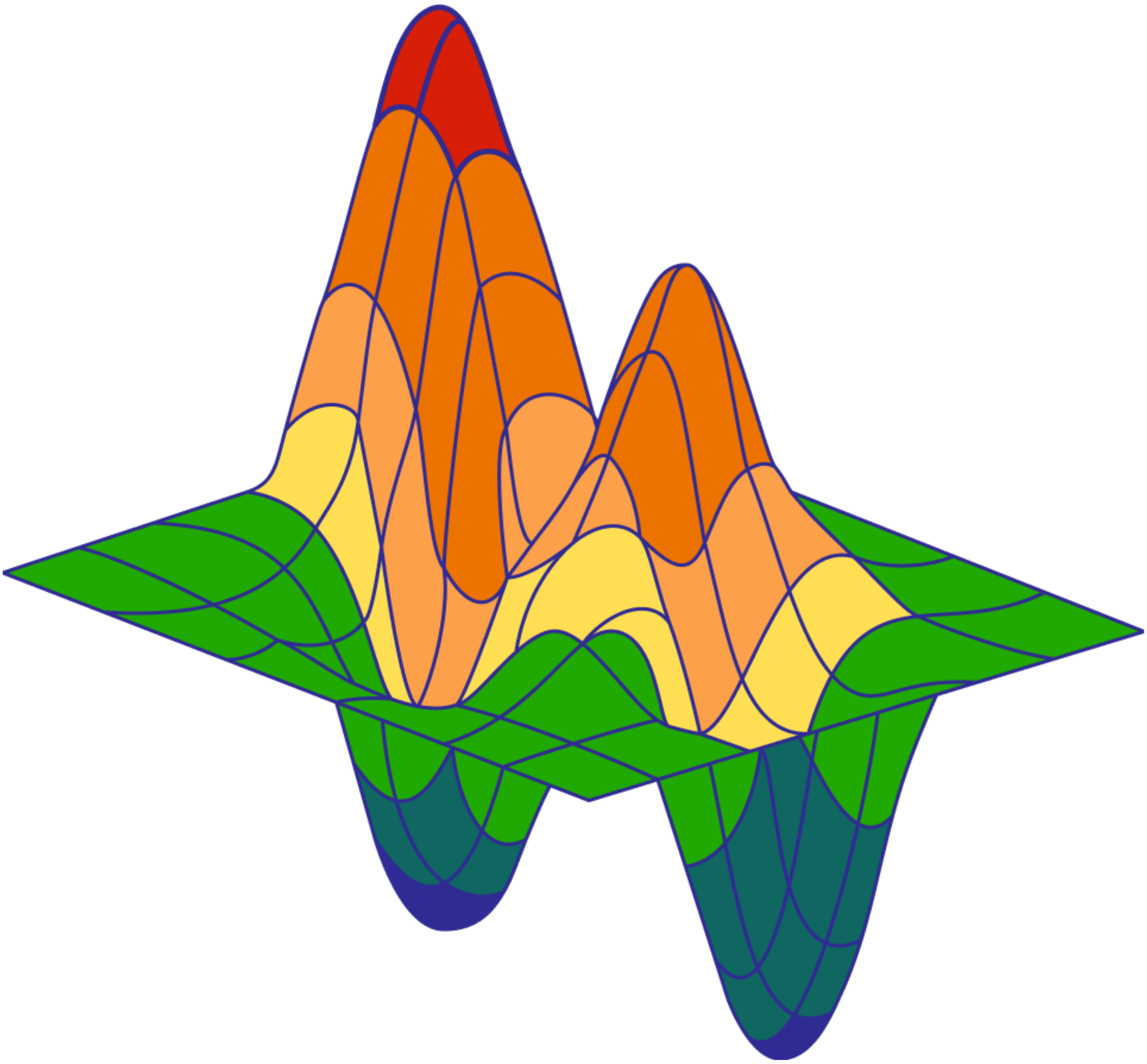
Multivariable Calculus with Proofs

MAT237 Course Notes 2024-25

Asif Zaman



https://www.math.utoronto.ca/zaman

Compiled on 2024-09-25 at 20:58:00

Contents

Appendix

Acknowledgements

I am grateful for the involvement of Sam De Abreu, Victoria Allder, Sirui Chen, Kevin Didi, Edric Liu, Raymond Liu, Lucas Prates, Katrina Sha, Grace Shang, Ryan Shi, Sarah Verreault, Amy Wang, Sarah Xie, and William Young as part of the summer Work Study programs in 2021, 2022, 2023, and 2024. They diligently contributed to the index, examples, figures, exercises, hints, and proofreading. I am also indebted to all the former MAT237 students who kindly pointed out many typographical errors in earlier drafts. I also thank Belal Abuelnasr, Semeon Artamonov, Thad Janisse, Caleb Jonker, Maryam Khaqan, and Fabian Parsch for their contributions and suggestions as instructors for this course.

List of notation

Sets

$\mathbb{N}$ is the set of natural numbers $\{0, 1, 2, 3, \dots\}$.
$\mathbb{N}^+$ is the set of positive natural numbers $\{1, 2, 3, \dots, \}$.
$\mathbb{Z}$ is the set of integers.
$\mathbb{Q}$ is the set of rational numbers.
$\mathbb{R}$ is the set of real numbers.
$\mathbb{C}$ is the set of complex numbers.
$A \cup B = \{x : x \in A \text{ or } x \in B\}$ is the union of sets A and B.
$A \cap B = \{x : x \in A \text{ and } x \in B\}$ is the intersection of sets A and B.
$A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ is the subtraction of the set B from the set A.
$A \times B = \{(a, b) : a \in A, b \in B\}$ is the Cartesian product of sets A and B.
$A^n = A \times \cdots \times A$ is the n-fold Cartesian product of the set A.

Topology and geometry

$A^c = \mathbb{R}^n \setminus A$ is the complement of a set $A \subseteq \mathbb{R}^n$.
A^o or $\operatorname{int}(A)$ is the interior of a set $A \subseteq \mathbb{R}^n$.
A^* is the set of limit points of a set $A \subseteq \mathbb{R}^n$.
$\overline{A}$ or $\operatorname{cl}(A)$ is the closure of a set $A \subseteq \mathbb{R}^n$.
∂A is the (topological) boundary of a set $A \subseteq \mathbb{R}^n$.
$B_r(x)$ is the open ball of radius $r > 0$ centred at $x \in \mathbb{R}^n$.
S^{n-1} is the unit sphere in $\mathbb{R}^n$ centred at the origin, where $n \in \mathbb{N}^+$.
$\|x\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ is the norm of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.
$x \cdot y = x_1 y_1 + \cdots x_n y_n$ is the dot product of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$.
$a \times b$ is the three-dimensional cross product for $a, b \in \mathbb{R}^3$.

Differential calculus

∂_i is the partial derivative in the ith component in $\mathbb{R}^n$.
D_v is the directional derivative with respect to $v \in \mathbb{R}^n$.
∇ is the gradient differential operator.
$T_p S$ is the tangent space of a set S at the point p.

Integral calculus

'length' is the Jordan measure in $\mathbb{R}$.
'area' is the Jordan measure in $\mathbb{R}^2$.
dA is the area element in $\mathbb{R}^2$.
'vol' is the Jordan measure in $\mathbb{R}^n$, where $n \in \mathbb{N}^+$.
dV is the volume element in $\mathbb{R}^n$, where $n \in \mathbb{N}^+$.

Vector calculus

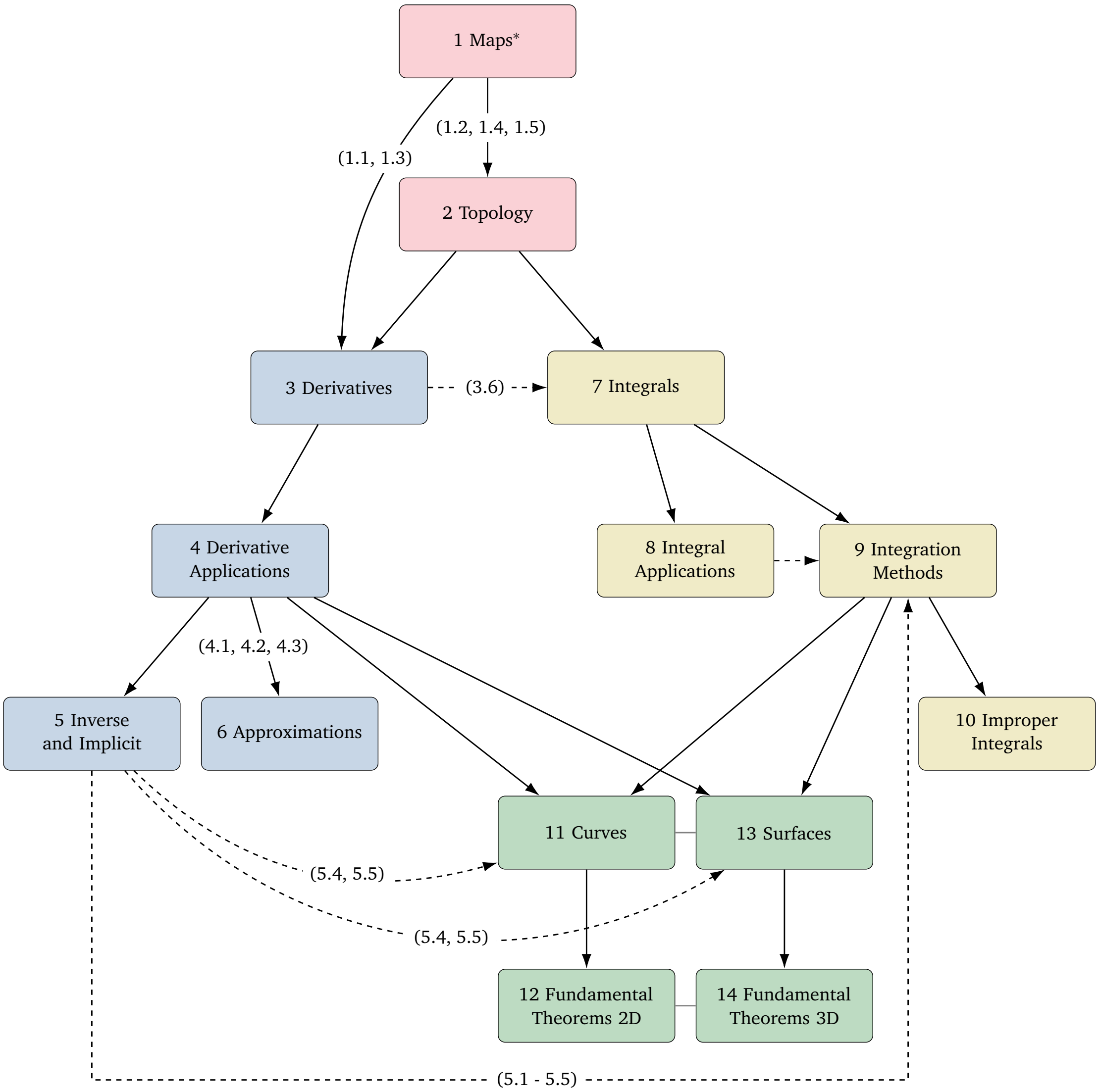
'grad' or ∇ is the gradient differential operator.
'div' or $\nabla\cdot$ is the divergence vector field operator.
'curl' is the curl vector field operator in $\mathbb{R}^2$ or $\mathbb{R}^3$, depending on context.
$\nabla\times$ is the curl vector field operator in $\mathbb{R}^3$.
T is the unit tangent vector of an oriented curve in $\mathbb{R}^n$.
N is the unit normal vector of an oriented curve in $\mathbb{R}^n$.
n is the unit normal of an oriented curve in $\mathbb{R}^2$, or an oriented surface in $\mathbb{R}^3$.

Student preface

(to be added)

Instructor preface

(to be added)



Foundations

0. Notation

When travelling the world, you will notice that social customs vary on different scales. Some practices are global; others are regional; and many are unique to an individual. If you are not familiar with these customs before arriving to a new place, you may find yourself confused or even offended.

The same patterns occur when studying conventions and notation in mathematics. Some are universal; others are subject-specific; many are unique to a particular author. Before you embark on your journey, it is worthwhile to be familiar with this textbook's conventions and notation to avoid confusion. We have tried to choose universal (or at least very common) conventions, but it is not always possible due to the vastness of multivariable calculus literature. When this happens, we often indicate common alternatives, but we will continue to adhere to our stated choices.

This chapter will concisely review sets, Euclidean space, and functions. It may not be thrilling, but it is fundamental to your learning. Conventions and notation are not simply trivialities in mathematics. Good practices will help you write unambiguously, and help you read smoothly. These characteristics form the bedrock for mathematical communication and comprehension.

0.1. Sets

Sets form the fundamental building blocks of mathematics. A fully rigorous introduction requires a careful construction of axiomatic logic. That is far more than necessary for this text, and is usually reserved for a senior undergraduate course on set theory and logic. Instead, the following informal description will suffice.

> *A **set** A is a collection of **elements**. Write $x \in A$ if the **element** x is a **member of** A. The symbol "$\in$" is the set membership symbol.*

The words "collection" or "family" are often used in place of "set" for linguistic diversity. Below is a list of canonical[1] sets that have a fixed notation throughout this text.

- $\mathbb{N}$ is the set of natural numbers $\{0, 1, 2, 3, \dots\}$.
- $\mathbb{N}^+$ is the set of positive natural numbers $\{1, 2, 3, \dots\}$.
- $\mathbb{Z}$ is the set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$.
- $\mathbb{Q}$ is the set of rational numbers $\{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{Z}, n \neq 0\}$.
- $\mathbb{R}$ is the set of real numbers.

A rigorous definition of these canonical sets takes substantial effort[2], but you can simply take them for granted. For a more detailed introduction to logic and sets, see [8, Chapters 1 and 2] or watch this MAT137 YouTube video playlist [9]. These are not fully rigorous either, but provide more details and examples.

0.1.1 Logic

Valid logical statements in this text will either be "true" or "false". In other words, they will have a binary truth value. There are several ways to combine or compare logical statements. Below is a truth table for standard logical operations with valid logical statements P and Q.

P	Q	$\neg P$	$P \land Q$	$P \lor Q$	$P \implies Q$	$P \iff Q$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Each logical operation has a colloquial name and pronounciation.

- $\neg P$ is **negation**, pronounced "not P".
- $P \land Q$ is **logical and**, pronounced "P and Q".
- $P \lor Q$ is **logical or**, pronounced "P or Q".
- $P \implies Q$ is **implication**, pronounced "P implies Q", "if P then Q", "Q if P", "P only if Q".
- $P \iff Q$ is **equivalence**, pronounced "P if and only if Q", or "P is equivalent to Q".

Remark 0.1.1 Mathematical definitions do *not* require the use of "if and only if". It is common to write a sentence like "define Q to hold if P holds". Before this sentence, the statement P must be defined but the statement Q is not yet defined. After this sentence, the phrase Q is literally a placeholder for the statement P so, by definition, Q holds if and only if P holds. This convention will used throughout this text for all definitions.

Quantifiers are used to construct more complex logical statements from a collection of statements. Let A be a set, possibly finite or infinite. For each element $x \in A$, let $P(x)$ be a

[1] The word "canonical" often carries a technical mathematical meaning. However, in this case, you can simply think of it as synonymous with the words "universal" or "standard".

[2] For details, see an introduction to real analysis.

logical statement. In other words, $P(x)$ is a collection of statements indexed by $x \in A$. There are two basic quantifiers which can construct more intricate statements.

The symbol $\forall$ is the **universal quantifier**, denoting "for all", "for", or "for every". For instance, the statement

$$\forall x \in A, P(x) \tag{0.1.1}$$

can be read aloud as "*For all elements x in the set A, the condition $P(x)$ holds.*" The comma used after the universal quantifier is standard. A well-written proof of (0.1.1) should look something like:

"*Fix $x \in A$.* [...] *Thus, $P(x)$ holds.*"

The phrase "*Fix $x \in A$*" can be replaced by "*Let $x \in A$ be arbitrary*". Note x must be arbitrary and cannot depend on any other quantities.

The symbol $\exists$ is the **existential quantifier**, denoting "there exists", "there is", or "for some". For instance, the statement

$$\exists x \in A \text{ s.t. } P(x) \tag{0.1.2}$$

can be read aloud as "*There exists an element x in the set A such that the condition $P(x)$ holds.*" The phrase "s.t." or "such that" after the existential quantifier is standard and will be this text's convention, but some texts use a comma or the symbol $\ni$ instead. A well-written proof of (0.1.2) should look something like:

"*Take $x = [...]$ so $x \in A$.* [...] *Thus, $P(x)$ holds.*"

The phrase "*Take*" can be replaced by "*Define*", "*Let*", or "*Choose*". Note x may depend on previously defined quantities.

The empty set, denoted $\varnothing$, is the set with no elements. It has a special role with these two quantifiers. If $A = \varnothing$ then (0.1.1) is *vacuously true*, and (0.1.2) is *vacuously false*. These cases may seem harmless but they can be pivotal in some contexts. For this reason, you should always consider the possibility of whether $A = \varnothing$ when proving statements like (0.1.1) or (0.1.2).

Statements with several quantifiers are even more complicated. Most importantly, the order of those quantifiers can sometimes be irrelevant and sometimes be life-altering. Let A and B be sets, possibly finite or infinite. For each $x \in A$ and $y \in B$, let $P(x, y)$ be a logical statement.

- Two consecutive $\forall$ quantifiers can be used in either order. That is,

$$\forall x \in A, \forall y \in B, P(x, y) \iff \forall y \in B, \forall x \in A, P(x, y).$$

- Two consecutive $\exists$ quantifiers can be used in either order. That is,

$$\exists x \in A, \exists y \in B \text{ s.t. } P(x, y) \iff \exists y \in B, \exists x \in A \text{ s.t. } P(x, y).$$

- The order of the two quantifiers $\forall$ and $\exists$ matters. That is, in most cases,

$$\forall x \in A, \exists y \in B \text{ s.t. } P(x, y) \not\Longleftrightarrow \exists y \in B \text{ s.t. } \forall x \in A, P(x, y).$$

Quantifiers are perhaps the most challenging hurdle when first learning logic, so you are encouraged to consult further sources to review this material, e.g. [8, 9].

While logical symbols are valuable in many contexts, they are often disruptive when reading mathematics. It is easier for most readers to parse natural language for logical arguments. Consequently, this text will never again use the symbols $\neg$, $\vee$, or $\wedge$ and instead opt for their natural language equivalents of "not", "or", and "and" respectively. The symbols $\implies$, $\iff$, $\forall$, and $\exists$ will be sparingly used compared to their colloquial names. You are strongly encouraged to do the same in your own writing.

0.1.2 Set relationships

There are three common ways to compare sets.

Definition 0.1.2 A set A is a **subset** of a set B if $\forall x \in A, x \in B$. If so, write $A \subseteq B$.

A well-written proof by definition that $A \subseteq B$ looks something like:

> "*Fix $x \in A$.* [...] *Hence, $x \in B$. As $x \in A$ was arbitrary, this proves $A \subseteq B$.*"

Definition 0.1.3 A set A is **equal** to a set B if $A \subseteq B$ and $B \subseteq A$. If so, write $A = B$.

A well-written proof by definition that $A = B$ looks something like:

> "*First, we show $A \subseteq B$.* [...] *This proves $A \subseteq B$.*"
> "*It remains to show $B \subseteq A$.* [...] *This proves $B \subseteq A$ and hence $A = B$.*"

Sometimes you may want to distinguish subset and equality.

Definition 0.1.4 A set A is a **proper subset** of a set B if $A \subseteq B$ and $A \neq B$. If so, write $A \subsetneq B$.

Remark 0.1.5 Note that $\subsetneq$ means "proper subset" whereas $\nsubseteq$ means "not a subset". These are completely different set relationships, so be careful to not mix them up.

A well-written proof by definition that $A \subsetneq B$ looks something like:

> "*First, we show $A \subseteq B$.* [...] *This proves $A \subseteq B$.*"
> "*It remains to show $B \nsubseteq A$. Take $x =$* [...] *Thus, $x \in B$ but $x \notin A$. This proves $B \nsubseteq A$ and hence $A \subsetneq B$.*"

Remark 0.1.6 Some texts write $A \subset B$ to mean $A \subseteq B$ while other texts mean $A \subsetneq B$. To avoid this confusion, this text will never use the $\subset$ symbol.

0.1.3 Set builder notation

Set builder notation is designed to define a set by picking out elements from another set satisfying certain conditions. This technique requires a collection of logical statements. For each element x in a set A, let $P(x)$ be a logical statement. You can define a subset of A as

$$\{\underbrace{x \in A}_{\text{elements}} : \underbrace{P(x) \text{ holds}}_{\text{condition}}\}, \tag{0.1.3}$$

which is read out loud as "*the set of elements x in A such that $P(x)$ holds.*" Some authors prefer to use a vertical bar $|$ instead of the colon $:$ in (0.1.3).

Example 0.1.7 The set of even integers is given by the set

$$\{n \in \mathbb{Z} : \exists m \in \mathbb{Z} \text{ s.t. } n = 2m\},$$

whereas

$$\{n \in \mathbb{Z} : \forall m \in \mathbb{Z} \text{ s.t. } n = 2m\} = \varnothing.$$

Note that set builder notation always uses a "dummy variable" for the elements. For example, (0.1.3) uses the letter x as a dummy variable. This means you can replace it by an entirely different letter and the set remains the same. For instance,

$$\{x \in A : P(x) \text{ holds}\} = \{y \in A : P(y) \text{ holds}\}.$$

To avoid abuse of notation when writing mathematics, this dummy variable should *never* be defined before the set, and should *not* be used after the set is defined[3].

Multiple conditions in set builder notation can be specified with a comma or "and".

Example 0.1.8 The following sets are equivalent:

$$\{x \in \mathbb{R} : x > 0, x < 9\} = \{x \in \mathbb{R} : x > 0 \text{ and } x < 9\}.$$

While (0.1.3) is the common mathematical standard for set builder notation, it is also acceptable to write

$$\{x : x \in A \text{ and } P(x) \text{ holds}\}. \tag{0.1.4}$$

Note it is critical that (0.1.4) specifies $x \in A$ because the letter x has no meaning otherwise. This variation on set builder notation leads to another flexible method for defining sets using functions; see Section 0.3 for details.

0.1.4 Set operations

Set operations are another common way to define sets. This method requires other sets to already be defined. The most basic operations are union and intersection.

Definition 0.1.9 Let I be a set and, for each $i \in I$, let A_i be a set.

- $\bigcup_{i \in I} A_i := \{x : \exists i \in I \text{ s.t. } x \in A_i\}$ is the **(arbitrary) union** of sets A_i over $i \in I$.
- $\bigcap_{i \in I} A_i := \{x : \forall i \in I, x \in A_i\}$ is the **(arbitrary) intersection** of sets A_i over $i \in I$.

Remark 0.1.10 The set I is often referred to as an *index set*.

There are several special cases that warrant their own notation: two sets, a finite list of sets, and a countably[4] infinite list of sets.

Example 0.1.11 For two sets A and B, their union and intersection are respectively

$$A \cup B = \{x : x \in A \text{ or } x \in B\} \quad \text{and} \quad A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Fix $n \in \mathbb{N}^+$. For a finite list of sets $A_1, A_2, \ldots, A_n$, you may write the finite union as

$$\bigcup_{i=1}^{n} A_i = A_1 \cup \cdots \cup A_n := \{x : x \in A_i \text{ for some } 1 \le i \le n\}$$

and the finite intersection as

$$\bigcap_{i=1}^{n} A_i = A_1 \cap \cdots \cap A_n := \{x : x \in A_i \text{ for all } 1 \le i \le n\}.$$

[3]If you have started a new argument where this set is no longer relevant, then it is permissible to re-use the same letter. However, it is always best practice to avoid abusing notation within the same proof.

[4]The phrase "countably infinite set" has a specific technical meaning. For your purposes, it is enough to think of this as "an infinite subset of $\mathbb{Z}$".

For the countably infinite list of sets $\{A_i : i \in \mathbb{N}^+\}$, you may write the infinite union as

$$\bigcup_{i=1}^{\infty} A_i := \{x : x \in A_i \text{ for some } i \in \mathbb{N}^+\}$$

and the infinite intersection as

$$\bigcap_{i=1}^{\infty} A_i := \{x : x \in A_i \text{ for all } i \in \mathbb{N}^+\}.$$

Another basic operation is subtraction.

Definition 0.1.12 $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ is the **subtraction** of the set B from the set A.

Remark 0.1.13 Some texts may write $A - B$ for set subtraction. This can have several other meanings so it is better to avoid it.

The union, intersection, and subtraction are natural operations and you may have already encountered them in previous courses on single variable calculus and linear algebra. For multivariable calculus, there is another fundamental set operation.

Definition 0.1.14 $A \times B = \{(a,b) : a \in A, b \in B\}$ is the **Cartesian product** of sets A and B.

There are several ways to refer to elements of $A \times B$. By definition, if $z \in A \times B$ then there exists $a \in A$ and $b \in B$ such that $z = (a,b)$. This observation suggests that the following two phrases can be seen as equivalent:

"*Let* $(a,b) \in A \times B$"
"*Let* $x \in A \times B$ *so there exists* $a \in A$ *and* $b \in B$ *such that* $x = (a,b)$."

The latter is the completely unambiguous interpretation of the former. This equivalence will be a convention adopted by this text.

While infinite Cartesian products are interesting, they will not be required to study multivariable calculus. You will, however, need finite Cartesian products.

Definition 0.1.15 Fix $n \in \mathbb{N}^+$. The ***n*-fold Cartesian product** of sets $A_1, \ldots, A_n$ is given by

$$A_1 \times \cdots \times A_n := \{(a_1, \ldots, a_n) : a_1 \in A_1, \ldots, a_n \in A_n\}.$$

Moreover, the ***n*-fold Cartesian power** of a set A is given by

$$A^n := \underbrace{A \times \cdots \times A}_{n \text{ times}} = \{(a_1, \ldots, a_n) : a_1 \in A, \cdots, a_n \in A\}.$$

Again, there are two equivalent ways to refer to elements of $A_1 \times \cdots \times A_n$.

"*Let* $(a_1, \ldots, a_n) \in A_1 \times \cdots \times A_n$"
"*Let* $x \in A_1 \times \cdots \times A_n$ *so there exists* $a_1 \in A_1, \ldots, a_n \in A_n$ *such that* $x = (a_1, \ldots, a_n)$."

This text will interchange these two phrases without mention.

Overall, this concludes a quick review of logic and sets. The conventions and notation will be present throughout this text and you are encouraged to adopt them in your own writing for the sake of consistency. As a friendly reminder, you are welcome to dig deeper for more examples and exercises by consulting more introductory sources such as [8, 9].

0.2. Euclidean space

One set will have a special prominence throughout this text.

Definition 0.2.1 Fix $n \in \mathbb{N}^+$. The set

$$\mathbb{R}^n = \big\{(x_1, \dots, x_n) : x_1 \in \mathbb{R}, \dots, x_n \in \mathbb{R}\big\}$$

is n**-dimensional (Euclidean) space**.

Almost everything will be related to this set in some way. Your prior experiences with linear algebra might overlap with this section but pay attention to the differences. Some conventions may differ dramatically. Some material may be entirely new, especially Sections 0.2.4 and 0.2.5.

0.2.1 Elements, subsets, and Cartesian products

Members of $\mathbb{R}^n$ can be written in two forms: **element-wise** or **component-wise**. Namely,

$$\underbrace{x \in \mathbb{R}^n}_{\text{element-wise}} \quad \text{or} \quad \underbrace{(x_1, \dots, x_n) \in \mathbb{R}^n}_{\text{component-wise}}.$$

These forms are related by the following paragraph.

> "*Let $x \in \mathbb{R}^n$. By definition of the set $\mathbb{R}^n$, for $i \in \{1, \dots, n\}$, there exists $x_i \in \mathbb{R}$ such that x_i is the i^{th} component of x. Thus, $x = (x_1, \dots, x_n)$.*"

This formally correct argument allows you to switch between elements and components, but it is too verbose for frequent use. The above paragraph will be considered equivalent to any of the following acceptable forms:

> "*Let $x \in \mathbb{R}^n$. Let $x_i \in \mathbb{R}$ be the i^{th} component of x for $1 \leq i \leq n$.*"
> "*Let $x \in \mathbb{R}^n$. Write $x = (x_1, \dots, x_n)$.*"
> "*Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.*"

The choice of an element-wise versus component-wise argument depends on how much you need to refer to specific components. If an argument does not really require components or becomes harder to read with components, then avoid them to keep things concise.

The components for $\mathbb{R}, \mathbb{R}^2$, and $\mathbb{R}^3$ have distinguished variables by convention. The word "variable" has no formal meaning, but it is usually mentioned when a specific letter is understood to correspond to a specific component. The word "distinguished" means that there are common letters which are often used by convention.

- An arbitrary element of $\mathbb{R}$ is often written as x, so the letter x refers to its only component.
- An arbitrary element of $\mathbb{R}^2$ is often written as (x, y), so the variables x and y refers to the first and second component respectively.
- An arbitrary element of $\mathbb{R}^3$ is often written as (x, y, z), so the variables x, y, and z refer to the first, second, and third components respectively.
- An arbitrary element of $\mathbb{R}^n$ is often written as $(x_1, \dots, x_n)$, so the variable x_i refer to the i^{th} component.

These are helpful when informally referring to a set using distinguished variables.

Example 0.2.2 In $\mathbb{R}^2$, the equation $x + y = 0$ corresponds to the line

$$\{(x, y) \in \mathbb{R}^2 : x + y = 0\} \subseteq \mathbb{R}^2.$$

In $\mathbb{R}^3$, the equation $x+y=0$ corresponds to the plane

$$\{(x,y,z)\in\mathbb{R}^3 : x+y=0\}\subseteq\mathbb{R}^3.$$

This distinction demonstrates that the context matters when you write down an informal equation with variables. If in doubt, write the formal set.

If you are writing arguments involving elements from multiple different spaces, e.g. $\mathbb{R}^2$ and $\mathbb{R}^3$, then there can be some ambiguity as to which variable refers to which component. In that case, you should use distinct variables or clearly indicate when you are switching from one convention to another.

Now, it is a common misconception that a lower-dimensional Euclidean space is a subset of a higher-dimensional Euclidean space. This statement is false. Indeed,

$$\mathbb{R}\not\subseteq\mathbb{R}^2\not\subseteq\mathbb{R}^3\not\subseteq\mathbb{R}^4\not\subseteq\cdots$$

A specific example can help clarify these issues.

Example 0.2.3 Notice that

$$\mathbb{R}^2\times\{0\}=\{(x,y,z): x\in\mathbb{R}, y\in\mathbb{R}, z=0\}\subseteq\mathbb{R}^3$$

yet

$$\mathbb{R}^2=\{(x,y): x\in\mathbb{R}, y\in\mathbb{R}\}\not\subseteq\mathbb{R}^3.$$

This example highlights another concern about the difference between $\mathbb{R}^3$ and $\mathbb{R}^2\times\mathbb{R}$.

Example 0.2.4 Technically speaking,

$$\mathbb{R}^3=\{(x,y,z): x\in\mathbb{R}, y\in\mathbb{R}, z\in\mathbb{R}\}$$

so an element of $\mathbb{R}^3$ has 3 components; each component is an element of $\mathbb{R}$. Yet,

$$\mathbb{R}^2\times\mathbb{R}=\{((x,y),z): (x,y)\in\mathbb{R}^2, z\in\mathbb{R}\}$$

so an element of $\mathbb{R}^2\times\mathbb{R}$ has 2 components; the first is an element of $\mathbb{R}^2$ and the second is element of $\mathbb{R}$. Similarly,

$$\mathbb{R}\times\mathbb{R}^2=\{(x,(y,z)): x\in\mathbb{R}, (y,z)\in\mathbb{R}^2\}.$$

This creates a weird situation since technically speaking $(2,3,7)\neq((2,3),7)\neq(2,(3,7))$.

To avoid this nuisance, we adopt a general convention.[5]

For any $m,n\in\mathbb{N}^+$, consider the sets $\mathbb{R}^{m+n}$ and $\mathbb{R}^m\times\mathbb{R}^n$ as equal. Thus,

$(x_1,\ldots,x_m,y_1,\ldots,y_n)\in\mathbb{R}^{m+n}$ is the same as $((x_1,\ldots,x_m),(y_1,\ldots,y_n))\in\mathbb{R}^m\times\mathbb{R}^n$.

For instance, this choice implies that

$$\mathbb{R}^3=\mathbb{R}^2\times\mathbb{R}=\mathbb{R}\times\mathbb{R}^2=\mathbb{R}\times\mathbb{R}\times\mathbb{R}.$$

[5]This choice is also referred to as a *canonical isomorphism* between $\mathbb{R}^{m+n}$ and $\mathbb{R}^m\times\mathbb{R}^n$. This fancy language is reserved for more advanced algebraic contexts. For us, it is simply a convention.

0.2.2 Matrices, vectors, and scalars

Euclidean space is often thought of as a space of vectors in linear algebra. Before discussing this perspective in detail, some linear algebra notation must first be fixed.

Definition 0.2.5 Fix $m, n \in \mathbb{N}^+$. An $m \times n$ **matrix** A is a rectangular array of real numbers with m rows and n columns. It is written as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} = \left[A_{ij}\right]_{i,j},$$

where the (i, j)**-entry of** A is denoted $A_{ij} \in \mathbb{R}$ for $i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}$.

Definition 0.2.6 Fix $m, n \in \mathbb{N}^+$. Let $A = \left[A_{ij}\right]_{i,j}$ be an $m \times n$ matrix. The **transpose** of A, denoted A^T, is the $n \times m$ matrix whose (i, j)-entry is A_{ji} for $i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}$. That is,

$$A^T = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{bmatrix} = \left[A_{ji}\right]_{i,j}.$$

Column and row vectors are defined using matrices.

Definition 0.2.7 Fix $m, n \in \mathbb{N}^+$. A **(m-dimensional) column vector** v is an $m \times 1$ matrix, written as

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \quad \text{where } v_1, \ldots, v_m \in \mathbb{R}$$

A **(n-dimensional) row vector** w is a $1 \times n$ matrix, written as

$$w = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \quad \text{where } w_1, \ldots, w_n \in \mathbb{R}.$$

A **scalar** λ is a real number, so $\lambda \in \mathbb{R}$.

Remark 0.2.8 By default, a "vector" will refer to a column vector. When possible, vectors will usually be denoted with Roman letters like u, v, w whereas scalars will usually be denoted with Greek letters like α, β, λ.

Remark 0.2.9 Some may prefer to write $\vec{x} \in \mathbb{R}^n$ instead of $x \in \mathbb{R}^n$. The arrow is a visual cue to remind a reader that the quantity is a vector, but this text will never use this notation[6].

Now, the space $\mathbb{R}^n$ can be viewed as a set of $n \times 1$ column vectors, or $1 \times n$ row vectors, or even as $a \times b$ matrices provided $n = ab$. This creates ambiguity if you want to perform some

[6]There are several reasons to avoid writing $\vec{x}$ instead of x. First, writing arrows will become tedious and cluttered; many quantities will be a vector. Second, good writing is unambiguous and good reading does not require an unnecessary visual cue. Third, omitting arrows is the mathematical norm for texts beyond introductory undergraduate courses. Although it may take some time to get comfortable with it, you are strongly encouraged to adopt the same practice in your own writing. It will be well worth it.

matrix operations, like multiplication or transpose. To remove all ambiguity, you must fix a permanent convention:

For this entire text, elements of $\mathbb{R}^n$ are always $n \times 1$ column vectors.

Notationally speaking, this implies that if $x \in \mathbb{R}^n$ then

$$x = (x_1, \dots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The leftmost and center notation will be used often, whereas the rightmost notation will be rarely utilized. This choice allows for more compact typesetting. Also, notice that $(x_1, \dots, x_n)$ is *not* a row vector, implying that

$$x = (x_1, \dots, x_n) \neq \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = x^T$$

Remark 0.2.10 The notation $[x_1, \cdots, x_n]$ is not defined and confusing because matrices never have commas. Do not use it.

The **zero vector** will be denoted as $0 \in \mathbb{R}^n$. This choice implies that

$$0 = (0, \dots, 0) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n,$$

which might seem really confusing to read. The symbol 0 is being used to denote the scalar $0 \in \mathbb{R}$ and the vector $0 \in \mathbb{R}^n$ simultaneously. This issue is called an *abuse of notation*. Normally, you should avoid this practice but it is acceptable when the context allows for only one plausible interpretation. For instance, if you say that $v, w \in \mathbb{R}^n$ satisfy $v + w = 0$, then the symbol 0 can only be sensibly read as the vector $0 \in \mathbb{R}^n$. Good writing will ensure that the meaning of the symbol 0 is understood without additional clarification.

Finally, you can express matrices in terms of vectors. Fix $m, n \in \mathbb{N}^+$.

- If $v_1, \dots, v_n \in \mathbb{R}^m$ then the $m \times n$ matrix with columns $v_1, \dots, v_n$ can be written as

$$\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} = (v_1, \dots, v_n)^T.$$

- If $w_1, \dots, w_m \in \mathbb{R}^n$ then the $m \times n$ matrix with rows $w_1, \dots, w_m$ can be written as

$$\begin{bmatrix} — & w_1 & — \\ & \vdots & \\ — & w_m & — \end{bmatrix} = (w_1, \dots, w_m).$$

These last two pieces of notation will be used infrequently and, if you choose to use it, you are recommended to remind the reader of any equivalent forms.

0.2.3 Dot product and norm

Geometry of $\mathbb{R}^n$ is often described using the dot product.

Definition 0.2.11 The **dot product** of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ is defined to be

$$x \cdot y := \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}.$$

The vector x is **orthogonal** to y if $x \cdot y = 0$.

Remark 0.2.12 The dot symbol $\cdot$ is used for dot products as well as scalar products like $2 \cdot 3 = 6$. It depends on context.

The dot product has beautiful geometric interpretations which you may have already seen in a course on linear algebra; for details or a quick review, you may wish to view this 3Blue1Brown video or read [14, Section 1.2].

Below is a standard list of algebraic properties that the dot product satisfies.

Lemma 0.2.13 Let $x, y, z \in \mathbb{R}^n$ and let $\lambda \in \mathbb{R}$.

(a) $x \cdot x \geq 0$ with equality if and only if $x = 0$. (Non-negativity)
(b) $x \cdot y = y \cdot x$. (Symmetry)
(c) $x \cdot (y + \lambda z) = x \cdot y + \lambda(x \cdot z)$. (Linearity)

These properties will be applied without mention throughout this text, along with other standard theorems of linear algebra.

The notion of distance in $\mathbb{R}^n$ is defined by the norm.

Definition 0.2.14 The **norm** of a point $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ is given by

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

Remark 0.2.15 Many other texts only use one vertical bar $|$ instead of two $||$ for the norm. In this text, the expression $|x|$ will only be defined as the absolute value of a real number $x \in \mathbb{R}$.

The norm has some standard properties which will be used without mention.

Lemma 0.2.16 Let $x, y \in \mathbb{R}^n$ and let $\lambda \in \mathbb{R}$.

(a) $||x|| \geq 0$ with equality if and only if $x = 0$. (Non-negativity)
(b) $||\lambda x|| = |\lambda| \cdot ||x||$
(c) $||x + y||^2 = ||x||^2 + ||y||^2 + 2(x \cdot y)$ (Law of cosines)
(d) $|x \cdot y| \leq ||x|| \cdot ||y||$ (Cauchy-Schwarz)

One property of the norm will be fundamental to your study.

Lemma 0.2.17 (Triangle inequality) Fix $k \in \mathbb{N}^+$. For $x_1, x_2, \dots, x_k \in \mathbb{R}^n$,

$$||x_1 + \cdots + x_k|| \leq ||x_1|| + \cdots + ||x_k||.$$

Proof. (Sketch) Use induction on k and the law of cosines. ■

The triangle inequality will be your bread-and-butter for countless proofs.

0.2.4 Balls and spheres

There are infinitely many subsets of $\mathbb{R}^n$, and some are extra important. Here you will see the different ways to generalize the idea of "intervals" in $\mathbb{R}$, one way via the norm and another via

Cartesian products. Pay attention to how these definitions compare in $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$, and $\mathbb{R}^n$. You will start by using the norm to define fundamental sets in $\mathbb{R}^n$: balls and spheres.

Definition 0.2.18 Let $r > 0$ and $a \in \mathbb{R}^n$.

- The **open ball of radius r centred at a** is the set

$$B_r(a) = \{x \in \mathbb{R}^n : ||x - a|| < r\}.$$

- The **closed ball of radius r centred at a** is the set

$$\{x \in \mathbb{R}^n : ||x - a|| \leq r\}.$$

- The **sphere of radius r centred at a** is the set

$$\{x \in \mathbb{R}^n : ||x - a|| = r\}.$$

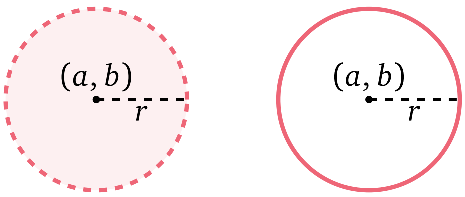
Remark 0.2.19 Other notations for the open ball include $B(a, r), B(a; r)$, or $B_a(r)$. None of these will appear in this text, but they may appear in other sources.

Remark 0.2.20 An open ball (or closed ball) is **punctured** if it excludes the centre. For example, $B_r(a) \setminus \{a\}$ is a punctured open ball.

The words "ball" and "sphere" are inspired by the three-dimensional case. Balls are always solid and spheres are always hollow. In other words, the Earth (and all its layers) is roughly a ball and the surface of the Earth is roughly a sphere.

Example 0.2.21 One-dimensional "balls" reduce to familiar intervals. Let $a \in \mathbb{R}$ and let $r > 0$. Then open ball $B_r(a) = (a - r, a + r)$ is the open interval of length $2r$ centred at a. The closed ball of radius r centered at a is the closed interval $[a - r, a + r]$. The sphere of radius r centred at a is the finite set $\{a - r, a + r\}$, which are the endpoints of those intervals.

Example 0.2.22 Two-dimensional "balls" are **disks** and "spheres" are **circles**. Let $(a, b) \in \mathbb{R}^2$ and let $r > 0$. Then $B_r((a, b)) = \{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 < r^2\}$ is an open disk and the sphere of radius r centred at (a, b) is the circle $\{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 = r^2\}$.



Spheres are special to residents of Earth, so a distinguished sphere will often arise.

Definition 0.2.23 The **$(n-1)$-dimensional unit sphere** in $\mathbb{R}^n$, denoted S^{n-1}, is the sphere of radius 1 centred at the origin. In other words,

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

Example 0.2.24 Note S^1 is the unit circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ in $\mathbb{R}^2$. The 2-dimensional unit sphere S^2 is the familiar surface $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$.

0.2.5 Rectangles and cubes

The other way to generalize intervals in $\mathbb{R}$ is using Cartesian products.

Definition 0.2.25 A **(closed) rectangle** in $\mathbb{R}^n$ is a set R of the form

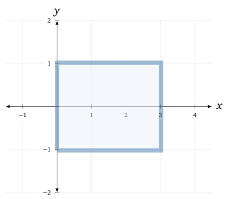
$$R = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \dots, x_n) : x_i \in [a_i, b_i] \text{ for } 1 \leq i \leq n\},$$

where $a_1, b_1, \dots, a_n, b_n \in \mathbb{R}$ and $a_i < b_i$ for all $1 \leq i \leq n$.

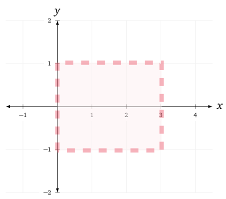
Remark 0.2.26 The set $(a_1, b_1) \times \cdots \times (a_n, b_n)$ is an open rectangle. In general, a "rectangle" refers to a closed rectangle unless specified otherwise.

Example 0.2.27 One-dimensional rectangles are closed intervals, like $[a, b]$ where $a < b$. Open rectangles are open intervals like (a, b). The singleton $\{a\}$ is not a rectangle.

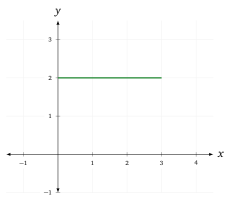
Example 0.2.28 Two-dimensional rectangles are "rectangles" in the usual colloquial sense. The set $[0, 3] \times [-1, 1]$ is a rectangle in $\mathbb{R}^2$. The set $(0, 3) \times (-1, 1)$ is an open rectangle. The set $[0, 3] \times \{2\}$ is not a rectangle. The set $[0, 3] \times (-1, 1)$ is neither an open rectangle nor a closed rectangle.



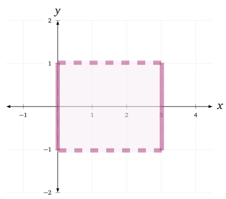
$[0, 3] \times [-1, 1]$



$(0, 3) \times (-1, 1)$



$[0, 3] \times \{2\}$



$[0, 3] \times (-1, 1)$

Example 0.2.29 Three-dimensional rectangles are also called "rectangular prisms" but that is not common lingo in multivariable calculus.

Rectangles where all the sides have the same length have a special name.

Definition 0.2.30 A **(n-dimensional) hypercube** is a set in $\mathbb{R}^n$ of the form

$$[a, b]^n = \underbrace{[a, b] \times \cdots \times [a, b]}_{n \text{ times}}$$

The **(n-dimensional) unit hypercube** is the set $[0, 1]^n$.

Example 0.2.31 A 2-dimensional hypercube is a **square** and $[0, 1]^2$ is the **unit square**. A 3-dimensional hypercube is a **cube** and $[0, 1]^3$ is the **unit cube**. They are all "solid".

This caps off a short summary of Euclidean space, its elements, standard operations, and its common subsets. While you learn the material, you are encouraged to adopt the same conventions and notation in your own writing.

0.3. Functions

Functions are at the centre of your study for multivariable calculus. Below is a general but somewhat informal definition.

> *Let A and B be sets. A* ***function*** *(or* ***map****) $f : A \to B$ assigns to each element $a \in A$ exactly one element $f(a) \in B$. The sets A and B are respectively the* ***domain*** *and* ***codomain*** *of f.*

The domain consists of "all possible inputs" and the codomain consists of "all possible outputs". For a more detailed introduction to abstract functions, see [8, Chapter 3].

This text will be focused on maps whose domain and codomain are subsets of some higher dimensional Euclidean space. These are referred to as maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ or, equivalently,

- "*f is a map from n dimensions to m dimensions.*"
- "*f is a $\mathbb{R}^m$-valued function of n real variables.*"
- "*f is a function of n real variables with m-dimensional output.*"
- "*f is a vector-valued multivariable function.*"

Formally, all of these statements mean that f is a function $f : A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ for some $m, n \in \mathbb{N}^+$. In this section, you will review some standard notation and conventions for such functions.

0.3.1 Domain and codomain

To unambiguously define a function, you must first specify its domain and codomain, and then you can specify the rule to assign elements. This will generally look something like:

> "*Define the sets $A = [...]$ and $B = [...]$.*
> *Define the map $f : A \to B$ by $f(x) = [...]$ for $x \in A$.*"

The order of the sentences is crucial. You must first define the domain A and codomain B. Otherwise, the function rule itself cannot be defined. Moreover, for the function rule, it is better to write the phrase "*for $x \in A$*" so that the dummy variable x is necessarily defined.

For maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, there are some special conventions. First, functions can be defined element-wise or component-wise.

Example 0.3.1 The norm function can defined element-wise by

> "*Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = ||x||$ for $x \in \mathbb{R}^n$.*"

It can also be defined component-wise by

> "*Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, \ldots, x_n) = (x_1^2 + \cdots + x_n^2)^{1/2}$ for $(x_1, \ldots, x_n) \in \mathbb{R}^n$.*"

Second, if you do not need to restrict your codomain, then it should be all of $\mathbb{R}^m$. One of the rare moments where you might restrict the codomain occurs when studying inverse functions; see Section 0.3.3 for details.

Example 0.3.2 The norm function could be defined in two different ways:

> "*Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = ||x||$ for $x \in \mathbb{R}^n$.*"
> "*Define $g : \mathbb{R}^n \to [0, \infty)$ by $g(x) = ||x||$ for $x \in \mathbb{R}^n$.*"

Both are valid but, unless you have a good reason, it is better to define f. Note $f \neq g$ since the codomains differ.

Third, if you do not specify the domain, then it is assumed to be the set of all inputs where the function rule is defined.

Example 0.3.3 The norm function can be defined in this simplified form:

"*Define* $f(x_1,\dots,x_n) = (x_1^2+\cdots+x_n^2)^{1/2}$".

This statement is equivalent to the unambiguous ones appearing in Example 0.3.1, since the expression $(x_1^2+\cdots+x_n^2)^{1/2}$ is defined for all $(x_1,\dots,x_n) \in \mathbb{R}^n$.

Example 0.3.4 Here is a completely unambiguous definition of a function.

"*Define* $A = \{(x,y) \in \mathbb{R}^2 : xy > 0\}$.
Define $h : A \to \mathbb{R}$ *by* $h(x,y) = \log(xy)$ *for* $(x,y) \in A$."

Since the expression $\log(xy)$ is defined if and only if $(x,y) \in A$, this can be equivalently defined in a simplified form.

"*Define* $h(x,y) = \log(xy)$."

Notice this sentence also carries two implicit assumptions. The letters x and y are the distinguished variables for $\mathbb{R}^2$ and $h(x,y)$ is defined only when $(x,y) \in A$ so the domain of h is A. Moreover, the expression $\log(xy)$ is a real number so the codomain of h is $\mathbb{R}$. In other words, this sentence is more concise but you must interpret it carefully.

This last convention is a shortcut and can lead to poor writing habits if you are not careful. If in doubt, you are encouraged to define your functions without any ambiguity whatsoever.

0.3.2 Image and preimage

There are two different ways to build sets using functions. The first method is via the image. Given a function $f : A \to B$ and a subset $S \subseteq A$, you can define a subset of B by

$$\{y \in B : \exists x \in S \text{ s.t. } y = f(x)\} \tag{0.3.1}$$

This definition is a bit clunky with the existential quantifier so mathematicians often prefer a more compact variant. The set (0.3.1) can be rewritten as (0.3.2) below.

Definition 0.3.5 Let A and B be sets. Let $f : A \to B$ be a function. For any set $S \subseteq A$, the **image of S under f**, denoted $f(S)$, is the set defined as

$$f(S) = \{f(x) : x \in S\}. \tag{0.3.2}$$

In particular, the **image of f**, denoted $\mathrm{im}(f)$, is the image of A under f, that is,

$$\mathrm{im}(f) = f(A).$$

Remark 0.3.6 The image of f is also commonly called the **range of f**.

Images under f satisfy some standard properties.

Lemma 0.3.7 Let A and B be sets. Let $f : A \to B$ be a function. Let S and T be subsets of A.
(a) $f(\varnothing) = \varnothing$.
(b) $f(S) \subseteq B$.
(c) If $S \subseteq T$, then $f(S) \subseteq f(T)$.
(d) $f(S \cup T) = f(S) \cup f(T)$.
(e) $f(S \cap T) \subseteq f(S) \cap f(T)$.

Proof. These are left as an exercise. They all follow quickly from the definitions. ■

The second method for building sets using functions is via the preimage.

Definition 0.3.8 Let A and B be sets. Let $f : A \to B$ be a function. For any set $U \subseteq B$, the **preimage of U under f**, denoted $f^{-1}(U)$, is the set defined as

$$f^{-1}(U) = \{x \in A : f(x) \in U\}. \tag{0.3.3}$$

Remark 0.3.9 The preimage is only defined on *subsets* of the codomain B. In other words, if $b \in B$ then $f^{-1}(\{b\})$ is defined and is a subset of A, but $f^{-1}(b)$ is not defined. The notation f^{-1} is commonly used for the inverse function of f, but that may not necessarily exist. Preimages under f always exist, so f^{-1} will be used with this meaning by default.

Preimages under f also satisfy some standard properties.

Lemma 0.3.10 Let A and B be sets. Let $f : A \to B$ be a function. Let U and V be subsets of B.
(a) $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(B) = A$.
(b) $f^{-1}(U) \subseteq A$.
(c) If $U \subseteq V$, then $f^{-1}(U) \subseteq f^{-1}(V)$.
(d) $f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$.
(e) $f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V)$.

Proof. These are left as an exercise. They all follow quickly from the definitions. ■

Composing image and preimage yield different results depending on the order of composition.

Lemma 0.3.11 Let A and B be sets. Let $f : A \to B$ be a function. Let $S \subseteq A$ and let $U \subseteq B$. Then

$$S \subseteq f^{-1}\big(f(S)\big) \quad \text{and} \quad f\big(f^{-1}(U)\big) \subseteq U.$$

Proof. These are left as an exercise. They follow quickly from the definition. To convince yourself of their truth, draw some pictures. ■

This captures the key facts about images and preimages of maps. Unlike domains and codomains, there are no special conventions for maps from $\mathbb{R}^n$ to $\mathbb{R}^m$.

0.3.3 Inverse functions

The existence of inverse functions will be a major question later in this text.

Definition 0.3.12 Let $f : A \to B$ be a function. A function $g : B \to A$ is an **inverse** of f provided both of the following hold:
- $\forall x \in A, g(f(x)) = x$
- $\forall y \in B, f(g(y)) = y$

Remark 0.3.13 If an inverse of f exists, then you can prove it is unique and it is usually denoted $f^{-1} : B \to A$. Moreover, for $b \in B$, the quantity $f^{-1}(b)$ is defined when the inverse of f exists. Remember you must be careful not to confuse this notation with the preimage.

There are three standard properties of functions that relate to the existence of an inverse.

Definition 0.3.14 A function $f : A \to B$ is **injective** (or **one-to-one**) provided for every $a_1, a_2 \in A$, if $f(a_1) = f(a_2)$ then $a_1 = a_2$.

Definition 0.3.15 A function $f : A \to B$ is **surjective** (or **onto**) provided for every $b \in B$, there exists $a \in A$ such that $f(a) = b$.

Definition 0.3.16 A function $f : A \to B$ is **bijective** if f is both injective and surjective.

These definitions give an equivalent characterization for the existence of the inverse.

Lemma 0.3.17 The inverse of a function $f : A \to B$ exists if and only if f is bijective.

Proof. Left as an exercise. This follows from a methodical application of the definitions. ∎

Most functions are not usually bijective but, by restricting their domains and codomains, you can produce a new bijective function.

Example 0.3.18 The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is neither surjective nor injective. However, the function $g : (-\infty, 0) \to (0, \infty)$ given by $g(x) = x^2$ is bijective.

0.3.4 Function operations

There are many ways to construct new functions from old functions using **function operations**. Many of these are only defined when the domain and codomain are subsets of a Euclidean space. Thus, for the remainder of this section, all functions will be vector-valued maps with vector variables. You must pay attention to the domains and codomains, especially the dimensions of each space. Otherwise, some operations do not make sense.

First, the most basic function operation is composition.

Definition 0.3.19 Let $f : A \to B$ and $g : B \to C$ be functions. The **composition of g with f** is the function $(g \circ f) : A \to C$ given by

$$(g \circ f)(x) = g(f(x)) \quad \text{for } x \in A.$$

Remark 0.3.20 Note there are three functions here, namely f, g, and $g \circ f$. This means the statement $(g \circ f)(x) = g(f(x))$ is a definition; it is not simply notation.

Second, you can take linear combinations of functions provided their domains are the same, and their codomains are both equal to the same Euclidean space.

Definition 0.3.21 Let $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$ be functions. Fix a scalar $\lambda \in \mathbb{R}$. A **linear combination of f and g** is a function $(f + \lambda g) : A \to \mathbb{R}^m$ given by

$$(f + \lambda g)(x) = f(x) + \lambda g(x) \quad \text{for } x \in A.$$

Third, you can define the dot product of two functions provided their domains are the same, and their codomains are both equal to the same Euclidean space.

Definition 0.3.22 Let $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$ be functions. The **dot product of f and g** is the real-valued function $(f \cdot g) : A \to \mathbb{R}$ given by

$$(f \cdot g)(x) = f(x) \cdot g(x) \quad \text{for } x \in A.$$

Fourth, you can define the product of a real-valued function with a vector-valued map.

Definition 0.3.23 Let $f : A \to \mathbb{R}^m$ be a vector-valued function. Let $\phi : A \to \mathbb{R}$ be a real-valued function. The **scalar product of ϕ and f** is the function $(\phi f) : A \to \mathbb{R}^m$ given by

$$(\phi f)(x) = \phi(x)f(x) = \begin{bmatrix} \phi(x)f_1(x) \\ \vdots \\ \phi(x)f_m(x) \end{bmatrix} \qquad \text{for } x \in A.$$

Fifth and finally, you can define the quotient of real-valued maps.

Definition 0.3.24 Let $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ be real-valued functions. Define $S \subseteq A$ by $S = \{x \in A : g(x) \neq 0\}$. The **quotient** of f over g is the function $(\frac{f}{g}) : S \to \mathbb{R}$ given by

$$\Big(\frac{f}{g}\Big)(x) = \frac{f(x)}{g(x)} \qquad \text{for } x \in S.$$

There are certainly more operations than what is listed here, but this will suffice for now. These standard operations will be vital when studying how they interact with more complex calculus phenomena like limits, derivatives, and integrals.

0.3.5 Supremum and infimum

Finally, the supremum and infimum of functions will be prevalent throughout this text. A very quick will be summarized in this subsection without any proofs. These ideas are quite subtle so you may wish consult other sources for a more detailed discussion; see, for example, this pair of MAT137 videos [9].

Now, recall the definition for a supremum of a set.

Definition 0.3.25 (Supremum of a set) Let $S \subseteq \mathbb{R}$ be a subset of real numbers.
- An element $b \in \mathbb{R}$ is an **upper bound for** S if $\forall x \in A, x \leq b$.
- An element $s \in \mathbb{R}$ is the **least upper bound of** S (or **supremum of** S) if s is an upper bound for S and, for every upper bound b of S, $s \leq b$. Denote $s = \sup(S)$.
- An element $m \in \mathbb{R}$ is the **maximum of** S if $m = \sup(S)$ and $m \in S$.

If the supremum of S exists, then $\sup(S) \in \mathbb{R}$ and, if not, write $\sup(S) = \infty$.

The infimum of a set is defined similarly.

Definition 0.3.26 (Infimum of a set) Let $S \subseteq \mathbb{R}$ be a set.
- An element $b \in \mathbb{R}$ is a **lower bound for** S if $\forall x \in S, b \leq x$.
- An element $i \in \mathbb{R}$ is the **greatest lower bound of** S (or **infimum of** S) if i is a lower bound for S and, for every lower bound b of S, $b \leq i$. Denote $i = \inf(S)$.
- An element $m \in \mathbb{R}$ is the **minimum of** S if $m = \inf(S)$ and $m \in S$.

If the infimum of S exists, then $\inf(S) \in \mathbb{R}$ and, if not, write $\inf(S) = -\infty$.

Two fundamental principles describe precisely when these quantities exist.

Theorem 0.3.27 (Least upper bound & greatest lower bound principles) Let $S \subseteq \mathbb{R}$ be a subset of real numbers. Both of the following hold.

i) The supremum of S exists if and only if there exists an upper bound for S.
ii) The infimum of S exists if and only if there exists a lower bound for S.

There are several equivalent definitions for the supremum of a set.

Lemma 0.3.28 Let $S \subseteq \mathbb{R}$ be a set. Fix $s \in \mathbb{R}$. All of the following are equivalent.

(a) s is the supremum of S.

(b) $-s$ is the infimum of $\{-x : x \in S\}$.

(c) $\big(\forall x \in S,\ x \leq s\big)$ and $\big(\forall \varepsilon > 0, \exists x \in S$ such that $x > s - \varepsilon\big)$.

The supremum (resp. infimum) of a real-valued function on a subset of its domain is a shorthand for the supremum (resp. infimum) of its image.

Definition 0.3.29 (Supremum and infimum of a function) Let $f : D \to \mathbb{R}$ be a real-valued function. Let A be a subset of its domain D. The **supremum of f on A** and the **infimum of f on A** are given by

$$\sup_{x \in A} f(x) := \sup\{f(x) : x \in A\}, \qquad \inf_{x \in A} f(x) := \inf\{f(x) : x \in A\}.$$

There are some standard properties of suprema and infima of functions.

Lemma 0.3.30 Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be real-valued functions. Let A and B be subsets of their domain D. Fix $\lambda \in \mathbb{R}$. All of the following hold:

(a) If $A \subseteq B$ then $\sup_{x \in A} f(x) \leq \sup_{x \in B} f(x)$ and $\inf_{x \in A} f(x) \geq \inf_{x \in B} f(x)$.

(b) $\sup_{x \in A}(f+g)(x) \leq \sup_{x \in A} f(x) + \sup_{x \in A} g(x)$ and $\inf_{x \in A}(f+g)(x) \geq \inf_{x \in A} f(x) + \inf_{x \in A} g(x)$.

(c) If $\lambda \in \mathbb{R}$ satisfies $\lambda > 0$ then $\sup_{x \in A}(\lambda f)(x) = \lambda \sup_{x \in A} f(x)$ and $\inf_{x \in A}(\lambda f)(x) = \lambda \inf_{x \in A} f(x)$.

(d) If $\lambda \in \mathbb{R}$ satisfies $\lambda < 0$ then $\sup_{x \in A}(\lambda f)(x) = \lambda \inf_{x \in A} f(x)$ and $\inf_{x \in A}(\lambda f)(x) = \lambda \sup_{x \in A} f(x)$.

Remark 0.3.31 If the suprema/infima do not exist, then the above statements can be suitably interpreted with $\pm\infty$ but you should be careful when writing proofs in such cases.

This concludes the review of functions and, more broadly, the chapter on notation and conventions. There was an overwhelming amount of symbols and you probably found it somewhat dry, but you are at least familiar with the basics for reading more advanced mathematics. Whenever you find yourself perplexed by some notation or expositional style, you may find it helpful to search this chapter for an explanation. More importantly, you are ready to launch into the exciting world of maps from $\mathbb{R}^m$ to $\mathbb{R}^n$. Enjoy the start of your ride!

1. Maps

You will begin this story by learning about the cast of characters: maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. These are so ubiquitous that there are many names and ways of referring to them, such as:

- *"f is a map from n dimensions to m dimensions."*
- *"f is a $\mathbb{R}^m$-valued function of n real variables."*
- *"f is a function of n real variables with m-dimensional output."*
- *"f is a vector-valued multivariable function."*

The word "real" is often omitted when it is understood. Formally, all of these statements mean that f is a function $f : A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. These maps (or functions)

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

are the centre of our study. Before embarking into the theory, you will want to address some motivational questions.

> *What do these maps represent? How can I interpret them? What are some common examples? How can they be used to model complex situations?*

This foundational chapter is exploratory in nature so it will be quite casual with few formal definitions and little mathematical rigour. You will become fluent with standard terminology and notation. You will explore key examples of multivariable functions, and practice using functions to model scientific phenomena. Most importantly, you will learn to visualize and describe maps using multiple viewpoints.

Deep theory will come soon but too much rigidity will prevent you from making connections and appreciating the scientific context of this mathematics. These maps lurk everywhere in your everyday life, so try searching for them. Allow your mind to be flexible!

1.1. Parametric curves

This section is about maps of the form

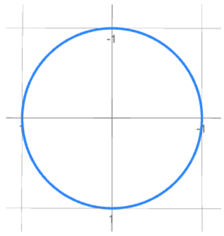
$$\mathbb{R} \longrightarrow \mathbb{R}^n.$$

These are sometimes called **vector-valued functions of a real variable**, especially when $n \geq 2$. They are more commonly called **parametric curves** because that is exactly what they describe physically. The “parameter” is the input variable and is often referred to as **time**.

