 \in T_0\mathbb{R}^n$ such that*

$$\mathcal{V}_0 f = (D_0 f)(v_0) \tag{12.64}$$

for all $f \in C^\infty(\mathbb{R}, \mathbb{R}^n)$.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. By Hadamard's Lemma, there exist smooth functions $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i \in \{1, \ldots, n\}$ such that

$$f = f(0) + \sum_{i=1}^{n} \pi_i g_i \tag{12.65}$$

and $g_i(0) = (\partial_i f)(0)$. Since $\mathcal{V}_0$ is a derivation at 0, $\mathcal{V}\big(f(0)\big) = 0$ by Exercise 12.50. Furthermore, by linearity and the Leibniz property,

$$\mathcal{V}_0 f = \sum_{i=1}^{n} \mathcal{V}_0(\pi_i g_i) = \sum_{i=1}^{n} \Big((\mathcal{V}_0 \pi_i) g_i(0) + \pi_i(0)(\mathcal{V}_0 g_i) \Big). \tag{12.66}$$

Since $g_i(0) = (\partial_i f)(0)$ and $\pi_i(0) = 0$, this becomes

$$\mathcal{V}_0 f = \sum_{i=1}^{n} (\mathcal{V}_0 \pi_i)(\partial_i f)(0). \tag{12.67}$$

Therefore, the vector

$$v_0 := \sum_{i=1}^{n} (\mathcal{V}_0 \pi_i) e_i, \tag{12.68}$$

where e_i is the i-th unit vector (the subscript 0 has been left off) satisfies the required properties. Namely,

$$(D_0 f)(v_0) = (D_0 f)\left(\sum_{i=1}^{n} (\mathcal{V}_0 \pi_i) e_i \right) = \sum_{i=1}^{n} (\mathcal{V}_0 \pi_i)(D_0 f)(e_i) = \sum_{i=1}^{n} (\mathcal{V}_0 \pi_i)(\partial_i f)(0) = \mathcal{V}_0 f. \tag{12.69}$$

■

Exercise 12.70. Prove the following slight generalization of Theorem 12.63: "Let $c \in \mathbb{R}^n$ and let $\mathcal{V}_c$ be a derivation on $\mathbb{R}^n$ at c. Then there exists a vector $v_c \in T_c\mathbb{R}^n$ such that

$$\mathcal{V}_c f = (D_c f)(v_c) \tag{12.71}$$

for all $f \in C^\infty(\mathbb{R}, \mathbb{R}^n)$." [Advice: It might be possible to use the results of Hadamard's Lemma and/or Theorem 12.63 and slightly modifying them for this case.]

Using this idea for vector fields in Euclidean space, the following exercise asks to prove a similar statement for smooth functions on smooth manifolds. However, it restricts attention to derivations concentrated at a single point (we have not yet even defined what we mean by vector fields on manifolds, so it would not make sense to talk about them yet).

Exercise 12.72. Let M be a smooth manifold and let $C^\infty(\mathbb{R}, M)$ be the set of smooth functions from M to $\mathbb{R}$ (see Definition 4.19). Let $c \in M$.

(a) Briefly describe the natural algebra structure on $C^\infty(\mathbb{R}, M)$ (see Exercises 12.28 and 12.44).

(b) In Theorem 6.30 and Exercise 12.48, it was shown that vector fields when viewed as operators (known as derivations) satisfy certain conditions. In Exercise 12.24, it was shown that T_cM is isomorphic to certain equivalence classes of curves in M. In this exercise, we will extend Theorem 6.30 and Theorem 12.63 to manifolds. Show that for any vector $V_c \in T_cM$, which can be represented by a differentiable curve $\gamma : \mathbb{R} \to M$ satisfying $\gamma(0) = c$ and $\gamma'(0) = V_c$, the function $\mathcal{V}_c : C^\infty(\mathbb{R}, M) \to \mathbb{R}$ defined by

$$C^\infty(\mathbb{R}, M) \ni f \mapsto \mathcal{V}_c(f) := (f \circ \gamma)'(0) \tag{12.73}$$

satisfies the conditions of Definition 12.45, i.e.

i. $\mathcal{V}_c(f + g) = \mathcal{V}_c f + \mathcal{V}_c g$,

ii. $\mathcal{V}_c(\lambda f) = \lambda \mathcal{V}_c f$, and

iii. $\mathcal{V}_c(fg) = (\mathcal{V}_c f)g(c) + f(c)(\mathcal{V}_c g)$

for all smooth functions $f, g \in C^\infty(\mathbb{R}, M)$ and all constants $\lambda \in \mathbb{R}$. A function $\mathcal{V}_c : C^\infty(\mathbb{R}, M) \to \mathbb{R}$ satisfying the three conditions in part (b) is known as a *derivation on M at c*. The set of derivations on M at c is denoted by $\mathfrak{Der}_c(M)$.

(c) In part (a), an assignment $T_cM \to \mathfrak{Der}_c(M)$ was provided. Construct an inverse to this function. Namely, associate to a derivation $\mathcal{V}_c : C^\infty(\mathbb{R}, M) \to \mathbb{R}$ at c a vector $V_c \in T_cM$ and show that these functions are (set-theoretic) inverses of each other. [Hint: use a local chart and then use Hadamard's Lemma and/or Theorem 12.63.]

Hopefully, these many different ways of thinking about tangent vectors will provide a more thorough understanding of what a tangent vector is. We now move back to properties of the tangent space and the differential of a differentiable function between manifolds.

Exercise 12.74. See Exercise 12.72 for background.

(a) Prove that the set of derivations on $C^\infty(\mathbb{R}, M)$ at c is a vector subspace of the vector space of all functions $\mathbb{R}^{C^\infty(\mathbb{R},M)}$.

(b) Prove that there is a vector space isomorphism $T_cM \xrightarrow{\cong} \mathfrak{Der}_c(M)$. [Hint: most of the work for this was already done in earlier results.]

The next few exercises explore the tangent space of Lie groups at the identity. This tangent space is known as the Lie algebra associated to the Lie group.

Definition 12.75. Let G be a Lie group. The *Lie algebra of G*, denoted by $\mathfrak{g}$, is the tangent space of G at the identity $e \in G$,

$$\mathfrak{g} := T_eG. \tag{12.76}$$

The reason for the word "algebra" is because there is an algebraic structure associated with the tangent space. We will discuss this algebraic structure later after we define vector fields on manifolds. The word "Lie" is used to distinguish this algebraic structure from an associative algebraic structure. Such an algebraic structure does not naturally appear for arbitrary manifolds at their tangent spaces. To give some examples, we recall what a Cauchy sequence.

Definition 12.77. Let $A \subseteq \mathbb{R}^n$. A sequence $a : \mathbb{N} \to A$ in A is a *Cauchy sequence* iff for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$\|a_n - a_m\| < \epsilon \qquad \forall\, n, m \geq N. \tag{12.78}$$

Exercise 12.79. Prove that a sequence $a : \mathbb{N} \to \mathbb{R}^n$ is Cauchy if and only if it converges.

Exercise 12.80. Let A be an $m \times m$ matrix. Prove that the sequence of partial sums

$$\begin{aligned} \mathbb{N} &\to \mathbb{R}^{m^2} \\ n &\mapsto \sum_{k=0}^{n} \frac{A^k}{k!} \end{aligned} \tag{12.81}$$

converges to a matrix. In this notation $A^0 := \mathbb{1}$ is the identity matrix and

$$k! := \begin{cases} 1 & \text{if } k = 0 \\ \prod_{i=1}^{k} i & \text{if } k \geq 1 \end{cases} \tag{12.82}$$

is the factorial of a non-negative number. The limit of the sequence of partial sums from above is called the *exponential* of the matrix A

$$\exp(A) := \sum_{k=0}^{\infty} \frac{A^k}{k!}. \tag{12.83}$$

[Hint: Use Problem 1.10 in Spivak [13] and Exercise 12.79. You may also assume the exponential of a *number* converges.]

Exercise 12.84. Let A be an $m \times m$ matrix. Prove that $\exp(A)$ is invertible. [Hint: Show that $\exp(-A)\exp(A) = \mathbb{1}$ by using the (incredibly useful)[22] identity $\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} a_{k,n} = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_{k,n-k}$.]

Exercise 12.85. Recall that $O(3)$ is the Lie group of orthogonal 3×3 matrices (see Exercise 11.71).

(a) Prove that the tangent space of $O(3)$ at the identity is the set of skew-symmetric 3×3 matrices, i.e. matrices A such that $A^T = -A$. [Hint: let $\gamma : \mathbb{R} \to O(3)$ be a path such that $\gamma(0) = \mathbb{1}$, let $v \in \mathbb{R}^3$ and consider the function $\mathbb{R} \to \mathbb{R}$ given by sending $t \in \mathbb{R}$ to $\langle \gamma(t)v, \gamma(t)v \rangle$.]

(b) Let A, B be skew-symmetric 3×3 matrices. Show that the commutator $[A, B]$ of A and B is again a skew-symmetric 3×3 matrix. Here, the commutator of A and B is defined to be $[A, B] := AB - BA$.

(c) Let A, B be skew-symmetric 3×3 matrices. Is it always true that AB is a skew-symmetric 3×3 matrix? Explain.

Level I problems.
From these notes: Exercises 12.10, 12.12, 12.14, 12.28, 12.44, 12.48, 12.50, 12.74

Level II problems.
From these notes: Exercises 12.24, 12.70, 12.72, 12.79, 12.80, 12.84, 12.85

[22] Seriously, you have no idea how many times I've used this identity in my life. Note that the $a_{k,n}$ is this summation formula are *not* the components of the matrix A. You do not need to express A in terms of its components to do this problem.

13 March 23: Level Sets for Manifolds

There is an incredibly important point that needs to be made about the tangent space T_cM to a manifold M at a point $c \in M$ that we took for granted for open subsets of Euclidean space.

There is no canonical basis for the vector space T_cM.

To make sure you believe this statement, look at all of the previous examples and try to provide a canonical choice. Yes, there is always *a* basis, but you might choose a different one than I might. This will imply that we will not be able to write down a canonical "Jacobian matrix" for a given differentiable function $f : N \to M$ between manifolds at a point. Nevertheless, there *is* a notion of a *differential* when viewed as a linear transformation *without* having chosen a basis. This is what the following theorem is about.

Theorem 13.1. *Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be differentiable m- and n-dimensional manifolds, respectively, let $c \in N$, and let $f : N \to M$ be function that is differentiable at c. Let $\tilde{U} \subseteq \mathbb{R}^l$ be an open neighborhood of c and let $\tilde{f} : \tilde{U} \to \mathbb{R}^k$ be a function that is differentiable at c and equals f when restricted to $N \cap \tilde{U}$. Then the following two conditions hold:*

(a) $(D_c\tilde{f})(T_cN) \subseteq T_{f(c)}M$ and

(b) if $\overline{U} \subseteq \mathbb{R}^l$ is another open neighborhood of c and $\overline{f} : \overline{U} \to \mathbb{R}^k$ is another function that is differentiable at c and agrees with f on $\overline{U} \cap M$, then $D_c\tilde{f}\big|_{T_cN} = D_c\overline{f}\big|_{T_cN}$, i.e. the restrictions of $D_c\tilde{f}$ and $D_c\overline{f}$ to T_cN both agree.

In this case, it is appropriate to denote the restrict of either $D_c\tilde{f}$ or $D_c\overline{f}$ to T_cN by D_cf.

Proof. The proof will be broken into two parts using the notation from each respective part.

(a) Choose parametrizations

$$\varphi : W \to N \qquad \& \qquad \psi : V \to M \tag{13.2}$$

of N and M about the point c and $f(c)$ respectively. Here $W \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open subsets chosen to satisfy the additional condition that[23] $\varphi(W) \subseteq \tilde{U}$ and $f\big(\varphi(W)\big) \subseteq \psi(V)$. Therefore, the composition

$$V \xleftarrow{\psi^{-1}} \psi(V) \xleftarrow{f} \varphi(W) \xleftarrow{\varphi} W \tag{13.3}$$

is differentiable on W since each function in the composition is differentiable. Furthermore, the diagram

$$\begin{tikzcd} \mathbb{R}^k & \tilde{U} \arrow[l, "\tilde{f}"'] \\ V \arrow[u, "\psi"] & W \arrow[l, "\psi^{-1}\circ f\circ\varphi"] \arrow[u, "\varphi"'] \end{tikzcd} \tag{13.4}$$

[23]See Exercise 13.9 for a justification.

of differentiable functions commutes. Hence, by the Chain Rule,[24] the diagram

$$
\begin{tikzcd}
\mathbb{R}^k & & \mathbb{R}^l \arrow[ll, "D_c\tilde{f}"'] \\
\mathbb{R}^m \arrow[u, "D_{\psi^{-1}(f(c))}\psi"] & & \mathbb{R}^n \arrow[ll, "D_{\varphi^{-1}(c)}(\psi^{-1}\circ f\circ\varphi)"] \arrow[u, "D_{\varphi^{-1}(c)}\varphi"']
\end{tikzcd}
\tag{13.5}
$$

of linear transformations commutes. Note that by starting at the bottom right corner of this diagram, it follows that T_cN, which is by definition $(D_{\varphi^{-1}(c)}\varphi)(\mathbb{R}^n) \subseteq \mathbb{R}^l$, gets mapped to a subspace $(D_c\tilde{f})(T_cN)$ of $\mathbb{R}^k$ by following the top and right part of this diagram. However, $\mathbb{R}^n$ also gets mapped to a subspace of $\mathbb{R}^m$ along the bottom arrow $D_{\varphi^{-1}(c)}(\psi^{-1}\circ f\circ\varphi)$. Since $T_{f(c)}M$ is by definition the image of $\mathbb{R}^m$ under $D_{\varphi^{-1}(c)}\varphi$ and $D_{\varphi^{-1}(c)}(\psi^{-1}\circ f\circ\varphi)(\mathbb{R}^n) \subseteq \mathbb{R}^m$, this implies $(D_c\tilde{f})(T_cN) \subseteq T_{f(c)}M$.

(b) Choosing V and W as in the proof of part (a) to satisfy the additional condition that $\varphi(W) \subseteq \overline{U}$ we obtain the commutative diagram (by combining two of the diagrams like (13.4))

$$
\begin{tikzcd}
\mathbb{R}^k \arrow[rr, equal, "\mathrm{id}"] & & \mathbb{R}^k & & & \\
& & & \tilde{U} \arrow[ulll, "\tilde{f}"'] & & \overline{U} \arrow[ulll, "\overline{f}"'] \\
& V \arrow[uul, "\psi"] \arrow[uur, "\psi"'] & & & & \\
& & & & W \arrow[ulll, "\psi^{-1}\circ f\circ\varphi"] \arrow[uul, "\varphi"] \arrow[uur, "\varphi"'] &
\end{tikzcd}
\tag{13.6}
$$

of differentiable functions. Applying the differential gives another commutative diagram (not all linear transformations have been labelled to avoid clutter)

$$
\begin{tikzcd}
\mathbb{R}^k \arrow[rr, equal, "\mathrm{id}"] & & \mathbb{R}^k & & & \\
& & & \tilde{U} \arrow[ulll, "D_c\tilde{f}"'] & & \overline{U} \arrow[ulll, "D_c\overline{f}"'] \\
& V \arrow[uul, "D_{\psi^{-1}(f(c))}\psi"] \arrow[uur] & & & & \\
& & & & W \arrow[ulll, "D_{\varphi^{-1}(c)}(\psi^{-1}\circ f\circ\varphi)"] \arrow[uul] \arrow[uur, "D_{\varphi^{-1}(c)}\varphi"'] &
\end{tikzcd}
\tag{13.7}
$$

By definition of T_cN, for any vector $v \in T_cN$, there exists a unique vector $u \in \mathbb{R}^n$ such that $v = (D_{\varphi^{-1}(c)}\varphi)(u)$. Hence, by commutativity of this diagram, it follows that $(D_c\tilde{f})(v) = (D_c\overline{f})(v)$. Since $v \in T_cN$ was arbitrary, this shows that $D_c\tilde{f}$ restricted to T_cN agrees with $D_c\overline{f}$ restricted to T_cN.

■

[24] Notice that we are using the ordinary Chain Rule and not some version for manifolds. This is because all differentials are defined on open subsets of Euclidean space. We will discuss a version for manifolds shortly—it will follow from this theorem.

Definition 13.8. Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be differentiable m- and n-dimensional manifolds, respectively, let $c \in N$, and let $f : N \to M$ be function that is differentiable at c. The *differential of f at c* is the linear transformation $D_c f : T_c N \to T_{f(c)} M$ from Theorem 13.1.

This result and construction should surprise you. After Definition 4.20, it was stated that for a function $f : A \to \mathbb{R}$ defined on an arbitrary subset $A \subseteq \mathbb{R}^k$, although it makes sense to ask if f is differentiable at a point $a \in A$, it is meaningless to ask what its differential is since it might not be unique. This theorem shows that if $A \subseteq \mathbb{R}^k$ is a manifold, then the differential *can* be defined, but only on a *subspace* of $\mathbb{R}^k$ (technically a subspace of $T_a\mathbb{R}^k$) namely, a subspace $T_a A$ whose dimension coincides with the dimension of the manifold A.

Exercise 13.9. Refer to the proof of part (a) of Theorem 13.1. Explain why the open sets V and W can be chosen to satisfy the conditions that $\varphi(W) \subseteq \tilde{U}$ and $f\big(\varphi(W)\big) \subseteq \psi(V)$.

In addition, this differential satisfies the following familiar properties.

Theorem 13.10 (Chain Rule for manifolds)**.** *Let $M \subseteq \mathbb{R}^j$, $N \subseteq \mathbb{R}^k$, and $P \subseteq \mathbb{R}^l$ be m-, n-, and p-dimensional manifolds of class C^r, respectively. Let $c \in P$ and let*

$$M \xleftarrow{f} N \xleftarrow{g} P \tag{13.11}$$

be functions with g differentiable at c and f differentiable at $g(c)$. Then $f \circ g$ is differentiable at c and the diagram

$$\begin{tikzcd} & T_{g(c)}N \arrow[dl, "D_{g(c)}f"'] & \\ T_{f(g(c))}M & & T_c P \arrow[ul, "D_c g"'] \arrow[ll, "D_c(f\circ g)"] \end{tikzcd} \tag{13.12}$$

of linear transformations commutes.

Exercise 13.13.

(a) Prove Theorem 13.10.

(b) Let $\mathrm{id}_M : M \to M$ be the identity map on a manifold $M \subseteq \mathbb{R}^k$. Prove that $D_c\mathrm{id} = \mathrm{id}_{T_c M}$ for all $c \in M$.

(c) Let M and N be manifolds and let $f : N \to M$ be a diffeomorphism (f is differentiable with a differentiable inverse). Show that $D_{f(c)} f^{-1} = (D_c f)^{-1}$ for all $c \in N$. Note that this implies, in particular, the dimension of M is equal to the dimension of N.

Exercise 13.14. This exercise explores the product of manifolds and differentiable functions on them.

(a) Let $N \subseteq \mathbb{R}^j$ and $Q \subseteq \mathbb{R}^l$ be manifolds of dimensions n and q respectively. Show that $N \times Q \subseteq \mathbb{R}^{j+l}$ is a manifold of dimension $n + q$.

(b) Let $a \in N$ and $b \in Q$. Show that

$$T_{(a,b)}(N \times Q) = T_a N \oplus T_b Q. \tag{13.15}$$

Here $\oplus$ stands for the direct sum of vector spaces. Recall, if V and W are subspaces of some vector space U satisfying $V \cap W = \{0\}$, then $V \oplus W$ is the subspace of U given by

$$V \oplus W = \big\{ v + w \in U \ : \ v \in V, \ w \in W \big\}. \tag{13.16}$$

(c) Let $P \subseteq \mathbb{R}^k$ and $M \subseteq \mathbb{R}^i$ be manifolds of dimension p and m, respectively. Let $f : N \to M$ and $g : Q \to P$ be differentiable functions. Show that

$$D_{(a,b)}(f \times g) = (D_a f) \oplus (D_b g). \tag{13.17}$$

Here $\oplus$ stands for the direct sum of linear transformations. Recall, if Y and Z are subspaces of another vector space X satisfying $Y \cap Z = \{0\}$ and $S : V \to Y$ and $T : W \to Z$ are linear transformations satisfying the condition that $S(V) \cap T(W) = \{0\}$, then $S \oplus T : V \oplus W \to Y \oplus Z$ is the linear transformation defined by[25]

$$V \oplus W \ni (v, w) \mapsto (S \oplus T)(v, w) := S(v) + T(w). \tag{13.18}$$

Example 13.19. The *(unit) n-torus in* $\mathbb{R}^{2n}$ is the n-dimensional manifold

$$\mathbb{T}^n := \overbrace{S^1 \times \cdots \times S^1}^{n \text{ times}}, \tag{13.20}$$

where $S^1 \subseteq \mathbb{R}^2$ is the unit circle in $\mathbb{R}^2$.

Let us come back to the point made earlier about there not being a canonical basis of vectors at the tangent space to an arbitrary manifold. But first, we need to be sure we understand a concept from linear algebra about bases and how a linear transformation $T : V \to W$ together with bases on V and W determines a matrix.

Definition 13.21. Let W and V be finite-dimensional vector spaces of dimensions n and m respectively, let $T : W \to V$ be a linear transformation, let $\mathcal{C} := \{w_1, \ldots, w_n\}$ be an ordered basis for W, and let $\mathcal{B} := \{v_1, \ldots, v_m\}$ be an ordered basis for V. Let $\Phi_{\mathcal{C}} : \mathbb{R}^n \to W$ and $\Psi_{\mathcal{B}} : \mathbb{R}^m \to V$ be the linear transformations uniquely determined by the conditions

$$\Psi_{\mathcal{B}}(e_i) = w_i \qquad \forall\, i \in \{1, \ldots, m\} \qquad \& \qquad \Phi_{\mathcal{C}}(e_j) = v_j \qquad \forall\, j \in \{1, \ldots, n\}. \tag{13.22}$$

[25]You already know what this is: this is just the block sum of matrices if we have a basis. Namely,

$$S(v) + T(w) = \begin{bmatrix} [S] & 0 \\ 0 & [T] \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$$

in terms of components.

The *matrix associated to the linear transformation T and the ordered bases $\mathcal{B}$ and $\mathcal{C}$* is the matrix associated to the linear transformation ${}_{\mathcal{B}}T^{\mathcal{C}} : \mathbb{R}^n \to \mathbb{R}^m$ determined by the condition that the diagram

$$\begin{tikzcd} V & W \arrow[l, "T"'] \\ \mathbb{R}^m \arrow[u, "\Psi_{\mathcal{B}}"] & \mathbb{R}^n \arrow[l, "{}_{\mathcal{B}}T^{\mathcal{C}}"] \arrow[u, "\Phi_{\mathcal{C}}"'] \end{tikzcd} \tag{13.23}$$

commutes. In other words,

$${}_{\mathcal{B}}T^{\mathcal{C}} := \Psi_{\mathcal{B}}^{-1} \circ T \circ \Phi_{\mathcal{C}}. \tag{13.24}$$

The reason for the notation is explained by the following exercise.

Exercise 13.25. Let U, V, and W be finite-dimensional vector spaces, let $U \xleftarrow{S} V \xleftarrow{T} W$ be linear transformations, and let $\mathcal{B}, \mathcal{C}$, and $\mathcal{D}$ be ordered bases for U, V, and W, respectively. Prove that

$${}_{\mathcal{B}}S^{\mathcal{C}} \circ {}_{\mathcal{C}}T^{\mathcal{D}} = {}_{\mathcal{B}}(S \circ T)^{\mathcal{D}}. \tag{13.26}$$

[Hint: You do not need to write down explicit vectors for the bases to do this problem.]

Definition 13.27. Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be differentiable m- and n-dimensional manifolds, respectively, let $c \in N$, and let $f : N \to M$ be function that is differentiable at c. Let $\mathcal{C}$ and $\mathcal{B}$ be ordered bases for T_cN and $T_{f(c)}M$, respectively. The *Jacobian of f at c with respect to $\mathcal{C}$ and $\mathcal{B}$* is the matrix $[{}_{\mathcal{B}}(D_cf)^{\mathcal{C}}]$ associated to the linear transformation ${}_{\mathcal{B}}(D_cf)^{\mathcal{C}} : \mathbb{R}^n \to \mathbb{R}^m$.

So if we can always choose a basis for each tangent space, why not just do that and call it a day? For one, we do not want to choose bases at points arbitrarily. For instance, if we pick a basis $\mathcal{C}_c$ for T_cN, and if $x \in N$ that is close to c (in some suitably small open neighborhood of c), we should choose another basis $\mathcal{C}_x$ of T_xN that is also sufficiently close to $\mathcal{C}_c$ in some precise sense (we will say in what precise sense soon). In fact, we might want the basis to vary smoothly in a neighborhood of c. We were able to do this in $\mathbb{R}^n$ because we had the vector fields $\mathcal{E}_i$ for each $i \in \{1, \ldots, n\}$, which were not only smoothly varying, they were constant. This certainly does not happen on an arbitrary manifold. For example, just look at the unit circle. $e_1 \in T_{(0,1)}S^1$ and $e_1 \in T_{(0,-1)}$ but e_1 is not in any other tangent space on the circle.

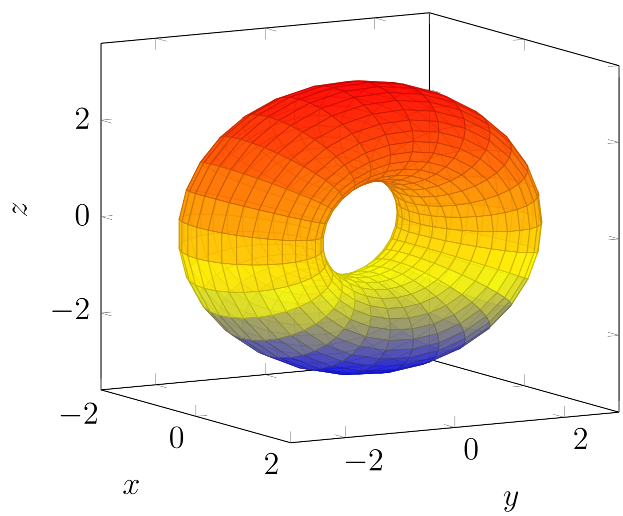
It turns out (and we will prove this) that the following statement: "There exists a smoothly varying basis of tangent vectors on every manifold" is *false*! This is closely related to the fact that not every n-dimensional manifold admits a non-vanishing vector field (a smoothly varying basis of tangent vectors would require n such linearly independent non-vanishing vectors, so if you don't even have a single non-vanishing vector field, then you can't possibly have n of them!). We will take this discussion momentarily, but first let us explore some more facts about implicitly defined manifolds, their tangent spaces, and differentiable functions between them. We can obtain manifolds as level sets of differentiable functions between manifolds provided their dimensions are appropriate. We repeat the definitions of critical points/values and regular points/values this time for maps between manifolds.

Definition 13.28. Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be smooth m- and n-dimensional manifolds, respectively, and let $f : N \to M$ be a smooth function. A *critical point* of f is a point $a \in N$ such that $D_a f : T_a N \to T_{f(a)} M$ is not onto (has rank less than m). Let C_f denote the critical points of f. A *regular value* of f is a point in $M \setminus f(C_f)$, the complement of the image of all the critical points of f. This set is denoted by $R_f := M \setminus f(C_f)$. A *regular point* of f is an element of $f^{-1}(R_f)$.

Example 13.29. Let $X : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}^3$ be defined by the following component functions sending $(s, t) \in [0, 2\pi) \times [0, 2\pi)$ to

$$\begin{aligned} X_1(s,t) &:= \sin(t) \\ X_2(s,t) &:= \big(2 + \cos(t)\big)\cos(s) \\ X_3(s,t) &:= \big(2 + \cos(t)\big)\sin(s). \end{aligned} \tag{13.30}$$

Let M be the image of X.



Let $\pi : \mathbb{R}^3 \to \mathbb{R}$ be the projection onto the third component

$$\mathbb{R}^3 \ni (x, y, z) \mapsto \pi(x, y, z) := z \tag{13.31}$$

and let $h : M \to \mathbb{R}$ be the restriction of π to M. π is smooth everywhere and has no critical points (check this!). Even though h is the restriction of π, h *does* have critical points. In fact, there are four critical points on M given by

$$C_h = \big\{(0,0,3), (0,0,1), (0,0,-1), (0,0,-3)\big\} \subseteq M. \tag{13.32}$$

The set of regular values is given by

$$R_f = \mathbb{R} \setminus \big\{3, 1, -1, -3\big\} = (-\infty, -3) \cup (-3, -1) \cup (-1, 1) \cup (1, 3) \cup (3, \infty). \tag{13.33}$$

Hence, the set of regular points on M is given by

$$M \setminus \Big(h^{-1}(3) \cup h^{-1}(1) \cup h^{-1}(-1) \cup h^{-1}(-3)\Big). \tag{13.34}$$

A visualization of this set can be obtained from the figures below, but before we do this, we should prove all of these claims with our currently available techniques. Fortunately, we have a parametrization of M by extending the function X to all of $\mathbb{R}^2$ and then restricting to local patches. As long as these patches are contained in squares of side length 2π, they will form a coordinate system for M about some point. Note that $X_3 = h \circ X$ and the tangent space to any point $c \in M$ can be computed by finding (s,t) such that $X(s,t) = c$ and then calculating the image of $D_{(s,t)}X$. Let's work this out explicitly. First, calculating the differential $D_{(s,t)}X$ as a matrix using the standard basis gives

$$[D_{(s,t)}X] = \begin{bmatrix} 0 & \cos(t) \\ -\big(2+\cos(t)\big)\sin(s) & -\sin(t)\cos(s) \\ \big(2+\cos(t)\big)\cos(s) & -\sin(t)\sin(s) \end{bmatrix}.$$

The span of these column vectors is precisely $T_{X(s,t)}M$. In fact, these two column vectors form a basis for this tangent space. $c = X(s,t)$ is a critical point of h if and only if $D_{X(s,t)}h$ is the 0 linear transformation (since the codomain of h is 1-dimensional). But h is precisely the restriction of π to M so

$$D_{X(s,t)}h = D_{X(s,t)}\pi\Big|_{T_{X(s,t)}M} = D_{X(s,t)}\pi\Big|_{(D_{(s,t)}X)(\mathbb{R}^2)},$$

where the bar notation means the restriction to the subspace in the subscript. This linear transformation vanishes precisely if the image of both basis vectors go to the zero vector because applying this linear transformation $D_{X(s,t)}\pi = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ to these vectors yields

$$\big(2+\cos(t)\big)\cos(s) \qquad \& \qquad -\sin(t)\sin(s), \tag{13.35}$$

respectively, under the differential of π restricted to T_cM. Setting these terms to zero enforces the conditions

$$s \in \left\{\frac{\pi}{2} + n\pi \ : \ n \in \mathbb{Z}\right\} \qquad \& \qquad t \in \{m\pi \ : \ m \in \mathbb{Z}\}. \tag{13.36}$$

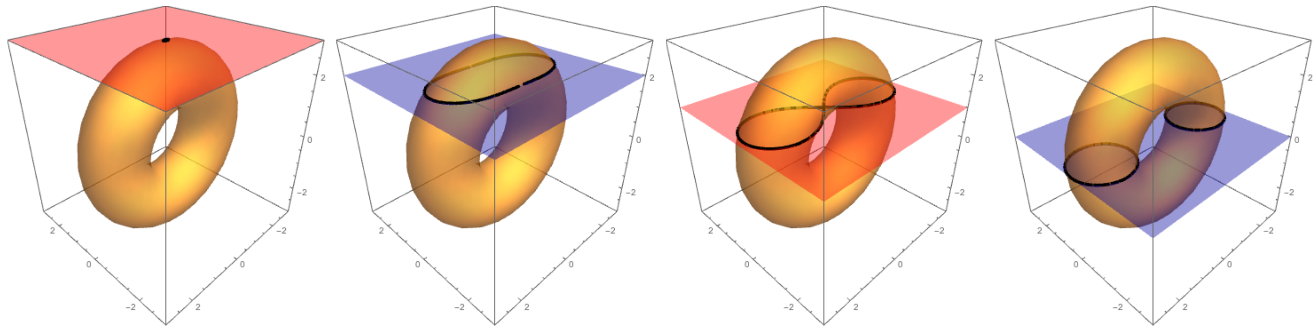
In the domain $[0,2\pi) \times [0,2\pi)$, this gives us only two possible values for s and t

$$s \in \left\{\frac{\pi}{2}, \frac{3\pi}{2}\right\} \qquad \& \qquad t \in \big\{0, \pi\big\} \tag{13.37}$$

which combined yield four critical points in M. They are given by plugging in these four possible combinations into the function X. The result is

$$\begin{aligned} X\left(\frac{\pi}{2}, 0\right) &= (0,0,3) \\ X\left(\frac{\pi}{2}, \pi\right) &= (0,0,1) \\ X\left(\frac{3\pi}{2}, \pi\right) &= (0,0,-1) \\ X\left(\frac{3\pi}{2}, 0\right) &= (0,0,-3). \end{aligned} \tag{13.38}$$

Because $X : [0,2\pi) \times [0,2\pi) \to M$ is one-to-one and onto, this exhausts all possible critical points. Now, let us look at $h^{-1}(\ell) \subseteq M$ for some values of $\ell \in \{-3,-2,-1,0,1,2,3\}$.



Notice that $h^{-1}(r)$ is a one-dimensional manifold for all $r \in R_f$ (this is depicted in the figure as a blue plane intersecting M). This manifold is diffeomorphic to one of the following three possibilities[26]

$$\varnothing, \qquad S^1, \qquad \text{or} \qquad S^1 \amalg S^1, \tag{13.39}$$

where $\amalg$ is the disjoint union. Also notice that $h^{-1}\big(h(c)\big)$ is *not* a one-dimensional manifold for all critical points $c \in C_f$ (this is depicted in the figure as a red plane intersecting M).

Exercise 13.40. This exercises refers to Example 13.29 and uses its notation. Incidentally, expressing the curves obtained from $h^{-1}(\ell)$ for various values of $\ell \in (-3,3)$ explicitly in terms of a single variable involves heavy use of the Implicit Function Theorem. For instance, one way to do this is to solve for s in the expression

$$X_3(s,t) = \big(2+\cos(t)\big)\sin(s) = \ell \tag{13.41}$$

and then plug this back into the expressions of X_1 and X_2. Of course, one can solve for $\sin(s)$ without any problem:

$$\sin(s) = \frac{\ell}{2+\cos(t)} \tag{13.42}$$

because the denominator never vanishes. Furthermore, only $\cos(s)$ appears in the other expressions for X_1 and X_2. One could therefore try to solve using trigonometric identities

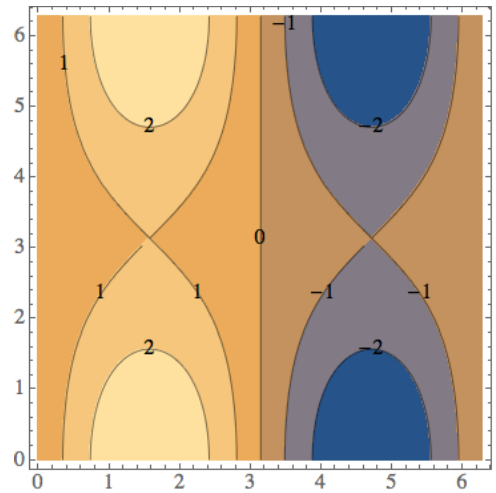
$$\cos(s) = \frac{\sqrt{4-\ell^2+2\cos(t)+\cos^2(t)}}{2+\cos t} \tag{13.43}$$

and plug this into the expressions to obtain a parametric curve description of the curve $h^{-1}(\ell) \subseteq M \subseteq \mathbb{R}^3$

$$(0,2\pi) \ni t \mapsto X^{\ell}(t) = \left(\sin(t), \sqrt{4-\ell^2+2\cos(t)+\cos^2(t)}, \frac{\ell}{2+\cos(t)}\right). \tag{13.44}$$

However, there is something wrong with my argument. It is not always possible to actually do this as the following contour plot illustrates showing the level sets of $X_3^{-1}(\ell)$.

[26]The empty set is a manifold of every dimension.



The horizontal axis here is for the s variable and the vertical axis is for the t variable.

(a) Let $\ell = 2$ and let $(s_0, t_0) \in (0, 2\pi) \times (0, 2\pi)$ be a point that satisfies $X_3(s_0, t_0) = \ell$. For which values of $t_0 \in (0, 2\pi)$ can you *not* find an open neighborhood $B \subseteq (0, 2\pi)$ containing t_0 and a smooth function $g : B \to \mathbb{R}$ such that $g(t_0) = s_0$ and $X_3\big(g(t), t\big) = \ell$ for all $t \in B$? In other words, for which points (s_0, t_0) can you *not* solve the equation $X_3(s, t) = 2$ for the variable s in terms of t near the point t_0 that contains the point s_0 in its image?

(b) Again, set $\ell = 2$. Pick any point (s_0, t_0) for which you *can* solve for one variable in terms of the other. Find the largest domain for which you can do this and for which your formula is well-defined.

Example 13.29 leads us directly to a deeper understanding of level sets of smooth functions from regular values and there relationship to lower-dimensional manifolds, but before stating this result, we describe the inverse function theorem for manifolds.

Theorem 13.45 (Inverse Function Theorem for Manifolds)**.** *Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be smooth m-dimensional manifolds, let $c \in N$, and let $f : N \to M$ be of class C^r on M. If $D_c f : T_c N \to T_{f(c)} M$ is an isomorphism of vector spaces, then there exists open sets $U \subseteq N$ and $V \subseteq M$ and a function $g : V \to U$ such that*

(a) $c \in U$ and $f(c) \in V$,

(b) $f : U \to V$ and $g : V \to U$ are inverses of each other,

(c) g is of class C^r on V with differential given by $D_{f(c)} g = (D_c f)^{-1}$.

Proof. Let $\Phi : N \cap B \xrightarrow{\cong} A$ be a coordinate system of N about c and $\Psi : M \cap Y \xrightarrow{\cong} X$ a coordinate system of M about $f(c)$. Here $A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^l, X \subseteq \mathbb{R}^m, Y \subseteq \mathbb{R}^k$ are all open. This uniquely defines a function $F : A \to X$ such that the diagram

$$\begin{tikzcd} M \cap Y \arrow[d, "\Psi"'] & N \cap B \arrow[l, "f"'] \arrow[d, "\Phi"] \\ X & A \arrow[l, "F"] \end{tikzcd} \tag{13.46}$$

commutes. Applying D to this diagram gives

$$\begin{tikzcd} T_{f(c)}M \arrow[d, "D_{f(c)}\Psi"'] & T_cN \arrow[l, "D_cf"'] \arrow[d, "D_c\Phi"] \\ \mathbb{R}^m & \mathbb{R}^n \arrow[l, "D_{\Phi(c)}F"] \end{tikzcd} \tag{13.47}$$

By assumption, $D_cf, D_c\Phi$, and $D_{f(c)}\Psi$ are all invertible linear transformations. Hence, $D_{\Phi(c)}F$ is invertible as well since it is given by

$$D_{\Phi(c)}F = (D_{f(c)}\Psi) \circ (D_cf) \circ (D_{\Phi(c)}F)^{-1} \tag{13.48}$$

since the diagram commutes and each linear transformation in this composition is invertible. In particular, this forces $m = n$. Therefore, by the ordinary Inverse Function Theorem (Theorem 8.11) and Exercise 8.42, there exist open sets $Z, W \subseteq \mathbb{R}^m$ and a function $G : Z \to W$ of class C^r that is an inverse of F. Since Ψ and Φ are smooth diffeomorphisms, the image of Z under Ψ^{-1} is an open subset of M. Hence, there exists an open set $V \subseteq \mathbb{R}^k$ such that $M \cap V = \Psi^{-1}(Z)$. Similarly, there exists an open set $U \subseteq \mathbb{R}^l$ such that $\Phi^{-1}(W) = N \cap U$. In other words, we have so far constructed the diagram

$$\begin{tikzcd} M \cap V \arrow[d, "\Psi"'] & N \cap U \arrow[d, "\Phi"] \\ Z \arrow[r, "G"'] & W \end{tikzcd} \tag{13.49}$$

Since Φ is invertible, there exists a unique map $g : M \cap V \to N \cap U$ such that the diagram

$$\begin{tikzcd} M \cap V \arrow[r, "g"] \arrow[d, "\Psi"'] & N \cap U \arrow[d, "\Phi"] \\ Z \arrow[r, "G"'] & W \end{tikzcd} \tag{13.50}$$

commutes, namely

$$g := \Phi^{-1} \circ G \circ \Psi. \tag{13.51}$$

This g satisfies all of the conditions of the theorem. ■

The following theorem is a generalization of Theorem 11.45 [8].

Theorem 13.52. *Let $M \subseteq \mathbb{R}^k$ and $N \subseteq \mathbb{R}^l$ be smooth m- and n-dimensional manifolds with $n \geq m$. Let $f : N \to M$ be a smooth function, and let $y \in M$ be a regular value. Then $f^{-1}(y) \subseteq N$ is a smooth manifold of dimension $n - m$.*

Proof. Let $c \in f^{-1}(y)$. The goal is to construct an open neighborhood around c and a chart on this neighborhood intersecting with $f^{-1}(y)$. Because y is a regular value, $D_cf : T_xN \to T_yM$ is onto. Let $K := \ker(D_cf)$ denote the kernel of this map, which is a vector subspace of T_cN. By the rank-nullity theorem (from linear algebra), $\dim K = n-m$. Choose an isomorphism $L : K \to \mathbb{R}^{n-m}$ and an extension $\tilde{L} : T_c\mathbb{R}^l \to \mathbb{R}^{n-m}$ of L (an infinite number of such choices are available—it does not matter what we choose). Such a function $\tilde{L}$ corresponds uniquely to a linear transformation $\mathcal{L} : \mathbb{R}^l \to \mathbb{R}^{n-m}$ given by ignoring the c. Then define $F : N \to M \times \mathbb{R}^{n-m}$ to be the function

$$N \ni x \mapsto F(x) := \big(f(x), \mathcal{L}(x)\big), \tag{13.53}$$

where the second component uses the fact that $N \subseteq \mathbb{R}^l$. F is differentiable for all $x \in N$ and its differential $D_cF : T_cN \to T_yM \oplus T_{\mathcal{L}(c)}\mathbb{R}^{n-m}$ at $c \in N$ is given by

$$T_cN \ni v_c \mapsto (D_cF)(v_c) = \big((D_cf)(v_c), L(v_c)\big). \tag{13.54}$$

D_cF has no non-trivial kernel. To see this, let $v_c \in T_cN$ satisfy $(D_cF)(v_c) = 0$. Then this means $(D_cf)(v_c) = 0$, but since L is non-vanishing on $\ker(D_cf) \setminus \{0\}$ and $L(v_c) = 0$ as well, this forces $v_c = 0$. Hence, D_cF is an isomorphism. By the Inverse Function Theorem for Manifolds, there exist open sets $U \subseteq \mathbb{R}^l$ and $V \subseteq \mathbb{R}^k \times \mathbb{R}^{n-m}$ such that $F : N \cap U \to (N \times \mathbb{R}^{n-m}) \cap V$ is a diffeomorphism. Furthermore, $F\big(f^{-1}(y) \cap U\big) = (\{y\} \times \mathbb{R}^{n-m}) \cap V$. In other words, F restricted to $f^{-1}(y) \cap U$ provides a coordinate system for $f^{-1}(y)$ about the point c proving that it is a manifold of dimension $n - m$. ∎

The following exercise comes from [9].

Exercise 13.55. Let $f, g : \mathbb{R}^3 \to \mathbb{R}$ be of class C^r. Under what conditions can you be sure that the solution se of the *system* of equations

$$\begin{aligned} f(x, y, z) &= 0 \\ g(x, y, z) &= 0 \end{aligned} \tag{13.56}$$

is a smooth curve (i.e. a one-dimensional manifold)?

Exercise 13.57. Let $M \subseteq \mathbb{R}^3$ be the graph of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2y^2, \tag{13.58}$$

let $\pi : \mathbb{R}^3 \to \mathbb{R}$ be the projection onto the third component, and let $h : M \to \mathbb{R}$ be the restriction of π to M (this is the height function on M).

(a) Explain why h is differentiable.

(b) Calculate $C_h \subseteq M$ and $R_h \subseteq \mathbb{R}$, the set of critical points and regular values of f, respectively.

(c) Is $h^{-1}(\ell)$ a manifold for all $\ell > 0$? What about for al $\ell < 0$? Do you expect these to be manifolds? Finally, what about for $\ell = 1$?

(d) Describe the set $h^{-1}(0)$ explicitly.

Exercise 13.59. Let $M \subseteq \mathbb{R}^3$ be the graph of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{R}^2 \ni (x, y) \mapsto f(x, y) := x^2, \tag{13.60}$$

let $\pi : \mathbb{R}^3 \to \mathbb{R}$ be the projection onto the third component, and let $h : M \to \mathbb{R}$ be the restriction of π to M (this is the height function on M).

(a) Explain why h is differentiable.

(b) Calculate $C_h \subseteq M$ and $R_h \subseteq \mathbb{R}$, the set of critical points and regular values of f, respectively.

(c) Is $h^{-1}(\ell)$ a manifold for all $\ell > 0$? What about for al $\ell < 0$? Do you expect these to be manifolds? Finally, what about for $\ell = 1$?

(d) Describe the set $h^{-1}(0)$ explicitly. Does this result contradict Theorem 13.52? Explain.

You might wonder why we always look at height functions to analyze problems like this. This is unnecessary as the following examples illustrate.

Exercise 13.61. Let $S^1 \subseteq \mathbb{R}^2$ be the unit circle and let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by sending $(x, y) \ni \mathbb{R}^2$ to $f(x, y) := x^2 - y^2$. Let $h : S^1 \to \mathbb{R}$ be the restriction of f to S^1.

(a) Explain why h is differentiable.

(b) Calculate $C_h \subseteq S^1$ and $R_h \subseteq \mathbb{R}$, the set of critical points and regular values of f, respectively.

(c) For what values of ℓ is $h^{-1}(\ell)$ a manifold?

Exercise 13.62. Let $S^2 \subseteq \mathbb{R}^3$ be the unit 2-sphere and let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function defined by sending $(x, y, z) \ni \mathbb{R}^2$ to $f(x, y, z) := x^2 + y^2 - z^2$. Let $h : S^2 \to \mathbb{R}$ be the restriction of f to S^2.

(a) Explain why h is differentiable.

(b) Calculate $C_h \subseteq S^2$ and $R_h \subseteq \mathbb{R}$, the set of critical points and regular values of f, respectively.

(c) For what values of ℓ is $h^{-1}(\ell)$ a manifold? What manifold is it? Classify the possibilities into cases.

We know that if we have a parametrization of a manifold, we can calculate the tangent space at a point by using the parametrization and simply calculating it. However, this task is not so simple when we have manifolds described by some level set. To do this, we introduce normal vectors to tangent spaces and then prove a theorem relating the tangent space to a level set with the normal space.

Definition 13.63. Let $c \in \mathbb{R}^k$ and define an inner product on $T_c\mathbb{R}^k$ by

$$T_c\mathbb{R}^k \times T_c\mathbb{R}^k \ni (u_c, v_c) \mapsto \langle u_c, v_c \rangle_c := \langle u, v \rangle, \tag{13.64}$$

where the latter notation is the usual Euclidean inner product on $\mathbb{R}^k$. Similarly, if $M \subseteq \mathbb{R}^k$ is an m-dimensional manifold and $c \in M$, then $\langle \cdot\,, \cdot \rangle_c$ restricts to T_cM so that T_cM is an inner product space.

Definition 13.65. Let $M \subseteq \mathbb{R}^k$ be a manifold of dimension m and let $c \in M$. The set

$$\nu_c M := \left\{ v_c \in T_c\mathbb{R}^k \ : \ \langle v_c, u_c \rangle_c = 0 \ \forall \ u_c \in T_cM \right\} \tag{13.66}$$

is called the <u>*normal space*</u> of M in $\mathbb{R}^k$ at c (or sometimes called the <u>*space of normal vectors to M in $\mathbb{R}^k$ at c*</u>).

Theorem 13.67. *Let $M \subseteq \mathbb{R}^k$ and let $c \in M$. Then*

$$T_c\mathbb{R}^k = T_cM \oplus \nu_cM, \tag{13.68}$$

i.e. for any vector $v_c \in T_c\mathbb{R}^k$, there exist unique vectors $u_c \in T_cM$ and $w_c \in \nu_cM$ such that $v_c = u_c + w_c$.

Exercise 13.69. Prove Theorem 13.67.

The reason for the specific phrasing in the definition of normal vectors is because the definition depends on the ambient space. To be more consistent, we define the notion of a submanifold.

Definition 13.70. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold. A subset $N \subseteq M$ is an <u>*n-dimensional submanifold of M*</u> iff for every point $c \in N$, there exists a coordinate system

$$U \underset{\psi}{\overset{\varphi}{\rightleftarrows}} V \tag{13.71}$$

of M about c such that

$$\varphi(U \cap N) = V \cap \big(\mathbb{R}^n \times \{0\}\big), \tag{13.72}$$

where the latter set $\mathbb{R}^n \times \{0\} \subseteq \mathbb{R}^k$ is explicitly described as

$$\mathbb{R}^n \times \{0\} := \{y \in \mathbb{R}^k \ : \ y_{n+1}, y_{n+2}, \dots, y_k = 0\}. \tag{13.73}$$

Exercise 13.74. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold, let $N \subseteq M$ be an n-dimensional submanifold of M, and let $c \in N$.

(a) Prove that the inclusion map $N \hookrightarrow M$ is differentiable.

(b) Show that $T_cN \subseteq T_cM$.

Definition 13.75. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold, let $N \subseteq M$ be an n-dimensional submanifold of M, and let $c \in N$. The set

$$\nu_c^M N := \big\{v_c \in T_cM \ : \ \langle v_c, u_c\rangle_c = 0 \ \forall \ u_c \in T_cN\big\} \tag{13.76}$$

is called the <u>*normal space*</u> of N in M at c (or sometimes called the <u>*space of normal vectors to N in M at c*</u>).

Unlike the tangent space, the normal space depends on the ambient space in which the manifold lives. To be consistent with earlier notation, for a manifold $M \subseteq \mathbb{R}^k$, the space of normal vectors to M at c might also be written as $\nu_c^{\mathbb{R}^l} M$ instead of ν_cM. However, if it is clear from context, we may leave out this additional superscript.

Theorem 13.77. *Let M and N be m- and n-dimensional manifolds respectively, let $f : N \to M$ be differentiable, and let $y \in R_f \subseteq M$ be a regular value of f, denote $S := f^{-1}(y)$, and let $x \in S$. Then*

$$\ker\Big(T_xN \xrightarrow{D_xf} T_yM\Big) = T_xS, \tag{13.78}$$

i.e.

$$\nu_x^N S \xrightarrow{D_xf} T_yM \tag{13.79}$$

is an isomorphism.

Proof. Applying D to the commutative diagram

$$\begin{tikzcd} N \arrow[d, "f"'] & S \arrow[l, hook'] \arrow[d, "f|_S"] \\ M & \{y\} \arrow[l, hook'] \end{tikzcd} \tag{13.80}$$

of differentiable maps induces the commutative diagram

$$\begin{tikzcd} T_xN \arrow[d, "D_xf"'] & T_xS \arrow[l, hook'] \arrow[d, "D_xf|_{T_xS=0}"] \\ T_yM & T_y\{y\} \arrow[l, hook'] \end{tikzcd} \tag{13.81}$$

of linear transformations by Exercise 13.74 (and the Chain Rule). Since $T_y\{y\}$ is just a zero-dimensional subspace, $(D_xf)(T_xS) = \{0\}$ as a subspace of T_yM. This shows that $T_xS \subseteq \ker(D_xf)$. To see the other containment, first note that by Theorem 13.52, $\dim(T_xS) = n - m$. Since D_xf is surjective and $T_xN = T_xS \oplus \nu_x^N S$ it follows that $(D_xf)(\nu_x^N S) = T_yM$. Because $\dim(\nu_x^N S) = m$, this shows that $D_xf : \nu_x^N S \to T_yM$ is an isomorphism and $\ker(D_xf) = T_xS$. ∎

Remark 13.82. By popular demand, I've decided to include the following problem, which could have been given in Lecture 6. It is contained in a remark because it is tangential to what we are doing (unless I change my mind sometime later in the future).

Exercise 13.83. Recall from Lecture 6 that given a smooth function $f : \mathbb{R}^n \to \mathbb{R}^m$, the differential of f at various points defines a smooth function $D_\square f : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$. The *differential of this* is the function $D_\square D_\square f : \mathbb{R}^n \to \mathrm{Hom}\big(\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n), \mathbb{R}^n\big)$, which we denoted by $D^2_\square f$ instead (since the input variable for $\square$ is the same). Such a function looks complicated, but it actually isn't thanks to a wonderful relationship between linear transformations and tensor products (an analogous relationship holds for functions and Cartesian products).

(a) Let U, V, W be finite-dimensional vector spaces. Show that there is a canonical isomorphism of vector spaces

$$\mathrm{Hom}\big(\mathrm{Hom}(U, V), W\big) \cong \mathrm{BiHom}\big(U, V \times W\big), \tag{13.84}$$

where $\mathrm{BiHom}\big(U, V \times W\big)$ is the set of functions $U \xleftarrow{L} V \times W$ that are bilinear, i.e.

$$L(v_1 + cv_2, w) = L(v_1, w) + cL(v_2, w) \qquad \& \qquad L(v, w_1 + cw_2) = L(v, w_1) + cL(v, w_2) \tag{13.85}$$

for all $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, and $c \in \mathbb{R}$.

It is a fact (I do not expect you to prove this) that there exists a unique (up to canonical isomorphism) vector space $V \otimes W$ satisfying the condition that there is a one-to-one correspondence between $\mathrm{BiHom}\big(U, V \times W\big)$ and $\mathrm{Hom}\big(U, V \otimes W\big)$.[27]

[27]Stating this more precisely goes way beyond what we should be focusing on. $\otimes$ is known as the tensor product of vector spaces. My personal understanding of it involves some advanced diagram theory.

(b) This fact implies that $D^2_{\Box}f : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n \otimes \mathbb{R}^n)$, which means that for every $x \in \mathbb{R}^n$, we get a linear transformation $D^2_x f : \mathbb{R}^n \otimes \mathbb{R}^n \to \mathbb{R}^m$. Concretely, this means $D^2_x f$ can be viewed as a function (the same notation is used abusively) $D^2_x f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ that is bilinear. Prove that

$$(D^2_x f_k)(e_i, e_j) = \partial_i \partial_j f_k(x) \tag{13.86}$$

for all $i, j \in \{1, \dots, n\}$ and for all $k \in \{1, \dots, m\}$.

Notice that when $m = 1$, this gives n^2 such functions $\partial_i \partial_j f$ (technically, because f is smooth, it gives

$$\frac{n^2 - n}{2} + n = \frac{n(n+1)}{2} \tag{13.87}$$

independent functions because partial derivatives commute). Using the standard Euclidean basis,[28] one can put these functions in the form of a matrix

$$Hf := \begin{bmatrix} \partial_1\partial_1 f & \partial_1\partial_2 f & \cdots & \partial_1\partial_n f \\ \partial_2\partial_1 f & \partial_2\partial_2 f & \cdots & \partial_1\partial_2 f \\ \vdots & \vdots & \ddots & \vdots \\ \partial_n\partial_1 f & \partial_n\partial_2 f & \cdots & \partial_n\partial_n f \end{bmatrix} \tag{13.88}$$

known as the <u>*Hessian matrix of f*</u>. However, one uses the Euclidean inner product to construct this matrix. It is more natural to think of $D^2 f$ simply as a multilinear function $D^2 f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Such an object is known as a 2-tensor.

Exercise 13.89. Let $f : N \to M$ be a smooth map between smooth manifolds. Let $\Gamma := \big\{(x, y) \in N \times M \ : \ y = f(x)\big\}$ be the graph of f.

(a) Prove that Γ is a smooth submanifold of $N \times M$.

(b) Prove that the tangent space

$$T_{(x,y)}\Gamma \subseteq T_x N \times T_y M \tag{13.90}$$

is equal to the graph of the linear transformation $D_x f$.