Example 1.1.1 Define the map $\gamma_1 : [-1,1] \to \mathbb{R}$ as $\gamma_1(t) = 2t^2$. This is simply a function that maps $\mathbb{R}$ to $\mathbb{R}$, which you have seen many times before. The image of γ_1 is the set $[0,2]$. As the time t varies from $t = -1$ to $t = 0$ to $t = 1$, you can imagine walking along the real number line from $\gamma(-1) = 2$ to $\gamma(0) = 0$ back to $\gamma(1) = 2$.

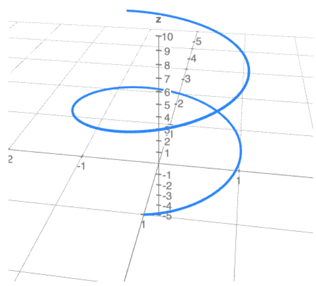
When we normally plot a map $\mathbb{R} \to \mathbb{R}$, we plot both the input space (x-axis) and the output space (y-axis) simultaneously in a single graph. However, when plotting parametric curves, we only care about the output, so a map $\mathbb{R} \to \mathbb{R}^n$ is illustrated by plotting its image in n dimensions.

Example 1.1.2 Define the map $\gamma_2 : [0,2\pi] \to \mathbb{R}^2$ as $\gamma_2(t) = (\cos(t), \sin(t))$. The image of γ_2 is the unit circle in $\mathbb{R}^2$, namely the set $\gamma_2([0,2\pi]) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.



View this Math3D demo to watch how the path is traced out over time.

Example 1.1.3 Consider the map $\gamma_3 : [0,\infty) \to \mathbb{R}^3$ given by $\gamma_3(t) = (\cos(t), \sin(t), t)$. Its image is a helix. As time increase, the helix is traced out in an upward fashion.



View this Math3D demo to watch how the path is traced out over time.

Example 1.1.4 Not all parametric curves can be easily illustrated. For example, the image of the map $\gamma_4(t) = (t^2, \sin(t), e^{-t}, t)$ lies in $\mathbb{R}^4$ so we cannot draw it. Similarly, the parametric curve $\gamma_n : \mathbb{R} \to \mathbb{R}^n$ given by $\gamma_n = (t, t^2, \dots, t^n)$ cannot be plotted for $n \geq 4$.

Example 1.1.5 Let $p, q \in \mathbb{R}^n$. A straight line path $\gamma : [0,1] \to \mathbb{R}^n$ from point p to point q is defined by $\gamma(t) = (1-t)p + tq$. You can also view this formula as $\gamma(t) = p + t(q-p)$ since p is the starting point and $(q-p)$ is the difference between the end and start.

1.1.1 Motion

A parametric curve $\gamma : I \to \mathbb{R}^n$ for some interval $I \subseteq \mathbb{R}$ describes the motion of an object moving in $\mathbb{R}^n$. The **position** of the object in $\mathbb{R}^n$ at time t is precisely $\gamma(t)$. What is its velocity? A natural suggestion might be: for $t \in I$, the **velocity** of the object at time t is given by the vector

$$\gamma'(t) = \lim_{h\to 0} \frac{\gamma(t+h)-\gamma(t)}{h},$$

which is a limit of average velocities. But be careful! The variable h is a *scalar* whereas the quantity $\frac{\gamma(t+h)-\gamma(t)}{h}$ is a *vector*. You have only seen limits of scalar quantities with scalar limit variables, so the above notion of a limit is not the same! You will formally study these later but do not worry about questions of limits or differentiability right now.

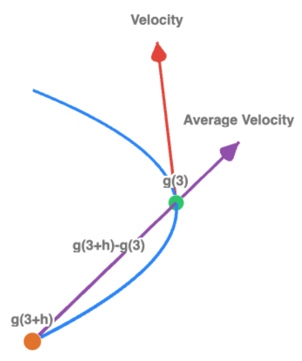
Example 1.1.6 An object in 3-dimensional space is moving along the path described by

$$g(t) = \left(2t-6,\ 5(t-3)^2,\ \frac{1}{2}t^2-4\right).$$

To illustrate the idea of a limit of average velocities, you can analyze the velocity $g'(3)$ using the limit definition. Similarly to the one-dimensional derivative definition, the approximation improves by averaging over shorter time intervals. Here the time interval is $[3, 3+h]$. Using a table of values, notice how our average velocities (resp. average speeds) approach a specific vector quantity (resp. scalar quantity) as $h \to 0$:

h	$\frac{g(3+h)-g(3)}{h}$	$\left\|\frac{g(3+h)-g(3)}{h}\right\|$
1	$(2,5,3.5)$	6.42
0.1	$(2,0.5,3.05)$	3.68
0.01	$(2,0.05,3.005)$	3.61
0.001	$(2,0.005,3.0005)$	3.61
0.0001	$(2,0.00005,3.00005)$	3.60

You might guess that $g'(3) \approx (2,0,3)$ and $||g'(3)|| \approx 3.60$ as that is what our average velocities and average speeds seem to approach. This limiting process also has a nice geometric representation. View this Math3D demo to visualize the process. Press the play button on the h slider to begin.



Notice that as $h \to 0$, the orange point $g(3+h)$ approaches the green point $g(3)$, so the distance $||g(3+h)-g(3)||$ is decreasing. The purple arrow is the average velocity and the red arrow is the velocity $g'(3)$. Notice how our average velocity approaches the velocity when $h \to 0$ as we observed with our table of values.

You can write

$$\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$$

for n single variable functions $\gamma_i : I \to \mathbb{R}$. These are the **component functions** of γ. Since they are single variable functions, you can apply your single variable calculus tools to compute their derivatives $\gamma_i'(t)$. This reduction to single variable calculus will be a recurring theme throughout multivariable calculus. As you will later see, the derivative of γ can be written as:

$$\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_n'(t)).$$

The **speed** of the object at time t is the magnitude of the velocity $||\gamma'(t)||$, which satisfies

$$||\gamma'(t)|| = \sqrt{\gamma_1'(t)^2 + \cdots + \gamma_n'(t)^2},$$

and the direction of motion is the **unit tangent vector** , denoted as $T = T(t)$ and defined by

$$T(t) = \frac{\gamma'(t)}{||\gamma'(t)||}.$$

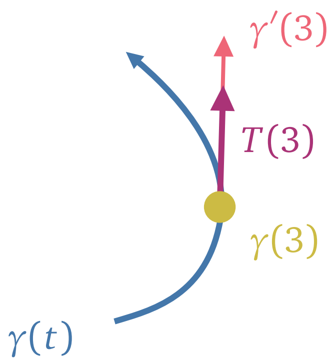
Example 1.1.7 Now, you can explicitly compute the velocity in Example 1.1.6 by finding γ'. To take the derivative of γ, you calculate the derivative of each component function:

$$\gamma'(t) = \left(\frac{d}{dt}(2t-6),\ 5\frac{d}{dt}(t-3)^2, \frac{d}{dt}\left(\frac{1}{2}t^2 - 4\right)\right) = (2, 10(t-3), t).$$

Then the velocity at $\gamma(3)$ is given by $\gamma'(3) = (2, 0, 3)$ and speed $||\gamma'(3)|| = \sqrt{13} \approx 3.60$ which matches the guesses in the previous example. The direction of motion at $\gamma(3)$ is given by the unit tangent vector:

$$T(3) = \frac{\gamma'(3)}{||\gamma'(3)||} = \frac{1}{\sqrt{13}}(2, 0, 3).$$

Below is an illustration of these quantities:



Notice that $\gamma'(3)$ and $T(3)$ are in the same direction except $T(3)$ is normalized by definition.

Similarly, the **acceleration** of the object is the second derivative γ'' so

$$\gamma''(t) = \lim_{h \to 0} \frac{\gamma'(t+h) - \gamma'(t)}{h} = (\gamma_1''(t), \ldots, \gamma_n''(t)).$$

This measures how your velocity is changing.

Example 1.1.8 When you throw a ball as a far as you can, what trajectory does it follow? Using your new tools for dealing with motion in $\mathbb{R}^3$ and some intuition, you can derive such a formula. Let $\gamma : [0,\infty) \to \mathbb{R}^3$ describe the trajectory of a ball. At $t = 0$, say we throw our ball with initial velocity $\gamma'(0) = (v_x, v_y, v_z)$ from the origin. Assuming the projectile only accelerates downwards in the vertical direction due to gravity, you might guess that its velocity is given by

$$\gamma'(t) = (v_x, v_y, v_z - gt)$$

where $g = 9.81\text{m/s}^2$ is the acceleration due to gravity. Since $\gamma'(t)$ is just the derivative of $\gamma(t)$ then it is tempting to write

$$\gamma(t) = \int_0^t \gamma'(u)du,$$

but γ' outputs a vector! Ignore rigour for now and assume that integrating over a column vector means integrating over each of its components. Then,

$$\gamma(t) = \int_0^t \gamma'(u)du = \left(\int_0^t v_x du, \int_0^t v_y du, \int_0^t (v_z - gu)du\right) = \left(v_x t, v_y t, v_z t - \frac{1}{2}gt^2\right).$$

These are the classic kinematics equations governing projectile motion in three dimensions! Play with this Math3D demo to test out this new expression. Note our motion is entirely dependent on what values we assign the initial velocities v_x, v_y, and v_z.

1.1.2 Frenet frame in three dimensions

Describing the motion of an object *relative to its frame* is very useful in physics and engineering. For example, an airline pilot needs to have very precise information about how the plane is moving, but there is nobody outside the plane to measure this information. The measurements must be done by the plane's systems, so they must all be relative to the pilot's frame.

As you have seen, the direction of motion is given by the unit tangent vector T. To more accurately describe the motion of an object, you want to know: how is my direction of motion changing? Naturally, you may compute the derivative $T'(t)$ but this is not usually a unit vector and a direction should be a unit vector. Instead, define the **(principal) unit normal** to be[1]

$$N(t) = \frac{T'(t)}{||T'(t)||}.$$

Somewhat surprisingly, T and N are *orthogonal*.

Example 1.1.9 Define $\gamma : \mathbb{R} \to \mathbb{R}^3$ as $\gamma(t) = (\cos(t), \sin(t), t)$ for all $t \in \mathbb{R}$. You can compute γ', γ'', T, and N. Its derivative is calculated by taking each component function's derivative:

$$\gamma'(t) = \left(\frac{d}{dt}\cos(t), \frac{d}{dt}\sin(t), \frac{d}{dt}t\right) = (-\sin(t), \cos(t), 1).$$

Similarly to the first derivative, to find $\gamma''(t)$, take the derivative of each component of $\gamma'(t)$:

$$\gamma''(t) = \left(-\frac{d}{dt}\sin(t), \frac{d}{dt}\cos(t), \frac{d}{dt}1\right) = (-\cos(t), -\sin(t), 0).$$

[1]You will later encounter a different definition of the unit normal in Section 12.3.1, which is specific in $\mathbb{R}^2$. For now, you may ignore this difference.

Since $||\gamma'(t)|| = \sqrt{\sin^2(t) + \cos^2(t) + 1} = \sqrt{2}$, it follows that

$$T(t) = \frac{\gamma'(t)}{||\gamma'(t)||} = \frac{1}{\sqrt{2}}(-\sin(t), \cos(t), 1)$$

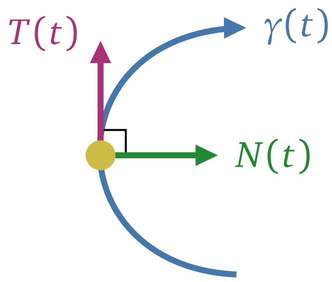
in which case

$$T'(t) = \frac{1}{\sqrt{2}}(-\cos(t), -\sin(t), 0).$$

As $||T'|| = \frac{1}{\sqrt{2}}$, we see that N is given by

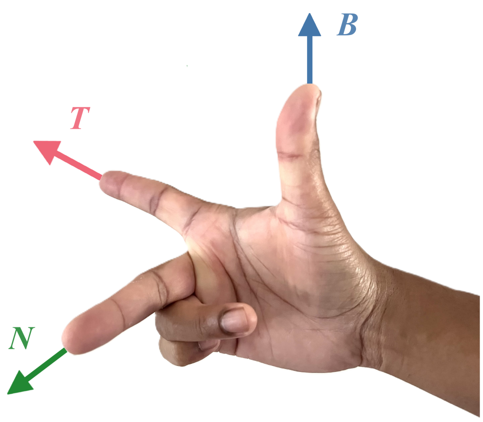
$$N(t) = \frac{T'(t)}{||T'(t)||} = \frac{\sqrt{2}}{\sqrt{2}}(-\cos(t), -\sin(t), 0) = (-\cos(t), -\sin(t), 0).$$

Below is an illustration of how the unit normal is oriented on the curve traced by $\gamma(t)$.



Remark 1.1.10 Note $\gamma'(t)$ is always a scalar multiple of $T(t)$. In the above example, $\gamma''(t)$ was a scalar multiple of $N(t)$ but this is a coincidence. It happened because $||\gamma'(t)||$ was just a scalar and not a function of t. If $||\gamma'(t)||$ was some function of t then γ'' is not necessarily a scalar multiple of $N(t)$.

The unit tangent T and unit normal N span a two-dimensional plane in $\mathbb{R}^3$ so, based on our linear algebra intuition, these cannot be enough to span all kinds of motion in $\mathbb{R}^3$. One more vector is required. Since T and N are already orthogonal unit vectors, you can choose another unit vector B which is orthogonal to both of them. However, both B and $-B$ will be orthogonal to T and N, so this choice is ambiguous. To remove this ambiguity, define the **binormal unit vector** B to be the *unique* unit vector such that $\{T, N, B\}$ form a positively-oriented ordered orthogonal basis in $\mathbb{R}^3$. Geometrically, this means T, N, B satisfies the right-hand rule.



This ordered basis $\{T, N, B\}$ forms the **Frenet frame** (or Frenet–Serret frame or TNB frame) describing the motion of an object in three dimensions.

Example 1.1.11 Continuing with Example 1.1.9, recall you have calculated the unit tangent T and the unit normal N for the path $\gamma(t) = (\cos(t), \sin(t), t)$. To find B algebraically, you can use the cross product

$$B = T \times N.$$

This special product only works for vectors in $\mathbb{R}^3$ and it has a neat memory trick. The cross product $a \times b$ of two vectors $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ can be calculated by expressing it as a "determinant". That is, if $\{e_1, e_2, e_3\}$ is the standard basis in $\mathbb{R}^3$, then

$$a \times b = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}.$$

This is not really a determinant (since some components are vectors) but by naively following the rules of calculating determinants, you will somehow end up with the correct expression[2]. In this example, you get that

$$B = T \times N = \frac{1}{\sqrt{2}} \det \begin{bmatrix} e_1 & e_2 & e_3 \\ -\sin(t) & \cos(t) & 1 \\ -\cos(t) & -\sin(t) & 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \sin(t) e_1 - \frac{1}{\sqrt{2}} \cos(t) e_2 + \frac{1}{\sqrt{2}} e_3,$$

so $B(t) = \frac{1}{\sqrt{2}}(\sin t, -\cos t, 1)$. View this Math3D demo of the TNB frame for γ. The green arrow is $T(t)$, the red is $N(t)$ and the orange is $B(t)$. Notice how they remain orthogonal to each other throughout the motion while being positively oriented.

1.1.3 Geometry of curves

The **trace** of a parametric curve $\gamma : I \to \mathbb{R}^n$ is the image of γ, that is, $\gamma(I)$. This is also referred to as the path traced out by γ.

Example 1.1.12 The unit circle in $\mathbb{R}^2$, $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, can be described as the trace of a parametric curve as you have seen in Example 1.1.2. That parametric curve however traced the unit circle only once. You could instead define a map $\gamma : [0, 4\pi] \to \mathbb{R}^2$ as $\gamma(t) = (\cos(t), \sin(t))$ that has an identical trace to the map defined in Example 1.1.2, but the circle loops around twice.

Notice then that many different parametric curves γ can have the same trace.

Example 1.1.13 There are many ways to trace the unit circle in $\mathbb{R}^2$:

- Define $\gamma_1 : [0, \pi] \to \mathbb{R}^2$ by $\gamma_1(t) = (\cos(2t), \sin(2t))$. Then γ_1 traces the unit circle twice as fast.
- Define $\gamma_2 : [0, 2\pi] \to \mathbb{R}^2$ by $\gamma_2(t) = (\cos(t - \pi), \sin(t))$. Then γ_2 traces the unit circle but starts at π instead of 0 radians of rotation.
- Define $\gamma_3 : [0, 6\pi] \to \mathbb{R}^2$ by $\gamma_3(t) = (\cos(t), -\sin(t))$. Then γ_3 traces the unit circle three times in the opposite direction.
- Define $\gamma_4 : [0, 14.1] \to \mathbb{R}^2$ by $\gamma_4(t) = \left(\cos\left(\frac{t}{4}\sin(t)\right), \sin\left(\frac{t}{4}\sin(t)\right)\right)$. Then γ_4 traces the unit circle in an...interesting way.

View this Desmos animation for a visual demonstration of the maps.

[2]The formal explanation for this coincidence requires some advanced graduate level math.

The trace of a parametric curve is simply a set, so you may be tempted to introduce the following definition:

*A set $C \subseteq \mathbb{R}^n$ is a **curve** if C is the trace of a continuous*[3] *parametric curve $\gamma : I \to \mathbb{R}^n$.*

Somewhat surprisingly, this attempted definition has many pitfalls. You will explore these issues much later in this text but, for now, you can experiment with this initial guess.

Geometers are often concerned with the shape of a curve C but not how C is traced out. How many times does C cross itself? How curvy is C? In other words, you may want to study curves C without worrying how it is described. There are also ways to describe a curve without directly using parametric curves.

Example 1.1.14 Define the set

$$C = \{(x, y) \in \mathbb{R}^2 : y = x^2, -2 \le x \le 2\}$$

so C describes the graph of the parabola $y = x^2$ on the domain $[-2, 2]$. Intuitively, you would consider C to be a curve and you can quickly prove it. Define the parametric curve $\gamma : [-2, 2] \to \mathbb{R}^2$ as $\gamma(t) = (t, t^2)$ which yields a trace $\gamma([-2, 2]) = C$. Since γ is continuous and $\gamma([-2, 2]) = C$, it follows that C is a curve.

Example 1.1.15 Define the set

$$U = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$

so U is the unit circle in $\mathbb{R}^2$. You have already seen the numerous ways to trace U in Example 1.1.9. Since all of those parametrizations are continuous then any of them would satisfy our definition. Therefore, U is a curve.

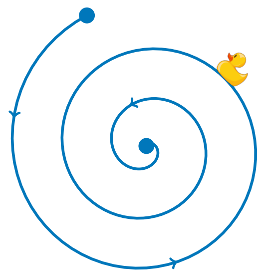
This brief foray into curves hopefully gives you a taste of the many interesting applications, questions, and problems to come!

[3] Continuity has not yet been defined. As usual for this chapter, focus on intuition and worry about details later.

Exercises for Section 1.1

Concepts and definitions

1.1.1 Belal wants to study how whirlpools work. He takes his rubber duck to a swimming pool with a slow-moving whirlpool. He places the duck on the edge of the whirlpool and, using his drone, he tracks its movement. It takes 2 minutes until the rubber duck reaches the centre. Afterwards, he downloads the video data to obtain a function $\gamma(t)$ which is the position of the duck t seconds after he placed it in the whirlpool. All distances are in metres.



(a) What should be the domain of γ?

(b) What should be the codomain of γ?

(c) If Belal plays the video data at double speed, how would the range of γ change?

1.1.2 Let $\gamma : [0,15] \to \mathbb{R}^3$ be the position of Maryam as she rides her bike from home to work. Time is measured in minutes, and distances are measured in kilometres. For each physical description, match the corresponding expression(s).

(a) The distance between Maryam's position at 6 minutes and her position at 6.1 minutes.

$$\gamma(6.1)-\gamma(6) \qquad \gamma(6)-\gamma(6.1) \qquad ||\gamma(6.1)-\gamma(6)|| \qquad ||\gamma(6.1)||-||\gamma(6)||$$

(b) The displacement from Maryam's position at 6 minutes to her position at 6.1 minutes.

$$\gamma(6.1)-\gamma(6) \qquad \gamma(6)-\gamma(6.1) \qquad ||\gamma(6.1)-\gamma(6)|| \qquad ||\gamma(6.1)||-||\gamma(6)||$$

(c) Maryam's average velocity over the time interval $[6, 6.1]$.

$$\frac{\gamma(6.1)-\gamma(6)}{0.1} \qquad \frac{\gamma(6)-\gamma(6.1)}{-0.1} \qquad \frac{||\gamma(6.1)-\gamma(6)||}{0.1} \qquad \frac{||\gamma(6)-\gamma(6.1)||}{-0.1}$$

(d) Maryam's average speed over the time interval $[5.9, 6]$.

$$\frac{\gamma(6)-\gamma(5.9)}{0.1} \qquad \frac{\gamma(5.9)-\gamma(6)}{-0.1} \qquad \frac{||\gamma(6)-\gamma(5.9)||}{0.1} \qquad \frac{||\gamma(5.9)-\gamma(6)||}{-0.1}$$

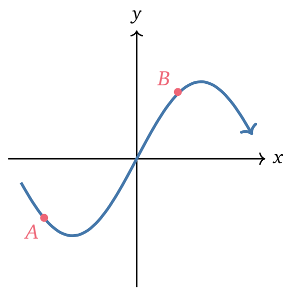
(e) Maryam's instantaneous velocity at 6 minutes.

$$\lim_{h\to 0}\frac{\gamma(6+h)-\gamma(6)}{h} \qquad \lim_{h\to 0}\frac{||\gamma(6+h)-\gamma(6)||}{h} \qquad \lim_{h\to 0}\frac{||\gamma(6+h)-\gamma(6)||}{|h|} \qquad \lim_{h\to 0}\left|\left|\frac{\gamma(6+h)-\gamma(6)}{h}\right|\right|$$

(f) Maryam's instantaneous speed at 6 minutes.

$$\lim_{h\to 0}\frac{\gamma(6+h)-\gamma(6)}{h} \qquad \lim_{h\to 0}\frac{||\gamma(6+h)-\gamma(6)||}{h} \qquad \lim_{h\to 0}\frac{||\gamma(6+h)-\gamma(6)||}{|h|} \qquad \lim_{h\to 0}\left|\left|\frac{\gamma(6+h)-\gamma(6)}{h}\right|\right|$$

1.1.3 Here you can develop your geometric intuition with motion in two-dimensions. A particle is moving along the drawn path.

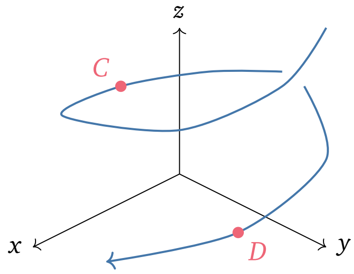


(a) Sketch the unit tangent and unit normal at point A.

(b) Sketch the unit tangent and unit normal at point B.

(c) The unit tangent and unit normal may not actually be defined at A and B. Why not?

1.1.4 Next, you can practice your geometric intuition with motion in three-dimensions. A particle is moving along the path shown on the righthand side. Sketch the Frenet frame $\{T, N, B\}$ at points C and D, that is, the unit tangent, unit normal, and unit binormal.



Computations

1.1.5 Each of the following are parameterizations of a circle. Describe their motion in plain terms. Sketch each of them along with a typical velocity vector and acceleration vector. Calculate their velocity and acceleration.

(a) $\gamma_1(t) = (\cos t, \sin t)$ for $0 \le t \le 2\pi$

(b) $\gamma_2(t) = (3\sin(2t), 3\cos(2t))$ for $0 \le t \le 2\pi$

(c) $\gamma_3(t) = 3\gamma_1(-2t)$

1.1.6 Define two different parametrized curves in $\mathbb{R}^2$ which go from $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ in a straight line.

1.1.7 A particle moves in $\mathbb{R}^3$ along the path $\gamma(t) = (\frac{1}{2}t^2 + 3t, 2t + 6, t - 2)$ where t measures time in seconds and distance is in meters.

(a) Calculate the particle's velocity, acceleration, and speed as functions of time.

(b) Calculate the particle's unit tangent, unit normal, and unit binormal as functions of time.

Applications and beyond

1.1.8 Here you will model the motion of the Earth, Moon, and Sun. To get a basic feel for any phenomenon, your first model should be really simple. Let's assume all orbits are circular, all orbits lie in the same 2-dimensional plane and all objects are moving at a constant speed. Explain any choices you must make or constants you must introduce.

(a) Assume the Earth orbits the Sun counterclockwise. Define a function $\gamma_1 : \mathbb{R} \to \mathbb{R}^2$ describing this orbit with the sun at the origin.

(b) Define a function $\gamma_2 : \mathbb{R} \to \mathbb{R}^2$ describing the orbit of the Moon with the Earth at the origin. Ensure the Moon orbits around the Earth in a way that is consistent with the assumption that the Earth is orbiting the Sun counterclockwise.

(c) Define a function $\gamma_3 : \mathbb{R} \to \mathbb{R}^2$ describing the orbit of the Moon around the Sun. Sketch this orbit using Desmos and compare it with the orbit of the Earth around the Sun.

1.1.9 You are driving up a spiral parking ramp in Eaton's centre (see figure [4]) at uniform speed. It takes you about 90 seconds to drive from the bottom floor to the top. Give a precise mathematical function which models the motion of your car as a function from the time it starts at the bottom of the ramp until it reaches the top floor. Explain any choices you must make or any constants you must introduce.

Hint: Assume the ramp makes 4 full loops, which takes you 90 seconds to drive up. What else do you need to know to model your location on the ramp at a given time?

[4] Image retrieved from Wikimedia Commons on 2024-07-30 licensed under CC BY-SA

1.2. Real-valued functions

This section is about maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}$$

usually for $n \geq 2$. In mathematics, the full names of these maps are usually **real-valued functions of n variables**, **multivariable real-valued functions**, or **real-valued functions of several real variables**. The common shorthand is simply **real-valued functions**.

Example 1.2.1 Here are some algebraic examples of real-valued functions.

- A multivariable polynomial such as $f(x,y) = x^2 + 3xy - y^2 - y^4$.
- A piecewise function such as

$$g(x,y,z) = \begin{cases} \frac{xyze^{xyz}}{x^2+y^2+z^2} & \text{if } (x,y,z) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- The norm function $N : \mathbb{R}^n \to \mathbb{R}$ defined by

$$N(x_1, \ldots, x_n) = ||x|| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

1.2.1 Scalar fields and densities

In physics, real-valued functions are called **scalar fields** or **scalar functions** or **potentials**.

Example 1.2.2 Meteorologists use temperature to help predict the weather as it can allude to cloud formation and the movement of huge climate systems. Temperature itself is a scalar value (measured in Celsius) and varies depending on where you are on Earth. At any given moment, the temperature $T(x,y,z)$ depends on your position (x,y,z) on Earth. For example, suppose (x_A, y_A, z_A) is a position in the Arctic and (x_D, y_D, z_D) is a position in the Sahara Desert. You might guess that $T(x_A, y_A, z_A) < T(x_D, y_D, z_D)$ since the Sahara Desert is always much hotter than the Arctic.

Example 1.2.3 Mass is never distributed uniformly among objects because there are always more dense regions and less dense regions in a mass. For example, a block of swiss cheese has regions where there is no mass (holes) and regions where cheese is tightly packed. Since the density of the cheese depends on where in the cheese you are looking then we can describe the density with a scalar field $\varphi : C \to [0, \infty)$ where $C \subseteq \mathbb{R}^3$ is the set of points in the cheese. Our scalar field $\varphi(x,y,z)$ outputs the density in units $\mathrm{kg/m^3}$ at (x,y,z). For instance, suppose $p \in C$ is a point inside a hole. Then one would expect $\varphi(p) \approx 0\,\mathrm{kg/m^3}$ as there's no mass in the hole.

Example 1.2.4 There are forces in physics that have a special connection with potentials, namely *conservative forces*. For example, the electrostatic force for a point charge is related to the scalar field $V : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$V(x,y,z) = \frac{Q}{4\pi\varepsilon_0\sqrt{x^2+y^2+z^2}}.$$

The function V is often referred to as the electric potential where ε_0 is the permittivity of free space constant and Q is the charge constant. As you will see much later in vector calculus, this function can describe how the electrostatic forces of two different charges influence each other as they move nearby.

Real-valued functions arise in many other fields of study, too.

Example 1.2.5 Economists and businesses strive to maximize profit or minimize costs subject to many constraints. They must account for many parameters before making decisions resulting in real-valued functions being highly important. For example, suppose you are CEO of a company named CHAYR and must produce 20,000 chairs. The number of chairs they can produce is given by the Cobb–Douglas function:

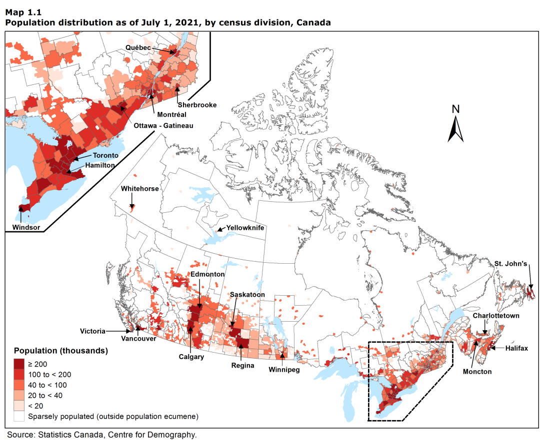
$$P(K,L) = \frac{1}{25}K^{1/4}L^{1/3}$$

where K is their capital expenditure and L is their labour costs. The total cost is therefore $C(K,L) = K + L$. You must decide how to spend your money the minimizes costs and still produces 20,000 chairs. In other words, you must minimize $C(K,L)$ subject to the constraint that $P(K,L) = 20,000$. You will learn how to solve such multivariable optimization problems.

Example 1.2.6 A streaming site FLYX uses very complicated algorithms to find content to recommend to you. Every movie or show you watch generates data points that FLYX stores and uses to associate a rough categorization of the type of viewer you are. An example of such a data point would be how many hours you watched a specific genre/style. Then, before *Netflix* recommends you a movie or show, it evaluates the data, represented in n variables $x_1,\ldots,x_n$, using some function $E:\mathbb{R}^n \to \mathbb{R}$ that outputs a score $E(x_1,\ldots,x_n)$. FLYX recommends the shows with the highest scores because they believe you will enjoy the movie or show. But how did FLYX create the magical real-valued function E? Real-valued functions are very important for data analysis.

Real-valued functions which are non-negative are special and often referred to as **densities**. Often, you will simply write $f \geq 0$ to mean that the real-valued function f is non-negative. Many densities are defined by counting a quantity and dividing by a unit of measurement. For instance, population density is the number of people divided by unit area.

Example 1.2.7 Population density, the number of people per unit area, of a country is valuable information for societal statistics as well as future development. Let $C \subseteq \mathbb{R}^2$ be the set of points in Canada (ideally we would need 3 variables, but let's assume C is the set of points on a 2D map of Canada). If $\varphi : C \to [0,\infty)$ is Canada's population density function measured in persons per square kilometre, then $\varphi(x,y)$ should be approximately the number of people in a 1 km by 1 km square centred at (x,y). Below is a heatmap of φ. [5]

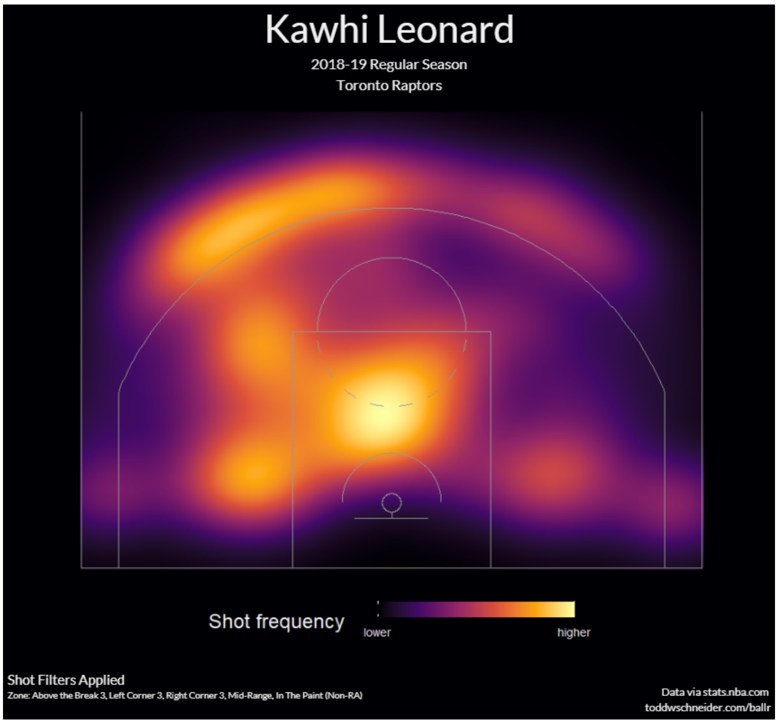


As one would expect, the GTA (Greater Toronto Area) is quite red implying high population density because of the limited space and large number of people. Northern regions like Nunavut remain less shaded (low population density) due to the large amount of land, but few inhabitants.

Example 1.2.8 Charge density describes the amount of charges (electrons/coulombs) per unit of measurement. The unit of measurement depends on the situation. Length is commonly used for current flowing through a wire. Area is best for current flowing through a conducting sheet. Volume is generally used for any scenario in $\mathbb{R}^3$.

Probability is also a great source of density functions.

Example 1.2.9 Sports analytics now includes much more sophisticated probabilistic methods to study games like basketball. Kawhi Leonard, a famous basketball player, had his shot frequency analyzed over his championship winning 2018–19 season.[6]



From this data, you can define Kawhi Leonard's probability density function $\varphi : C \to [0, \infty)$ for taking a shot anywhere on the court $C \subseteq \mathbb{R}^2$. What is the probability that Leonard takes a shot from beyond the 3-point line? It should be the integral of φ over the region T beyond the 3-point line. Analyzing this data occupies the minds of many sports executives.

Not all densities can be described by counting. For example, energy density describes the energy (joules) per unit volume (meters cubed). This quantity is mainly used in electromagnetism where energy can be stored in electric and magnetic fields. From a macroscopic viewpoint, this energy is not something you can count because it varies continuously in space.

Example 1.2.10 How do you define densities for things you cannot count? Think of a block of cheese again. If the mass of the entire cheese was *uniformly* distributed across the region

[5] Adapted from Statistics Canada, Annual Demographic Estimates: Subprovincial Areas, July 1, 2021, January 13, 2022. This does not constitute an endorsement by Statistics Canada of this product.
[6] Image created with code modified from Schneider licensed under MIT.

then then its density ρ is

$$\rho = \frac{m}{V},$$

where m is the total mass of the region and V is the total volume of the region. But you bought a fancy cheese with varying density, so ρ is actually the *average* mass density.

So what is the density $\varphi(x, y, z)$ at a specific point (x, y, z) of cheese? This question is strange since a point has zero volume and zero mass. To make sense of "density at a point", you need a limit. Imagine looking at our non-uniformly mass distributed cheese, but very zoomed in. Chop off a very small piece, for example. The variation is less apparent and the piece of cheese seems more uniform. The average density of the small piece of cheese can be used to approximate the density function itself! This approximation becomes exact if you take the limit as volume goes to zero. A bit more formally, if $C_\varepsilon(x, y, z) \subseteq \mathbb{R}^3$ is the cube of sidelength $\varepsilon > 0$ centred at (x, y, z) then

$$\varphi(x, y, z) = \lim_{\varepsilon \to 0^+} \frac{\mathrm{mass}\big(C_\varepsilon(x, y, z)\big)}{\mathrm{volume}\big(C_\varepsilon(x, y, z)\big)}.$$

This story will continue much further when studying integration in $\mathbb{R}^n$.

1.2.2 Graphs, level sets, and slices

There are many ways to visualize multivariable functions and you may need to relate these visuals. These are formally defined as sets in $\mathbb{R}^n$. The most basic is a graph.

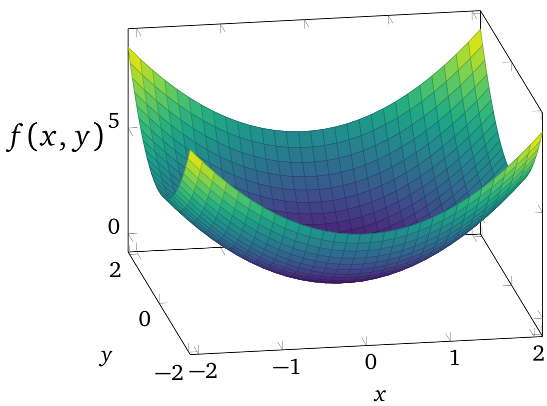
Definition 1.2.11 Let $A \subseteq \mathbb{R}^n$. The **graph of a function** $f : A \to \mathbb{R}$ is the set in $\mathbb{R}^{n+1}$ given by

$$\{(x, f(x)) : x \in A\}.$$

Remark 1.2.12 Notice $(x, f(x))$ belongs to $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$ since $x \in \mathbb{R}^n$ and $f(x) \in \mathbb{R}$.

These sets are the main tool for visualizing a multivariable function f.

Example 1.2.13 Define the function $g : [-1, 1] \to \mathbb{R}$ as $g(x) = x^3$. The graph of g is the set $\{(x, x^3) : x \in [-1, 1]\}$, which you can view on Desmos. Notice the graph lies in $\mathbb{R}^2$ and f is a single-variable real-valued function. What about a two-variable real-valued function? Its graph is also called a **surface plot**. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2$. The graph of f is therefore $\{(x, y, x^2 + y^2) : (x, y) \in \mathbb{R}^2\}$, which is plotted below. This specific surface is referred to as a "paraboloid".



What about three-variable real-valued functions? Consider the function $h : \mathbb{R}^3 \to \mathbb{R}$ given by $h(x,y,z) = xyz$ so its graph is the set $\{(x,y,z,xyz) : (x,y,z) \in \mathbb{R}^3\}$ lying in $\mathbb{R}^4$. Since the graph exists in four-dimensional space, you cannot directly visualize this set as a plot. You can only plot graphs of functions with one or two variables for input.

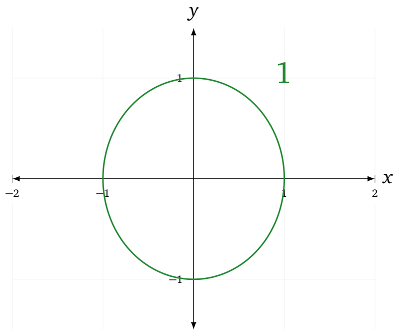
It is often helpful to "reduce dimensions" when trying to visualize a graph or prove a lemma. There are several ways of doing so.

Definition 1.2.14 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$ be a real-valued function. Fix $k \in \mathbb{R}$. The **level set** of f at k is the set $\{x \in \mathbb{R}^n : f(x) = k\}$. This is also called the k-**level set**.

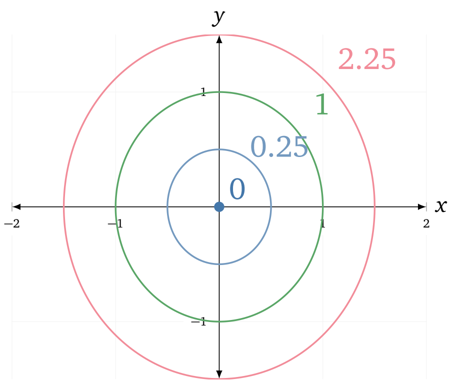
Remark 1.2.15 A level set in $\mathbb{R}^2$ is also called a **contour**.

For graphs of 2-variable functions, you can create a **contour plot** by plotting the level sets for a few different values.

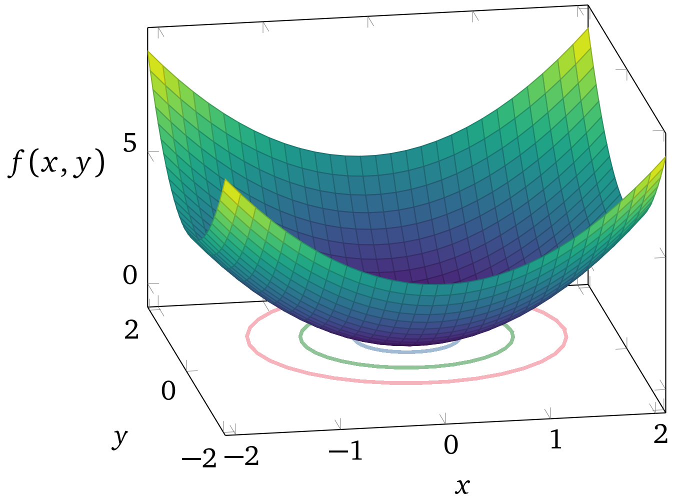
Example 1.2.16 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ as $f(x,y) = x^2 + y^2$. You can visualize f and its graph using only its level sets. Begin with the 1–level set, 0–level set and (-1)–level set. By definition, the 1–level set of f is the set $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, which is the unit circle in $\mathbb{R}^2$. This single contour is plotted below.



The "1" indicates that for any (x,y) on this contour, $f(x,y) = 1$. The 0–level set is given by $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 0\} = \{(0,0)\}$, since only the origin satisfies $x^2 + y^2 = 0$. Hence, this contour is a single point. The (-1)–level set is the set $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = -1\} = \varnothing$, since there are no points (x,y) satisfying $x^2 + y^2 = -1$. Hence, this contour is empty. By plotting a few more contours, you obtain a contour plot. Here each contour is coloured differently, but that is not always done.

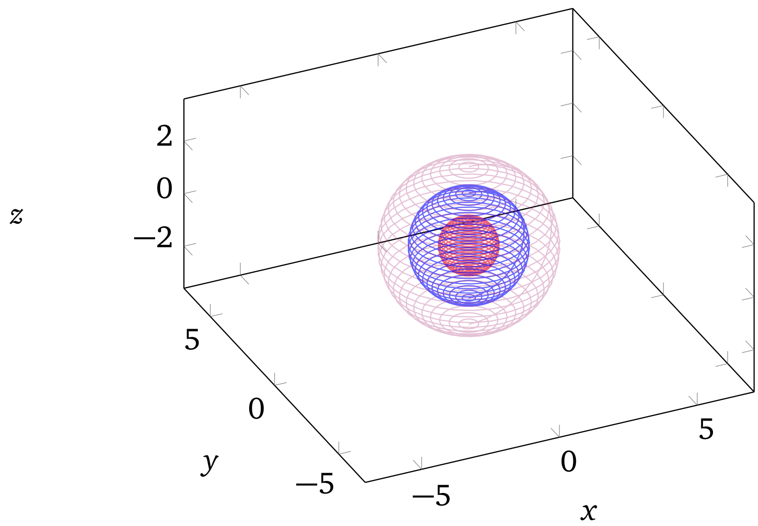


How does this contour plot correspond to the graph of f? Imagine raising each contour by their value out of the page. For example, the 0 point remains on the page (the xy-plane) and the 0.5 circle raises out of the page by 0.5. This recreates a "skeleton" graph of f, as illustrated below.



Play with this Math3D demo to see how the contours correspond to the graph of f. Toggle switches in order. Also, watch this Math3D demo to see how the k-level set relates to the graph of f as k varies from -1 to 4.

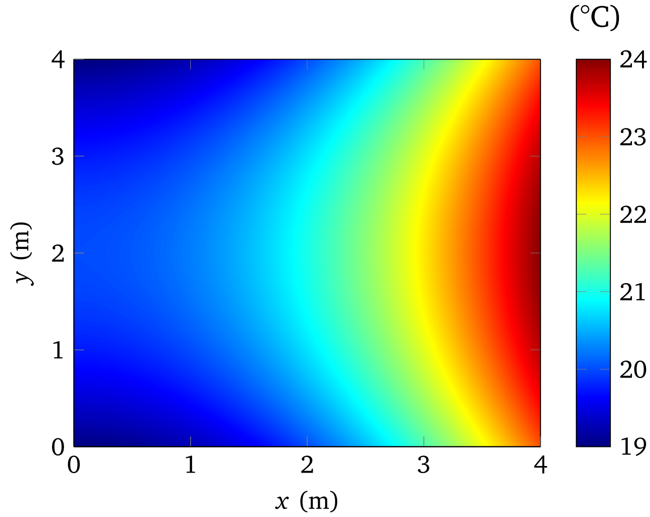
Example 1.2.17 What about the level sets of a 3-variable function? Define the function $f(x, y, z) = x^2 + y^2 + z^2$. The graph of f lies in $\mathbb{R}^4$ and cannot be plotted, but its level sets lie in $\mathbb{R}^3$ so you can plot them. Notice that the k-level set of f is just a sphere of radius $\sqrt{k}$ for $k \geq 0$. Below are a of couple level sets drawn on the same plot.



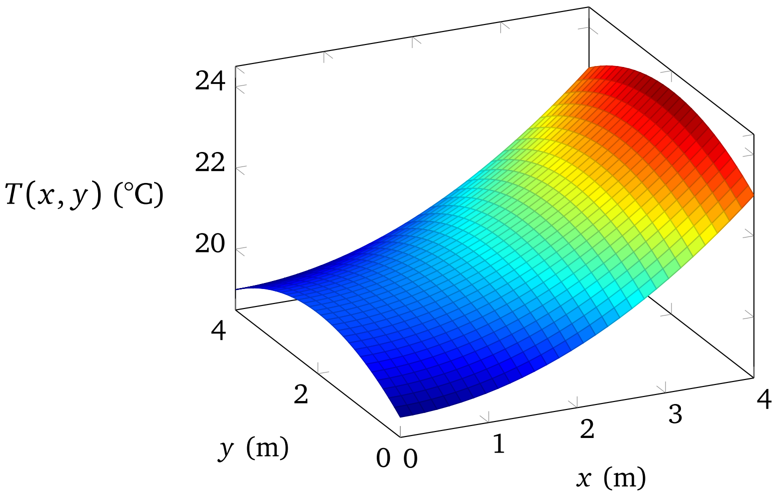
The innermost red sphere is the 1-level set, the middle blue sphere is the 4-level set, and the outermost pink sphere is the 9-level set. Three-dimensional level sets may seem abstract and not that helpful at a glance; however, they can offer valuable interpretations. For example, if f describes the magnitude of force felt by a mass at position (x, y, z), then the k-level set describes all positions in three-dimensional space that a mass would feel a force of k.

Another way of “reducing dimensions” for 2-variable functions is to use a colour gradient that corresponds to the values of the function f. These are called **heat maps**. These are like a continuous version of a contour plot.

Example 1.2.18 You are sitting in a 4 m by 4 m room with a window. The sun radiates heat through the window increasing the temperature of the room. You model the temperature in Celsius using the function $T : [0,4]^2 \to [0,\infty)$ defined as $T(x,y) = 0.25(x^2 - (y-2)^2) + 20$. The input, x and y, are measured in meters and describe your position in the room. To visualize T on its domain, you can create a heat map:



The highest temperature, indicated by the red, is in the vicinity of the window as expected. You can see how this corresponds to the actual graph of T:



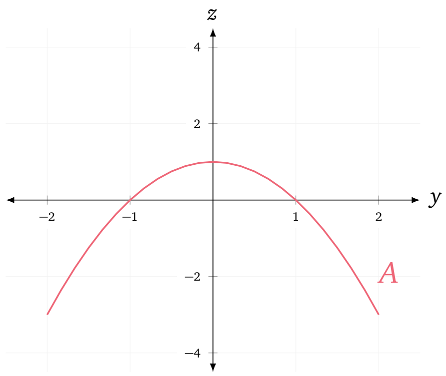
Yet another way of "reducing dimension" is **slicing**, or equivalently fixing a variable. Here is the definition for 2-variable functions.

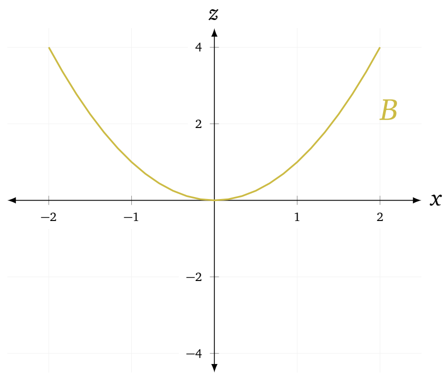
Definition 1.2.19 Let $A \subseteq \mathbb{R}^2$ and $f : A \to \mathbb{R}$ be a real-valued function.

- For fixed $a \in \mathbb{R}$, the **x-slice at a of the graph of f** is the set
$$\{(y,z) \in \mathbb{R}^2 : (a,y) \in A,\ z = f(a,y)\}.$$
- For fixed $b \in \mathbb{R}$, the **y-slice at b of the graph of f** is the set
$$\{(x,z) \in \mathbb{R}^2 : (x,b) \in A,\ z = f(x,b)\}.$$

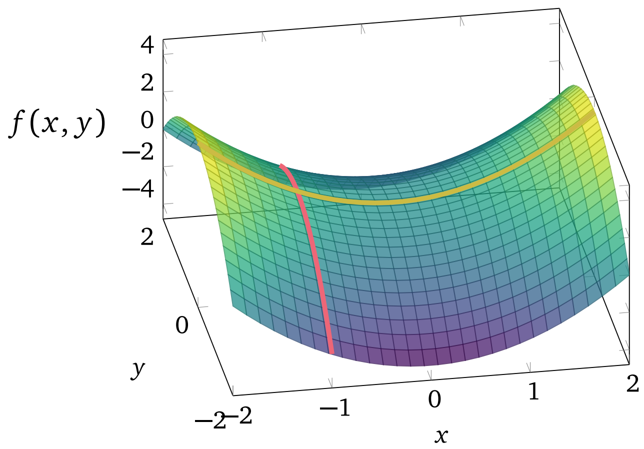
The phrase "slicing" comes from viewing these sets as slices of the graph of f.

Example 1.2.20 Here you will plot the graph of $f(x,y) = x^2 - y^2$. The x-slice at $x = -1$ is the set $A = \{(y,z) : z = 1 - y^2\}$ and the y-slice at $y = 0$ is the set $B = \{(x,z) : z = x^2\}$. Notice both A and B are sets in $\mathbb{R}^2$. These are plotted separately below.





How do these 2D slices correspond to the graph of f? You can visualize them below.



Note the illustration above shows sets in $\mathbb{R}^3$ which correspond to the slices A and B lying in $\mathbb{R}^2$. Play with this Math3D demo to view different slices.

You similarly define slices for 3-variable functions.

Definition 1.2.21 Let $A \subseteq \mathbb{R}^3$ and $f : A \to \mathbb{R}$ be a real-valued function.

- For fixed $a \in \mathbb{R}$, the **x-slice at a of the graph of f** is the set
$$\{(y,z,w) \in \mathbb{R}^3 : (a,y,z) \in A,\ w = f(a,y,z)\}.$$
- For fixed $b \in \mathbb{R}$, the **y-slice at b of the graph of f** is the set
$$\{(x,z,w) \in \mathbb{R}^3 : (x,b,z) \in A,\ w = f(x,b,z)\}.$$
- For fixed $c \in \mathbb{R}$, the **z-slice at c of the graph of f** is the set
$$\{(x,y,w) \in \mathbb{R}^3 : (x,y,c) \in A,\ w = f(x,y,c)\}.$$

These slices are sets in $\mathbb{R}^3$ so, like Example 1.2.17, you can plot them. These definitions continue to generalize to higher dimensions, but this will be enough for now. This illustrates many ways to visualize real-valued functions of two or three variables. You should become familiar with all of them and their relationships.

Exercises for Section 1.2

Concepts and definitions

1.2.1 Let $f(x,y) = e^{-x^2-y^2}$.

(a) Sketch the level sets of f at $4, 1, 1/4$ and 0 on a single plot.

(b) Confirm your plot is correct by using Desmos.

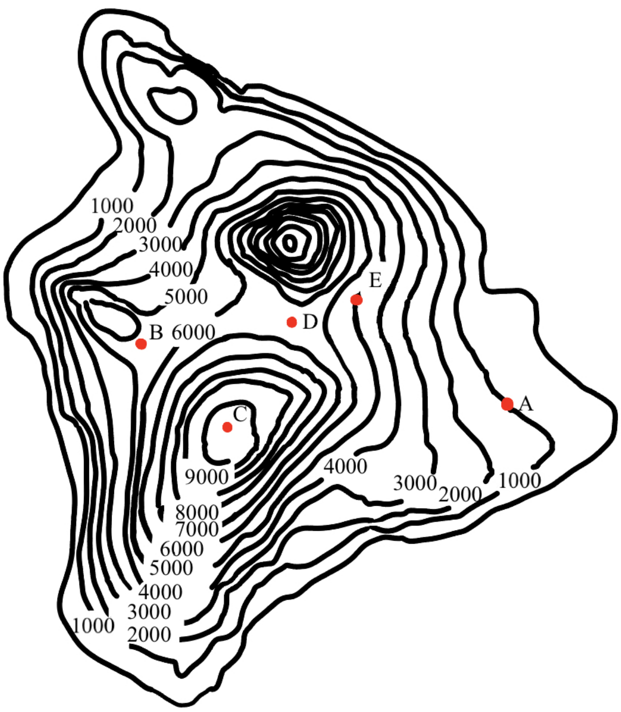
(c) Plot the graph of f using Math3D.

1.2.2 Graphing software[7] is an incredible asset in multivariable calculus. You should use it frequently to explore any questions that you encounter. Now, consider the graph of $f(x,y) = x^2 - y^2$ which can be viewed at https://www.math3d.org/6nylL0VA. It takes a moment to load. Ignore any specific equations that you see.

(a) Describe the yellow line as a level set, x-slice, or y-slice of f. Write the formal set.

(b) Describe the red line as a level set, x-slice, or y-slice of f. Write the formal set.

1.2.3 A topographic map of the island of Hawaii is below. Contour plot values are measured in feet.



Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a function describing the height of the island at a point

(a) On which level set of h does the point A belong?

(b) What can you deduce about $h(B)$ from the diagram?

(c) You are holding a boulder steady at point E. When you let go, you watch the boulder roll down the mountain. Sketch what path it takes on the map.

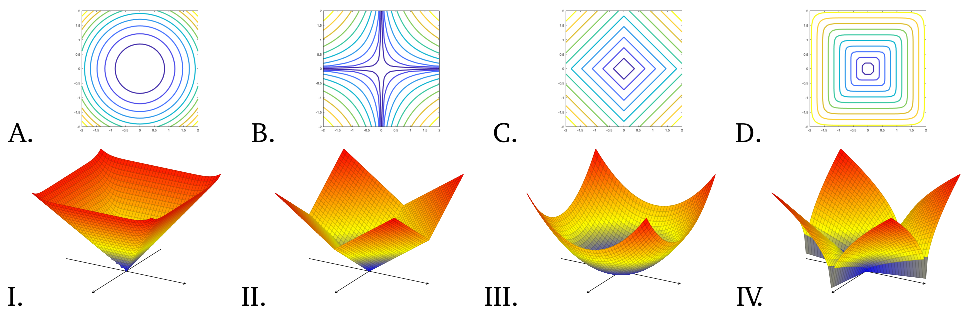
[7] Desmos and Math3D are quick-to-use browser calculators. Octave online is a more powerful graphing calculator and requires a bit more time to figure out. There are plenty of examples to copy-and-paste so you can get started quickly. It is based on the scientific computing software Octave.

(d) Identify on the map where the local extrema for h exist.

(e) Which spot on the island has the steepest ascent? Explain why.

(f) The diagram actually has an error. Find the error and explain what is wrong.
Hint: Look carefully at the contour on which E lies.

1.2.4 A **contour plot** for a function $f : \mathbb{R}^2 \to \mathbb{R}$ is a finite collection of level sets plotted on a single plane. Match the contour plot to the graph of the function.



1.2.5 The definitions of graphs, level sets, and slices exists for functions $f : \mathbb{R}^3 \to \mathbb{R}$ of three variables. Since it's similar to the two variable case, it's your job to create them.

(a) Define the graph of f

(b) Define a level set of f.

(c) Define a x-slice of f, a y-slice of f, and a z-slice of f.

(d) Which of these sets can be plotted in the 2D plane? And in 3D space?

1.2.6 Now, create definitions of graphs, level sets, and slices for functions $f : \mathbb{R}^n \to \mathbb{R}$ of n-variables.

(a) Define the graph of f.

(b) Define a level set of f.

(c) Fix $i \in \{1, 2, \dots, n\}$. Define an x_i-slice of f.

(d) For $n \geq 4$, which of these sets can be plotted in three dimensional space?

Proofs

1.2.7 Level sets can be weird.

(a) Construct a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that its level set at 0 is equal to $B = \{(x, -x) : x \in \mathbb{R}\}$.
Do not prove your claim.

(b) Construct a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that its level set at 0 is equal to $A \cup B$ where

$$A = \{(x, x) : x \in \mathbb{R}\}, \qquad B = \{(x, -x) : x \in \mathbb{R}\}.$$

Notice $A \cup B$ is the union of two intersecting lines. Prove your example is valid.

Applications and beyond

1.2.8 Functions $f : \mathbb{R}^n \to [0, \infty)$ with non-negative values have a special interpretation as **density**. There are two intuitions for the concept of density. The physical intuition is:

> *The mass of an object $S \subseteq \mathbb{R}^n$ is equal to the volume of S times the average value of its density $f : \mathbb{R}^n \to [0, \infty)$ over S.*

These terms do not have a rigorous definition yet but you can still use this intuition. Aang's body is described by a set $A \subseteq \mathbb{R}^3$. Let $\varphi : A \to [0, \infty)$ be Aang's (mass) density function.

(a) Specify units for φ.

(b) Is φ constant? If not, suggest regions where you expect it is roughly constant.

(c) Approximate the maximum and minimum values for φ. Explain where they should occur.

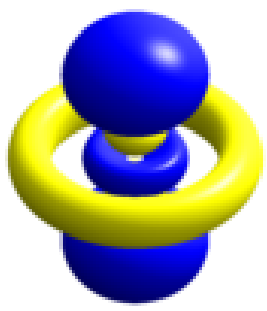
(d) Express the surface of his skin as a set using φ.

1.2.9 Densities $f : \mathbb{R}^n \to [0, \infty)$ have a second intuition via probability. Probability applies in all kinds of scenarios but let's use a specific example to keep things concrete. Quantum mechanics is founded on the idea that the position of an electron is a probabilistic notion.

> *The probability that an electron lies in a bounded region $S \subseteq \mathbb{R}^3$ is the volume of S times the average value of the electron's* ***probability density function*** *$f : \mathbb{R}^3 \to [0, \infty)$ over S.*

Again, nothing here is rigorous but you can always use this to guide your thinking.

(a) These are two images modeling an electron's position in a hydrogen orbital [8]. Ignore the colours. One interpretation is non-probabilistic (i.e. classical) and the other is probabilistic (i.e. quantum). Which one uses a probability density function $f : \mathbb{R}^3 \to [0, \infty)$?



(b) There seems to be 6 basic regions where the electron can lie. You want to determine the probability that the electron lies in one of those six regions. What features of the diagram would you use to compare the probabilities for each region? *Hint:* There are two such features.

(c) If the maximum of f occurs at $p \in \mathbb{R}^3$, what is the probability that the electron is at point p?

[8] Images retrieved from Wikimedia Commons (hydrogen eigenstate and atomic orbital cloud) on 2024-07-23 licensed under CC BY-SA.

1.3. Vector fields

This section is about maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^n.$$

This is a special case since the dimension of the domain and codomain are the same. There are two major interpretations of these maps. Your viewpoint will depend on the context and here you will focus on vector fields. As usual, the cases $n = 2$ and $n = 3$ dominate the discussion because they can be reasonably visualized and have natural physical analogies. Your understanding of those scenarios will strongly motivate how you think about higher dimensions.

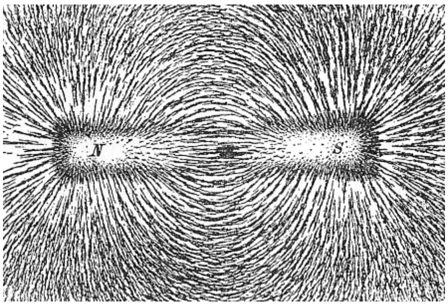
Definition 1.3.1 A **(n-dimensional) vector field** is a function F with domain and codomain lying in $\mathbb{R}^n$.

You may also say F is a **vector field in** $\mathbb{R}^n$. Heuristically, this viewpoint says:

$F(x) \in \mathbb{R}^n$ *is a vector at the point* $x \in \mathbb{R}^n$.

The name "vector field" is inspired by physical examples and it is common practice to use capital letters for vector fields. The vector $F(x)$ often represents the velocity of a fluid (or a force) at the point x so it is sometimes called **velocity field** or **force field** in those contexts.

Example 1.3.2 The ocean is an example of a (time-dependent) vector field; at a given moment in time, each point x in the ocean has a velocity $F(x)$. The same is true of atmospheric winds and weather patterns, like a hurricane[9].

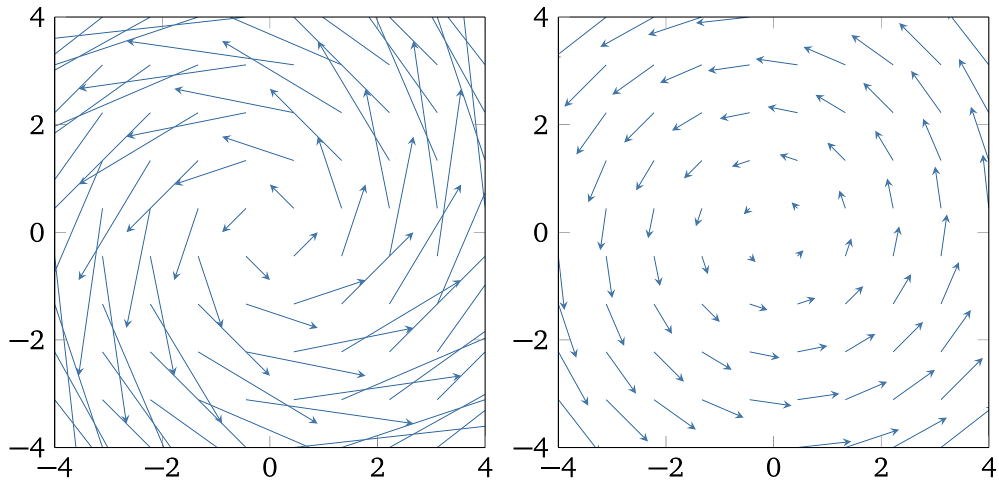


The magnetic field generated by a magnet is another example of a vector field; each point x is influenced by a force $F(x)$ imposed by the magnet. The same is true of the gravitational force fields, like planet Earth.

Vector fields can be visualized by assigning a vector $F(x) \in \mathbb{R}^n$ to each point x in the domain. As usual, this can be plotted for vector fields in $\mathbb{R}^2$ or $\mathbb{R}^3$ only.

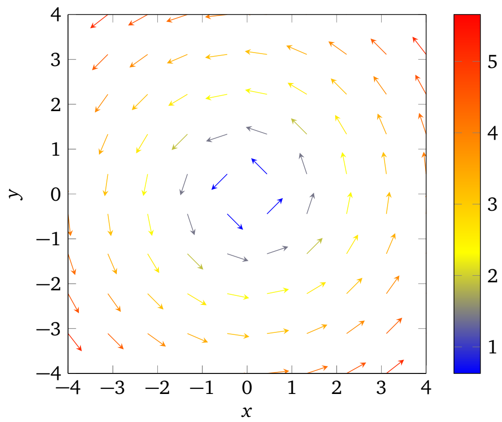
Example 1.3.3 How do you plot a two-dimensional vector field? Define the vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ by $F(x, y) = (-y, x)$. The point $(1, 2)$ has associated vector $F(1, 2) = (-2, 1)$ so at the point $(1, 2)$, you can draw the vector $(-2, 1)$. By doing this process for many points on a grid, you can produce a **vector field plot** like the one below on the left.

[9] Images retrieved from Wikimedia Commons (hurricane and magnet) on 2024-07-23 licensed under PD.



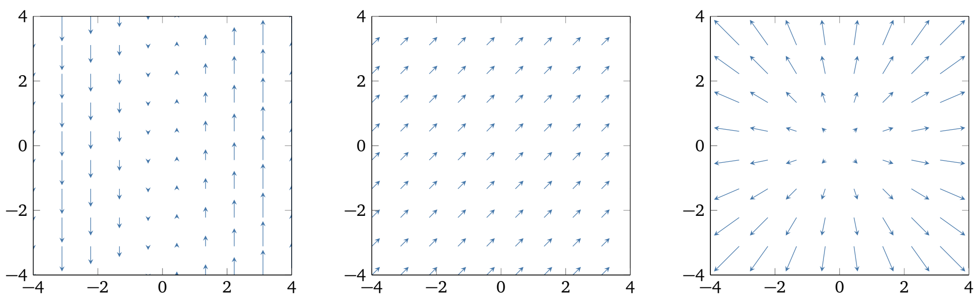
The picture on the left is messy because the vectors are plotted *according to scale*. For example, the vector $(-2, 1)$ is really drawn to scale with length $\sqrt{5}$. To avoid this ugly appearance, it is convention to proportionally rescale the vectors so they are smaller. Rescaling vectors to 25% of their original size produces the plot on the right. That's much better! By comparing the *relative* size of vectors, you can still see where the "fluid" is moving faster.

Alternatively, you can scale all the vectors to be the *same size* and use a colour gradient to represent the magnitude. One possibility is to colour the arrows according to their magnitude. A common convention is that "hot" arrows have large magnitude and "cold" arrows have small magnitude, but you should always include a colour bar. This produces the plot below.



Vector fields with different equations can illustrate very different phenomena.

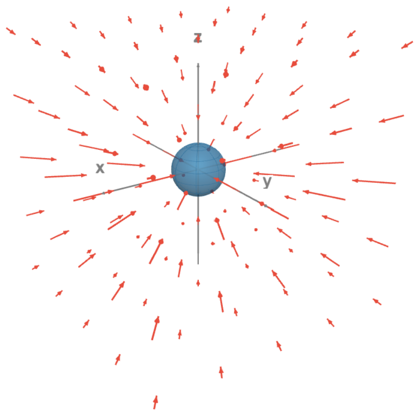
Example 1.3.4 Define $F(x, y) = (x, y)$, $G(x, y) = (0, x)$, and $H(x, y) = (1, 1)$. Each of these vector fields are plotted below.



Which vector fields correspond to which plot? You can check by simply evaluating them at some points but it is better to associate some geometric properties. Notice H is constant so the vector field should be constant. Notice G does not depend on y. Notice the magnitude of F grows larger as (x, y) moves away from the origin. This Geogebra demo gives a simple way of plotting 2D vector fields, so you can confirm your guesses with it.

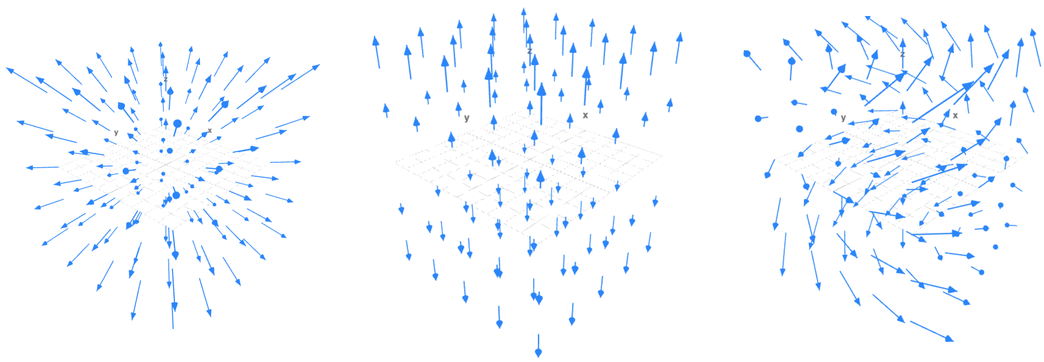
Three-dimensional vector fields can be visualized in a similar fashion.

Example 1.3.5 The gravitational force field for Earth is an example of a vector field in $\mathbb{R}^3$.



Notice the vectors are coloured according to the magnitude of their force.

Example 1.3.6 Sketching 3D vector fields is much more difficult, so you will usually use software to do so. Plots for $F(x,y,z) = (-y,x,z), G(x,y,z) = (x,y,z)$, and $H(x,y,z) = (0,0,z)$ are below.



Which vector fields correspond to which plot? Use this Math3D demo to confirm your guess. Remember to try and determine some geometric properties from the equations.

Notation

You should be familiar with some equivalent notation for vector fields, such as

$$F(x,y,z) = (x^2, yx, -z), \qquad F(x,y,z) = \langle x^2, yx, -z\rangle, \qquad F = [x^2, yx, -z], \qquad F = x^2\hat{\imath} + yx\hat{\jmath} - z\hat{k}.$$

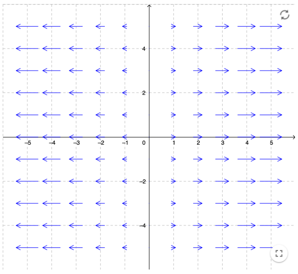
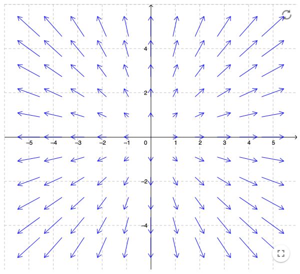
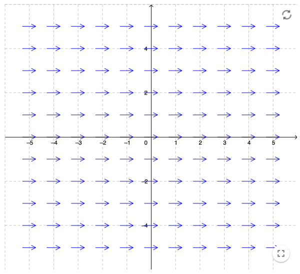
In this text, you will see the leftmost notation as the standard convention.

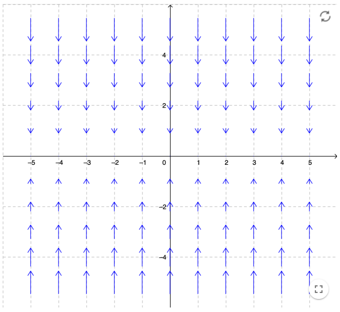
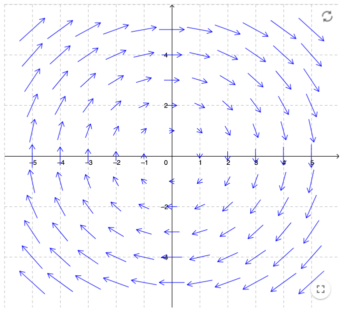
Exercises for Section 1.3

Concepts and definitions

1.3.1 Match the picture of the vector field to its algebraic definition.

$A(x,y)=(x,0)$ $\quad B(x,y)=(0,-y)$ $\quad C(x,y)=(3,0)$ $\quad D(x,y)=(x,y)$ $\quad E(x,y)=(y,-x)$

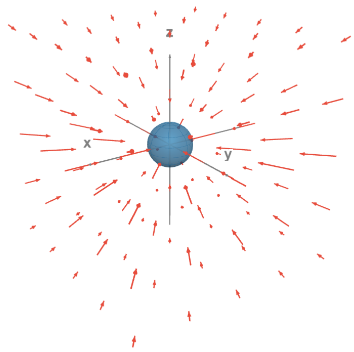
I. 
II. 
III. 

IV. 
V. 

1.3.2 Gravitational fields are described using vector fields $F:\mathbb{R}^3\to\mathbb{R}^3$ since a vector can represent force. Newton's law of gravitation states that

$$F(x,y,z)=\frac{-Gm_1m_2}{||(x,y,z)||^2}\cdot\frac{(x,y,z)}{||(x,y,z)||}$$

where F is the gravitational force exerted by an object at the origin with mass m_1 on an object at (x,y,z) with mass m_2. Here G is the universal gravitational constant. List features of this picture of Earth's gravitational field that are reflected in Newton's law.



1.4. Coordinate transformations

This section continues to study maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^n.$$

You will view them as transformations instead of vector fields. The word "transformation" is widely used in mathematics and science, referring to almost any kind of map, but no common technical definition really exists. In this text[10], a **transformation** will usually refer to any map with domain and codomain in $\mathbb{R}^n$. In other words, a transformation is a map between two subsets lying in the same dimension. It is usually continuous.

Example 1.4.1 The transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(u,v) = (u+v, u-v)$ is a linear transformation. The transformations $g : \mathbb{R}^2 \to \mathbb{R}^2$ and $h : \mathbb{R}^2 \to \mathbb{R}^2$ given by $g(u,v) = (u^2+v^2, v)$ and $h(u,v) = (u+1, 0)$ are non-linear transformations. You can show that f is invertible, whereas g and h are not.

A **coordinate transformation** $f : A \to B$ will refer to a continuous transformation that is usually (but not always) bijective. The domain A and the map f form a **coordinate system** for the codomain B. For example, if $b = f(a)$ then you might informally say any one of:

"The point b can be written as the point a in the coordinate system defined by f."

"The point b in B-space corresponds to the point a in A-space".

This language arises when you are plotting subsets of B but you wish to describe these subsets using the coordinate system defined by f and A.

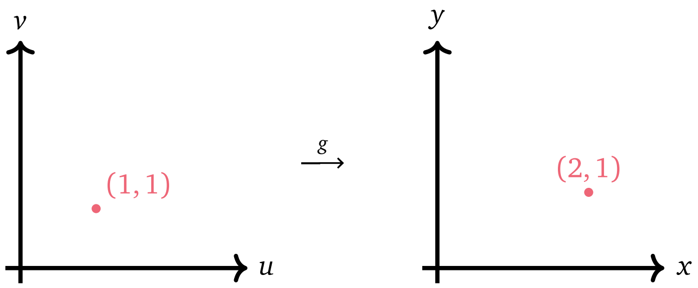
Example 1.4.2 Consider the transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $g(u,v) = (u^2+v^2, v)$. The pair (u,v) denotes a point in the domain $\mathbb{R}^2$ so, to distinguish points in the codomain $\mathbb{R}^2$, it is a common convention to use another pair, usually (x,y). Then you may write

$$(x,y) = (u^2+v^2, v)$$

to equivalently describe the coordinate transformation g. Regarding notation, there is nothing special about the choice of letters (u,v) and (x,y). Introducing them makes it easier to describe whether you are referring to the domain or codomain of g, but that is all.

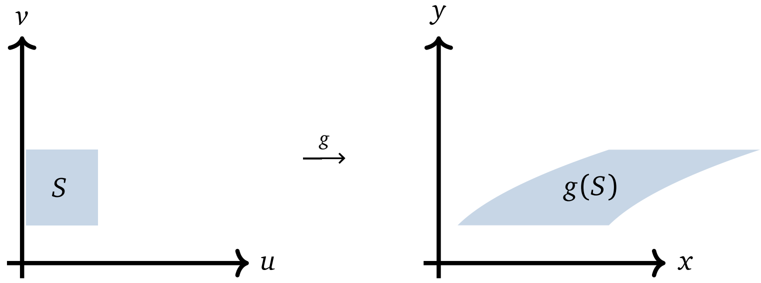
Now, notice $g(1,1) = (2,1)$ for example. This means the point $(2,1)$ in (x,y)–space corresponds to the point $(1,1)$ in (u,v)–space. Notice you also have that $g(-1,1) = (2,1)$ so the point $(2,1)$ can be written as $(1,1)$ or $(-1,1)$ in (u,v)–coordinates.

Transformations in $\mathbb{R}^2$ are commonly visualized using two planes. For example, you can plot the point $(1,1)$ in the (u,v)–plane and its image $g(1,1) = (2,1)$ in the (x,y)–plane.

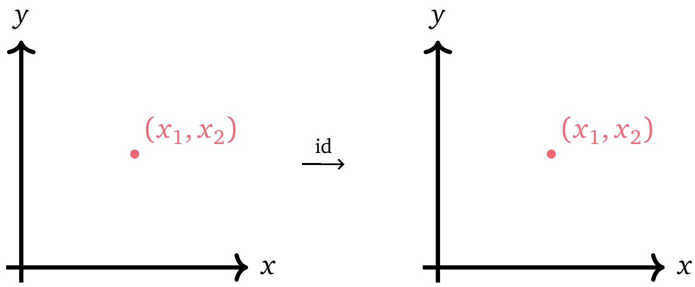


[10] Beware that this convention is not widely used in mathematics.

This idea generalizes to any set $S \subseteq \mathbb{R}^2$ in the (u,v)–plane and its image $g(S) \subseteq \mathbb{R}^2$ in the (x,y)–plane. For example, you can use $S = [0,1] \times [0.5, 1.5]$ and obtain the plot below.



Example 1.4.3 The standard coordinate system for $\mathbb{R}^n$ is **rectangular coordinates**. The coordinate transformation is the identity map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ where $\mathrm{id}(x_1, \dots, x_n) = (x_1, \dots, x_n)$. The plot for $\mathbb{R}^2$ is included below.



As you can see, this is the boring familiar way to plot points using the standard basis of $\mathbb{R}^2$. The point (x_1, x_2) corresponds to itself. Whoop-dee-doo.

Coordinate transformations are wonderful because they allow you to describe the same set in many ways. By choosing your coordinate system wisely, you can dramatically simplify many scenarios. This feature mirrors "change-of-basis" in linear algebra: choose the right basis and your linear algebra problem can be a piece of cake! You will study three fundamental examples: polar coordinates in $\mathbb{R}^2$, cylindrical coordinates in $\mathbb{R}^3$, and spherical coordinates in $\mathbb{R}^3$.

1.4.1 Polar coordinates

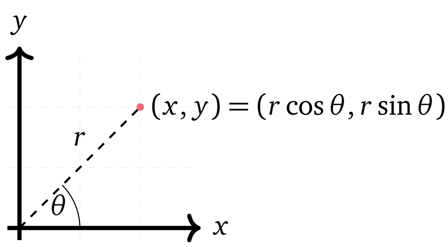
A point in the (x,y)–plane can be equivalently described using its distance from the origin and the polar angle. Formally, define the **polar coordinate transformation** $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(r,\theta) = (r\cos\theta, r\sin\theta).$$

The variable r is the *radius* and the variable θ is the *polar angle*. Notice this allows the radius to be negative! Informally, you can write

$$(x,y) = (r\cos\theta, r\sin\theta)$$

and geometrically this can be viewed as follows.

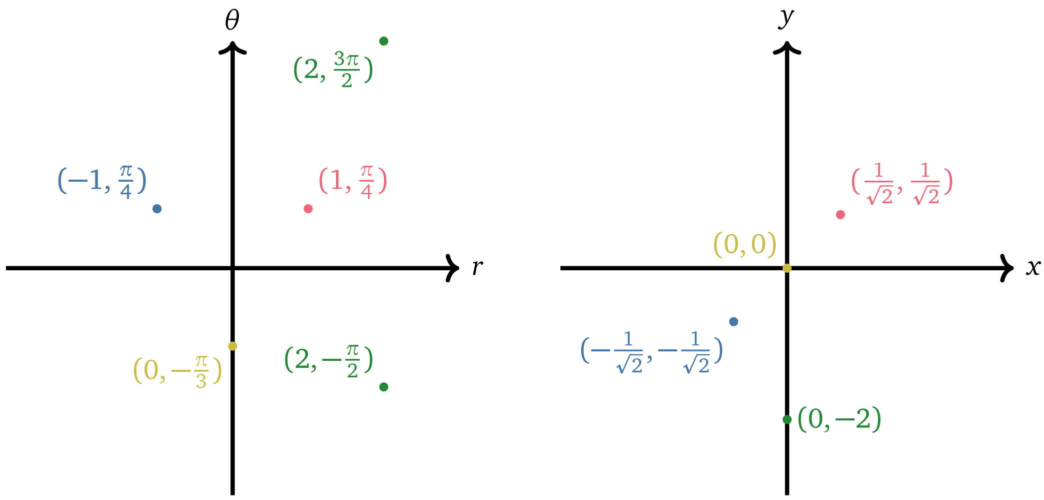


The picture is a bit deceiving since the polar angle θ is not unique.

Example 1.4.4 The polar coordinates transformation T has a lot of symmetries. By direct calculation, notice that

$$T(1, \tfrac{\pi}{4}) = (\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}), \quad T(-1, \tfrac{\pi}{4}) = (-\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}), \quad T(2, \tfrac{-\pi}{2}) = T(2, \tfrac{3\pi}{2}) = (0, -2), \quad T(0, -\tfrac{\pi}{3}) = 0.$$

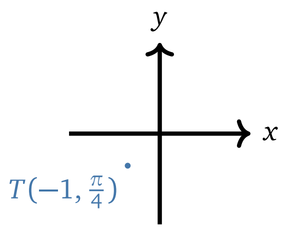
You can plot the transformation of these points.



Notice the radius can be *negative* and that infinitely many points in the (r, θ)–plane correspond to the same point in the (x, y)–plane.

If $(x, y) = T(r, \theta)$ then you may say the point (x, y) can be written as (r, θ) in **polar coordinates**. This language can be convenient since it allows us to *plot* points in rectangular coordinates but *label* them in polar coordinates.

Example 1.4.5 Continuing with the previous example, the point $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ can be written as $(-1, \frac{\pi}{4})$ in polar coordinates. This observation leads to the following plot.



Many authors only write $(-1, \frac{\pi}{4})$ instead of $T(-1, \frac{\pi}{4})$. This can be confusing since the point is actually $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, but you must interpret the meaning from context.

The rectangular coordinate system describes "rectangular" shapes quite nicely, but other shapes (e.g. circles, hyperbolas, and ellipses) can be described more simply in polar coordinates.

Example 1.4.6 What does the polar equation $r = 2$ represent? Informally, notice that

$$x^2 + y^2 = (r\cos\theta)^2 + (r\sin\theta)^2 = r^2.$$

Thus, the equation $r = 2$ in polar coordinates should represent the circle $x^2 + y^2 = 4$.

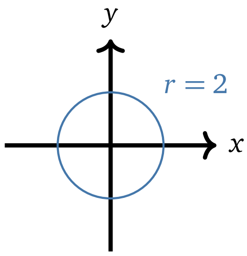
Formally speaking, the polar equation $r = 2$ corresponds to the image of the set

$$A = \{(r,\theta) : r = 2, \theta \in \mathbb{R}\} = \{(2,\theta) : \theta \in \mathbb{R}\}$$

under the polar coordinate transformation T. Notice θ is *any* real number so

$$T(A) = \{(2\cos\theta, 2\sin\theta) : \theta \in \mathbb{R}\}$$

is precisely the circle of radius 2 centered at $(0,0)$. This can be plotted in the (x,y)–plane with its polar coordinate equation.



Example 1.4.7 What does the polar equation $\theta = \frac{\pi}{3}$ represent? Informally, this is trickier since this suggests you must solve the equations

$$x = r\cos\theta, \qquad y = r\sin\theta$$

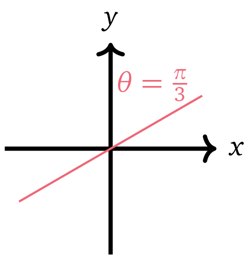
for θ. This will involve some kind of inverse cosine, inverse sine, or inverse tangent but remember that the correct choice of trigonometric inverse depends on the range of θ. The formal route is better. Formally speaking, the polar equation $\theta = \pi/3$ corresponds to the image of the set

$$B = \{(r,\theta) : r \in \mathbb{R}, \theta = \tfrac{\pi}{3}\} = \{(r, \tfrac{\pi}{3}) : r \in \mathbb{R}\}$$

under the polar coordinate transformation T. Notice r is *any* real number! That will be our convention[11]unless you specify r must be positive. Many other sources will always assume r is positive if you write the equation "$\theta = \pi/3$", so you must be careful! Now, observe that

$$T(B) = \{(r\cos\tfrac{\pi}{3}, r\sin\tfrac{\pi}{3}) : r \in \mathbb{R}\} = \{(\tfrac{r}{2}, \tfrac{r\sqrt{3}}{2}) : r \in \mathbb{R}\}$$

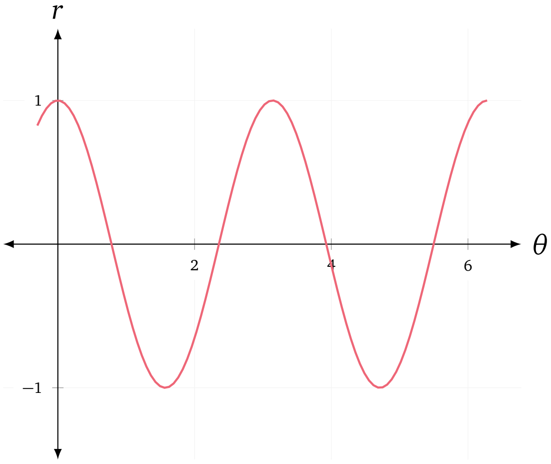
which is a line through the origin. This is plotted below.



Example 1.4.8 Even simple polar equations, like $r = \cos(2\theta)$, are quite tricky to draw. Software helps but you should be familiar with how to sketch and identify basic examples. First, you can plot $r = \cos(2\theta)$ in the (r,θ)–plane as you usually would. Use a table of values

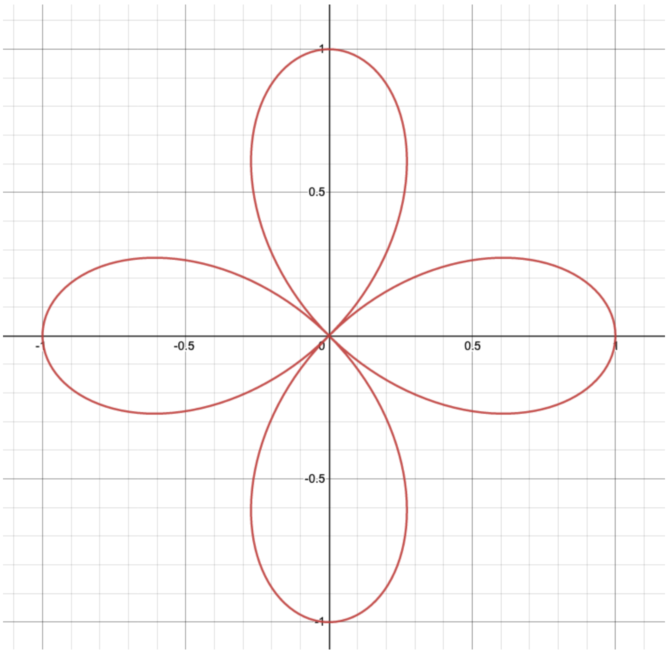
[11]This convention may feel strange since you normally think of a "radius" as positive, but the flexibility and symmetry of allowing negative radii is actually helpful. Embrace the symmetry!

to pay attention to special points like extreme values of r and values of θ where $r = 0$. Since $\cos(2\theta)$ is periodic with period π, it is enough to plot it for $0 \le \theta \le \pi$.



θ	r
0	1
0.25π	0
0.50π	-1
0.75π	0
1.00π	1

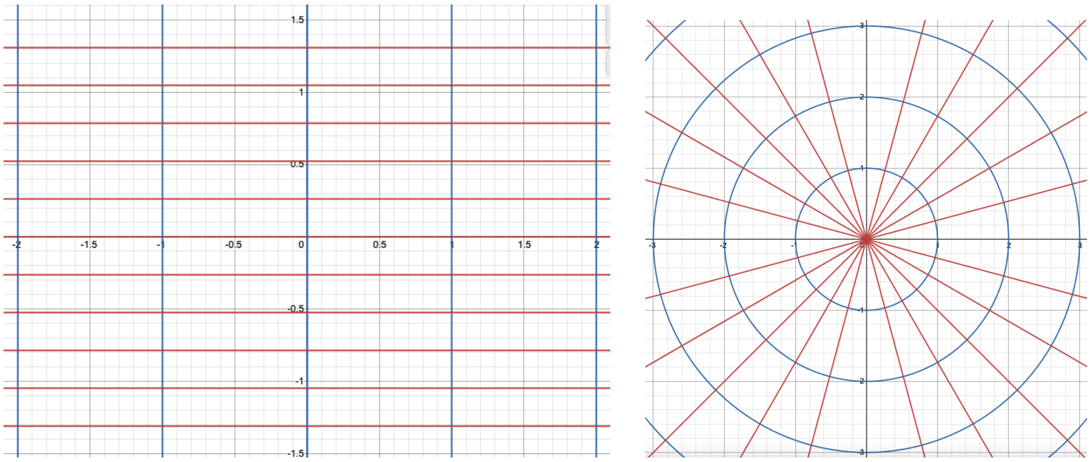
You can transform this into the (x, y)–plane using the polar coordinate transformation. Your table of values can help you produce the following plot.



View this Desmos animation to watch how this polar curve is traced out. Pay attention to when the radius is negative in the range $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$.

Coordinate transformations typically preserve some kind of geometric properties or modify them in a predictable manner. A nice visual in $\mathbb{R}^2$ involves grid lines. The domain $\mathbb{R}^2$ is plotted with the usual vertical and horizontal grid lines (with different colours) and then those lines are transformed to the codomain (keeping the same colours).

Example 1.4.9 On the left-hand side, the grid lines in the (r, θ)–plane are plotted. The blue vertical grid lines correspond to $r = a$ for various constants a and the red horizontal grid lines correspond to $\theta = b$. These are mapped under the polar coordinate transformation to produce the beautiful picture on the righthand side in the (x, y)–plane.



Compare with Examples 1.4.6 and 1.4.7 to see why this produces circles and lines.

You have seen that the polar coordinate transformation is not bijective on its entire domain. However, by restricting its domain to a subset, you can obtain a bijection.

Lemma 1.4.10 The polar coordinate transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(r, \theta) = (r \cos\theta, r \sin\theta)$$

maps the subset $(0, \infty) \times (-\pi, \pi)$ bijectively to the subset $\mathbb{R}^2 \setminus \{(x, 0) : x \le 0\}$.

Proof. Let $A = (0, \infty) \times (-\pi, \pi)$ and $B = \mathbb{R}^2 \setminus \{(x, 0) : x \le 0\}$. First, we prove that $T(A) \subseteq B$. Fix $(r, \theta) \in A$. Assume by contradiction that $T(r, \theta) = (r \cos\theta, r \sin\theta) \notin B$, which implies $r \sin\theta = 0$ and $r \cos\theta \le 0$. This only happens if θ is an odd integer multiple of π but that is impossible when $(r, \theta) \in A$. This proves the claim. Thus, we may define $f : A \to B$ by $f(r, \theta) = T(r, \theta)$ for all $(r, \theta) \in A$. We must prove f is bijective.

Next, we prove that f is injective, i.e. one-to-one. Let $(r_1, \theta_1), (r_2, \theta_2) \in A$ be such that $f(r_1, \theta_1) = f(r_2, \theta_2)$. Then $r_1 \cos\theta_1 = r_2 \cos\theta_2$ and $r_1 \sin\theta_1 = r_2 \sin\theta_2$. It follows that

$$r_1^2 = (r_1 \cos\theta_1)^2 + (r_1 \sin\theta_1)^2 = (r_2 \cos\theta_2)^2 + (r_2 \sin\theta_2)^2 = r_2^2.$$

so $r_1 = r_2$ as both r_1 and r_2 are positive. This implies $(\cos\theta_1, \sin\theta_1) = (\cos\theta_2, \sin\theta_2)$. As $\theta_1, \theta_2 \in (-\pi, \pi)$, notice $\cos\theta_1 = \cos\theta_2$ implies $\theta_1 = \pm\theta_2$ and, similarly, $\sin\theta_1 = \sin\theta_2$ implies $\theta_1 = \theta_2$ or $\theta_1 = \pi - \theta_2$ or $\theta_1 = -\pi - \theta_2$. These conditions are consistent only when $\theta_1 = \theta_2$ so we may conclude that $(r_1, \theta_1) = (r_2, \theta_2)$ as desired.

Finally, we prove that f is surjective. Let $(x, y) \in B$. Define $r = \sqrt{x^2 + y^2}$ and

$$\theta = \begin{cases} \arccos\left(\frac{x}{r}\right) & \text{if } y \ge 0, \\ -\arccos\left(\frac{x}{r}\right) & \text{if } y < 0, \end{cases}$$

so $r > 0$ and $\theta \in (-\pi, \pi)$ by definition. By checking cases depending on the sign of y, you can verify that $f(r, \theta) = (x, y)$. This proves f is surjective and hence bijective, as required. ∎

The proof of Lemma 1.4.10 shows that defining an "inverse" for the polar coordinate transformation is tricky. Many texts suggest the following relationship

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \arctan(y/x),$$

but you must be careful! This relationship only holds for $r \in (0, \infty)$ and $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

1.4.2 Cylindrical coordinates

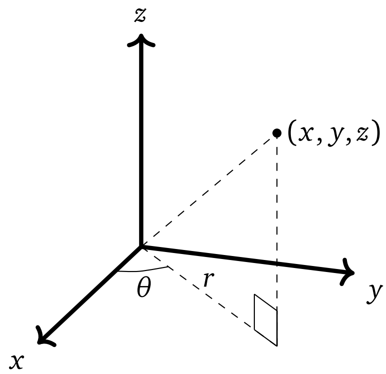
You can extend the idea of polar coordinates in $\mathbb{R}^2$ to a coordinate system in $\mathbb{R}^3$ in two ways. Define the **cylindrical coordinate transformation** $T : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$T(r,\theta,z) = (r\cos\theta, r\sin\theta, z).$$

The variable r is the *polar radius*, the variable θ is the *polar angle*, and the variable z is the usual rectangular coordinate. Informally, you may write

$$(x,y,z) = (r\cos\theta, r\sin\theta, z)$$

and say the point (x,y,z) can be written as (r,θ,z) in **cylindrical coordinates**. The geometry of this coordinate system is captured by the figure below.



This allows us to *plot* points in rectangular coordinates but *label* them in cylindrical coordinates.

Example 1.4.11 The points $(1,\frac{\pi}{2},3)$ and $(-1,-\frac{\pi}{2},3)$ in the (r,θ,z)–space both correspond to the point $(0,1,3)$ in (x,y,z)–space. In other words, $(0,1,3)$ can be written as $(1,\frac{\pi}{2},3)$ or $(-1,-\frac{\pi}{2},3)$ in cylindrical coordinates. There are actually infinitely many ways to write $(0,1,3)$ in cylindrical coordinates.

Much like the name suggests, cylindrical coordinates are useful for describing objects with rotational symmetry about the z-axis in simpler terms.

Example 1.4.12 What does the cylindrical equation $r = 2$ represent in $\mathbb{R}^3$? Like Example 1.4.6, you can informally calculate in cylindrical coordinates that

$$x^2 + y^2 = r^2 = 4,$$

but this shape is *not* a circle in $\mathbb{R}^3$. The equation $r = 2$ does not restrict the z-coordinate (or the θ-coordinate). It helps to consider the formal interpretation. The cylindrical equation $r = 2$ corresponds to the image of the set

$$A = \{(r,\theta,z) : r = 2, \theta \in \mathbb{R}, z \in \mathbb{R}\} = \{(2,\theta,z) : \theta \in \mathbb{R}, z \in \mathbb{R}\}$$

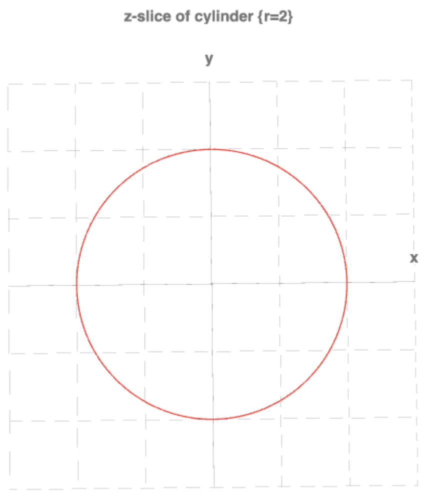
under the cylindrical coordinate transformation T. Then

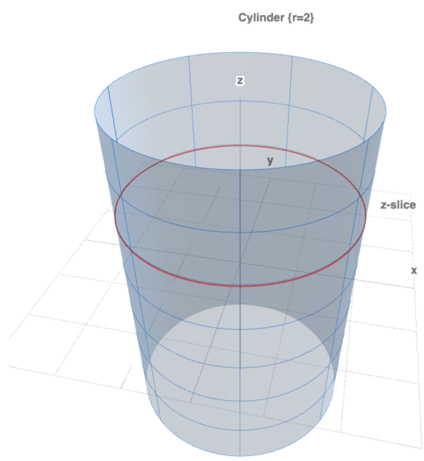
$$T(A) = \{(2\cos\theta, 2\sin\theta, z) : \theta \in \mathbb{R}, z \in \mathbb{R}\},$$

but what is the shape of this set? For any *fixed* value of z, the z-slice of $T(A)$ is the set

$$\{(2\cos\theta, 2\sin\theta) : \theta \in \mathbb{R}\} \subseteq \mathbb{R}^2,$$

which is the circle of radius 2. Since this is true for all values of z, you can imagine stacking a bunch of unit circles on top of each other. This creates a cylinder! In other words, $r = 2$ is a cylinder in $\mathbb{R}^3$.





View this Math3D demo to better visualize this process.

Example 1.4.13 The cylindrical equation $\theta = \pi/4$ corresponds to a plane passing through the z-axis. Namely, if $B = \{(r, \frac{\pi}{4}, z) : r, z \in \mathbb{R}\}$ then its image under the cylindrical coordinate transformation T is

$$T(B) = \{(r\cos\tfrac{\pi}{4}, r\sin\tfrac{\pi}{4}, z) : r, z \in \mathbb{R}\} = \{(r/\sqrt{2}, r/\sqrt{2}, z) : r, z \in \mathbb{R}\}.$$

Remember r is allowed to be negative! View this Math3D demo to see a plot of this example.

Example 1.4.14 The cylindrical equation $z = -1$ is a (flat) plane, just like in rectangular coordinates. Formally speaking, if $C = \{(r, \theta, -1) : r, \theta \in \mathbb{R}\}$ then

$$T(C) = \{(r\cos\theta, r\sin\theta, -1) : r, \theta \in \mathbb{R}\} = C$$

so indeed it remains a flat plane. This is not surprising since there is no constraint on r or θ.

The cylindrical coordinate transformation is a bijection once you restrict its domain.

Lemma 1.4.15 The cylindrical coordinate transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$$

maps the subset $(0, \infty) \times (-\pi, \pi) \times \mathbb{R}$ bijectively to the subset $\mathbb{R}^3 \setminus \{(x, 0, z) : x \leq 0, z \in \mathbb{R}\}$.

Proof. This is left as an exercise. Follow the arguments in Lemma 1.4.10. ■

There are many other ways to restrict the domain and obtain a bijection. You can explore a lot more with cylindrical coordinates but, for now, these examples will be enough. The key takeaway is that this coordinate system can nicely describe objects with rotational symmetry about the z-axis.

1.4.3 Spherical coordinates

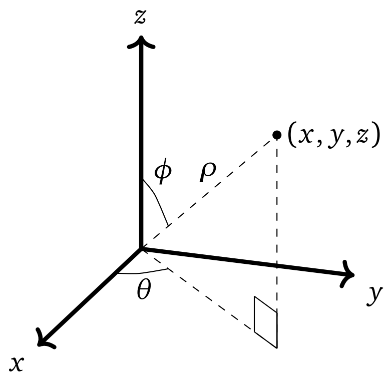
The second way of extending polar coordinates in $\mathbb{R}^2$ is through spherical coordinates in $\mathbb{R}^3$. Define the **spherical coordinate transformation** $T : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$T(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi).$$

The variable ρ is the *(spherical) radius*, the variable θ is the *polar angle* (or *azimuthal angle*), and the variable ϕ is the *inclination angle* (or *zenith angle*)[12]. These Greek letters are pronounced as 'rho', 'thay-ta', and 'fai' (or 'fee'). Remember ρ, θ, and ϕ can be *any* real numbers, including negative ones! Many sources restrict their values, but the transformation is defined for $\mathbb{R}^3$. Informally, you may write

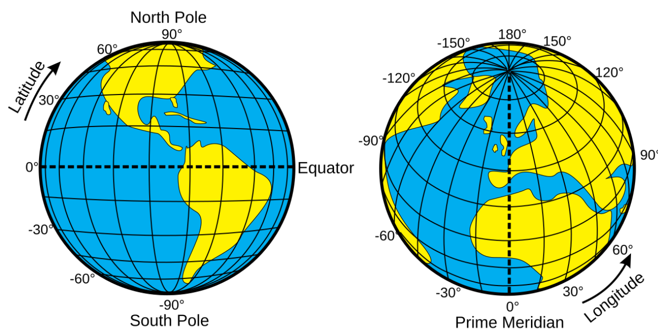
$$(x, y, z) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$$

and say the point (x, y, z) can be written as (ρ, θ, ϕ) in **spherical coordinates**. The geometry of this coordinate system is captured by the figure below.



Play with this Math3D demo to get a feeling for how this system works. Spherical coordinates can be used to describe objects with rotational symmetry about the origin.

Example 1.4.16 Locations on the Earth's surface are described with latitude and longitude. This geographical system is an example of spherical coordinates[13].



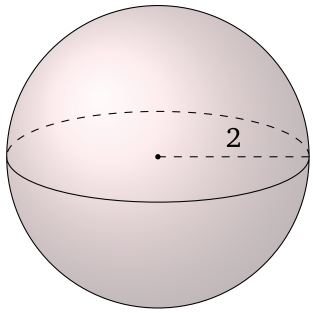
Roughly speaking, longitude relates to the polar angle θ and latitude relates to the inclination angle ϕ. Although the Earth is not a perfect sphere, its spherical radius ρ is $\approx 6{,}378$ km.

[12] These choices of names and variables are fairly common mathematical usage. Be careful with notations in other sources, especially in physics. The roles of θ and ϕ may be switched and ρ may be replaced with r. Some write φ instead of ϕ. The inclination angle can also be called the polar angle. It is truly a sad state of affairs across the literature, so always check your source's conventions.

[13] Image retrieved from Wikimedia Commons on 2024-07-23 licensed under CC0.

Example 1.4.17 The points $(\sqrt{2}, \frac{\pi}{4}, \frac{\pi}{2})$ and $(-\sqrt{2}, \frac{\pi}{4}, \frac{3\pi}{2})$ in (ρ, θ, ϕ)–space both correspond to the same point $(1,1,0)$ in (x,y,z)–space. In other words, $(1,1,0)$ can be written as $(\sqrt{2}, \frac{\pi}{4}, \frac{\pi}{2})$ or $(-\sqrt{2}, \frac{\pi}{4}, \frac{3\pi}{2})$ in spherical coordinates. As usual, there are actually infinitely many ways to do so.

Example 1.4.18 What does the spherical equation $\rho = 2$ represent in $\mathbb{R}^3$? Since ρ is the distance from the origin, this should be the set of points which are distance 2 away from the origin, so that's a sphere with radius 2.



You can verify this informally and formally. Informally, notice

$$\begin{aligned} x^2+y^2+z^2 &= \rho^2\cos^2\theta\sin^2\phi + \rho^2\sin^2\theta\sin^2\phi + \rho^2\cos^2\phi \\ &= \rho^2\sin^2\phi + \rho^2\cos^2\phi \\ &= \rho^2 = 4, \end{aligned}$$

so indeed every point satisfying $\rho = 2$ satisfies $x^2+y^2+z^2 = 4$. The equation $\rho = 2$ also does not restrict the θ-coordinate or the ϕ-coordinate. Formally, the spherical equation $\rho = 2$ corresponds to the image of the set

$$A = \{(\rho,\theta,\phi) : \rho = 2, \theta \in \mathbb{R}, \phi \in \mathbb{R}\} = \{(2,\theta,\phi) : \theta \in \mathbb{R}, \phi \in \mathbb{R}\}$$

under the spherical coordinate transformation T. Then

$$T(A) = \{(2\cos\theta\sin\phi, 2\sin\theta\sin\phi, 2\cos\phi) : \theta \in \mathbb{R}, \phi \in \mathbb{R}\},$$

but what is the shape of this set? From the calculation above, every point $(x,y,z) \in T(A)$ satisfies $x^2+y^2+z^2 = 4$ so $T(A)$ is a subset of the sphere of radius 2. You must show $T(A)$ is actually equal to the sphere of radius 2, but this requires a bit of work. If you prove Lemma 1.4.20 below, you can use the same ideas to prove this fact. This is left as an exercise.

Example 1.4.19 The spherical equation $\theta = \pi/4$ represents the same object as it does in cylindrical coordinates. Informally speaking, this is not surprising since θ should be the polar angle in both cases. Formally, this conclusion is not obvious. If $B = \{(\rho, \frac{\pi}{4}, \phi) : \rho, \phi \in \mathbb{R}\}$ then its image under the spherical coordinate transformation is

$$\begin{aligned} T(B) &= \{(\rho\cos\tfrac{\pi}{4}\sin\phi, \rho\sin\tfrac{\pi}{4}\sin\phi, \rho\cos\phi) : \rho,\phi \in \mathbb{R}\} \\ &= \{(\tfrac{1}{\sqrt{2}}\rho\sin\phi, \tfrac{1}{\sqrt{2}}\rho\sin\phi, \rho\cos\phi) : \rho,\phi \in \mathbb{R}\}. \end{aligned}$$

Indeed, it is not obvious that this matches Example 1.4.13. The proof is left as an exercise.

You can restrict the domain of the spherical coordinate transformation to obtain a bijection. There are many other ways to do so, as well.

Lemma 1.4.20 The spherical coordinate transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$$

maps the subset $(0, \infty) \times (-\pi, \pi) \times (0, \pi)$ bijectively to the subset $\mathbb{R}^3 \setminus \{(x, 0, z) : x \leq 0, z \in \mathbb{R}\}$.

Proof. This is left as an exercise. Adapt the ideas from Lemma 1.4.10. ■

You can explore a lot more with spherical coordinates but, for now, these examples will be enough. The key takeaway is that this coordinate system can nicely describe objects with rotational symmetries about the origin.

Exercises for Section 1.4

Concepts and definitions

1.4.1 Often you will want to understand how quantities vary under a transformation, such as volume or area. However, linear transformations are simple and non-linear ones are not.

(a) Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a **linear** transformation given by $T(x) = Ax$ for some 2x2 matrix A. If $U \subseteq R^2$ is a unit square, what is the area of $T(U)$? And what are the possible shapes of $T(U)$?

(b) You have already seen an example of a **non-linear** transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$, namely the polar coordinate transformation $f(r, \theta) = (r\cos\theta, r\sin\theta)$. Let $U = [0, 1] \times [\frac{\pi}{6}, \frac{\pi}{4}]$ and $V = [1, 2] \times [\frac{\pi}{6}, \frac{\pi}{4}]$. Sketch the transformation of U and V under f. Remember to draw two planes: the (r, θ)-plane and the (x, y)-plane.

(c) What is the area of U and V in $\mathbb{R}^2$? And the area of $f(U)$ and $f(V)$ in $\mathbb{R}^2$?

1.4.2 Start with polar coordinates in $\mathbb{R}^2$. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinates transformation given by

$$T(r, \theta) = (r\cos\theta, r\sin\theta).$$

The domain of T is referred to as the (r, θ)–plane and its codomain is the (x, y)–plane. When asked to draw transformations under T, remember to draw these two sets of 2D axes.

(a) Draw the points $(r, \theta) = (1, \frac{\pi}{4}), (1, -\frac{\pi}{4}), (-1, \frac{\pi}{4})$ and $(-1, -\frac{\pi}{4})$ in the (r, θ)–plane and in the (x, y)–plane.

(b) Is T a linear transformation?

(c) Consider the polar equation $r = 2$. Write down two sets describing this equation: one set A in the (r, θ)–plane and the other set $B = T(A)$ in the (x, y)–plane.

(d) Below are four equations defining curves in $\mathbb{R}^2$.

$$\text{(a) } r = 2 \qquad \text{(b) } \theta = \frac{\pi}{4} \qquad \text{(c) } \theta = \frac{\pi}{4}, r > 0 \qquad \text{(d) } r = 0$$

Draw all of them in the (r, θ)–plane and the (x, y)–plane. Use different colours.

(e) What does a vertical translation by $+\frac{\pi}{3}$ in the (r, θ)–plane correspond to in the (x, y)–plane?

(f) I have a set A in the (r, θ)–plane and its image $T(A)$ in the (x, y)–plane. I want to dilate the image outward by a factor of 2 from the origin. Define a map $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ so that $T_2(A)$ is the image of this transformation.

(g) Is the polar coordinate transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ one-to-one? If so, explain. If not, restrict it to a region S where it is one-to-one and onto and find its inverse on that region.

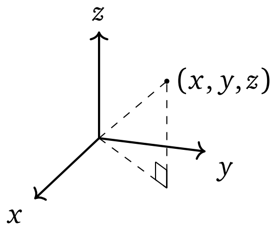
1.4.3 Next, we continue to cylindrical coordinates. A lot of the same features from polar coordinates carry over but now we're in 3-dimensions.

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the cylindrical coordinates transformation given by

$$T(r, \theta, z) = (r\cos\theta, r\sin\theta, z).$$

The domain is called (r, θ, z)-space and the codomain is called (x, y, z)-space. When asked to draw transformations under T, remember to draw these two sets of 3D axes.

(a) Label the diagram below to geometrically explain the formula defining $T(r,\theta,z)$. You will need to label all sides of the triangle.



(b) What does a θ-translation by $-\frac{\pi}{2}$ in the (r,θ,z)-space correspond to in the (x,y,z)-space? Describe it carefully since this is now 3-dimensions.

(c) What does a vertical translation by -3 in the (x,y,z)–space correspond to in the (r,θ,z)–space?

(d) I have a set A in the (r,θ,z)–space and its image $T(A)$ in the (x,y,z)-space. Consider the function $f:\mathbb{R}^3\to\mathbb{R}^3$ defined by $f(r,\theta,z)=T(3r,\theta,z-3)$. Describe in plain terms how the image $f(A)$ is related to the image $T(A)$.

(e) Below are three equations defining objects in $\mathbb{R}^3$ using cylindrical coordinates.

(a) $r=2$ (b) $\theta=\dfrac{\pi}{4}, r>0$ (c) $0\le r\le z$

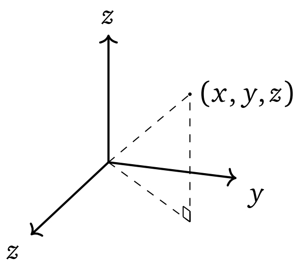
Sketch each of them in (r,θ,z)-space and (x,y,z)-space.

1.4.4 Finally, there are spherical coordinates in $\mathbb{R}^3$. Let $T:\mathbb{R}^3\to\mathbb{R}^3$ be the spherical coordinates transformation given by

$$T(\rho,\theta,\phi)=(\rho\cos\theta\sin\phi,\rho\sin\theta\sin\phi,\rho\cos\phi).$$

The domain is called (ρ,θ,ϕ)-space and the codomain is called (x,y,z)-space. When asked to draw transformations under T, remember to draw these two sets of 3D axes.

(a) Label the diagram to geometrically explain the formula defining $T(\rho,\theta,\phi)$.



(b) Give an algebraic description for a solid sphere of radius 7 using spherical coordinates.

(c) Provide at least 3 different values of (ρ,θ,ϕ) such that $T(\rho,\theta,\phi)=(0,1,0)$. *Hint:* Use geometry.

(d) Below are three equations defining objects in $\mathbb{R}^3$ using spherical coordinates.

(a) $\dfrac{\pi}{4}\le\theta\le\dfrac{\pi}{3}$ (b) $1\le\rho\le 2$ (c) $\phi=\dfrac{\pi}{4}$

Sketch each of them in (ρ,θ,ϕ)-space and (x,y,z)-space.

(e) Assume the Earth is a perfect sphere. Give the position of Toronto using spherical coordinates. *Hint:* You will need to look up some information about the Earth.

(f) Assume the Earth is a perfect sphere. Describe the Tropic of Capricorn using spherical coordinates.

(g) Describe a spherical coordinate transformation that corresponds to inflating a ball to three times its size.

(h) Exhibit a region S where $T : \mathbb{R}^3 \to \mathbb{R}^3$ is one-to-one and onto. No proof is necessary.

Applications and beyond

1.4.5 Drawing transformations $\mathbb{R}^2 \to \mathbb{R}^2$ is a surprisingly useful tool when trying to understand how a proof works or figure out a problem in higher dimensions. You need to develop this skill in order to think about abstract functions and proofs. Let's practice.

(a) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$. Suppose f maps a disk $A \subseteq \mathbb{R}^2$ inside a blob $B \subseteq \mathbb{R}^2$. Sketch this transformation and label all relevant regions.

(b) Does your drawing seem to assume f is "continuous"? Explain why or why not

(c) Draw the same situation and assume f has the properties of a continuous function. Draw a sequence of points $x_1, x_2, \ldots$ in A that approach some $x \in A$. Draw the corresponding images. Label everything you can.

(d) Now, let's draw a more complicated scenario.

- Pick a point $a \in \mathbb{R}^2$ in the domain and let $b = f(a) \in \mathbb{R}^2$.
- Draw a disk $B_\varepsilon(b) = \{x \in \mathbb{R}^2 : ||x - b|| < \varepsilon\}$ of radius ε centered at b.
- Draw a disk $B_\delta(a) = \{x \in \mathbb{R}^2 : ||x - a|| < \delta\}$ of radius δ centered at a.

Assume that all of these quantities satisfy:

$$\forall x \in \mathbb{R}^2, x \in B_\delta(a) \implies f(x) \in B_\varepsilon(f(a))$$

Draw this situation along with the image $f(B_\delta(a))$. Label everything you can.

1.5. Parametric, explicit, and implicit form

Linear algebra has created a language for identifying lower dimensional objects in higher dimensional spaces. Indeed you can use it to answer questions of the following form.

If a set $S \subseteq \mathbb{R}^n$ is described by ***linear*** *equations, then what is its dimension?*

The concept of "dimension" in linear algebra is only defined for subspaces of $\mathbb{R}^n$, and it is calculated by finding the size of a basis. Thus, depending on how these linear equations are described, you can compute the dimension of S by analyzing an appropriate subspace of $\mathbb{R}^n$. Notice, however, linear algebra only provides tools to analyze *linear* objects. This means you can study "flat" objects like lines, planes, and subspaces because these are defined by linear equations. Most geometric objects are not defined this way, thus spawning a golden question.

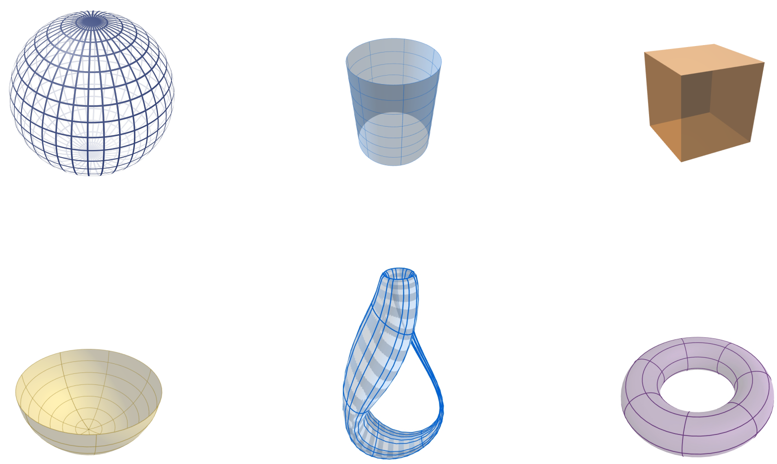
If a set $S \subseteq \mathbb{R}^n$ is described by ***nonlinear*** *equations, then what is its "dimension"? How do you even define "dimension"?*

This question will be a central problem in your multivariable calculus journey. You will need to develop some significant theory before even approaching it. At the moment, you can speculate with some low dimensional cases to guide your intuition.

The first case is the idea of a "curve". A curve should presumably be created by bending a straight line. Since a straight line is a 1-dimensional (linear) object in higher dimensional space, a curve should probably be thought of as a 1-dimensional (nonlinear) object living in a higher dimensions. You have already been introduced to curves, and these are rather special. It is better to think about a slightly more complicated situation.

The second case is the idea of a “surface”. This will capture the core issues of the golden question. Intuitively speaking, a surface should presumably be created by bending a piece of a plane in $\mathbb{R}^3$. Since a plane is a 2-dimensional (linear) object living in 3-dimensional space, a surface should probably be thought of as a 2-dimensional (nonlinear) object living in 3-dimensional space. You can imagine this idea with some visual examples.

Example 1.5.1 Here are pictures of classic surfaces: a sphere [14], a cylinder, a cube, a hemisphere, a Klein bottle, and a torus.



As you can see, there is a tremendous diversity in their shapes and smoothness. Formally describing some of them can be really quite difficult. This variation indicates how tricky it will be to formalize the concept of “2-dimensional object in 3-dimensional space”.

[14]Image retrieved from Wikimedia Commons on 2024-07-23 licensed under CC BY.

The general case is the idea of a "**manifold**".[15] Fix $k, n \in \mathbb{N}^+$ with $k < n$. Intuitively speaking, a k-dimensional manifold in $\mathbb{R}^n$ should presumably be created by bending a piece of a k-dimensional plane in $\mathbb{R}^n$. This loosely described idea sounds really interesting, especially in light of the golden question above. It will take many chapters before you can study this problem and rigorously define manifolds (and hence curves and surfaces). Right now, you need to take a step back and ask more foundational questions.

How can a set $S \subseteq \mathbb{R}^n$ be described by nonlinear equations? Is there more than 1 way?

It was historically an enormous struggle to come up with a satisfactory definition of manifolds, and these foundational questions were some of the core struggles. There are unfathomably many ways to describe sets with nonlinear equations. In this section, you will explore the three fundamental forms for describing sets: parametric form, explicit form, and implicit form. Each form involves maps from $\mathbb{R}^n \to \mathbb{R}^m$ where $m \neq n$.

1.5.1 Parametric form

One natural way to describe sets with nonlinear equations is using maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

when $n < m$. You have already seen the case $n = 1$ (parametric curves $\mathbb{R} \to \mathbb{R}^m$) but these are very special. You will usually think of $n \geq 2$ in which case the maps $\mathbb{R}^n \to \mathbb{R}^m$ are new to you. The visual focus here will be on the special case $n = 2$ and $m = 3$, that is, maps of the form

$$\mathbb{R}^2 \longrightarrow \mathbb{R}^3.$$

These maps will presumably create surfaces, i.e. 2-dimensional manifolds in $\mathbb{R}^3$.

Example 1.5.2 The **unit sphere** is the sphere of radius 1 centred at the origin. Let S be the unit sphere in $\mathbb{R}^3$. It can be described using 2 "parameters". Namely, define $g : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$g(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi).$$

This is the spherical coordinate transformation with the spherical radius fixed, so

$$S = \{g(\theta, \phi) : (\theta, \phi) \in \mathbb{R}^2\} = \operatorname{im}(g).$$

It seems reasonable to guess that S will be a “2-dimensional manifold in $\mathbb{R}^3$” (i.e. a surface) because it can be described as the image of a map from $\mathbb{R}^2$ to $\mathbb{R}^3$. In other words, it can be described in $\mathbb{R}^3$ with two parameters, the polar angle θ and the inclination angle ϕ.

Unfortunately, this calculation is not quite good enough for a definition of a surface. For example, you could be silly and define $h : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$h(\theta, \phi) = (\cos\theta, \sin\theta, 0),$$

so the image of h is a unit circle lying in the $z = 0$ plane. Although h has two inputs, the output in $\mathbb{R}^3$ is a curve, i.e. “1-dimensional manifold”! How can you find a formal way of distinguishing these two scenarios? That is a tough question for much later.

Despite the challenges presented by this example, you can extend the idea to any dimension.

[15] This textbook will use the word "manifold" to really mean "embedded manifold". This choice is not standard in mathematical literature, but you will not explore the distinction here. You will simply use the word "manifold". The full modern definition is better reserved for a course in differential geometry.

Definition 1.5.3 Let $m, n \in \mathbb{N}^+$ with $n < m$. A set $S \subseteq \mathbb{R}^m$ can be written in **parametric form (with n-variables)** if there exists a set $A \subseteq \mathbb{R}^n$ and a continuous map $g : A \to \mathbb{R}^m$ such that

$$S = \{g(x) : x \in A\} = \operatorname{im}(g).$$

Equivalently, we say the set S is **parametrized by** g.

If a set in $\mathbb{R}^m$ is parametrized by a map with n inputs, then you might guess that it should be an n-dimensional manifold in $\mathbb{R}^m$. However, Example 1.5.2 illustrates that the set could be anything! It could be a curve, a point, or a total mess. Thus, the definition above is only a starting point and does not fully match your intuitive understanding of an n-dimensional manifold. Definition 1.5.3 simply gives one way to describe sets with nonlinear equations.

1.5.2 Explicit form

Another way to describe sets is actually a special case of parametric form. In particular, you have already encountered a large class of sets which are easy to parametrize: graphs.

Example 1.5.4 The graph of the real-valued function $f(x, y) = x^2 + y^2$ is the set

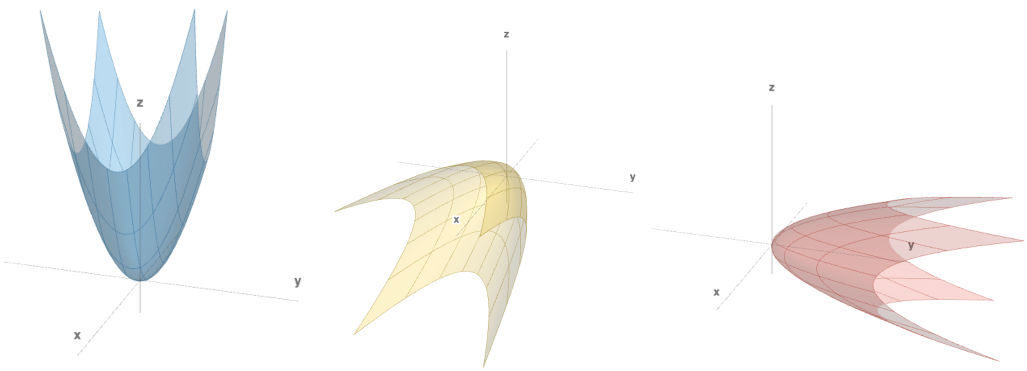
$$S = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\} = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\}$$

in $\mathbb{R}^3$. How can you parametrize this graph? Define $g : \mathbb{R}^2 \to \mathbb{R}^3$ by $g(x, y) = (x, y, x^2 + y^2)$ so $\operatorname{im}(g) = g(\mathbb{R}^2) = S$ implying S is parametrized by g. This produces the graph below.

Now, by convention, *the* graph of f is always defined using the z-coordinate but there are other graphs produced by f. As long as you can express one coordinate as a function of the others, the set will be a graph. For example, the sets

$$S_1 = \{(x, y, z) \in \mathbb{R}^3 : x = y^2 + z^2\} \qquad S_2 = \{(x, y, z) \in \mathbb{R}^3 : y = x^2 + z^2\}$$

are also graphs of f which can be parametrized. From left to right, the plots of S, S_1, and S_2 are below.



These common sets are all called **paraboloids** due to the shapes of their slices.

The above example illustrates that if $f : \mathbb{R}^2 \to \mathbb{R}$ is a real-valued function, then any of the following sets

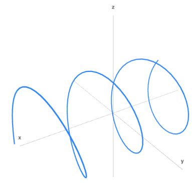
$$\{(x, y, z) \in \mathbb{R}^3 : x = f(y, z)\}, \quad \{(x, y, z) \in \mathbb{R}^3 : y = f(x, z)\}, \quad \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\},$$

are graphs of f. The third set is commonly referred to as *the* graph of f by convention.

Example 1.5.5 Real-valued functions are not the only maps that create graphs. You can use vector-valued functions, too. For example, the graph of the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ defined by $\gamma(t) = (\cos t, \sin t)$ is the set

$$S = \{(t, \gamma(t)) : t \in \mathbb{R}\} = \{(t, \cos t, \sin t) : t \in \mathbb{R}\}$$

lying in $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$. This produces a helix along the x-axis.



Notice the y- and z-variables are functions of the x-variable. This is expected since S is parametrized by the function $g(t) = (t, \gamma(t)) = (t, \cos t, \sin t)$.

Graphs can therefore be formalized to any dimension.

Definition 1.5.6 Let $m, n \in \mathbb{N}^+$ with $n < m$. Let $A \subseteq \mathbb{R}^n$. Let $f : A \to \mathbb{R}^{m-n}$ be a continuous function. The set $S \subseteq \mathbb{R}^n \times \mathbb{R}^{m-n} = \mathbb{R}^m$ given by

$$S = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^{m-n} : x \in A, y = f(x)\}$$

is ***the* graph of** f. A set $\widetilde{S} \subseteq \mathbb{R}^m$ is ***a* graph of** f if $\widetilde{S}$ is the same as S after reordering variables.

Remark 1.5.7 "Reordering variables" can be formally expressed as $\widetilde{S} = \pi(S)$ for a linear transformation $\pi : \mathbb{R}^m \to \mathbb{R}^m$ given by $\pi(x) = Px$ where P is a $m \times m$ permutation matrix.

As the examples demonstrate, graphs can always be written in parametric form. This suggests a new definition.

Definition 1.5.8 Let $m, n \in \mathbb{N}^+$ with $n < m$. A set $S \subseteq \mathbb{R}^m$ can be written in **explicit form (in n variables)** if S is a graph of a continuous function $f : A \to \mathbb{R}^{m-n}$ where $A \subseteq \mathbb{R}^n$.

Example 1.5.9 The set $\{(x, y, z) \in \mathbb{R}^3 : x = y^2 + z^2\}$ is written in explicit form since it is a graph of the function $f(y, z) = y^2 + z^2$.

Example 1.5.10 By Example 1.5.2, the unit sphere S in $\mathbb{R}^3$ can be written in parametric form, but not in explicit form. Since there are many different possible graphs, there are many cases to consider. Here is a sketch of the proof.

Proof. (Sketch) Suppose, for a contradiction, that S can be written in explicit form. There are 6 cases to consider. First, assume you can write

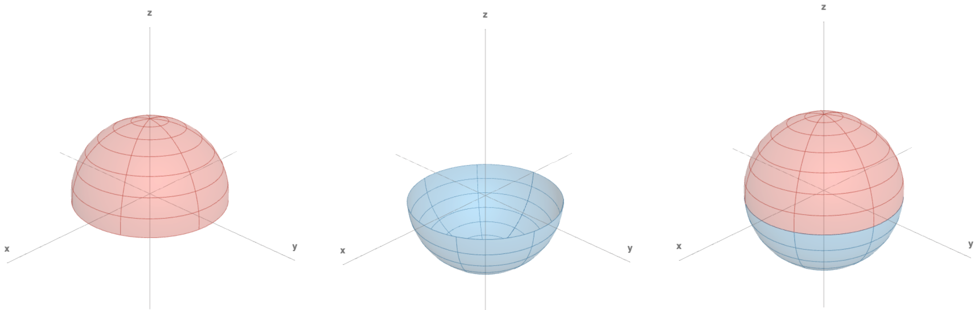
$$S = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\}$$

for some continuous real-valued two variable function f. Notice that $(0, 0, \pm 1) \in S$ which implies $f(0, 0) = 1$ and $f(0, 0) = -1$, which is a contradiction. It is possible that S could be written as $x = f(y, z), y = f(x, z)$ or $(x, y) = g(z), (x, z) = g(y), (y, z) = g(x)$, but you can obtain a contradiction in a similar fashion since $(0, \pm 1, 0) \in S$ and $(\pm 1, 0, 0) \in S$. ■

Although the unit sphere S cannot be written in explicit form, it can be written as a *finite union* of sets in explicit form! In particular,

$$S = \left\{(x,y,z) \in \mathbb{R}^3 : z = \sqrt{1-x^2-y^2}\right\} \cup \left\{(x,y,z) \in \mathbb{R}^3 : z = -\sqrt{1-x^2-y^2}\right\}$$

so S is the union of the graphs of $f(x,y) = \sqrt{1-x^2-y^2}$ and of $g(x,y) = -\sqrt{1-x^2-y^2}$. These two sets in explicit form are the upper and lower hemisphere.