Level I problems.
From these notes: Exercises 13.9, 13.25, 13.55, 13.61

Level II problems.
From these notes: Exercises 13.13, 13.40, 13.57, 13.59, 13.62, 13.69, 13.74

Level III problems.
From these notes: Exercises 13.14, 13.83

[28] When $m = 1$, $D^2 f$ is originally of the form $D^2_x f : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}, \mathbb{R}^n)$. The codomain is the dual space of $\mathbb{R}^n$ (by definition). Since $\mathbb{R}^n$ has an inner product, there is an isomorphism $\varphi : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}, \mathbb{R}^n)$ given by sending the vector $v \in \mathbb{R}^n$ to the linear function $\langle v, \ \cdot \ \rangle : \mathbb{R}^n \to \mathbb{R}$. Therefore, *using* this isomorphism, we get a linear transformation $\varphi \circ D^2_x f : \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}, \mathbb{R}^n) \to \mathbb{R}^n$. This linear transformation is the Hessian of f at x.

14 March 28: Brief recap (buffer for exam review)

Level I problems.
From Spivak [13]:
From these notes: Exercises

Level II problems.
From Spivak [13]:
From these notes: Exercises

Do not forget that there is a midterm on March 30th covering Lectures 8, 10, 11, 12, and 13!

15 April 4: The Index of a Vector Field in Euclidean Space

In the study of vector fields in Euclidean space, we saw how they gave rise to differential equations. Without solving these differential equations, one can deduce a lot of the behavior of such a system by examining the critical points of the vector field and the integral curves near such critical points. In particular, certain aspects of the integral curves are unchanged under small perturbations of the vector fields. For example, under a perturbation of a source or a sink, the critical point remains a source or a sink, respectively. We would like to have a rigorous and mathematically precise way of distinguishing different isolated critical points of vector fields. Furthermore, we would like this procedure to be robust enough to work on vector fields on manifolds as well (which we must also define). To do this, we can enclose a given critical point in a disk. The boundary of this disk is a sphere, and the vector field restricted to this sphere defines a function to the standard sphere of that same dimension. The set of continuous functions $f : S^n \to S^n$ from a sphere to itself is actually a topological space[29] in its own right (though it is not a finite-dimensional manifold in general). In fact, it is a disconnected space with $\mathbb{Z}$-many connected (in fact, path-connected) components. Therefore, a continuous function $f : S^n \to S^n$ must be in one of these components. This component is an invariant of f that does not change when f is perturbed by continuous deformations. It is known as the *degree* of f. We will first study this notion for vector fields in the Euclidean plane before moving on to the general study. Most of the material from the next two lectures can be found in Milnor's book [8].

Definition 15.1. Let V be a vector space of dimension $m \in \{0\} \cup \mathbb{N}$. Two bases on V, which are described by two isomorphisms $\Psi, \Phi : \mathbb{R}^m \to V$, are said to have the <u>*same orienatation*</u> off $\det(\Psi^{-1} \circ \Phi) > 0$. Otherwise, they have the <u>*opposite orientation*</u>.

Having the same orientation on a vector space V is an equivalence relation on the set of isomorphisms from $\mathbb{R}^m$ to V.

Definition 15.2. Let V be a vector space of dimension $m \in \{0\} \cup \mathbb{N}$. An <u>*orientation*</u> on V is a choice of an equivalence class of isomorphisms $\mathbb{R}^m \to V$. Let V and W be two m-dimensional vector spaces with orientations described by equivalence classes of isomorphisms $\Phi : \mathbb{R}^m \to V$ and $\Psi : \mathbb{R}^m \to W$. A linear isomorphism $S : V \to W$ is <u>*orientation-preserving*</u> iff $\det(\Psi^{-1} \circ S \circ \Phi) > 0$. Otherwise, it is <u>*orientation-reversing*</u>.

Note that the notion of orientation-preserving/reversing does not depend on the choice of isomorphisms Φ and Ψ and only depends on the orientations on V and W.

Definition 15.3. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold. An <u>*orientation on M*</u> consists of a choice of orientations on T_cM for all $c \in M$ satisfying the following condition: for every $c \in M$, there exists an open set $U \subseteq \mathbb{R}^m$, an open set $V \subseteq \mathbb{R}^k$, and a parametrization $\varphi : U \to M \cap V$ of M about c such that $D_x\varphi : T_x\mathbb{R}^m \to T_{\varphi(x)}M$ is orientation-preserving for all $x \in U$ (Here $T_x\mathbb{R}^m$ is equipped with the standard basis coming from $\mathbb{R}^m$). A manifold is <u>*orientable*</u> iff there

[29] We briefly talked about topological spaces last semester. We do not need to know the definition to get an idea of what's going on. Simply use your general idea of what a space is and what it means to be connected or disconnected.

exists an orientation on M. An *oriented manifold* is a manifold equipped with an orientation. A diffeomorphism $f : N \to M$ from an oriented m-dimensional manifold N to an oriented m-dimensional manifold M is *orientation-preserving* iff $D_x f : T_x N \to T_{f(x)} M$ is orientation-preserving for all $x \in N$.

The idea behind this definition is that an orientation on a manifold is a choice of smoothly varying orientations on all of the tangent spaces. Note the subtle difference between a manifold *being* orientable versus a manifold *having* an orientation. The first is a property while the second is extra datum.

Exercise 15.4. Let M be a manifold. Prove that M is path-connected if and only if M is connected.

Exercise 15.5. Let M be an orientable manifold.

(a) Prove that if M is connected, then there exists exactly two possible orientations on M.

(b) What happens if M is not connected?

Example 15.6. The m-dimensional sphere $S^m \subseteq \mathbb{R}^{m+1}$ is orientable. In the process of describing an orientation, we will prove that S^m is a manifold and also construct its tangent space. Let

$$V_S := \mathbb{R}^m \times (-\infty, 1) \qquad \& \qquad V_N := \mathbb{R}^m \times (-1, \infty) \tag{15.7}$$

so that $S^m \cap V_S$ is all of S^m except the "north pole" and $S^m \cap V_N$ is all of S^m except the "south pole." Define a coordinate chart on $S^m \cap V_S$ by[30]

$$\begin{aligned} S^m \cap V_S &\xrightarrow{\psi_S} \mathbb{R}^m \\ (x_1, x_2, \dots, x_m, x_{m+1}) &\mapsto \left(\frac{-x_1}{1 - x_{m+1}}, \frac{x_2}{1 - x_{m+1}}, \dots, \frac{x_m}{1 - x_{m+1}} \right). \end{aligned} \tag{15.8}$$

ψ_S is smooth on all of V_S by simply extending the formula algebraically. This is almost the stereographic projection with the minor difference that the first coordinate has an additional minus sign. This minus sign will be so that we get the correct orientation on the sphere later on towards the end of this example. The inverse of this coordinate chart is[31]

$$\begin{aligned} \mathbb{R}^m &\xrightarrow{\varphi_S} S^{m-1} \cap V_S \\ (y_1, y_2, \dots, y_m) &\mapsto \left(\frac{-2y_1}{1 + R(y)^2}, \frac{2y_2}{1 + R(y)^2}, \dots, \frac{2y_m}{1 + R(y)^2}, \frac{-1 + R(y)^2}{1 + R(y)^2} \right), \end{aligned} \tag{15.9}$$

where

$$R(y) \equiv R(y_1, \dots, y_m) := \sqrt{\sum_{i=1}^m y_i^2}. \tag{15.10}$$

[30] This is obtained by drawing the straight line from the north pole through the point $(-x_1, x_2, \dots, x_m, x_{m+1})$ on S^m and assigning the point that intersects the $\mathbb{R}^m \times \{0\}$ plane in $\mathbb{R}^{m+1}$.

[31] This is obtained by squaring both sides of the equality $y_i = \frac{x_i}{1 - x_{m+1}}$ and summing over all $i \in \{1, \dots, m\}$. This gives $\sum_{i=1}^m y_i^2 = \frac{\sum_{i=1}^m x_i^2}{(1 - x_{m+1})^2}$, but since $(x_1, \dots, x_m, x_{m+1})$ must lie on the unit sphere, $\sum_{i=1}^m x_i^2 = 1 - x_{m+1}^2$ which, upon canceling a common $1 - x_{m+1}$, gives $\sum_{i=1}^m y_i^2 = \frac{1 + x_{m+1}}{1 - x_{m+1}}$.

φ_S is smooth because its components are the ratios of smooth functions and the denominator never vanishes. For later calculations, we compute the partial derivatives of these components, which are mostly of the form

$$\partial_j \left(\frac{2y_k}{1 + \sum_{i=1}^m y_i^2} \right) = \frac{2\delta_{jk} \left(1 + \sum_{i=1}^m y_i^2\right) - 4y_j y_k}{\left(1 + \sum_{i=1}^m y_i^2\right)^2} \tag{15.11}$$

except for the last term which is

$$\partial_j \left(\frac{-1 + \sum_{i=1}^m y_i^2}{1 + \sum_{i=1}^m y_i^2} \right) = \frac{2y_j \left(1 + \sum_{i=1}^m y_i^2\right) - 2y_j \left(-1 + \sum_{i=1}^m y_i^2\right)}{\left(1 + \sum_{i=1}^m y_i^2\right)^2} = \frac{4y_j}{\left(1 + \sum_{i=1}^m y_i^2\right)^2}. \tag{15.12}$$

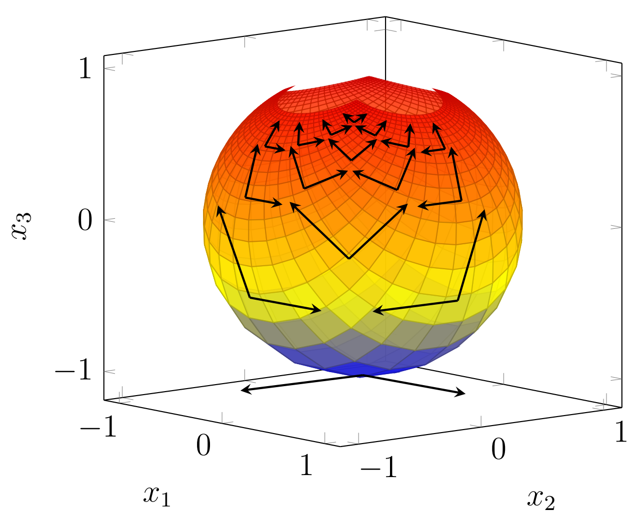
Since φ_S is a parametrization of $S^m \cap V_S$, it can be used to describe the tangent space to all points in $S^m \cap V_S$. The differential of φ_S at $y := (y_1, \ldots, y_m) \in \mathbb{R}^m$ with respect to the standard Euclidean basis in $\mathbb{R}^m$ viewed as a linear transformation $D_y\varphi_S : T_y\mathbb{R}^m \to T_{\varphi_S(y)}\mathbb{R}^k$ is given by

$$[D_y\varphi_S] = \frac{1}{\left(1 + R(y)^2\right)^2} \begin{bmatrix} 4y_1^2 - 2\left(1 + R(y)^2\right) & 4y_2y_1 & \cdots & 4y_my_1 \\ -4y_1y_2 & 2\left(1 + R(y)^2\right) - 4y_2^2 & \cdots & -4y_my_2 \\ \vdots & \vdots & \ddots & \vdots \\ -4y_1y_m & -4y_2y_m & \cdots & 2\left(1 + R(y)^2\right) - 4y_m^2 \\ 4y_1 & 4y_2 & \cdots & 4y_m \end{bmatrix} \tag{15.13}$$

The columns of this matrix form a basis for the tangent space to S^m at the point $\varphi_S(y)$. In the case that $m = 2$, this simplifies to

$$[D_y\varphi_S] = \frac{1}{\left(1 + y_1^2 + y_2^2\right)^2} \begin{bmatrix} 2(y_1^2 - y_2^2 - 1) & 4y_2y_1 \\ -4y_2y_1 & 2(1 + y_1^2 - y_2^2) \\ 4y_1 & 4y_2 \end{bmatrix}. \tag{15.14}$$

A visualization of the image of $[-3, 3] \times [-3, 3] \subseteq \mathbb{R}^2$ under the parametrization φ_S together with a plot of some of the columns of the matrix $[D_y\varphi_S]$ on this image looks like the following plot (the vectors are oriented in such a way so that appending the outward normal to the sphere gives a right-handed coordinate system).



Similarly, working with V_N, we get the coordinate chart

$$\begin{aligned} S^m \cap V_N &\xrightarrow{\psi_N} \mathbb{R}^m \\ (x_1, x_2, \dots, x_m, x_{m+1}) &\mapsto \left(\frac{x_1}{1 + x_{m+1}}, \frac{x_2}{1 + x_{m+1}}, \dots, \frac{x_m}{1 + x_{m+1}} \right). \end{aligned} \tag{15.15}$$

The inverse of this coordinate chart is (we use z's to distinguish these coordinates from the y's)

$$\begin{aligned} \mathbb{R}^m &\xrightarrow{\varphi_N} S^{m-1} \cap V_N \\ (z_1, z_2, \dots, z_m) &\mapsto \left(\frac{2z_1}{1 + R(z)^2}, \frac{2z_2}{1 + R(z)^2}, \dots, \frac{2z_m}{1 + R(z)^2}, \frac{1 - R(z)^2}{1 + R(z)^2} \right), \end{aligned} \tag{15.16}$$

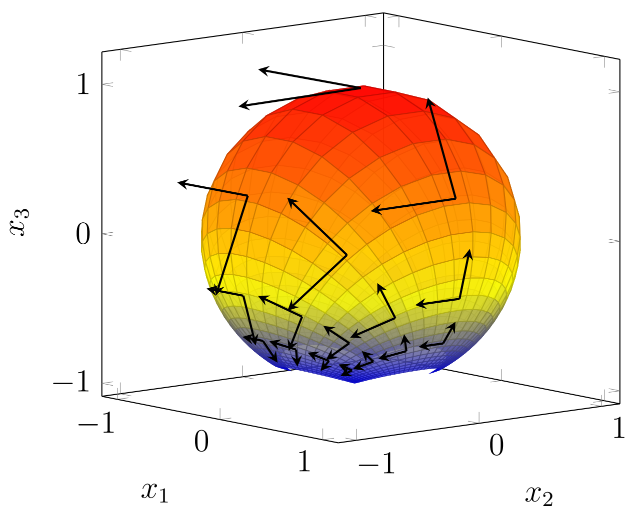
The differential of φ_N at $z \in \mathbb{R}^m$ is

$$[D_z \varphi_N] = \frac{1}{\left(1 + R(z)^2\right)^2} \begin{bmatrix} 2\left(1 + R(z)^2\right) - 4z_1^2 & -4z_2 z_1 & \cdots & -4z_m z_1 \\ -4z_1 z_2 & 2\left(1 + R(z)^2\right) - 4z_2^2 & \cdots & -4z_m z_2 \\ \vdots & \vdots & \ddots & \vdots \\ -4z_1 z_m & -4z_2 z_m & \cdots & 2\left(1 + R(z)^2\right) - 4z_m^2 \\ -4z_1 & -4z_2 & \cdots & -4z_m \end{bmatrix} \tag{15.17}$$

In the case that $m = 2$, this simplifies to

$$[D_z \varphi_N] = \frac{1}{\left(1 + z_1^2 + z_2^2\right)^2} \begin{bmatrix} 2(1 - z_1^2 + z_2^2) & -4z_2 z_1 \\ -4z_2 z_1 & 2(1 + z_1^2 - z_2^2) \\ -4z_1 & -4z_2 \end{bmatrix}. \tag{15.18}$$

A visualization of the image of $[-3, 3] \times [-3, 3] \subseteq \mathbb{R}^3$ under the parametrization φ_N together with a plot of some of the columns of the matrix $[D_z \varphi_N]$ on this image looks like the following plot.



To check that the two induced orientations agree on all of S^m, we must check that the differential of the coordinate transformation[32] $\psi_N \circ \varphi_S : \mathbb{R}^m \setminus \{0\} \to \mathbb{R}^m \setminus \{0\}$ is orientation-preserving. By

[32]The reason we remove the point 0 is because $\varphi_S(0)$ is not in the domain of ψ_N.

plugging in the definitions, this gives

$$
\begin{aligned}
\psi_N\big(\varphi_S(y)\big) &= \psi_N\left(\frac{-2y_1}{1+R(y)^2}, \frac{2y_2}{1+R(y)^2}, \dots, \frac{2y_m}{1+R(y)^2}, \frac{-1+R(y)^2}{1+R(y)^2}\right) \\
&= \left(\frac{-y_1}{R(y)^2}, \frac{y_2}{R(y)^2}, \dots, \frac{y_m}{R(y)^2}\right)
\end{aligned}
\tag{15.19}
$$

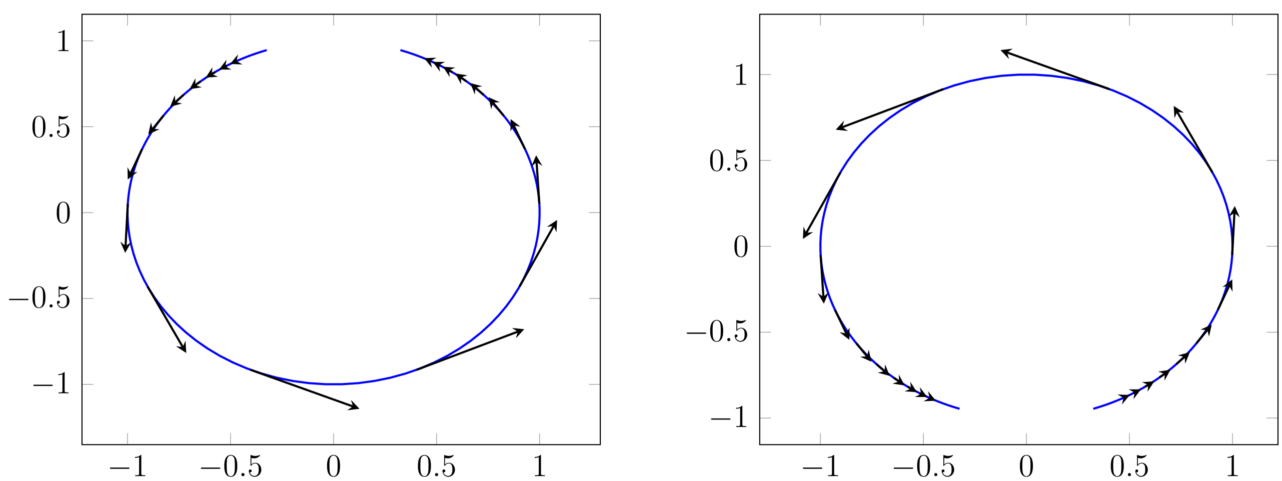
and the differential of this map is

$$
\big[D_y(\psi_N \circ \varphi_S)\big] = \frac{1}{R(y)^4}\begin{bmatrix} 2y_1^2 - R(y)^2 & 2y_1y_2 & \cdots & 2y_1y_m \\ -2y_2y_1 & R(y)^2 - 2y_2^2 & \cdots & -2y_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ -2y_my_1 & -2y_my_2 & \cdots & R(y)^2 - 2y_m^2 \end{bmatrix}
\tag{15.20}
$$

We do not need to calculate the determinant of this matrix at an arbitrary point $y \in \mathbb{R}^m \setminus \{0\}$ to check if it is orientation-preserving. By continuity and the fact that the determinant can never vanish, it suffices to calculate it at a single point. Therefore, calculate it at, say, $y = (1, 0, \dots, 0)$. At this point, the differential becomes

$$
\big[D_{(1,0,\dots,0)}(\psi_N \circ \varphi_S)\big] = \mathbb{1}_m,
\tag{15.21}
$$

the identity matrix, which has determinant 1. For consistency, let us also check what this gives when $m = 1$. For $\varphi_S : \mathbb{R} \to S^1 \cap V_S$ and $\varphi_N : \mathbb{R} \to S^1 \cap V_N$, we get



respectively, which agrees with the standard counter-clockwise orientation of S^1.

Exercise 15.22. Prove that if M and N are orientable manifolds then $M \times N$ is orientable.

Exercise 15.23. Let $M \subseteq \mathbb{R}^k$ and N be m-dimensional manifolds with N compact and let $f : N \to M$ be a smooth map.

(a) Prove that $f^{-1}(r)$ is a finite set of points (possibly empty) for every regular value $r \in R_f$.

(b) For each $r \in R_f$, let $\#f^{-1}(r)$ denote the number of points in $f^{-1}(r)$. Prove that for every $r \in R_f$, there exists an open set $U \subseteq \mathbb{R}^k$ with $r \in U$ such that

$$
\#f^{-1}(y) = \#f^{-1}(r) \qquad \forall\, y \in R_f \cap U,
\tag{15.24}
$$

i.e. prove that $\#f^{-1} : R_f \to \mathbb{Z}$ is a locally constant function.

Remark 15.25. Although $\#f^{-1} : R_f \to \mathbb{Z}$ is a locally constant function, it is *not* necessarily constant, even if N is assumed to be connected. However, if N is connected and one passes to $\mathbb{Z}_2$, then the composition (note that we restrict ourselves now to regular values that are also points in M) $R_f \cap M \xrightarrow{\#f^{-1}} \mathbb{Z} \to \mathbb{Z}_2$ *is* constant. It would take us a bit far afield to prove this, but its proof can be found in Chapter 4 of Milnor's book [8]. Two facts that are sufficient to prove this result are Sard's theorem and the existence and uniqueness to solutions of differential equations. We will address the latter soon and the former only if we have time later on.

Definition 15.26. Let $M \subseteq \mathbb{R}^k$ and N be oriented m-dimensional manifolds with N compact. Let $f : N \to M$ be a smooth map and let $x \in N$ be a regular *point* of f (so that $D_x f : T_x N \to T_r M$ is an isomorphism). The *sign* of $D_x f$ is the function

$$\operatorname{sign}(D_x f) := \begin{cases} +1 & \text{if } D_x f \text{ is orientation-preserving} \\ -1 & \text{if } D_x f \text{ is orientation-reversing.} \end{cases} \tag{15.27}$$

For a regular *value* $r \in R_f$, the *degree of f at r* is

$$\deg(f; r) := \sum_{x \in f^{-1}(r)} \operatorname{sign}(D_x f). \tag{15.28}$$

Since $\det(D_x f)$ never vanishes at the set of regular points and by Exercise 15.23, $\deg(f; r)$ is well-defined and finite. Note that it can also be defined as

$$\operatorname{sign}(D_x f) = \frac{\det([D_x f])}{\big|\det([D_x f])\big|}, \tag{15.29}$$

where the matrix $[D_x f]$ is evaluated using representative ordered bases at $x \in N$ and $f(x) = r$ in M. However, one must be careful to not miscalculate this determinant using the standard basis in which the manifolds are embedded in.

Exercise 15.30. Using the notation from Definition 15.26, prove that $\deg(f;\ \cdot\ ) : R_f \to \mathbb{Z}$ is a locally constant function, i.e. for every $r \in R_f$, there exists an open set $U \subseteq \mathbb{R}^k$ with $r \in U$ such that

$$\deg(f; y) = \deg(f; r) \qquad \forall\, y \in R_f \cap U. \tag{15.31}$$

When the codomain manifold N is connected, a more powerful result holds.

Theorem 15.32. *Let M and N be oriented m-dimensional manifolds with N compact and M connected and let $f : N \to M$ be a smooth map. Then $\deg(f;\ \cdot\ ) : R_f \cap M \to \mathbb{Z}$ is a constant function.*

Proof. The proof of this theorem is a bit involved, but not too difficult. The exposition in Chapter 5 of Milnor's book is simple and enlightening [8]. Time permitting, we might revisit this result and prove it, though it involves defining manifolds with boundaries, homotopies, and some important and non-trivial results about critical points and measure zero sets (Sard's Theorem). ■

This theorem allows the following definition to be made.

Definition 15.33. Let M and N be oriented m-dimensional manifolds with N compact and M connected and let $f : N \to M$ be a smooth map. The *degree of f* is the number

$$\operatorname{def}(f) := \deg(f; r) \tag{15.34}$$

for any regular value $r \in R_f \cap M$.

We will use the concept of degrees to study vector fields in Euclidean space. Before stating the definition, we provide some intuition.

Definition 15.35. Let $A \subseteq \mathbb{R}^m$ be open and let $V : A \to \mathbb{R}^m$ be a smooth vector field. A critical point c of V is *isolated* iff there exists an $\epsilon > 0$ such that V restricted to the closed disc $\overline{V_\epsilon(c)}$ vanishes nowhere except at c.

We have dealt with several vector fields with isolated critical earlier in this course. In all of these cases, by continuity of vector fields, we know there exists an open disc around every critical point in which the vector field vanishes nowhere (except at that critical point). The boundary of this disc is a sphere (of an appropriate dimension) and the vector field defines a function from this sphere to a sphere of the same dimension by rescaling the vector field. The degree of this function encodes some of the information of the critical point and, as we will see later, information about the topology of the space on which the vector field is defined. However, to have a well-defined quantity independent of the choice of small neighborhood around the critical point, we should provide sufficient reason to know that the actual choice of the neighborhood does not alter the degree. To properly make this statement mathematically, we must introduce the notion of a homotopy.

Definition 15.36. Let M and N be manifolds and let $f, g : N \to M$ be two smooth functions. f is *smoothly homotopic* to g iff there exists a smooth function $H : [0,1] \times N \to M$ such that

$$H(0, \,\cdot\,) = f \qquad \& \qquad H(1, \,\cdot\,) = g. \tag{15.37}$$

H is known as a *smooth homotopy* from f to g.

A smooth homotopy as above will often be depicted as

$$\begin{tikzcd} M & N \arrow[l, bend right, "g"'] \arrow[l, bend left, "f"] \arrow[l, phantom, "\Downarrow H"] \end{tikzcd} \tag{15.38}$$

because it provides a smooth path

$$\begin{tikzcd} f(x) \arrow[d, rightsquigarrow, "H(\,\cdot\,{,}x)"] \\ g(x) \end{tikzcd} \tag{15.39}$$

in M for every $x \in N$ from $f(x)$ to $g(x)$. In fact, a smooth homotopy can be viewed as a function $H : [0,1] \to M^N$, i.e. a path of functions from N to M though defining smoothness of H from this perspective is tricky (in fact, any reasonable definition of smoothness here reduces to $H : [0,1] \times N \to M$ being smooth anyway, so there is no point in complicating matters).

Example 15.40. Consider the two functions $\mathrm{id}_{S^1} : S^1 \to S^1$, the identity, and $R : S^1 \to S^1$, rotation by π radians. Then these two smooth functions are smoothly homotopic and a homotopy H from id_{S^1} to R is given explicitly by restricting the family of linear transformations on $\mathbb{R}^2$ given by

$$[0,1] \times \mathbb{R}^2 \ni (t,x,y) \mapsto H(t,x,y) := \big(x\cos(\pi t) - y\sin(\pi t), x\sin(\pi t) + y\cos(\pi t)\big) \tag{15.41}$$

to S^1, i.e. rotations by arbitrary angles. Setting $t = 0$ gives the identity and setting $t = 1$ gives the rotation by π radians.

Exercise 15.42. Prove that the two maps ϕ (use different notation to distinguish the two) in parts (c) and (d) of Example 15.59 are smoothly homotopic.

Theorem 15.43. *Smooth homotopy provides an equivalence relation on the set of smooth smooth functions from a manifold N to a manifold M.*

Proof. One must prove the three defining assumptions of an equivalence relation.

(a) Every smooth map $f : N \to M$ is smooth homotopic to itself through the constant homotopy. This proves reflexivity.

(b) Let $H : f \Rightarrow g$ be a smooth homotopy from $f : N \to M$ to $g : N \to M$. Then $\overline{H} : [0,1] \times N \to M$ defined by

$$[0,1] \times N \ni (t,x) \mapsto \overline{H}(t,x) := H(1-t,x) \tag{15.44}$$

is a smooth homotopy from g to f. This proves symmetry.

(c) Let $H : f \Rightarrow g$ and $G : g \Rightarrow h$ be two smooth homotopies. By Problem 2.26 in Spivak [13], there exists a smooth function $p : \mathbb{R} \to [0,1]$ such that $p(t) = 0$ for all $t \leq 0$ and $p(t) = 1$ for all $t \geq \frac{2}{3}$.

Then $G \bullet H : [0,1] \times N \to M$ defined by

$$[0,1] \times N \ni (t,x) \mapsto (G \bullet H)(t,x) := \begin{cases} H\big(p(2t),x\big) & \text{if } 0 \leq t \leq \frac{1}{2} \\ G\big(p(2t-1),x\big) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases} \tag{15.45}$$

is a smooth homotopy from f to h. This proves transitivity.

∎

Caution: One could not have simply defined $G \bullet H$ by

$$[0,1] \times N \ni (t,x) \mapsto \begin{cases} H(2t,x) & \text{if } 0 \leq t \leq \frac{1}{2} \\ G(2t-1,x) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases} \tag{15.46}$$

because this function need not be smooth at $t = \frac{1}{2}$. It is, however, always continuous (for this reason, the study of homotopies is much simpler if one restricts oneself to only continuous functions instead of smooth functions).

Exercise 15.47. Prove that $G \bullet H$ as defined in (15.45) is smooth.

Exercise 15.48. Let M be a smooth m-dimensional manifold.

(a) Prove that M is connected if and only if every two smooth maps $f, g : \{\bullet\} \to M$ are smoothly homotopic. Here, $\{\bullet\}$ denotes a one-element subset of some Euclidean space.

(b) What is the meaning of the cardinality of the set of all smooth homotopy classes of maps $\{\bullet\} \to M$?

Remark 15.49. Let M be a manifold. A *multiplication* on M consists of a smooth function $\mu : M \times M \to M$. A Lie group is an example of a manifold with multiplication. In fact, of particular interest are certain spheres. S^1 and S^3 have natural multiplications coming from viewing S^1 as a subset of the complex numbers and S^3 as a subset of the quaternions. In fact, S^7 has a natural multiplication coming from viewing it as a subset of the octonions [4]. However, while the multiplications on S^1 and S^3 are associative (in fact, they are Lie groups), the multiplication on S^7 is *not* associative. But perhaps, you might wonder if it is associative up to homotopy. A multiplication on M is called *homotopy associative* iff there exists a smooth homotopy

$$\begin{tikzcd} M \times M \arrow[d, "\mu"'] \arrow[dr, Rightarrow] & M \times M \times M \arrow[l, "\mu \times \mathrm{id}_M"'] \arrow[d, "\mathrm{id}_M \times \mu"] \\ M & M \times M \arrow[l, "\mu"] \end{tikzcd} \tag{15.50}$$

Research Project (part of remark). Does the product induced on S^7 from the octonions provide a homotopy associative multiplication? Prove or disprove your claim.

Theorem 15.51. *Let M and N be m-dimensional manifolds with N compact and M connected. Let $f, g : N \to M$ be two smooth maps. If f is smoothly homotopic to g, then*

$$\deg(f) = \deg(g). \tag{15.52}$$

Proof. See Chapter 5 of [8].[33] ■

Now that we know that the degree is independent of homotopy, we immediately get the following corollary, which is lengthy and awkward to state precisely.

Theorem 15.53. *Let $A \subseteq \mathbb{R}^m$ be open and let $V : A \to \mathbb{R}^m$ be a smooth vector field with an isolated critical point at $c \in A$. Let $\epsilon, \delta > 0$ be real numbers such that the vector field V restricted to $V_\epsilon(c)$ and $V_\delta(c)$ vanishes nowhere except at c. Let $\phi_\epsilon, \phi_\delta : S^{m-1} \to S^{m-1}$ be the maps defined by*

$$S^{m-1} \ni x \mapsto \phi_\epsilon(x) := \frac{V_{c+\epsilon x}}{|V_{c+\epsilon x}|} \qquad \& \qquad S^{m-1} \ni x \mapsto \phi_\delta(x) := \frac{V_{c+\delta x}}{|V_{c+\delta x}|}. \tag{15.54}$$

[33] Actually proving this is quite involved and would take us too far afield. By this point, you should be more than capable of reading Milnor's book and convince yourself of the validity of this claim.

Then

$$\deg(\phi_\epsilon) = \deg(\phi_\delta). \tag{15.55}$$

Note that in the statement of this theorem, $\partial\overline{V_\epsilon(c)}$ (and similarly for δ) and S^{m-1} are oriented in the standard way as in Example 15.6 ($\partial\overline{V_\epsilon(c)}$ is just an $(m-1)$-dimensional sphere of radius ϵ centered at c and its orientation can be obtained by translating to the origin, scaling the sphere by $\frac{1}{\epsilon}$, and then using the orientation of the standard sphere).

Proof. Let $H : [0,1] \times S^{m-1} \to S^{m-1}$ be the function

$$[0,1] \ni t \mapsto \phi_{(1-t)\epsilon+t\delta}. \tag{15.56}$$

Then H defines a smooth homotopy from ϕ_ϵ to ϕ_δ since V is smooth in the open disc $V_{\max\{\epsilon,\delta\}}(c)$. Hence, by Theorem 15.51, the conclusion is reached. ■

Definition 15.57. Let $A \subseteq \mathbb{R}^m$ be open and let $V : A \to \mathbb{R}^m$ be a smooth vector field with an isolated critical point at $c \in A$. The <u>*index*</u> $\iota(V; c)$ of V at c is the degree of the function

$$\begin{aligned} \partial\overline{V_\epsilon(c)} &\xrightarrow{\phi_\epsilon} S^{m-1} \\ x &\mapsto \frac{V_x}{|V_x|}, \end{aligned} \tag{15.58}$$

where $\epsilon > 0$ is small enough so that the vector field V vanishes nowhere on $V_\epsilon(c) \setminus \{c\}$.

The index of a vector field provides is with a concrete mathematical object associated to isolated critical points of vector fields that is robust to smooth perturbations/homotopies. We will spend the rest of this lecture working through examples.

Example 15.59. Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field of the form

$$V(x,y) = (\alpha x, \beta y) \tag{15.60}$$

for some constants $\alpha, \beta \in \mathbb{R}$. 0 is an isolated zero provided that α and β are nonzero. For the following calculations, set $\epsilon := 1$.

(a) When $\alpha = \beta = 1$, the function $\phi : \partial\overline{V_\epsilon(0)} \to S^1$ is the identity map. Hence, its degree is 1 so that the index of V at 0 is 1.

(b) When $\alpha = \beta = -1$, the function $\phi : \partial\overline{V_\epsilon(0)} \to S^1$ is rotation by π radians. Hence, every point of $\partial\overline{V_\epsilon(0)}$ is a regular point. Take for example $(1,0)$. Then the matrix associated to the differential of ϕ is the restriction of the matrix

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{15.61}$$

(which comes from extending ϕ to a function, namely the rotation, in an open neighborhood of $\partial\overline{V_\epsilon(0)}$, and calculating its differential at $(1,0)$, which is the same function since rotation is a linear transformation) to the tangent space $T_{(1,0)}\partial\overline{V_\epsilon(0)}$. Since e_2 is a representative for the (standard) orientation on $T_{(1,0)}\partial\overline{V_\epsilon(0)}$, the image of this vector is

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \tag{15.62}$$

which is in the tangent space at $T_{(-1,0)}S^1$. This vector agrees with the basis vector chosen for the (standard) orientation on S^1 and so ϕ is orientation-preserving. Hence, the index of V at 0 is 1.

(c) When $\alpha = 1$ and $\beta = -1$, the function $\phi : \partial\overline{V_\epsilon(0)} \to S^1$ is reflection through the horizontal axis. Since this extends to a linear transformation on $\mathbb{R}^2$, the differential is

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{15.63}$$

restricted to the tangent space at any point on $\partial\overline{V_\epsilon(0)}$. At $(1,0)$ (a regular point of ϕ), the image of the basis vector e_2 gets sent to

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \tag{15.64}$$

which is also a vector at $(1,0)$ since this point is fixed under ϕ. Since this vector is oriented oppositely to the standard orientation, the degree of ϕ is -1. Hence, the index of V at 0 is -1.

(d) An analogous situation to part (c) occurs when $\alpha = -1$ and $\beta = 1$ and the index of V at 0 is -1.

Exercise 15.65. Calculate the index of the following vector fields $V : \mathbb{R}^2 \to \mathbb{R}^2$ at the point 0.

(a) $V(x,y) := (-y, x)$.

(b) $V(x,y) := (y, -x)$.

(c) $V(x,y) := (-y - x, x - y)$.

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

Exercise 15.66. Let $V : [0,\infty) \times [0,\infty) \to \mathbb{R}^2$ be the vector field given by

$$[0,\infty) \times [0,\infty) \ni (x,y) \mapsto V(x,y) := \big(0.4x - 0.4xy, -0.1y + 0.2xy\big) \tag{15.67}$$

from Exercise 7.47. Find the isolated critical points of V and calculate their respective indices. Note: you may need to extend V to a larger domain to do this.

Exercise 15.68. Let $V : [0,\infty) \times [0,\infty) \to \mathbb{R}^2$ be the vector field given by

$$[0,\infty) \times [0,\infty) \ni (x,y) \mapsto V(x,y) = \left(2x\left(1 - \frac{x}{2}\right) - xy, 3y\left(1 - \frac{y}{3}\right) - 2xy\right). \tag{15.69}$$

Find the critical points of V and calculate their respective indices. Note: you may need to extend V to a larger domain to do this.

Exercise 15.70. Find the critical points of the following vector fields and calculate their respective indices.

(a) Let $V_2 : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) := \big(x^2 - y^2, 2xy\big). \tag{15.71}$$

(b) Let $V_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$\mathbb{R}^2 \ni (x, y) \mapsto V(x, y) := \big(x^3 - 3xy^2, 3x^2y - y^3\big). \tag{15.72}$$

(c) Fix $n \in \mathbb{Z} \setminus \{0\}$. Using the bijection $\mathbb{R}^2 \to \mathbb{C}$ given by $(x, y) \mapsto x + y\sqrt{-1}$ representing ordered pairs as complex numbers with inverse given by sending $z \in \mathbb{C}$ to $\left(\frac{z+z\sqrt{-1}}{2}, \frac{z-z\sqrt{-1}}{2}\right)$, let $V_n : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$\mathbb{R}^2 \ni z \mapsto V(z) := z^n. \tag{15.73}$$

Let us now consider a two-dimensional example.

Example 15.74. Let $V : \mathbb{R}^3 \to \mathbb{R}^3$ be the vector field given by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto V(x, y, z) := \big(-y, x, z\big). \tag{15.75}$$

On S^2, this vector field restricts to a function from S^2 to S^2 since the vectors are already normalized. Furthermore, this induced function $\phi : \partial\overline{V_1(0)} \to S^2$ is rotation by $\frac{\pi}{2}$ radians in the xy-plane so that ϕ extends to this rotation in $\mathbb{R}^3$. Since this is linear, the differential at any point agrees with this function, and as a matrix, this is given by

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{15.76}$$

At the point $(1, 0, 0)$, an oriented basis agreeing with the orientation on S^2 is (e_2, e_3). The images of these two vectors under the above transformation are

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{15.77}$$

which give the ordered basis $(-e_1, e_3)$ at the point $(0, 1, 0)$. This ordered basis agrees with the standard orientation at this point. Hence, the degree is 1. Hence, the index of V at 0 is 1.

Exercise 15.78. Let $V : \mathbb{R}^3 \to \mathbb{R}^3$ be the vector field given by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto V(x, y, z) = \left(10(y - x), x(28 - z) - y, xy - \frac{8}{3}z\right) \tag{15.79}$$

from Exercise 7.56. Find the critical points of V and calculate their respective indices.

Exercise 15.80. Let M, N, and P be oriented m-dimensional manifolds with M and N connected and N and P compact. Let $f : N \to M$ and $g : P \to N$ be smooth. Prove that

$$\deg(f \circ g) = \big(\mathrm{def}(f)\big)\big(\mathrm{def}(g)\big). \tag{15.81}$$

Problems.

From these notes: Exercises 15.4, 15.5, 15.22, 15.23, 15.30, 15.42, 15.47, 15.48, 15.65, 15.66, 15.68, 15.70, 15.78, 15.80

16 April 6: Vector Fields and Flows on Manifolds

Before defining vector fields on manifolds, we revisit what a vector field is in Euclidean space. The following is not really a new definition but is a rephrasing of the definition in terms of tangent spaces.

Definition 16.1. Let $A \subseteq \mathbb{R}^m$ be an open subset of $\mathbb{R}^m$. A smooth (tangent) *vector field* on A is a smooth function $V : A \to \mathbb{R}^m$ such that $V(x) \equiv V_x \in T_x A$ for all $x \in A$.

This justifies our earlier perspective on drawing vector fields as vectors attached at points of the domain.

Definition 16.2. Let $M \subseteq \mathbb{R}^k$ be a manifold. A smooth tangent *vector field* on M is a smooth function $V : M \to \mathbb{R}^k$ such that $V(x) \equiv V_x \in T_x M$ for all $x \in M$.

The adjective "tangent" might often be used to distinguish these vector fields from other kinds of smoothly varying vectors (such as normal vectors or otherwise).

Example 16.3. Let $X : [0, 2\pi) \times [0, \pi) \to \mathbb{R}^3$ be the function defined by

$$[0, 2\pi) \times [0, \pi) \ni (\phi, \theta) \mapsto X(\phi, \theta) := \big(\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta)\big) \tag{16.4}$$

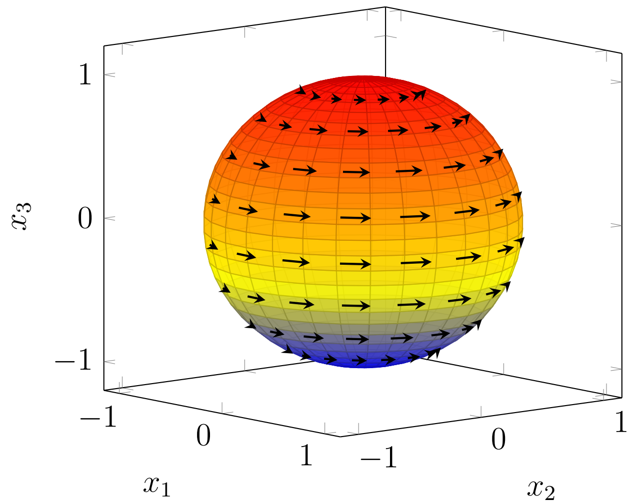
and let M be the image of X. Then $M = S^2$ (though X is *not* a bijective parametrization of S^2). The differential of X at (ϕ, θ) with respect to the standard basis is given by

$$[D_{(\phi,\theta)}X] = \begin{bmatrix} -\sin(\phi)\sin(\theta) & \cos(\phi)\cos(\theta) \\ \cos(\phi)\sin(\theta) & \sin(\phi)\cos(\theta) \\ 0 & -\sin(\theta) \end{bmatrix} \tag{16.5}$$

and the image of the unit vector e_1 under this linear transformation is just the first column of this matrix. It is not a-priori obvious that this is a smooth vector field on S^2. This is because the image of a vector field under a smooth map need not be a vector field on the image (this will be illustrated in an example later). Nevertheless, the vector field is smooth because it is the restriction of the smooth vector field whose value at $(x, y, z) \in \mathbb{R}^3$ is given by the derivative of the rotation of $\mathbb{R}^3$ along the z axis, namely

$$\mathbb{R}^3 \ni (x, y, z) \mapsto V_{(x,y,z)} = \frac{d}{dt}\left(\begin{bmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right)\Bigg|_{t=0} = \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix} \tag{16.6}$$

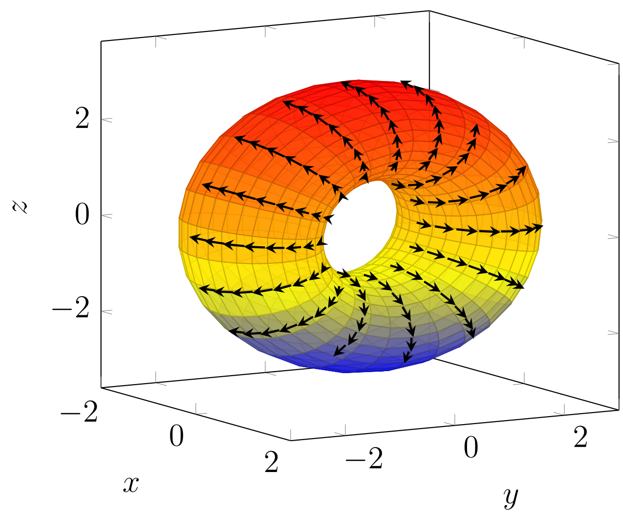
which is smooth on $\mathbb{R}^3$ and restricts to the vector field discussed above on S^2.

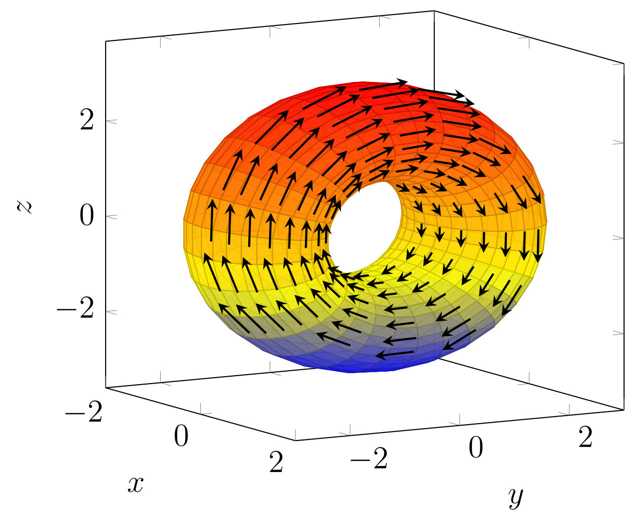


Example 16.7. Let $X : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}^3$ be defined by

$$[0, 2\pi) \times [0, 2\pi) \ni (s, t) \mapsto X(s, t) := \Big(\sin(t), \big(2 + \cos(t)\big) \cos(s), \big(2 + \cos(t)\big) \sin(s) \Big) \tag{16.8}$$

and let M be the image of X. The differential of X at an arbitrary point $(s, t) \in [0, 2\pi) \times [0, 2\pi)$ was calculated in Example 13.29. The images of the standard Euclidean vector fields provide two examples of vector fields on M (these are the two columns of the matrix $[D_{(s,t)}X]$ in that example).





Notice that these vector fields do not vanish anywhere on the torus.

Exercise 16.9. Prove that the two vector fields from Example 16.7 are smooth vector fields on the torus M.

An example of a vector field that is not a *tangent* vector field is the following.

Example 16.10. Let $V : \mathbb{R}^3 \to \mathbb{R}^3$ be the tangent vector field on $\mathbb{R}^3$ given by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto V(x, y, z) := (x, y, z). \tag{16.11}$$

The restriction of this vector field to S^2 is not a tangent vector field on S^2. This is because all the vectors are in the normal space on S^2.

A particular kind of homotopy that often occurs is one in which the constituent smooth maps are diffeomorphisms.

Definition 16.12. Let M and N be m-dimensional manifolds and let $f, g : N \to M$ be two diffeomorphisms. A *smooth isotopy from f to g* is a smooth homotopy $H : f \Rightarrow g$ such that $H(t, \,\cdot\,) : N \to M$ is a diffeomorphism for all $t \in [0, 1]$. In this case, the diffeomorphisms f and g are said to be *smoothly isotopic*.

A closely related concept is that of a one-parameter group of diffeomorphisms, also known as a flow.

Definition 16.13. Let M be a smooth manifold. A *flow* on M consists of a smooth map $\varphi : \mathbb{R} \times M \to M$ such that

(a) $\varphi(t, \,\cdot\,) : M \to M$ is a diffeomorphism for all $t \in \mathbb{R}$,

(b) $\varphi(0, \,\cdot\,) = \mathrm{id}_M$, and

(c) $\varphi(s + t, \,\cdot\,) = \varphi(s, \,\cdot\,) \circ \varphi(t, \,\cdot\,)$ for all $s, t \in \mathbb{R}$.

The axioms of a flow can be phrased in the language of groups. The set, $\mathrm{Diff}(M)$, of diffeomorphisms of a manifold M is a group. A flow is therefore a smooth function $\varphi : \mathbb{R} \times M \to M$ such that the naturally associated function $\tilde{\varphi} : \mathbb{R} \to \mathrm{Diff}(M)$ is a group homomorphism (this function assigns $t \in \mathbb{R}$ to the diffeomorphism $\varphi(t, \,\cdot\,)$).

Theorem 16.14. *Let $M \subseteq \mathbb{R}^k$ be a smooth manifold and $\varphi : \mathbb{R} \times M \to M$ a flow on M. Then the assignment*

$$M \ni x \mapsto V_x := D_0\varphi(\,\cdot\,, x) \tag{16.15}$$

is a smooth tangent vector field on M. The vector field V is called the infinitesimal generator of φ.

Proof. First note that the derivative

$$D_0\varphi(\,\cdot\,, x) = \frac{\partial}{\partial t}\varphi(t, x)\Big|_{t=0} := \lim_{t \to 0} \frac{\varphi(t, x) - \varphi(0, x)}{t} \tag{16.16}$$

makes sense because we are viewing the manifold M as a subset of some Euclidean space, say $\mathbb{R}^k$. The fact that $V_x \in T_xM$ follows immediately from the fact that for every fixed $x \in M$, the assignment $t \mapsto \phi(t, x)$ is a smooth path in M. Therefore, the derivative at 0 is an element of T_xM. Smoothness of V follows from smoothness of φ. In more detail, since $\varphi : [0, 1] \times M \to M$ is smooth, for each $x \in M$, there exists an open rectangle $U := U_1 \times U_2 \subseteq \mathbb{R} \times \mathbb{R}^k$ around the point $(0, x) \in [0, 1] \times M$ and a smooth function $\Phi : U \to \mathbb{R}^k$ that restricts to φ on $U \cap \big([0, 1] \times M\big)$. Since Φ is smooth, the assignment

$$U_2 \ni x \mapsto \frac{\partial}{\partial t}\Phi(t, x)\Big|_{t=0} \tag{16.17}$$

is a smooth vector field on U_2 and agrees with the vector field V on $U_2 \cap M$. Therefore, V is smooth at $x \in M$. Since $x \in M$ was arbitrary, V is a smooth tangent vector field on M. ■

We will discuss the converse of this theorem in a few lectures.

Example 16.18. Let $\Phi : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ be the function defined by

$$\mathbb{R} \ni t \mapsto \Phi(t, \cdot\,) := \begin{bmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{16.19}$$

which is rotation by angle t along the z axis. This is a linear isomorphism and is therefore a diffeomorphism of $\mathbb{R}^3$. Since $\Phi(t, \cdot\,)$ takes any point on S^2 to S^2, this restricts to a diffeomorphism of S^2 for all $t \in \mathbb{R}^3$. When $t = 0$, this is the identity. Finally, $\Phi(s+t, \cdot\,) = \Phi(s, \cdot\,) \circ \Phi(t, \cdot\,)$ follows from basic trigonometry. Hence, Φ is a flow. Its infinitesimal generator vector field was calculated in Example 16.3. Namely, this infinitesimal generator at $(x, y, z) \in S^2$ is given by $(-y, x, 0)$.

Exercise 16.20. Let $L : \mathbb{R}^m \to \mathbb{R}^m$ be an orientation-preserving linear isomorphism. Prove that L is smoothly isotopic to the identity.

Theorem 16.21. *Any orientation preserving diffeomorphism of $\mathbb{R}^m$ is smoothly isotopic to the identity.*

Proof. Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be such a diffeomorphism. Since translating by a constant vector does not alter these properties, we can assume $f(0) = 0$. For each $t \in (0, 1]$, set $H(t, \cdot\,) : \mathbb{R}^m \to \mathbb{R}^m$ to be the function

$$\mathbb{R}^m \ni x \mapsto H(t, x) := \frac{f(tx)}{t}. \tag{16.22}$$

Set $H(0, \cdot\,) := D_0 f$, the differential of f at 0. Since

$$(D_0 f)(x) = D_0(f \circ \gamma_{0;x}) = \frac{d}{dt} f(tx)\big|_{t=0} = \lim_{t\to\infty} \frac{f(tx) - f(0)}{t} = \lim_{t\to\infty} \frac{f(tx)}{t}, \tag{16.23}$$

$H : [0, 1] \times \mathbb{R}^m \to \mathbb{R}^m$ is continuous. To see that H is also smooth, by Hadamard's Lemma, there exists a smooth function $g : \mathbb{R}^m \to \mathbb{R}^m$ such that

$$f(x) = \sum_{i=1}^{m} x_i g_i(x) \tag{16.24}$$

for all $x \in \mathbb{R}^m$. Therefore,

$$H(t, x) = \sum_{i=1}^{m} x_i g_i(tx) \tag{16.25}$$

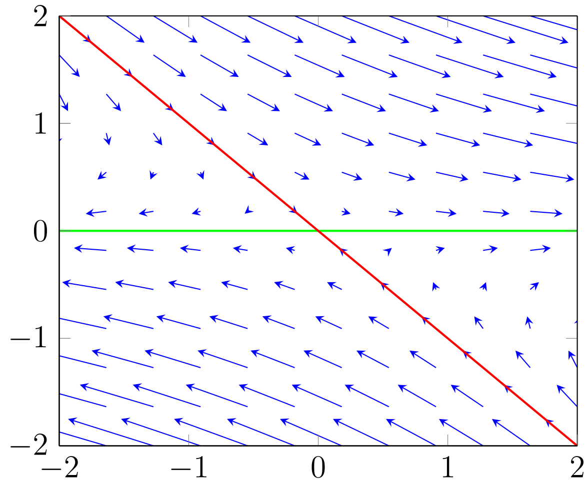
for all $t \in [0, 1]$ and all $x \in \mathbb{R}^m$. Thus, H is a smooth homotopy from $D_0 f$ to f. Since $D_0 f$ is a linear isomorphism and f is a diffeomorphism, $H(t, \cdot\,)$ is a diffeomorphism for all $t \in [0, 1]$ so that H is a smooth isotopy from $D_0 f$ to f. Since f is an orientation-preserving diffeomorphism, $D_0 f : \mathbb{R}^m \to \mathbb{R}^m$ is an orientation-preserving linear isomorphism. By Exercise 16.20, $D_0 f$ is smoothly isotopic to the identity. Since isotopy is an equivalence relation by Theorem 15.43 and is in particular transitive, f is smoothly isotopic to the identity. ■

Definition 16.26. Let M and N be manifolds (not necessarily of the same dimension), let V and W be tangent vector fields on M and N, respectively, and let $f : N \to M$ be a smooth map. The tangent vector fields V and W <u>*correspond (to each other) under f*</u> iff

$$(D_x f)(W_x) = V_{f(x)} \qquad \forall\, x \in N. \tag{16.27}$$

Example 16.28. Let $W : \mathbb{R}^2 \to \mathbb{R}^2$ be the vector field given by

$$\mathbb{R}^2 \ni (x, y) \mapsto V_{(x,y)} := \big(x + 2y, -y\big). \tag{16.29}$$

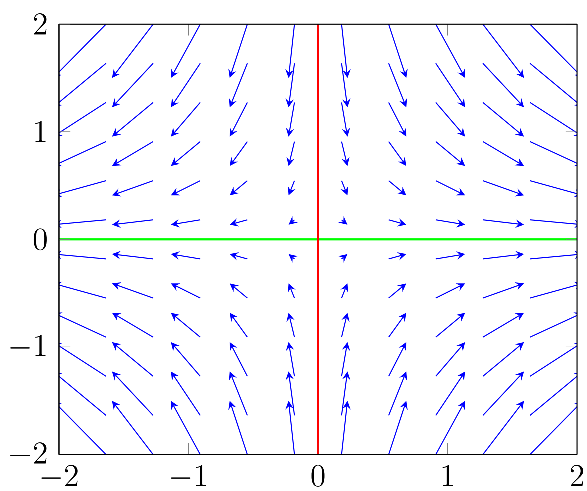


The arrows are scaled to 1/8 their actual lengths for easier view and a green line has been drawn to depict the eigenspace for the positive eigenvalue and a red line has been drawn to depict the eigenspace for the negative eigenvalue. The eigenvalues and eigenvector pairs for the vector field V at the origin are given by

$$\left(\lambda_1 = 1 \quad \& \quad \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \qquad \& \qquad \left(\lambda_2 = -1 \quad \& \quad \vec{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right). \tag{16.30}$$

Compare this vector field V to the vector field $W : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbb{R}^2 \ni (x, y) \mapsto W_{(x,y)} := (x, -y). \tag{16.31}$$



The arrows are scaled to 1/4 their actual lengths for easier view and a green line has been drawn to depict the eigenspace for the positive eigenvalue and a red line has been drawn to depict the eigenspace for the negative eigenvalue. The vectors V and W have similar features. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the smooth function that is given by the linear transformation

$$f = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}. \tag{16.32}$$

Furthermore, since f is already linear,

$$(D_{(x,y)}f)(W_{(x,y)}) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} x+y \\ -y \end{bmatrix} \tag{16.33}$$

while

$$V_{f(x,y)} = \begin{bmatrix} f_1(x,y) + 2f_2(x,y) \\ -f_2(x,y) \end{bmatrix} = \begin{bmatrix} (x-y) + 2(y) \\ -y \end{bmatrix} = \begin{bmatrix} x+y \\ -y \end{bmatrix} \tag{16.34}$$

showing that W corresponds to V under f. Therefore, correspondence of vector fields is basically a change of coordinates that preserves the structure of the vector field. We will illustrate this with another example later.