Informally, you can see that any vertical line crosses the lower hemisphere at most once, and the upper hemisphere at most once. But there are vertical lines that pass the entire sphere *twice*! Colloquially speaking, the sphere "fails the vertical line test", but this only means that S cannot be written as a graph of the form $z = f(x,y)$. You also need to exclude the other 5 cases, but the argument is similar.

Thus, sets in explicit form can always be written in parametric form, but the converse is not necessarily true. Sets in explicit form are very special, because some variables can be *explicitly* expressed as a function of the others. This feature will be masterfully exploited in your eventual definition of manifolds. For now, explicit form is just another way to describe sets.

1.5.3 Implicit form

Sets can also be naturally described by nonlinear equations using maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

when $n > m$. For instance, a real-valued function $\mathbb{R}^3 \to \mathbb{R}$ can create a surface via its level sets.

Example 1.5.11 The unit sphere S in $\mathbb{R}^3$ is defined to be the set of points that are distance 1 away from the origin. That is,

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\},$$

which is the 1-level set of the function $f(x,y,z) = x^2 + y^2 + z^2$. The *implicit* equation

$$x^2 + y^2 + z^2 = 1$$

does not *explicitly* express one variable in terms of the others, like $z = g(x,y)$.

Other sets can also be created by looking at maps $\mathbb{R}^n \to \mathbb{R}^m$ when $n > m \geq 2$.

Example 1.5.12 Define $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $g(x,y,z) = (z - x^2 - y^2, x^2 + y^2 + z^2)$. Consider the set C in $\mathbb{R}^3$ defined by

$$C = \{(x,y,z) \in \mathbb{R}^3 : g(x,y,z) = (0,9)\}.$$

In other words, C is the set of points satisfying the two non-linear equations

$$z - x^2 - y^2 = 0, \qquad x^2 + y^2 + z^2 = 9.$$

Luckily, you can solve these to find that C is the circle $x^2 + y^2 = a$ lying in the plane $z = a$ where $a = \frac{1}{2}(\sqrt{37} - 1)$. In other words, C is a curve implicitly defined by 2 non-linear equations. View this Math3D demo to see how C is the intersection of two surfaces.

These examples suggest another way to describe sets by nonlinear equations.

Definition 1.5.13 Let $m, n \in \mathbb{N}^+$ with $n > m$. A set $S \subseteq \mathbb{R}^n$ can be written in **implicit form (with m equations)** if there exists a constant $c \in \mathbb{R}^m$, a set $A \subseteq \mathbb{R}^n$, and a continuous function $f : A \to \mathbb{R}^m$ such that

$$S = f^{-1}(\{c\}) = \{x \in \mathbb{R}^n : f(x) = c\}.$$

Remark 1.5.14 The notation f^{-1} is *not* the inverse function of f unless specified otherwise. It is an operation on sets. The **preimage** of a set C under a function $f : A \to B$ is the set

$$f^{-1}(C) = \{x \in A : f(x) \in C\}.$$

If the set $C = \{c\}$ is a singleton, then the condition $f(x) \in C$ becomes $f(x) = c$. It is incorrect to write $f^{-1}(c)$ instead of $f^{-1}(\{c\})$ because the expression $f^{-1}(c)$ only makes sense if f^{-1} is the inverse function of f. See Section 0.3 for details.

Example 1.5.15 The unit sphere S in $\mathbb{R}^3$ can be written in implicit form, because it is the 1-level set of the continuous function $f(x, y, z) = x^2 + y^2 + z^2$, that is, $S = f^{-1}(\{1\})$. Similarly, the set C in Example 1.5.12 can be written in implicit form since $C = g^{-1}(\{(0, 9)\})$ where $g(x, y, z) = (z - x^2 - y^2, x^2 + y^2 + z^2)$.

Example 1.5.16 The paraboloid $\{(x, y, z) \in \mathbb{R}^3 : z - x^2 - y^2 = 0\}$ is written in implicit form, since it is the 0-level set of $f(x, y, z) = z - x^2 - y^2$. The same trick can be used for any graph, so sets in explicit form can always be written in implicit form.

If a set in $\mathbb{R}^n$ is written in implicit form with m nonlinear equations, then what would you guess to be its "dimension" (if it were a manifold)? You have not yet defined manifolds but you can use linear algebra to formulate an educated guess. A system with n variables and m linear equations can be represented by an equation of the form $Ax = 0$, where A is an $m \times n$ matrix and $x \in \mathbb{R}^n$ is unknown. The set of solutions x to this linear system is the null space of A. If $m < n$ and *the m linear equations (i.e. rows of A) are linearly independent*, then the null space is $(n - m)$-dimensional. While sets in implicit form are defined using *nonlinear* equations, a similar principle appears to hold true in some sense.

Example 1.5.17 The unit sphere S in $\mathbb{R}^3$ is written in implicit form with 1 nonlinear equation $x^2 + y^2 + z^2 = 1$ in 3 variables x, y, z. This means the sphere should presumably be $3 - 1 = 2$ dimensional, which it is! Similarly, the set C in Example 1.5.12 is defined by 2 nonlinear equations in 3 variables, so C should presumably be $3 - 2 = 1$ dimensional, which it is! This acts as evidence that the principles of linear algebra may carry over to nonlinear systems.

This investigation is the beginning of something much greater, namely the implicit function theorem. You will explore that in depth later. For now, the key takeaway is that sets can have three different descriptions: parametric, explicit, and implicit. Each has its own advantages and disadvantages, and you will need to discover their connections to truly understand manifolds.

Exercises for Section 1.5

Concepts and definitions

1.5.1 Define the following functions:

- $f : \{(x,z) \in \mathbb{R}^2 : x^2+z^2 \leq 1\} \to \mathbb{R}$ by $f(x,z) = \sqrt{1-x^2-z^2}$
- $g : \{(s,t) \in \mathbb{R}^2 : s^2+t^2 \leq 1\} \to \mathbb{R}^3$ by $g(s,t) = (s, \sqrt{1-s^2-t^2}, t)$
- $h : \mathbb{R}^3 \to \mathbb{R}$ by $h(x,y,z) = x^2+y^2+z^2$

Identify whether the set is written in parametric form, explicit form, implicit form, or none.

(a) $A = \mathrm{im}(g)$

(b) $B = h^{-1}(\{1\})$

(c) $C = \{(x,y,z) \in \mathbb{R}^3 : y = f(x,z)\}$

(d) $D = \{(x,y,z) \in \mathbb{R}^3 : y = \pm f(x,z)\}$

1.5.2 Let $S \subseteq \mathbb{R}^3$ be the upper hemisphere of radius 3 centred at the origin, so

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 9, z \geq 0\}.$$

Which of the following maps parametrize S?

(a) $g_A : [0,2\pi] \times [0,\frac{\pi}{2}] \to \mathbb{R}^3$ given by $g_A(\theta,\phi) = (3\cos\theta\sin\phi, 3\sin\theta\sin\phi, 3\cos\phi)$

(b) $g_B : [0,3] \times [0,2\pi] \times [0,\frac{\pi}{2}] \to \mathbb{R}^3$ given by $g_B(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$

(c) $g_C : [0,4\pi] \times [0,\frac{\pi}{2}] \to \mathbb{R}^3$ given by $g_C(\theta,\phi) = (3\cos\theta\sin\phi, 3\sin\theta\sin\phi, 3\cos\phi)$

(d) $g_D : [0,3] \times [0,2\pi] \to \mathbb{R}^3$ given by $g_D(r,\theta) = (r\cos\theta, r\sin\theta, \sqrt{9-r^2})$

(e) $g_E : [0,2\pi] \times [0,3] \to \mathbb{R}^3$ given by $g_E(\theta,z) = (\sqrt{9-z^2}\cos\theta, \sqrt{9-z^2}\sin\theta, z)$

(f) $g_F : \mathbb{R}^2 \to \mathbb{R}^3$ given by $g_F(x,y) = (x,y,\sqrt{9-x^2-y^2})$

Computations

1.5.3 Parametrizing sets is an essential skill throughout multivariable calculus. Algebraic manipulations are not usually enough to find them. It takes some creativity, persistence, and experience. You may need to sketch pictures and use geometry to your advantage.

(a) Parametrize the half cone $A = \{(x,y,z) \in \mathbb{R}^3 : z = \sqrt{x^2+y^2}, -1 \leq x \leq 1, -1 \leq y \leq 1\}$.

(b) Parametrize the cylinder $B = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 = 9\}$.

(c) Parametrize the paraboloid $C = \{(x,y,z) \in \mathbb{R}^3 : x = y^2+z^2, y^2+z^2 \leq 2\}$ in two ways.

(d) Parametrize the (two-sided) cone $D = \{(x,y,z) \in \mathbb{R}^3 : y^2 = x^2+z^2, -3 \leq y \leq 3\}$ in two ways.

1.5.4 Determine whether each set is written in parametric form, explicit form, implicit form, or none of them. If it is one of the three forms, prove your assertion by defining an appropriate map.

(a) $A = \{(x,y,z) \in \mathbb{R}^3 : xyz = 237\}$

(b) $B = \left\{\left(\frac{237}{yz}, y, z\right) : (y,z) \in \mathbb{R}^2, yz \neq 0\right\}$

(c) $C = \left\{(x,y,z) \in \mathbb{R}^3 : x = \frac{237}{yz}\right\}$

(d) $D = \left\{(x,y,z) \in \mathbb{R}^3 : x = \frac{237}{yz} \text{ or } y = \frac{237}{xz} \text{ or } z = \frac{237}{xy}\right\}$

1.5.5 Fix $a, b, c > 0$. Define the ellipsoid

$$S = \left\{(x,y,z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\}.$$

(a) Write S in implicit form.

(b) Write S as a union of sets in explicit form.

(c) Explain why S cannot be written in explicit form. Outline an argument but do not prove it.

(d) Write S in parametric form using a "stretched" variant of spherical coordinates.

(e) Write S in parametric form using a "stretched" variant of cylindrical coordinates.

Proofs

1.5.6 Let $S \subseteq \mathbb{R}^3$ be the sphere of radius 3 centred at the origin. Determine if the statement is true or false. If true, exhibit 2 different expressions. If not, formally state what you need to prove.

(a) The set S can be written in implicit form.

(b) The set S can be written in parametric form.

(c) The set S can be written in explicit form.

1.5.7 Sets in explicit form are rather special. Here you will study its relationship to the other 2 forms with a specific example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be any continuous map. Define the set

$$S = \{(x,y,z) \in \mathbb{R}^3 : z = f(x,y)\}.$$

Determine which statements are true or false. Prove your assertion.

(a) The set S in explicit form can also be written in parametric form.

(b) The set S in explicit form can also be written in implicit form.

1.5.8 Determine which of the following statements are true or false. If true, prove it. If false, state a counterexample without proof.

(a) If a set S can be written in explicit form, then S can be written in implicit form.

(b) If a set S can be written in implicit form, then S can be written in explicit form.

(c) If a set S can be written in explicit form, then S can be written in parametric form.

(d) If a set S can be written in parametric form, then S can be written in explicit form.

Applications and beyond

1.5.9 Define the map $g : \mathbb{R}^2 \to \mathbb{R}^3$ by $g(x,y) = (x,0,0)$. Let $S = \operatorname{im}(g) \subseteq \mathbb{R}^3$. Should S be considered a surface, i.e. a 2-dimensional manifold in $\mathbb{R}^3$? Explain in your own words.

1.5.10 The concept of "dimension" will be a key hurdle to defining manifolds. Right now, you can build some intuition by borrowing from linear algebra. There are 3 heuristic principles:

> *1) If a set can be described by n "independent" variables, then the set should presumably be a n-dimensional manifold.*

2) If a set can be described by $m-n$ "dependent" variables and n "independent" variables, then the set should presumably be a n-dimensional manifold.

3) If a set can be described by m variables and n "independent" equations, then the set should presumably be a $(m-n)$-dimensional manifold.

Let $m, n \in \mathbb{N}^+$ with $n < m$. For each set S below, identify whether it is written in parametric, explicit, or implicit form. Assuming each set is a manifold, guess its "dimension" using one of the above heuristic principles.

(a) $S = \{(xy, \sin x, y^2, x, y) : x, y \in \mathbb{R}\}$

(b) $S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}$

(c) $S = \{(x_1, x_2, \cdots, x_{n-1}, x_1 x_2 x_3 \cdots x_{n-1}) : (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}\}$

(d) There exists $U \subseteq \mathbb{R}^n$ and continuous $f : U \to \mathbb{R}^m$ such that

$$S = \{f(x_1, x_2, \ldots, x_n) : (x_1, \ldots, x_n) \in U\} = f(U).$$

(e) There exists $V \subseteq \mathbb{R}^n$ and continuous $h : V \to \mathbb{R}^{m-n}$ such that

$$S = \{(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_m) \in \mathbb{R}^m : (x_{n+1}, \ldots, x_m) = h(x_1, \ldots, x_n), (x_1, \ldots, x_n) \in V\}.$$

(f) There exists $W \subseteq \mathbb{R}^m$ and continuous $g : W \to \mathbb{R}^{m-n}$ such that

$$S = \{(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_m) \in W : g(x_1, \ldots, x_m) = \underbrace{(0, 0, \ldots, 0)}_{m-n \text{ times}}\}.$$

1.6. Dimension reductions

This section is about maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

when $n > m$. These can be used to write sets in implicit form, but they have other uses. While these maps do not usually have their own name, they are sometimes referred to as **dimension reductions**[16], because they can push higher dimensional objects into a lower dimensional space. Large datasets in high dimensions are widespread nowadays, appearing in problems for signal processing, machine learning, data science, and bioinformatics. Unfortunately, high-dimensional spaces are difficult and computationally expensive to analyze. Dimension reduction transforms data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains meaningful properties of the original data.

It has widespread scientific applications and some classical mathematical examples, which you will briefly explore in this section. As usual, the special case $\mathbb{R}^3 \to \mathbb{R}^2$ has visual advantages.

Example 1.6.1 Creating a map of the Earth is a classical projection problem. You must push a sphere, an object in $\mathbb{R}^3$, into a rectangle, an object in $\mathbb{R}^2$. The *Mercator projection*, invented in 1569, is the modern standard, which is still used in Google Maps to this day[17].

This dimension reduction preserves local directions so its popularity rose due to its use for ocean navigation. However, by reducing dimensions, it loses some geometric information. The Mercator projection distorts distances dramatically. Greenland and the African continent appear to be the same size on the map, but Africa is actually 14 times larger!

Linear projections are classic mathematical examples of dimension reductions.

Example 1.6.2 For $i \in \{1,\dots,n\}$, the map $\pi_i : \mathbb{R}^n \to \mathbb{R}$ given by

$$\pi_i(x_1,\dots,x_n) = x_i$$

is the i^{th} **coordinate map**. These are convenient for proofs when you are trying to express some quantities in terms of continuous maps. Notice π_i is a linear transformation.

[16] This linguistic choice is not standard in mathematics. However, it can be found in other scientific areas such as statistics, computer graphics, and data science.

[17] Images retrieved from Wikimedia Commons (Earth and Mercator projection) on 2024-07-23 licensed under PD (Earth) and CC BY-SA (Mercator projection).

Remark 1.6.3 The Greek letter π is often used for projections. Since π is used for other reasons, this conflict is an *abuse of notation*. You are permitted to abuse notation provided the context makes your notation unambiguous.

Example 1.6.4 For $i \in \{1, \ldots, n\}$, the map $\Pi_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ given by

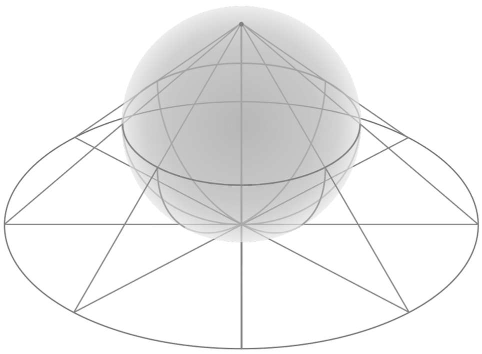
$$\Pi_i(x_1, \ldots, x_n) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

is the i^{th} **coordinate plane projection**. These are also linear transformations which are convenient for proofs. They also have a natural interpretation as the "shadow" of an object.

In particular, for $n = 3$, the map $\Pi_3(x, y, z) = (x, y)$ is also referred to as the **projection into the xy-plane**. The image of the paraboloid $P = \{(x, y, z) : z = x^2 + y^2 \leq 1\}$ under Π_3 is the unit disk $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ in $\mathbb{R}^2$. Informally speaking, Π_3 produces the "shadow" of an object in $\mathbb{R}^3$ when viewed from above.

Both of these examples are linear projections. Nonlinear projections are more intricate.

Example 1.6.5 The unit sphere S^2 in $\mathbb{R}^3$ can be projected onto the plane in many ways. A famous way is *stereographic projection*. Each point (x, y, z) on the sphere except $(0, 0, 1)$ is mapped onto the plane $z = -1$ by drawing a line through the north pole $(0, 0, 1)$ and the point (x, y, z). The line intersects the plane $z = -1$ at a point $(\pi(x, y, z), -1) \in \mathbb{R}^3$ for some function $\pi : S^2 \setminus \{(0, 0, 1)\} \to \mathbb{R}^2$. This process is illustrated in the left diagram below[18].



It is left as an exercise to find an algebraic expression for this projection π. The formula is actually quite simple. Stereographic projection is used in some fisheye lenses for wide-angle views. The right hand image is an example of this fun photography technique.

This concludes our study of maps $\mathbb{R}^n \to \mathbb{R}^m$. You have explored examples from across the spectrum: parametric curves ($n = 1$); real-valued functions ($m = 1$); vector fields, transformations, and coordinate systems ($m = n$); sets written in parametric or explicit form ($n < m$); and sets written in implicit form, or dimension reductions ($n > m$). Each one has a vast jungle of applications and mysterious phenomena that warrant further investigation. This motivational introduction sets the stage for building a rich mathematical theory. Topology will be the next step in your multivariable calculus journey.

[18]Images retrieved from Wikimedia Commons (stereographic projection and fisheye) on 2024-07-23 licensed under PD (stereographic projection) and CC BY (fisheye).

Exercises for Section 1.6

Concepts and definitions

1.6.1 You learned about the Mercator projection of the Earth which is a map $\mathbb{R}^3 \to \mathbb{R}^2$. What information does this dimension reduction retain? What information does it lose?

1.6.2 Your friend says

> *"I have a solid object S which looks like a 2×2 square centered at the origin when viewed from above."*

Formally interpret your friend's claim using a linear projection. *Hint:* First, define the projection π associated with "viewed from above". Then write $\pi(S)$ in set builder notation.

Computations

1.6.3 Let $S \subseteq \mathbb{R}^3$ be the sphere of radius 1 centered at the origin. Let $\widetilde{S} = S \setminus \{(0,0,1)\}$. The stereographic projection $\pi : \widetilde{S} \to \mathbb{R}^2$ of the sphere onto the plane is defined as:

> *For every $(x,y,z) \in \widetilde{S}$, let $\pi(x,y,z)$ be the straight line projection from the north pole through (x,y,z) onto the plane tangent to the bottom of the sphere.*

This projection preserves angles between curves on a sphere but distorts distances. You will transform this geometric description into an algebraic one.

(a) Let $p = (x,y,z) \in \widetilde{S}$. Define the straight line $\gamma_p : \mathbb{R} \to \mathbb{R}^3$ going through p and $(0,0,1)$.

(b) Find the point $q \in \mathbb{R}^3$ at which γ_p intersects the plane tangent to the bottom of the sphere.

(c) Give an algebraic definition for $\pi : \widetilde{S} \to \mathbb{R}^2$.

1.6.4 Let $m, n \in \mathbb{N}^+$ and $n > m$. A **linear projection** is a map $\pi : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$\forall x \in \mathbb{R}^n,\ \pi(x) = Ax$$

for some $m \times n$ matrix A. Linear projections are the simplest and most common kind of dimension reduction. This sequence of exercises is good review for some linear algebra facts.

(a) Is π necessarily injective? If not, what are the possible values for its nullity?

(b) Is π necessarily surjective? If not, what are the possible values for its rank?

(c) Let $\pi : \mathbb{R}^n \to \mathbb{R}^m$ denote the projection onto the first m coordinates of $\mathbb{R}^n$. Write an explicit formula for π and describe its associated matrix.

2. Topology

You have explored some examples of maps in higher dimensions but this only represents the tip of an iceberg. Hopefully the sample of topics has piqued your interest since it will take time to deeply explore some of the exciting questions raised by the first chapter. Now, you begin the journey of building a modern rigorous foundation for calculus in higher dimensions.

As you develop the theory of functions $\mathbb{R}^n \to \mathbb{R}^m$, you will need to more thoroughly understand n-dimensional space

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{R}\}$$

and rigorously define the notion of limits in this space. There will be some recurring themes:

- Simple definitions in $\mathbb{R}$ can often be naturally generalized to $\mathbb{R}^n$.
- Theorems over $\mathbb{R}$ can be quickly transferred to theorems over $\mathbb{R}^n$.
- New subtle properties arise in $\mathbb{R}^n$ that were indistinguishable or trivial in $\mathbb{R}$.

Pay attention to these features as you continue to dig deeper.

After some substantial efforts, you will finally be able to explore the first profound question of multivariable calculus.

For a given optimization problem, when does a solution exist?

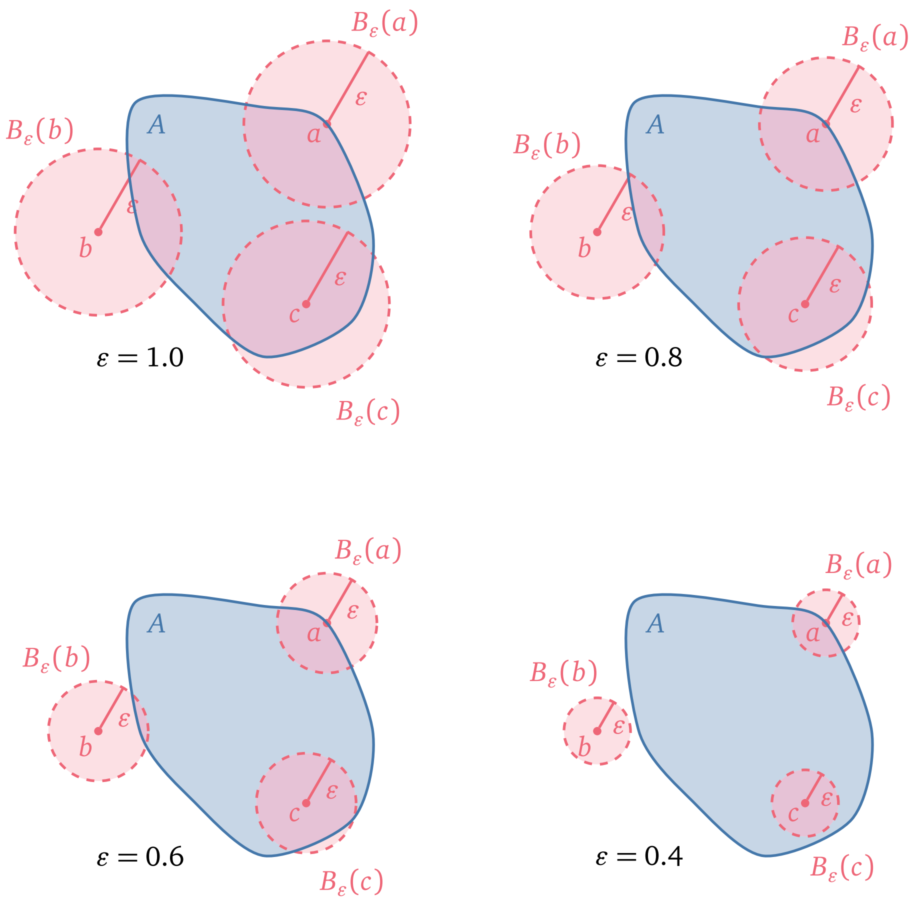
Your careful work with the topology of $\mathbb{R}^n$ will produce a powerful theorem that resolves this question for a wide range of optimization problems. The end of this chapter will therefore mark the beginning of your journey with optimization. Onward!

2.1. Interior, boundary, and closure

A mango has an inside (the juicy fruit), an outside (the air), and an edge (the skin). Given a solid object in 3-dimensions, you have an intuitive physical understanding of its inside, outside, and edge. This creates an immediate theoretical obstacle.

> *How can you formalize the ideas of "inside", "outside", and "edge" for any set in $\mathbb{R}^n$?*

The simple yet brilliant solution will be to *zoom in* near a point. The rigorous formulation of this technique will be achieved with shrinking open balls. Illustrated below is a set $A \subseteq \mathbb{R}^2$ and three different points $a, b, c \in \mathbb{R}^2$ pictured four times. Each time, balls with smaller and smaller radii are drawn at each point.



This picture will serve as the inspiration for nearly all the definitions that you will discover in this section. As you read each definition, you are encouraged to refer back to this picture and reflect on how a new topological concept relates to it.

2.1.1 Interior

A point p should be "inside" a region A if, once you zoom in close enough on p, all you see is the region A. This idea can be formalized using a sufficiently small open ball.

Definition 2.1.1 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is an **interior point** of A if there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq A$.

Example 2.1.2 The point $p = 2$ is an interior point of the interval $A = [1,4)$ since $B_{1/2}(2) = (1.5, 2.5)$ is a subset of A. Intuitively, you can see that 5 is not an interior point of A. Proving this formally requires verifying the negation:

$$\forall \varepsilon > 0, B_\varepsilon(5) = (5-\varepsilon, 5+\varepsilon) \not\subseteq [1,4) = A.$$

Fix $\varepsilon > 0$. Set $x = 5$. Then $x \in (5-\varepsilon, 5+\varepsilon)$ yet $x \notin [1,4)$. Hence, $B_\varepsilon(5) \not\subseteq [1,4)$ so 5 is not an interior point of A.

You can also check that the endpoint 1 is not an interior point of $[1,4)$. You can follow the same argument but your choice of x will be different since $1 \in [1,4)$. In particular, x should depend on ε.

Example 2.1.3 Let $c, d \in \mathbb{R}$ with $c < d$. Let $p \in \mathbb{R}$. If $c < p < d$ then you can prove that p is an interior point of the closed interval $[c,d]$. This builds upon the previous example and is left as an exercise.

Proving an open ball is a subset of another set is trickier in higher dimensions.

Example 2.1.4 The point $p = (1,0)$ is an interior point of the left halfplane

$$A = \{(x,y) \in \mathbb{R}^2 : x \leq 2\}.$$

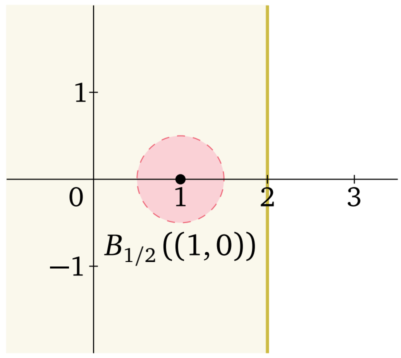
How do you prove this by definition? As the picture below illustrates, it suffices to show that the open ball $B_{1/2}((1,0))$ is a subset of A. Formally, this means you must prove that for all points $(x,y) \in \mathbb{R}^2$,

$$(x,y) \in B_{1/2}((1,0)) \implies (x,y) \in A.$$

Proof. Let $(x,y) \in B_{1/2}((1,0))$ so $(x-1)^2 + y^2 < (1/2)^2$. Since $y^2 \geq 0$, this implies that

$$(x-1)^2 < \left(\frac{1}{2}\right)^2 \implies x-1 < \frac{1}{2} \implies x < 1+\frac{1}{2} \leq 2.$$

Thus, $(x,y) \in A$ as required. Therefore, $(1,0)$ is an interior point of A. ■



The definition of interior point leads to the concept of the "inside" of a set.

Definition 2.1.5 Let $A \subseteq \mathbb{R}^n$ be a set. The **interior** of A, denoted A^o or int(A), is the set of interior points of A.

Remark 2.1.6 Other texts use $\mathring{A}$ or A^{int} to denote the interior.

Example 2.1.7 The interior of $A = [1,4)$ is the open interval $A^o = (1,4)$. First, you can prove that $(1,4) \subseteq A^o$ or, more formally,

$$\forall x \in (1,4), \exists \varepsilon > 0 \text{ s.t. } (x-\varepsilon, x+\varepsilon) \subseteq [1,4).$$

Let $x \in (1,4)$. Set $\varepsilon = \min\{\frac{x-1}{2}, \frac{4-x}{2}\} > 0$. As $x > 1$ and $\varepsilon \leq \frac{x-1}{2}$, it follows that

$$x - \varepsilon \geq x - \frac{x-1}{2} = \frac{x+1}{2} > \frac{1+1}{2} = 1.$$

Similarly, $x < 4$ and $\varepsilon \leq \frac{4-x}{2}$ implies that

$$x + \varepsilon \leq x + \frac{4-x}{2} = \frac{4+x}{2} < \frac{4+4}{2} = 4.$$

Overall, this implies that $(x-\varepsilon, x+\varepsilon) \subseteq (1,4)$ as desired.

It remains to prove that $A^{\circ} \subseteq (1,4)$. Since $A^{\circ} \subseteq A = [1,4)$ (see Lemma 2.1.11 below), it suffices to prove that $1 \notin A^{\circ}$. Let $\varepsilon > 0$ arbitrary. Take $x = 1 - \frac{\varepsilon}{2}$ so $x \in (1-\varepsilon, 1+\varepsilon)$ yet $x \notin [1,4)$. As $\varepsilon > 0$ was arbitrary, this proves that $1 \notin A^{\circ}$ by definition, completing the proof.

Example 2.1.8 The interior of $A = \{(x,y) \in \mathbb{R}^2 : x \leq 2\}$ is the set $\{(x,y) \in \mathbb{R}^2 : x < 2\}$. The proof is left as an exercise. You will need to combine the ideas of Examples 2.1.4 and 2.1.7.

Example 2.1.9 Here is a list of examples of interiors of sets. Let $a \in \mathbb{R}^n$ and $r > 0$.
- The interior of $\mathbb{R}^n$ is $\mathbb{R}^n$.
- The interior of any finite set is empty.
- The interior of the open ball $B_r(a)$ is the open ball itself.
- The interior of the closed ball $\{x \in \mathbb{R}^n : ||x-a|| \leq r\}$ is the open ball $B_r(a)$.
- The interior of any sphere is empty.
- The interior of the hypercube $[c,d]^n$ is the open hypercube $(c,d)^n$, where $c < d$.

Verifying these statements (even in special cases like $n = 1$) are excellent exercises.

Example 2.1.10 Not all sets have nice geometric interpretations. The set of rational numbers, $\mathbb{Q}$, is quite a nightmare. What is its interior? It is *empty*! A fully formal proof is reserved for a course in topology or analysis[1], but here is a nearly complete argument. Let $q \in \mathbb{Q}$ so q is a rational number. Fix $\varepsilon > 0$. The interval $B_\varepsilon(q) = (q-\varepsilon, q+\varepsilon)$ must contain irrational numbers, so $B_\varepsilon(q)$ cannot be a subset of $\mathbb{Q}$. Thus, q is not an interior point and the interior of $\mathbb{Q}$ is empty.

Sets like $\mathbb{Q}$ are notorious counterexamples to many statements that you shall later see, so it is good to keep in your back pocket. That said, these types of insidious examples are not the focus of this text, so you will often wave your hands and assume various properties about them without a formal proof.

The interior satisfies natural properties with respect to other set operations. These should not be memorized; instead, sketch some pictures in $\mathbb{R}^2$ to convince yourself why each of them should be true.

Lemma 2.1.11 Let A and B be sets of $\mathbb{R}^n$. Then
(a) $A^{\circ} \subseteq A$
(b) $(A \cup B)^{\circ} \supseteq A^{\circ} \cup B^{\circ}$
(c) $(A \cap B)^{\circ} = A^{\circ} \cap B^{\circ}$
(d) $(A \times B)^{\circ} = A^{\circ} \times B^{\circ}$

Proof. This is left as an exercise. ■

With the "inside" of a set formalized as the interior, you can study the "edge" of a set.

[1] See a course in topology or an introduction to analysis. Look for "irrationals (and rationals) are dense in $\mathbb{R}$".

2.1.2 Boundary

A point p should be at the "edge" of a region A if, no matter how close you zoom into p, you can see points nearby inside A and outside A. This idea can be formalized as follows.

Definition 2.1.12 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is a **boundary point** of A if for every $\varepsilon > 0$, the sets $B_\varepsilon(p) \cap A$ and $B_\varepsilon(p) \cap A^c$ are both non-empty.

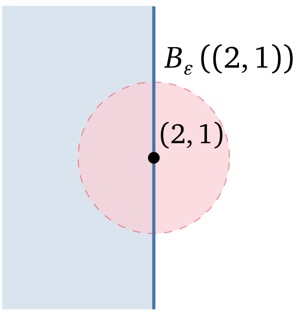
Example 2.1.13 The point $p = 1$ is a boundary point of the interval $A = [1,4)$. Fix $\varepsilon > 0$. Define the points $x = 1$ and $y = 1 - \frac{\varepsilon}{2}$ so both $x, y \in B_\varepsilon(1) = (1-\varepsilon, 1+\varepsilon)$. Moreover, $x \in [1,4)$ and $y \notin [1,4)$ so $x \in A$ and $y \in A^c$. Therefore, $x \in B_\varepsilon(1) \cap A$ and $y \in B_\varepsilon(1) \cap A^c$ so both sets are non-empty. Thus, 1 is a boundary point of $[1,4)$.

On the other hand, the point $q = 2$ is not a boundary point of the interval $A = [1,4)$ because $B_{1/2}(2) = (1.5, 2.5)$ is disjoint from the complement A^c.

Example 2.1.14 The point $p = (2,1)$ is a boundary point of the set $A = \{(x,y) \in \mathbb{R}^2 : x \leq 2\}$. To prove this, fix $\varepsilon > 0$. Then

$$B_\varepsilon(p) = \{(x,y) \in \mathbb{R}^2 : (x-2)^2 + (y-1)^2 < \varepsilon^2\}$$

so $(2,1) \in B_\varepsilon(p)$ and $(2+\frac{\varepsilon}{2}, 1) \in B_\varepsilon(p)$. By definition of A, notice $(2,1) \in A$ and $(2+\frac{\varepsilon}{2}, 1) \notin A$. Therefore, $B_\varepsilon(p) \cap A$ and $B_\varepsilon(p) \cap A^c$ respectively contain $(2,1)$ and $(2+\frac{\varepsilon}{2}, 1)$, so both sets are non-empty as required.



Boundary points produce the concept for the "edge" of a set.

Definition 2.1.15 Let $A \subseteq \mathbb{R}^n$ be a set. The **(topological) boundary** of A, denoted ∂A, is the set of boundary points of A.

Remark 2.1.16 The word 'topological' will almost always be omitted when referring to the boundary of a set. It is included here to distinguish this notion from the *relative* boundary of a surface; that definition that will be introduced much later.

On another note, the symbol ∂ is called 'curly dee' or 'del'. It goes by many other names (see Wikipedia) and will appear again when defining partial derivatives. There are very good reasons for this abuse of notation, but the explanation is a bit too long to tell now.

Example 2.1.17 The boundary of $A = [1,4)$ is the pair of endpoints $\partial A = \{1,4\}$.

Example 2.1.18 The boundary of the left half plane $A = \{(x,y) \in \mathbb{R}^2 : x \leq 2\}$ is the vertical line $\partial A = \{(2,y) : y \in \mathbb{R}\}$.

Example 2.1.19 Here is a list of examples of boundaries of sets. Let $a \in \mathbb{R}^n$ and $r > 0$.

- The boundary of $\mathbb{R}^n$ is empty.
- The boundary of any finite set A is A itself.
- The boundary of the closed interval $[c, d]$ is the finite set $\{c, d\}$.
- The boundary of the open ball $B_r(a)$ is the sphere $\partial B_r(a) = \{x \in \mathbb{R}^n : ||x - a|| = r\}$
- The boundary of $B_1(0)$ is the unit sphere $\partial B_1(0) = S^{n-1}$.
- The boundary of any sphere is the sphere itself.

You should verify these properties both intuitively and formally. Use $n = 2$.

Example 2.1.20 Return to your rational nightmare, $\mathbb{Q}$. What is its boundary? It is all of $\mathbb{R}$! Again, a fully formal proof is reserved for a course in topology or analysis, but here's a nearly complete proof. Fix $x \in \mathbb{R}$ and $\varepsilon > 0$. The open interval $B_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$ contains both rational and irrational points since $\mathbb{Q}$ and $\mathbb{Q}^c = \mathbb{R} \setminus \mathbb{Q}$ are "dense" in $\mathbb{R}$. In other words, $B_\varepsilon(x) \cap \mathbb{Q}$ and $B_\varepsilon(x) \cap \mathbb{Q}^c$ are both non-empty so $x \in \partial Q$. Since x was arbitrary, this proves that $\partial\mathbb{Q} = \mathbb{R}$. Pause for a moment to notice how really strange this is. The set of rationals $\mathbb{Q}$ has empty interior, yet its boundary is everything!

The boundary also satisfies natural properties with respect to many set operations. As before, these are not worth memorizing. If you can re-discover them yourself by sketching some pictures in $\mathbb{R}$ or $\mathbb{R}^2$, then you will be able to reproduce them on your own.

Lemma 2.1.21 Let A and B be sets in $\mathbb{R}^n$. Then

(a) $\partial(A \cup B) \subseteq \partial A \cup \partial B$
(b) $\partial(A \cap B) \subseteq \partial A \cup \partial B$
(c) $\partial(A \times B) = (\partial A \times B) \cup (A \times \partial B) \cup (\partial A \times \partial B)$

Proof. These are left as exercises.[2] ■

Now, for a given set $S \subseteq \mathbb{R}^n$, the "inside" is the interior S°, the "edge" is the boundary ∂S, and the "outside" is the interior of the complement $(\mathbb{R}^n \setminus S)^\circ$. This addresses the original motivation of this section, but there is a closely related concept you need to investigate.

2.1.3 Closure

The sets you must deeply understand are the domains of functions because calculus is all about the limit of function values. Defining limits on a domain in $\mathbb{R}^n$ presents a new issue. Over $\mathbb{R}$, you may have been satisfied with taking a limit of a function at a point p provided an open interval of p, like $(p - \varepsilon, p + \varepsilon)$, belonged to the domain of the function. This simplicity in one dimension exists because there are only two ways of approaching a point on the real number line: from the left or from the right.

This feature abruptly changes in $\mathbb{R}^n$ for $n \geq 2$. Suddenly, you can approach a point in infinitely many ways. If the domain of your function is a complicated set A in $\mathbb{R}^n$, how do you decide where you can take limits of your function? You must determine which points p in $\mathbb{R}^n$ can be approached by points only from A. Intuitively, a point p can be "approached" via a region A if, no matter how close you zoom into p, you can see points of A not including p itself. This idea needs to be formalized if you want to rigorously study limits.

[2] It might be both comical and frustrating how often statements are left as an exercise. Rest assured that this choice is pedagogical and not due to laziness. These are genuinely excellent exercises and it would spoil your learning to reveal them. Also, some proofs can be quite technical and not so enlightening. When a statement is left as an exercise, it is better to try and heuristically convince yourself of the statement, e.g. with a drawing or a special case; a formal proof is a more advanced step which you can take after completing basic exercises.

Definition 2.1.22 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is a **limit point** of A if for every $\varepsilon > 0$, the set $B_\varepsilon(p) \setminus \{p\}$ contains points in A.

Example 2.1.23 The point $p = 2$ is a limit point of $A = [1,4)$. To prove this, fix $\varepsilon > 0$. Set $x = \max\{1, 2 - \frac{\varepsilon}{2}\}$ so $x \in B_\varepsilon(2) = (2-\varepsilon, 2+\varepsilon)$ yet $1 \leq x < 2$. Hence, $x \in [1,4)$ so the set $B_\varepsilon(2) \setminus \{2\}$ contains points in $[1,4)$, establishing that 2 is a limit point of $[1,4)$. A similar argument can also show that 1 or 4 is a limit point of A.

Example 2.1.24 The point $(0,0)$ is not a limit point of

$$A = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(0,0)\}$$

because the punctured open disk

$$B_{1/2}((0,0)) \setminus \{(0,0)\} = \{(x,y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < (1/2)^2\}$$

does not contain any points of A. The set of limit points of A is $\{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\}$.

Example 2.1.25 Here is a list of examples of sets of limit points. Let $a \in \mathbb{R}^n$ and $r > 0$.
- The set of limit points of $\mathbb{R}^n$ is $\mathbb{R}^n$.
- The set of limit points of any finite set is empty.
- The set of limit points of the open ball $B_r(a)$ is the closed ball $\{x \in \mathbb{R}^n : ||x - a|| \leq r\}$
- The set of limit points of any sphere is the sphere itself.
- The set of limit points of any (closed) rectangle is the rectangle itself.

You should verify these properties both intuitively and formally. Use $n = 2$ if it helps.

Now, as demonstrated by Example 2.1.24, the set of limit points of a set $A \subseteq \mathbb{R}^n$ does not actually contain A itself. This caveat will be a bit of a nuisance when A is the domain of a function, since you may want to evaluate a function at all points of A. You can resolve this issue with a simple definition.

Definition 2.1.26 Let $A \subseteq \mathbb{R}^n$. The **closure** of A, denoted $\overline{A}$ or cl(A), is the union of the set A and the set of limit points of A.

Remark 2.1.27 The set of limit points of A is often denoted A^* so $\overline{A} = A \cup A^*$. Other texts may use A' for the set of limit points though you will not see that convention here.

Example 2.1.28 The closure of $A = [1,4)$ is the set $\overline{A} = [1,4]$ since its limit points are $A^* = [1,4]$. On the other hand, the closure of $B = [1,4) \cup \{5\}$ is the set $\overline{B} = [1,4] \cup \{5\}$ even though its set of limit points $B^* = [1,4]$ is the same as A^*.

Example 2.1.29 The closure of the set

$$A = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(0,0)\}$$

is the set

$$\overline{A} = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\} \cup \{(0,0)\}.$$

Example 2.1.30 Here is a list of examples of closures. Let $a \in \mathbb{R}^n$ and $r > 0$.
- The closure of $\mathbb{R}^n$ is $\overline{\mathbb{R}^n} = \mathbb{R}^n$.

- The closure of any finite set is the set itself.
- The closure of the open ball $B_r(a)$ is the closed ball $\overline{B_r(a)} = \{x \in \mathbb{R}^n : ||x - a|| \leq r\}$.
- The closure of any sphere is the sphere itself.
- The closure of any (closed) rectangle is the rectangle itself.
- The closure of $\mathbb{Q}$ is $\overline{\mathbb{Q}} = \mathbb{R}$.

As usual, verify these properties both intuitively and formally. Use $n = 2$ if it helps.

Again, closure respects natural set operations.

Lemma 2.1.31 Let A and B be sets in $\mathbb{R}^n$. Then
(a) $A \subseteq \overline{A}$
(b) $\mathrm{cl}(A \cup B) = \overline{A} \cup \overline{B}$
(c) $\mathrm{cl}(A \cap B) \subseteq \overline{A} \cap \overline{B}$
(d) $\mathrm{cl}(A \times B) = \overline{A} \times \overline{B}$

Proof. This is left as an exercise. You may find some easier to prove once you are equipped with an alternate definition of limit points from Section 2.2. ■

Despite starting with independent definitions, the interior, boundary, and closure of a set possess all the relationships that you might guess.

Theorem 2.1.32 For any set $A \subseteq \mathbb{R}^n$, all of the following hold:

$$A^{\circ} \subseteq A \subseteq \overline{A}, \qquad A^{\circ} \cap \partial A = \varnothing, \qquad \overline{A} = A^{\circ} \cup \partial A, \qquad \partial A = \overline{A} \setminus A^{\circ}.$$

Proof. The first statement follows from Lemmas 2.1.11 and 2.1.31. The second statement follows quickly from the definitions of interior and boundary, so it is left as an exercise. The fourth statements follows quickly from the second and third as well as basic properties of sets. It remains to prove the third statement. We proceed by proving both set containments.

We begin by proving that $A^{\circ} \cup \partial A \subseteq \overline{A}$. Since $A^{\circ} \subseteq A \subseteq \overline{A}$, it suffices to show that $\partial A \setminus A \subseteq \overline{A}$. Fix $p \in \partial A$ with $p \notin A$. This implies for every $\varepsilon > 0$, the set $B_\varepsilon(p) \setminus \{p\}$ contains points of A. Therefore, p is a limit point of A and hence $p \in \overline{A}$ by definition. This proves $A^{\circ} \cup \partial A \subseteq \overline{A}$. We next prove that $\overline{A} \subseteq A^{\circ} \cup \partial A$ by showing $A \subseteq A^{\circ} \cup \partial A$ and $A^* \setminus A \subseteq \partial A$.

Fix $p \in A$. If there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq A$ then $p \in A^{\circ}$ as required. Otherwise, for every $\varepsilon > 0$, the ball $B_\varepsilon(p)$ is not a subset of A, but $B_\varepsilon(p) \cap A$ is non-empty as it contains p. Therefore, for every $\varepsilon > 0$, both $B_\varepsilon(p) \cap A$ and $B_\varepsilon(p) \cap A^c$ are non-empty so $p \in \partial A$ by definition. This proves $A \subseteq A^{\circ} \cup \partial A$.

Fix $q \in A^*$ with $q \notin A$. By definition of a limit point, for every $\varepsilon > 0$, the punctured open ball $B_\varepsilon(q) \setminus \{q\}$ contains points in A in which case $B_\varepsilon(q) \cap A$ is non-empty. Moreover, for every $\varepsilon > 0$, the set $B_\varepsilon(q) \cap A^c$ is non-empty since $q \in A^c$. By definition, $q \in \partial A$ which proves that $A^* \setminus A \subseteq \partial A$ as desired. ■

Theorem 2.1.32 is an elegant collection of relationships between these foundational topological concepts in $\mathbb{R}^n$. It represents a triumph of visual intuition and excellent definition crafting. Now, these definitions were described in terms of open balls, but there is another equally valuable perspective in terms of sequences. Mathematics thrives on building multiple equivalent perspectives, so sequences are next on your topology bucket list.

Exercises for Section 2.1

Concepts and definitions

2.1.1 Let $A \subseteq \mathbb{R}^n$ be a set and let $p \in \mathbb{R}^n$ be a point. Recall A^o is the interior of A, ∂A is the boundary of A, A^* is the set of limit points of A, $\overline{A}$ is the closure of A, and A^c is the complement of A.

I.	$p \in A^o$	III.	$p \in \partial A$	V.	$p \in A^*$	VII.	$p \in \overline{A}$
II.	$p \notin A^o$	IV.	$p \notin \partial A$	VI.	$p \notin A^*$	VIII.	$p \notin \overline{A}$

Each of statement below is *equivalent* to exactly one of statement above. Match them.

(a) $p \in A^o \cup \partial A$

(b) $p \notin A^* \cup A$

(c) $\exists \varepsilon > 0$ s.t. $B_\varepsilon(p) \subseteq A$

(d) $\exists \varepsilon > 0$ s.t. $B_\varepsilon(p) \subseteq A$ or $B_\varepsilon(p) \subseteq A^c$

(e) $\exists \varepsilon > 0$ s.t. $(B_\varepsilon(p) \setminus \{p\}) \cap A = \varnothing$

(f) $\forall \varepsilon > 0, B_\varepsilon(p) \cap A^c \neq \varnothing$

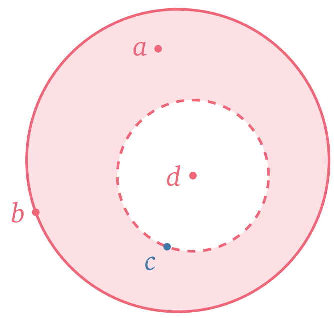
(g) $\forall \varepsilon > 0, (B_\varepsilon(p) \setminus \{p\}) \cap A \neq \varnothing$

(h) $\forall \varepsilon > 0, B_\varepsilon(p) \cap A \neq \varnothing$ and $B_\varepsilon(p) \cap A^c \neq \varnothing$

2.1.2 Consider the following set $S \subseteq \mathbb{R}^2$ and the points a, b, c, d. **Assume** $a, b, d \in S$ **and** $c \notin S$.

(a) For each of the four points a, b, c, d, identify whether it is an interior point, a boundary point, or a limit point of S. Identify all that apply.

(b) Sketch the interior of S, the boundary of S, and the closure of S.



2.1.3 For this question, use the diagram at the beginning of Section 2.1 with four pictures of a set $A \subseteq \mathbb{R}^2$ and three points $a, b, c \in \mathbb{R}^2$. Assume $a, c \in A$ but $b \notin A$. Remember each picture draws an ε-ball around each points with different values of ε.

(a) The point a is a boundary point of A. Which value(s) of ε prove this claim?

(b) The point b is not a boundary point of A. Which value(s) of ε prove this claim?

(c) The point c is an interior point of A. Which value(s) of ε prove this claim?

(d) The point a is not an interior point of A. Which value(s) of ε prove this claim?

(e) The point a is a limit point of A. Which value(s) of ε prove this claim?

(f) The point b is not a limit point of A. Which value(s) of ε prove this claim?

2.1.4 For each of the following sets, write their interior S^o, their boundary ∂S, their set of limit points S^*, and their closure $\overline{S}$ using set builder notation or canonical sets. No justification is required.

(a) $S = [137, 237]$

(b) $S = B_r(a)$ where $a \in \mathbb{R}^n$ and $r > 0$]

(c) $S = B_r(a) \setminus \{a\}$

(d) $S = \mathbb{R}^n$

(e) $S = \mathbb{Z}^n$

(f) $S = \mathbb{Q}^n$

2.1.5 Let $A \subseteq \mathbb{R}^n$ be a set. Let $p \in \mathbb{R}^n$ be a point. Which of the following statements are true or false? If true, justify it with a lemma or short proof. If false, give a counterexample.

(a) If p is not an interior point of A then p does not belong to A.

(b) If p is a boundary point of A then p is in the closure of A.

(c) If p is a limit point of A but not an interior point, then p is a boundary point of A.

(d) If p belongs to A then p is a limit point of A.

Proofs

2.1.6 The essence of many proofs in higher dimensions can be discovered with an illustrative sketch, so you should practice drawing a "picture proof" whenever possible. It helps to discover the key ideas. Consider the upper half plane $S = \{(x, y) \in \mathbb{R}^2 : y > -1\}$.

(a) Write the formal open ball definition of "$(2,0)$ is an interior point of S."

(b) Draw a "picture proof" that $(2,0)$ is an interior point of S. Label it using notation from your formal statement.

(c) Prove that $(2,0)$ is an interior point of S. Use your "picture proof" to guide you.

2.1.7 Define again the upper half plane $S = \{(x, y) \in \mathbb{R}^2 : y > -1\}$.

(a) Below is an incorrect proof that $(2,-1)$ is a boundary point of S.

> 1. *Set* $\varepsilon = 0.5$.
> 2. *Define the points* $a = (2,-0.75)$ *and the point* $b = (2,-1.25)$, *so* $a \in S$ *and* $b \in S^c$.
> 3. *Notice* $||a-(2,-1)|| = ||b-(2,-1)|| = 0.25 < 0.5 = \varepsilon$ *so* $a, b \in B_\varepsilon((2,-1))$.
> 4. *Thus, there is an open ball at* $(2,-1)$ *with points in* S *and in* S^c.
> 5. *This implies the point* $(2,-1)$ *is a boundary point of* S.

One line has a critical flaw. Identify the line and explain what is wrong with the proof.

(b) Prove that $(2,-1)$ is a boundary point of S. Include a “picture proof” of your argument.

2.1.8 Let $A \subseteq \mathbb{R}^n$. Prove that every interior point of A is also a limit point of A.

2.1.9 The interior, boundary, and closure of a set have some fundamental relationships. Theseusually follow in a few lines from the definitions. You establish a couple here. Let $A \subseteq \mathbb{R}^n$ be a set.

(a) Prove that $A^\circ \subseteq A \subseteq \overline{A}$.

(b) Prove that $A^\circ \cap \partial A = \varnothing$.

2.1.10 Interiors, boundaries, and closures respect many set operations. You can practice proving almost any of them. Only a couple are highlighted here. Let $A, B \subseteq \mathbb{R}^n$ be sets.

(a) Prove that $(A \cup B)^\circ \supseteq A^\circ \cup B^\circ$.

(b) Prove that $(A \cap B)^\circ = A^\circ \cap B^\circ$.

2.1.11 Let $S \subseteq \mathbb{R}^n$. A point p is an **isolated point** of S if $p \in S$ and p is not a limit point.

(a) Write the definition of an isolated point using quantifiers.

(b) Which of the following sets contain isolated points? No justification is necessary.

$$A = [-\pi, \pi] \qquad B = B_1(0) \setminus \{0\} \qquad C = B_1((0,0)) \cup \{(2,2)\} \qquad D = \mathbb{Z}^n \qquad E = \mathbb{Q} \times \mathbb{Z}$$

(c) Prove that the set of isolated points of S is $\partial S \setminus S^*$, where S^* is the set of limit points of S.

2.1.12 Here is a terse proof of one direction of Theorem 2.1.32, namely that $A^o \cup \partial A \subseteq \overline{A}$.

> 1. *Let* $p \in A^o \cup \partial A$.
> 2. *Since* $A \subseteq \overline{A}$, *it suffices to assume that* $p \notin A$.
> 3. *If* $p \notin A$ *then* $p \notin A^o$ *so* $p \in \partial A$.
> 4. *Thus, for every* $\varepsilon > 0$, *the set* $B_\varepsilon(p) \setminus \{p\}$ *contains points of* A.
> 5. *This implies* $p \in \overline{A}$ *and therefore* $A^o \cup \partial A \subseteq \overline{A}$.

The proof is essentially correct, but you must fill in more details and clarify some items.

(a) Why does the author say *"it suffices to assume"* in line 2? Explain what this means.

(b) How does line 4 follow from the previous lines? Fill in the details.

(c) How does line 5 follow from line 4? Fill in the details.

2.2. Sequences

The limit is the fundamental notion which drives calculus. You have seen a description of limit points in terms of open balls. There are equivalent definitions with sequences, and these will prove to be quite useful for practical and theoretical purposes. It is also an intuitive perspective.

Definition 2.2.1 A **sequence** in $\mathbb{R}^n$ is a function with domain $\{k \in \mathbb{Z} : k \geq k_0\}$ for some fixed $k_0 \in \mathbb{Z}$ and codomain $\mathbb{R}^n$.

Remark 2.2.2 Since these are so ubiquitous, there are lots of different conventions and choices of notation. There are many ways to specify a sequence including

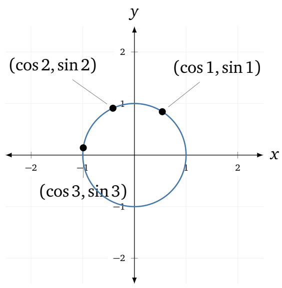
$$(x(k)) \qquad (x(k))_k \qquad (x(k))_{k=k_0}^{\infty} \qquad \{x(k)\} \qquad \{x(k)\}_k \qquad \{x(k)\}_{k=k_0}^{\infty}$$

It is also common to write x_k instead of $x(k)$. Unless stated otherwise, you may assume the domain of a sequence is $\mathbb{N}^+ = \{1, 2, 3, \ldots, \}$.

Example 2.2.3 The first few terms of the sequence $x(k) = (\cos k, \sin k)$ in $\mathbb{R}^2$ are

$$x(1) = (\cos 1, \sin 1), \quad x(2) = (\cos 2, \sin 2), \quad x(3) = (\cos 3, \sin 3), \quad \ldots$$

Each point in the sequence is a 1-radian counterclockwise rotation of the previous point about the origin. This rotation continues as k grows, so it appears the sequence will not converge to any point.



As with sequences in $\mathbb{R}$, you can specify what it means to “pick terms from a sequence”.

Definition 2.2.4 Let $x : \mathbb{N}^+ \to \mathbb{R}^n$ be a sequence and let $m : \mathbb{N}^+ \to \mathbb{N}^+$ be a strictly increasing function. The sequence $\{x(m(k))\}_{k=1}^{\infty}$ is a **subsequence** of the sequence $\{x(k)\}_{k=1}^{\infty}$.

Remark 2.2.5 This definition can be modified to allow for domains for x and m that are subsets of $\mathbb{Z}$ which are not necessarily $\mathbb{N}^+$. At minimum, the domain of x must equal the codomain of m. This version with $\mathbb{N}^+$ is stated for simplicity but you are free to define subsequences and sequences with different domains.

Example 2.2.6 Consider the sequence $x(k) = k^2$ for $k \in \mathbb{N}^+$. The subsequence of even terms $x(2), x(4), x(6), \ldots$, is defined using the strictly increasing function $m : \mathbb{N}^+ \to \mathbb{N}^+$ given by $m(k) = 2k$. Then $x(m(k)) = x(2k) = 4k^2$ for all $k \in \mathbb{N}^+$.

Subsequences need not follow some nice pattern like even numbers. Define $p : \mathbb{N}^+ \to \mathbb{N}^+$ so that $p(k)$ is the kth prime. The first few terms of the subsequence $\{x(p(k))\}_{k=1}^{\infty}$ are $4, 9, 25, 49, \ldots$ and this does not follow some obviously nice pattern.

These core definitions lead to a natural generalization of convergence.

2.2.1 Convergence of sequences

A term $x(k)$ in a sequence often represents a finite approximation to some desired value. The next term $x(k+1)$ in a sequence is usually a refined version of the previous term. Intuitively, the limit $p \in \mathbb{R}^n$ (if it exists) is the result of refining these finite approximations to arbitrary accuracy. Using this intuition and the definition of convergent sequences in $\mathbb{R}$, you can generalize the notion of convergent sequences to $\mathbb{R}^n$. The core distinction is that distance in $\mathbb{R}$ is measured using the absolute value $|\cdot|$ whereas the distance in $\mathbb{R}^n$ is measured with the norm $||\cdot||$.

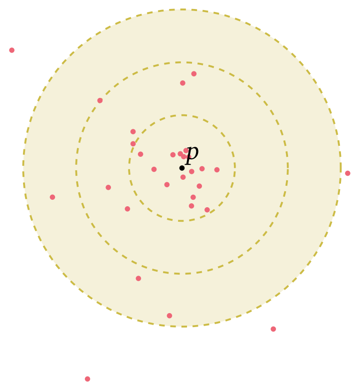
Definition 2.2.7 Fix a point $p \in \mathbb{R}^n$. A sequence $\{x(k)\}_k$ in $\mathbb{R}^n$ **converges to** p provided

$$\forall \varepsilon > 0, \exists K \in \mathbb{N} \text{ s.t. } \forall k \in \mathbb{N}, k \geq K \implies ||x(k) - p|| < \varepsilon.$$

If the above holds, write $\lim_{k\to\infty} x(k) = p$ or equivalently $x(k) \to p$.

Remark 2.2.8 You can equivalently say the limit of the sequence $\{x(k)\}_k$ is p.

This definition can be visualized in $\mathbb{R}^2$ with the diagram below.



In fact, this picture inspires a geometric interpretation with shrinking balls in $\mathbb{R}^n$.

> *A sequence $\{x(k)\}_k$ in $\mathbb{R}^n$ converges to p if and only if every open ball centered at p contains all but finitely many points of the sequence $\{x(k)\}_k$.*

This can be rephrased formally using ε-balls, which is part of an exercise at the end of this section. Right now, a routine example applying Definition 2.2.7 will demonstrate how closely arguments for sequences in $\mathbb{R}^n$ mimic arguments for sequences in $\mathbb{R}$.

Example 2.2.9 The sequence $\{x(k)\}_{k=1}^{\infty}$ given by $x(k) = (2 + \frac{1}{k}, \frac{\sin k}{k})$ converges to $(2, 0)$.

Proof. Let $\varepsilon > 0$ be arbitrary. Take $K = \lceil \sqrt{2}/\varepsilon \rceil + 1$. Let $k \in \mathbb{N}^+$ satisfy $k \geq K$. Then

$$\begin{aligned} ||x(k) - p|| = ||\left(2 + \tfrac{1}{k}, \tfrac{\sin k}{k}\right) - (2,0)|| &= \sqrt{\left(2 + \tfrac{1}{k} - 2\right)^2 + \left(\tfrac{\sin k}{k}\right)^2} \\ &= \sqrt{\tfrac{1 + \sin^2 k}{k^2}} \\ &\leq \sqrt{\tfrac{2}{k^2}} && \text{as } 0 \leq \sin^2 k \leq 1 \\ &< \varepsilon && \text{as } k > \sqrt{2}/\varepsilon, \end{aligned}$$

so $||x(k) - p|| < \varepsilon$. By the definition, $\lim_{k\to\infty} x(k) = p$ as desired. ■

It will be helpful to refer to convergence without knowing the specific limiting value.

Definition 2.2.10 Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$. The sequence $\{x(k)\}_k$ **converges** if there exists $p \in \mathbb{R}^n$ such that $\lim_{k\to\infty} x(k) = p$. Otherwise, the sequence $\{x(k)\}_k$ **diverges**.

Remark 2.2.11 If $\{x(k)\}_k$ converges (resp. diverges), then you can equivalently say the limit of the sequence $\{x(k)\}_k$ exists (resp. does not exist).

Many of the same facts about limits of sequences over $\mathbb{R}$ carry over with almost identical arguments. First and foremost, limits of sequences are unique.

Lemma 2.2.12 (Uniqueness of limits) Fix points $p, q \in \mathbb{R}^n$. Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$. If $\lim_{k\to\infty} x(k) = p$ and $\lim_{k\to\infty} x(k) = q$, then $p = q$.

Proof. The argument is very similar to the uniqueness of limits over $\mathbb{R}$, so this is left as an exercise. Remember the triangle inequality and that $x(k) - p = x(k) - q + q - p$. ■

Moreover, subsequences of convergent sequences also converge to the same value.

Lemma 2.2.13 Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$ converging to $p \in \mathbb{R}^n$. Every subsequence of $\{x(k)\}$ also converges to p.

Proof. This is left as an exercise. It follows quickly from definitions. ■

There is also a standard list of limit laws for sequences in higher dimensions.

Theorem 2.2.14 Fix $p, q \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Let $\{x(k)\}_k$ and $\{y(k)\}_k$ be sequences in $\mathbb{R}^n$.

(a) If $x(k) = p$ for all except finitely many $k \in \mathbb{N}^+$, then $x(k) \to p$. (Constant)

(b) If $x(k) \to p$ and $y(k) \to q$ then $x(k) + \lambda y(k) \to p + \lambda q$. (Linearity)

(c) If $x(k) \to p$ and $y(k) \to q$ then $x(k) \cdot y(k) \to p \cdot q$. (Dot Product)

Proof. These are left as exercises for proofs by definition. Item (a) follows quickly. For (b) and (c), use the triangle inequality. For (c), add zero to the expression $x(k) \cdot y(k) - p \cdot q$. ■

This correspondence between $|\cdot|$ in $\mathbb{R}$ and $||\cdot||$ in $\mathbb{R}^n$ demonstrates that, with good definitions, generalizing ideas from $\mathbb{R}$ to $\mathbb{R}^n$ is often a matter of notation. As you shall see in the next subsection, this connection can be formalized.