In general, you cannot push-forward vector fields under arbitrary smooth maps.

Example 16.35. Let $f : \mathbb{R} \to \mathbb{R}$ be the function sending $x \in \mathbb{R}$ to $f(x) = x^2$ and let $V : \mathbb{R} \to \mathbb{R}$ be the constant vector field $V(x) = 1$ for all $x \in \mathbb{R}$. Then $(D_x f)(V_x) = 2x$ for all $x \in \mathbb{R}$. So while $f(-1) = 1 = f(1)$, we have

$$(D_{-1}f)(V_{-1}) = -2 \neq 2 = (D_1 f)(V_1). \tag{16.36}$$

Even if the function f is one-to-one, you still might not be able to push-forward smooth tangent vector fields as the following example illustrates.

Example 16.37. Let $\varphi_S : \mathbb{R} \to S^1$ be one part of the stereographic projections given by

$$\mathbb{R} \ni y \mapsto \left(\frac{-2y}{1+y^2}, \frac{-1+y^2}{1+y^2} \right) \tag{16.38}$$

and consider the vector field V on $\mathbb{R}$ given by[34]

$$\mathbb{R} \ni y \mapsto V_y := y^3 \tag{16.39}$$

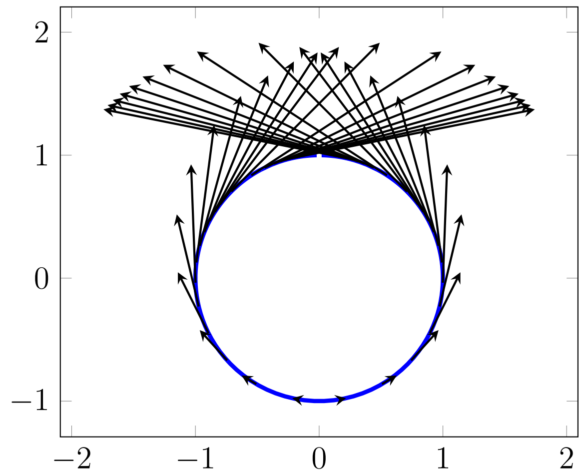
Then, the differential of φ_S at $y \in \mathbb{R}$ is

$$[D_y \varphi_S] = \frac{1}{(1+y^2)^2} \begin{bmatrix} 2(y^2-1) \\ 4y \end{bmatrix} \tag{16.40}$$

The pointwise push-forward of this vector field under φ_S is therefore

$$(D_y \varphi_S)(V_y) = \frac{1}{(1+y^2)^2} \begin{bmatrix} 2(y^2-1) \\ 4y \end{bmatrix} \begin{bmatrix} y^3 \end{bmatrix} = \frac{1}{(1+y^2)^2} \begin{bmatrix} 2(y^5-y^3) \\ 4y^4 \end{bmatrix} \tag{16.41}$$

[34]The stereographic projection sends points off at $\pm\infty$ in $\mathbb{R}$ towards the same point on the unit circle, namely the north pole. Therefore, we should try to put a vector field on $\mathbb{R}$ that has very different values as you tend towards $\pm\infty$. Furthermore, it should grow quickly enough so that in the image it does not tend to zero. This is the motivation behind choosing the vector field in this manner.



The vectors above have been slightly scaled so that they fit on the page. The problem that prevents the pushforward from being a smooth vector field occurs at the north pole, where no value of the vector field can be given so that the vector field is continuous at that point. This is because the direction of the vector fields from both sides as you approach the north pole differ but their magnitude does not tend to zero. In this case, the magnitude diverges, but by rescaling (as in the picture), one can make it so that the magnitude tends to a finite constant.

However, if f is a diffeomorphism, then the image of a tangent vector field is a tangent vector field.

Exercise 16.42. Let $f : N \to M$ be a diffeomorphism of m-dimensional manifolds and let W be a smooth tangent vector field on N.

(a) Prove that the assignment

$$M \ni y \mapsto V_y := (D_{f^{-1}(y)}f)(W_{f^{-1}(y)}) \tag{16.43}$$

defines a smooth tangent vector field on M. This vector field is called the *push-forward of W along f* and is denoted by f_*W.

(b) Show that if c is an isolated critical point of W on N, then $f(c)$ is an isolated critical point of V on M.

The following example shows us how corresponding vector fields under a change of coordinates can turn a difficult ODE into a manageable one.

Example 16.44. Consider the Lotka-Volterra model given by the ODE

$$\begin{aligned} \dot{x} &= \lambda x - \eta xy \\ \dot{y} &= -\mu y + \xi xy \end{aligned} \tag{16.45}$$

on $[0,\infty) \times [0,\infty)$ from Example 7.47 but with coefficients kept more general, i.e. $\lambda, \eta, \mu, \xi > 0$. One of the critical points of the associated vector field

$$[0,\infty) \times [0,\infty) \ni (x,y) \mapsto W(x,y) = (\lambda x - \eta xy, -\mu y + \xi xy) \tag{16.46}$$

is given by

$$c := \left(\frac{\mu}{\xi}, \frac{\lambda}{\eta}\right) \tag{16.47}$$

(the other critical point is $(0,0)$ but we will not analyze this critical point presently). Consider the change of coordinates given by the function

$$\begin{aligned} (0,\infty)\times(0,\infty) &\xrightarrow{f} \mathbb{R}^2 \\ (x,y) &\mapsto \left(\ln\left(\frac{\eta}{\lambda}y\right), \ln\left(\frac{\xi}{\mu}x\right)\right), \end{aligned} \tag{16.48}$$

which gives us a coordinate chart around the critical point (16.47). Let us denote $f_1 := q$ and $f_2 := p$ (for reasons that will become apparent soon). Notice that f is a diffeomorphism. Therefore, we can find a vector field V on $\mathbb{R}^2$ that corresponds to W under f by calculating the differential of f at (x,y) and applying it to $W_{(x,y)}$ (Exercise 16.42 discusses the general situation)

$$V_{f(x,y)} = (D_{(x,y)}f)(W_{(x,y)}) = \begin{bmatrix} 0 & \frac{1}{y} \\ \frac{1}{x} & 0 \end{bmatrix} \begin{bmatrix} \lambda x - \eta xy \\ -\mu y + \xi xy \end{bmatrix} = \begin{bmatrix} -\mu + \xi x \\ \lambda - \eta y \end{bmatrix} \tag{16.49}$$

With respect to the variables q and p in $\mathbb{R}^2$, this vector field V is given by

$$V_{(q,p)} = \big(\mu(e^p - 1), \lambda(1 - e^q)\big). \tag{16.50}$$

Therefore, our differential equation becomes

$$\begin{aligned} \dot{q} &= \mu(e^p - 1) \\ \dot{p} &= \lambda(1 - e^q) \end{aligned} \tag{16.51}$$

as you can check by substitution. The critical point is now at $(q = 0, p = 0)$. While we can certainly check what the index of this vector field is, we should already have the intuition that the index is 1 and that this index is preserved under f. Let us explicitly check this. Notice that the critical point is now at $(0,0)$. Using the expansion of the exponential, the vector field V can be expressed as

$$V_{(q,p)} = \left(\mu\left(p + \frac{p^2}{2!} + \frac{p^3}{3!} + \cdots\right), -\lambda\left(q + \frac{q^2}{2!} + \frac{q^3}{3!} + \cdots\right)\right). \tag{16.52}$$

Notice that the constant term cancelled. For $\epsilon > 0$ sufficiently small, this vector field is approximately given by

$$V_{(q,p)} = (\mu p, -\lambda q) + \mathcal{O}(\epsilon^2) \qquad \forall\ (q,p) \in V_\epsilon(0,0), \tag{16.53}$$

where $\mathcal{O}(\epsilon^2)$ means higher order terms in ϵ. In other words, since $\mu, \lambda > 0$, this vector field is a center with counter-clockwise motion. Therefore, it has index 1. While we are discussing this example, we should mention that the form that the differential equation is now in is actually quite convenient. We can explicitly integrate it and obtain a Hamiltonian for it. One such Hamiltonian is[35]

$$\mathbb{R}^2 \ni (q,p) \mapsto H(q,p) := \mu(e^p - p - 1) + \lambda(e^q - q - 1). \tag{16.54}$$

[35]The $-\mu$ and $-\lambda$ terms are simply added so that $H(0,0) = 0$ and so that $H(q,p) \geq 0$ for all other $(q,p) \in \mathbb{R}^2$.

and one can check that $\dot{q} = \frac{\partial H}{\partial p}$ and $\dot{p} = -\frac{\partial H}{\partial q}$. You might say: so what? Well, recall Theorem 7.68 which guarantees that for any integral curve γ of V, which is a Hamiltonian vector field, $H\big(\gamma(t)\big)$ is constant for all $t \in \mathbb{R}$. Furthermore, notice that the differential of H is given by the 1×2 matrix

$$[D_{(q,p)}H] = \begin{bmatrix} \lambda(e^q - 1) & \mu(e^p - 1) \end{bmatrix} \tag{16.55}$$

Therefore, (q, p) is a regular point provided that $(q, p) \neq (0, 0)$. Furthermore, $H(q, p) \geq 0$ for all $(q, p) \in \mathbb{R}^2$. This says that if $\ell > 0$, then $H^{-1}(\ell)$ is a 1-dimensional manifold in $\mathbb{R}^2$. Since f is a diffeomorphism, we can pull this manifold back to $(0, \infty) \times (0, \infty)$. Can you guess what this manifold is? It's an integral curve to our initial differential equation! At first, our non-linear differential equation looked incredibly complicated and difficult to solve explicitly. By a change of coordinates, we've viewed it as a Hamiltonian system, solved for an integral of motion, and used that integral of motion to describe the solutions explicitly, which we are able to do by the Regular Value Theorem. In terms of the original variables $(x, y) \in (0, \infty) \times (0, \infty)$, the equation describing this integral curve is

$$\mu\left(\frac{\xi}{\mu}x - \ln\left(\frac{\xi}{\mu}x\right) - 1\right) + \lambda\left(\frac{\eta}{\lambda}y - \ln\left(\frac{\eta}{\lambda}y\right) - 1\right) = \ell, \tag{16.56}$$

which, as you can see, is quite complicated and would have been a challenge to solve otherwise.

We now make sense of preservation of the index of a vector field at an isolated critical point under a diffeomorphism.

Theorem 16.57. *Let $A, B \subseteq \mathbb{R}^m$ be open subsets, let $f : A \to B$ be a diffeomorphism, let V be a vector field on A, and W a vector field on B, and suppose that V and W correspond under f. Then*

$$\iota(V, c) = \iota\big(W, f(c)\big) \tag{16.58}$$

for all isolated critical points $c \in A$ of V.

Proof. See Milnor [8]. ■

Note that in the above theorem, the standard orientation on $\mathbb{R}^m$ is used to calculate the index. This result allows us to make the following definition of the index of a vector field on a manifold.

Definition 16.59. Let $M \subseteq \mathbb{R}^k$ be an m-dimensional manifold, let V be a vector field on M, and let c be an isolated critical point of V. Let $U \subseteq \mathbb{R}^m$ and $W \subseteq \mathbb{R}^k$ be open sets and $\varphi : U \to M \cap V$ be a parametrization of M about c with associated coordinate chart $\psi : M \cap W \to U$. Assume that U and W are chosen small enough so that $V_x \neq 0$ for all $x \in (M \cap W) \setminus \{c\}$. The *index of V at c* is

$$\iota(V, c) := \iota\big(\psi_* V, \psi(c)\big), \tag{16.60}$$

where $\psi_* V$ is the push-forward vector field (see Exercise 16.42).

Exercise 16.61. Use the results of Theorem 16.57 to prove that $\iota(V, c)$ is well-defined, i.e. independent of the choice of parametrization of M about c.

The sum of the indices of a vector field on a manifold is actually independent of the vector field. This is an incredibly surprising theorem, which we will not prove, but will state. To properly state it, we should define what it means to triangulate a manifold. This is where our old friend, the simplex, makes a comeback (see Exercise 2.46).

Definition 16.62. A *simplicial complex* is a set K consisting of simplices (of possibly various dimensions) satisfying

(a) every face of a simplex in K is also a simplex in K and

(b) the intersection of any two simplices σ and τ in K is either empty or is both a face of σ and τ.

Let K_i denote the subset of K consisting of all the i-simplices of K. Let $|K_i|$ denote the cardinality of K_i, i.e. the number of (distinct) i-simplices of K.

Example 16.63. Every simplex is a simplicial complex. Namely, let K be the set of all simplices of Δ^m. The number of i-simplices is given by

$$|K_i| = \binom{m+1}{i+1} \equiv \frac{(m+1)!}{(i+1)!(m-i)!} \qquad \forall\, i \in \{0,1,\ldots,m\}. \tag{16.64}$$

Example 16.65. The boundary of a simplex is a simplicial complex. Namely, let K be the set of all simplices of $\partial\Delta^m$. The number of i-simplices is given by

$$|K_i| = \binom{m+1}{i+1} \qquad \forall\, i \in \{0,1,\ldots,m-1\}. \tag{16.66}$$

Definition 16.67. A *triangulation of a manifold* M consists of a simplicial complex K together with a homeomorphism $h : K \to M$.

Example 16.68. For any $m \in \mathbb{N}$, let K be the boundary of Δ^{m+1} with its center, which is $\left(\frac{1}{m+2},\ldots,\frac{1}{m+2}\right) \in \mathbb{R}^{m+2}$, translated to the origin. Let $r : \mathbb{R}^{m+2} \setminus \{0\} \to \mathbb{R}^{m+2} \setminus \{0\}$ be the function defined by sending $x \in \mathbb{R}^{m+2} \setminus \{0\}$ to $\frac{x}{|x|}$. Restricting r to K provides a map from K to S^{m+1}. By a suitable rotation Θ of S^{m+1}, this will map $r(K) \subseteq S^{m+1}$ to the equator, which is S^m. This provides a homeomorphism of $\partial\Delta^{m+1}$ to S^m.

Given a smooth manifold M, it is not at all obvious whether or not a triangulation of it exists. However, it is true. Furthermore, if M is compact, then there exists a finite triangulation.

Definition 16.69. Let K be a finite simplicial complex and let $\dim K := n$ be the largest dimension of simplices in K. For each $i \in \{0,1,\ldots,n\}$, let K_i denote the subset of K consisting of i-simplices. The *Euler characteristic of K* is the number

$$\chi(K) := \sum_{i=0}^{n} (-1)^i |K_i|. \tag{16.70}$$

The Euler characteristic is a topological invariant in the sense that given two finite simplicial complexes K and L for which there exists a simplicial homeomorphism $K \to L$, then $\chi(K) = \chi(L)$. This allows one to define the Euler characteristic for smooth manifolds.

Definition 16.71. Let M be a smooth compact manifold and $(K, h : K \to M)$ a triangulation of M. The <u>*Euler characteristic of M*</u> is the Euler characteristic of K, i.e.

$$\chi(M) := \chi(K). \tag{16.72}$$

Example 16.73. A triangulation of the sphere S^m was described in Example 16.68. It was shown that $\partial\Delta^{m+1} \cong S^m$. The number of i-simplices of $\partial\Delta^{m+1}$ was calculated in Example 16.65. Since the superscript is slightly difference, we write this number here explicitly

$$\big|(\partial\Delta^{m+1})_i\big| = \binom{m+2}{i+1} \qquad \forall\, i \in \{0, 1, \ldots, m\}. \tag{16.74}$$

Therefore,

$$\chi(S^m) = \chi\big(\partial\Delta^{m+1}\big) = \sum_{i=0}^{m+1}(-1)^i\binom{m+2}{i+1} = \begin{cases} 2 & \text{if } m \text{ even} \\ 0 & \text{if } m \text{ odd} \end{cases}. \tag{16.75}$$

Exercise 16.76. Draw a triangulation of the two-dimensional torus. Then, use this to calculate the Euler characteristic of the torus.

Theorem 16.77 (Poincaré-Hopf)**.** *Let M be a compact manifold, let W be a smooth tangent vector field on M with isolated critical points, and let* $\mathrm{CP}(W)$ *denote the set of critical points of W. Then*

$$\sum_{c \in \mathrm{CP}(W)} \iota(W, c) = \chi(M). \tag{16.78}$$

Proof. A part of this proof is provided in Milnor [8]. ∎

This theorem says a lot. First of all, the Euler characteristic has no reference to any vector fields whatsoever. Therefore, the sum of the indices of the critical points of *any* smooth tangent vector field on M (whose critical points are isolated) is always the *same*! Furthermore, not only are they the same, but they don't even depend on the smooth structure of M. Only the topology, namely the combinatorial structure used to describe a triangulation of M, are used in defining the Euler characteristic. This theorem is an example of a theorem that relates different fields of mathematics. In this case, this theorem relates analysis and smooth manifolds to combinatorics and topology. Much of modern mathematics is built on the interplay between different fields.

Exercise 16.79. In Theorem 16.77, it was implicitly assumed that $\mathrm{CP}(W)$ is a finite set. Prove that this is true under the assumptions in the statement of that theorem.

Theorem 16.80. *Fix $m \in N$. Every tangent vector field on S^{2m} has at least one critical point.*

Proof. Suppose, to the contrary, that V is a tangent vector field on S^{2m} for which $V_x \neq 0$ for all $x \in S^{2m}$. Then, the set of critical points $\mathrm{CP}(V)$ is empty and

$$\sum_{c \in \mathrm{CP}(V)} \iota(V, c) = 0. \tag{16.81}$$

However, according to Example 16.73, an even-dimensional sphere has Euler characteristic of 2. Hence, by Theorem 16.77,

$$\sum_{c \in \mathrm{CP}(V)} \iota(V, c) = \chi(S^{2m}) = 2. \tag{16.82}$$

This is a contradiction since $2 \neq 0$. ∎

Exercise 16.83. Fix $m \in \mathbb{N}$. Construct a smooth non-vanishing tangent vector field on S^{2m-1}.

Exercise 16.84. Fix $m \in \mathbb{N}$. The antipodal map $a : S^m \to S^m$ is defined by

$$S^m \ni (x_1, x_2, \ldots, x_{m+1}) \mapsto (-x_1, -x_2, \ldots, -x_{m+1}). \tag{16.85}$$

(a) Prove that the antipodal map $a : S^m \to S^m$ is smoothly homotopic to the identity map $\mathrm{id} : S^m \to S^m$ whenever m is odd.

(b) Is a smoothly homotopic to id if m is even? Explain.

Theorem 16.86 (Hopf)**.** *Let M be a connected, oriented, compact m-dimensional manifold. Two smooth maps $f, g : M \to S^m$ are smoothly homotopic if and only if* $\deg(f) = \deg(g)$.

Proof. See Chapter 7 in Milnor's book [8]. ■

Remark 16.87. As innocent as this theorem seems, the general study of homotopy classes of maps into or out of spheres is an open problem.

Exercise 16.88. Let M be a smooth m-dimensional manifold and let $f, g : M \to S^p$ be two smooth maps satisfying

$$\big|f(x) - g(x)\big| < 2 \qquad \forall\, x \in M. \tag{16.89}$$

Prove that f is smoothly homotopic to g.

Problems.

From these notes: Exercises 16.9, 16.20, 16.42, 16.61, 16.76, 16.79, 16.83, 16.84, 16.88

17 April 11: Complete Metric Spaces

We will start using parts of Kolmogorov and Fomin [7] now with occasional other references such as Bryant [5].

Definition 17.1. A *metric space* consists of a set X together with a function $d : X \times X \to \mathbb{R}$ satisfying

(a) $d(x, y) = 0$ if and only if $x = y = 0$,

(b) $d(x, y) = d(y, x)$ for all $x, y \in X$ (*reflexivity*), and

(c) $d(x, z) \le d(x, y) + d(y, z)$ for all $x, y, z \in X$ (*triangle inequality*).

A metric space as above will be denoted by the pair (X, d).

Note that it follows from these conditions that the distance is always non-negative, namely $d(x, y) \ge 0$ for all $x, y \in X$ (*positivity*). As the notation suggests, a set X may be equipped with several different kinds of (inequivalent) metrics.

Definition 17.2. Let (X, d) be a metric space, let $x \in X$, and let $r > 0$. The *sphere of radius r centered at x* is the set

$$S_r(x) := \big\{y \in X \; : \; d(x, y) = r\big\}. \tag{17.3}$$

The *open ball of radius r centered at x* is the set

$$B_r(x) := \big\{y \in X \; : \; d(x, y) < r\big\}. \tag{17.4}$$

We will often use spheres to give us an idea of what the distance in a metric space looks like. Kolmogorov and Fomin offer several examples of metric spaces [7]. In addition to those examples, we also have the following.

Example 17.5. Let $K \subseteq \mathbb{R}^k$ be a compact subset of Euclidean space and let $C(K)$ denote the set of real-valued continuous functions on K. For any two such functions $f, g : K \to \mathbb{R}$, set

$$d_{\text{sup}}(f, g) := \sup_{x \in K} \Big\{\big|f(x) - g(x)\big|\Big\}. \tag{17.6}$$

Note that because K is compact and the functions f and g are continuous, $d(f, g)$ exists. Then $\big(C(K), d_{\text{sup}}\big)$ is a metric space. The first two conditions are immediately satisfied. For the last condition, note that for any three functions $f, g, h : K \to \mathbb{R}$,

$$\big|f(x) - h(x)\big| = \big|f(x) - g(x) + g(x) - h(x)\big| \le \big|f(x) - g(x)\big| + \big|g(x) - h(x)\big| \qquad \forall\, x \in K \tag{17.7}$$

by the triangle inequality in $\mathbb{R}$. Hence, by applying $\sup_{x \in K}$ to both sides,

$$\begin{aligned} \sup_{x \in K} \Big\{\big|f(x) - h(x)\big|\Big\} &\le \sup_{x \in K} \Big\{\big|f(x) - g(x)\big| + \big|g(x) - h(x)\big|\Big\} \\ &= \sup_{x \in K} \Big\{\big|f(x) - g(x)\big|\Big\} + \sup_{x \in K} \Big\{\big|g(x) - h(x)\big|\Big\} \end{aligned} \tag{17.8}$$

since the supremum of the sum of two non-negative quantities is the sum of the suprema (because both exist). Since we will describe other metrics on spaces of continuous functions, we denote this one by d_{sup} instead of just d.

Exercise 17.9. Let $X := \mathbb{R}^2$ but with a distance defined for $x = (x_1, x_2)$ and $y = (y_1, y_2)$ by

$$d(x,y) := \begin{cases} |x_1 - y_1| & \text{if } x_2 = y_2 \\ |x_1| + |x_2 - y_2| + |y_1| & \text{if } x_2 \neq y_2. \end{cases} \tag{17.10}$$

(a) Calculate the distances $d(x,y)$ between the following points $x, y \in \mathbb{R}^2$.

i. $x = (-1,1)$ and $y = (1,1)$.

ii. $x = (-1,1)$ and $y = (1,-1)$.

iii. $x = (0,1)$ and $y = (0,-1)$.

(b) Draw the unit sphere centered at $(0,0)$ with respect to this metric.

(c) Draw the unit sphere centered at $(1,1)$ with respect to this metric.

(d) Prove that (X,d) is a metric space.

Exercise 17.11. Let $X := \mathbb{R}^2$ but with a distance defined for $x = (x_1, x_2)$ and $y = (y_1, y_2)$ by

$$d(x,y) := \begin{cases} 0 & \text{if } x = y \\ \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2} & \text{if } x \neq y. \end{cases} \tag{17.12}$$

(a) Draw the unit sphere centered at $(0,0)$ with respect to this metric.

(b) Draw the unit sphere centered at $\left(\frac{1}{2}, 0\right)$ with respect to this metric. More generally, let $x \in B_1(0)$. Describe the unit sphere centered at x.

(c) Let $x \in \mathbb{R}^2 \setminus B_1(0)$. Describe the unit sphere centered at x.

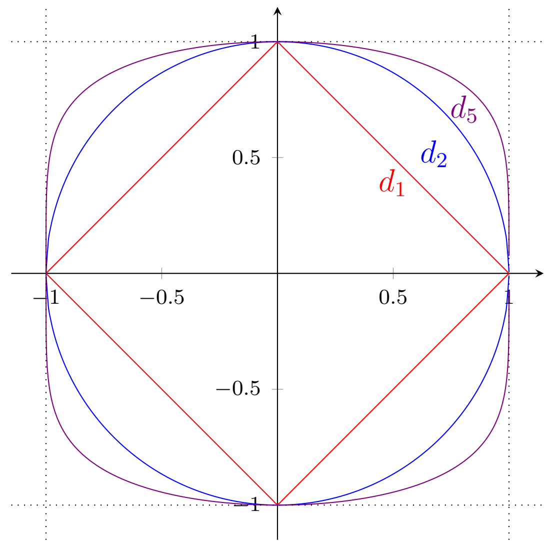
(d) Prove that (X,d) is a metric space.

(e) Find two points $x, y \in X$ and two numbers $r, s > 0$ with $r > s$ but with $B_r(x) \subsetneq B_s(y)$.

Example 17.13. Fix $p \geq 1$ and $n \in \mathbb{N}$. Let $X := \mathbb{R}^n$ but with a distance defined by

$$X \times X \ni (x,y) \mapsto d_p(x,y) := \left(\sum_{i=1}^{n} |y_i - x_i|^p \right)^{1/p}. \tag{17.14}$$

Kolmogorov and Fomin prove that this is a metric space in Example 10 in Section 6 [7]. Instead of redoing this proof, let us illustrate the unit spheres centered at 0 in (X, d_p) for various values of p.



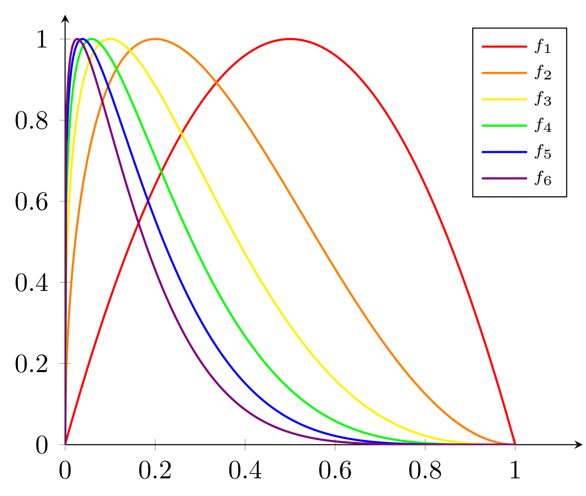
There is even a “limit” of these metrics as $p \to \infty$. Namely,

$$X \times X \ni (x, y) \mapsto d_\infty(x, y) := \max_{i \in \{1,\ldots,n\}} \big\{ |x_i - y_i| \big\}. \tag{17.15}$$

What does the unit sphere look like with respect to the metric d_∞?

Example 17.16. Let $C([0,1])$ be the set of continuous functions on the closed interval $[0,1]$. We have already described one type of metric that can be put on $C([0,1])$ in Example 17.5. One of the properties of this metric, however, is that the distance between the function 0 and functions from the following sequence of functions

$$\mathbb{N} \times [0,1] \ni (n, x) \mapsto f_n(x) := \left(\frac{n^2+1}{n^2} \right)^n \sqrt[n]{n^2+1}\, \sqrt[n]{x}(1-x)^n. \tag{17.17}$$



does not tend to zero even though the functions converge pointwise to the zero function. In fact, with respect to d_{sup}, the distance between 0 and these functions is always 1:

$$d_{\text{sup}}(0, f_n) = 1 \qquad \forall\, n \in \mathbb{N} \tag{17.18}$$

even though we know that as functions

$$\lim_{n\to\infty}\big(f_n(x)\big) = 0 \qquad \forall\, x \in [0,1]. \tag{17.19}$$

The reason for this is that $d_{\sup}$ is the metric in which sequences (of functions) converge uniformly. We will make this more precise when discussing sequences. Instead, let us define another metric $d_{\int}$ on $C([0,1])$ given by

$$C([0,1]) \times C([0,1]) \ni (f,g) \mapsto d_{\int}(f,g) := \int_0^1 \big|f(x) - g(x)\big| dx. \tag{17.20}$$

Exercise 17.21. Prove that $\big(C([0,1]), d_{\int}\big)$ is a metric space and show that

$$\lim_{n\to\infty}\big(d_{\int}(0, f_n)\big) = 0, \tag{17.22}$$

where the $\{f_n\}$ are given by (17.17).