2.2.2 Components of sequences

A sequence in $\mathbb{R}^n$ can be thought of as n sequences in $\mathbb{R}$. Namely, a sequence $\{x(k)\}_k$ can be written as

$$x(k) = (x_1(k), x_2(k), \ldots, x_n(k)) \in \mathbb{R}^n$$

so the n sequences in $\mathbb{R}$ are given by $\{x_1(k)\}_k, \ldots, \{x_n(k)\}_k$.

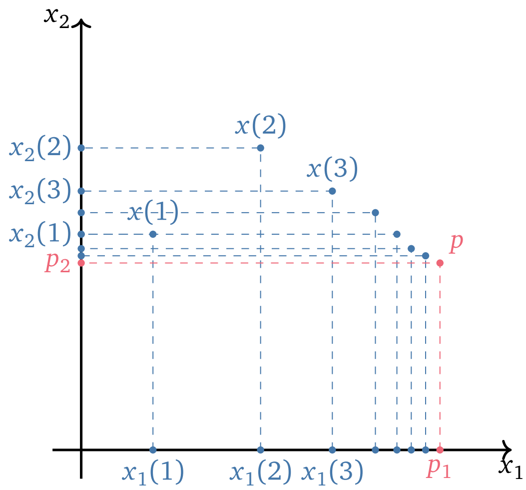
Example 2.2.15 Consider the sequence $\{(\cos k, \sin k)\}_k$ in $\mathbb{R}^2$. In this case, $x(k) = (\cos k, \sin k)$ so $x_1(k) = \cos k$ and $x_2(k) = \sin k$. The first few terms of $\{x(k)\}_{k=1}^{\infty}$ are:

$$x(1) = (\cos 1, \sin 1), \quad x(2) = (\cos 2, \sin 2), \quad x(3) = (\cos 3, \sin 3), \quad \ldots$$

This viewpoint allows you to formally connect the notions of convergence between $\mathbb{R}$ and $\mathbb{R}^n$ in a groundbreaking way.

Theorem 2.2.16 (Convergence component-wise) Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$ with $x(k) = (x_1(k), \ldots, x_n(k))$. Fix $p = (p_1, \ldots, p_n) \in \mathbb{R}^n$. The sequence $\{x(k)\}_k$ converges to p if and only if $\{x_i(k)\}_k$ converges to p_i for all $i \in \{1, 2, \ldots, n\}$.

The truth of this theorem can be seen in the following 2D illustration. The sequence of points in $\mathbb{R}^2$ converges to a point and the corresponding component sequences in $\mathbb{R}$ converge to the corresponding components of that point.



Its proof illuminates a way to relate distance in $\mathbb{R}^n$ to the corresponding n distances in $\mathbb{R}$.

Proof. ($\Longrightarrow$) Assume $\{x(k)\}_k$ converges to $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Fix $i \in \{1, 2, \ldots, n\}$ and fix $\varepsilon > 0$. It suffices to prove $x_i(k) \to a_i$ as $k \to \infty$. Since $x(k) \to a$, there exists $K \in \mathbb{N}$ such that

$$k \geq K \implies ||x(k) - a|| < \varepsilon.$$

Take this same K and let $k \in \mathbb{N}$ satisfy $k \geq K$. Then

$$|x_i(k) - a_i| = \sqrt{|x_i(k) - a_i|^2} \leq \sqrt{|x_1(k) - a_1|^2 + \cdots + |x_n(k) - a_n|^2} = ||x(k) - a|| < \varepsilon$$

Since $\varepsilon > 0$ was arbitrary, this proves that $x_i(k) \to a_i$ as $k \to \infty$. Since $i \in \{1, 2, \ldots, n\}$ was arbitrary, this proves the desired implication.

($\Longleftarrow$) Assume $\{x_i(k)\}$ converges for all $i = 1, 2, \ldots, n$. For each $i \in \{1, 2, \ldots, n\}$, let $a_i \in \mathbb{R}$ be the limit of $\{x_i(k)\}_k$. Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. It suffices to prove that $x(k) \to a$. Let $\varepsilon > 0$. For $i \in \{1, \ldots, n\}$, since $x_i(k) \to a_i$, there exists $K_i \in \mathbb{N}$ such that

$$k \geq K_i \implies |x_i(k) - a_i| < \frac{\varepsilon}{\sqrt{n}}. \tag{2.2.1}$$

Take $K = \max\{K_1, \ldots, K_n\}$. Then

$$\begin{aligned} ||x(k) - a|| &= \sqrt{|x_1(k) - a_1|^2 + \cdots + |x_n(k) - a_n|^2} \\ &\leq \sqrt{n \max_{1 \leq i \leq n} |x_i(k) - a_i|^2} \\ &< \sqrt{n \cdot \frac{\varepsilon^2}{n}} = \varepsilon, \end{aligned}$$

as $k \geq K \geq K_i$ for all $1 \leq i \leq n$ and (2.2.2) holds. This completes the proof. ■

Theorem 2.2.16 marks the beginning for using components to reduce theorems over $\mathbb{R}^n$ to theorems over $\mathbb{R}$. For example, limit laws in $\mathbb{R}^n$ (Theorem 2.2.14) can be swiftly established by appealing to the limit laws in $\mathbb{R}$ instead of directly through the formal definition of the limit. This counts as a major victory.

2.2.3 Interior points, boundary points, and limit points

A sequence of points $\{x(k)\}_k$ in $\mathbb{R}^n$ often arises in optimization problems or search algorithms. In those cases, your approximations may only naturally lie in a particular set $A \subseteq \mathbb{R}^n$ such as the domain of your optimizing function. Sometimes these approximations will converge to a point inside your set A, that is, the interior of A. Sometimes the limit of this sequence may "fall outside" your set A, that is, the boundary of A. This perspective leads to an equivalent formulation for interior points, boundary points, and limit points.

Theorem 2.2.17 Let $A \subseteq \mathbb{R}^n$ be a set. Let $p \in \mathbb{R}^n$ be a point.

(a) The point p is an interior point of A if and only if for every sequence $\{x(k)\}_k$ of points converging to p, there exists $K \in \mathbb{N}^+$ such that $\{x(k)\}_{k=K}^{\infty} \subseteq A$.

(b) The point p is a boundary point of A if and only if there exists a sequence of points in A converging to p and there exists a sequence of points in A^c converging to p.

(c) The point p is a limit point of A if and only if there exists a sequence of points in $A \setminus \{p\}$ which converges to p.

This theorem is a boon. For instance, it is easy to exhibit limit points in explicit examples.

Example 2.2.18 Let $A = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(0,0)\}$. You can quickly check that $(1,1)$ is a limit point of A using the sequential definition. Define the sequence

$$x(k) = (1 - \tfrac{1}{4k}, 1 - \tfrac{1}{4k})$$

for $k \geq 1$ so $x(k) \to (1,1)$ as $k \to \infty$ by Theorem 2.2.16. Moreover, $x(k) \in A \setminus \{(1,1)\}$ since for $k \geq 1$,

$$2 > (1 - \tfrac{1}{4k})^2 + (1 - \tfrac{1}{4k})^2 \geq 2(\tfrac{3}{4})^2 > 1.$$

Thus, $(1,1)$ is a limit point of A by Theorem 2.2.17.

A similar argument can work for the limit point $(-1,0)$. On the other hand, $(0,0)$ and $(3,0)$ are not limit points of S but how would you prove it? Try the definition with open balls (Definition 2.1.22) instead.

The proof of this theorem requires a clever idea.

Proof of Theorem 2.2.17. We only prove (c). As (a) and (b) are similar, they are left as exercises.

$(\Longleftarrow)$ Let $\{x(k)\}_n$ be a sequence of points in $A \setminus \{p\}$ which converge to p. Fix $\varepsilon > 0$. By definition of convergence, there exists $K \in \mathbb{N}$ such that

$$\forall k \in \mathbb{N}, k \geq K \implies ||x(k) - p|| < \varepsilon.$$

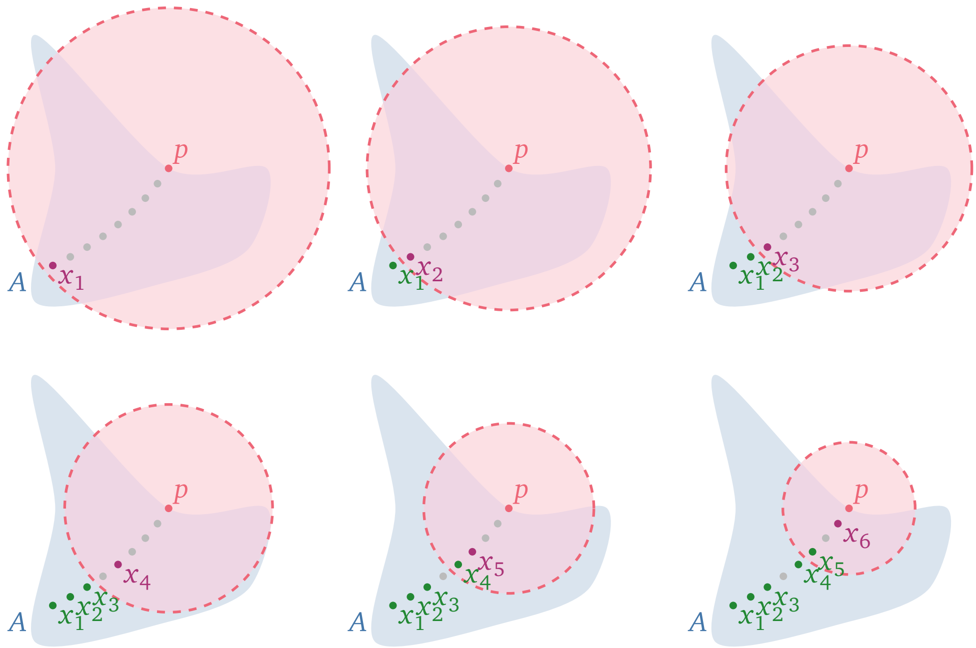
In particular, $x(k) \in B_\varepsilon(p)$ and, by assumption, $x(k) \in A \setminus \{p\}$. Thus, $B_\varepsilon(p) \setminus \{p\}$ contains points of A. This proves p is a limit point of A.

$(\Longrightarrow)$ Assume $p \in \mathbb{R}^n$ is a limit point of A. It suffices to construct a sequence $\{x(k)\}_k$ such that $\forall k \in \mathbb{N}^+, x(k) \in A \setminus \{p\}$ and $x(k) \to p$. For each $k \in \mathbb{N}^+$, choose a point $x(k) \in B_{1/k}(p) \cap A$ with $x(k) \neq p$. Such a point exists by assumption. It remains to show $x(k) \to p$ as $k \to \infty$. Fix $\varepsilon > 0$. Take $K = \lceil \frac{1}{\varepsilon} \rceil$. Let $k \in \mathbb{N}$ satisfy $k \geq K$. Then

$$\begin{aligned}
||x(k) - p|| &< \frac{1}{k} && \text{since } x(k) \in B_{1/k}(p) \\
&\leq \frac{1}{K} && \text{since } k \geq K \\
&\leq \varepsilon && \text{since } K \geq 1/\varepsilon.
\end{aligned}$$

This completes the proof. ∎

A picture illustrating this proof is provided below. It shows how each point $x_k = x(k) \in A$ in the constructed sequence is selected using successively smaller open balls.



These definitions are another tool for verifying whether a point lies on the boundary or interior. Sometimes the sequential definitions are easier, and sometimes the open ball definitions are easier. It is up to you to choose. With many good definitions, the parallels between $\mathbb{R}^n$ and $\mathbb{R}$ allow you to rapidly establish new theorems via sequences. This theme will continue for many more chapters to come.

Exercises for Section 2.2

Concepts and definitions

2.2.1 Let $\{x(k)\}_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n$. Let $p \in \mathbb{R}^n$. Which statements are equivalent to "$x(k) \to p$"?

(a) $\forall \varepsilon > 0, \exists K \in \mathbb{N}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| < \varepsilon$

(b) $\forall \varepsilon > 0, \exists K \in \mathbb{N}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| \leq 2\varepsilon$

(c) $\forall \varepsilon > 0, \exists K \in \mathbb{R}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| < \varepsilon$

(d) $\forall \varepsilon \in (0,1), \exists K \in \mathbb{N}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| < \varepsilon$

(e) $\forall \varepsilon > 0, \exists K \in \mathbb{N}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| < \dfrac{1}{\varepsilon}$

(f) $\forall m \in \mathbb{N}^+, \exists K \in \mathbb{N}$ s.t. $\forall k \in \mathbb{N}, \quad k \geq K \implies ||x(k) - p|| < \dfrac{1}{m}$

2.2.2 Write the formal definition for divergence of a sequence in $\mathbb{R}^n$ with quantifiers.

2.2.3 Recall this equivalent viewpoint of a convergent sequence:

A sequence $\{x(k)\}_k$ in $\mathbb{R}^n$ converges to $p \in \mathbb{R}^n$ if every open ball centered at p contains all but finitely many points of the sequence $\{x(k)\}_k$.

Let $\{x(k)\}_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n$. Let $p \in \mathbb{R}^n$. Which statements are equivalent to $x(k) \to p$?

(a) $\forall \varepsilon > 0, \exists K \in \mathbb{N}^+$ s.t. $\forall k \in \mathbb{N}^+, k \geq K \implies x(k) \in B_\varepsilon(p)$

(b) $\forall \varepsilon > 0, \exists K \in \mathbb{N}^+$ s.t. $\{x(k) : k \in \mathbb{N}^+, k \geq K\} \subseteq B_\varepsilon(p)$

(c) $\forall \varepsilon > 0, \exists K \in \mathbb{N}^+$ s.t. $\{x(k) : k \in \mathbb{N}^+, 1 \leq k \leq K\} \cap B_\varepsilon(p) = \varnothing$

(d) $\forall \varepsilon > 0$, the set $\{x(k) : k \in \mathbb{N}^+, x(k) \notin B_\varepsilon(p)\}$ is finite.

(e) $\forall \varepsilon > 0$, the set $\{k \in \mathbb{N}^+ : x(k) \notin B_\varepsilon(p)\}$ is finite

Proofs

2.2.4 Use Theorem 2.2.16 for both of the following parts.

(a) Below is an attempted proof that the sequence $\{(\cos(\frac{1}{k}), \sin(\frac{1}{k}))\}_{k=1}^{\infty}$ converges.

> 1. *Note* $\lim_{k\to\infty} \cos(\frac{1}{k}) = 1$ *and* $\lim_{k\to\infty} \sin(\frac{1}{k}) = 0$.
> 2. *Thus,* $\left\{(\cos(\frac{1}{k}), \sin(\frac{1}{k})\right\}_{k=1}^{\infty}$ *converges to* $(1,0)$.

Each line is missing a key justification. Revise them by adding the justifications.

(b) Prove that the sequence $\{(\cos k, \sin k)\}_{k=1}^{\infty}$ diverges.

2.2.5 Recall two equivalent definitions for a limit point $p \in \mathbb{R}^n$ of a set $S \subseteq \mathbb{R}^n$:

I. *There exists a sequence $\{x(k)\}_k$ of points in $S \setminus \{p\}$ such that $x(k) \to p$.*

II. *Every open ball centered at p contains points in $S \setminus \{p\}$.*

Let $S = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(0,0)\}$. Choose the most convenient definition for each of the following exercises.

(a) Show $(1,1)$ is a limit point of S.

(b) Show that $(0,0)$ is not a limit point of S.

2.2.6 Define

$$A = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 < 2\} \cup \{(0,0)\}.$$

(a) Show that $(0,0)$ is not an interior point of A using the sequential definition.

(b) Show that $(0,0)$ is a boundary point of A using the sequential definition.

(c) Consider how you might prove (a) and (b) using the open ball definitions instead. The sequential definitions are somewhat easier to use in both cases. Briefly explain why.

2.2.7 Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$. Prove or disprove the following claims.

(a) If $\{x(k)\}_k$ converges, then any subsequence of $\{x(k)\}_k$ also converges.

(b) If $\{x(k)\}_k$ diverges, then any subsequence of $\{x(k)\}_k$ also diverges.

Applications and beyond

2.2.8 The proof of Theorem 2.2.16 in the ($\Longrightarrow$) direction is below. Review it again.

1. *Assume $\{x(k)\}_k$ converges to $a = (a_1, \dots, a_n) \in \mathbb{R}^n$.*
2. *Fix $i \in \{1, 2, \dots, n\}$ and fix $\varepsilon > 0$.*
3. *It suffices to prove $x_i(k) \to a_i$ as $k \to \infty$.*
4. *Since $x(k) \to a$, there exists $K \in \mathbb{N}$ such that $k \ge K \implies ||x(k) - a|| < \varepsilon$.*
5. *Take this same K and let $k \in \mathbb{N}$ satisfy $k \ge K$.*
6. *Then*

$$\begin{aligned} |x_i(k) - a_i| &= \sqrt{|x_i(k) - a_i|^2} \\ &\le \sqrt{|x_1(k) - a_1|^2 + \cdots + |x_n(k) - a_n|^2} \\ &= ||x(k) - a|| \\ &< \varepsilon \end{aligned}$$

7. *Since $\varepsilon > 0$ was arbitrary, this proves that $x_i(k) \to a_i$ as $k \to \infty$.*
8. *Since $i \in \{1, 2, \dots, n\}$ was arbitrary, this proves the desired implication.*

After Line 1, the author could have explicitly written what they want to show, but they did not. This choice is acceptable and can be more elegant, but the author needs to implicitly reference what they want to show at different steps of the proof. Here you will identify and clarify those implicit references.

(a) Write down the WTS[3] as a formal logical statement with quantifiers. Imagine you are adding this line between Lines 1 and 2.

(b) Line 3 implicitly references the WTS by saying the phrase *"it suffices to show"*. Identify the corresponding part of your WTS.

(c) Lines 7 and 8 implicitly reference the WTS by saying the phrases *"since [..] was arbitrary"*. Identify the correspond parts of your WTS.

(d) If you omitted these phrases from Lines 3, 7, and 8, how would the proof writing suffer?

[3] This stands for "Want To Show" or "Want To Prove".

2.2.9

The proof of Theorem 2.2.16 in the ($\Longleftarrow$) direction is below. Review it again.

1. *Assume* $\{x_i(k)\}$ *converges for all* $i = 1, 2, \ldots, n$.
2. *For each* $i \in \{1, 2, \ldots, n\}$, *let* $a_i \in \mathbb{R}$ *be the limit of* $\{x_i(k)\}_k$.
3. *Let* $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. *It suffices to prove that* $x(k) \to a$.
4. *Let* $\varepsilon > 0$. *For* $i \in \{1, \ldots, n\}$, *since* $x_i(k) \to a_i$, *there exists* $K_i \in \mathbb{N}$ *such that*

$$k \geq K_i \implies |x_i(k) - a_i| < \frac{\varepsilon}{\sqrt{n}}. \tag{2.2.2}$$

5. *Take* $K = \max\{K_1, \ldots, K_n\}$.
6. *Then*

$$\begin{aligned} ||x(k) - a|| &= \sqrt{|x_1(k) - a_1|^2 + \cdots + |x_n(k) - a_n|^2} \\ &\leq \sqrt{n \max_{1 \leq i \leq n} |x_i(k) - a_i|^2} \\ &< \sqrt{n \cdot \frac{\varepsilon^2}{n}} = \varepsilon, \end{aligned}$$

as $k \geq K \geq K_i$ *for all* $1 \leq i \leq n$ *and* (2.2.2) *holds. This completes the proof.*

Longer proofs can be confusing to read with so many technical details, so it is worthwhile to informally summarize the key ideas. Here is a two-sentence attempt:

I. *As each sequence* $x_i(k)$ *in* $\mathbb{R}$ *converges, each coordinate is eventually close to its limit* a_i.

II. *Since the distance between two points in* $\mathbb{R}^n$ *is at most* n *times the maximum distance between the coordinates, the sequence* $x(k)$ *in* $\mathbb{R}^n$ *must be eventually close to its limit* $a = (a_1, \ldots, a_n)$.

Identify which sentence corresponds to which line(s) of the proof.

2.2.10

Summarizing proofs with a picture or concise informal sentences helps reveal the big picture and illustrate formal arguments. You can practice this with the proof of Theorem 2.2.17(c).

(a) Draw a picture of the ($\Longrightarrow$) proof and summarize the idea in a single well-written sentence.

(b) Draw a picture of the ($\Longleftarrow$) proof and summarize the idea in two well-written sentences.

2.3. Open sets and closed sets

Given a set $A \subseteq \mathbb{R}^n$, you will often only consider sequences converging to points inside A. For example, if the set A is the domain of a map, you may want a sequence of approximations $\{x(k)\}_k$ lying inside A to actually converge to a point p within the domain A. Otherwise, you cannot necessarily evaluate the map at the point p.

You can ensure this desirable feature in two different ways.

1) If a sequence in $\mathbb{R}^n$ converges to $a \in A$, then the tail of the sequence belongs to A.

2) If a sequence in A converges to $a \in \mathbb{R}^n$, then a must belong to A.

Many sets A will not satisfy either property, so each of these two properties of sets in $\mathbb{R}^n$ warrants their own definition. As you shall see, these properties are fundamental topological notions and they are intimately related.

2.3.1 Open sets

First, you want to define a property of a set $A \subseteq \mathbb{R}^n$ satisfying the following.

If a sequence in $\mathbb{R}^n$ converges to $a \in A$, then the tail of the sequence belongs to A.

In other words, no matter how you approach $a \in A$ you must eventually lie inside A. That's precisely the sequential definition of an interior point! This suggests a definition.

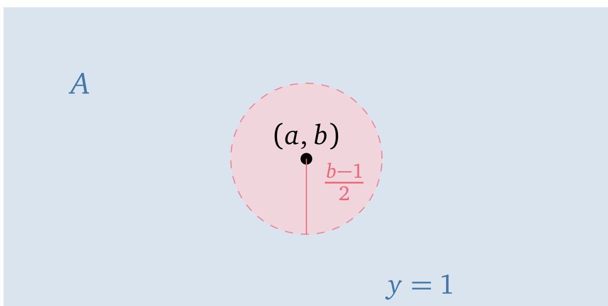
Definition 2.3.1 A set $A \subseteq \mathbb{R}^n$ is **open** if every point of A is an interior point of A.

If a is an interior point of A, then you can approach a however you wish; that is, the set A itself does not really restrict how the sequence approaches a. On the other hand, if a is a boundary point of A then you can only approach a along certain sequences while still lying inside A; that is, the set A restricts how the sequence approaches a. This informal comparison indicates that our definition of open sets is probably a good one.

Example 2.3.2 Here are some standard examples of open sets:
- The empty set $\varnothing$ is (vacuously) open.
- The set $\mathbb{R}^n$ is open.
- For $a, b \in \mathbb{R}$ with $a < b$, the open interval (a, b) in $\mathbb{R}$ is open.
- For $p \in \mathbb{R}^n$ and $r > 0$, the open ball $\{x \in \mathbb{R}^n : ||x - p|| < r\}$ in $\mathbb{R}^n$ is open.

The latter two statements are not trivial[4]. They require a proof, albeit a short one.

Example 2.3.3 The set $A = \{(x, y) \in \mathbb{R}^2 : y > 1\}$ is open. Below is a "picture proof".



[4]The word *trivial* has a precise mathematical meaning and should not be used outside this context; it means that the definition is automatically satisfied without any argument. For example, the proof that the empty set $\varnothing$ is open is trivial. It does not mean "it is too easy to write down". Some authors may choose to use it that way but you should not. Proof by authority is not a proof.

Using this drawing, you can construct a formal proof by definition.

Proof. Let $(a,b) \in A$ be given. Fix $r = \frac{b-1}{2}$ so $(a,b) \in A$ implies $r > 0$. It suffices to show that $B_r((a,b)) \subseteq A$. For $(x,y) \in B_r((a,b))$, it follows that

$$\begin{aligned} |y-b| \leq \|(x,y)-(a,b)\| < r &\implies |y-b| < \frac{b-1}{2} \\ &\implies b - \frac{b-1}{2} < y < b + \frac{b-1}{2} \\ &\implies y > \frac{b+1}{2} > 1 \qquad \text{as } b > 1. \end{aligned}$$

Therefore, $(x,y) \in A$, which proves that $B_r((a,b))$ is contained in A, as desired. ■

The easiest way to produce an open set is using the interior of a set.

Lemma 2.3.4 The interior of a set $A \subseteq \mathbb{R}^n$ is open.

Proof. Let $a \in A^\circ$. By definition, there exists $\varepsilon > 0$ such that $B_\varepsilon(a) \subseteq A$. Fix $x \in B_\varepsilon(a)$. It suffices to show that $x \in A^\circ$. Since the open ball $B_\varepsilon(a)$ is open and $x \in B_\varepsilon(a)$, there exists $\delta > 0$ such that $B_\delta(x) \subseteq B_\varepsilon(a)$. As $B_\varepsilon(a) \subseteq A$, it follows that $B_\delta(x) \subseteq A$ so $x \in A^\circ$. ■

This gives rise to equivalent definitions of an open set.

Lemma 2.3.5 Let $A \subseteq \mathbb{R}^n$. All of the following are equivalent:
(a) A is open.
(b) $A = A^\circ$.
(c) $A \cap \partial A = \varnothing$.

Proof. This is left as an exercise. To show (a) implies (b), modify the proof of Lemma 2.3.4. To show (b) implies (c), use Theorem 2.1.32. To show (c) implies (a), you must show a point $a \in A$ is either a boundary point of A or an interior point of A; this fact is required in the proof of Theorem 2.1.32. ■

In addition to expanding your geometric intuition, these equivalent definitions can be used to verify whether a set is open.

Example 2.3.6 Consider the interval $I = [137, 237)$ in $\mathbb{R}$. By Theorem 2.2.17, the point 137 is a boundary point of I because the sequence $\{137 + \frac{1}{k}\}_{k=1}^\infty$ in I, and the sequence $\{137 - \frac{1}{k}\}_{k=1}^\infty$ in I^c both converge to 137. Therefore, $137 \in I \cap \partial I$ so $I \cap \partial I$ is non-empty. By Lemma 2.3.5, I is not open.

2.3.2 Closed sets

Second, you want to define a property of a set $A \subseteq \mathbb{R}^n$ satisfying the following.

If a sequence in A converges to a point $a \in \mathbb{R}^n$, then a must belong to A.

In other words, A must contain all of its limit points. This suggests a definition.

Definition 2.3.7 A set $A \subseteq \mathbb{R}^n$ is **closed**[5] if every limit point of A belongs to A.

[5]The word *closed* comes from a common mathematical phrasing. You might say "a set is closed under taking limits" kind of like how you might say "a subspace is closed under vector addition" in linear algebra. A darker viewpoint is that a sequence cannot escape the set. There is nowhere it can go. The way is shut. The door is *closed*!

Any convergent sequence in a closed set A must converge to a point in A; the sequence cannot “fall outside” the set. That is a lovely property to have.

Example 2.3.8 Here are some common examples of closed sets.
- The empty set $\varnothing$ is (vacuously) closed.
- The set $\mathbb{R}^n$ is closed.
- For $a, b \in \mathbb{R}$ with $a < b$, the closed interval $[a, b]$ in $\mathbb{R}$ is closed.
- For $p \in \mathbb{R}^n$ and $r > 0$, the closed ball $\{x \in \mathbb{R}^n : ||x - p|| \leq r\}$ is closed.

Again, the latter two statements are not trivial. They require a proof.

Example 2.3.9 The set $A = \{(x, y) \in \mathbb{R}^2 : y \geq 1\}$ is closed. Here is a proof by definition.

Proof. Let (a, b) be a limit point of A. Let $\{(x(k), y(k))\}_k$ be a sequence in $A \setminus \{(a, b)\}$ converging to (a, b) so $x(k) \to a$ and $y(k) \to b$ by Theorem 2.2.16. As $y(k) \geq 1$ for all $k \in \mathbb{N}$ and $y(k) \to b$, it follows by a limit law over $\mathbb{R}$ that $b = \lim_{k\to\infty} y(k) \geq 1$. Thus, $(a, b) \in A$. ∎

Closed sets are often produced via the closure.

Lemma 2.3.10 The closure of a set A is closed.

Proof. Let p be a limit point of $\overline{A}$. There exists a sequence $\{x(k)\}_{k=1}^{\infty}$ in $\overline{A} \setminus \{p\}$ converging to p. For each $k \in \mathbb{N}^+$, note $x(k) \in \overline{A}$ implies that there exists $y(k) \in A$ satisfying $||x(k) - y(k)|| < 1/k$. Thus, it suffices to show that the sequence $\{y(k)\}_{k=1}^{\infty}$ in A converges to p, as this implies $p \in \overline{A}$.

Fix $\varepsilon > 0$. As $x(k) \to p$, there exists $K \in \mathbb{N}$ such that

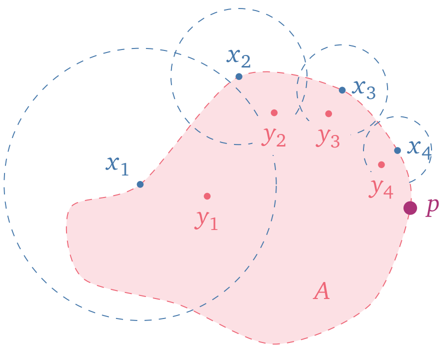
$$\forall k \in \mathbb{N}, k \geq K \implies ||x(k) - p|| < \frac{\varepsilon}{2}.$$

Set $K' = \max\{K, \lceil \frac{2}{\varepsilon} \rceil\}$. By the triangle inequality, for $k \in \mathbb{N}$ with $k \geq K'$,

$$\begin{aligned} ||y(k) - p|| = ||y(k) - x(k) + x(k) - p|| &\leq ||y(k) - x(k)|| + ||x(k) - p|| \\ &< \frac{1}{k} + \frac{\varepsilon}{2} \\ &\leq \frac{1}{K'} + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Hence, $y(k) \to p$ as required. ∎

Remark 2.3.11 Why was the sequence $\{y(k)\}_k$ constructed that way? If $\{x(k)\}_k$ was a sequence in $A \setminus \{p\}$ instead of $\overline{A} \setminus \{p\}$, then $x(k) \to p$ implies p is a limit point of A and hence $p \in \overline{A}$. Thus, the "worst case scenario" is that $x(k) \in \overline{A} \setminus A$ for every k, in which case $x(k) \in \partial A$ for every k. The figure below demonstrates this with $x_k = x(k)$ and $y_k = y(k)$.



Here you need to use $\{x(k)\}_k$ in $\overline{A}$ to construct a sequence $\{y(k)\}_k$ in A converging to p. By choosing $y(k)$ in shrinking balls centred at $x(k)$, this forces $y(k)$ to converge to p.

The above lemma gives rise to equivalent definitions of a closed set.

Lemma 2.3.12 Let $A \subseteq \mathbb{R}^n$. All of the following are equivalent:
(a) A is closed.
(b) $A = \overline{A}$.
(c) $\partial A \subseteq A$.

Proof. This is left as an exercise. To show (a) implies (b), modify the proof in Lemma 2.3.10. To show (b) implies (c), use Theorem 2.1.32. To show (c) implies (a), you must show that a limit point of A is either a boundary point of A or an interior point of A. ■

Again, you can use these equivalent definitions to verify whether a set A is closed.

Example 2.3.13 Consider again the interval $I = [137, 237)$ in $\mathbb{R}$. By an argument similar to Example 2.3.6, the point 237 is a boundary point of I. On the other hand, $237 \notin I$ so $\partial I \nsubseteq I$. Hence, I is not closed by Lemma 2.3.12.

2.3.3 Set operations

As you have seen, there were similar but dual motivations for the definitions of open sets and closed sets. Somewhat surprisingly, their definitions are literally complementary.

Theorem 2.3.14 A set $A \subseteq \mathbb{R}^n$ is open if and only if its complement $A^c = \mathbb{R}^n \setminus A$ is closed.

Proof. Assume $A \subseteq \mathbb{R}^n$ is an open set. Let p be a limit point of the complement A^c, so there exists a sequence $\{x(k)\}_{k=1}^{\infty}$ in $A^c \setminus \{p\}$ satisfying $x(k) \to p$. Since A is open, if $p \in A$ then the tail of this sequence must lie in A but this is impossible since $x(k) \in A^c$ for all $k \geq 1$. Therefore, p must lie in the complement $\mathbb{R}^n \setminus A = A^c$. This shows A^c is closed.

Conversely, assume $B = A^c$ is a closed set so we must show $B^c = A$ is open. Let $p \in B^c$ be arbitrary. Since B is closed, the point p cannot be a limit point of B. Therefore, there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \setminus \{p\}$ does not contain any points of B. In other words, $B_\varepsilon(p) \setminus \{p\} \subseteq B^c$. Again, since B is closed, the point p does not belong to B so it follows that $B_\varepsilon(p) \subseteq B^c$. This shows p is an interior point of B^c and hence B^c is open. ■

If you draw a few pictures, then Theorem 2.3.14 is not so surprising.

Example 2.3.15 The closed interval $[223, 224]$ is closed and its complement, the union of open intervals $(-\infty, 223) \cup (224, \infty)$, is open. Similarly, the open disk $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ is open and its complement $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \geq 1\}$ is closed.

Example 2.3.16 The empty set $\varnothing$ and the set $\mathbb{R}^n$ are both open and closed in $\mathbb{R}^n$, which is sometimes called **clopen**. You can prove that these are the only two clopen sets in $\mathbb{R}^n$.

Example 2.3.17 There are examples of sets which are *neither* open nor closed.

- The interval $[137, 237)$ is neither open nor closed as shown in previous examples.
- The set $A = \{\frac{1}{n} : n \in \mathbb{N}\}$ is neither open nor closed. You can show it is not open because any open ball around any point is not contained in the set. You can show it is not closed because 0 is a limit point and does not belong to A.
- The set $B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, y \geq 0\}$ is the upper half of an open disk, including the points along the x-axis. You can check it is not open because $(0, 0)$ is not an interior point. You can check it is not closed because $(0, 1)$ is a boundary point

that does not belong to B.

- The set of rationals $\mathbb{Q}$ is neither open nor closed since $\mathbb{Q}^{\circ} = \varnothing$ and $\overline{\mathbb{Q}} = \mathbb{R}$.

Open sets and closed sets respect other basic set operations, too.

Lemma 2.3.18 All of the following are true for sets in $\mathbb{R}^n$:
(a) A finite intersection of open sets is open.
(b) A finite or infinite union of open sets is open.
(c) A finite union of closed sets is closed.
(d) A finite or infinite intersection of closed sets is closed.
(e) A finite Cartesian product of open sets is open.
(f) A finite Cartesian product of closed sets is closed.

Proof. These are left as exercises. Your proof should be sensitive to the distinction between finite and infinite unions (or intersections). For example, with (a), you will need to take a minimum of a finite list of numbers whereas with (b), you will not need to take a minimum. ■

Example 2.3.19 An infinite intersection of open sets may not necessarily be open, such as:

- $\displaystyle\bigcap_{\varepsilon>0}(-\varepsilon, \varepsilon)$. This is equal to the singleton $\{0\}$, which is not open.
- $\displaystyle\bigcap_{n=1}^{\infty}\left(0, 1+\frac{1}{n}\right)$. This is equal to $(0,1]$, which is not open.

An infinite union of closed sets may not necessarily be closed, such as:

- $\displaystyle\bigcup_{0<\varepsilon<1}[-\varepsilon, \varepsilon]$. This is equal to the interval $(-1,1)$, which is not closed.
- $\displaystyle\bigcup_{n=1}^{\infty}\left[0, 1-\frac{1}{n}\right]$. This is equal to $[0,1)$, which is not closed.

The properties of open and closed are two fundamental topological properties of sets. They capture desirable features about the behaviour of limit points, but they do not capture everything you may want. Both of them tell you what happens *if* you have a convergent sequence in your set A. However, does there even exist a convergent sequence in A? You will need to develop a new topological property to address this critical issue and, as you shall later see, this is key to solving multivariable optimization problems.

Exercises for Section 2.3

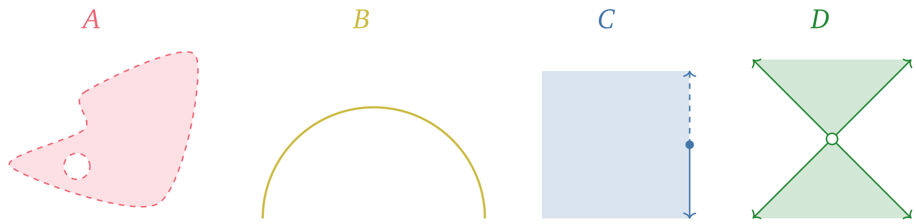
Concepts and definitions

2.3.1

(a) Identify whether A, B, C, D are open, closed, both, or neither.

(b) Sketch the interior and closure of A, B, C, D.

(A dotted line means the points are not contained in the set; an edge without a dotted or solid line means the set is unbounded in that direction.)



2.3.2 Which sets are open, closed, both, or neither?

(a) $\{(x,y) \in \mathbb{R}^2 : x < 0\}$

(b) $\varnothing$

(c) $\mathbb{R}^n$

(d) $B_1((1,0)) \cup B_1((-1,0))$

(e) $\{(x,y,z) \in \mathbb{R}^3 : 4 \le x^2 + y^2 + z^2 \le 9\}$

(f) $\mathbb{Z} \times \mathbb{R}$

(g) $\{(x,y) \in \mathbb{R}^2 : 0 < x < 2, 3 \le y \le 5\}$

(h) $[0,1]^n = \underbrace{[0,1] \times \cdots \times [0,1]}_{n \text{ times}}$

2.3.3 Let $A \subseteq \mathbb{R}^n$. Some of the following statements are *equivalent* to "A is open" or "A is closed" or neither. Decide which statements are equivalent to which property.

(a) $A = A^o$

(b) $\partial A \subseteq A$

(c) $A^* \subseteq A$

(d) A^c is not open.

(e) A^c is closed.

(f) $\partial A = \partial(A^c)$

(g) $\overline{A} = A$

(h) $\overline{\overline{A}} = \overline{A}$

(i) $A \cap \partial A = \varnothing$

(j) $\partial A = \overline{A} \setminus A^o$

(k) Every point of A is an interior point of A.

(l) Every point of A is a limit point of A.

Proofs

2.3.4 Consider again the left half plane $S = \{(x,y) \in \mathbb{R}^2 : x < 2\}$. You will prove S is open. The essence of many proofs in higher dimensions can be discovered with an illustrative sketch, so it important that you practice trying to draw a "picture proof" whenever possible. It helps to discover the key ideas.

(a) Write the open ball definition of "S is open." Express it with as many quantifiers as you can.

(b) Draw a "picture proof" that S is open. Label it using notation from your formal statement.

(c) Prove that S is open. Use your "picture proof" to guide you.

2.3.5 Since there are so many equivalent definitions of many things, you will often have several options available for a proof. You will need to get in the habit of listing those options and trying to find the simplest approach. It is not always obvious so sometimes you just have try every method. Let P be the plane $z = 0$ in $\mathbb{R}^3$. You will show the plane is closed in $\mathbb{R}^3$.

(a) List at least three different options available to prove P is closed.

(b) Choose one of them to prove that P is closed.

2.3.6 Topology uses natural language and intuitive pictures. This is a great benefit but, if you are not careful with definitions, you can easily fool yourself into believing a flawed argument.

(a) Here is a flawed proof of the true statement: *The interior of a set A is open.*

1. *Fix $a \in A^{\circ}$.*
2. *By definition, there exists $\varepsilon > 0$ such that $B_{\varepsilon}(a) \subseteq A$.*
3. *Since a was arbitrary, this proves that A° is open.*

This attempted proof has a fatal mistake. Identify it and briefly explain.

(b) Here is a flawed proof of the true statement: *The closure of a set A is closed.*

1. *Let a be a limit point of A.*
2. *By definition, $\overline{A} = A \cup A^*$ where A^* is the set of limit points of A.*
3. *Thus, $a \in \overline{A}$.*
4. *Since a was arbitrary, this proves that $\overline{A}$ is closed.*

This attempted proof has a fatal mistake. Identify it and briefly explain.

2.3.7 The open sets in $\mathbb{R}^n$ form a **topology**, meaning that they satisfy three conditions:

- The empty set and $\mathbb{R}^n$ are open.
- The union of any arbitrary collection of open sets is open.
- The intersection of any finite collection of open sets is open.

You will investigate some simple cases of this scenario.

(a) Suppose A, B are open subsets of $\mathbb{R}^n$. Prove that $A \cup B$ is open.

(b) Show that the union of any collection of open subsets is open. *Hint:* Generalize ideas from (a).

(c) Suppose A, B are open subsets of $\mathbb{R}^n$. Prove that $A \cap B$ is open.

(d) The intersection of any collection of open sets is **not** necessarily open. Give a counterexample and explain why the above argument would fail here.

2.3.8 Recall Lemma 2.3.10 establishes that the closure of a set A is closed. Here is an essentially correct proof. Some details are missing which you will fill in.

1. *Let p be a limit point of $\overline{A}$.*
2. *There exists a sequence $\{x(k)\}_{k=1}^{\infty}$ in $\overline{A}\setminus\{p\}$ converging to p.*
3. *For each $k \in \mathbb{N}^+$, note $x(k) \in \overline{A}$ implies that there exists $y(k) \in A$ satisfying $||x(k)-y(k)|| < 1/k$.*
4. *It suffices to show that the sequence $\{y(k)\}_{k=1}^{\infty}$ in A converges to p.*
5. *Fix $\varepsilon > 0$. As $x(k) \to p$, there exists $K \in \mathbb{N}$ such that*

$$\forall k \in \mathbb{N}, k \geq K \implies ||x(k)-p|| < \frac{\varepsilon}{2}.$$

6. *Set $K' = \max\{K, \lceil \frac{2}{\varepsilon} \rceil\}$.*
7. *By the triangle inequality, for $k \in \mathbb{N}$ with $k \geq K'$,*

$$\begin{aligned}||y(k)-p|| = ||y(k)-x(k)+x(k)-p|| &\leq ||y(k)-x(k)|| + ||x(k)-p|| \\ &< \frac{1}{k} + \frac{\varepsilon}{2} \\ &\leq \frac{1}{K'} + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

8. *Hence, $y(k) \to p$ as required.*

(a) Why does line 2 follow from line 1?

(b) Add a bit more justification to line 3. Namely, why do these $y(k)$ exist?

(c) Why does the claim in line 4 imply the lemma?

2.4. Compact sets

Suppose a set $A \subseteq \mathbb{R}^n$ is the domain of a real-valued function f. You want to determine its maximum value on A, but how can you find it? Here is a brilliant strategy. Construct a sequence of points $\{x(k)\}_{k=1}^{\infty}$ in A, where each value $f(x(k))$ is attempting to approximate the maximum value of f. Ideally, as $k \to \infty$, the sequence $\{x(k)\}_{k=1}^{\infty}$ converges to some point p inside A. Then $f(p)$ will presumably be the maximum value of f. This is a great idea!

However, there are two critical issues.

1) How can you ensure the sequence $\{x(k)\}_k$ in A converges to a point $p \in \mathbb{R}^n$?

2) How can you ensure the limiting point p belongs to A?

As you have seen, you can successfully address the second question if you assume the set A is closed. Is this also enough to address the first question? Unfortunately, it is not.

Example 2.4.1 Consider the closed set $A = [-1, 1]$. The sequence $x(k) = (-1)^k$ for $k \in \mathbb{N}^+$ lying in A does not converge. On the other hand, notice that $x(2k) = 1$ for all $k \in \mathbb{N}^+$ so the subsequence of even terms $x(2), x(4), x(6), \ldots$ converges to $1 \in A$.

Consider the closed set $B = [0, \infty)$. The sequence $y(k) = 2^k$ for $k \in \mathbb{N}^+$ lying in B does not converge and, no matter how you pick terms from this sequence, you cannot form a new subsequence which will converge. The issue here is that the set B is not bounded, so the terms of the sequence $x(k)$ can always stay far apart.

It is silly to ask every sequence in a closed set to converge; you can always construct an "alternating" sequence like the example above which will not converge. On the other hand, this illustrates that you can pick infinitely terms from a sequence and form a new subsequence which will converge. If your set A has the property that you can *always* do this *for any* sequence in A then you will have addressed the first question!

2.4.1 Definitions of compactness

This preliminary discussion inspires a mysterious yet magical definition, which is designed to address the core question of this section.

Definition 2.4.2 A set $A \subseteq \mathbb{R}^n$ is **compact**[6] if every sequence of A has a subsequence which converges to a point lying inside A.

Notice compact sets are literally defined to address the initial two questions.

Example 2.4.3 Here are some common examples of non-compact sets.

- The interval $A = (0, 1]$ in $\mathbb{R}$ is not compact. Verify this with the sequence $\{\frac{1}{k}\}_{k=1}^{\infty}$ since it (and any subsequence) converges to 0 which does not belong to $A = (0, 1]$.
- The set $B = \{\frac{1}{n} : n \in \mathbb{N}\}$ in $\mathbb{R}$ is not compact for the same reason as A.
- The set $\mathbb{R}^n$ is not compact since, for example, the sequence $x(k) = (k, k, \ldots, k)$ and any subsequence of it does not converge.
- The unit open ball $B_1(0)$ in $\mathbb{R}^n$ is not compact because you can construct a sequence which converges to a point on the unit sphere $\partial B_1(0) = S^{n-1}$.

The non-examples of $(0, 1]$ and $\mathbb{R}^n$ illustrate two necessary conditions for compactness. First, the set must be closed. Second, the set cannot be arbitrarily large.

[6] Truthfully speaking, this definition is for *sequentially compact* sets. See a course in topology or an introduction to real analysis for the standard definition of *compact* sets. These notions coincide for $\mathbb{R}^n$.

Definition 2.4.4 A set $A \subseteq \mathbb{R}^n$ is **bounded** if there is $R > 0$ such that $A \subseteq \{x \in \mathbb{R}^n : ||x|| < R\}$. If a set is not bounded, then it is **unbounded**.

Example 2.4.5 The set $\mathbb{R}^n$ is unbounded. The interval $[1,4]$ is bounded since $[1,4] \subseteq (-5,5)$.

If a set A is unbounded, then it cannot be compact because you can construct a sequence $\{x(k)\}_k$ lying in A where $\|x(k)\| \to \infty$. As a result, the terms will never get close to each other, so no subsequence can converge. Now, which sets are compact?

Example 2.4.6 Here are some common examples of compact sets.
- The empty set $\varnothing$ is (vacuously) compact.
- Any finite set A is compact since every sequence must repeat at least one element a infinitely often. The constant subsequence of this element trivially converges to $a \in A$.
- The closed interval $[a,b]$ in $\mathbb{R}$ is compact but this is not easy to see from the definition.

These examples illustrate how difficult it is to verify whether a set A is compact by definition. You must check *every* subsequence of *every* sequence in A. That seems ludicrous! Luckily, there is a miraculously natural equivalent[7] definition of compactness.

Theorem 2.4.7 (Bolzano–Weierstrass[8]) A set in $\mathbb{R}^n$ is compact if and only if it is both closed and bounded.

Proof. Postponed to Section 2.4.3. ∎

This powerful theorem makes it dramatically easier to verify compactness.

Example 2.4.8 Let $a, b \in \mathbb{R}$ with $a < b$. The interval $[a,b]$ is closed and bounded, so it is compact by the Bolzano–Weierstrass theorem. The interval $[a,\infty)$ is closed but not bounded, so it is not compact. The open interval (a,b) is bounded but not closed, so it is not compact. The interval $(-\infty, b)$ is neither bounded nor closed, so it is not compact. The details to check whether these sets are bounded or closed are left as exercises.

Example 2.4.9 Any finite set of $\mathbb{R}^n$ (including the empty set) is closed and bounded. Hence it is compact. The set $\mathbb{R}^n$ is not bounded and hence not compact.

Example 2.4.10 Fix $r > 0$ and $a \in \mathbb{R}$. The open ball $B_r(a)$ is not closed so it is not compact. The closed ball $\overline{B_r(a)} = \{x \in \mathbb{R}^n : ||x-a|| \le r\}$ and the sphere $\partial B_r(a) = \{x \in \mathbb{R}^n : ||x-a|| = r\}$ are closed and bounded, so both sets are compact.

Equipped with this tool, you can do quite a lot.

2.4.2 Set operations and subsets

Like open sets and closed sets, compact sets respect set operations in a natural way.

Lemma 2.4.11 All of the following are true for sets in $\mathbb{R}^n$.
(a) A finite union of compact sets is compact.
(b) A finite or infinite intersection of compact sets is compact.
(c) A finite Cartesian product of compact sets is compact.

[7] This equivalence is special to $\mathbb{R}^n$. It is not true for all topological spaces. See a course in topology for details.
[8] This statement often appears as Heine-Borel instead of Bolzano-Weierstrass. This disparity occurs because other texts use a more general definition of compact than Definition 2.4.2. The Bolzano-Weierstrass theorem shows that *sequentially* compact sets are the same as closed and bounded sets, so Theorem 2.4.7 is properly named.

Proof. This is left as an exercise. Check that the same statements are true if you replace "compact" with "bounded". Then apply Lemma 2.3.18 and the Bolzano–Weierstrass theorem. ■

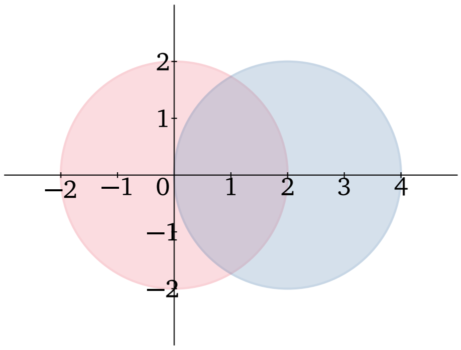
You can use this lemma to produce compact sets using simpler compact sets.

Example 2.4.12 For $a, b \in \mathbb{R}$ with $a < b$, since the interval $[a, b]$ is compact, the hypercube $[a, b]^n = [a, b] \times \cdots [a, b]$ is compact by Lemma 2.4.11.

Example 2.4.13 The set

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, (x-2)^2 + y^2 \leq 4\}.$$

is the intersection of $B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$ and $C = \{(x, y) \in \mathbb{R}^2 : (x-2)^2 + y^2 \leq 4\}$. Both B and C are compact so, by Lemma 2.4.11, the set $A = B \cap C$ is compact.



Subsets of compact sets also satisfy a nice property.

Lemma 2.4.14 Let A be a compact set in $\mathbb{R}^n$. If $B \subseteq A$ and B is closed then B is compact.

Proof. This follows directly from the Bolzano–Weierstrass theorem but the proof here will be from the definition of compactness. Let $\{x(k)\}_k$ be a sequence in B. Since $B \subseteq A$, the sequence also lies in A so, by the compactness of A, it has a subsequence $\{x(m(k))\}_{k=1}^{\infty}$ converging to some $p \in A$. The subsequence $\{x(m(k))\}_{k=1}^{\infty}$ also lies in B. As B is closed and $x(m(k)) \to p$, this implies $p \in B$. Hence, B is compact. ■

This allows you to produce even more compact sets from other compact sets.

Example 2.4.15 Consider the following subset of $\mathbb{R}^3$:

$$S = \{(x, y, z) \in \mathbb{R}^3 : 1 \leq x^2 + y^2 \leq 2, -1 \leq z \leq 1\}$$

Intuitively, this is a hollowed out cylinder in $\mathbb{R}^3$. You can prove that this is a compact set using the above two lemmas. First, the set

$$A = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\}$$

is a closed subset of the closed ball $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 2\}$. By Lemma 2.4.14, A is compact in $\mathbb{R}^2$. The closed interval $B = [-1, 1]$ is compact in $\mathbb{R}$. Thus, by Lemma 2.4.11,

$$A \times B = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\} \times [-1, 1] = S$$

is compact in $\mathbb{R}^3$, as desired.

All that remains is to prove the Bolzano–Weierstrass theorem.

2.4.3 Proof of Bolzano–Weierstrass

Both directions of the proof are tough, but it is easier to show the conditions of closed and bounded are necessary for compactness. The serious challenge is to show that they are sufficient. The proof begins with the easier direction.

Proof. ($\Longrightarrow$) Assume $A \subseteq \mathbb{R}^n$ is compact.

First, we show A is closed. Let p be a limit point of A. Then there exists a sequence $\{x(k)\}_k$ lying in A such that $x(k) \to p$. You can verify that any subsequence of $\{x(k)\}_k$ must therefore converge to p as well. Since A is compact, this implies $p \in A$ and hence A is closed.

Second, we show A is bounded by contrapositive. If A is not bounded, then $\forall k \in \mathbb{N}^+$, the set A is not a subset of $B_k(0)$ so there exists a point $x(k) \in A$ with $||x(k)|| > k$. Let $\{x(m(k))\}_{k=1}^{\infty}$ be any subsequence of $\{x(k)\}_{k=1}^{\infty}$, so $m : \mathbb{N}^+ \to \mathbb{N}^+$ is a strictly increasing function. It suffices to show that $x(m(k))$ does not converge to any point in $\mathbb{R}^n$. Fix $p \in \mathbb{R}^n$. We must verify that

$$\exists \varepsilon > 0 \text{ s.t. } \forall K \in \mathbb{N}^+, \exists k \in \mathbb{N}^+ \text{ s.t. } k \geq K \text{ and } ||x(m(k)) - p|| \geq \varepsilon.$$

Take $\varepsilon = 1$. Let $K \in \mathbb{N}^+$. Since $m : \mathbb{N}^+ \to \mathbb{N}^+$ is a strictly increasing function, set k to be large enough so that $m(k)$ is greater than K and the norm $||p||$. By rearranging the triangle inequality, we see by our choice of the sequence x that

$$||x(m(k)) - p|| \geq ||x(m(k))|| - ||p|| > m(k) - ||p|| \geq 1 = \varepsilon.$$

This shows $x(m(k))$ does not converge to p, as required. Since the subsequence and point p were arbitrary, the set A is not compact. This completes the proof of this direction.

($\Longleftarrow$) Assume $A \subseteq \mathbb{R}^n$ is closed and bounded. Let $\{x(k)\}_{k=1}^{\infty}$ be a sequence lying in A. We claim that $\{x(k)\}_k$ has a convergent subsequence. Assuming this claim, the subsequence must converge to a point p in A, since A is closed, and this would prove A is compact.

It remains to prove the claim. Express the sequence $x(k) = (x_1(k), \dots, x_n(k))$ in terms of its components, so $x_i(k) \in \mathbb{R}$ for all $k \in \mathbb{N}^+$ and all $i \in \{1, \dots, n\}$. For $i \in \{1, \dots, n\}$, we show that each sequence $\{x_i(k)\}_{k=1}^{\infty}$ of real numbers is bounded in $\mathbb{R}$. Since A is bounded, there exists $R > 0$ such that $A \subseteq B_R(0)$ so $\forall a \in A$, we have $||a|| < R$. Writing $a = (a_1, \dots, a_n)$, it follows that $|a_i| \leq ||a|| < R$ for all $i \in \{1, \dots, n\}$. This implies that $A \subseteq (-R, R)^n$ so $x(k) \in (-R, R)^n$ for every $k \in \mathbb{N}^+$. Thus, for $i \in \{1, \dots, n\}$, the coordinate sequence $\{x_i(k)\}_{k=1}^{\infty}$ is bounded within $(-R, R)$.

Every bounded sequence in $\mathbb{R}$ has a convergent subsequence[9]. Thus, the first coordinate sequence $\{x_1(k)\}_{k=1}^{\infty}$ in $\mathbb{R}$ has a convergent subsequence, say $\{x_1(m_1(k))\}_{k=1}^{\infty}$ for some strictly increasing function $m_1 : \mathbb{N}^+ \to \mathbb{N}^+$. It converges to some real number $p_1 \in \mathbb{R}$.

The second coordinate sequence $\{x_2(k)\}_{k=1}^{\infty}$ also has the subsequence $\{x_2(m_1(k))\}_{k=1}^{\infty}$ in $\mathbb{R}$ and this subsequence is also a bounded sequence $x_2 \circ m_1 : \mathbb{N}^+ \to \mathbb{R}^n$. Thus, this *subsequence* has a convergent subsequence $\{(x_2 \circ m_1)(m_2(k))\}_{k=1}^{\infty}$ where $m_2 : \mathbb{N}^+ \to \mathbb{N}^+$ is a strictly increasing function. This subsequence converges to some real number $p_2 \in \mathbb{R}$, so

$$(x_2 \circ m_1)(m_2(k)) = x_2(m_1(m_2(k))) = x_2((m_1 \circ m_2)(k)) \to p_2$$

as $k \to \infty$. Any subsequence of a convergent sequence still converges, so the first coordinate sequence still converges to p along $(m_1 \circ m_2)(k)$. That is, $x_1((m_1 \circ m_2)(k)) \to p_1$.

You can repeat this process for the third coordinate sequence and so on. Ultimately, this gives strictly increasing functions $m_1, \dots, m_n$ from $\mathbb{N}^+ \to \mathbb{N}^+$ and $p_1, \dots, p_n \in \mathbb{R}$ such that

$$\forall i \in \{1, \dots, n\},\, x_i(m(k)) \to p_i \quad \text{as } k \to \infty,$$

where $m = m_1 \circ m_2 \circ \cdots \circ m_n : \mathbb{N}^+ \to \mathbb{N}+$ is strictly increasing. By Theorem 2.2.16, the subsequence $\{x(m(k))\}_{k=1}^{\infty}$ converges to $p = (p_1, \dots, p_n) \in \mathbb{R}^n$. This completes the proof. ■

[9] This property of $\mathbb{R}$ is *completeness*, and you can assume it. See an introduction to real analysis for details.

Exercises for Section 2.4

Concepts and definitions

2.4.1 Which of the following sets are compact? Briefly explain why or why not.

(a) $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$

(b) $\{(x,y) \in \mathbb{R}^2 : x^2 - y^2 \leq 1\}$

(c) $B_2(0)$

(d) $\overline{B_2(0)} \setminus B_1(0)$

(e) $\left\{\frac{1}{n} : n \in \mathbb{N}^+\right\}$

(f) $\{a\}$ for a fixed $a \in \mathbb{R}^n$

2.4.2 Which statements are equivalent to "S is a compact set"? Briefly justify why or why not.

(a) S contains all its limit points.

(b) There exists $r > 0$ so that $S \subseteq \{x \in \mathbb{R}^n : \|x\| \leq r\}$.

(c) Any subset of S has a limit point.

(d) Any sequence in S converges with its limit lying in S.

(e) Any sequence in S has a convergent subsequence with its limit lying in S.

(f) S is bounded and closed.

(g) S is bounded and not open.

(h) S^c is unbounded and open.

2.4.3 You've been whacked with a bunch of theorems and definitions related to compact sets. You must navigate through this maze and recognize new relationships. Which statements are true or false? If true, briefly justify. If false, state a counterexample.

(a) Every subset A of a compact set B is closed.

(b) Every subset A of a compact set B is bounded.

(c) Every bounded subset A of a closed set B is compact.

(d) If A is closed and B is not compact then $A \cap B$ is not compact.

(e) If A and B are bounded then $A \cup B$ is bounded.

(f) If S is compact then $\overline{S}$ is compact.

(g) If S is compact (in $\mathbb{R}^n$) then $S \times [0,1]$ is compact (in $\mathbb{R}^{n+1}$).

Proofs

2.4.4 Divergent sequences can still have convergent subsequences.

(a) Show that the divergent sequence $x(k) = (-1)^k$ has a convergent subsequence.

(b) Show that the divergent sequence $\{(\cos k, \sin k)\}_k$ has a convergent subsequence.

2.4.5 Let $\{x(k)\}_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n$ converging to some point $p \in \mathbb{R}^n$. Prove that any subsequence of $\{x(k)\}_{k=1}^{\infty}$ also converges to p.

2.4.6 You can construct compact sets out of other simple sets using set operations and topological properties of compact sets. Show that each set below is compact.

(a) Show that $A = \{(x,y,z) \in \mathbb{R}^3 : 0 \leq x,y,z \leq 1 \text{ or } (x-1)^2 + (y-1)^2 + (z-1)^2 \leq 1\}$ is compact.

(b) Show that $B = \{(x,y,z) \in \mathbb{R}^3 : -9 \leq x,y,z \leq 9 \text{ and } x^2 + y^2 + z^2 \geq 4\}$ is compact.

2.4.7 The Bolzano–Weierstrass theorem also makes it easier to prove properties of compact sets. Use this powerful theorem and lemmas about closed sets to prove facts about compact sets.

(a) Let A be a compact set in $\mathbb{R}^n$. If $B \subseteq A$ and B is closed then B is compact.

(b) Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Prove that if A and B are compact then the union $A \cup B$ is compact.

2.4.8 Many authors write terse proofs or condense difficult proofs, like the Bolzano–Weierstrass theorem. The ideas are there but each step can feel like a big leap without enough detail. You will learn to identify the missing details and fill them in yourself. Here is a terse proof that if a set is compact then it is closed.

1. *Assume $A \subseteq \mathbb{R}^n$ is compact. Let p be a limit point of A.*
2. *There exists a sequence $\{x(k)\}_k$ lying in A such that $x(k) \to p$.*
3. *You can verify that any subsequence of $\{x(k)\}_k$ must therefore converge to p as well.*
4. *Since A is compact, this implies $p \in A$.*
5. *Hence, A is closed.*

(a) By definition of limit points, the sequence should lie in $A \setminus \{p\}$. Why can the author omit this?

(b) How does Line 4 follow from Lines 2 and 3? Fill in the details by adding an extra line or two.

(c) The conclusion in Line 5 appears out of nowhere. What was the author trying to show?

2.4.9 Continuing the previous exercise, you will fill in the missing details of a big proof. Here is a proof that if a set is compact then it is bounded. The proof is by contrapositive.

1. *Assume $A \subseteq \mathbb{R}^n$ is not bounded.*
2. *For all $k \in \mathbb{N}^+$, there exists a point $x(k) \in A$ with $||x(k)|| > k$.*
3. *Let $m : \mathbb{N}^+ \to \mathbb{N}^+$ be strictly increasing, so $\{x(m(k))\}_{k=1}^{\infty}$ is a subsequence of $\{x(k)\}_{k=1}^{\infty}$.*
4. *We must verify that*

$$\exists \varepsilon > 0 \text{ s.t. } \forall K \in \mathbb{N}^+, \exists k \in \mathbb{N}^+ \text{ s.t. } k \geq K \text{ and } ||x(m(k)) - p|| \geq \varepsilon.$$

5. *Fix $p \in \mathbb{R}^n$. Take $\varepsilon = 1$. Let $K \in \mathbb{N}^+$.*
6. *Set k to be large enough so that $m(k)$ is greater than K and the norm $||p||$.*
7. *It follows that*

$$||x(m(k)) - p|| \geq ||x(m(k))|| - ||p|| > m(k) - ||p|| \geq 1 = \varepsilon.$$

8. *Hence, $x(m(k)) \not\to p$.*
9. *Therefore no subsequence of $\{x(k)\}_{k=1}^{\infty}$ converges, so the set A is not compact.*

(a) Explain why Line 2 follows from Line 1.

(b) Justify Line 6 with more detail. In other words, why can you choose k that way?

(c) The conclusion in Line 9 seems to appear out of nowhere. What was the author trying to show? Add an extra line or two to explain how Line 9 follows.

2.5. Limits

You have studied sets in $\mathbb{R}^n$ and developed tools to study their topological properties. All of your labour will help generalize a fundamental notion of calculus: *limits*. Luckily, the underlying formalities and proofs are very similar to single variable calculus, but there are some new subtleties in higher dimensions.

Limits in $\mathbb{R}$ can only approach a real number from either the left or the right. Furthermore, limits near the endpoints of an interval $[a, b]$ in $\mathbb{R}$ are simple to understand; they can only be approached from one side so these are often separately defined as one-sided limits. Having a different definition of limits for endpoints versus interior points can be rather annoying and there is rarely any meaningful difference in proofs with each type of limit.