Definition 17.23. Let (X,d) be a metric space and $x : \mathbb{N} \to X$ a sequence. x is a *Cauchy sequence* in (X,d) iff for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$d(x_n, x_m) < \epsilon \qquad \forall\, n, m \geq N. \tag{17.24}$$

x *converges* in (X,d) to a point $\lim x \in X$ iff for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$d\left(\lim x, x_n\right) < \epsilon \qquad \forall\, n \geq N. \tag{17.25}$$

A sequence that converges is said to be *convergent*. (X,d) is *complete* iff every Cauchy sequence in (X,d) converges.

Theorem 17.26. *Let (X,d) be a metric space and let $x : \mathbb{N} \to X$ be a convergent sequence in (X,d). Then x is a Cauchy sequence.*

Proof. See Theorem 1 in Section 7 of [7]. ■

Example 17.27. Let $\big(C([0,1]), d_{\sup}\big)$ be the metric space from Example 17.5. As one may have guessed from the discussion in Example 17.16, a sequence of continuous functions $f : \mathbb{N} \to C([0,1])$ converges to $\lim f \in C([0,1])$ if and only if f converges to $\lim f$ *uniformly* (recall [1]). Furthermore, $\big(C([0,1]), d_{\sup}\big)$ is complete. This result follows from the Cauchy Criterion for uniform convergence and the Continuous Limit Theorem, which implies that the limit of a sequence of continuous functions is continuous with respect to the metric $d_{\sup}$ (recall [1]).

We know from Analysis I that $\mathbb{R}$ is a complete metric space and that $\mathbb{Q}$ is not complete. We just saw that $C([0,1])$ is complete with respect to the sup metric. What about if we measure the distance between two functions by the area between their graphs?

Exercise 17.28. Prove that $\big(C([0,1]), d_{\int}\big)$ appearing in Example 17.16 is not a complete metric space by going through the following outline.

(a) Consider the sequence of functions $f : \mathbb{N} \to C([0,1])$ given by

$$\mathbb{N} \times [0,1] \ni (x,n) \mapsto f_n(x) := \begin{cases} 0 & \text{if } 0 \le x \le \frac{1}{2} \\ (n+1)\left(x - \frac{1}{2}\right) & \text{if } \frac{1}{2} \le x \le \frac{1}{2} + \frac{1}{1+n} \\ 1 & \text{if } \frac{1}{2} + \frac{1}{1+n} \le x \le 1 \end{cases} \tag{17.29}$$

Compute $d_{\int}(f_n, f_m)$ for arbitrary $n, m \in \mathbb{N}$ and prove that f is a Cauchy sequence.

(b) Let g be the function

$$[0,1] \ni x \mapsto g(x) := \begin{cases} 0 & \text{if } 0 \le x \le \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} < x \le 1 \end{cases} \tag{17.30}$$

Note that g is not in $C([0,1])$ but it is still integrable. Show that

$$\lim_{n\to\infty} \int_0^1 \big|f_n(x) - g(x)\big| dx = 0. \tag{17.31}$$

(c) Using these results, prove, by contradiction, that there does not exist a continuous function $h \in C([0,1])$ such that $\lim_{n\to\infty} d_{\int}(f_n, h) = 0$.

We have so far describe the collection of mathematical objects known as metric spaces. There is a natural notion of structure-preserving function between such mathematical objects.

Definition 17.32. Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $f : X \to Y$ is an *isometry* iff

$$d_Y\big(f(x), f(x')\big) = d_X(x, x') \qquad \forall\, x, x' \in X. \tag{17.33}$$

In other words, a function f is an isometry iff the diagram

$$\begin{tikzcd} Y \times Y \arrow[dr, "d_Y"'] & & X \times X \arrow[ll, "f\times f"'] \arrow[dl, "d_X"] \\ & \mathbb{R} & \end{tikzcd} \tag{17.34}$$

commutes. f is an *isometric isomorphism* if, in addition, there exists an isometry $g : Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$.

Note that an isometry is automatically one-to-one. Also be aware that my definition of isometry is different from that of Kolmogorov and Fomin (I do not require an isometry to be surjective). If a metric space is not complete, there is a canonical way to enlarge it so that the resulting metric space is complete and the original space sits inside the large one isometrically. To describe this more precisely, we define open, closed, and dense subsets of a metric space.

Definition 17.35. Let (X, d) be a metric space. A subset $A \subseteq X$ is *open* iff for every $x \in A$, there exists an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq A$. A subset $C \subseteq X$ is *closed* iff $X \setminus C$ is open.

Note that an open ball $B_r(x)$ is indeed open for all $x \in X$ and $r > 0$. The collection of open and closed subsets of a metric space obey the usual conditions of a topology.

Theorem 17.36. *Let (X, d) be a metric space.*

(a) X and $\varnothing$ are open subsets of X.

(b) The union of an arbitrary collection of open sets is open.

(c) The intersection of a finite collection of open sets is open.

Similarly,

(a) X and $\varnothing$ are closed subsets of X.

(b) The union of an finite collection of closed sets is closed.

(c) The intersection of an infinite collection of closed sets is closed.

Exercise 17.37. Prove Theorem 17.36.

Definition 17.38. Let (X, d) be a metric space. The *closure* of a set $A \subseteq X$ is the intersection of all closed subsets of X containing A. It is denoted by $\overline{A}$.

From Theorem 17.36, it immediately follows that $\overline{A}$ is closed. In fact, for any other closed set C that contains A, $\overline{A} \subseteq C$. Hence, $\overline{A}$ is the smallest closed set containing A. There is another way to describe the closure in terms of limit points.[36]

Definition 17.39. Let (X, d) be a metric space and let $A \subseteq X$. A *limit point* of A is a point $x \in X$ such that

$$A \cap B_\epsilon(x) \setminus \{x\} \neq \varnothing \qquad \forall\, \epsilon > 0. \tag{17.40}$$

Theorem 17.41. *Let (X, d) be a metric space and let $A \subseteq X$. Then*

$$\overline{A} = \big\{ x \in X \ : \ x \in A \text{ or } x \text{ is a limit point of } A \big\}. \tag{17.42}$$

Proof. Informal exercise. ■

Note that from this theorem and the fact that the closure of a closed set is itself, it follows that A is closed if and only if A contains its limit points.

Definition 17.43. Let (X, d) be a metric space. A *completion* of (X, d) consists of a complete metric space $(\tilde{X}, \tilde{d})$ together with an isometry $\eta : (X, d) \to (\tilde{X}, \tilde{d})$ satisfying the following universal property. for any other complete metric space (Z, ρ) with an isometry $\psi : (X, d) \to (Z, \rho)$, there exists a unique isometry $h : (\tilde{X}, \tilde{d}) \to (Z, \rho)$ such that the diagram

$$\begin{tikzcd} (X,d) \arrow[dd, "\psi"'] \arrow[dr, "\eta"] & \\ & (\tilde{X}, \tilde{d}) \arrow[dl, "h"] \\ (Z,\rho) & \end{tikzcd} \tag{17.44}$$

commutes.

[36]Warning: Kolmogorov and Fomin call use the terminology "contact point" to refer to what I called "limit point" last semester. They use the terminology "limit point" to what some people call an "accumulation point." I want to stick to my notes from last semester, so I will use the terminology "limit point" as I have meant it here. Therefore, my terminology conflicts with Kolmogorov and Fomin. Be warned.

It is immediately from this definition that if there are two completions, then they are canonically isometrically isomorphic (this follows from the universal property). Therefore, one can call it "the" completion of a metric space. In particular, it immediately follows from this definition that if (X, d) is already complete, $(\tilde{X}, \tilde{d})$ is canonically isomorphic to (X, d).

Theorem 17.45. *Let (X, d) be a metric space. A completion $\eta : (X, d) \to (\tilde{X}, \tilde{d})$ of (X, d) exists. Furthermore, $\overline{\eta(X)} = \tilde{X}$.*

In other words, the completion is the "smallest" complete metric space containing the original metric space.

Proof. See Theorem 4 in Section 7 of Kolmogorov and Fomin [7]. However, one of the main differences between Theorem 17.45 and Theorem 4 in [7] is the uniqueness property described by the universal property in Definition 17.43. One still constructs $\tilde{X}$ as equivalence classes of Cauchy sequences in X and $\tilde{d}$ in terms of d acting on representatives. For any other completion $\psi : (X, d) \to (Z, \rho)$, we still have to construct a unique isometry $h : (\tilde{X}, \tilde{d}) \to (Z, \rho)$ satisfying the required diagram. Let $[x] \in \tilde{X}$ be such an equivalence class with a representative Cauchy sequence $x : \mathbb{N} \to X$. Define

$$h\big([x]\big) := \lim_{n\to\infty} \psi(x_n). \tag{17.46}$$

Then h is well-defined and is the required isometry. The details are requested in the following exercise. ∎

Exercise 17.47. Using the notation of the proof of Theorem 17.45, complete the following tasks.

(a) Prove that $\big(\psi(x_1), \psi(x_2), \dots\big)$ is a Cauchy sequence in (Z, ρ) and therefore converges.

(b) Prove that h is well-defined. Namely, for any other representative $y : \mathbb{N} \to X$ of $[x]$ (which means that $\lim_{n\to\infty}(y_n - x_n) = 0$), prove that the sequences $\big(\psi(x_1), \psi(x_2), \dots\big)$ and $\big(\psi(y_1), \psi(y_2), \dots\big)$ converge to the same point in Z.

(c) Prove that h is an isometry.

(d) Prove that the diagram (17.44) commutes.

(e) Explain why h is the unique isometry such that the diagram (17.44) commutes.

Exercise 17.48. Describe the completion of $\big(C([0, 1]), d_{\int}\big)$ more concretely. For example, can it be described as a collection of functions on $[0, 1]$ satisfying some property? Can the distance between such functions be described more explicitly? Then show that your metric space is indeed the completion of $\big(C([0, 1]), d_{\int}\big)$. [Warning: I do not know the answer to this, and therefore do not know how difficult this is.]

For many purposes, isometries are too rigid in the sense that they preserve the distance between points. Even continuous functions $\mathbb{R} \to \mathbb{R}$ rarely satisfy this property. There is a weaker notion of structure-preserving function between metric spaces that also appears frequently. This comes from the fact that $\mathbb{R}$ has an ordering and a notion of discs of varying radii at points.

Definition 17.49. Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $f : X \to Y$ is continuous at $x \in X$ iff for every $\epsilon > 0$, there exists a $\delta > 0$ such that if for any $x' \in X$ satisfying $d_X(x, x') < \delta$, then $d_Y\big(f(x), f(x')\big) < \epsilon$. f is continuous iff f is continuous at x for all $x \in X$.

Of course, one might come up with other definitions of structure-preserving functions between metric spaces as well. The following one will be useful in the next lecture.

Definition 17.50. Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $f : X \to Y$ is Lipschitz iff there exists a number $M \geq 0$ such that

$$d_Y\big(f(x), f(x')\big) \leq M d_X(x, x') \qquad \forall\, x, x' \in X. \tag{17.51}$$

f is a contraction iff there exists a number $M \in [0, 1)$ such that (17.51) holds. f is distance decreasing iff

$$d_Y\big(f(x), f(x')\big) < d_X(x, x') \qquad \forall\, (x, x') \in (X \times X) \setminus \Delta(X). \tag{17.52}$$

Using Problem 6.4 [7], one can easily show that a Lipschitz is always continuous. A continuous function of metric spaces preserves only the *topology* of a metric space (and not the distance). This can be made precise using the notion of open sets.

Exercise 17.53. Let (X, d_X) and (Y, d_Y) be two metric spaces. Prove that $f : (X, d_X) \to (Y, d_Y)$ is continuous if and only if for any open set $U \subseteq Y$, the subset $f^{-1}(U)$ is an open subset of X.

Exercise 17.54. Let (X, d) be a metric space.

(a) Prove that $\big|d(x, z) - d(y, u)\big| \leq d(x, y) + d(z, u)$ for all $x, y, z, u \in X$.

(b) Prove that $\big|d(x, z) - d(y, z)\big| \leq d(x, y)$ for all $x, y, z \in X$.

(c) Prove that if $x, y : \mathbb{N} \to X$ are sequences converging to $\lim x$ and $\lim y$, respectively, then

$$\lim_{n\to\infty} d(x_n, y_n) = d(\lim x, \lim y). \tag{17.55}$$

Remark 17.56. The completion of a metric space is an example of an adjunction of categories. The details will not be provided here, but the words will be provided in case the reader is curious. Let MSI denote the category of metric spaces with isometries. Let CMSI denote the category of complete metric spaces with isometries. Let F : CMSI $\to$ MSI denote the functor that simply ignores whether a metric space is complete. The completion is then a functor C : MSI $\to$ CMSI together with natural transformations $\eta : \mathrm{id}_{\mathrm{MSI}} \Rightarrow F \circ C$ and $\epsilon : C \circ F \Rightarrow$ CMSI satisfying what are known as the "zig-zag" identities. η is nothing but the η we have provided above, but for a specific metric space (X, d) so it is more precise to denote it by $\eta_{(X,d)} : (X, d) \to (\tilde{X}, \tilde{d}) \equiv F\big(C(X, d)\big)$. ϵ in our case is actually invertible because if a metric space is already complete, the isometry is an isomorphism. Since adjunctions are unique up to unique isomorphism, the definition of the completion of a metric space can therefore simply be given as "the left adjoint of the forgetful functor F."

Exercise 17.57. Look up the definition of an adjunction of categories and explain the above Remark in such a way so that you understand it. In particular, prove that the completion described in this lecture satisfies all the conditions of an adjunction.

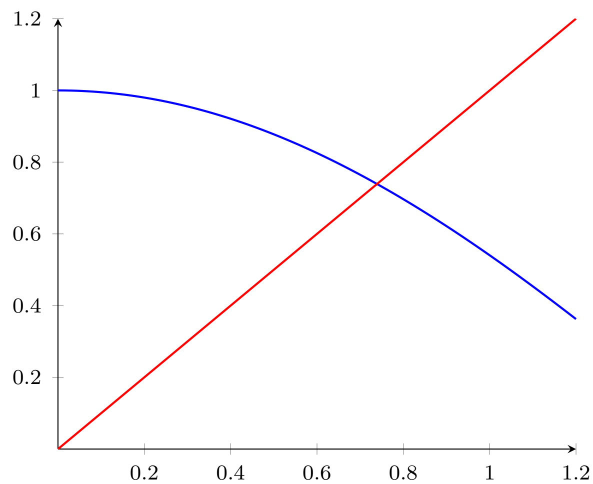
Problems.
From Kolmogorov and Fomin [7]: Problems 6.4, 6.11, 7.7, 7.8, 7.9
From these notes: Exercises 17.9, 17.11, 17.21, 17.28, 17.37, 17.47, 17.48, 17.53, 17.54, 17.57

18 April 13: The Contraction Mapping Theorem

In this section, we will state and prove three variants of the contraction mapping theorem for metric spaces. This section mostly follows [5]. We will see several applications of this theorem in later lectures.

Definition 18.1. Let $f : X \to X$ be a function from a set X to itself. A *fixed point* for f is a point $x \in X$ such that $f(x) = x$.

Example 18.2. The fixed points of the function $\cos : \mathbb{R} \to \mathbb{R}$ can be obtained by solving $\cos(x) = x$ for the variable x. One way to solve this system is to first guess. Let x_0 be a guess, some value slightly less than $\frac{\pi}{2}$, say.



An observation shows that consecutively applying the function cos to the initial guess brings us closer and closer to the actual solution. In fact, the sequence $x : \{0\} \cup \mathbb{N} \to \mathbb{R}$ given by

$$(x_0, \cos(x_0), \cos(\cos(x_0)), \cos(\cos(\cos(x_0))), \dots) \tag{18.3}$$

is a Cauchy sequence and converges to the fixed point! If y is the solution to $\cos(y) = y$, then from this picture and we would expect that x_{n+1} is closer to y than x_n, i.e.

$$|y - x_{n+1}| < |y - x_n| \tag{18.4}$$

for all $n \in \mathbb{N}$. Plugging in $y = \cos(y)$ and $x_{n+1} = \cos(x_n)$, this inequality becomes

$$\big|\cos(y) - \cos(x_n)\big| < |y - x_n| \tag{18.5}$$

for all $n \in \mathbb{N}$. However, this deduction came from a single guess. Since we do not know what the fixed point is, it is natural to require that this property holds for all distinct points, namely

$$\big|\cos(y) - \cos(x)\big| < |x - y| \qquad \forall\ (x, y) \in (\mathbb{R} \times \mathbb{R}) \setminus \Delta(\mathbb{R}). \tag{18.6}$$

Note that here $\Delta(\mathbb{R}) := \{(x, x) \in \mathbb{R}^2\}$. Is this true of the cosine function? It turns out the answer is yes.

Exercise 18.7.

(a) Prove equation (18.6). [Hint: you will need to use some trigonometric identities.]

(b) Prove that sin is distance decreasing as well.

The goal of today's lecture is to abstract the key properties of the previous example that allows one to be sure that a fixed point exists. One might guess that a function $f : (X, d) \to (X, d)$ on a metric space satisfying a condition such as

$$d\big(f(x), f(y)\big) < d(x, y) \qquad \forall\ (x, y) \in (X \times X) \setminus \Delta(X) \tag{18.8}$$

would be enough to guarantee the existence of a fixed point of f. Unfortunately, these assumptions are not enough as the following Exercise indicates [5].

Exercise 18.9. Let $f : [1, \infty) \to [1, \infty)$ be the function defined by

$$[1, \infty) \ni x \mapsto f(x) := x + \frac{1}{x} \tag{18.10}$$

and let $g : [1, \infty) \to [1, \infty)$ be the function defined by

$$[1, \infty) \ni x \mapsto g(x) := \frac{25}{26}\left(x + \frac{1}{x}\right). \tag{18.11}$$

(a) Prove that f is distance decreasing, i.e. satisfies

$$\big|f(x) - f(y)\big| < |x - y| \qquad \forall\ (x, y) \in \big([1, \infty) \times [1, \infty)\big) \setminus \Delta[1, \infty). \tag{18.12}$$

(b) Prove that f does not have a fixed point.

(c) Show that g is a contraction, and in fact, satisfies

$$\big|g(x) - g(y)\big| \le \frac{25}{26}|x - y| \qquad \forall\ x, y \in [1, \infty). \tag{18.13}$$

(d) Find the fixed point of g. Note that it is unique (we will prove a general uniqueness theorem later).

This previous example illustrates that maybe instead of requiring a condition such as 18.12, a condition that might guarantee the existence of a unique fixed point is that of a contraction. What about the possible spaces? Metric spaces have a notion of a distance, so perhaps all we need is the notion of a contraction on a metric space?

Exercise 18.14. Let $X := \mathbb{Q} \cap [1, \infty)$ and let $f : X \to X$ be the function defined by

$$X \ni x \mapsto f(x) := \frac{x}{2} + \frac{1}{x}. \tag{18.15}$$

(a) Prove that f satisfies

$$\big|f(x) - f(y)\big| \le \frac{1}{2}|x - y| \qquad \forall\ x, y \in X \tag{18.16}$$

so that f is a contraction on X.

(b) Prove that f does not have a fixed point.

The problem with the previous example is that the domain is not a complete metric space. However, assuming we have a contraction on a complete metric space, then a unique fixed point is guaranteed to exist.

Theorem 18.17 (Contraction Mapping Theorem I)**.** *Let (X, d) be a complete metric space and let $f : (X, d) \to (X, d)$ be a contraction. Then there exists a unique fixed point of f, i.e. a unique point $x \in X$ such that $f(x) = x$. In fact, for any $x_0 \in X$, the sequence*

$$\mathbb{N} \ni n \mapsto x_n := f^n(x_0) \tag{18.18}$$

converges to x.

This is one of those theorems whose proof is useful on its own. Please understand it well.

Proof. See Section 8 Theorem 1 of [7]. ■

Exercise 18.19. Let $\Big(C\big([0, \tfrac{1}{2}]\big), d_{\text{sup}}\Big)$ be the metric space of continuous functions on the closed interval $\big[0, \tfrac{1}{2}\big]$ equipped with the sup norm. Let $\Phi : C\big([0, \tfrac{1}{2}]\big) \to C\big([0, \tfrac{1}{2}]\big)$ be the function defined by sending a function $f \in C\big([0, \tfrac{1}{2}]\big)$ to the function $\Phi(f)$ defined by

$$\left[0, \frac{1}{2}\right] \ni x \mapsto x\big(f(x) + 1\big). \tag{18.20}$$

(a) Prove that Φ is a contraction on $C\big([0, \tfrac{1}{2}]\big)$.

(b) Let $p_1 \in C\big([0, \tfrac{1}{2}]\big)$ be the function defined by sending $x \in \big[0, \tfrac{1}{2}\big]$ to $p_1(x) = x$. Calculate the functions in the sequence $p : \mathbb{N} \to C\big([0, \tfrac{1}{2}]\big)$ given by

$$\mathbb{N} \ni n \mapsto p_n := \Phi^{n-1}(p_1) \tag{18.21}$$

explicitly.

(c) Prove that the sequence p converges to the function $\lim p$ (with respect to the metric d_{sup}) defined by

$$\left[0, \frac{1}{2}\right] \ni x \mapsto \frac{x}{1 - x} \tag{18.22}$$

and prove that $\lim p$ is the unique fixed point of Φ.

Differentiable functions have a characterizing property for being contractions.

Exercise 18.23. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to [a, b]$. Prove that f is a contraction if and only if there exists a number $k < 1$ such that $\big|f'(x)\big| \le k$ for all $x \in [a, b]$. [Hint: the Mean Value Theorem is helpful here.]

Exercise 18.24.

(a) Prove that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := \cos(x)$ is not a contraction.

(b) Prove that cos restricted to the domain $[0,1]$ has range contained in $[0,1]$ and is in fact a contraction (Hint: use the Mean Value Theorem).[37]

(c) Prove that the function $g : \mathbb{R} \to \mathbb{R}$ given by $h(x) := \cos(\cos(x))$ is a contraction.

(d) Prove that the function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) := \sin(\sin(\cdots\sin(x)\cdots))$ is not a contraction for any number of sines.

This previous example motivates the following generalization of the Contraction Mapping Theorem.

Theorem 18.25 (Contraction Mapping Theorem II). *Let (X,d) be a complete metric space and let $f : (X,d) \to (X,d)$ be a function for which there exists a natural number $N \in \mathbb{N}$ such that f^N (the N-th iterate of f) is a contraction. Then there exists a unique fixed point of f, i.e. a unique point $x \in X$ such that $f(x) = x$. In fact, for any $x_0 \in X$, the sequence*

$$\mathbb{N} \ni n \mapsto x_n := f^n(x_0) \tag{18.26}$$

converges to x.

Proof. See Section 8 Theorem 1′ of [7] (though, see Theorem 4.3 in [5] for a more intuitive proof). ■

Exercise 18.27. Let $f : \mathbb{R} \to \mathbb{R}$ be the function given by $f(x) := e^{-x}$.

(a) Prove that f is not a contraction.

(b) Prove that f^2, given by $f^2(x) := e^{-e^{-x}}$, is a contraction.

(c) Use an iterative technique, modeled after the proof of Theorem 18.25 to approximate the unique fixed point of the expression

$$-\log(x) = x \tag{18.28}$$

to three decimal places.

Even though sin and none of its iterates are contractions, it is quite obvious that $x = 0$ is the unique fixed point of the equation $x = \sin(x)$. So far, none of our theorems cover this case. This conclusion will follow from a sufficient condition that guarantees the existence of fixed points on *compact* metric spaces.

Theorem 18.29. *Let (X,d) be a metric space and let $A \subseteq X$ be a subset. The the following are equivalent.*

(a) Every sequence in A has a convergence subsequence whose limit is contained in A.

(b) Every open cover of A has a finite subcover.

[37] I'd like to thank Keith Conrad for pointing this out. See Example 2.6 in his notes http://www.math.uconn.edu/~kconrad/blurbs/analysis/contraction.pdf for a neat proof.

Proof. Informal exercise. ■

The definitions of the above expressions are naturally generalized from the ones in Euclidean space.

Definition 18.30. Let (X, d) be a metric space. A subset $A \subseteq X$ satisfying any (and hence all) of the equivalent conditions of Theorem is said to be *compact*.

In $\mathbb{R}^n$, we had a theorem that said a subset is compact if and only if it is closed and bounded. This was called the Heine-Borel Theorem (see Theorem 2.41). This Theorem is not true for metric spaces.

Definition 18.31. Let (X, d) be a metric space. A subset $A \subseteq X$ is *bounded* iff there exists a number $D > 0$ such that $d(a, b) \leq D$ for all $a, b \in A$.

Exercise 18.32. Let $A \subseteq C\big([0,1]\big)$ be the subset of continuous functions $f : [0,1] \to \mathbb{R}$ satisfying $f([0,1]) \subseteq [0,1]$.

(a) With respect to the sup metric d_{sup}, prove that $d_{\text{sup}}(f, g) \leq 1$ for all $f, g \in A$.

(b) Prove that A is a closed subset of $C\big([0,1]\big)$.

(c) Hence, A is closed and bounded. Let $h : \mathbb{N} \to A$ be the sequence of functions given by

$$\mathbb{N} \times [0,1] \mapsto f_n(x) := \begin{cases} 2^n x & \text{if } 0 \leq x \leq \frac{1}{2^n} \\ 1 & \text{if } \frac{1}{2^n} < x \leq 1 \end{cases} \tag{18.33}$$

Use this sequence to prove that A is not compact.

We know that the image of a compact subset of Euclidean space under a continuous map is compact. The same is true for continuous functions on compact subsets of metric spaces.

Exercise 18.34. Let (X, d_X) and (Y, d_Y) be two metric spaces, let $A \subseteq X$ be a compact subset of X, and let $f : A \to Y$ be a continuous map. Prove that $f(A)$ is a compact subset of Y.

With compactness now defined for metric spaces, we may state another theorem for fixed points.

Theorem 18.35. *Let (X, d) be a compact metric space and let $f : (X, d) \to (X, d)$ be a function satisfying*

$$d\big(f(x), f(y)\big) < d(x, y) \qquad \forall\, (x, y) \in (X \times X) \setminus \Delta(X). \tag{18.36}$$

Then f has a unique fixed point. In fact, for any $x_0 \in X$, the sequence

$$\mathbb{N} \ni n \mapsto x_n := f^n(x_0) \tag{18.37}$$

converges to x.

Proof. See Theorem 4.4 in [5]. ■

Exercise 18.38. Let $f : [-\pi, \pi] \to [-\pi, \pi]$ be the function given by $f(x) := 1 + \sin(x)$.

(a) Show that $\left|f'(x)\right| < 1$ for all $x \in (-\pi, \pi)$.

(b) Prove that $\left|f(x) - f(y)\right| < |x - y|$ for all distinct $x, y \in [-\pi, \pi]$.

(c) What can you deduce about the equation $1 + \sin(x) = x$? Namely, do any of the theorems in this section guarantee that f has a unique fixed point?

(d) Approximate this fixed point to three decimal places using the method of successive approximations.

Problems.
From Kolmogorov and Fomin [7]: Problems 8.1, 8.2, 8.3, 8.4, 8.5, 8.7
From these notes: Exercises 18.7, 18.9, 18.14, 18.19, 18.23, 18.24, 18.27, 18.32, 18.34, 18.38

19 April 18: Linear Systems

In the next two lectures, we will study differential equations from a more systematic perspective. We begin with solutions to autonomous linear systems and then apply our analysis to the study of an illustrative example. We then use the exponential of an operator to solve non-autonomous linear ordinary differential equations. In the next lecture, we will prove existence and uniqueness for *arbitrary* non-autonomous systems. Due to time constraints, we may occasionally be a bit imprecise and refer to our knowledge from multivariable calculus that we have not yet covered (such as integrals over polygonal regions in Euclidean space). Our references for this material include [2] and [3].