Limits in $\mathbb{R}^n$, on the other hand, need to be able to approach a point from any possible direction and in any weird way. It is therefore not reasonable to come up with a separate definition for limits at boundary points, since there is no way to classify how you might approach a boundary point. Thus, a good definition of a limit in $\mathbb{R}^n$ should be sensitive to the function's domain $A \subseteq \mathbb{R}^n$, yet be indistinguishable for boundary points versus interior points of A. As you shall see, you have prepared the necessary ingredients to solve this problem using the definition of limit points with open balls or with sequences.

2.5.1 Formal definitions

The formal definition of a limit in $\mathbb{R}$ generalizes almost effortlessly by replacing the absolute value with the corresponding norm (of the domain or codomain).

Definition 2.5.1 Let $f : A \to \mathbb{R}^m$ be a function with $A \subseteq \mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be a limit point of A and let $b \in \mathbb{R}^m$. Define b to be the **limit of f at a** provided

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \varepsilon.$$

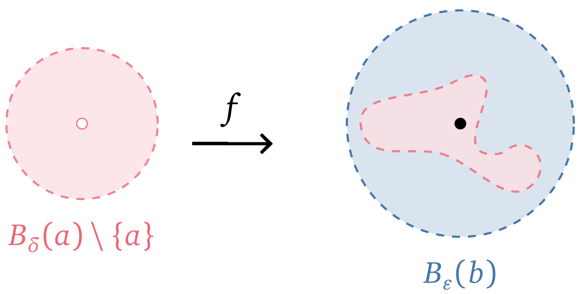
If the above holds, then write $\lim_{x \to a} f(x) = b$ or write $f(x) \to b$ as $x \to a$.

Remark 2.5.2 If there exists $b \in \mathbb{R}^m$ such that $\lim_{x \to a} f(x) = b$ then the **limit of f at a exists**. Otherwise, the limit does not exist.

Informally, this definition says

> *"The value $f(x)$ can be made ε-close to b provided x is δ-close to a."*

Geometrically, this is illustrated with the following sketch.



Since the definition is so similar to limits over $\mathbb{R}$, the proof structure with the formal definition for multivariable limits is unsurprisingly similar to single variable limits.

Example 2.5.3 Here is a proof that $\lim_{(x,y)\to(2,3)}(x+y) = 5$ using the formal definition of a limit.

Proof. Fix $\varepsilon > 0$. Take $\delta = \varepsilon/2$. Let $(x, y) \in \mathbb{R}^2$. Assume $0 < \|(x, y) - (2,3)\| < \delta$. This implies that

$$|x-2| \leq \sqrt{(x-2)^2 + (y-3)^2} = \|(x,y) - (2,3)\| < \delta \tag{2.5.1}$$

and similarly $|y-3| < \delta$. Then

$$\begin{aligned}
\|(x+y) - 5\| = |x+y-5| = |(x-2)+(y-3)| & \\
\leq |x-2| + |y-3| & \qquad \text{by the triangle inequality} \\
< 2\delta & \qquad \text{as } |x-2| < \delta \text{ and } |y-3| < \delta \\
= \varepsilon & \qquad \text{as } \delta = \varepsilon/2.
\end{aligned}$$

This completes the proof. ■

There are a few new ideas required to estimate the norms that arise in the proof. One idea is to estimate the distance between a fixed coordinate of two points by the (norm) distance between the two points. This idea was displayed in (2.5.1) of the previous example. Another idea is to use intermediate approximations by "fixing one variable at a time". This occurs in the previous example with the triangle inequality. The next example again shows both ideas.

Example 2.5.4 Here is a proof that $\lim_{(x,y)\to(2,3)} xy = 6$ using the formal definition of a limit.

Proof. Fix $\varepsilon > 0$. Take $\delta = \min\{\varepsilon/6, 1\}$. Let $(x, y) \in \mathbb{R}^2$. Assume $0 < \|(x, y) - (2,3)\| < \delta$. Note that

$$|x-2| \leq \|(x,y)-(2,3)\| < \delta \quad \text{and} \quad |y-3| \leq \|(x,y)-(2,3)\| < \delta.$$

Moreover, as $\delta \leq 1$, this implies that $2 < y < 4$. Then

$$\begin{aligned}
\|xy - 6\| = |xy-6| = |xy - 2y + 2y - 6| & \\
\leq |xy-2y| + |2y-6| & \qquad \text{by the triangle inequality} \\
\leq |x-2| \cdot |y| + 2|y-3| & \\
\leq 4|x-2| + 2|y-3| & \qquad \text{as } 2 < y < 4 \\
< 6\delta & \qquad \text{as } |x-2| < \delta \text{ and } |y-3| < \delta \\
\leq \varepsilon & \qquad \text{as } \delta \leq \varepsilon/6.
\end{aligned}$$

This completes the proof. ■

Yet another idea is to actually use the existence of single variable limits as an input.

Example 2.5.5 Here is a proof that $\lim_{(x,y)\to(0,0)} \cos(x+y) = 1$.

Proof. Fix $\varepsilon > 0$. As $\cos(t)$ is continuous at $t = 0$ and $\cos 0 = 1$, there exists $\delta_1 > 0$ such that

$$\forall t \in \mathbb{R}, |t| < \delta_1 \implies |\cos(t) - 1| < \varepsilon. \tag{2.5.2}$$

Take $\delta = \delta_1/2$. Let $(x, y) \in \mathbb{R}^2$. Assume $0 < \|(x,y)\| < \delta$ so this implies, by the triangle inequality, that

$$|x+y| \leq |x| + |y| \leq 2\|(x,y)\| < 2\delta = \delta_1.$$

By (2.5.2), it follows that $|\cos(x+y) - 1| < \varepsilon$ as required. ■

After a careful inspection of Definition 2.5.1, notice there are two key differences. First, limits of a function are *only defined* at limit points of the domain. The remaining points of a function's domain deserve their own name.

Definition 2.5.6 Let $A \subseteq \mathbb{R}^n$ be a set. A point $a \in \mathbb{R}^n$ is an **isolated point of** A if $a \in A$ and a is not a limit point of A.

In particular, a limit at an isolated point of a function's domain is not defined.

Example 2.5.7 Let $A = [1,3) \cup \{7\}$ in $\mathbb{R}$. Define the function $f : A \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & 1 \le x < 2, \\ 100 & 2 \le x < 3, \\ e^{\pi} & x = 7. \end{cases}$$

The limit of f at a for any $a \in [1,3)$ is defined; it may or may not exist, but it is defined. The limit of f at 7 is *not* defined since 7 is an isolated point of the domain A.

Second, you only consider points belonging to the punctured ball *and the function's domain*; you cannot evaluate the function at points outside the function's domain.

Example 2.5.8 Define $f(x,y) = \log(1 - x^2 - y^2)$. Although f is only defined inside the open ball $B_1((0,0))$, the limit of f as $(x,y) \to (1,0)$ is still defined according to Definition 2.5.1. It does not exist, but it is defined.

Proving a limit *does not* exist using Definition 2.5.1 is possible but rather tedious. The negation of this formal open ball definition is quite a mouthful. Instead, you can use an equivalent sequential definition of the limit.

Theorem 2.5.9 (Sequential definition of limits) Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}^m$ be a function. Let $a \in \mathbb{R}^n$ be a limit point of A and let $b \in \mathbb{R}^m$. Then $\lim\limits_{x \to a} f(x) = b$ if and only if for every sequence of points $\{x(k)\}_k$ in $A \setminus \{a\}$ with $x(k) \to a$, the sequence of points $\{f(x(k))\}_k$ in $\mathbb{R}^m$ converges to b, i.e. $f(x(k)) \to b$.

Proof. This is left as a challenging exercise. See Theorem 2.2.17 for the key ideas. ■

This equivalent definition is your main tool for proving the non-existence of limits.

Example 2.5.10 Here is a proof that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ does not exist.

Proof. Define two sequences in $\mathbb{R}^2 \setminus \{(0,0)\}$ by $x(k) = (0, 1/k)$ and $y(k) = (1/k, 1/k)$ for $k \in \mathbb{N}^+$. By Theorem 2.2.16, these both converge to $(0,0)$. Notice that for $k \in \mathbb{N}^+$,

$$f(x(k)) = \frac{0}{0^2 + 1/k^2} = 0, \quad \text{and} \quad f(y(k)) = \frac{1/k^2}{1/k^2 + 1/k^2} = \frac{1}{2}.$$

Thus, $f(x(k)) \to 0$ and $f(y(k)) \to 1/2$ as $k \to \infty$ by Theorem 2.2.16. Since $\lim\limits_{k\to\infty} f(x(k)) \neq \lim\limits_{k\to\infty} f(y(k))$ where $\{x(k)\}_k$ and $\{y(k)\}_k$ both converge to $(0,0)$, the desired limit does not exist by Theorem 2.5.9. ■

This concludes the formal definitions of the limit of a function. Like single variable calculus, these definitions capture the same key ideas but they are not efficient for computation.

2.5.2 Basic properties

Some basic properties of limits are needed to speed up calculations, and to prove other theorems. The first and most fundamental property allows you to reduce limits of vector-valued functions to limits of real-valued functions.

Theorem 2.5.11 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Let a be a limit point of A and let $b = (b_1, \dots, b_m) \in \mathbb{R}^m$. Let $f_1, \dots, f_m$ be the component functions of f so $f = (f_1, \dots, f_m)$. Then

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

if and only if for all $i \in \{1, \dots, m\}$,

$$\lim_{x \to a} f_i(x) = b_i$$

Proof. ($\implies$) Assume $\lim_{x \to a} f(x) = b$. Fix $i \in \{1, \dots, m\}$. Let $\varepsilon > 0$. By assumption, there exists $\delta > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies \|f(x) - b\| < \varepsilon.$$

For any $x \in A$, we have that $|f_i(x) - b_i| \leq \|f(x) - b\|$ so the above implies with the same δ that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies |f_i(x) - b_i| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows that $f_i(x) \to b_i$ as $x \to a$.

($\impliedby$) Assume $\lim_{x \to a} f_i(x) = b_i$ for every $i \in \{1, \dots, m\}$. Let $\varepsilon > 0$. By assumption, for each $i \in \{1, \dots, m\}$, there exists $\delta_i > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta_i \implies |f_i(x) - b_i| < \frac{\varepsilon}{\sqrt{m}}. \tag{2.5.3}$$

Set $\delta = \min\{\delta_1, \dots, \delta_m\}$. Let $x \in A$. Assume $0 < \|x - a\| < \delta$. For each $i \in \{1, \dots, m\}$, we have $\delta \leq \delta_i$ so (2.5.3) implies that $|f_i(x) - b_i| < \varepsilon / \sqrt{m}$. It follows that

$$\begin{aligned}
\|f(x) - b\| = \sqrt{|f_1(x) - b_1|^2 + \cdots + |f_m(x) - b_m|^2} &\leq \sqrt{m \max_{1 \leq i \leq m} |f_i(x) - b_i|^2} \\
&< \sqrt{m \cdot \frac{\varepsilon^2}{m}} \\
&= \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this shows that $f(x) \to b$ as $x \to a$. ■

You can see how Theorem 2.5.11 is used in practice.

Example 2.5.12 Define $f(x, y) = (x + y, xy)$ so $f = (f_1, f_2)$ can be written using its coordinate functions $f_1(x, y) = x + y$ and $f_2(x, y) = xy$. By Examples 2.5.3 and 2.5.4, it follows that

$$\lim_{(x,y) \to (2,3)} f_1(x, y) = 5 \quad \text{and} \quad \lim_{(x,y) \to (2,3)} f_2(x, y) = 6.$$

Therefore, by Theorem 2.5.11,

$$\lim_{(x,y) \to (2,3)} f(x, y) = \left(\lim_{(x,y) \to (2,3)} f_1(x, y), \lim_{(x,y) \to (2,3)} f_2(x, y) \right) = (5, 6).$$

The second property is limits are unique.

Lemma 2.5.13 (Uniqueness of limits) Let a be a limit point of $A \subseteq \mathbb{R}^n$. Let $f : A \to \mathbb{R}^m$ and fix points $b_1, b_2 \in \mathbb{R}^m$. If $\lim_{x\to a} f(x) = b_1$ and $\lim_{x\to a} f(x) = b_2$, then $b_1 = b_2$.

Proof. This is left as an exercise. The proof is nearly identical to the single variable case. ■

Many standard properties also hold along with a few more since limits can be vector-valued.

Theorem 2.5.14 Let $A \subseteq \mathbb{R}^n$ be a set and let a be a limit point of A. Let f and g be $\mathbb{R}^m$-valued functions defined on A. Let ϕ be a real-valued function defined on A. Let $\lambda \in \mathbb{R}$ and $b \in \mathbb{R}^m$ be constants. All of the following hold:

(a) (*Constants*) $\lim_{x\to a} b = b$ and $\lim_{x\to a} x = a$.

(b) (*Linearity*) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist then $\lim_{x\to a}(f(x) + \lambda g(x))$ exists and

$$\lim_{x\to a}(f(x) + \lambda g(x)) = \lim_{x\to a} f(x) + \lambda \lim_{x\to a} g(x).$$

(c) (*Scalar product*) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} \phi(x)$ exist then $\lim_{x\to a}(\phi(x)f(x))$ exists and

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

(d) (*Dot product*) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist then $\lim_{x\to a}(f(x) \cdot g(x))$ exists and

$$\lim_{x\to a}(f(x) \cdot g(x)) = \left(\lim_{x\to a} f(x)\right) \cdot \left(\lim_{x\to a} g(x)\right).$$

Proof. These are all left as exercises. They are quite similar to the single variable case. You will not, however, be able to use single variable limit laws because these are all multivariable limits. Remember to write your proofs while keeping in mind which quantities are vectors and which are scalars. For (a) and (b), use the formal definition. For (c), use Theorem 2.5.11 to reduce to the case $m = 1$. Prove the case $m = 1$ using the formal definition. For (d), use (b) and (c). ■

Finally, there is a multivariable squeeze theorem for real-valued functions.

Theorem 2.5.15 (Squeeze theorem) Let $A \subseteq \mathbb{R}^n$ be a set and let a be a limit point of A. Let f, g, h be real-valued functions with domain A. Assume there exists $\delta > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies f(x) \leq g(x) \leq h(x)$$

If $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = b$ for some $b \in \mathbb{R}$ then $\lim_{x\to a} g(x) = b$.

Proof. This is left as an exercise. The proof is nearly identical to the single variable case. ■

This collection of limit properties will form the basis for quantities defined via limits.

2.5.3 Limits with infinity

The concept of infinity changes in higher dimensions. It is still defined using a limit but, since there are countlessly many ways to diverge outward in $\mathbb{R}^n$ for $n \geq 2$, the notion of $\pm\infty$ no longer makes any sense. Instead, for $\mathbb{R}^n$ with $n \geq 2$, you only consider whether the norm of a sequence grows arbitrarily large. This motivates a definition for limits at infinity.

Definition 2.5.16 Let $A \subseteq \mathbb{R}^n$ be unbounded. Let $f : A \to \mathbb{R}^m$ and let $b \in \mathbb{R}^m$. Define b to be the **limit of** $f(x)$ **as** $||x|| \to \infty$ provided

$$\forall \varepsilon > 0, \exists M > 0 \text{ s.t. } \forall x \in A, ||x|| > M \implies ||f(x) - b|| < \varepsilon.$$

If the above holds, then write $\lim_{||x|| \to \infty} f(x) = b$ or write $f(x) \to b$ as $||x|| \to \infty$.
If the above does not hold, the limit $\lim_{||x|| \to \infty} f(x)$ does not exist.

Limits at infinity do not behave quite the same in higher dimensions.

Example 2.5.17 In the two-dimensional plane, you can check that

$$\lim_{||(x,y)|| \to \infty} \frac{1}{x^2 + y^2} = 0 \quad \text{yet} \quad \lim_{||(x,y)|| \to \infty} \frac{1}{x^2} \text{ does not exist.}$$

This may seem strange at first but notice that the latter function $\frac{1}{x^2}$ does not tend to zero along *all possible sequences* $\{(x(k), y(k))\}_k$ with $||(x(k), y(k))|| \to \infty$. That feature is equivalent to Definition 2.5.16.

You can similarly define limits of *real-valued* functions diverging to infinity.

Definition 2.5.18 Let $A \subseteq \mathbb{R}^n$ be a set. Let a be a limit point of A. Let $f : A \to \mathbb{R}$ be a real-valued function. The **limit of** f **at** a **diverges to** $+\infty$ provided

$$\forall M > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, 0 < ||x - a|| < \delta \implies f(x) > M.$$

If the above holds, then write $\lim_{x \to a} f(x) = +\infty$ or write $f(x) \to +\infty$ as $x \to a$.
The definition for $f(x) \to -\infty$ as $x \to a$ is similar.

These limits also do not behave the same in higher dimensions.

Example 2.5.19 In the two-dimensional plane, you can check that

$$\lim_{(x,y) \to (0,0)} \frac{1}{x^2 + y^2} = +\infty \quad \text{yet} \quad \lim_{(x,y) \to (0,0)} \frac{x}{x^2 + y^2} \text{ does not exist and is not } \pm\infty.$$

Again, you can check the latter limit by using the sequential definition of the limit.

Despite these differences, these two definitions of limits involving infinity are quite similar to the one-dimensional case. You can prove analogous limit laws and formulate equivalent sequential versions. There are actually more definitions of limits with infinity but, instead of writing them all down, you can create the necessary definitions yourself using these two as inspiration. Definitions 2.5.16 and 2.5.18 will be enough for now.

Exercises for Section 2.5

Concepts and definitions

2.5.1 For each $S \subseteq \mathbb{R}^2$, state without proof its set of limit points S^* and set of isolated points $S \setminus S^*$.

(a) $A = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\}$.

(b) $B = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\} \cup \{(0, 0)\}$.

(c) $C = \{(x, y) \in \mathbb{R}^2 : x \neq 1\}$

(d) $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 2\} \setminus \{(0, 0)\}$

(e) $E = \mathbb{R}^2$

(f) $F = \mathbb{Z} \times \mathbb{Z}$

2.5.2 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be a limit point of A and let $b \in \mathbb{R}^m$. Which statements are equivalent to "$\lim_{x \to a} f(x) = b$"?

(a) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $||x - a|| < \delta \implies ||f(x) - b|| < \varepsilon$

(b) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $0 < ||x - a|| < \delta \implies ||f(x) - b|| < \varepsilon$

(c) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall x \in A,\ 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \varepsilon$

(d) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $x \in B_\delta(a) \implies f(x) \in B_\varepsilon(b)$

(e) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $x \in B_\delta(a)$ and $x \neq a \implies f(x) \in B_\varepsilon(b)$

(f) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $x \in B_\delta(a) \cap A$ and $x \neq a \implies f(x) \in B_\varepsilon(b)$

Computations

2.5.3 Theorem 2.5.11 reduces limits of a vector-valued function to limits of real-valued functions. Here you will investigate the power and limitations of this theorem.

(a) Consider $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (\cos t, \sin t)$.
Can you apply this theorem to evaluate $\lim_{t \to 0} f(t)$? If so, do it.

(b) Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x, y) = \left(\frac{\cos(x+y)}{x^2+y^2}, \sin(x + y)\right)$.
Can you apply this theorem to evaluate $\lim_{(x,y) \to (0,0)} f(x, y)$? If so, do it.

(c) Replace the ??? with the most useful values you can.

If I know how to determine limits and their existence for functions $\mathbb{R}^{???} \to \mathbb{R}^{???}$ then I can use this theorem to determine limits and their existence for functions $\mathbb{R}^n \to \mathbb{R}^m$.

2.5.4 Proving a limit does not exist by the ε-δ definition can be annoying to write down properly. Instead, you will take a different approach to show that

$$\lim_{(x,y) \to (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$$

does not exist. Set $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$.

(a) Aloy and Beta are reading a commonly found argument.

1. *Along the line* $(x,y)=(t,t)$*, we have that* $f(t,t)=0$ *for* $t\neq 0$.
2. *Thus,* $\lim_{t\to 0} f(t,t)=\lim_{t\to 0} 0=0$.
3. *Along the line* $(x,y)=(0,t)$*, we have that* $f(0,t)=\frac{0-t^2}{0+t^2}=-1$ *for* $t\neq 0$.
4. *Thus,* $\lim_{t\to 0} f(0,t)=\lim_{t\to 0}(-1)=-1$.
5. *Note both* (t,t) *and* $(0,t)$ *approach* $(0,0)$ *as* $t\to 0$.
6. *Since* $\lim_{t\to 0} f(t,t)\neq \lim_{t\to 0} f(0,t)$*, it follows that* $\lim_{(x,y)\to(0,0)} f(x,y)$ *does not exist.*

They have differing opinions on the quality of this argument.

> Aloy says *"How are they justifying Line 6? I do not know any lemma which shows that."*
>
> Beta says *"That's the definition of limits. If the limit exists, then it must be the same no matter how you approach it."*

Who is closer to the truth? Briefly explain.

(b) Prove $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist using the sequential definition of the limit.

Proofs

2.5.5 Here is a proof that $\lim_{(x,y)\to(2,3)} xy=6$ using the formal definition of a limit. There are several good ideas which you will uncover.

1. *Fix* $\varepsilon>0$*. Take* $\delta=\min\{\varepsilon/6,1\}$.
2. *Let* $(x,y)\in\mathbb{R}^2$*. Assume* $0<\|(x,y)-(2,3)\|<\delta$*. Note that*

$$|x-2|\leq\|(x,y)-(2,3)\|<\delta \quad \text{and} \quad |y-3|\leq\|(x,y)-(2,3)\|<\delta.$$

3. *Moreover, as* $\delta\leq 1$*, this implies that* $2<y<4$.
4. *Then*

$$\begin{aligned}\|xy-6\|=|xy-6|&=|xy-2y+2y-6|\\ &\leq|xy-2y|+|2y-6| && \text{by the triangle inequality}\\ &\leq|x-2|\cdot|y|+2|y-3|\end{aligned}$$

5. *As* $2<y<4$*, the righthand side is*

$$\begin{aligned}&\leq 4|x-2|+2|y-3|\\ &<6\delta && \text{as } |x-2|<\delta \text{ and } |y-3|<\delta\\ &=\varepsilon && \text{as } \delta\leq\varepsilon/6.\end{aligned}$$

Define $f:\mathbb{R}^2\to\mathbb{R}$ by $f(x,y)=xy$. The core of the proof is to estimate $|f(x,y)-6|$.

(a) Line 4 introduces an "intermediate approximation" to break up the estimation into two parts. This idea is extremely valuable and can be informally described as follows:

> *The error between $f(x,y)$ and 6 can be approximated by the error between $f(x,y)$ and ______ and the error between ______ and 6.*

Fill in the blanks with the same "intermediate approximation".

(b) How does Line 4 split the approximation into two parts? *Hint:* There are 2 mini-ideas.

(c) The choice of intermediate approximation is inspired by *fixing a variable*. Using Line 2, informally explain why this strategy is successful here.

2.5.6 Using the formal definition of the limit, prove that $\lim_{(x,y)\to(1,2)} (3x-5y) = -7$.

2.5.7 The style and structure of ε-δ proofs for limits is very similar to single variable calculus. However, the rough work and manipulation with inequalities can be trickier in the multivariable case.

(a) What is wrong with this "proof"[10]? Briefly explain.

1. *Consider the function $f(x,y):\mathbb{R}^2\to\mathbb{R}$ defined by $f(x,y)=2x^2+3xy+7y$. We want to show that*
$$\lim_{(x,y)\to(1,2)} f(x,y)=22$$
2. *Fix an arbitrary $\varepsilon>0$. We want to find $\delta>0$ such that if $\|(x,y)-(1,2)\|<\delta$, then $|f(x,y)-22|<\varepsilon$.*
3. *Notice,*
$$\begin{aligned}|f(x,y)-22| &= |2x^2+3xy+7y-22|\\ &= |2(x^2-1)+3(xy-2)+7(y-2)|\\ &\le 2|x+1||x-1|+3|xy-2|+7|y-2|\\ &= 2|x+1||x-1|+3|xy-y+y-2|+7|y-2|\\ &\le 2|x+1||x-1|+3|y||x-1|+10|y-2|\end{aligned}$$
4. *We need to bound coefficients $|x+1|$ and $|y|$ of $|x-1|$.*
First let us impose that $\delta\le 1$, then
$$\begin{aligned}\|(x,y)-(1,2)\|<\delta\le 1 &\implies \sqrt{(x-1)^2+(y-2)^2}<1\\ &\implies |x-1|<1 \quad \textit{and} \quad |y-2|<1\\ &\implies 0<x<2 \quad \textit{and} \quad 1<y<3\end{aligned}$$
Hence $|x+1|<3$ and $|y|<3$.
5. *Now $|f(x,y)-3|<6|x-1|+9|x-1|+10|y-2|$, clearly, take $\delta=\min\{1,\frac{\varepsilon}{25}\}$.*
It follows that
$$\|(x-y)-(1,2)\|<\delta \implies |x-1|<\frac{\varepsilon}{25} \quad \textit{and} \quad |y-2|<\frac{\varepsilon}{25}$$
6. *Thus $|f(x,y)-22|<6\cdot\frac{\varepsilon}{25}+9\cdot\frac{\varepsilon}{25}+10\cdot\frac{\varepsilon}{25}=\varepsilon$, as desired.* ■

(b) Rewrite this argument into a well-written formal proof. That is, show

$$\lim_{(x,y)\to(1,2)} f(x,y) = 22$$

where $f(x,y) = 2x^2 + 3xy + 7y$.

2.5.8 Use the ε-δ definition of a limit to show that $\displaystyle\lim_{(x,y)\to(1,2)} 2^{x+y} = 8$.

2.5.9 Use the ε-δ definition of a limit to show that $\displaystyle\lim_{t\to 0}(\cos t, \sin t) = (1,0)$.

2.5.10 Prove the linearity property of limits, namely Theorem 2.5.14(b).

2.5.11 Prove the squeeze theorem, namely Theorem 2.5.15.

Applications and beyond

2.5.12 Ideas and arguments for limits can be naturally extended to limits at infinity.

(a) Prove that $\displaystyle\lim_{||(x,y)||\to\infty} \frac{1}{x^2+y^2} = 0$ using the formal definition.

(b) Let $A \subseteq \mathbb{R}^n$ be unbounded. Let $f : A \to \mathbb{R}^m$ and let $b \in \mathbb{R}^m$. Conjecture an equivalent sequential definition for $\displaystyle\lim_{||x||\to\infty} f(x) = b$.

(c) Use your new definition to prove that the limit $\displaystyle\lim_{||(x,y)||\to\infty} \frac{1}{x^2}$ does not exist.

2.5.13 Limits can be used to compare rates of convergence.

(a) Prove that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^{137}y^{223}}{||(x,y)||^{360}}$ does not exist.

(b) Prove that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^{137}y^{223}}{||(x,y)||^{237}} = 0$

2.5.14 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be an **interior** point of A and let $b \in \mathbb{R}^m$. Prove that

$$\forall \varepsilon > 0, \exists \delta_1 > 0 \text{ s.t. } \forall x \in A,\ 0 < ||x-a|| < \delta_1 \implies ||f(x) - b|| < \varepsilon$$

is equivalent to

$$\forall \varepsilon > 0, \exists \delta_2 > 0 \text{ s.t. } \forall x \in \mathbb{R}^n,\ 0 < ||x-a|| < \delta_2 \implies ||f(x) - b|| < \varepsilon.$$

This equivalence holds only for interior points, but not limit points in general.

2.5.15 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be an **isolated** point of A. For *every* $b \in \mathbb{R}^m$, prove

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A,\ 0 < ||x-a|| < \delta \implies ||f(x) - b|| < \varepsilon.$$

This explains why the definition of a limit (Definition 2.5.1) excludes isolated points; otherwise, limits would not be unique!

2.5.16 Prove that sequential definition of the limit is equivalent to the open ball definition of the limit, namely Theorem 2.5.9.

[10] It is inspired by a proof written as a standard example in another textbook. Beware when reading other sources.

2.6. Continuity

With the foundations for limits of maps $\mathbb{R}^n \to \mathbb{R}^m$, the next natural mathematical step is to explore continuity in higher dimensions. There are many similarities with continuity for maps $\mathbb{R} \to \mathbb{R}$ from single variable calculus. The definitions and basic properties parallel each other in almost every way, but you must grapple with the subtlety of limits in higher dimensions whereby you can approach a point countlessly many ways. This feature (not a bug) leads to many possible ways that a function can be discontinuous.

Now, the desire to generalize calculus concepts to higher dimensions is an admirable one, but you can dig deeper beyond this motivation. What does "continuity" really mean? As you shall see, its true purpose is to preserve topological properties of sets.

2.6.1 Formal definitions

Inspired by single variable calculus, you may guess that a function $f : A \to \mathbb{R}^m$ is continuous at a point $a \in A$ whenever

$$\lim_{x \to a} f(x) = f(a). \tag{2.6.1}$$

This is nearly correct but there is an annoying issue with this attempted definition: the limit is only defined if a is a limit point of A. If a is an isolated point of A, then the limit is not defined at all! If you expand (2.6.1) with quantifiers, you will write

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon.$$

There is no harm in replacing "$0 < ||x - a|| < \delta$" with "$||x - a|| < \delta$" since f is defined at a and $||f(x) - f(a)|| = 0$ when $x = a$. This leads to a better definition of continuity.

Definition 2.6.1 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. Let $a \in A$ be a point. The function f is **continuous at** a provided

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, ||x - a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon.$$

This definition is better than (2.6.1) because it includes isolated points and also gives the expected criterion (2.6.1) for limit points.

Lemma 2.6.2 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. Let $a \in A$ be a point. Both of the following hold:

(a) If a is an isolated point of A, then f is continuous at a.
(b) If a is a limit point of A, then f is continuous at a if and only if $\lim_{x \to a} f(x) = f(a)$.

Proof. (Sketch) (a) If a is an isolated point of A then there exists $\delta > 0$ such that $A \cap B_\delta(a) = \{a\}$. Thus, the implication in Definition 2.6.1 is automatically satisfied for any $\varepsilon > 0$ with this fixed choice of δ. (b) You can verify this with the arguments preceding Definition 2.6.1. ■

That is enough fussing over the definition. Here are some basic examples.

Example 2.6.3 Set $f(x, y) = x + y$ and $g(x, y) = xy$. From Examples 2.5.3 and 2.5.4,

$$\lim_{(x,y) \to (2,3)} f(x, y) = 5 = f(2, 3), \qquad \text{and} \quad \lim_{(x,y) \to (2,3)} g(x, y) = 6 = g(2, 3).$$

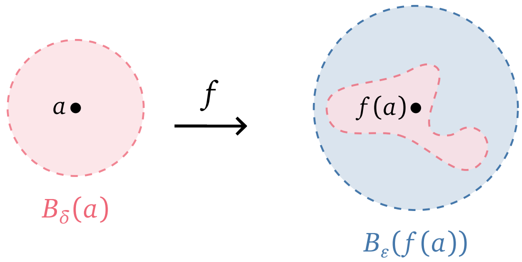
Thus, both f and g are continuous at $(2, 3)$.

Example 2.6.4 Define $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ by $f(x) = 1/x$. Note the statement "*f is discontinuous at* 0" is invalid and nonsensical. The function f is not defined at 0, and Definition 2.6.1 only permits you to discuss continuity at points inside the domain. It is correct to instead say "*f is not defined at* 0." On the other hand, if $g : \mathbb{R} \to \mathbb{R}$ is given by $g(0) = 0$ and $g(x) = 1/x$ for $x \neq 0$, then it is correct and valid to say "*g is discontinuous at* 0."

There is an elegant geometric interpretation of continuity using subsets and open balls. Keeping with the notation in Definition 2.6.1, notice f is continuous at a is equivalent to

$$\begin{aligned} &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, x \in B_\delta(a) \implies f(x) \in B_\varepsilon(f(a)) \\ \iff &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A \cap B_\delta(a), f(x) \in B_\varepsilon(f(a)) \\ \iff &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } f(A \cap B_\delta(a)) \subseteq B_\varepsilon(f(a)). \end{aligned}$$

This newly formulated definition can be illustrated with the following diagram.



While this framing may not make a proof easier, it helps to visualize the underlying ideas.

As you might expect, there is also a sequential definition of continuity.

Theorem 2.6.5 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. Let $a \in A$ be a point. Then f is continuous at a if and only if for every sequence $\{x(k)\}_k$ in A converging to a, the sequence $\{f(x(k))\}_k$ in $\mathbb{R}^m$ converges to $f(a)$.

Proof. Left as an exercise. First consider isolated points. For limit points, use Theorem 2.5.9. ■

This theorem can useful in many situations, including verifying a discontinuity. Next, you can introduce some terminology to conveniently discuss continuity on subsets of a domain.

Definition 2.6.6 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. For a subset $S \subseteq A$, the function f is **continuous on** S if f is continuous at a for every $a \in S$. The function f is **continuous** if f is continuous on its domain A.

Checking continuity on subsets is not much different than checking continuity at a point.

Example 2.6.7 The function $f(x,y) = x + y$ is continuous on its domain $\mathbb{R}^2$. Here is a proof.

Proof. Fix $(a,b) \in \mathbb{R}^2$. Let $\varepsilon > 0$. Take $\delta = \frac{\varepsilon}{2}$. Let $(x,y) \in \mathbb{R}^2$. Assume $||(x,y) - (a,b)|| < \delta$ so, as usual, $|x - a| < \delta$ and $|y - b| < \delta$. Then

$$\begin{aligned} ||f(x,y) - f(a,b)|| = |(x+y) - (a+b)| &= |x - a + y - b| \\ &\le |x-a| + |y-b| && \text{by the triangle inequality} \\ &< \delta + \delta && \text{as } |x-a| < \delta \text{ and } |y-b| < \delta \\ &= 2\delta = \varepsilon, \end{aligned}$$

since $\delta = \varepsilon/2$. This completes the proof. ■

Remark 2.6.8 Notice the function $f(x,y)=x+y$ is linear and therefore simple enough so δ only depends on ε. For other functions, the parameter δ may depend on (a,b) as well as ε.

Example 2.6.9 Every function $f:\mathbb{Z}^2\to\mathbb{R}$ is continuous, because all points in $\mathbb{Z}^2$ are isolated.

Example 2.6.10 The function $f:\mathbb{R}\setminus\{0\}\to\mathbb{R}$ given by $f(x)=1/x$ is continuous, because f is continuous at every point in its domain. However, the function $g:\mathbb{R}\to\mathbb{R}$ given by $g(0)=0$ and $g(x)=1/x$ for $x\neq 0$ is not continuous, because g is not continuous at 0.

Example 2.6.11 Define

$$F(x,y)=\begin{cases}x+y & \text{if } (x,y)\neq(2,3)\\ 237 & \text{otherwise.}\end{cases}$$

By Example 2.6.7, F is continuous on $\mathbb{R}^2\setminus\{(2,3)\}$, but F is not continuous at $(2,3)$ since

$$\lim_{(x,y)\to(2,3)} F(x,y)=5\neq 237=F(2,3).$$

This discontinuity is a *removable discontinuity*. It is harmless and straightforward to describe.

Recall in single variable calculus, there were 3 common categories for discontinuities: removable, jump, or infinite. Discontinuities in higher dimensions are much more diverse.

Example 2.6.12 Define $G(x,y)=\frac{xy}{x^2+y^2}$. Notice G is not continuous at $(0,0)$ because it is not even defined at $(0,0)$. You can fix this by defining

$$H(x,y)=\begin{cases}\frac{xy}{x^2+y^2} & \text{if } (x,y)\neq(0,0)\\ 0 & \text{otherwise,}\end{cases}$$

so H is defined at $(0,0)$. However, H is not continuous at $(0,0)$ since $\lim\limits_{(x,y)\to(0,0)} H(x,y)$ does not exist by Example 2.5.10. The graph of H illustrates how strange this discontinuity looks; view this Math3D demo of its graph. It is like a pinched piece of paper.

Example 2.6.13 For $(x,y)\neq(0,0)$, define

$$f(x,y)=\frac{1}{x^2+y^2},\qquad g(x,y)=\frac{x}{x^2+y^2},\qquad h(x,y)=\frac{x}{x+y},$$

and set $f(0,0)=g(0,0)=h(0,0)=0$. The functions f, g, and h are all discontinuous for the same reason as the function H in Example 2.6.12, namely, their limits at $(0,0)$ do not exist. However, the discontinuities all look completely different! View the graphs of f,g,h in this Math3D demo by toggling each surface one at a time. Notice $f(x,y)\to+\infty$ as $(x,y)\to(0,0)$ but the same is not at all true for g or h.

The above set of examples illustrates how more dimensions dramatically complicate the nature of discontinuities. Notice these examples are only about maps $\mathbb{R}^2\to\mathbb{R}$ yet they are already so wild! It only gets worse as the dimensions increase. There are entire fields of study dedicated to studying the surfaces created by discontinuities. Your goals will be more modest. You need to identify whether or not a function is continuous at a point; even that is not so easy.

2.6.2 Basic properties

Checking continuity by the definition is tedious; you will need the help of some theorems. As with limits, the first and most fundamental property allows you to reduce checking continuity of vector-valued functions to continuity of real-valued functions.

Theorem 2.6.14 The map $f = (f_1, \ldots, f_m) : A \to \mathbb{R}^m$ is continuous at $a \in A$ if and only if for each $i \in \{1, \ldots, m\}$, the component function f_i is continuous at a.

Proof. This follows almost immediately from Theorem 2.5.11 and Lemma 2.6.2. ■

Example 2.6.15 By Theorem 2.6.14, the function $F(x, y) = (x + y, xy)$ is continuous at $(2, 3)$ because each of its component functions is continuous at $(2, 3)$ from Example 2.6.3.

This property also gives a proof that some of your favourite functions are continuous.

Lemma 2.6.16 Every linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ is continuous.

Proof. By Theorem 2.6.14, it suffices to check linear maps of the form $\mathbb{R}^n \to \mathbb{R}$. That is, for fixed $c_1, \ldots, c_n \in \mathbb{R}$, you must show $f(x_1, \ldots, x_n) = c_1x_1 + \cdots + c_nx_n$ is continuous on $\mathbb{R}^n$. This is left as an exercise to prove by the definition. ■

Example 2.6.17 Lemma 2.6.16 shows all of the following maps are continuous.

- The identity map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\mathrm{id}(x) = x$ is linear and hence continuous.
- Fix $i \in \{1, 2, \ldots, n\}$. The ith coordinate projection map $\pi_i : \mathbb{R}^n \to \mathbb{R}$ defined by $\pi_i(x) = x_i$ is linear and hence continuous.
- For any $m \times n$ matrix A, the linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(x) = Ax$ is continuous. By a theorem in linear algebra, this actually describes all possible linear maps.

There is also the standard list of properties that follows from limits.

Theorem 2.6.18 Let $A \subseteq \mathbb{R}^n$ and let $a \in A$. Let f and g be $\mathbb{R}^m$-valued functions defined on A. Let ϕ be a real-valued function defined on A. Let $\lambda \in \mathbb{R}$. All of the following hold:

(a) If f and g are continuous at a then the function $f + \lambda g$ is continuous at a.
(b) If f and ϕ are continuous at a then their scalar product ϕf is continuous at a.
(c) If f and g are continuous at a then their dot product $f \cdot g$ is continuous at a.

Proof. Again, by Theorem 2.6.14, it suffices to check all of these properties for real-valued functions, that is, when $m = 1$. Notice (b) and (c) are the same in this case. Each statement can then be shown directly by the definition of continuity. This is left as an exercise. ■

This allows you to create more continuous functions.

Example 2.6.19 Define the map $f : \mathbb{R}^n \to \mathbb{R}$ by $N(x) = ||x||^2 = x_1^2 + \cdots + x_n^2$. Notice f is the dot product of the identity map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ with itself. That is, for $x \in \mathbb{R}^n$,

$$\mathrm{id}(x) \cdot \mathrm{id}(x) = x \cdot x = ||x||^2 = f(x).$$

Since id is a linear map and hence continuous, the map f is continuous by Theorem 2.6.18(c).

Another fundamental property of continuity relates to composition of functions.

Corollary 2.6.20 Let $f : A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Let $g : B \to \mathbb{R}^k$. Let $a \in A$. If f is continuous at a and g is continuous at $f(a)$ then $g \circ f$ is continuous at a.

This corollary follows immediately from the following slightly more general theorem.

Theorem 2.6.21 Let $f: A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Let $g: B \to \mathbb{R}^k$. Let a be a limit point of A and let $b \in B$. If $\lim_{x\to a} f(x) = b$ and g is continuous at b then $\lim_{x\to a} g \circ f(x) = g(b)$.

Remark 2.6.22 The assumption of continuity is vital; see this MAT137 video for details.

Proof. Fix $\varepsilon > 0$. Since g is continuous at b, there exists $\eta > 0$ such that

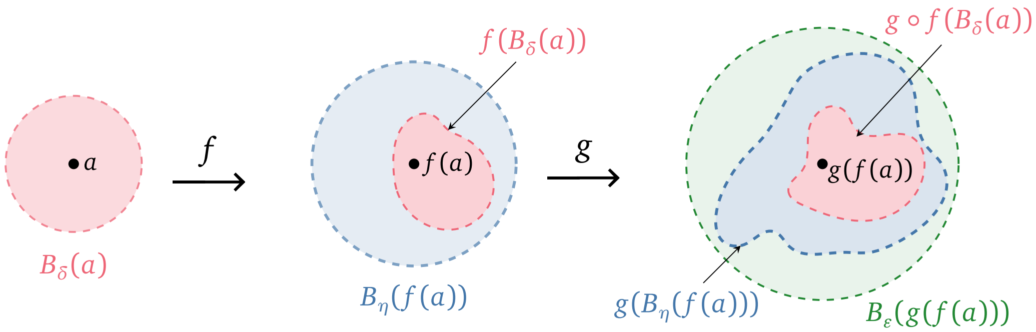
$$\forall y \in B, ||y-b|| < \eta \implies ||g(y)-g(b)|| < \varepsilon.$$

Since $f(x) \to b$ as $x \to a$, there exists $\delta > 0$ such that

$$\forall x \in A, 0 < ||x-a|| < \delta \implies ||f(x)-b|| < \eta.$$

Fix $x \in A$. Assume $0 < ||x-a|| < \delta$. By the above, $||f(x)-b|| < \eta$ so the previous implication with $y = f(x)$ implies that $||g(f(x))-g(b)|| < \varepsilon$, as required. ■

The above proof is illustrated with the following figure.



Corollary 2.6.20 with Theorem 2.6.18 can construct even more continuous functions.

Example 2.6.23 You have that $g(t) = \cos(t)$ is continuous on $\mathbb{R}$. From Example 2.6.7, you have seen that the function $f(x,y) = x+y$ is continuous on $\mathbb{R}^2$. Therefore, $g \circ f(x,y) = g(x+y) = \cos(x+y)$ is continuous on $\mathbb{R}^2$ by Corollary 2.6.20.

Like linear maps, multivariable polynomials[11] are another class of continuous functions.

Definition 2.6.24 A **monomial** in the n variables $x_1,\dots,x_n$ is a function of the form $x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ for some $\alpha_1,\dots,\alpha_n \in \mathbb{N}$. A **polynomial** in the n variables $x_1,\dots,x_n$ is a linear combination of monomials in n variables with real coefficients.

Example 2.6.25 The function $p(x,y,z) = xy + 3z^4$ is a polynomial in the 3 variables x,y,z. It is a linear combination of the monomials xy and z^4 which respectively correspond to $(\alpha_1,\alpha_2,\alpha_3) = (1,1,0)$ and $(0,0,4)$ in Definition 2.6.24. You can show p is continuous by writing it as a composition of functions. Namely, define the functions

- $f: \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = xy$, and $g: \mathbb{R} \to \mathbb{R}$ by $g(x) = x^4$.
- $\pi_3: \mathbb{R}^3 \to \mathbb{R}$ by $\pi_3(x,y,z) = z$, and $\pi_{1,2}: \mathbb{R}^3 \to \mathbb{R}^2$ by $\pi_{1,2}(x,y,z) = (x,y)$

You can check f is continuous by definition. From single-variable calculus, the function g is continuous. By Lemma 2.6.16, the projections π_3 and $\pi_{1,2}$ are continuous. Notice that

$$\forall (x,y,z) \in \mathbb{R}^3, \quad p(x,y,z) = xy + 3z^4 = (f \circ \pi_{1,2})(x,y,z) + 3(g \circ \pi_3)(x,y,z).$$

[11] These functions are your friends.

From Theorem 2.6.21 and our previous observations, $f \circ \pi_{1,2}$ and $g \circ \pi_3$ are continuous. Thus, by Theorem 2.6.18(a), the polynomial p is continuous.

These ideas can be extended to all multivariable polynomials.

Lemma 2.6.26 All polynomials in n variables are continuous on $\mathbb{R}^n$.

Proof. By Theorem 2.6.18(a), it suffices to show all monomials in n variables are continuous. This is left as an exercise. All you need is Lemma 2.6.16, continuity of $f(x,y) = xy$, and continuity of single-variable powers x^n. Use induction on the degree of the monomial. ■

These basic properties of continuous functions are repeatedly used, often without mention later. It is good to be familiar with them since continuity forms the bread and butter of many arguments. Yet there is still much more to the idea of continuity.

2.6.3 Topological properties

Topological properties of sets (like open, closed, bounded, and compact) help characterize the behaviour of functions on sets, such as the existence of extrema. There is a dual perspective: how do functions affect the properties of these sets? Without more restrictions, anything can happen. If you restrict to continuous functions, something magical happens.

*The **preimage** of a continuous map preserves the properties of open and closed.*

Even more remarkably, this phenomenon is an equivalent definition.

Theorem 2.6.27 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. The following are equivalent:
(a) f is continuous on $\mathbb{R}^n$
(b) The preimage $f^{-1}(U)$ is open for every open set $U \subseteq \mathbb{R}^m$
(c) The preimage $f^{-1}(V)$ is closed for every closed set $V \subseteq \mathbb{R}^m$.

Remark 2.6.28 You can prove a similar theorem for functions with domains other than $\mathbb{R}^n$ but the conclusions have quite a few topological subtleties (see *relative topology*). These nuances are better left for a course in topology.

Proof. The proofs that (a) implies (b), and (a) implies (c) are included below. The converse statements are left as exercises. Assume f is continuous on $\mathbb{R}^n$. By definition,

$$\begin{aligned} &\forall a \in \mathbb{R}^n, \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } f(B_\delta(a)) \subseteq B_\varepsilon(f(a)) \\ \iff &\forall a \in \mathbb{R}^n, \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a))). \end{aligned} \tag{2.6.2}$$

Here f^{-1} denotes the preimage. To prove (b), fix an open set $U \subseteq \mathbb{R}^m$. Let $a \in f^{-1}(U)$ so $f(a) \in U$. Since U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(f(a)) \subseteq U$. By (2.6.2), there exists $\delta > 0$ such that

$$B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a))) \subseteq f^{-1}(U)$$

as required. This establishes that $f^{-1}(U)$ is open and hence (b) holds. To prove (c), fix a closed set $V \subseteq \mathbb{R}^m$ so V^c is open. Note $f^{-1}(V^c) = \big(f^{-1}(V)\big)^c$ is open since (b) already follows from (a). You can therefore conclude that $f^{-1}(V)$ is closed, so (c) also holds. ■

Theorem 2.6.27 opens up many new ideas in topology, but that is another story[12]. You can instead wield this powerful tool to swiftly show many sets are open or closed.

[12] Insert another shameless plug for a course in topology.

Example 2.6.29 Let $A = \{(x,y,z) \in \mathbb{R}^3 : xy + 3z^4 \le 8\}$. The polynomial $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x,y,z) = xy + 3z^4$ is continuous by Lemma 2.6.26. Notice that

$$f^{-1}((-\infty, 8]) = \{(x,y,z) \in \mathbb{R}^3 : f(x,y,z) \le 8\} = A.$$

Since the interval $(-\infty, 8]$ is closed, it follows by Theorem 2.6.27 that the set A is closed.

You can also use it to prove that a function is not continuous.

Example 2.6.30 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by $f(x,y) = 237$ for $x^2 + y^2 < 1$ and $f(x,y) = 0$ otherwise. By definition, you have that

$$f^{-1}(\{237\}) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\} = B_1((0,0)),$$

which is an open ball. Thus, $f^{-1}(\{237\})$ is open, yet the set $\{237\}$ is closed. By Theorem 2.6.27, f cannot be continuous on $\mathbb{R}^n$.

You might wonder how the preimage affects boundedness and compactness. Unfortunately, neither are preserved under the preimage of a continuous map.

Example 2.6.31 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a constant map, say $f(x) = 223$ for every $x \in \mathbb{R}^n$. The set $\{223\}$ is bounded but the preimage $f^{-1}(\{223\}) = \mathbb{R}^n$ is unbounded. The same example shows that the preimage of a compact set is not necessarily compact.

Now, what about the *image* of a continuous map? Sadly, it does not respect the properties of open, closed, or bounded. There are many such counterexamples but it is enough to find one-dimensional examples.

Example 2.6.32 Let $f : \mathbb{R} \to \mathbb{R}$ be the continuous map given by $f(x) = \sin x$. You can prove that $f((0, 2\pi)) = [-1, 1]$ so the image of the open set $(0, 2\pi)$ is a closed set $[-1, 1]$.

Example 2.6.33 Let $g : \mathbb{R} \to \mathbb{R}$ be the continuous map given by $g(x) = 1/(x^2+1)$. Since $g(x)$ attains a maximum at $x = 0$, and $g(x) \to 0^+$ as $x \to \pm\infty$, you can check that $g(\mathbb{R}) = (0, 1]$ so the image of the clopen set $\mathbb{R}$ is neither open nor closed.

Example 2.6.34 Let $h : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ be the continuous map given by $h(x) = \tan x$. You can check that $h((-\frac{\pi}{2}, \frac{\pi}{2})) = \mathbb{R}$ so the image of the bounded set $(-\frac{\pi}{2}, \frac{\pi}{2})$ is the unbounded set $\mathbb{R}$.

Compactness, on the other hand, is a different story.

Theorem 2.6.35 If A is a compact subset of $\mathbb{R}^n$ and f is an $\mathbb{R}^m$-valued function that is continuous on A, then $f(A)$ is a compact subset of $\mathbb{R}^m$.

Informally speaking, this theorem says:

> *The* ***image*** *of a continuous map preserves compactness.*

The proof is a natural combination of their definitions.

Proof. If A is empty, then $f(A)$ is empty and hence compact. Assume A is non-empty. Let $\{y(k)\}_{k=1}^{\infty}$ be a sequence in $f(A)$. For each $k \in \mathbb{N}^+$, the set $f^{-1}(y(k))$ is non-empty since $y(k) \in f(A)$ and A is non-empty. Thus, you can choose some $x(k) \in f^{-1}(y(k))$ for each $k \in \mathbb{N}^+$. The sequence $\{x(k)\}_k$ lies in A so, since A is compact, it follows by definition that there exists some strictly increasing $m : \mathbb{N}^+ \to \mathbb{N}^+$ such that the subsequence $\{x(m(k))\}_k$ converges to some

$a \in A$. By continuity of f and Theorem 2.6.5, you can conclude that

$$\lim_{k\to\infty} y(m(k)) = \lim_{k\to\infty} f\big(x(m(k))\big) = f\Big(\lim_{k\to\infty} x(m(k))\Big) = f(a).$$

Thus, the subsequence $\{y(m(k))\}_k$ converges to $f(a) \in f(A)$. Hence, $f(A)$ is compact. ■

Theorem 2.6.35 will be decisive in the proof of the extreme value theorem. For now, you can use it to show seemingly complicated sets are compact.

Example 2.6.36 Consider the set

$$B = \{(xy, yz, xz) : 0 \le x, y, z \le 1\}.$$

The map $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by $f(x, y, z) = (xy, yz, xz)$ satisfies

$$f([0,1]^3) = \{f(x, y, z) : 0 \le x, y, z \le 1\} = B.$$

As f is a polynomial in each component, it is continuous by Theorem 2.6.14 and Lemma 2.6.26. Since $[0,1]^3$ is compact, $B = f([0,1]^3)$ is compact by Theorem 2.6.35.

Example 2.6.37 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x) = 1/x$ if $x \neq 0$ and $f(0) = 0$. For any $0 < \varepsilon < 1$, the image of the set $A = [\varepsilon, 1]$ is the set $f(A) = [1, 1/\varepsilon]$, which is compact. On the other hand, the image of the set $B = [0, 1]$ is the set $\{0\} \cup [1, \infty)$, which is not compact. Hence, f is not continuous by Theorem 2.6.35.

Continuous functions are fundamental maps in topology. Their formal definition is familiar to single-variable calculus, as well as their many basic properties. Now, with your foundations in open sets, closed sets, and compact sets, you can also appreciate how continuous functions are the glue holding topological ideas together.

Exercises for Section 2.6

Concepts and definitions

2.6.1 Let $A \subseteq \mathbb{R}^n$ be a set. Let $f : A \to \mathbb{R}^m$. Fix $a \in A$. Which statements are true or false?

(a) If f is continuous at a then $\lim_{x \to a} f(x) = f(a)$.

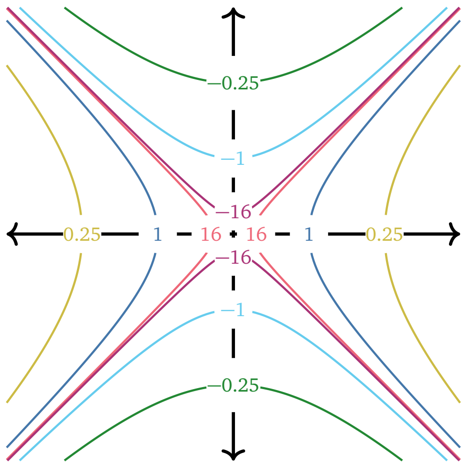
(b) If f is discontinuous at a then a is either a removable, infinite, or jump discontinuity.

(c) If there exists a $m \times n$ matrix M such that $f(x) = Mx$, then f is continuous everywhere.

(d) If $f(x) = (f_1(x), \cdots, f_m(x))$ where each of $f_1, \ldots, f_m$ is a polynomial, then f is continuous.

2.6.2 The contour map of a real-valued map $f : \mathbb{R}^2 \to \mathbb{R}$ is plotted on the right. Assume f is discontinuous somewhere in $\mathbb{R}^2$.

Where does f appear to be discontinuous?



2.6.3 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Fix $a \in A$. The definition of continuity has more subtleties than you might expect. Consider the following 6 statements.

A. $\lim_{x \to a} f(x) = f(a)$

B. $\forall \varepsilon > 0, \exists \delta > 0$ such that $\forall x \in \mathbb{R}^n$, $\|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon$

C. $\forall \varepsilon > 0, \exists \delta > 0$ such that $\forall x \in A$, $\|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon$

D. $\forall \varepsilon > 0, \exists \delta > 0$ such that $\forall x \in A$, $0 < \|x - a\| < \delta \implies \|f(x) - f(a)\| < \varepsilon$

E. $\forall \varepsilon > 0, \exists \delta > 0$ such that $f\big(B_\delta(a)\big) \subseteq B_\varepsilon(f(a))$.

F. $\forall \varepsilon > 0, \exists \delta > 0$ such that $f\big(A \cap B_\delta(a)\big) \subseteq B_\varepsilon(f(a))$.

(a) If a is a limit point of A, which of these statements are equivalent to "f is continuous at a"?

(b) If a is an isolated point of A, which of these statements are equivalent to "f is continuous at a"?

2.6.4 Assume $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous. Which of these statements are true or false? If true, briefly justify and cite any theorems used. If false, state a counterexample.

(a) If $A \subseteq \mathbb{R}^n$ is open then the image of A under f is open.

(b) If $A \subseteq \mathbb{R}^n$ is closed then the image of A under f is closed.

(c) For every $r > 0$ and $y \in \mathbb{R}^m$, the set $f^{-1}(B_r(y))$ is open.

(d) If $A \subseteq \mathbb{R}^n$ is open and bounded then $f(A)$ must be open.

(e) If $A \subseteq \mathbb{R}^n$ is closed and bounded then $f(A)$ must be closed.

(f) The preimages of compact sets under f must be compact.

Computations

2.6.5 Identify the largest possible set S on which each function is continuous. Briefly explain why.

(a) $g(x,y) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

(b) $h(x_1,\ldots,x_n) = x_1x_2^2 + x_2^2x_3^3 + x_3^3x_4^4 + \cdots + x_{n-1}^{n-1}x_n^n$

(c) $p(x,y,z) = \left(\dfrac{x}{1-z}, \dfrac{y}{1-z}\right).$

(d) $q(x,y) = \dfrac{x^2+y^2}{xy+3}$

(e) $f(x,y,z) = \dfrac{\sin(xy+z)}{1+z^2}$

2.6.6 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(0,0) = 0$ and $f(x,y) = \dfrac{x^2y}{x^4+y^2}$.

(a) Show that if (x,y) approaches $(0,0)$ along any line through $(0,0)$, then $f(x,y)$ approaches 0.

(b) Is f continuous at the origin? Prove your assertion.

2.6.7 Theorem 2.6.27 can be used in both directions. For example, you can verify sets are open or closed or you can verify that a function is not continuous on its domain.

(a) Prove $S = \{(x,y) \in \mathbb{R}^2 : y^2 - x^5 + xy > 1\}$ is open in $\mathbb{R}^2$.

(b) Prove $S = \{(w,x,y,z) \in \mathbb{R}^4 : x^2 - y^3 = 237z^2 \text{ and } xy = w\}$ is closed in $\mathbb{R}^4$.

(c) Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 1$ if $x \in \mathbb{Z}$ and $f(x) = 0$ otherwise. Use the theorem to show that f is not continuous on $\mathbb{R}$.

2.6.8 Theorem 2.6.35 can be used to quickly show sets are compact.

(a) Show that $\big\{(x_1^2+x_2^2, x_1^2-x_2^2) : x_1, x_2 \in [-1,1]\big\}$ is compact

(b) Show that the unit sphere in $\mathbb{R}^3$ is compact. *Hint:* Use a suitable parametrization.

(c) Show that the line segment from $p \in \mathbb{R}^n$ to $q \in \mathbb{R}^n$ is compact.

2.6.9 Prove that $q : \mathbb{R}^2 \to \mathbb{R}$ given by

$$q(x,y) = \frac{x}{y}$$

is continuous on $\{(x,y) \in \mathbb{R}^2 : y \neq 0\}$. Use properties of continuous functions and the continuity of certain single variable functions.

2.6.10 Prove that

$$f(x,y,z,w) = \frac{xe^{yz}}{w^2-1}$$

is continuous. Use the properties of continuous functions, the continuity of certain single variable functions, and the previous question.

Proofs

2.6.11 Prove that $p(x, y) = xy$ is continuous on $\mathbb{R}^2$ by using the open ball definition of continuity.

2.6.12 You can also prove the function $p(x, y) = xy$ is a continuous function on $\mathbb{R}^2$ using linear maps (Lemma 2.6.16) and basic properties of continuity (Theorem 2.6.18). These are both commonly used to construct continuous functions.

(a) Here is an attempted proof with the lemma and theorem.

> 1. *The functions $f(x) = x$ and $g(y) = y$ are continuous by the lemma.*
> 2. *By the theorem, their product $p(x, y) = f(x)g(y) = xy$ is continuous on $\mathbb{R}^2$.*

The reasoning is flawed. Explain what is wrong.

(b) Fix the attempted proof with the lemma and theorem.

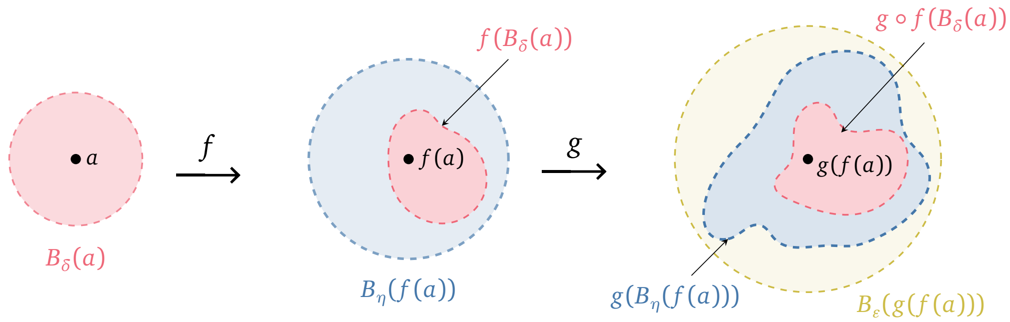
(c) Extend the ideas of that proof to show that $q(x, y) = x^j y^k$ is continuous for any fixed $j, k \in \mathbb{N}$.

2.6.13 Give three proofs that the sum of continuous functions is continuous: one with limit laws, one from the open ball definition, and one with sequences.

2.6.14 The textbook provides a formal proof of Theorem 2.6.21 from the definition of the limit and a geometric illustration involving balls. Understanding how to translate between the formal proof and the picture is an important skill to develop. The formal proof is below.

> 1. *Fix $\varepsilon > 0$.*
> 2. *Since g is continuous at b, there exists $\eta > 0$ such that*
>
> $$\forall y \in B, ||y - b|| < \eta \implies ||g(y) - g(b)|| < \varepsilon.$$
>
> 3. *Since $f(x) \to b$ as $x \to a$, there exists $\delta > 0$ such that*
>
> $$\forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \eta.$$
>
> 4. *Fix $x \in A$. Assume $0 < ||x - a|| < \delta$.*
> 5. *By the above, $||f(x) - b|| < \eta$.*
> 6. *The previous implication with $y = f(x)$ implies that $||g(f(x)) - g(b)|| < \varepsilon$, as required.*

Below is a picture proof **assuming** f is continuous at a in which case $b = f(a)$.



Indicate which line(s) in the proof correspond to which features of the figure.

2.6.15 Recall the alternate attempted approach to show that

$$\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$$

does not exist. Set $g(x,y) = \frac{x^2-y^2}{x^2+y^2}$. Here was a common flawed proof.

> 1. *Along the line* $(x,y)=(t,t)$, *we have that* $g(t,t)=0$ *for* $t\neq 0$.
> 2. *Thus,* $\lim_{t\to 0} g(t,t) = \lim_{t\to 0} 0 = 0$.
> 3. *Along the line* $(x,y)=(0,t)$, *we have that* $g(0,t) = \frac{0-t^2}{0+t^2} = -1$ *for* $t\neq 0$.
> 4. *Thus,* $\lim_{t\to 0} g(0,t) = \lim_{t\to 0}(-1) = -1$.
> 5. *Note both* (t,t) *and* $(0,t)$ *approach* $(0,0)$ *as* $t\to 0$.
> 6. *Since* $\lim_{t\to 0} g(t,t) \neq \lim_{t\to 0} g(0,t)$, *it follows that* $\lim_{(x,y)\to(0,0)} g(x,y)$ *does not exist.*

Line 6 is unjustified, but you can now fix it. Use Theorem 2.6.21 to justify line 6.

2.6.16 Prove that any level set of a continuous map $f:\mathbb{R}^n\to\mathbb{R}$ is closed.

2.6.17 Prove that if $A\subseteq\mathbb{R}^n$ is compact, then the graph of a continuous map $f:A\to\mathbb{R}^m$ is compact.

Applications and beyond

2.6.18 Prove that all polynomials $P(x,y)$ in two variables are continuous on $\mathbb{R}^2$. Use only that single variable polynomials are continuous, that linear combinations of continuous functions are continuous, and that products of continuous functions are continuous (i.e. Theorem 2.6.18).

2.6.19 Let S be a subset of $\mathbb{R}^n$. Define the **indicator function** of S to be the function $\chi_S:\mathbb{R}^n\to\mathbb{R}$ given by $\chi_S(x)=1$ if $x\in S$ and $\chi_S(x)=0$ if $x\notin S$. Show that if $S\neq\varnothing$ and $S\neq\mathbb{R}^n$ then χ_S is not continuous on $\mathbb{R}^n$.

2.6.20 Here is a shorter proof that (a) implies (b) in Theorem 2.6.27.

> 1. *Suppose* f *is continuous. Fix an open set* $U\subseteq\mathbb{R}^m$
> 2. *Let* $a\in f^{-1}(U)$.
> 3. *There exists* $\varepsilon>0$ *such that* $B_\varepsilon(f(a))\subseteq U$.
> 4. *Hence, there exists* $\delta>0$ *such that* $f(B_\delta(a))\subseteq B_\varepsilon(f(a))\subseteq U$.
> 5. *This implies that* $B_\delta(a)\subseteq f^{-1}(U)$.
> 6. *This shows that* $f^{-1}(U)$ *is open.*

This proof could use pictures and improve the writing.

(a) Line 2 has a minor error. Identify the error and fix it.

(b) Line 3 is not justified. State the missing justification.

(c) Line 4 is not justified. State the missing justification.

(d) Line 5 is not justified. State the exact implication used by the author.

(e) Draw a picture proof which captures lines 3 to 5. Label it with corresponding quantities.

2.6.21 Prove that (b) implies (a) in Theorem 2.6.27.

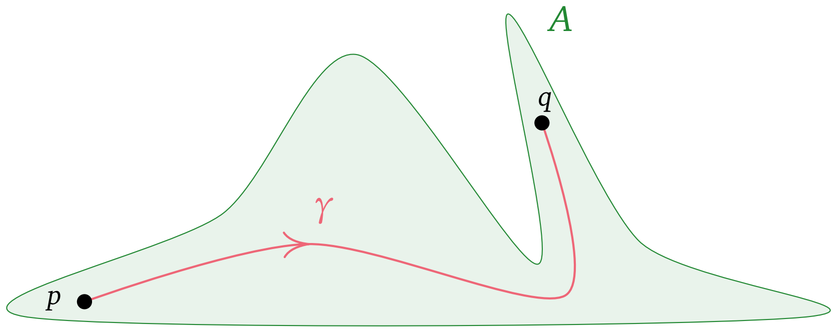
2.6.22

(a) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. Let $B \subseteq \mathbb{R}^m$. Prove that $f^{-1}(B^c) = \left(f^{-1}(B)\right)^c$.

(b) Prove that (c) implies (a) in Theorem 2.6.27.

2.7. Path-connected sets

You have investigated some key topological properties of sets: open, closed, and compact. These each capture their own intuitive concept, but none of them ensure that the set is "connected" in a natural sense. There are examples of open sets, closed sets, and compact sets which are disjoint unions of sets. How can you capture the concept of "connected" for a set $S \subseteq \mathbb{R}^n$? Informally, you want to be able to walk from one point $p \in S$ to any other point $q \in S$ without leaving S, much like the figure below.



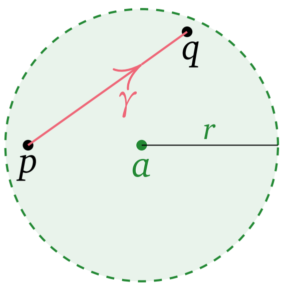
Continuity and parametric curves allow you to nicely formalize this idea.

Definition 2.7.1 A set $S \subseteq \mathbb{R}^n$ is **path-connected**[13] if for every pair of points $p, q \in S$ there exists a continuous function $\gamma : [a,b] \to \mathbb{R}^n$ such that $\gamma(a) = p$ and $\gamma(b) = q$ and $\operatorname{im}(\gamma) \subseteq S$.

Remark 2.7.2 Later, you may also refer to a set S as C^k **path-connected** if γ must also be continuously k-times differentiable on (a,b). This requires derivatives, so the concept is postponed.

A few simple examples can demonstrate that this definition is sensible.

Example 2.7.3 Fix $a \in \mathbb{R}^n$ and $r > 0$. The open ball $B_r(a)$ is path-connected.



The "picture proof" is illustrated above and the formal proof is below.

Proof. Fix $p, q \in B_r(a)$. Define $\gamma : [0,1] \to \mathbb{R}^n$ by

$$\gamma(t) = (1-t)p + tq,$$

so $\gamma(0) = p$ and $\gamma(1) = q$. Note γ is the straight line segment from p to q. Each component of γ is a linear single-variable polynomial t, so it is continuous by Theorem 2.6.14. It remains to check $\operatorname{im}(\gamma)$ is contained in the open ball.

[13]The definition of a *connected* set is different than path-connected; see a course in topology for details.

Fix $t \in [0,1]$. Observe that

$$\begin{aligned}
\|\gamma(t) - a\| &= \|(1-t)p + tq - a\| \\
&= \|(1-t)(p-a) + t(q-a)\| \\
&\leq \|(1-t)(p-a)\| + \|t(q-a)\| && \text{by the triangle inequality} \\
&= |1-t|\|p-a\| + |t|\|q-a\| && \text{as } t \in \mathbb{R} \\
&< |1-t|r + |t|r = r && \text{as } p,q \in B_r(a).
\end{aligned}$$

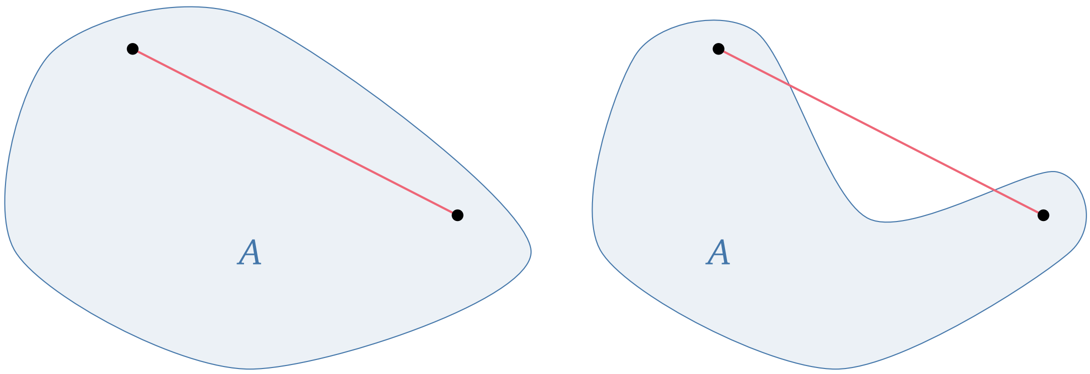
This implies that $\gamma(t) \in B_r(a)$ so $\mathrm{im}(\gamma)$ is in the open ball. ■

The open ball is a special example of a path-connected set since you can connect any two points by a line segment. This common property warrants its own terminology.

Definition 2.7.4 A set $S \subseteq \mathbb{R}^n$ is **convex** if the line segment between any two points $p, q \in S$ lies inside S.

Convex sets are an especially nice case of path-connected sets; you will not delve into them too deeply in multivariable calculus, but you may encounter them in future courses on optimization, statistics, or economics.

Example 2.7.5 Balls, cubes, planes, and regular polygons are all convex; however, anything that has an indent or a part jutting out (for example, a crescent moon shape) is not convex.



All convex sets are path-connected, but the converse is not true. For instance, the union of any 2 closed balls which touch at a single point a is not convex, but is path-connected. For any 2 points p and q in different balls, the straight line from p to a and then a to q is a path.

Proving sets are not path-connected requires careful use of the intermediate value theorem.

Example 2.7.6 The set $S = \{(x,y) \in \mathbb{R}^2 : x \neq 0\}$ is the 2-dimensional plane minus the vertical axis. It is not path-connected. You can illustrate the proof with a picture.

Proof. Take $p = (-1,0)$ and $q = (1,0)$. Let $\gamma : [a,b] \to \mathbb{R}^2$ be any continuous function such that $\gamma(a) = p = (-1,0)$ and $\gamma(b) = q = (1,0)$. Write $\gamma = (\gamma_1, \gamma_2)$ in terms of its component functions, each of which are continuous by Theorem 2.6.14. Note $\gamma_1(a) = -1$ and $\gamma_1(b) = 1$ and $\gamma_1 : [a,b] \to \mathbb{R}$ is continuous. By the intermediate value theorem, there exists $c \in (a,b)$ such that $\gamma_1(c) = 0$. This implies that $\gamma(c)$ lies on the vertical axis, so $\gamma(c) \notin S$. This proves $\mathrm{im}(\gamma) \nsubseteq S$ and hence S is not path-connected. ■

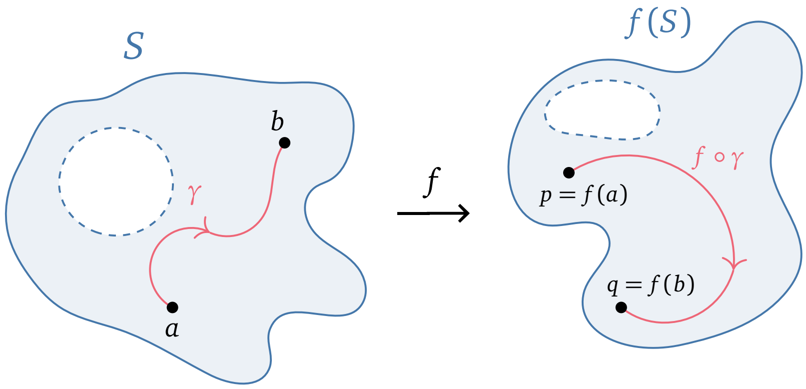
How do continuous functions interact with path-connected sets? The *preimage* does not preserve path-connectedness.

Example 2.7.7 Let $f : \mathbb{R} \to \mathbb{R}$ be the continuous map given by $f(x) = x^2$. The set $[1,4]$ is path-connected, but the preimage $f^{-1}([1,4]) = [-2,-1] \cup [1,2]$ is not path-connected.

However, the *image* of a continuous map preserves path-connectedness.

Theorem 2.7.8 Let $S \subseteq \mathbb{R}^n$ be a path-connected set. Let f be a $\mathbb{R}^m$-valued function defined on S. If f is continuous on S then $f(S)$ is path-connected.

Proof. Let $p, q \in f(S)$ be arbitrary. By definition, there exists $a, b \in S$ such that $f(a) = p$ and $f(b) = q$. Since S is path-connected, there exists a continuous map $\gamma : [0,1] \to \mathbb{R}^n$ such that $\gamma(0) = a$ and $\gamma(1) = b$ and the range of γ lies inside S. Since the range of γ lies inside the domain of f, we may define the map $f \circ \gamma : [0,1] \to \mathbb{R}^m$. An illustration is below.



It suffices to show that $f \circ \gamma$ is continuous, has range lying in $f(S)$, starts at p, and ends at q. Since f and γ are continuous, it follows by Theorem 2.6.21 that $f \circ \gamma$ is continuous. The range of γ lying in S implies that the range of $f \circ \gamma$ lies in $f(S)$. Finally, $f \circ \gamma(0) = f(\gamma(0)) = f(a) = p$ and $f \circ \gamma(1) = f(\gamma(1)) = f(b) = q$ so $f \circ \gamma$ indeed starts at p and ends at q. ∎

You can see this theorem in action with a difficult-to-describe set.

Example 2.7.9 Recall the set $B = \{(xy, yz, xz) : 0 \leq x, y, z \leq 1\}$ from Example 2.6.36 and the continuous function $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by $f(x,y,z) = (xy, yz, xz)$. Since the cube $[0,1]^3$ is path-connected, it follows by Theorem 2.7.8 that $B = f([0,1]^3)$ is also path-connected.

You have actually stumbled across a great result. By taking $S = [a,b]$ and $m = n = 1$ in Theorem 2.7.8, you recover the intermediate value theorem over $\mathbb{R}$!

Corollary 2.7.10 (Intermediate value theorem) Let f be a real-valued function defined on $[a,b]$. If f is continuous on $[a,b]$ then $f([a,b])$ is path-connected.

Remark 2.7.11 As $[a,b]$ is compact, $f([a,b])$ is a compact path-connected set by Theorems 2.6.35 and 2.7.8. You can verify that the only compact path-connected sets in $\mathbb{R}$ are closed intervals, so $f([a,b]) = [c,d]$ for some $c, d \in \mathbb{R}$ with $c \leq d$.

Generalizing a core theorem from calculus like the intermediate value theorem is a major achievement. It signifies the culmination of serious effort and a careful construction of mathematical theory. In the next section, you will conclude this chapter on topology with another grand generalization from single variable calculus: the extreme value theorem. This feat will be your first real step towards optimization in higher dimensions.

Exercises for Section 2.7

Concepts and definitions

2.7.1 Examples guide our intuition for most concepts. It is helpful to list basic examples of open, closed, and compact sets in $\mathbb{R}^n$. These should be sets where it is straightforward to use basic results and definitions to verify the desired topological property.

(a) List basic examples of open sets in $\mathbb{R}^n$. Identify which are bounded.

(b) List basic examples of closed sets in $\mathbb{R}^n$. Identify which are bounded.

(c) List basic examples of compact sets in $\mathbb{R}^n$.

(d) List basic examples of path-connected sets in $\mathbb{R}^n$.

2.7.2 Which of the following sets are path-connected?

(a) $A = \mathbb{R}^n \setminus \{0\}$

(b) $B = [0,1]^2 \cup [2,3]^2$

(c) $C = \{(x, x^2) : 0 < x < 137\}$

(d) $D = \{(x,y,z) \in \mathbb{R}^3 : x \neq 0\}$

Proofs

2.7.3 Draw a "picture proof" of the fact:

> *"The 2-dimensional plane minus the line $x = 0$ is not path-connected."*

Sketch a path which would be typical in a formal written proof. Label your drawing and provide 1 well-written sentence summarizing the proof idea. Afterwards, write a formal proof.

2.7.4 Prove that the graph of any continuous map $f : [-5,5]^2 \to \mathbb{R}$ is path-connected.

2.7.5 Prove that the washer $\{(x,y) \in \mathbb{R}^2 : 4 \leq x^2 + y^2 \leq 9\}$ is path-connected in two different ways.

2.7.6 Here is a multivariable intermediate value theorem for real-valued functions.

> **Theorem.** *Let $A \subseteq \mathbb{R}^n$ be path-connected and let $f : A \to \mathbb{R}$ be a continuous real-valued function. Let $K \in \mathbb{R}$. If there exists $a, b \in A$ such that $f(a) < K < f(b)$ then there exists $c \in A$ such that $f(c) = K$.*

Below is an incomplete proof. The key ideas are there but missing details and is poorly written.

1. *There exists a continuous $\gamma : [0,1] \to \mathbb{R}^m$ such that $\gamma(0) = a$ and $\gamma(1) = b$ and $\operatorname{im}(\gamma) \subseteq A$.*
2. *Thus, the function $f \circ \gamma : [0,1] \to \mathbb{R}$ is continuous.*
3. *Moreover, $(f \circ \gamma)(0) = f(\gamma(0)) = f(a)$ and $(f \circ \gamma)(1) = f(\gamma(1)) = f(b)$.*
4. *By the intermediate value theorem, there exists $t \in [0,1]$ such that $(f \circ \gamma)(t) = K$.*
5. *Taking $c = \gamma(t)$ completes the proof.*

You will identify the missing details and improve the writing.

(a) The author should fix several quantities before the first line. What should be fixed?

(b) The author uses that A is path-connected without mention. Where do they use it? Fix that line.

(c) Why does line 2 follow from line 1? Fix that line to make it clear.

(d) Why does line 5 complete the proof? Fix that line to make it clear.

2.8. Global extrema

Multivariable optimization is one of the most fundamental applications of calculus. It has far-reaching consequences in mathematics, economics, statistics, physics, data science, computer science, biology, and almost every corner of science. For example, you may want to maximize revenue for a company based on several cost constraints, or you may want to minimize error in a statistical prediction over a set of parameters. Describing real-life optimization requires many variables and the equations can get messy. Over several chapters of this textbook, you will develop basic techniques for solving these optimization problems. You must first address a simple yet deceptively challenging question:

For a given optimization problem, does a solution exist?

Of course, one answer is to actually find the solution itself and prove it is optimal. This may be possible for a given problem, but it is impossible to execute in general. Even worse, you may be trying to search for a solution when one does not even exist! In this section, you will use your advances in topology to characterize a broad class of optimization problems where you are *guaranteed* the existence of a solution.

2.8.1 Definitions of global extrema

First, you must define extrema of real-valued functions and introduce some terminology.

Definition 2.8.1 Let $A \subseteq \mathbb{R}^n$ and let f be a real-valued function defined on A.

- A point $p \in A$ is a **(global) maximum point of** f **on** A if $f(p) \geq f(x)$ for all $x \in A$. If so, the value $f(p)$ is the **(global) maximum value of** f **on** A.
- If a maximum point of f on A exists, then f **attains a (global) maximum on** A.

The definitions of minimum point, minimum value, and attaining a minimum are similar.

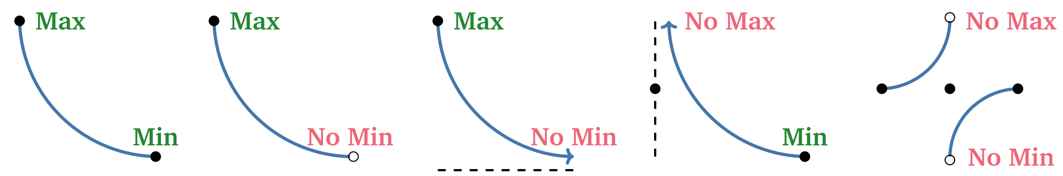
Remark 2.8.2 The word "global" is often omitted. The phrase "on A" is often omitted when A is the domain of the function. A maximum value is also called a **maximum**. The word **extremum** is used to refer to either a minimum or a maximum.

With these firm definitions, you can explore a few examples beginning with a reminder of the extreme value theorem over $\mathbb{R}$.

Example 2.8.3 Recall the extreme value theorem from single variable calculus.

Extreme value theorem. *If the function* $f : [a,b] \to \mathbb{R}$ *is continuous then* f *attains a maximum and a minimum on* $[a,b]$.

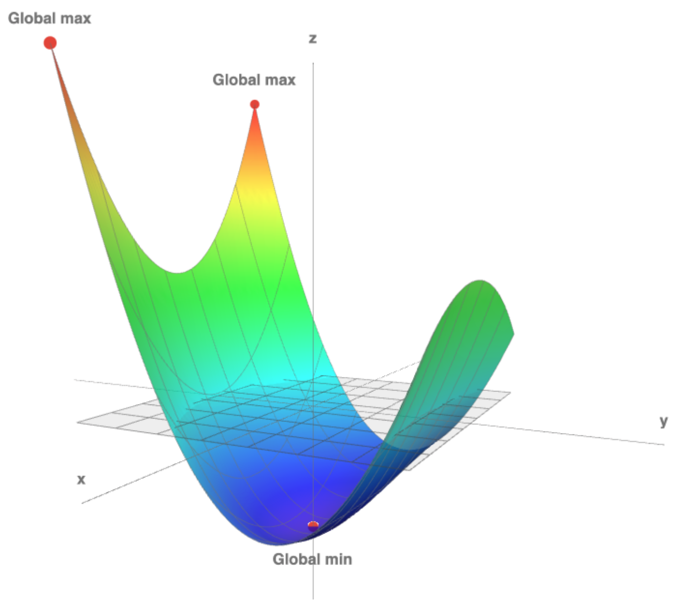
Remember it is crucial that the function is continuous *and* that the interval $[a,b]$ is closed and bounded. A few pictures below illustrate why.



$f:[a,b]\to\mathbb{R}$	$f:[a,b)\to\mathbb{R}$	$f:[a,\infty)\to\mathbb{R}$	$f:[a,b]\to\mathbb{R}$	$f:[a,b]\to\mathbb{R}$
Continuous	Continuous	Continuous	NOT Continuous	NOT Continuous

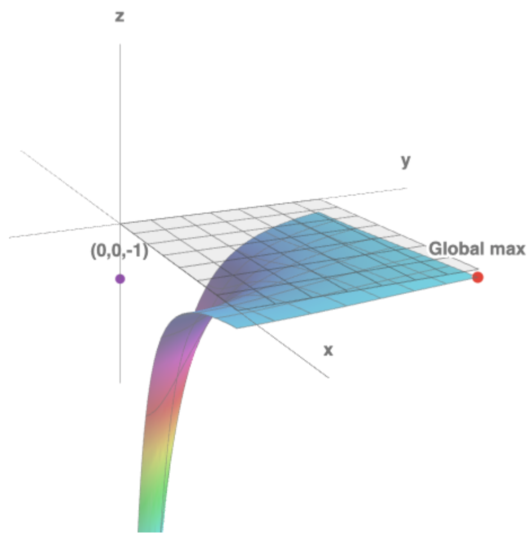
For more details, view this MAT137 video.

Example 2.8.4 Consider $f(x,y) = x^2+4y^2-2x^2y-2$ on the rectangle $A = [-1,1]\times[-1,1]$. The graph of f is below which you can also view on Math3D.



From looking at the graph, you can see that f attains its minimum value at $(0,0)$ and its maximum values at $(1,-1)$ and $(-1,-1)$. Note that there may be multiple points of A at which f attains its minimum or maximum. Moreover, notice that the minimum is attained on the interior of A, whereas the maxima are attained on the boundary of A. If you consider f on the subset $B = (-1,1)^2$ then f still attains its minimum on B since $(0,0)\in B$ but it does *not* attain a maximum on B. You can get arbitrarily close to $(-1,1)$ or $(-1,-1)$ on B but you cannot reach either point. While f is continuous on A and continuous on B, the issue is that B is not closed whereas A is closed (and hence contains its boundary).

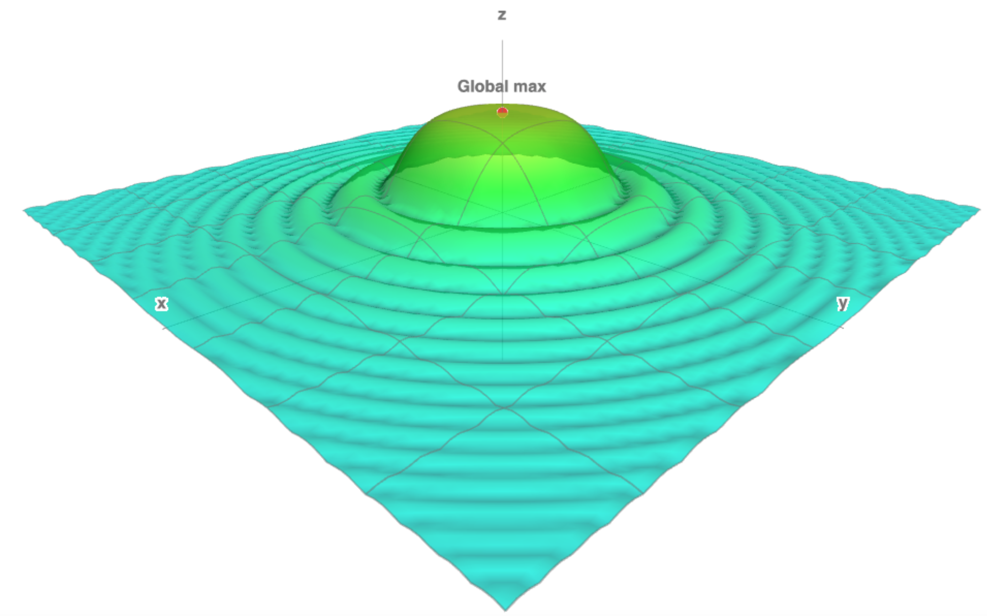
Example 2.8.5 Let $A = [0,2]\times[0,2]$. Define $f : A \to \mathbb{R}$ by $f(0,0) = -1$ and $f(x,y) = \dfrac{-1}{x^2+y^2}$ otherwise. While A is closed, notice f is not continuous on A. In fact, as $(x,y)\to(0,0)$, $f(x,y)\to-\infty$ so f does not attain a minimum on A. It does, however, attain its maximum at $(2,2)$. You can view these features in its graph, which you can also view on Math3D.



Example 2.8.6 Let $A = \mathbb{R}^2$. Define $f : A \to \mathbb{R}$ by

$$f(x,y) = \frac{3|\sin(x^2+y^2)|+3}{x^2+y^2+1}.$$

The domain $A = \mathbb{R}^2$ is unbounded and you can check that f is continuous. It attains its maximum on A at $(0,0)$. You can check that $f(x,y) > 0$ always yet $f(x,y) \to 0$ as $||(x,y)|| \to \infty$. You can prove that this implies that f has no minimum on A. These features can be seen in its graph which you can view on Math3D.



These investigations have hopefully convinced you of the key ingredients to guaranteeing the existence of extrema. The function should be continuous and the domain should be closed and bounded, i.e. compact.

2.8.2 Extreme value theorem

Collecting all of these observations, you can formulate the generalization of the extreme value theorem to $\mathbb{R}^n$. The statement is itself a massive accomplishment.

Theorem 2.8.7 (Extreme value theorem) If $A \subseteq \mathbb{R}^n$ is a non-empty compact set and the map $f : A \to \mathbb{R}$ is continuous, then f attains maximum and minimum values at points of A.

Its proof requires a defining property of the real numbers, which you may take for granted.[14]

Theorem 2.8.8 (Least upper bound principle) Let $S \subseteq \mathbb{R}$ be a non-empty subset of real numbers. If S has an upper bound, then the supremum of S exists.

Remark 2.8.9 Recall a real number $b \in \mathbb{R}$ is an *upper bound for* S if $\forall x \in S, x \le b$. Moreover, a real number $s \in \mathbb{R}$ is the *supremum* of a set $S \subseteq \mathbb{R}$, denoted $s = \sup(S)$, if s is an upper bound for S, and $s \le b$ for every upper bound $b \in \mathbb{R}$ of S. If the supremum of S does not exist, then write $\sup(S) = \infty$. See this MAT137 video for details.

[14]A rigorous proof of the least upper bound principle can be found in course on an introduction to analysis.

You are finally ready to prove the extreme value theorem in higher dimensions.

Proof of Theorem 2.8.7. It suffices to show a maximum is attained since the proof for the minimum is similar. Since f is continuous on A and A is compact, it follows that $f(A)$ is a compact by Theorem 2.6.35. By the Bolzano–Weierstrass theorem, this implies $f(A)$ is a bounded subset of $\mathbb{R}$ so, by the least upper bound principle (Theorem 2.8.8), the quantity

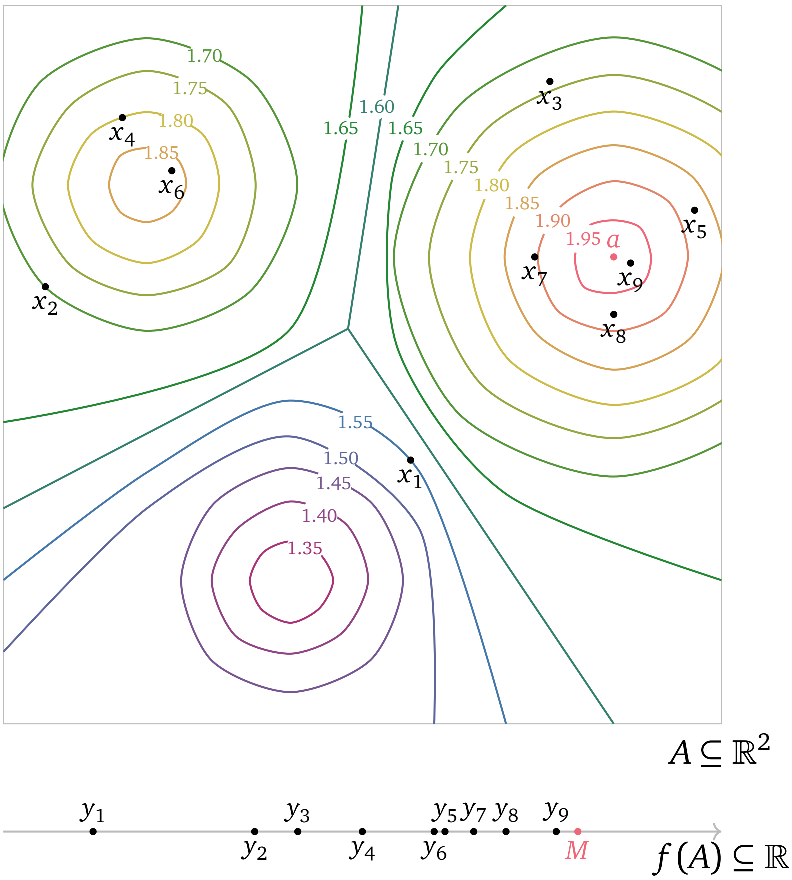
$$M = \sup f(A) = \sup\{y : y \in f(A)\}$$

is a finite real number. It suffices to show there exists $a \in A$ such that $f(a) = M$ or, equivalently, we must show $M \in f(A)$. By a definition of the supremum, for each $k \in \mathbb{N}^+$, there exists $y(k) \in f(A)$ such that

$$M - \frac{1}{k} < y(k) \leq M.$$

If $y(k) = M$ for some $k \in \mathbb{N}^+$, then $M \in f(A)$ as required. Otherwise, $y(k) \neq M$ for all $k \in \mathbb{N}^+$ and $y(k) \to M$ as $k \to \infty$, so M is a limit point of $f(A)$. Since $f(A)$ is compact and hence closed, it again follows that $M \in f(A)$ as required. ■

An illustration of this proof is below, assuming A is a rectangle in $\mathbb{R}^2$. For each point $y_k = y(k) \in f(A)$, there exists $x_k \in A$ such that $f(x_k) = y_k$. The sequence $\{x_k\}_k \subseteq A$ does not appear in the proof above, but they are included in the figure below to help demonstrate the intuition behind the argument. Take some time to compare the picture with the written proof.



The proof is a beautiful display of the tools you have developed on compactness, continuity, and their properties. It is a testament to the importance of topology in optimization. Now, you can investigate how the extreme value theorem may be used to guarantee the existence of a solution to an optimization problem.

Example 2.8.10 At a given moment, is there a hottest point on the Earth? Common sense suggests there should be and the extreme value theorem does, too! The Earth can be roughly viewed as the unit sphere S^2 in $\mathbb{R}^3$. Let $T(x,y,z)$ be the temperature in Celsius at a given point $(x,y,z) \in S^2$, so $T : S^2 \to \mathbb{R}$. Presumably, temperature should vary continuously across the Earth's surface so T is continuous. The sphere S^2 is compact so by the extreme value theorem T attains its maximum and minimum on S^2. A maximum point $p \in S^2$ of T corresponds to the hottest point on Earth.

Example 2.8.11 Recall Example 2.8.4. Can you find extrema for the function $f(x,y) = x^2 + 4y^2 - 2x^2y - 2$ on the closed rectangle $A = [-1,1]^2$? Since f is continuous and A is compact, the extreme value theorem implies that the answer is yes, there exists both a minimum and a maximum of f on A.

Can you find extrema for f on the open rectangle $B = (-1,1)^2$? The extreme value theorem guarantees you nothing here since B is not compact. Extrema may or may not exist; in this case, a minimum exists and a maximum does not but anything can happen in general.

Example 2.8.12 Recall Example 2.8.5 where $A = [0,2]^2$ and $f : A \to \mathbb{R}$ is defined by $f(0,0) = -1$ and $f(x,y) = -\frac{1}{x^2+y^2}$ otherwise. Can you find extrema for f on A? The extreme value theorem again gives no direct information since f is not continuous on A. Extrema may or may not exist; in this case, a maximum exists but a minimum does not.

Example 2.8.13 Recall Example 2.8.6 where $A = \mathbb{R}^2$ and $f : A \to \mathbb{R}$ is continuous. Can you find extrema for f on A? Once again, the extreme value theorem does not directly imply anything because A is not bounded and hence not compact.

Despite these setbacks, the extreme value theorem can still be used to prove the existence of extrema for non-compact sets. Here is one such result.

Lemma 2.8.14 Let $A \subseteq \mathbb{R}^n$ be closed and unbounded. Let f be a continuous real-valued function on A. If $f(x) \to -\infty$ as $||x|| \to \infty$ in A, then f attains a maximum on A.

Proof. This is left as a challenging exercise. Far away, the function is very small. Nearby, use the extreme value theorem for compact sets. ■

There are many more results of this type, but it is not important to list them. Rather you can formulate them yourself and prove them with a clever application of the extreme value theorem for compact sets.

This concludes the chapter on topology with many fantastic achievements. You can describe sets in higher dimensions along with their interior, boundary, and closure. You have characterized topological properties of sets (open, closed, compact, path-connected) and related them to sequences, limits, and continuity. You have extended the intermediate value theorem and the extreme value theorem from single variable calculus. In other words, you are ready to build differential calculus in higher dimensions.

Exercises for Section 2.8

Concepts and definitions

2.8.1 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$ be continuous. For each set A, does f always have a maximum and minimum on A? Briefly explain citing any theorems, if necessary.

(a) $A = \overline{B_1(0)}$ in $\mathbb{R}^n$

(b) $A = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 \leq 1\}$

(c) $A = B_1(0) \setminus \{0\}$ in $\mathbb{R}^n$

2.8.2 While the extreme value theorem (EVT) is powerful, it has limitations as you will see in these examples. Determine whether EVT applies to each of the following optimization scenarios. If it does not, restrict the domain A so that it does apply.

(a) $f(x, y, z) = x + yz + \dfrac{1}{1 - x^2 - y^2 - z^2}$ on $A = B_1(0)$.

(b) $g(x, y) = xy^2 + xy + x$ on $A = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 \leq 1\}$.

(c) $h(x, y) = e^{-x^2 - y^2}$ on $A = \mathbb{R}^2$.

Proofs

2.8.3 You can dissect proofs and statements of theorems by breaking them up into smaller seemingly independent pieces. One piece of the extreme value theorem is Theorem 2.6.35. Another piece is implicitly proved but not explicitly stated.

> **Lemma.** *If C is a non-empty compact subset of $\mathbb{R}$, then C admits a maximum and minimum value. That is, there exists $a, b \in C$ such that $a \leq x \leq b$ for all $x \in C$.*

(a) Using the lemma and Theorem 2.6.35, give a shorter proof of the extreme value theorem.

(b) Prove the lemma using the least upper bound principle.

2.8.4 The proof of the extreme value theorem only establishes the following.

> **Theorem I.** *If A is a non-empty compact subset of $\mathbb{R}^n$ and $f : A \to \mathbb{R}$ is a continuous function, then f attains its maximum value at a point of A.*

This proof is acceptable because the argument can be modified slightly to obtain the same result about minimums, which can be written as

> **Theorem II.** *If A is a non-empty compact subset of $\mathbb{R}^n$ and $f : A \to \mathbb{R}$ is a continuous function, then f attains its minimum value at a point of A.*

Instead of repeating this textbook's argument, prove Theorem II using Theorem I.
Hint: How can you turn a minimum into a maximum?

2.8.5 Prove that if $f : \mathbb{R}^n \to [0, \infty)$ is continuous and

$$\lim_{||x|| \to \infty} f(x) = 0,$$

then f attains a maximum. *Hint:* Use the extreme value theorem for compact sets.

Applications and beyond

2.8.6 A condensed proof of the extreme value theorem for maximum values only is written below.

1. *Suppose $f : A \to \mathbb{R}$ is continuous. Assume A is a non-empty compact subset of $\mathbb{R}^n$.*
2. *Then $f(A) \subseteq \mathbb{R}$ is compact.*
3. *So $f(A)$ is closed and bounded.*
4. *Then $f(A)$ has a least upper bound M.*
5. *It suffices to prove that $M \in f(A)$.*
6. *For all $k \in \mathbb{N}^+$ there exists $y_k \in f(A)$ such that $M - \frac{1}{k} < y_k \leq M$.*
7. *If $y_k = M$ for some $k \in \mathbb{N}^+$, then $M \in f(A)$.*
8. *Otherwise, $y_k \neq M$ for every $k \in \mathbb{N}^+$ and $y_k \to M$ as $k \to \infty$.*
9. *So M is a limit point.*
10. *Since $f(A)$ is closed, it follows that $M \in f(A)$ which completes the proof.*

You will fill in some details for this proof.

(a) Lines 2, 3, and 4 follow from three very powerful theorems. State them.

(b) Justify the claim in Line 5 with more details. You will need Definition 2.8.1.

(c) Justify the claims in Line 8 and 9 with more details. You will need a theorem from Section 2.2, a limit law over $\mathbb{R}$, and some previous lines.

2.8.7 View the illustration after the proof of the extreme value theorem (Theorem 2.8.7). The picture assumes A is a rectangle in $\mathbb{R}^2$. Moreover, for each point $y_k = y(k) \in f(A)$, there exists $x_k \in A$ such that $f(x_k) = y_k$. Here you will compare this picture with the general proof.

(a) Does the picture suggest $f(A)$ is an interval in $\mathbb{R}$? If so, approximate the interval.

(b) In general, does $f(A)$ necessarily need to be an interval? Explain why or why not.

(c) Does the picture suggest that the sequence $\{x_k\}_k$ in A converges?

(d) In general, does the sequence $\{x_k\}_k$ in A necessarily converge? Explain why or why not.

2.8.8 The extreme value theorem guarantees the existence of a solution to the optimization of a continuous function on a compact domain. However, many other situations naturally occur and you will need to recognize what conditions you require to guarantee the existence of a solution. The correct conditions are not always given to you, so you must discover it for yourself. Here you will conjecture theorems like the extreme value theorem.

For each conjecture below, fill in the blanks with reasonably general conditions that should guarantee the conclusion. Note you may need more than one condition and the answers are not unique. Remember to draw examples.

(a) **Conjecture A.** *Let $A \subseteq \mathbb{R}^n$ be closed and unbounded. Let $f : A \to \mathbb{R}$ be continuous.*

If ______________________________ *then f attains a maximum on A.*

(b) **Conjecture B.** *Let $A \subseteq \mathbb{R}^n$ be closed and unbounded. Let $f : A \to \mathbb{R}$ be continuous.*

If ______________________________ *then f attains a minimum on A.*

(c) **Conjecture C.** *Let $f : [0, \infty)^n \to [0, \infty)$ be continuous.*

If ______________________________ *then f attains a maximum on the interior of $[0, \infty)^n$.*

(d) **Conjecture D.** *Let $f : \overline{B_1(0)} \setminus \{0\} \to \mathbb{R}$ be continuous.*

If ______________________________ *then f attains a maximum on $\overline{B_1(0)} \setminus \{0\}$.*

II

Differential calculus

3. Derivatives

With the foundations of topology, you are ready to generalize the idea of derivatives from one variable to many variables. This chapter will be dedicated to unravelling the definition of the derivative for maps $\mathbb{R}^n \to \mathbb{R}^m$, and it will take a long and winding road. There will be many important versions of a derivative before you arrive at the ultimate definition. Initially, you will try to reduce the definition to differentiation of single variable maps $\mathbb{R} \to \mathbb{R}$, where you previously interpreted the derivative as an instantaneous rate of change or as a slope. This strategy will be quite successful for the special case of maps $\mathbb{R} \to \mathbb{R}^m$ and somewhat successful for real-valued maps $\mathbb{R}^n \to \mathbb{R}$.

However, this attempt to generalize "rates" and "slopes" to higher dimensions will not be enough, because those notions are inherently one dimensional. For example, a "rate of change" or a "slope" is really a scalar, not a vector. The crucial insight from *calculus* will be to interpret single variable derivatives via *linear* approximations.

Indeed, the ultimate definition of derivatives for any map $\mathbb{R}^n \to \mathbb{R}^m$ will require a miraculous blend of linear algebra and calculus. The underlying philosophy has a succinct description:

> *Nonlinear maps are well approximated by linear maps.*

This will be discovered by viewing derivatives from four different perspectives: physical, geometric, analytic, and algebraic. By translating between these overarching viewpoints, you will truly and deeply understand derivatives.

3.1. Derivatives of one variable

Equipped with a rigorous definition of limits, you can begin to generalize the derivative to higher dimensions with the very special case of derivatives for maps of one variable, that is, derivatives of parametric curves. These are maps of the form

$$\mathbb{R} \longrightarrow \mathbb{R}^m,$$

which were investigated in Section 1.1. These are a very special case, because the theory quickly reduces to single variable calculus. On the other hand, you can use this setting to build familiarity with the four viewpoints of the derivative: algebraic, analytic, physical, and geometric. These give a glimmer of the new ideas needed to fully generalize the derivative.

3.1.1 Definition

Motivated by single variable calculus and your newfound definition of limits, you can create a definition for the derivative of parametric curves.

Definition 3.1.1 Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. The **derivative of f at a** is defined to be

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

provided the limit exists. If the limit exists, then f is **differentiable at a**.

This definition visually matches the derivative in single variable calculus, but there are several subtle observations. First, notice which quantities are scalars and which are vectors. The limit variable $h \in \mathbb{R}$ is a scalar which approaches the real number 0. The limit quantity $\frac{f(a+h)-f(a)}{h}$ is the scalar $\frac{1}{h} \in \mathbb{R}$ multiplied by the vector $f(a+h)-f(a) \in \mathbb{R}^m$. This implies that $f'(a)$ is also an element of $\mathbb{R}^m$. Second, you can prove the limit is equivalent to

$$f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \tag{3.1.1}$$

by using a theorem on composition of limits. Third, the derivative is only defined at interior points of A. This choice is not necessary[1], but it is made for simplicity.

Example 3.1.2 Let $f : \mathbb{R} \to \mathbb{R}^2$ be defined by $f(x) = (\cos x, \sin x)$. You can verify by definition that f is differentiable at π. Notice

$$\begin{aligned}
f'(\pi) &= \lim_{h\to 0} \frac{(\cos(\pi+h), \sin(\pi+h)) - (\cos\pi, \sin\pi)}{h} \\
&= \lim_{h\to 0} \left(\frac{\cos(\pi+h)-\cos\pi}{h}, \frac{\sin(\pi+h)-\sin\pi}{h} \right) \\
&= \left(\lim_{h\to 0} \frac{\cos(\pi+h)-\cos\pi}{h}, \lim_{h\to 0} \frac{\sin(\pi+h)-\sin\pi}{h} \right) \qquad \text{by Theorem 2.5.11.}
\end{aligned}$$

These are the derivatives of $\cos x$ and $\sin x$ at $x = \pi$, so the above is $(-\sin\pi, \cos\pi) = (0, -1)$. Hence, f is differentiable at π and $f'(\pi) = (0, -1)$.

[1] You could define a derivative if a is a limit point of A provided $a \in A$. This would include boundary points where the derivative would be similar to defining a "one-sided derivative" in single variable calculus. This slightly more general definition is not necessary for us and it can be a bit of a nuisance in the proofs of lemmas. The choice to only consider interior points is for simplicity.

3.1.2 Basic properties

The prior example illustrates that derivatives of parametric curves reduce to differentiation from single variable calculus. All you need to do is differentiate component by component.

Lemma 3.1.3 Let $A \subseteq \mathbb{R}$ and let $f = (f_1, \ldots, f_m) : A \to \mathbb{R}^m$. Let a be an interior point of A. The function f is differentiable at a if and only if for every $i \in \{1, \ldots, m\}$, the component function f_i is differentiable at a. If so,

$$f'(a) = (f_1'(a), \ldots, f_m'(a)) = \begin{bmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{bmatrix}.$$

Remark 3.1.4 Recall $(v_1, \ldots, v_m)$ denotes a *column* vector. Elements of $\mathbb{R}^m$ are always considered column vectors by convention. A row vector can be denoted as the transpose of a column vector $(v_1, \ldots, v_m)^T$ or as the $1 \times m$ matrix $\begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix}$.

Proof. This is left as an exercise. It follows from Theorem 2.5.11 and Definition 3.1.1. ■

Example 3.1.5 Let $f : \mathbb{R} \to \mathbb{R}^2$ be defined by $f(x) = (\cos x, \sin x)$. Since $\cos x$ and $\sin x$ are differentiable everywhere, the function f is differentiable everywhere by Lemma 3.1.3 and, moreover, $f'(x) = (\frac{d}{dx} \cos x, \frac{d}{dx} \sin x) = (-\sin x, \cos x)$ at each $x \in \mathbb{R}$.

You can prove basic properties of the derivative for parametric curves using the definition and limit laws. Such proofs would mimic the proofs in single-variable calculus. Alternatively, you can exploit Lemma 3.1.3 and differentiation rules of single variable calculus.

Theorem 3.1.6 Let $A \subseteq \mathbb{R}$ and let f and g be $\mathbb{R}^m$-valued functions defined on A. Let a be an interior point of A. Let $\lambda \in \mathbb{R}$ and let φ be a real-valued function defined on A.

(a) (*Linearity*) If f and g are differentiable at a, then $f + \lambda g$ is differentiable at a and

$$(f + \lambda g)'(a) = f'(a) + \lambda g'(a).$$

(b) (*Scalar product*) If f and φ are differentiable at a, then φf is differentiable at a and

$$(\varphi f)'(a) = \varphi'(a) f(a) + \varphi(a) f'(a).$$

(c) (*Dot product*) If f and g are differentiable at a, then $f \cdot g$ is differentiable at a and

$$(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a).$$

Proof. Items (b) and (c) are left as exercises. For (a), write $f = (f_1, \ldots, f_m)$ and $g = (g_1, \ldots, g_m)$. By Lemma 3.1.3, each of the components are differentiable at a and

$$(f + \lambda g)'(a) = \begin{bmatrix} (f_1 + \lambda g_1)'(a) \\ \vdots \\ (f_m + \lambda g_m)'(a) \end{bmatrix}.$$

By linearity of differentiation for maps $\mathbb{R} \to \mathbb{R}$, the above is equal to

$$\begin{bmatrix} f_1'(a) + \lambda g_1'(a) \\ \vdots \\ f_m'(a) + \lambda g_m'(a) \end{bmatrix} = \begin{bmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{bmatrix} + \lambda \begin{bmatrix} g_1'(a) \\ \vdots \\ g_m'(a) \end{bmatrix} = f'(a) + \lambda g'(a),$$

as required. ∎

Lemma 3.1.7 (Chain rule) Let $A, B \subseteq \mathbb{R}$ be sets. Let $\varphi : A \to B$ be a real-valued function, and let f be a $\mathbb{R}^m$-valued function defined on B. Let a be an interior point of A such that $\varphi(a)$ is an interior point of B. If φ is differentiable at a and f is differentiable at $\varphi(a)$, then

$$(f \circ \varphi)'(a) = f'(\varphi(a))\varphi'(a).$$

Proof. This is left as an exercise. Use Lemma 3.1.3 and the chain rule for maps $\mathbb{R} \to \mathbb{R}$. ∎

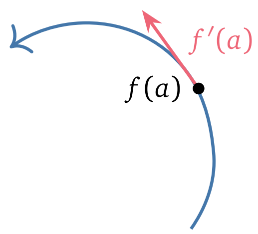
These quick and fast proofs are quite pleasing. Instead of building everything from scratch, you can rely on a simple lemma and your differentiation theorems from single variable calculus.

3.1.3 Four viewpoints

Now, you can study the derivative of a parametric curve from many viewpoints. Fluency and translation between these perspectives is vital for a full and complete understanding of the derivative. Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a parametric curve differentiable at $a \in A$.

Physical viewpoint

If $f : A \to \mathbb{R}^m$ is the position of a particle, then the derivative $f'(a)$ is the **(instantaneous) velocity** of the particle at time a and position $f(a)$. This is motivated by Example 1.1.6.

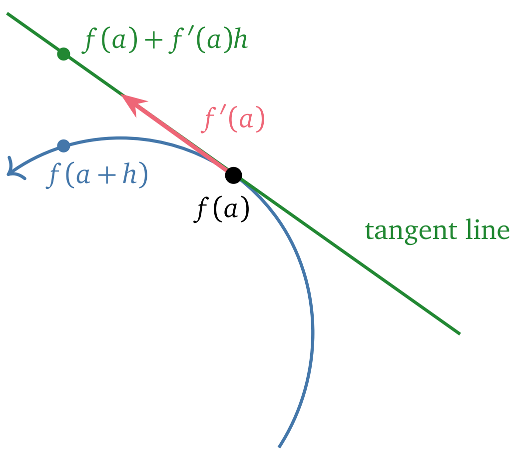


Geometric viewpoint

You can use one viewpoint to build another. The velocity vector points in the direction of motion, so it should presumably be "tangent" to the path of motion. This suggests that the derivative of a parametric curve can be used to define the tangent line at a point on the curve. The **tangent line** at the point $f(a)$ on the parametric curve f is presumably[2] given by the set

$$\{f(a) + hf'(a) : h \in \mathbb{R}\}.$$

Notice $f(a)$ is a point on the line and $f'(a)$ is a direction vector defining the line.



Analytic viewpoint

Fix a point a. How can you approximate $f(x)$ for x near a? Heuristically speaking, from (3.1.1), you may expect that

$$f'(a) \approx \frac{f(x)-f(a)}{x-a}$$

Multiplying both sides by the scalar $x-a$ and solving for the vector $f(x)$ suggests that

$$f(x) \approx f(a)+f'(a)(x-a)$$

when $x \approx a$. The righthand side is a linear polynomial in the scalar x, so these informal manipulations suggest a definition. The **linear approximation of f at a** is the function $\ell : \mathbb{R} \to \mathbb{R}^m$ defined by

$$\ell(x) = f(a)+f'(a)(x-a),$$

so you expect that $f(x) \approx \ell(x)$ for x near a, or equivalently $f(a+h) \approx \ell(a+h)$ for h near 0. It is worthwhile to compare this relation with the diagram from the geometric viewpoint.

Example 3.1.8 Let $f(x) = (\cos x, \sin x)$. Since $f(\pi) = (-1,0)$ and $f'(\pi) = (0,-1)$, the linear approximation of f at π is the function $\ell : \mathbb{R} \to \mathbb{R}^2$ given by

$$\ell(x) = (-1,0)+(0,-1)(x-\pi) = (-1,\pi-x) = \begin{bmatrix} -1 \\ \pi - x \end{bmatrix}.$$

For instance, $f(3)$ is approximately $\ell(3) = (-1, \pi-3)$.

Algebraic viewpoint

By incorporating linear algebra, you can view the derivative from a completely novel perspective.

Theorem 3.1.9 Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. The function f is differentiable at a if and only if there exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{f(a+h)-f(a)-L(h)}{h} = 0,$$

in which case $L(h) = f'(a)h$.

Proof. Assume f is differentiable at a, so $f'(a)$ exists. Then define $L : \mathbb{R} \to \mathbb{R}^m$ by $L(h) = f'(a)h$ for $h \in \mathbb{R}$. It follows that

$$\frac{f(a+h)-f(a)-L(h)}{h} = \frac{f(a+h)-f(a)-f'(a)h}{h} = \frac{f(a+h)-f(a)}{h} - f'(a).$$

Taking $h \to 0$, it follows by Theorem 2.5.14 that

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

Conversely, assume such a linear map $L : \mathbb{R} \to \mathbb{R}^m$ exists so $L(h) = Bh$ for some $m \times 1$ matrix B. Then

$$\begin{aligned} f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} &= \lim_{h\to 0} \frac{f(a+h)-f(a)-L(h)+L(h)}{h} \\ &= \lim_{h\to 0} \left(\frac{f(a+h)-f(a)-L(h)}{h} + B \right). \end{aligned}$$

By Theorem 2.5.14 and the defining property of L, this limit exists and is equal to B. ■

[2]The correct definition of a tangent line to a curve is quite subtle, and is postponed until the next chapter. For now, it is better to ignore those subtleties and assume everything works out exactly as you would hope.

This equivalent definition may appear innocuous at first glance, but it is revolutionary! By reframing the derivative in terms of a linear transformation, it creates an avenue for blending *linear algebra* with *calculus*. This linear map warrants its own definition.

Definition 3.1.10 Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is differentiable at a, then the linear map $df_a : \mathbb{R} \to \mathbb{R}^m$ defined by $df_a(h) = f'(a)h$ is the **differential of f at a**.

The differential has a nice relationship with the analytic viewpoint of linear approximations. Namely, for h near 0, you expect

$$f(a+h) \approx f(a) + f'(a)h = f(a) + df_a(h)$$

so $df_a(h)$ approximates the difference between $f(a)$ and $f(a+h)$.

Differentials can elegantly encapsulate properties of the derivative. A showcase example is the chain rule from single variable calculus. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be differentiable everywhere. Fix $a \in \mathbb{R}$. The chain rule in single variable calculus states that

$$(g \circ f)'(a) = g'(f(a))f'(a)$$

in classical notation. In terms of differentials, this implies that

$$\begin{aligned} d(g \circ f)_a(h) = (g \circ f)'(a)h &= g'(f(a))\big(f'(a)h\big) \\ &= g'(f(a))df_a(h) \\ &= dg_{f(a)}(df_a(h)) \\ &= \big(dg_{f(a)} \circ df_a\big)(h). \end{aligned}$$

The chain rule is therefore equivalent to the following elegant statement about linear maps:

$$d(g \circ f)_a = dg_{f(a)} \circ df_a.$$

In other words, the differential of the composition is the composition of differentials! All four viewpoints are critical and this fourth viewpoint will ultimately lead to a breakthrough when generalizing the derivative to higher dimensions.

Exercises for Section 3.1

Concepts and definitions

3.1.1 Let $f : \mathbb{R} \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}$. For $h \in \mathbb{R}$, you have the following relationship:

$$df_a(h) = f'(a)h = \begin{bmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{bmatrix} h.$$

(a) For each quantity, identify its name (if possible) and the type of mathematical object it is.

i) $f'(a)$ ii) $f_1'(a)$ iii) $df_a(h)$ iv) df_a v) f'

(b) One equality is by definition while the other follows from a theorem. Identify which is which.

3.1.2 Let $f : \mathbb{R} \to \mathbb{R}^m$ and let $a \in \mathbb{R}$. Which statements are equivalent to "f is differentiable at a"?

(a) The limit $\lim\limits_{h\to 0} \dfrac{f(a+h)-f(a)}{h}$ exists.

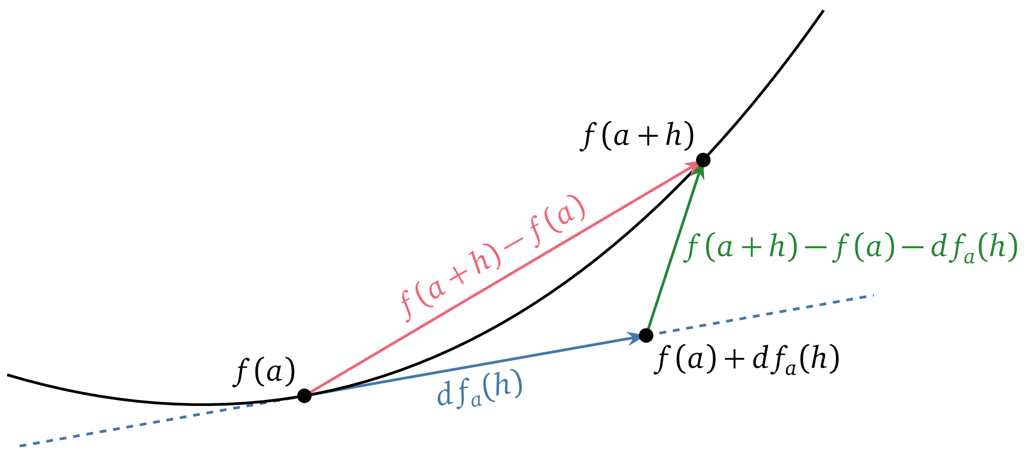
(b) There exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{f(a+h)-f(a)-L(h)}{h} = 0$.

(c) There exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{f(a+h)-L(h)}{h} = 0$.

(d) There exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \left|\dfrac{f(a+h)-f(a)-L(h)}{h}\right| = 0$.

(e) There exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \left\|\dfrac{f(a+h)-f(a)-L(h)}{h}\right\| = 0$.

3.1.3 The four viewpoints of the derivative are broad mathematical themes that appear in many contexts. A picture is a useful way to remember the relationships between these viewpoints for the derivative. You will explore each perspective. Let $f : \mathbb{R} \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}$.



For parts (b)–(d), use the picture above.

(a) The **algebraic** viewpoint is:

The derivative defines a linear map which gives a first order approximation to f at a.

Write this down precisely using a limit. Why is the phrase “first order approximation” used?

(b) The **analytic** viewpoint is:

> *The derivative linearly approximates the average rate of change of f near a with small error.*

How does the picture relate? Label each quantity using the context of linear approximation.
Hint: Label the "exact value", "approximate value", "exact difference", "approximate difference", and "error".

(c) The **geometric** viewpoint is:

> *The derivative is a tangent vector to the curve f at a which defines the tangent line.*

How does the same picture relate? If possible, label each quantity using a geometric context.
Hint: Label the "base point", "curve", "tangent vector", and "tangent line".

(d) The **physical** viewpoint is:

> *The derivative is the (instantaneous) velocity of the particle f at time a.*

How does the same picture relate? If possible, label each quantity using a physical context.
Hint: Label the "current position", "future position", "path", "exact displacement", "approximate displacement", and "velocity". Introduce 1 new arrow.

Computations

3.1.4 Let $g : \mathbb{R} \to \mathbb{R}^3$ be given by $g(t) = (t\cos t, t\sin t, t)$. View its Math3D plot.

(a) Compute the derivative of g at 2π.

(b) Compute the velocity of g at 2π.

(c) Compute the unit tangent vector of g at 2π.

(d) Compute the differential of g at 2π.

(e) Compute the linear approximation of g at 2π.

(f) Compute the tangent line to the curve g at $g(2\pi)$.

Proofs

3.1.5 Prove Lemma 3.1.3, i.e. you differentiate a parametric curve component-by-component.

3.1.6 Prove Lemma 3.1.7, i.e. a simplified version of the chain rule.

3.1.7 Let $f : \mathbb{R} \to \mathbb{R}^m$ and $g : \mathbb{R} \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}$. Prove that $f + g$ is differentiable at a and

$$d(f+g)_a = df_a + dg_a.$$

3.1.8

(a) Let $I \subseteq \mathbb{R}$ be an open interval. Let $g : I \to \mathbb{R}^n$ be differentiable and assume that $||g(t)|| = 1$ for all $t \in I$. Prove that $g(t) \cdot g'(t) = 0$ for all $t \in I$.

(b) Let $I \subseteq \mathbb{R}$ be an open interval. Let $\gamma : I \to \mathbb{R}^n$ be twice differentiable; in other words, $\gamma' : I \to \mathbb{R}$ is also differentiable. Assume its unit tangent vector $T(t)$ and unit normal vector $N(t)$ exist for all $t \in I$. Prove that $T(t) \cdot N(t) = 0$ for all $t \in I$.

3.2. Partial derivatives

Generalizing the derivative to maps of the form

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

will take several attempts, and each attempt will build on earlier ones. Any reasonable definition of a derivative should capture the following basic principle:

A derivative should describe how a function is changing near a point.

This goal presents a serious challenge for $n \geq 2$ because there are countless many ways to "change near a point" inside $\mathbb{R}^n$. This section restricts your attention to a limited number of such ways. The underlying idea is to reduce everything to a one-dimensional setting, so you can apply all of your single variable calculus tools and interpretations as rates of change.

3.2.1 Definition

How can you reduce high dimensional settings to a lower dimensional setting? From Chapter 1, one method is by *slicing*, which is also known as *fixing variables*. This basic idea is explored in the following motivational example from several viewpoints.

Example 3.2.1 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 - y^2$. You want to describe how f changes near the point $(a, b) = (1, 2)$. To reduce everything to single variable calculus, you can investigate how f changes with respect to a single variable *by fixing all other variables*.

Consider the *algebraic* viewpoint. Define the single variable functions $p : \mathbb{R} \to \mathbb{R}$ and $q : \mathbb{R} \to \mathbb{R}$ by

$$p(x) = f(x, 2) = x^2 - 4, \qquad q(y) = f(1, y) = 1 - y^2. \tag{3.2.1}$$

In other words, p is the function f with the y-variable fixed to $y = 2$ and q is the function f with the x-variable fixed to $x = 1$. Now that they are single variable functions, you can easily calculate their derivatives with single variable differentiation rules. Notice, however, only the derivatives $p'(1) = 2$ and $q'(2) = -4$ relate to the function f near the point $(1, 2)$. Observe that the limit definition of these single variable derivatives

$$\begin{aligned} p'(1) &= \lim_{h\to 0} \frac{p(1+h) - p(1)}{h} = \lim_{h\to 0} \frac{f(1+h, 2) - f(1, 2)}{h}, \\ q'(2) &= \lim_{h\to 0} \frac{q(2+h) - q(2)}{h} = \lim_{h\to 0} \frac{f(1, 2+h) - f(1, 2)}{h} \end{aligned} \tag{3.2.2}$$

suggests that these quantify how f changes near $(1, 2)$ but in different ways. The derivative $p'(1)$ quantifies how f changes in the x-direction at $(1, 2)$ whereas $q'(2)$ quantifies how f changes in the y-direction at $(1, 2)$. How does this quantification occur?

Consider the *analytic* viewpoint. From the limit definition in (3.2.2), you can heuristically see that for $h \approx 0$,

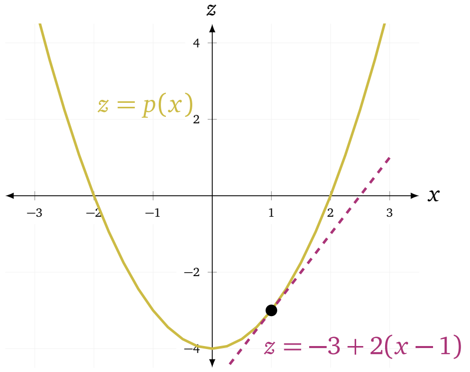
$$\begin{aligned} f(1+h, 2) &\approx f(1, 2) + p'(1)h = -3 + 2h, \\ f(1, 2+h) &\approx f(1, 2) + q'(2)h = -3 - 4h, \end{aligned}$$

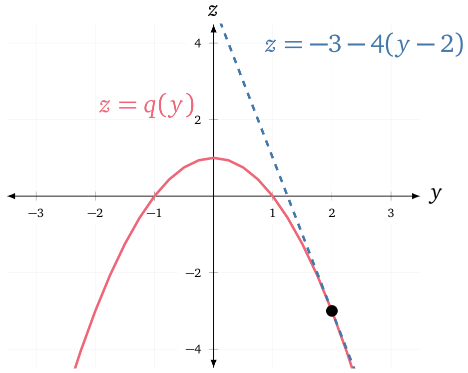
or equivalently for $x \approx 1$ and $y \approx 2$,

$$\begin{aligned} f(x, 2) &\approx f(1, 2) + p'(1)(x-1) = -3 + 2(x-1), \\ f(1, y) &\approx f(1, 2) + q'(2)(y-2) = -3 - 4(y-2). \end{aligned} \tag{3.2.3}$$

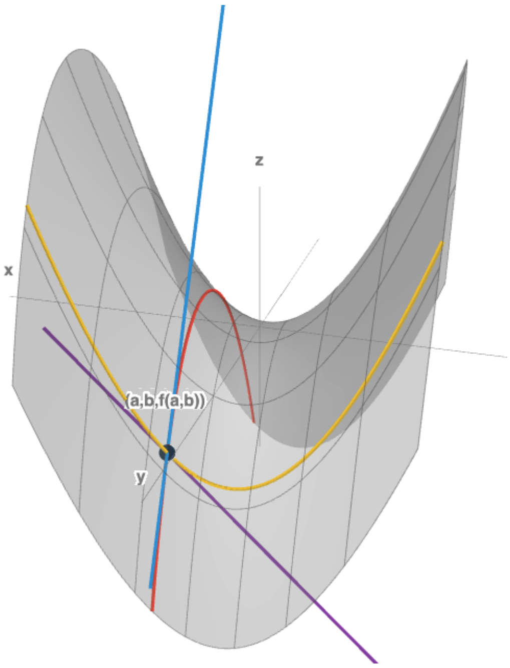
In other words, the linear approximation of $p(x)$ at $x = 1$ can be used to linearly approximate $f(x, 2)$ and similarly, the linear approximation of $q(y)$ at $y = 2$ can be used to linearly approximate $f(1,y)$. Notice *single variable* functions are estimating the change in a *multivariable* function in a *single* direction.