There is actually a general method to solving *any* linear differential equation. Recall from the end of lecture 12 where we introduced the exponential of a matrix.

Theorem 19.1. *Let A be an $m \times m$ matrix. The function $\varphi_A : \mathbb{R} \to M_m(\mathbb{R})$ defined by*

$$\mathbb{R} \ni t \mapsto e^{tA} \tag{19.2}$$

is infinitely differentiable and its derivative at $s \in \mathbb{R}$ is given by the matrix

$$(D_s \varphi_A)(1) \equiv \varphi_A'(s) = Ae^{sA}. \tag{19.3}$$

Proof. Exercise. ∎

Exercise 19.4. Prove Theorem 19.1. [Hint: Show that the sequence of partial sums (viewed as functions of t into $m \times m$ matrices) for the exponential matrix converges uniformly on any interval $[a, b]$ and use the Differentiable Limit Theorem from Analysis I.]

Theorem 19.5. *Let A be an $m \times m$ matrix and consider the linear differential equation given by*

$$\dot{x} = Ax \qquad \text{with} \qquad x(0) = x_0. \tag{19.6}$$

Then, the vector field $\mathbb{R}^m \ni x \mapsto Ax$ is integrable and the function

$$\mathbb{R} \ni t \mapsto e^{tA} x_0 \tag{19.7}$$

is the unique solution to (19.6).

Proof. Exercise. ∎

Exercise 19.8. Prove Theorem 19.5.

This theorem provides us with a formal result for solutions to autonomous linear ordinary differential equations. However, it is of very little practical significance because computing the exponential of a matrix is quite difficult if done directly. Instead, one often computes the exponential of a matrix in simple situations and then uses methods to extrapolate these results to less generic cases.

Definition 19.9. An $m \times m$ matrix J is in Jordan (normal) form iff

$$J = \bigoplus_{i=1}^{n} \begin{bmatrix} \lambda_i & * & 0 & 0 & \cdots & 0 \\ 0 & \lambda_i & * & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & \lambda_i & * & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_i & * \\ 0 & \cdots & \cdots & 0 & 0 & \lambda_i \end{bmatrix}, \tag{19.10}$$

where $\{\lambda_1, \ldots, \lambda_n\}$ are n distinct complex numbers and the $*$'s are either 0 or 1. The $*$'s are furthermore written so that the 1's appear first and the 0's afterwards (for any given fixed i).

Theorem 19.11 (Jordan decomposition for matrices)**.** *Let A be an $m \times m$ matrix. There exists an invertible matrix S and a Jordan matrix J such that $A = SJS^{-1}$.*

Proof. See any reasonable book on linear algebra. ■

Exercise 19.12. Compute the exponentials of the following $m \times m$ matrices.

(a) An $m \times m$ diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_m \end{bmatrix} \tag{19.13}$$

with $\{\lambda_i\}_{i \in \{1,\ldots,m\}}$ fixed complex numbers.

(b) An $m \times m$ nilpotent matrix of the form

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \tag{19.14}$$

(c) An $m \times m$ Jordan block matrix

$$\begin{bmatrix} t\lambda & t & 0 & \cdots & 0 \\ 0 & t\lambda & t & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t\lambda & t \\ 0 & \cdots & 0 & 0 & t\lambda \end{bmatrix} \tag{19.15}$$

where λ is a fixed complex number and t is a fixed real number.

(d) A block matrix of the form

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{19.16}$$

with A an $m \times m$ matrix and B an $n \times n$ matrix (the 0's are matrices of appropriate size).

(e) Prove that for an $m \times m$ matrix B and an invertible $m \times m$ matrix S,

$$\exp(SBS^{-1}) = S\exp(B)S^{-1}. \tag{19.17}$$

If it is not clear to you how to do these for arbitrary m, try to do the cases $m = 2$ and $m = 3$.

Example 19.18. Consider the ordinary differential equation in $\mathbb{R}^2$ of the form

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= -x - ky \end{aligned} \tag{19.19}$$

with initial condition (x_0, y_0) and $k > 0$. This system describe a one-dimensional oscillator with friction (the x variable is interpreted as the position and the y variable as the velocity). The matrix associated to this system is

$$A = \begin{bmatrix} 0 & 1 \\ -1 & -k \end{bmatrix}. \tag{19.20}$$

The eigenvalues of this system are therefore given by solving

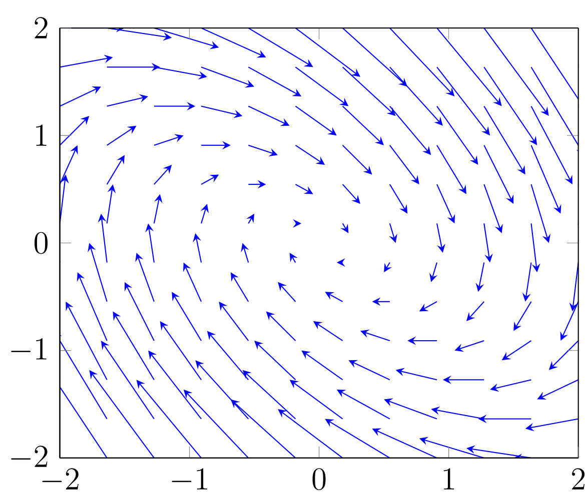
$$0 = \det \begin{bmatrix} -\lambda & 1 \\ -1 & -k-\lambda \end{bmatrix} = \lambda(k+\lambda) + 1 = 1 + k\lambda + \lambda^2, \tag{19.21}$$

which has solutions

$$\lambda = -\frac{k}{2} \pm \frac{1}{2}\sqrt{k^2 - 4}. \tag{19.22}$$

Since $k > 0$, there are three cases to consider.

i. ($k \in (0, 2)$) When $k = 1$, the vector field associated to this system looks like (after rescaling)



In this case, the eigenvalues are complex and distinct. A similarity matrix that diagonalizes A is given by

$$S = \begin{bmatrix} \frac{1}{2}\left(-k+\sqrt{k^2-4}\right) & \frac{1}{2}\left(-k-\sqrt{k^2-4}\right) \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_2 & \lambda_1 \\ 1 & 1 \end{bmatrix} \tag{19.23}$$

with a diagonal matrix given by

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \tag{19.24}$$

where we have used the shorthand notation

$$\lambda_1 := \frac{1}{2}\left(-k - i\sqrt{4-k^2}\right) \qquad \& \qquad \lambda_2 := \frac{1}{2}\left(-k + i\sqrt{4-k^2}\right) \tag{19.25}$$

with $i := \sqrt{-1}$. In this case, S^{-1} is given by

$$S^{-1} = \frac{1}{\lambda_2 - \lambda_1}\begin{bmatrix} 1 & -\lambda_1 \\ -1 & \lambda_2 \end{bmatrix}. \tag{19.26}$$

Indeed, you can check that $A = SDS^{-1}$. Therefore,

$$\begin{aligned} \exp(tA) &= \exp(tSDS^{-1}) = S\exp(tD)S^{-1} \\ &= \frac{1}{\lambda_2-\lambda_1}\begin{bmatrix} \lambda_2 & \lambda_1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^{t\lambda_1} & 0 \\ 0 & e^{t\lambda_2} \end{bmatrix}\begin{bmatrix} 1 & -\lambda_1 \\ -1 & \lambda_2 \end{bmatrix} \\ &= \frac{1}{\lambda_2-\lambda_1}\begin{bmatrix} \lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t} & \lambda_1\lambda_2\left(e^{\lambda_2 t} - e^{\lambda_1 t}\right) \\ e^{\lambda_1 t} - e^{\lambda_2 t} & \lambda_2 e^{\lambda_2 t} - \lambda_1 e^{\lambda_1 t} \end{bmatrix} \end{aligned} \tag{19.27}$$

for any $t \in \mathbb{R}$. It is convenient to let

$$\gamma := \frac{k}{2} \qquad \& \qquad \omega := \frac{\sqrt{4-k^2}}{2} \tag{19.28}$$

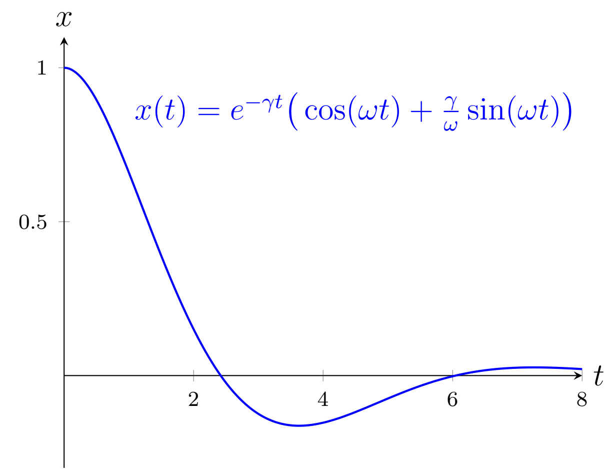
so that

$$\lambda_1 = -\gamma - i\omega \qquad \& \qquad \lambda_2 = -\gamma + i\omega. \tag{19.29}$$

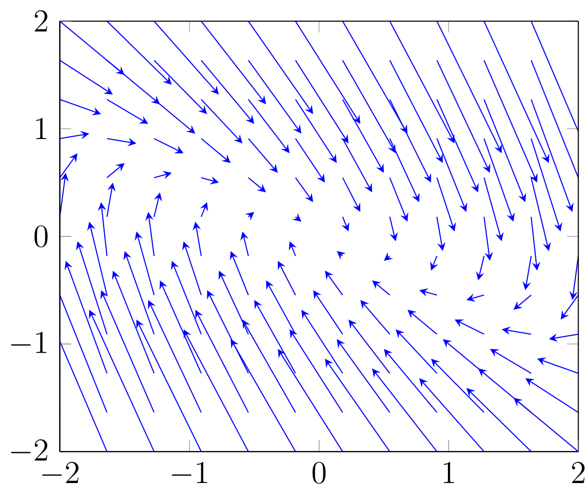
With these substitutions, the solution becomes (after some algebra)

$$\exp(tA) = \begin{bmatrix} e^{-\gamma t}\left(\cos(\omega t) + \frac{\gamma}{\omega}\sin(\omega t)\right) & \frac{\gamma^2+\omega^2}{\omega}e^{-\gamma t}\sin(\omega t) \\ -\frac{1}{\omega}e^{-\gamma t}\sin(\omega t) & e^{-\gamma t}\left(\cos(\omega t) - \frac{\gamma}{\omega}\sin(\omega t)\right) \end{bmatrix}. \tag{19.30}$$

For example, with initial conditions $x_0 = 1$ and $y_0 = 0$, which corresponds to letting the oscillator go without giving it an initial velocity, the position as a function of time is given by ($k = 1$ in this graph)



ii. ($k = 2$) This is the degenerate case where the matrix for the system is not diagonalizable. The vector field associated to this system looks like (after rescaling)



Nevertheless, it can be put into Jordan normal form, and a matrix that accomplishes this is

$$S = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix} \tag{19.31}$$

with a Jordan matrix given by

$$J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}. \tag{19.32}$$

The inverse of S is given by

$$S^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \tag{19.33}$$

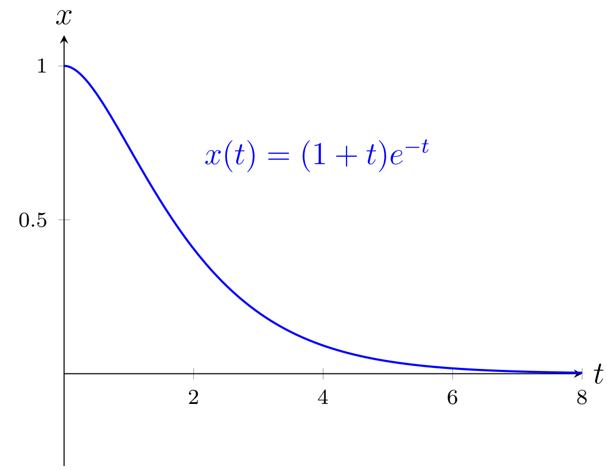
Indeed, you can check that $A = SJS^{-1}$. The exponential of tJ is given by

$$\exp(tJ) = \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \tag{19.34}$$

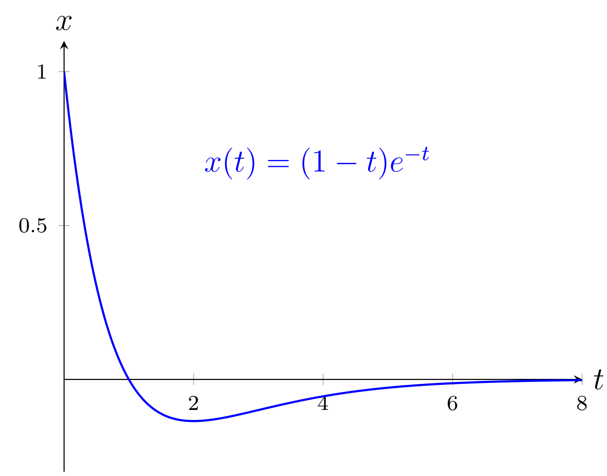
for all $t \in \mathbb{R}$. Therefore,

$$\exp(tA) = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} (1+t)e^{-t} & te^{-t} \\ -te^{-t} & (1-t)e^{-t} \end{bmatrix}. \tag{19.35}$$

Notice that there are no oscillatory terms and any initial configuration with zero velocity asymptotically approaches 0 without passing through 0.



However, if there is an initial sufficiently strong "kick," the trajectory will pass through the origin once (but only once!). Such a trajectory occurs, for instance, when $x_0 = 1$ and $y_0 = -2$ and is depicted here



iii. ($k > 2$) This case is left as Exercise 19.36.

Exercise 19.36. Refer to Example 19.18 in the case $k > 2$.

(a) Diagonalize the matrix A by providing a similarity matrix S and a diagonal matrix D such that $A = SDS^{-1}$.

(b) Solve the system given the initial conditions (x_0, y_0).

(c) Draw the phase diagram for this vector field and draw some integral curves (for example, when $(x_0, y_0) = (1, 0)$ and $(x_0, y_0) = (1, -2)$ for example). Describe the solutions in physical terms.

Consider, now, a non-autonomous differential equation on $\mathbb{R}^m$ of the form

$$\dot{x}(t) = A(t)x(t), \tag{19.37}$$

where $A : [a, b] \to M_m(\mathbb{R})$ is a continuous family of $m \times m$ matrices on a closed interval with $a < b$. The initial condition of this system is $x(a)$. Notice that this differential equation can be expressed as an integral equation of the form

$$x(t) = x(a) + \int_a^t A(s)x(s)\, ds, \tag{19.38}$$

where the integral here is interpreted via components, namely

$$x_i(t) = x_i(a) + \int_a^t \sum_{j=1}^{m} A_{ij}(s)x_j(s)\, ds \tag{19.39}$$

for all $i \in \{1, \ldots, m\}$. This is because, by the fundamental theorem of calculus, the derivative of (19.38) precisely reproduces (19.37). By using this integral equation recursively, we see that it can be expressed as

$$\begin{aligned} x(t) &= x(a) + \int_a^t A(s)x(s)\, ds \\ &= x(a) + \int_a^t A(s)\left(x(a) + \int_a^s A(r)x(r)dr\right) ds \\ &= x(a) + \int_a^t A(s)x(a)\, ds + \int_a^t A(s)\int_a^s A(r)x(r)\, dr\, ds. \end{aligned} \tag{19.40}$$

Note that one does not actually need to define integrals for functions of several variables to understand the right-hand-side. One merely integrates $A(r)x(r)$ with respect to r first to obtain a function that depends only on s. Although it seems like this has made the equation more complicated, let's just do this again, but his time, we will replace our dummy variables to keep track of how many times we do this.

$$\begin{aligned} x(t) &= x(a) + \int_a^t A(s_1)x(a)\, ds_1 + \int_a^t A(s_1)\int_a^{s_1} A(s_2)x(s_2)\, ds_2\, ds_1 \\ &= x(a) + \int_a^t A(s_1)x(a)\, ds_1 + \int_a^t A(s_1)\int_a^{s_1} A(s_2)x(a)\, ds_2\, ds_1 \\ &\quad + \int_a^t A(s_1)\int_a^{s_1} A(s_2)A(s_3)x(s_3)\, ds_3\, ds_2\, ds_1 \end{aligned} \tag{19.41}$$

By linearity properties of the integral, this can be expressed as

$$\begin{aligned} x(t) = &\left(\mathbb{1}_m + \int_a^t A(s_1)\, ds_1 + \int_a^t A(s_1)\int_a^{s_1} A(s_2)\, ds_2\, ds_1\right) x(a) \\ &+ \int_a^t A(s_1)\int_a^{s_1} A(s_2)\int_a^{s_2} A(s_3)x(s_3)\, ds_3\, ds_2\, ds_1. \end{aligned} \tag{19.42}$$

Notice how at every stage, the initial condition $x(a)$ is kept while the unknown solution is pushed "off to infinity." Therefore, one might guess that

$$x(t) = \left(\mathbb{1}_m + \int_a^t A(s_1)\, ds_1 + \int_a^t A(s_1) \int_a^{s_1} A(s_2)\, ds_2\, ds_1 + \cdots \right) x(a) \tag{19.43}$$

is a solution to the differential equation (19.37). This is indeed the case, as we will presently show.

Exercise 19.44. Let $M_m(\mathbb{R})$ be the algebra of all real $m \times m$ matrices. Let $d : M_m(\mathbb{R}) \times M_m(\mathbb{R}) \to \mathbb{R}$ be the function

$$M_m(\mathbb{R}) \times M_m(\mathbb{R}) \ni (A, B) \mapsto d_{\mathrm{op}}(A, B) := \|A - B\|, \tag{19.45}$$

where $\|\cdot\|$ is the operator norm from Exercise 8.7. Prove that $(M_m(\mathbb{R}), d_{\mathrm{op}})$ is a complete metric space.

Theorem 19.46. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $A : [a, b] \to M_m(\mathbb{R})$ be a continuous function. The (extended) sequence of partial sums associated to the (extended) sequence whose value at 0 is $\mathbb{1}_m$ and whose other values are*

$$\mathbb{N} \ni n \mapsto \int_a^t A(s_1)\, ds_1 \int_a^{s_1} A(s_2)\, ds_2 \int_a^{s_2} A(s_3)\, ds_3 \cdots \int_a^{s_{n-1}} A(s_n)\, ds_n \tag{19.47}$$

converges (with respect to the operator metric) for all $t \in [a, b]$, i.e.

$$\sum_{n=0}^{\infty} \int_a^t A(s_1)\, ds_1 \int_a^{s_1} A(s_2)\, ds_2 \int_a^{s_2} A(s_3)\, ds_3 \cdots \int_a^{s_{n-1}} A(s_n)\, ds_n \tag{19.48}$$

exists and is a linear transformation (note that it is understood here that the $n = 0$ term corresponds to the identity matrix and $s_0 := t$).

Proof. It suffices to show that the sequence of partial sums is a Cauchy sequence since $M_m(\mathbb{R})$ with the operator metric is a complete metric space by Exercise 19.44. First note that the integrals above can be expressed as an integral over an n-simplex

$$\begin{aligned} \int_a^t A(s_1)\, ds_1 \int_a^{s_1} A(s_2)\, ds_2 \cdots \int_a^{s_{n-1}} A(s_n)\, ds_n & \\ &= \int_a^t \int_a^{s_1} \cdots \int_a^{s_{n-1}} A(s_1) A(s_2) \cdots A(s_n)\, ds_n \cdots ds_2 ds_1 \\ &= \int\limits_{a \le s_n \le \cdots \le s_2 \le s_1 \le t} A(s_1) A(s_2) \cdots A(s_n)\, ds_n \cdots ds_2 ds_1. \end{aligned} \tag{19.49}$$

Setting

$$L := \max_{\substack{s \in [a,b], \\ i,j \in \{1,\ldots,m\}}} |A_{ij}(s)|, \tag{19.50}$$

the norm of the above integral is therefore bounded in the sense that

$$\begin{aligned}\left\|\int_a^t A(s_1)\,ds_1\int_a^{s_1}A(s_2)\,ds_2\cdots\int_a^{s_{n-1}}A(s_n)\,ds_n\right\|\\ &\le L^n\left\|\int\limits_{a\le s_n\le\cdots\le s_2\le s_1\le t}\mathbb{1}\,ds_n\cdots ds_2ds_1\right\|\\ &=\frac{L^n(b-a)^n}{n!}.\end{aligned}\tag{19.51}$$

Therefore, to show that the nested integrals are Cauchy, fix $\epsilon>0$. Set $N\in\mathbb{N}$ to be an integer large enough so that

$$\sum_{k=m}^{n}\frac{L^k(b-a)^k}{k!}<\epsilon\qquad\forall\,n\ge m\ge N.\tag{19.52}$$

Such a natural number exists because the series for the exponential, $e^{L(b-a)}$, converges (so is Cauchy). Thus, by the triangle inequality and the results above,

$$\begin{aligned}\left\|\sum_{k=m}^{n}\int_a^t A(s_1)\,ds_1\int_a^{s_1}A(s_2)\,ds_2\cdots\int_a^{s_{k-1}}A(s_k)\,ds_k\right\|\\ &\le\sum_{k=m}^{n}\left\|\int_a^t A(s_1)\,ds_1\int_a^{s_1}A(s_2)\,ds_2\cdots\int_a^{s_{k-1}}A(s_k)\,ds_k\right\|\\ &\le\sum_{k=m}^{n}\frac{L^k(b-a)^k}{k!}\\ &<\epsilon\qquad\forall\,n\ge m\ge N.\end{aligned}\tag{19.53}$$

Since the sequence of partial sums of integrals is Cauchy, it converges since $M_m(\mathbb{R})$ is complete by Exercise 19.44. ■

Definition 19.54. Let $a,b\in\mathbb{R}$ with $a<b$ and let $A:[a,b]\to M_m(\mathbb{R})$ be a continuous function. The linear transformation

$$P_a^tA:=\sum_{n=0}^{\infty}\int_a^t A(s_1)\,ds_1\int_a^{s_1}A(s_2)\,ds_2\int_a^{s_2}A(s_3)\,ds_3\cdots\int_a^{s_{n-1}}A(s_n)\,ds_n\tag{19.55}$$

is called the <u>*time-ordered exponential*</u> of A.

The time-ordered exponential has many wonderful properties, reduces to many well-known situations, and can be used to solve non-autonomous ODEs as the rest of this lecture will show.

Theorem 19.56. *Let $a,b\in\mathbb{R}$ with $a<b$ and let $A:[a,b]\to M_m(\mathbb{R})$ be a continuous function. Prove that*

$$P_a^tA=(P_s^tA)(P_a^sA)\tag{19.57}$$

for all $s,t\in(a,b)$ with $s<t$.

Exercise 19.58. Prove Theorem 19.56.

Theorem 19.59. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $A : [a, b] \to M_m(\mathbb{R})$ be a continuous function. Then,*

$$(a, b) \ni t \mapsto P_a^t A \tag{19.60}$$

is continuously differentiable for all $t \in (a, b)$ and its derivative is given by

$$\frac{d}{dt}\left(P_a^t A\right) = A(t) P_a^t A. \tag{19.61}$$

Exercise 19.62. Prove Theorem 19.59.

Theorem 19.63. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $A : [a, b] \to M_m(\mathbb{R})$ be a continuous function. Then*

$$[a, b] \ni t \mapsto x(t) := (P_a^t A) x(a) \tag{19.64}$$

is the unique C^1 solution to the differential equation

$$\dot{x}(t) = A(t) x(t) \tag{19.65}$$

(with initial condition $x(a)$).

Exercise 19.66. Prove Theorem 19.63.

Exercise 19.67. The time-ordered exponential of $A : [a, b] \to M_m(\mathbb{R})$ and its relation to linear differential equations reduces to several well-known special cases.

(a) Show that if $m = 1$, so that A is a family of 1×1 matrices, then

$$P_a^t A = \exp\left(\int_a^t A(s)\, ds\right). \tag{19.68}$$

Using this result, solve the (single variable) ODE

$$\dot{x}(t) = A(t) x(t) \qquad \text{with initial condition } x(a). \tag{19.69}$$

(b) Let A and B be arbitrary matrices that commute, i.e. $[A, B] = 0$ (here, $[A, B] := AB - BA$ is the commutator of A with B). Show that

$$\exp(A + B) = \exp(A)\exp(B) = \exp(B)\exp(A). \tag{19.70}$$

(c) Given an example of two 2×2 matrices A and B such that $\exp(A + B) \neq \exp(A)\exp(B)$.

(d) Let m now be arbitrary but suppose that

$$\big[A(r), A(s)\big] = 0 \tag{19.71}$$

for all $r, s \in [a, b]$. Show that

$$P_a^t A = \exp\left(\int_a^t A(s)\, ds\right). \tag{19.72}$$

In particular, if $A(r) = A(s)$ for all $r, s \in [a, b]$, show that

$$P_a^t A = \exp\Big((t - a)A\Big). \tag{19.73}$$

Exercise 19.74. There is another method of constructing the time-ordered exponential that does not actually use any integrals of several variables in its construction [10]. A *simple function* on a closed interval $[a,b]$ is a function $f:[a,b]\to\mathbb{R}$ for which there exists a partition $a=:a_0<a_1<\cdots<a_{n-1}<a_n:=b$ of $[a,b]$ such that f is constant on (a_i,a_{i+1}) for all $i\in\{0,\dots,n-1\}$. Let $S\big([a,b]\big)$ be the set of all simple functions. Given such a simple function f, define the *integral* of f on $[a,b]$ to be

$$\int_a^b f(s)\,ds := \sum_{i=0}^{n-1}(a_{i+1}-a_i)f\left(\frac{a_i+a_{i+1}}{2}\right). \tag{19.75}$$

The metric $d_{\sup}$ is a well-defined metric on $S\big([a,b]\big)$ though $S\big([a,b]\big)$ is not complete with respect to this metric. Let $R\big([a,b]\big):=\widetilde{S\big([a,b]\big)}$ be the completion of this metric space. One can construct $R\big([a,b]\big)$ in such a way so that it is a subset of all functions $\mathbb{R}^{[a,b]}$. Elements of $R\big([a,b]\big)$ are called *regulated functions* (on $[a,b]$).

In fact, one can similarly replace the codomain $\mathbb{R}$ in the definitions of simple and regulated functions and consider functions $f:[a,b]\to M_m(\mathbb{R})$ to $m\times m$ matrices. In this case, the sup distance uses the operator norm and is denoted by $d_{\mathrm{op}\sup}$. We write $S\big(M_m(\mathbb{R}),[a,b]\big)$ to denote the $M_m(\mathbb{R})$-valued simple functions and $R\big(M_m(\mathbb{R}),[a,b]\big)$ for their completion with respect to $d_{\mathrm{op}\sup}$. Let $A:[a,b]\to M_m(\mathbb{R})$ be a $M_m(\mathbb{R})$-valued simple function with partition $a=:a_0<a_1<\cdots<a_{n-1}<a_n:=b$. Define

$$\mathcal{P}_a^b A := e^{(a_n-a_{n-1})A_{n-1}}e^{(a_{n-1}-a_{n-2})A_{n-2}}\cdots e^{(a_1-a_0)A_0}, \tag{19.76}$$

where

$$A_i := A\left(\frac{a_i+a_{i+1}}{2}\right) \tag{19.77}$$

is the value of A on the i-th partition with $i\in\{0,1,\dots,n-1\}$. Note the order in which these matrices are multiplied (earlier times on the right).

(a) Show that the integral (19.75) defines a continuous function $S\big([a,b]\big)\to\mathbb{R}$ sending f to its integral over $[a,b]$. Note that this implies that the integral can be extended uniquely to all of $R\big(M_m(\mathbb{R}),[a,b]\big)$.

(b) Show that $\Big(S\big([a,b]\big),d_{\sup}\Big)$ is not complete.

(c) Show that every continuous functions on $[a,b]$ is regulated.

(d) Let A and B be $m\times m$ matrices. Prove that

$$\|e^A-e^B\|\le e^{\max\{\|A\|,\|B\|\}}d_{\mathrm{op}}(A,B). \tag{19.78}$$

[Hint: $A^n-B^n=A^{n-1}(A-B)+A^{n-2}(A-B)B+\cdots+A(A-B)B^{n-2}+(A-B)B^{n-1}$.]

(e) Prove that if $A,B:[a,b]\to M_m(\mathbb{R})$ are two $M_m(\mathbb{R})$-valued simple functions, then

$$\big\|\mathcal{P}_a^bA-\mathcal{P}_a^bB\big\|\le(b-a)e^{(b-a)\max\{\|A\|,\|B\|\}}d_{\mathrm{op}\sup}(A,B). \tag{19.79}$$

(f) Prove that the assignment

$$S\big(M_m(\mathbb{R}),[a,b]\big) \ni A \mapsto \mathcal{P}_a^b A \tag{19.80}$$

is continuous with respect to the metric $d_{\mathrm{op\,sup}}$ on the domain and d_{sup} on the codomain. Therefore, it extends to the completion on regulated functions, i.e. to $R\big(M_m(\mathbb{R}),[a,b]\big)$.

(g) Prove that for $A:[a,b]\to M_m(\mathbb{R})$ a continuous function, $\mathcal{P}_a^b A = P_a^b A$.

Remark 19.81. The time-ordered exponential is used in many important situations throughout mathematics and science. For example, in the context of Riemannian geometry, it is used to solve the geodesic equation. The geodesic equation is an ordinary differential equation describing the shortest distance between two points on a manifold as a subset of Euclidean space. Geodesics can then be used to define a metric on a manifold. In the more general context of gauge theory (connections on vector bundles), the time-ordered exponential is used to parallel transport vectors [3].

Problems.

From these notes: Exercises 19.4, 19.8, 19.12, 19.36, 19.44, 19.58, 19.62, 19.66, 19.67, 19.74

20 April 20: Existence and Uniqueness for Ordinary Differential Equations

Motivated by the time-ordered exponential solution of non-autonomous linear differential equations, we study arbitrary non-autonomous differential equations of the form

$$\dot{x}(t) = V(x,t), \tag{20.1}$$

where $V : [a,b] \to C(\mathbb{R}^m, \mathbb{R}^m)$ is a time-dependent continuous vector field on $\mathbb{R}^m$. Continuity of V in this case is defined to mean continuity of the closely related function

$$V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m. \tag{20.2}$$

We use the same notation because these perspectives are completely interchangeable. A special case considered in the previous lecture is when $V(t)$ is a *linear* transformation of $\mathbb{R}^m$ for all time $t \in [a,b]$. In this case, $V : [a,b] \to M_m(\mathbb{R}) \subseteq C(\mathbb{R}^m, \mathbb{R}^m)$ since linear transformations are continuous.

As we've also noticed, this differential equation with the initial condition $x(a)$ can be expressed as the integral equation

$$x(t) = x(a) + \int_a^t V\big(x(s), s\big)\, ds \tag{20.3}$$

provided that V is continuous. In fact, we can use this integral equation to define a convenient function on the set $C\big(\mathbb{R}^m, [a,b]\big)$. Given a continuous time-dependent vector field $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$, let $\Phi_V : C\big(\mathbb{R}^m, [a,b]\big) \to C\big(\mathbb{R}^m, [a,b]\big)$ be the function defined by sending a continuous function $x : [a,b] \to \mathbb{R}^m$ to the function $\Phi_V(x)$ whose value at $t \in [a,b]$ is written as

$$[a,b] \ni t \xmapsto{\Phi_V(x)} \Phi_V(x,t) := x(a) + \int_a^t V\big(x(s), s\big)\, ds. \tag{20.4}$$

A fixed point of Φ_V is therefore a solution to (20.3) which is therefore a solution to our ODE. Notice that

$$\Phi_V^2(x,t) = x(a) + \int_a^t V\left(s, x(a) + \int_a^s V\big(r, x(r)\big)\, dr\right) ds. \tag{20.5}$$

In other words, if we keep on iterating Φ_V, we will obtain something analogous to the situation that occurred in the previous lecture with V a time-dependent family of linear transformations. This leads us to an important question: what conditions on $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$ guarantee the existence and uniqueness of a fixed point of Φ_V? One method that we can employ is to find a metric on $C\big(\mathbb{R}^m, [a,b]\big)$ so that Φ_V, or some iterate of Φ_V, is a contraction on this metric space.

Theorem 20.6. *Let $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$ be a continuous family of continuous vector fields on $\mathbb{R}^m$ for which $V(t, \,\cdot\,)$ is Lipshitz for all $t \in [a,b]$, i.e. for each $t \in [a,b]$ there exists a number $L_t \geq 0$ such that*

$$\big|V(t,x) - V(t,y)\big| \leq L_t |x - y| \qquad \forall\, x, y \in \mathbb{R}. \tag{20.7}$$

Then there exists a constant $L \geq 0$ such that

$$\big|V(t,x) - V(t,y)\big| \leq L |x - y| \qquad \forall\, x, y \in \mathbb{R}^m,\ t \in [a,b]. \tag{20.8}$$

Exercise 20.9.

(a) Prove Theorem 20.6.

(b) Is it true that if a continuous function $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$ satisfies condition (20.8) then it is Lipschitz on $[a,b]\times\mathbb{R}^m$? Prove your claim. Recall, in this case, a function $V : [a,b]\times\mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz iff there exists an $M \geq 0$ such that

$$\big|V(s,x) - V(t,y)\big| \leq M\sqrt{|s-t|^2 + |x-y|^2} \qquad \forall\, x,y \in \mathbb{R}^m,\ s,t \in [a,b]. \tag{20.10}$$

The metric here used is the standard one on $[a,b] \times \mathbb{R}^m$ obtained from a subset of $\mathbb{R}^{1+m}$.

Definition 20.11. Let $S \subseteq \mathbb{R}$ be an arbitrary subset of $\mathbb{R}$. A continuous family of vector fields $V : S \times \mathbb{R}^m \to \mathbb{R}^m$ satisfying condition (20.7) will be called *fiberwise Lipschitz*. A continuous family of vector fields $V : S \times \mathbb{R}^m \to \mathbb{R}^m$ satisfying condition (20.8) for some $L \geq 0$ will be called *fiberwise globally Lipschitz*.

Hence, part (a) of Exercise 20.9 asks you to prove that if S is compact, these two notions are equivalent (since the converse is immediate).

Theorem 20.12. *Let $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$ be a fiberwise globally Lipschitz vector field on $\mathbb{R}^m$ and let*

$$\dot{x}(t) = V(x,t) \qquad x_a := x(a) \tag{20.13}$$

be the associated initial value problem (IVP). Then there exists a unique C^1 integral curve $x : [a,b] \to \mathbb{R}^m$ for this IVP.

Proof. Let $L \geq 0$ be the Lipschitz constant for V. Set Φ_V to be as in (20.4). Then, for any two functions $x, y \in C\big(\mathbb{R}^m, [a,b]\big)$.

$$\begin{aligned}
\Big|\Phi_V(x,t) - \Phi_V(y,t)\Big| &= \left|\int_a^t \Big(V\big(s,x(s)\big) - V\big(s,y(s)\big)\Big)ds\right| \\
&\leq \int_a^t \Big|V\big(s,x(s)\big) - V\big(s,y(s)\big)\Big| ds \\
&\leq L\int_a^t \big|x(s) - y(s)\big| ds \\
&\leq L\int_a^t \max_{s\in[a,b]}\Big\{\big|x(s)-y(s)\big|\Big\} ds \\
&\leq L d_{\sup}(x,y)(t-a).
\end{aligned} \tag{20.14}$$

Similarly,

$$\begin{aligned}
\Big|\Phi_V^2(x,t) - \Phi_V^2(y,t)\Big| &\leq L\int_a^t \Big|\Phi_V(x,s) - \Phi_V(y,s)\Big| \\
&\leq L^2 d_{\sup}(x,y)\int_a^t (s-a)ds \\
&\leq \frac{L^2(t-a)^2}{2} d_{\sup}(x,y).
\end{aligned} \tag{20.15}$$

In fact, more generally,

$$\Big|\Phi_V^n(x,t) - \Phi_V^n(y,t)\Big| \le \frac{L^n(t-a)^n}{n!} d_{\mathrm{sup}}(x,y) \tag{20.16}$$

for arbitrary $n \in \mathbb{N}$. Since

$$\lim_{n\to\infty} \frac{L^n(t-a)^n}{n!} = 0, \tag{20.17}$$

for $\epsilon := \frac{1}{2}$, there exists an $N \in \mathbb{N}$ such that

$$\frac{L^n(t-a)^n}{n!} < \epsilon \qquad \forall\, n \ge N. \tag{20.18}$$

Hence, Φ_V^N is a contraction on $\Big(C\big(\mathbb{R}^m,[a,b]\big), d_{\mathrm{sup}}\Big)$ so that there exists a unique fixed point $x \in C\big(\mathbb{R}^m,[a,b]\big)$ for Φ_V by the Iterative Mapping Contraction Theorem. The fact that x is continuously differentiable follows from the fact that it satisfies $x = \Phi_V(x)$ since $\dot{x}(t) = V\big(t, x(t)\big)$ for all $t \in [a,b]$ and by assumption V is continuous. ■

Notice that the proof of this theorem actually says more.

Corollary 20.19. *Let $V : [a,b] \times \mathbb{R}^m \to \mathbb{R}^m$ be a fiberwise globally Lipschitz vector field and let $\Phi_V : C\big(\mathbb{R}^m,[a,b]\big) \to C\big(\mathbb{R}^m,[a,b]\big)$ be the function in (20.4). Then for any $x \in C\big(\mathbb{R}^m,[a,b]\big)$, the sequence,*

$$\Big(x, \Phi_V(x), \Phi_V^2(x), \Phi_V^3(x), \ldots\Big) \tag{20.20}$$

converges to the unique fixed point of the ODE $\dot{x}(t) = V\big(t, x(t)\big)$.

Exercise 20.21. Solve the ODE

$$\frac{dx}{dt} = (x+t)t \qquad \forall\, t \in [0,1], \qquad x(0) := 0. \tag{20.22}$$

Exercise 20.23. Let $V : [1,3] \times \mathbb{R} \to \mathbb{R}$ be the vector field given by

$$[1,3] \times \mathbb{R} \ni (t,x) \mapsto (x+t^2)e^{t-1}. \tag{20.24}$$

(a) Prove that $L := e^2$ is a Lipschitz constant for V so that with this L, V is a fiberwise globally Lipschitz vector field.

(b) Find the smallest integer $N \in \mathbb{N}$ such that Φ_V^N is a contraction.

(c) If you have access to a computing device, try to numerically solve this differential equation to a reasonable degree of accuracy and compare your integral curves with a plot of the vector field (since this is an ODE of a single variable, you may plot $\dot{t} = 1$ as one of the components of the vector field).

An exercise that actually belonged in the notes from three lectures ago is the following.

Exercise 20.25. Let X be a set and let d and d' be two metrics on X for which there exist numbers $L, M \geq 0$ satisfying

$$Ld'(x,y) \leq d(x,y) \leq Md'(x,y) \qquad \forall\, x, y \in X. \tag{20.26}$$

When this occurs, d is said to be *squeezed* by d'. Let $x : \mathbb{N} \to X$ be a sequence in X.

(a) Show that x is Cauchy with respect to d if and only if x is Cauchy with respect to d'.

(b) Show that x converges to a point $\lim x \in X$ with respect to d if and only if x converges to $\lim x$ with respect to d'.

(c) Show that X is complete with respect to d if and only if X is complete with respect to d'.

(d) Let d_2 be the standard metric on $\mathbb{R}^m$ and let d_∞ be the max metric on $\mathbb{R}^m$ (recall Example 17.13). Show that

$$d_\infty(x,y) \leq d_2(x,y) \leq \sqrt{m}\, d_\infty(x,y). \tag{20.27}$$

(e) Show that the relation of being squeezed is an equivalence relation on the set of metrics on a set X.

These results have many important and convenient consequences. For one, if you have a function $f : X \to X$ on a set, it might be that f is a contraction with respect to d but not with respect to d'. Therefore, you will be able to prove the existence and uniqueness of a fixed point with respect to d but since the function f was just defined on the set, which does not care about the metric structure, you would have found our unique fixed point. In some sense, choosing the appropriate metric is part of the battle in proving the existence and uniqueness of fixed points. For instance, the following exercises use this idea to prove the existence and uniqueness of a stationary probability measure under a regular stochastic process, a very important theorem in probability theory.

Definition 20.28. Let X be a finite set. A *probability measure* on X is a function $\mu : X \to \mathbb{R}$ satisfying

$$\mu(x) \geq 0 \qquad \forall\, x \in X \qquad \& \qquad \sum_{x \in X} \mu(x) = 1. \tag{20.29}$$

Often, the notation $\mu_x := \mu(x)$ is used. Let $\Pr(X)$ denote the set of probability measures on X.

Definition 20.30. A *stochastic map/process* from a set X to a set Y is a function

$$\begin{aligned} X &\xrightarrow{f} \Pr(Y) \\ x &\mapsto f(x) \end{aligned} \tag{20.31}$$

whose evaluation on elements in Y is written as

$$Y \ni y \xmapsto{f(x)} f_{yx} \in \mathbb{R}_{\geq 0}. \tag{20.32}$$

A stochastic process is a generalization of a function. Instead of associating a unique element in the codomain to an element in the domain, a stochastic process associates a *probability distribution* of possible outcomes. Stochastic processes, just like functions, can be composed. To distinguish functions from stochastic processes, the latter will be denoted with squiggly arrows $f : X \rightsquigarrow Y$.

Definition 20.33. Let X, Y, and Z be finite sets. The *composition* of the stochastic map $f : X \rightsquigarrow Y$ followed by the stochastic map $g : Y \rightsquigarrow Z$, written as $g \circ f : X \rightsquigarrow Z$, is the stochastic map $g \circ f : X \to \Pr(Z)$ defined by sending x to the probability measure defined by

$$Z \ni z \xmapsto{(g \circ f)(x)} (g \circ f)_{zx} := \sum_{y \in Y} g_{zy} f_{yx}. \tag{20.34}$$

It follows immediately from these definitions that a stochastic process $\{\bullet\} \rightsquigarrow X$ from a single element set to X is itself a probability measure on X and all probability measures are of this form. Henceforth, it is convenient to view every probability measure on X as a stochastic process $\mu : \{\bullet\} \rightsquigarrow X$.

Definition 20.35. Let $f : X \rightsquigarrow X$ be a stochastic process from a finite set X to itself. A probability measure $\mu : \{\bullet\} \rightsquigarrow X$ on X is a *stationary measure* with respect to f iff $f \circ \mu = \mu$.

Notice that $f \circ \mu : \{\bullet\} \rightsquigarrow X$ is another stochastic process from $\{\bullet\}$ to X and is therefore a probability measure on X (regardless if it is stationary). The next exercise will give one proof that stationary measures for stochastic processes always exist.

Exercise 20.36. Let X be a finite set and let $d_{\text{tv}} : \Pr(X) \times \Pr(X) \to \mathbb{R}$ (the subscripts stand for "total variation") be the function defined by

$$\Pr(X) \times \Pr(X) \ni (\mu, \nu) \mapsto d_{\text{tv}}(\mu, \nu) := \frac{1}{2} \sum_{x \in X} \big|\mu_x - \nu_x\big|. \tag{20.37}$$

(a) Prove that d_{tv} is a metric on $\Pr(X)$.

(b) Prove that $\Pr(X)$ is compact.

(c) Let $\mu \in \Pr(X)$. Show that the sequence

$$\mathbb{N} \ni n \mapsto \frac{1}{n} \sum_{k=1}^{n} f^k \circ \mu \tag{20.38}$$

is a sequence $\overline{\mu} : \mathbb{N} \to \Pr(X)$ of probability measures on X. Note that $f^k \circ \mu := f \circ \cdots \circ f \circ \mu$ with k-many applications of f and the sum of two measures is defined pointwise.

(d) By compactness of $\Pr(X)$, let ν be a convergent subsequence of $\overline{\mu}$. Prove that $f \circ (\lim \nu) = \lim \nu$, i.e. ν is a stationary measure for f.

The previous exercise guarantees existence, but what guarantees uniqueness of stationary measures?

Exercise 20.39. Let X be a finite set and let $f : X \rightsquigarrow X$ be a stochastic process such that $f_{x'x} > 0$ for all $x, x' \in X$. Such a stochastic process is said to be *strictly positive*.

(a) Prove that f is a contraction on $\big(\mathrm{Pr}(X), d_{\mathrm{tv}}\big)$.

(b) Prove that if f is a strictly positive stochastic process then there exists a unique stationary measure on X with respect to f.

(c) A stochastic process $f : X \rightsquigarrow X$ is said to be *regular* iff f^N is strictly positive for some $N \in \mathbb{N}$. Prove that if f is regular, then there exists a unique stationary measure on X with respect to f.

The version of the statement for the existence and uniqueness of solutions to ODE's in [7] is slightly stronger than the one presented here. Nevertheless, Kolmogorov and Fomin's version is important to the study of vector fields and differential equations on manifolds. In fact, even the Lipschitz condition is slightly different from ours.

Exercise 20.40. Let $V : [a, b] \times \mathbb{R}^m \to \mathbb{R}^m$ be a continuous time-dependent vector field. Show that V is fiberwise globally Lipschitz if and only if for each $i \in \{1, \ldots, m\}$, there exists an $L_i \ge 0$ such that

$$\Big|V_i(t, x) - V_i(t, y)\Big| \le L_i \max_{j \in \{1,\ldots,m\}} |x_j - y_j| \qquad \forall\, x, y \in \mathbb{R}^m,\ t \in [a, b]. \tag{20.41}$$

Theorem 20.42. *See Theorem 2' in Section 8 of [7].*

Proof. See [7]. ■

The following two exercises are from [2].

Exercise 20.43. Let $C \subseteq \mathbb{R}^n$ be a convex compact subset of Euclidean space and let $f : C \to \mathbb{R}^n$ be a continuously differentiable function. Show that f is Lipschitz on C with Lipschitz constant

$$L := \sup_{x \in C} \|D_x f\|. \tag{20.44}$$

Exercise 20.45. Prove the inverse function theorem using the Contraction Mapping Theorem.

Problems.

From these notes: Exercises 20.9, 20.21, 20.23, 20.25, 20.36, 20.39, 20.40, 20.43, 20.45

21 April 25: Survey of results and applications

Rather than proving a lot of theorems, and since this is the last lecture, we will instead provide a survey of results that follow from the things we have learned this semester. In addition, references will be provided for those who are interested in pursuing certain topics further.

In the previous two lectures, we described how locally there exist solutions to ordinary differential equations. What we did not describe is how these integral curves depend on the initial conditions.

Theorem 21.1. *Let $U \subseteq \mathbb{R}^m$ be an open subset and let $V : \mathbb{R} \times U \to \mathbb{R}^m$ be a time-dependent continuously differentiable vector field. For every $x \in U$, there exists an $\epsilon > 0$ and an open neighborhood W of $x \in U$ and a continuously differentiable function $\varphi : (-\epsilon, \epsilon) \times W \to U$ such that $\varphi(\,\cdot\,, y)$ is an integrable curve of V through y for all $y \in W$. Furthermore, φ is unique in the following sense. If $x' \in U$, and $\varphi' : (-\epsilon', \epsilon') \times W' \to U$ is another C^1 function with $\epsilon' > 0$ and W' an open neighborhood of x', then φ' and φ agree on $(-\delta, \delta) \times (W \cap W')$, where $\delta := \min\{\epsilon, \epsilon'\}$.*

Proof. See either Section 32.6 in [2], Theorem 4 in Lecture 2 of [10] for an infinite-dimensional case, and/or Section 2.5 in [6] for a different, yet more direct, perspective. ■

One can use this theorem to prove the following result for vector fields on manifolds.

Theorem 21.2. *Let $M \subseteq \mathbb{R}^k$ be a C^r manifold with $r \geq 2$, and let $V : M \to \mathbb{R}^k$ be a C^r vector field on M such that there exists a compact $K \subseteq M$ for which $V = 0$ on $M \setminus \{K\}$ (i.e. V is supported on K). Then there exists a C^r flow $\varphi : \mathbb{R} \times M \to M$ such that*

$$\frac{d}{dt}\varphi(t, x) = V_{\varphi(t,x)} \qquad \forall\, t \in \mathbb{R},\ x \in M. \tag{21.3}$$

In particular, every C^r vector field on a compact manifold is integrable.

Proof. See Section 35 of [2]. The idea is to use local charts. ■

Remark 21.4. In some sense, the set of smooth vector fields on a compact manifold is the Lie algebra of the smooth group of diffeomorphisms of the manifold that are smoothly homotopic to the identity. It is very often an infinite-dimensional Lie algebra. There is a notion of an exponential map in this context as well and the exponential of a vector field is basically the flow associated to that vector field. One example that has been studied in depth is S^1 and the set of orientation-preserving and orientation-reversing diffeomorphisms. The Lie algebra of orientation vector fields is closely related to something called the Virasoro algebra, which is a Lie algebra that describes certain aspects of conformal field theory and string theory [11].

An immediate corollary of these theorems together with the results on indices of vector fields is the existence of fixed points for certain diffeomorphisms.

Theorem 21.5. *Let M be a smooth compact manifold with Euler characteristic $\chi(M) \neq 0$. Then every diffeomorphism $f : M \to M$ of M that is smoothly homotopic to the identity has at least one fixed point.*

Exercise 21.6. Prove Theorem 21.5 using the results from above and what we've learned in this course clearly referencing the theorems used.

The following theorem says that every vector field can be straightened out near a point where the vector field does not vanish.

Theorem 21.7. *Let $A \subseteq \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$ and let $V : A \to \mathbb{R}^n$ be a C^k vector field on A, with $k \geq 1$ and let $x_0 \in A$ be a point in A such that $V_{x_0} \neq 0$. Then there exists an open neighborhood $U \subseteq A$ of x, an open set $B \subseteq \mathbb{R}^n$, and a C^k diffeomorphism $f : U \to B$ such that the pushforward vector field $f_*(V)$ is constant. In other words, there exists an open neighborhood of x such that V corresponds to a constant vector field.*

Proof. See Theorem 1 in Section 3 of [10]. ■

There is a similar result near a critical point of a vector field. Under some additional assumptions, every vector field near an isolated critical point can be linearized. These conditions are not obvious and a thorough account of where they come from can be found in Section 3 of [10]. We merely state the result.

Definition 21.8. Let $A \subseteq \mathbb{R}^n$ be an open subset and let $V : A \to \mathbb{R}^n$ be a C^k vector field with $k \geq 1$ with an isolated critical point at $a \in A$ and let $\{\lambda_1, \ldots, \lambda_n\}$ be the eigenvalues of $D_aV : \mathbb{R}^n \to \mathbb{R}^n$. V is said to satisfy the *k-th order eigenvalue condition at a* iff for every $i \in \{1, \ldots, n\}$, there do *not* exist positive real numbers $\mu_1, \ldots, \mu_n\}$ satisfying

$$2 \leq \sum_{j=1}^{n} \mu_j \leq k \tag{21.9}$$

and

$$\lambda_i = \sum_{j=1}^{n} \mu_j \lambda_j. \tag{21.10}$$

Theorem 21.11. *Let $A \subseteq \mathbb{R}^n$ be an open subset and let $V : A \to \mathbb{R}^n$ be a C^k vector field with $k \geq 1$ with an isolated critical point at $a \in A$ and suppose that V satisfies the k-th order eigenvalue condition at a. Then there exists an open neighborhood $U \subseteq A$ of x, an open set $B \subseteq \mathbb{R}^n$, and a C^k diffeomorphism $f : U \to B$ such that the pushforward vector field $f_*(V)$ is linear.*

Proof. See Theorem 9 in Section 3 of [10]. ■

Vector fields can correspond to other vector fields under a variety of transformations. The previous results indicated that vector fields always correspond to simple types of vector fields under C^k diffeomorphisms. Such diffeomorphisms forget the actual eigenvalues of the initial vector field. Depending on the types of diffeomorphism, however, information about the eigenvalues can be preserved. A diffeomorphism can be linear or can be C^∞. It also makes sense to ask for a homeomorphism to preserve vector fields, but in this case, it is more appropriate to phrase the situation in terms of flows.

Definition 21.12. Two flows $\varphi, \psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ are equivalent iff there exists a bijection $h : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\begin{tikzcd} \mathbb{R} \times \mathbb{R}^n \arrow[d, "\phi"'] & \mathbb{R} \times \mathbb{R}^n \arrow[l, "\mathrm{id}\times h"'] \arrow[d, "\psi"] \\ \mathbb{R}^n & \mathbb{R}^n \arrow[l, "h"] \end{tikzcd} \tag{21.13}$$

Two flows are

(a) *linearly equivalent* iff h is a linear isomorphism,

(b) *differentiably equivalent* iff h is a C^∞ diffeomorphism,

(c) *topologically equivalent* iff h is a C^∞ homeomorphism.

Exercise 21.14. Show that the above notion of equivalence is an equivalence relation on the set of flows on $\mathbb{R}^n$. Prove that linear equivalence implies differentiable equivalence implies topological equivalence.

Theorem 21.15. *Let $A, B : \mathbb{R}^n \to \mathbb{R}^n$ be linear transformations all of whose eigenvalues have multiplicity* 1. *Then the flows associated to the linear systems*

$$\dot{x} = Ax \qquad \& \qquad \dot{x} = Bx \tag{21.16}$$

are linearly equivalent if and only if the eigenvalues of A and B coincide.

Proof. See Section 21.3 in [2]. ■

One might think that there are fewer differentiable equivalence classes of linear systems, but this turns out to be false.

Theorem 21.17. *Let $A, B : \mathbb{R}^n \to \mathbb{R}^n$ be linear transformations all of whose eigenvalues have multiplicity* 1. *Then the flows associated to the linear systems*

$$\dot{x} = Ax \qquad \& \qquad \dot{x} = Bx \tag{21.18}$$

are differentiably equivalent if and only if they are linearly equivalent.

Proof. See Section 21.4 in [2]. ■

Definition 21.19. Let $m_-, m_+ : M_n(\mathbb{R}) \to \{0\} \cup \mathbb{N}$ be the functions that assign to every real $n \times n$ matrix A the number of eigenvalues whose real parts are negative $m_-(A)$ and whose real parts are positive $m_+(A)$.

Theorem 21.20. *Let $A, B : \mathbb{R}^n \to \mathbb{R}^n$ be linear transformations. Then the flows associated to the linear systems*

$$\dot{x} = Ax \qquad \& \qquad \dot{x} = Bx \tag{21.21}$$

are topologically equivalent if and only if

$$m_-(A) = m_-(B) \qquad \& \qquad m_+(A) = m_+(B). \tag{21.22}$$

Proof. See Section 22 in [2]. ■

22 April 27: Review

Today is a review session. Please have questions ready.

Do not forget that there is a final on May 5th covering everything from the semester but focusing more so on Lectures 15 through 21!

References

[1] Stephen Abbott, *Understanding analysis*, 2nd ed., Undergraduate Texts in Mathematics, Springer, 2015.

[2] Vladimir I. Arnold, *Ordinary differential equations*, 1st ed., translated by R. Cooke, Universitext, Springer-Verlag Berlin Heidelberg, 1992.

[3] John Baez and Javier P. Muniain, *Gauge fields, knots and gravity*, Series on Knots and Everything, vol. 4, World Scientific, 1994.

[4] John C. Baez, *The octonions*, Bulletin of the American Mathematical Society **39** (2002), no. 2, 145–205, available at 0105155.

[5] Victor Bryant, *Metric spaces: iteration and application*, Cambridge University Press, 1985.

[6] Witold Hurewicz, *Lectures on ordinary differential equations*, The M.I.T. Press, 1958.

[7] A. N. Kolmogorov and S. V. and Fomin, *Introductory real analysis*, Dover Publications, Inc., New York, 1975. Translated from the second Russian edition and edited by Richard A. Silverman, Corrected reprinting.

[8] John W. Milnor, *Topology from the differentiable viewpoint*, Revised, Princeton Landmarks in Mathematics and Physics, Princeton University Press, 1997.

[9] James R. Munkres, *Analysis on manifolds*, Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1991. MR1079066

[10] Edward Nelson, *Topics in dynamics. I: Flows*, Mathematical Notes, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1969. MR0282379

[11] Andrew Pressley and Graeme Segal, *Loop groups*, Revisde, Oxford Mathematical Monographs, Clarendon Press, 1988.

[12] R. L. E. Schwarzenberger, *Elementary differential equations*, Chapman and Hall Mathematics Series, Spottiswoode, Ballantyne and Co. Ltd., 1969.

[13] Michael Spivak, *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*, W. A. Benjamin, Inc., New York-Amsterdam, 1965.