Consider the *geometric* viewpoint. You can refer to the function p as the y-slice of f at $y=2$ and refer to the function q as the x-slice of f at $x=1$. Their graphs in $\mathbb{R}^2$ are plotted below along with the tangent lines from (3.2.3).





The slopes of the tangent lines are $p'(1)$ and $q'(2)$ respectively. These graphs in $\mathbb{R}^2$ are precisely the *slices* of the graph of f as illustrated below.



View this Math3D demo to rotate this object and understand it in more detail. Toggle the objects to get a better view.

Consider the *physical* viewpoint. Think of yourself as a tiny person sitting on the graph of f at the point $(a,b,f(a,b)) = (1,2,-3)$. If you walk on this surface in the x-direction from $(a,b) = (1,2)$ towards $(a+0.5,b) = (1.5,2)$, then the "slope" in that direction is $p'(1) = 2$ so you expect to change your elevation by approximately $p'(1) \cdot 0.5 = 1$. If you walk on this surface in the y-direction form $(a,b) = (1,2)$ to $(a,b+0.5) = (1,2.5)$, then the "slope" in that direction is $q'(2) = -4$ so you expect to change your elevation by approximately $q'(2) \cdot 0.5 = -2$. Translating between all four perspectives in this example will solidify your understanding.

This motivational example suggests a formal definition for a new kind of derivative.

Definition 3.2.2 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Let a be an interior point of A. Fix $1 \le j \le n$. Let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$. The **jth partial derivative of f at a** is given by

$$\partial_j f(a) := \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$

provided the limit exists. The **jth partial derivative of f** is the function $\partial_j f : U \to \mathbb{R}^m$, where U is the set of points $a \in A$ such that $\partial_j f(a)$ exists.

Remark 3.2.3 Equivalent notation for $\partial_j f$ includes

$$\frac{\partial f}{\partial x_j} \qquad D_{e_j} f \qquad f_{x_j} \qquad D_j f \qquad \partial_{x_j} f$$

The first three versions will also be used in this text, but the notations D_j and ∂_{x_j} will not.

The standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$ is introduced, so you can *fix all variables except one*. For instance, fix $j \in \{1, \dots, n\}$. If $a = (a_1, \dots, a_{j-1}, a_j, a_{j+1}, \dots, a_n) \in \mathbb{R}^n$ then

$$\forall h \in \mathbb{R}, \qquad a + he_j = (a_1, \dots, a_{j-1}, a_j + h, a_{j+1}, \dots, a_n),$$

so all components except the jth component remain fixed.

Example 3.2.4 Recall $f(x, y) = x^2 - y^2$ at the point $(1, 2)$ from Example 3.2.1. By definition,

$$\begin{aligned}
\partial_1 f(1,2) &= \lim_{h \to 0} \frac{f((1,2) + h(1,0)) - f(1,2)}{h} = \lim_{h \to 0} \frac{f(1+h,2) - f(1,2)}{h} \\
\partial_2 f(1,2) &= \lim_{h \to 0} \frac{f((1,2) + h(0,1)) - f(1,2)}{h} = \lim_{h \to 0} \frac{f(1,2+h) - f(1,2)}{h}.
\end{aligned}$$

Notice these are exactly the same as in (3.2.2), so you can verify that $\partial_1 f(1,2) = p'(1) = 2$ and $\partial_2 f(1,2) = q'(2) = -4$. These can be calculated by single variable differentiation rules or by using the limit definition. For instance,

$$\begin{aligned}
\partial_1 f(1,2) = \lim_{h \to 0} \frac{f(1+h,2) - f(1,2)}{h} &= \lim_{h \to 0} \frac{\left((1+h)^2 - 2^2\right) - (-3)}{h} \\
&= \lim_{h \to 0} \frac{h^2 + 2h}{h} = 2.
\end{aligned}$$

Note there is no obvious relationship between partial derivatives at the same point.

Example 3.2.5 Let $f(x, y) = |x|y$. By definition,

$$f_x(0,3) = \lim_{h \to 0} \frac{f((0,3) + h(1,0)) - f(0,3)}{h} = \lim_{h \to 0} \frac{f(h,3)}{h} = \lim_{h \to 0} \frac{3|h|}{h}.$$

That limit does not exist so $f_x(0,3)$ does not exist. On the other hand,

$$f_y(0,3) = \lim_{h \to 0} \frac{f((0,3) + h(0,1)) - f(0,3)}{h} = \lim_{h \to 0} \frac{f(0,3+h)}{h} = \lim_{h \to 0} \frac{0}{h} = 0,$$

so $f_y(0,3)$ exists. View this Math3D demo to visually investigate the issue.

Now, what exactly does a partial derivative represent? For a real-valued function, it can be classically interpreted in many applied contexts as a *rate of change*.

Example 3.2.6 [3]Let $H(x,t)$ be the temperature (in Celsius) in a room as a function of distance x (in metres) from a heater and time t (in minutes) after the heater has been turned on. What does the quantity $H_x(1,30)$ represent? Remember that

$$H_x(1,30) = \lim_{\Delta x \to 0} \frac{H(1+\Delta x, 30) - H(1,30)}{\Delta x}. \tag{3.2.4}$$

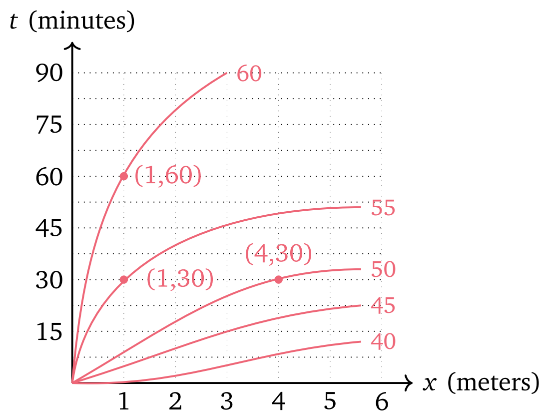
Since it is a limit of quantities measured in Celsius per metre, the units for $H_x(1,30)$ are also Celsius per metre. If $\Delta x = 1$, then this means that $H_x(1,30) \approx H(2,30) - H(1,20)$. In other words, if you are 1 metre away from the heater after 30 minutes, then the temperature changes by *approximately* $H_x(1,30)$ per metre farther away that someone else stands *at the exact same time*. This phrasing suggests that $H_x(1,30)$ is presumably *negative* since standing farther away from the heater should decrease the temperature.

Similarly, what does the quantity $H_t(1,30)$ represent? Again, remember that

$$H_t(1,30) = \lim_{\Delta t \to 0} \frac{H(1, 30+\Delta t) - H(1,30)}{\Delta t}, \tag{3.2.5}$$

so the units for $H_t(1,30)$ are Celsius per minute. If $\Delta t = 1$, then this roughly means that $H_t(1,30) \approx H(1,31) - H(1,30)$. In other words, if you are 1 metre away from the heater after 30 minutes, then the temperature changes by *approximately* $H_t(1,30)$ per minute longer you stand *at the exact same distance*. This phrasing suggests that $H_t(1,30)$ is presumably *positive* since it will get hotter as the heater remains on for a longer duration.

Now, assume you are given the following contour plot of $H(x,t)$.



This contour map suggests that $H(1,30) = 55$ Celsius. How could you estimate $H_x(1,30)$ from this contour plot? The idea is to use a nearby point on an appropriately chosen contour. Looking at the contour plot, notice that the point $(4,30)$ is one of the nearest points to $(1,30)$ on another contour with the time fixed at 30. The definition of $H_x(1,30)$ in (3.2.5) suggests that by taking $\Delta x = 3$, you can estimate

$$H_x(1,30) \approx \frac{H(1+3,30) - H(1,30)}{3} = \frac{50-55}{3} \approx -1.67 \text{ Celsius per metre.}$$

Similarly, you can estimate $H_t(1,30)$ using the point $(1,60)$ in the contour plot. This approach gives $H_t(1,30) \approx 0.17$ Celsius per minute, which is left as an exercise.

3.2.2 Computations

Partial derivatives reduce to derivatives of one variable. This means you can use all your existing tools from Section 3.1 and single variable calculus to quickly compute derivatives.

Lemma 3.2.7 Let $A \subseteq \mathbb{R}^n$ and let $f = (f_1, \dots, f_m) : A \to \mathbb{R}^m$. Let a be an interior point of A. The jth partial of f at a exists if and only if for every $i \in \{1, \dots, m\}$, the jth partial of the component function $f_i : A \to \mathbb{R}$ at a exists. If so,

$$\partial_j f(a) = (\partial_j f_1(a), \dots, \partial_j f_m(a)).$$

Theorem 3.2.8 Let $A \subseteq \mathbb{R}^n$. Let $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$. Let a be an interior point of A. Fix $1 \le j \le n$. Let $\lambda \in \mathbb{R}$ and let φ be a real-valued function defined on A.

(a) (*Linearity*) If $\partial_j f(a)$ and $\partial_j g(a)$ both exist, then $\partial_j(f + \lambda g)(a)$ exists and

$$\partial_j(f + \lambda g)(a) = \partial_j f(a) + \lambda \partial_j g(a).$$

(b) (*Scalar product*) If $\partial_j f(a)$ and $\partial_j \varphi(a)$ both exist, then $\partial_j(\varphi f)(a)$ exists and

$$\partial_j(\varphi f)(a) = f(a)\partial_j \varphi(a) + \varphi(a)\partial_j f(a).$$

(c) (*Dot product*) If $\partial_j f(a)$ and $\partial_j g(a)$ both exist, then $\partial_j(f \cdot g)(a)$ exists and

$$\partial_j(f \cdot g)(a) = \partial_j f(a) \cdot g(a) + f(a) \cdot \partial_j g(a).$$

Proof. This is left as an exercise. Use the definition with several limit laws. ∎

All of your other single variable calculus differentiation rules also apply. You should try to formally state them yourself as an exercise. For now, you must be able to compute with them. The calculations of partial derivatives are delightfully familiar. You essentially "fix all other variables as constants" and follow your differentiation rules with respect to one variable.

Example 3.2.9 Define the function

$$f(x, y, z) = \left(x^2 - y^2 + \frac{x}{x+z}, e^x - xy\right)$$

You can calculate $\partial_1 f$ by computing component-by-component via Lemma 3.2.7. Write $f = (f_1, f_2)$ so you must compute $\partial_1 f_1$ and $\partial_1 f_2$ separately. First,

$$\partial_1 f_1 = \frac{\partial f_1}{\partial x} = \frac{\partial}{\partial x}\left(x^2 - y^2 + \frac{x}{x+z}\right).$$

[3] This example is inspired by Hughes-Hallett et al. [12].

Treating y and z as constants, you calculate this by differentiating with respect to x. That is,

$$\begin{aligned}
\partial_1 f_1 &= \frac{\partial}{\partial x}(x^2) - \frac{\partial}{\partial x}(y^2) + \frac{\partial}{\partial x}\left(\frac{x}{x+z}\right) && \text{by linearity} \\
&= 2x - \frac{\partial}{\partial x}(y^2) + \frac{\partial}{\partial x}\left(\frac{x}{x+z}\right) && \text{by the power rule} \\
&= 2x + \frac{\partial}{\partial x}\left(\frac{x}{x+z}\right) && \text{since } y \text{ is constant} \\
&= 2x + \frac{z}{(x+z)^2} && \text{by the quotient rule}
\end{aligned}$$

Therefore, $\partial_1 f_1 = 2x + \dfrac{z}{(x+z)^2}$. You can calculate $\partial_1 f_2$ in a similar way.

$$\begin{aligned}
\partial_1 f_2 = \frac{\partial f_2}{\partial x} &= \frac{\partial}{\partial x}(e^x - xy) \\
&= \frac{\partial}{\partial x}(e^x) - \frac{\partial}{\partial x}(xy) && \text{by linearity} \\
&= e^x - \frac{\partial}{\partial x}(xy) && \text{by the exponent rule} \\
&= e^x - y && \text{by the power rule.}
\end{aligned}$$

Therefore, you can conclude that

$$\partial_1 f = \begin{bmatrix} \partial_1 f_1 \\ \partial_1 f_2 \end{bmatrix} = \begin{bmatrix} 2x + \dfrac{z}{(x+z)^2} \\ e^x - y \end{bmatrix}.$$

Note the domain of $\partial_1 f$ is the set $\mathbb{R}^3 \setminus \{(x,y,z) \in \mathbb{R}^3 : x+z=0\}$.

Example 3.2.10 Recall $f(x,y) = x^2 - y^2$ so $\partial_1 f(x,y) = 2x$ and $\partial_2 f(x,y) = -2y$. Hence,

$$\partial_1 f(1,2) = 2, \qquad \partial_2 f(1,2) = -4.$$

If you want to linearly approximate $f(1.1,2)$ using the information about f at $(1,2)$ then you can use the 1st partial since

$$f(1.1,2) \approx f(1,2) + \partial_1 f(1,2)(1.1-1) = -3 + 2(0.1) = -2.8.$$

Similarly, you can linearly approximate $f(1,1.9)$ using the 2nd partial since

$$f(1,1.9) \approx f(1,2) + \partial_2 f(1,2)(1.9-2) = -3 - 4(-0.1) = -2.6.$$

How can you linearly approximate $f(1.1,1.9)$ using partials of f at $(1,2)$? You cannot move from $(1,2)$ to the point $(1.1,1.9)$ by moving only along the x-direction or only along the y-direction. This suggests using one partial derivative to approximate $f(1.1,1.9)$ will not be accurate. However, you can move *in a linear combination of these two directions* to reach $(1.1,1.9)$ from $(1,2)$, so perhaps you can somehow use both partial derivatives. That idea is for the next section.

Partial derivatives are an excellent attempt at generalizing the derivative to higher dimensions. You fix all variables except one, and differentiate with respect to that remaining variable using calculus. It is a wonderfully natural idea and pervasive across all scientific disciplines but, as Example 3.2.10 demonstrates, it seems a bit mathematically restrictive. You could attempt to find the rate of change in *other directions*, not just by fixing variables. This ambition will carry you forward to the next attempt at generalizing the derivative.

Exercises for Section 3.2

Concepts and definitions

3.2.1 Let $A \subseteq \mathbb{R}^n$. Assume all the partial derivatives of $F = (F_1, \ldots, F_m) : A \to \mathbb{R}^m$ exist at every interior point of A. Fix $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. Let a be an interior point of A.

(a) For each quantity, identify the type of mathematical object that it is.

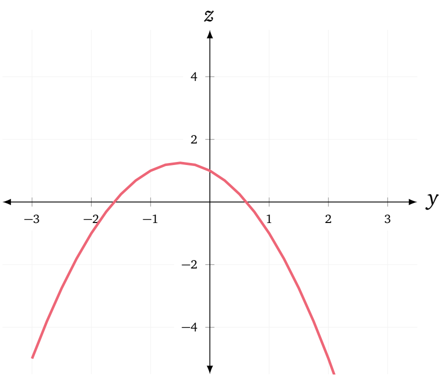
i) $\partial_j F(a)$ ii) $\partial_j F$ iii) $\partial_j F_i(a)$ iv) $\partial_j F_i$

(b) Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Consider the identity given by

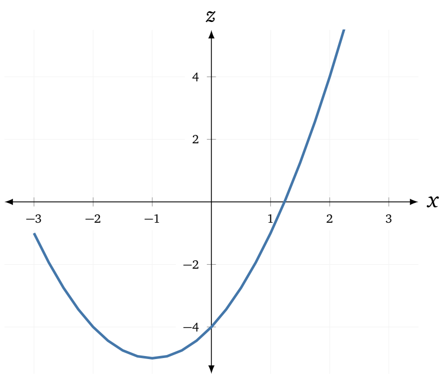
$$\lim_{h \to 0} \frac{F(a + he_j) - F(a)}{h} = \partial_j F(a) = \begin{bmatrix} \partial_j F_1(a) \\ \vdots \\ \partial_j F_m(a) \end{bmatrix} = (\partial_j F_1(a), \ldots, \partial_j F_m(a))$$

One equality holds by definition. One equality is an equivalent notation. One equality holds by a lemma. Identify which is which.

3.2.2 Below are slices of the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$.



x-slice at $x = -1$



y-slice at $y = 2$

If possible, determine whether each of the following values are positive, negative, or zero.

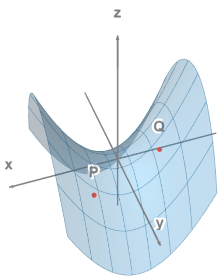
(a) $\dfrac{\partial f}{\partial x}(-1, 2)$

(b) $\dfrac{\partial f}{\partial x}(2, -1)$

(c) $\dfrac{\partial f}{\partial y}(-1, 2)$

(d) $\dfrac{\partial f}{\partial y}(2, -1)$

3.2.3 Consider the graph of a function $g : \mathbb{R}^2 \to \mathbb{R}$ whose partials exist everywhere (Math3D demo). The axes are labelled in the positive direction for x, y, and z.



(a) At $P = (a, b, g(a, b))$, identify if $g_x(a, b)$ and $g_y(a, b)$ are positive, negative, or zero. Select the best guess.

(b) At $Q = (c, d, g(c, d))$, identify if $g_x(c, d)$ and $g_y(c, d)$ are positive, negative, or zero. Select the best guess.

3.2.4 Let $f:\mathbb{R}^2\to\mathbb{R}$ and let $F:\mathbb{R}^2\to\mathbb{R}^2$. Fix $(a,b)\in\mathbb{R}^2$. Which statements are true or false?

(a) $g:\mathbb{R}\to\mathbb{R}$ given by $g(t)=f(a,t)$ is differentiable at $t=b$ if and only if $\partial_1 f(a,b)$ exists.

(b) $h:\mathbb{R}\to\mathbb{R}$ given by $h(t)=f(t,b)$ is differentiable at $t=a$ if and only if $\partial_1 f(a,b)$ exists.

(c) The partial $\partial_1 f(a,b)$ exists if and only if the partial $\partial_2 f(a,b)$ exists.

(d) $G:\mathbb{R}\to\mathbb{R}^2$ given by $G(t)=F(a,t)$ is differentiable at $t=b$ if and only if $\partial_2 F(a,b)$ exists.

(e) $H:\mathbb{R}\to\mathbb{R}^2$ given by $H(t)=F(t,b)$ is differentiable at $t=a$ if and only if $\partial_1 F(a,b)$ exists.

(f) The partial $\partial_1 F(a,b)$ exists if and only if the partial $\partial_2 F(a,b)$ exists.

Computations

3.2.5 A real-valued function $h(x,y)$ is represented by the table of data below.

$y\downarrow$ \ $x\rightarrow$	0	0.5	1	1.5	2	2.5
5	43	48	60	63	72	81
10	58	65	73	79	88	89
15	72	63	85	32	47	51
20	83	74	87	44	31	37

(a) Estimate the partial derivative $h_x(1,10)$ in two different ways.

(b) Estimate the partial derivative $h_y(1,10)$ in two different ways.

(c) Use the previous parts to estimate $h(0.8,12)$.

3.2.6 Let $f(x,y,z)=x^2y+e^{xy}$. Calculate $f_x(1,0,3)$ and $f_y(1,0,3)$ and $f_z(1,0,3)$.

3.2.7 Let $F(x,y)=(x\cos y, x\sin y)$. Compute $\partial_1 F(2,\pi)$ and $\partial_2 F(2,\pi)$.

3.2.8 Define $g(x,y,z)=xy^2z^3\arctan(xyz)$ and calculate g_x, g_y, and g_z.

3.2.9 Define $G(x,y,z)=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$ and calculate G_x, G_y, and G_z.

3.2.10 Define $f:\mathbb{R}^n\to\mathbb{R}$ by $f(x)=||x||^2$ and calculate $\partial_j f$ for each $1\le j\le n$.

3.2.11 Fix $a\in\mathbb{R}^n$. Define $g:\mathbb{R}^n\to\mathbb{R}$ by $g(x)=\dfrac{1}{||x-a||}$ and calculate $\partial_j g$ for each $1\le j\le n$.

3.2.12 Let A be an $m\times n$ matrix and define $F:\mathbb{R}^n\to\mathbb{R}^m$ by $F(x)=Ax$. Calculate $\partial_j F$ for $j\in\{1,\dots,n\}$.

Proofs

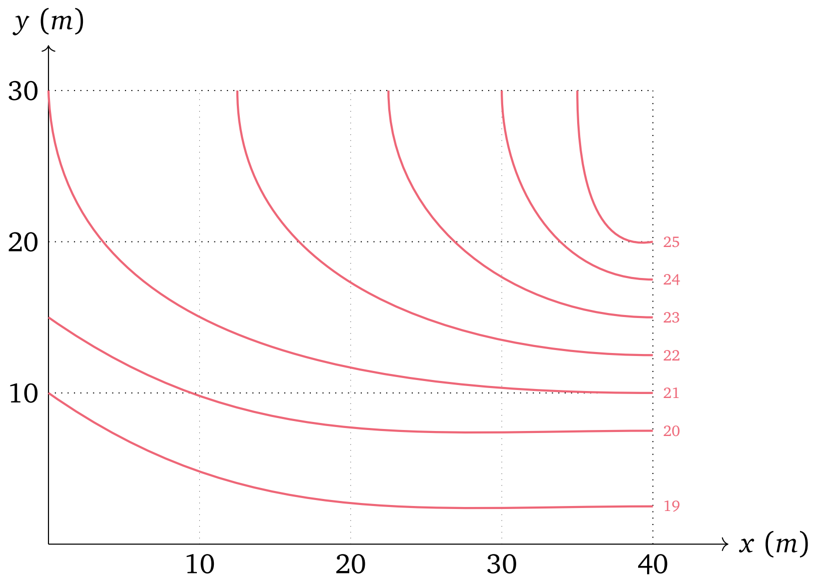
3.2.13 Differentiation rules for partial derivatives are straightforward because they are imported from single variable calculus. Other rules are similar. Fix $j\in\{1,\dots,n\}$ and $\lambda\in\mathbb{R}$.

(a) Let $f:\mathbb{R}^n\to\mathbb{R}$ and $g:\mathbb{R}^n\to\mathbb{R}$. Assume both of their partial derivatives $\partial_j f$ and $\partial_j g$ exist at $a\in\mathbb{R}^n$. Prove that $\partial_j(f+\lambda g)(a)$ exists and $\partial_j(f+g)(a)=\partial_j f(a)+\lambda\partial_j g(a)$.

(b) Let $F:\mathbb{R}^n\to\mathbb{R}^m$ and $G:\mathbb{R}^n\to\mathbb{R}^m$. Assume both $\partial_j F$ and $\partial_j G$ exist at $a\in\mathbb{R}^n$. Give **two** proofs that $\partial_j(F+\lambda G)(a)$ exists and $\partial_j(F+\lambda G)(a)=\partial_j F(a)+\lambda\partial_j G(a)$.

Applications and beyond

3.2.14 The contour map below [4] represents the temperature $T = T(x, y)$ measured in degrees Celsius, within a 40 m by 30 m space.



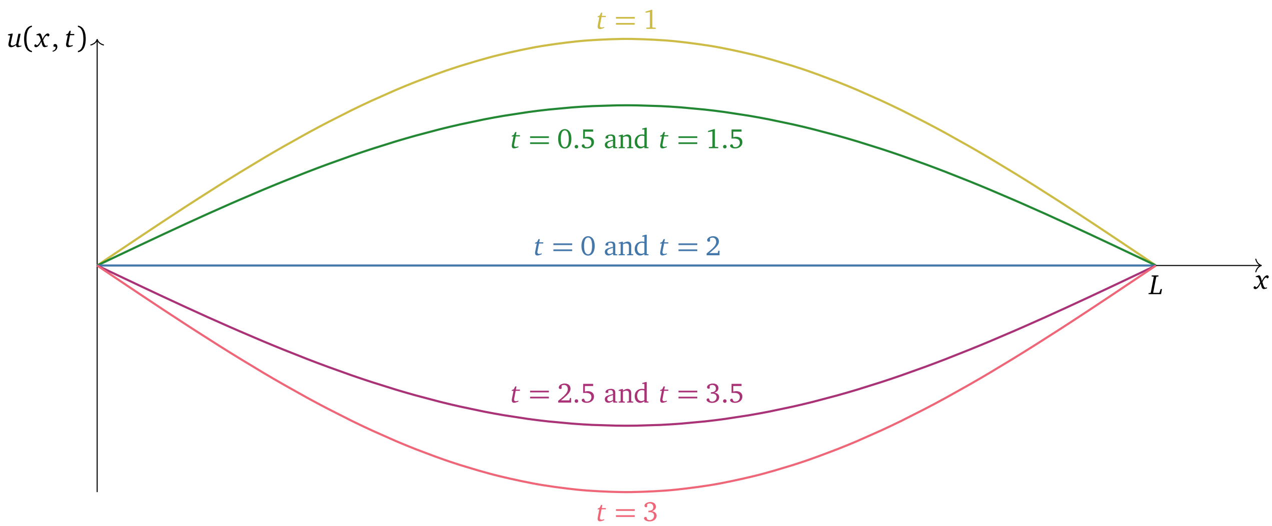
(a) Use plain language to describe what $T_y(10, 10)$ represents. Include units in your description.

(b) Without doing any calculations, is $T_y(10, 10)$ positive, zero, or negative?

(c) Approximate $T_y(10, 10)$.

(d) Approximate $T(10, 10.1)$.

3.2.15 A string of length 10 metres is fixed at both ends ($x = 0$ and $x = 10$). When plucked, the string forms a standing wave. The function $u(x, t)$ (measured in metres) is the **displacement of the string at point x (measured in metres) at time t (measured in seconds)**[5].



[4] Image created by Cindy Blois with permission.

Fix a point $x_0 \in (0, L)$ and time $t_0 \in (0, \infty)$.

(a) Use plain language to describe what $u_x(x_0, t_0)$ represents. Include units in your description.

(b) Use plain language to describe what $u_t(x_0, t_0)$ represents. Include units in your description.

(c) Determine whether each quantity is positive, zero, or negative. Choose the most plausible.

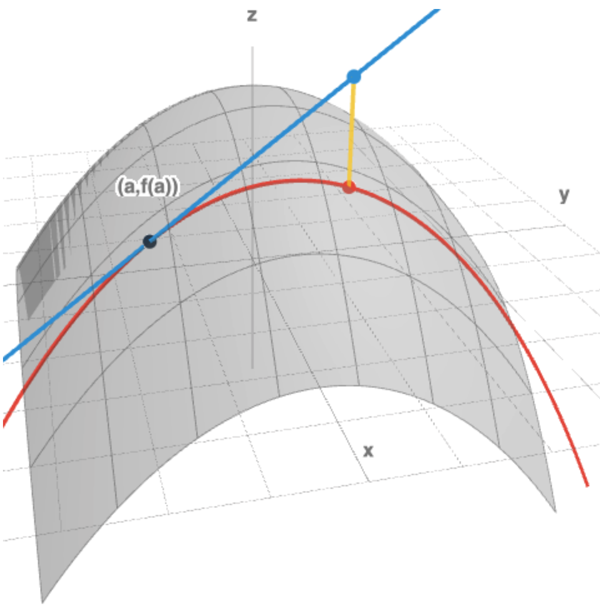
i) $u_x(5, 0)$ ii) $u_x(0, 1)$ iii) $u_t(0, 1)$ iv) $u_t(5, 1)$

v) $u_x(5, 0.5)$ vi) $u_x(2, 2.5)$ vii) $u_t(2, 2.5)$ viii) $u_t(5, 1.5)$

3.2.16 Given a job posting, the number of applicants n can be modeled as a function $f(w, d)$ of the amount of wage w in dollars/hour and the distance d from the city center in kilometers. Using the units of n, w, and d, describe the following statements in plain language.

$$\frac{\partial f}{\partial w}(20, 2) = 10.2, \qquad \frac{\partial f}{\partial d}(20.2) = -2.8.$$

3.2.17 You can apply the four viewpoints to the partial derivatives of a **real-valued** function. Real-valued functions have special interpretations since partial derivatives are precisely like differentiation in single variable calculus. Assume the jth partial of $f : \mathbb{R}^n \to \mathbb{R}$ exists at $a \in \mathbb{R}^n$.



Math3D demo

Use the figure above for parts (b) and (c).

(a) The **algebraic** viewpoint is:

$\partial_j f(a)$ is the instantaneous rate of change of f at a in the e_j direction.

Write this down precisely using a limit.

(b) The **analytic** viewpoint is:

$\partial_j f(a)$ defines a linear approximation of f at a in the e_j direction with small error.

First, label the two points with $(a+he_j, f(a+he_j)) \in \mathbb{R}^n \times \mathbb{R}$ and $(a+he_j, f(a)+h\partial_j f(a)) \in \mathbb{R}^n \times \mathbb{R}$. Second, label the "exact value", "approximate value", and "error" in the picture.

[5] Image created by Cindy Blois with permission.

(c) The **geometric** viewpoint is:

> $\partial_j f(a)$ *is the slope of the tangent line of* $f(a_1, \ldots, a_{j-1}, t, a_{j+1}, \ldots, a_n)$ *at* $t = a_j$.

How does the same picture of the graph of f relate? Include a plot in $\mathbb{R}^2$ to explain.

(d) A **physical/applied** viewpoint is:

> $\partial_j f(a)$ *approximates the change in* f *if* a_j *increases to* $a_j + 1$ *with other variables fixed.*

Let $R(p,q)$ be a company's revenue in dollars generated by selling PayStations at price p dollars per unit and Sintendos at price q dollars per unit. These are competing products so their prices affect each other's sales.

Use plain language to describe what $\dfrac{\partial R}{\partial p}(500, 200)$ represents in practical terms.

3.3. Directional derivatives

Partial derivatives were defined by fixing variables (or slicing). This was an attempt to reduce a high dimensional setting to a one-dimensional setting. Unfortunately, partial derivatives are restrictive since they only discuss changes in a function in a *fixed* direction. For a function with domain in $\mathbb{R}^n$, there are only n such directions, given by the standard basis $\{e_1, \dots, e_n\}$. Indeed, Definition 3.2.2 defines a limit in the direction e_j for a fixed $j \in \{1, \dots, n\}$.

How can you approximate changes in any other direction?

You will introduce another approach to define a higher dimensional derivative using one dimension. Namely, you will study *directional derivatives* which describe how a function changes along a line *in any given direction*. This investigation will suggest a relationship with partial derivatives, and a novel geometric perspective with "tangent vectors".

3.3.1 Definition

The change in f at $a \in \mathbb{R}^n$ in the direction of a standard basis vector is captured by Definition 3.2.2. You can generalize this idea to any direction $v \in \mathbb{R}^n$ along the line $\{a + hv : h \in \mathbb{R}\}$.

Definition 3.3.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. Fix a vector[6] $v \in \mathbb{R}^n$. The **directional derivative of f at a in the direction** v is given by

$$D_v f(a) = \lim_{h \to 0} \frac{f(a + hv) - f(a)}{h}$$

The **directional derivative of f in the direction** v is the function $D_v f : U \to \mathbb{R}^m$, where U is the set of points $a \in A$ such that $D_v f(a)$ exists.

Remark 3.3.2 Since a is an interior point of A, there exists $\varepsilon > 0$ such that the quantity $f(a + hv)$ is defined for $h \in (-\varepsilon, \varepsilon)$. It is a good exercise to formally prove this statement.

Remark 3.3.3 If $\{e_1, \dots, e_n\}$ is the standard basis for $\mathbb{R}^n$, then $D_{e_j} f = \partial_j f$ so partial derivatives are a special case of directional derivatives.

Example 3.3.4 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = 9 - x^2 - y^2$. You want to describe how f changes near the point $a = (1, -1)$ in the direction $v = (2, -1) \in \mathbb{R}^2$.

Consider the *algebraic* viewpoint. Define the single variable function $g : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} g(h) = f(a + hv) &= f((1, -1) + h(2, -1)) \\ &= f(1 + 2h, -1 - h) \\ &= 9 - (1 + 2h)^2 - (-1 - h)^2 = 7 - 6h - 5h^2. \end{aligned}$$

Again, g is a single variable function so you can calculate its derivative to find that $g'(h) = -6 - 10h$. Notice, however, that only the derivative $g'(0) = -6$ relates to the function f near the point $(1, -1)$. The limit definition implies that

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(a + hv) - f(a)}{h} = D_v f(a).$$

This implies that $D_{(2,-1)} f(1, -1) = g'(0) = -6$.

[6]Many texts assume that v is always a unit vector for a directional derivative. This assumption is not necessary.

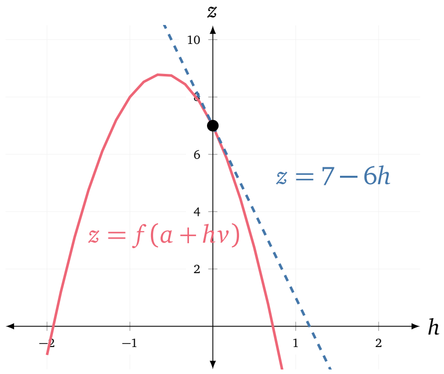
Consider the *analytic* viewpoint. Definition 3.3.1 heuristically implies for $h \approx 0$ that

$$f(a+hv) \approx f(a) + hD_v f(a) = f(1,-1) + hD_{(2,-1)} f(1,-1) = 7 - 6h.$$

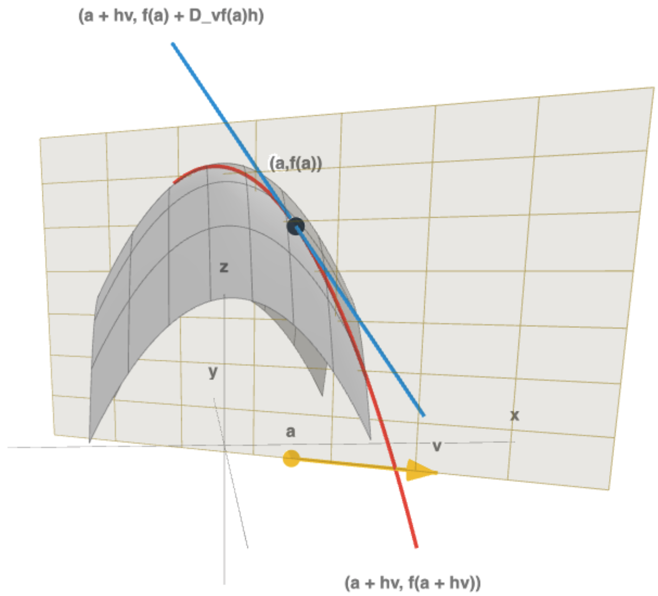
In other words, the function $\ell : \mathbb{R} \to \mathbb{R}$ given by $\ell(h) = f(a) + hD_v f(a)$ linearly approximates $g : \mathbb{R} \to \mathbb{R}$ given by $g(h) = f(a+hv)$. For instance,

$$f(1.2,-1.1) = f(a+0.1v) \approx f(a) + 0.1D_v f(a) = 7 - 6(0.1) = 6.4.$$

Consider the *geometric* viewpoint. You can plot $g(h) = f(a+hv) = 7 - 6h - 5h^2$ as usual.



The slope of the tangent line is $g'(0) = -6$. This graph corresponds to a slice of the graph of f as illustrated below. View the Math3D demo for a better visual.



Formally, the curve in the 3D plot is given by

$$\{(a+hv, f(a+hv)) : h \in \mathbb{R}\} = \{(1+2h, -1-h, 7-6h-5h^2) : h \in \mathbb{R}\}$$

and the line is given by

$$\{(a+hv, f(a) + hD_v f(a)) : h \in \mathbb{R}\} = \{(1+2h, -1-h, 7-6h) : h \in \mathbb{R}\}.$$

3.3.2 Computations

Directional derivatives have again reduced differentiation to single variable functions. This produces the usual basic lemmas and properties.

Lemma 3.3.5 Let $A \subseteq \mathbb{R}^n$ and let $f = (f_1, \ldots, f_m) : A \to \mathbb{R}^m$. Let a be an interior point of A. Fix $v \in \mathbb{R}^n$. Then $D_v f(a)$ exists if and only if $D_v f_i(a)$ exists for every $i \in \{1, \ldots, m\}$. If so,

$$D_v f(a) = (D_v f_1(a), \ldots, D_v f_m(a)).$$

Exercise 3.3.6 State and prove a theorem like Theorem 3.2.8 for directional derivatives.

Until partial derivatives, it was not clear how to use differentiation rules from single variable calculus to calculate directional derivatives. The first method is to use the limit definition.

Example 3.3.7 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = x^2 y$. You can calculate its directional derivative in the direction $v = (1,2)$ using the limit definition. For any $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned} D_{(1,2)} f(x,y) &= \lim_{h \to 0} \frac{f(x+h, y+2h) - f(x,y)}{h} \\ &= \lim_{h \to 0} \frac{(x+h)^2(y+2h) - x^2 y}{h} \\ &= \lim_{h \to 0} \frac{2xyh + yh^2 + 2x^2 h + 4xh^2 + 2h^3}{h} = 2xy + 2x^2. \end{aligned}$$

This computational obstacle is frustrating. It would be better to have a more efficient method the exploits the ease of calculating partial derivatives.

Linear algebra provides a critical insight. Recall any $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ can be written as a linear combination of the standard basis $\{e_1, \ldots, e_n\}$ in $\mathbb{R}^n$, namely, $v = \sum_{j=1}^n v_j e_j$. A linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ will send the basis $\{e_1, \ldots, e_n\}$ in $\mathbb{R}^n$ to a set of vectors $\{T(e_1), \ldots, T(e_n)\}$ in $\mathbb{R}^m$. These may not be a basis for $\mathbb{R}^m$ but this finite set will span the image $\operatorname{im}(T)$. That is,

$$T(v) = \sum_{j=1}^n v_j T(e_j). \tag{3.3.1}$$

Since partial derivatives correspond to directional derivatives with the standard basis, this suggests a pair of reasonable questions.

If every partial derivative $\partial_1 f, \ldots, \partial_n f$ exists, does every directional derivative $D_v f$ exist for all $v \in \mathbb{R}^n$? If so, can you write it as a linear combination of partials?

Without more assumptions, the answer to the first question is unfortunately negative.

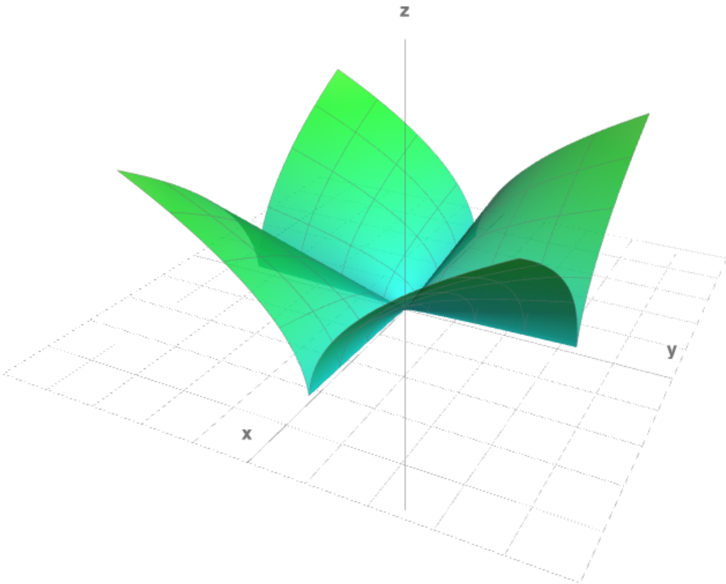
Example 3.3.8 Define the function $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(x,y) = |xy|^{1/2}$. By the limit definition,

$$\partial_1 g(0,0) = \lim_{h \to 0} \frac{g(h,0) - g(0,0)}{h} = \lim_{h \to 0} \frac{0-0}{h} = 0,$$

and similarly $\partial_2 g(0,0) = 0$. On the other hand,

$$D_{(1,1)} g(0,0) = \lim_{h \to 0} \frac{g(h,h) - g(0,0)}{h} = \lim_{h \to 0} \frac{|h|}{h}$$

does not exist. Both partials exist, yet a directional derivative does not. In fact, you can verify no other directional derivatives exist. You can see this issue with the graph of g below.



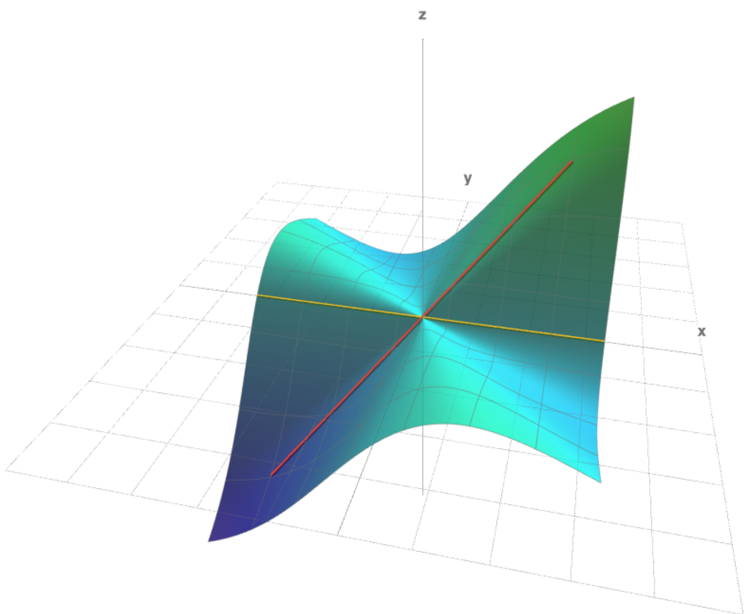
Play with this Math3D demo to better visualize this phenomenon.

Even if you assume that the directional derivative exists, the answer to the second question is also negative without further assumptions.

Example 3.3.9 Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = \frac{x^2y+xy^2}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. You can again verify that $\partial_1 f(0,0) = \partial_2 f(0,0) = 0$. However,

$$D_{(1,1)}f(0,0) = \lim_{h\to 0} \frac{f(h,h) - f(0,0)}{h} = \lim_{h\to 0} \frac{h-0}{h} = 1.$$

Thus, $D_{(1,1)}f(0,0)$ exists yet $D_{(1,1)}f(0,0) \neq \partial_1 f(0,0) + \partial_2 f(0,0)$. In fact, $D_{(1,1)}f(0,0)$ cannot be expressed as any linear combination of the partials!



Play with this Math3D demo for a better visual.

These two examples illustrate that you will require a more powerful assumption about a non-linear map f to "glue together" all directional derivatives $D_v f$ in a way that respects linearity like (3.3.1). This critical assumption will ultimately be *differentiability* of f.

Theorem 3.3.10 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is differentiable at a, then for all $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$,

$$D_v f(a) = \sum_{j=1}^{n} v_j \partial_j f(a).$$

Proof. The notion of "differentiable" has not yet been introduced and this lemma will follow from a more general theorem later. The proof is therefore postponed and you may take this for granted unless specified otherwise. ∎

Assuming functions are "differentiable" (whatever that means), you can use Theorem 3.3.10 to quickly compute their directional derivatives.

Example 3.3.11 Consider again $f(x,y) = x^2 y$. Its partial derivatives are $\partial_1 f(x,y) = 2xy$ and $\partial_2 f(x,y) = x^2$ by standard differentiation rules. By Theorem 3.3.10, if f is differentiable, then it follows that

$$D_{(1,2)} f(x,y) = \partial_1 f(x,y) + 2\partial_2 f(x,y) = 2xy + 2x^2.$$

This matches the calculation from the previous example, but it was way simpler to compute.

3.3.3 Geometry of directional derivatives

Directional derivatives, along with partial derivatives, can be geometrically interpreted using the graph of a function as in Example 3.3.4. This visualization strongly relied on the fact that the function was *real-valued*. Since it was real-valued, you could interpret the directional derivative as the "slope" along a slice in a given direction. Also, it was a map of the form $\mathbb{R}^2 \to \mathbb{R}$, so its graph lies in $\mathbb{R}^3$ which you can visualize.

How can you think about directional derivatives for vector-valued maps? There is a deeper interpretation from geometry.

> *The directional derivative of f at a in the direction v outputs a vector $D_v f(a)$ that is* ***tangent to the parametric curve*** *$\gamma(t) = f(a + tv)$ at $t = 0$.*

Informally speaking, this holds since

$$\gamma'(0) = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t} = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} = D_v f(a).$$

This alternate perspective will generalize beautifully to directional derivatives for any map of the form $\mathbb{R}^n \to \mathbb{R}^m$. You will investigate this feature in some examples.

Example 3.3.12 Recall the polar coordinate transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$f(r,\theta) = (r\cos\theta, r\sin\theta).$$

You can read Section 1.4 for more details on this transformation. Assume f is "differentiable" so by Theorem 3.3.10 all of its directional derivatives exist. Calculating its partial derivatives, you find that

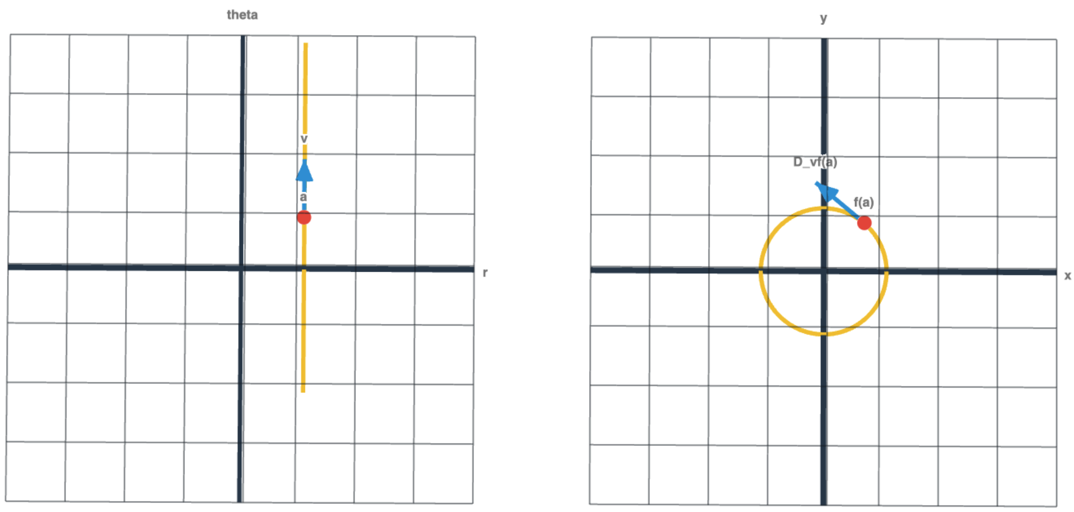
$$\partial_1 f(r,\theta) = (\cos\theta, \sin\theta), \qquad \partial_2 f(r,\theta) = (-r\sin\theta, r\cos\theta)$$

so by Theorem 3.3.10, it follows that for any vector $v = (v_1, v_2) \in \mathbb{R}^2$,

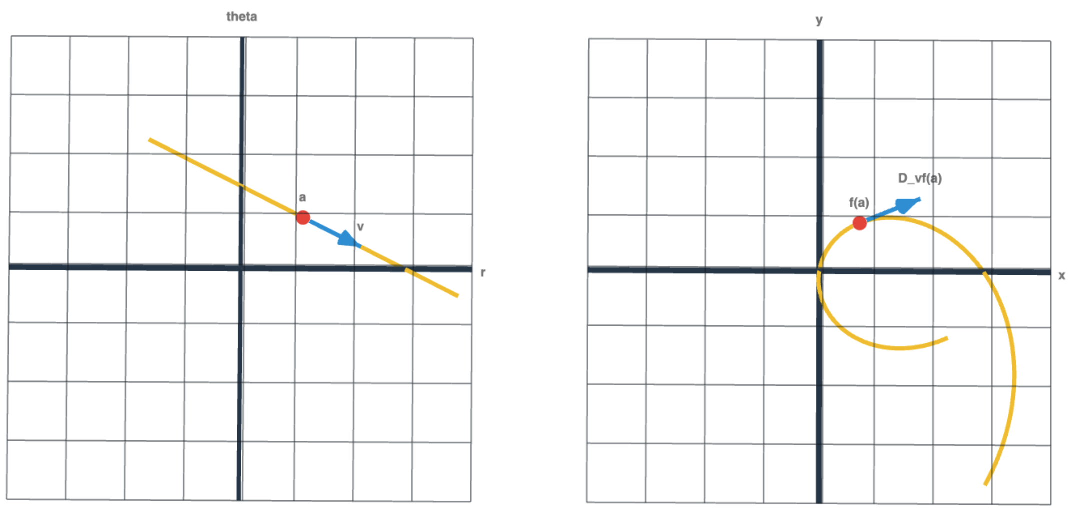
$$D_v f(r,\theta) = v_1 \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + v_2 \begin{bmatrix} -r\sin\theta \\ r\cos\theta \end{bmatrix} = \begin{bmatrix} v_1\cos\theta - v_2 r\sin\theta \\ v_1\sin\theta + v_2 r\cos\theta \end{bmatrix}. \tag{3.3.2}$$

This gives a nice algebraic formula, but what does it represent geometrically?

Fix a point $a \in \mathbb{R}^2$ and a vector $v \in \mathbb{R}^2$. Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ defined by $\gamma(t) = a + tv$, so γ is a straight line passing through a. Think of this line as lying in the (r,θ)-plane, which is the domain of the polar coordinate transformation. This line $\ell = \{a+tv : t \in \mathbb{R}\}$ in the (r,θ)-plane is mapped via f to the curve $f(\ell) = \{f(a+tv) : t \in \mathbb{R}\}$ lying in the (x,y)-plane. This is illustrated below for the choice $v = (0,1)$.



You can think of the direction vector $v \in \mathbb{R}^2$ as being mapped to the directional derivative $D_v f(a) \in \mathbb{R}^2$. Notice $D_v f(a)$ is *tangent to the curve* $f(\ell)$ at the point $f(a)$. This same phenomenon still holds if you choose different directions v.



Play with this Math3D demo by moving the sliders.

You can view the same phenomenon for a set parametrized by a map $\mathbb{R}^2 \to \mathbb{R}^3$.

Example 3.3.13 Define the vector-valued map $F : \mathbb{R}^2 \to \mathbb{R}^3$ by $F(x, y) = (x, y, 9 - x^2 - y^2)$. Notice F is a parametrization of the graph of the real-valued function $f(x, y) = 9 - x^2 - y^2$ from Example 3.3.4. This graph can be written in parametric form as

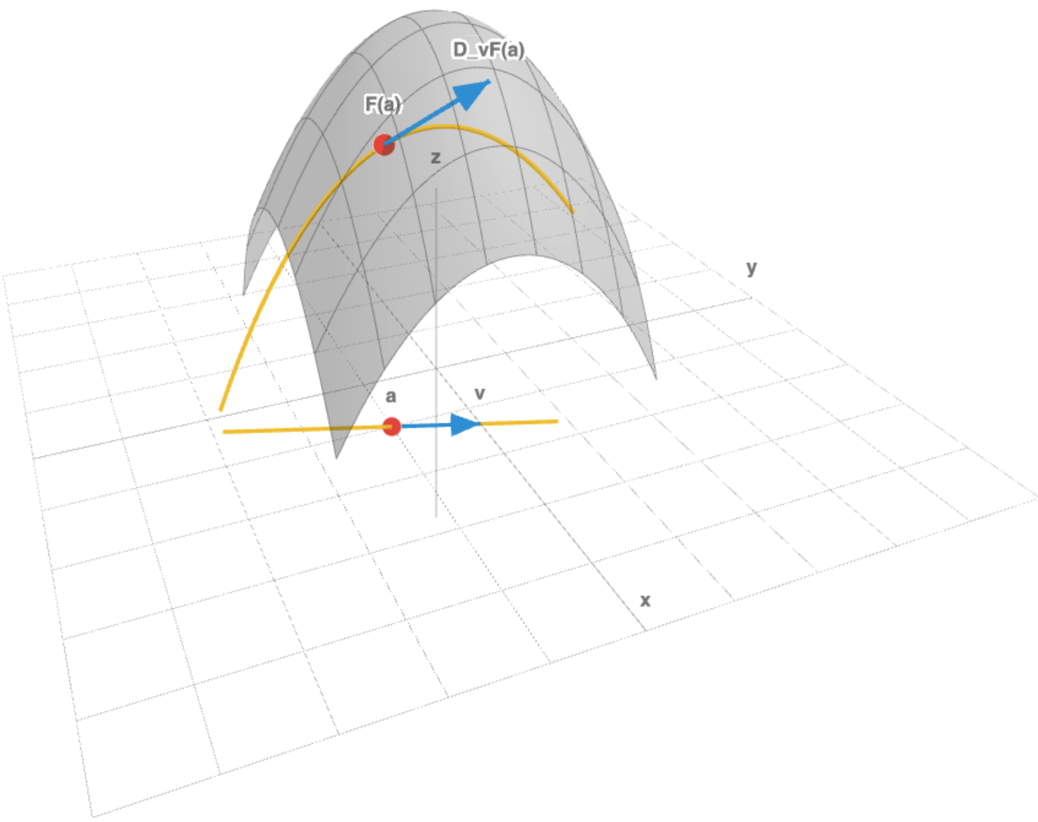
$$S = \{(x, y, 9 - x^2 - y^2) : x, y \in \mathbb{R}\} = F(\mathbb{R}^2).$$

Recall you interpreted a directional derivative of f as a slope of a tangent line for a single variable function. By instead studying directional derivatives of F, you will view the same picture from an entirely different perspective using vectors tangent to curves on S.

Assume F is differentiable so for any $v = (v_1, v_2) \in \mathbb{R}^2$ and any $(x, y) \in \mathbb{R}^2$,

$$D_v F(x, y) = v_1 \partial_1 F(x, y) + v_2 \partial_2 F(x, y) = (v_1, v_2, -2v_1 x - 2v_2 y) \in \mathbb{R}^3$$

by Theorem 3.3.10. Fix a point $a \in \mathbb{R}^2$ and a vector $v \in \mathbb{R}^2$. You can again think of how the line $\ell = \{a + tv : t \in \mathbb{R}\}$ in the (x, y)-plane is mapped via F to the curve $F(\ell) = \{F(a + tv) : t \in \mathbb{R}\}$ lying on the graph S in (x, y, z)-space. The direction $v \in \mathbb{R}^2$ maps to the directional derivative $D_v F(a) \in \mathbb{R}^3$. This description is illustrated below in this Math3D demo.



Comparing this idea with Example 3.3.4, you are hopefully convinced that viewing directional derivatives as “tangent vectors” is much more useful than as “slopes”.

This observation about the geometry of directional derivatives will later blossom into a deeper study of tangent planes and surfaces. Right now, everything hinges on understanding when you can express directional derivatives as a linear combinations of partials (Theorem 3.3.10), much like linear maps. This desirable property suggests derivatives allow non-linear maps to be treated like linear maps! That is a remarkable insight from linear algebra.

Exercises for Section 3.3

Concepts and definitions

3.3.1 Let $A \subseteq \mathbb{R}^n$. Assume $F = (F_1, \ldots, F_m) : A \to \mathbb{R}^m$ is differentiable[7] at every interior point of A. Fix $v \in \mathbb{R}^n$. Let a be an interior point of A. Fix $i \in \{1, \ldots, n\}$

(a) For each quantity, identify the type of mathematical object that it is.

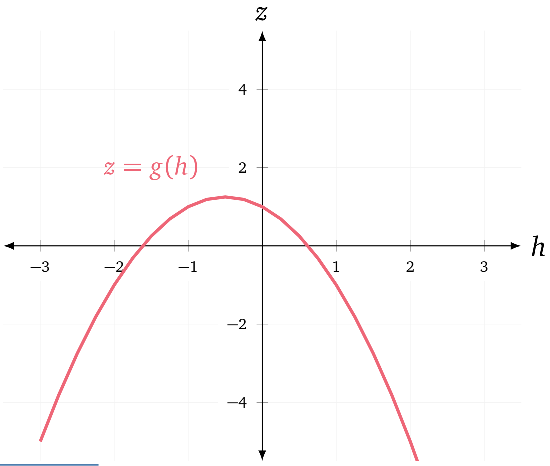
i) $D_v F(a)$ ii) $D_v F$ iii) $D_v F_i(a)$ iv) $D_v F_i$

(b) Write $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Since F is differentiable at a, the following identity holds:

$$\lim_{t \to 0} \frac{F(a + tv) - F(a)}{t} = D_v F(a) = \sum_{j=1}^{n} v_j \partial_j F(a) = v_1 \begin{bmatrix} \partial_1 F_1(a) \\ \vdots \\ \partial_1 F_m(a) \end{bmatrix} + \cdots + v_n \begin{bmatrix} \partial_n F_1(a) \\ \vdots \\ \partial_n F_m(a) \end{bmatrix}$$

One equality holds by definition. One equality holds by a lemma. One equality holds by a theorem. Identify which is which.

3.3.2 Let $f : \mathbb{R}^2 \to \mathbb{R}$ and assume it is "differentiable" everywhere. Fix $a = (-2, -1)$ and let $v = (3, 4)$. Below is a graph of the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(h) = f(a + hv)$.



If possible, determine whether each value is positive, negative, or zero.

(a) $f(-2, -1)$

(b) $\dfrac{\partial f}{\partial x}(-2, -1)$

(c) $\dfrac{\partial f}{\partial y}(-2, -1)$

(d) $D_v f(-2, -1)$

3.3.3 Multivariable calculus is so pervasive across many subjects that there is a lot of equivalent notation. For this course, you should use notation consistent with the textbook but you should also be aware of other usage. Let $f : \mathbb{R}^n \to \mathbb{R}^m$. Let $a, v \in \mathbb{R}^n$ and $i \in \{1, 2, \ldots, n\}$.

(a) The directional derivative with respect to v of f at a has equivalent expressions such as

$$D_v f(a) \qquad \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} \qquad \frac{d}{dt} f(a + tv)\Big|_{t=0} \qquad \partial_v f(a) \qquad \nabla_v f(a)$$

One of these is the textbook notation and one of these is the definition. Identify them.

(b) What is the difference between $D_v f$ and $D_v f(a)$?

(c) The ith partial derivative of f at a has many equivalent expressions such as

$$D_i f(a) \qquad D_{e_i} f(a) \qquad \frac{\partial f}{\partial x_i}(a) \qquad \partial_{e_i} f(a) \qquad \partial_i f(a) \qquad \frac{\partial f}{\partial x_i}\Big|_{x=a} \qquad f_{x_i}(a)$$

$$\lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h} \qquad \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_m) - f(a_1, \ldots, a_i, \ldots, a_m)}{h}$$

Three of these are the textbook notation and four of them are valid definitions. Identify them.

[7] You have not yet defined this term, but do not worry about it for now. Assume it and carry on.

(d) All of the following expressions represent the same thing:

$$D_i f \qquad D_{e_i} f \qquad \frac{\partial f}{\partial x_i} \qquad \partial_i f \qquad f_{x_i}$$

Describe what they represent in words.

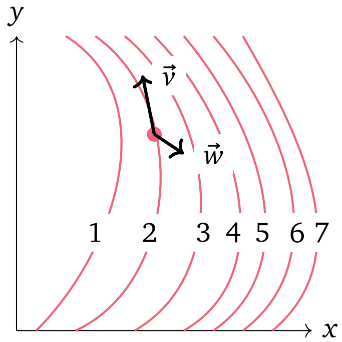
3.3.4 Let $F : \mathbb{R}^n \to \mathbb{R}^m$. Fix $a, w \in \mathbb{R}^n$. Which statements are true or false?

(a) $\gamma : \mathbb{R} \to \mathbb{R}^m$ given by $\gamma(t) = F(a + tw)$ is differentiable at $t = 0$ if and only if $D_w F(a)$ exists.

(b) If $D_v F(a)$ exists for every $v \in \mathbb{R}^n$ then $\partial_1 F(a), \ldots, \partial_n F(a)$ all exist.

(c) If $\partial_1 F(a), \ldots, \partial_n F(a)$ all exist, then $D_v F(a)$ exists for every $v \in \mathbb{R}^n$.

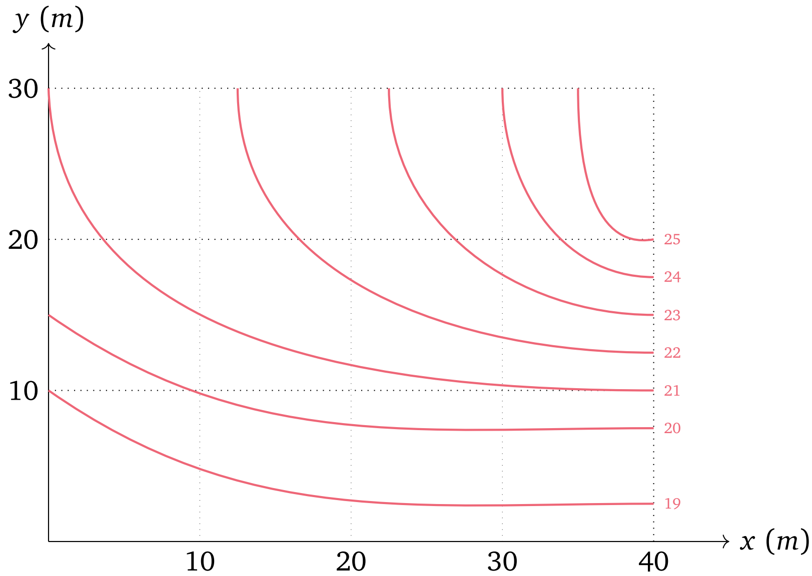
3.3.5 Consider the contour plot of the function $f : \mathbb{R}^2 \to \mathbb{R}$ below.



At the given point a, determine whether $D_v f(a)$ and $D_w f(a)$ are positive, zero, or negative.

Computations

3.3.6 The contour map below represents the temperature $T = T(x, y)$ measured in degrees Celsius, within a 40 m by 30 m space.



(a) Without doing any calculations, is $D_{(-1,-1)} T(10, 10)$ positive, zero, or negative?

(b) Approximate $D_{(-1,-1)} T(10, 10)$ in at least two different ways.

(c) Approximate $T(9.5, 9.5)$ using one of your estimates above.

3.3.7 Partial derivatives are computed rather easily through differentiation rules. Assuming Theorem 3.3.10 holds, the directional derivative can be computed using the partial derivatives. You have not yet seen the definition of "differentiable" but do not worry about that assumption now. You will see how to apply this theorem in an example. Define

$$f(x,y) = (x^2 - y^2, e^{xy} + xy).$$

Assume f satisfies the assumptions of Theorem 3.3.10 for all parts below.

(a) Compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

(b) Let $v = (2,-3)$. Express $D_v f$ in terms of the partial derivatives of f.

(c) Compute $D_v f(1,-1)$.

3.3.8 Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = \arctan(x) + y^2$. Assume f is differentiable at $(1,2)$.

(a) Find the average rate of change of f when you go from $(1,2)$ to $(3,1)$.

(b) Find the instantaneous rate of change of f as you leave the point $(1,2)$ heading toward $(3,1)$.

3.3.9 Assume $g(x,y,z) = xy^2z^3 \arctan(xyz)$ satisfies Lemma 3.3.10. For $v \in \mathbb{R}^3$, calculate $D_v g$.

3.3.10 Assume $f(x,y,z) = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ satisfies Theorem 3.3.10. For $v \in \mathbb{R}^3$, calculate $D_v f$.

3.3.11 Assume $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = ||x||^2$ satisfies Theorem 3.3.10. For $v \in \mathbb{R}^n$, find $D_v f$.

3.3.12 Let A be an $m \times n$ matrix. Assume $F : \mathbb{R}^n \to \mathbb{R}^m$ defined by $F(x) = Ax$ satisfies Theorem 3.3.10. For $v \in \mathbb{R}^n$, compute $D_v F$.

Proofs

3.3.13 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^n \to \mathbb{R}^m$. Fix $v \in \mathbb{R}^n$. Assume both of their directional derivatives $D_v F$ and $D_v G$ exist at $a \in \mathbb{R}^n$. Prove that $D_v(F+G)(a)$ exists and that

$$D_v(F+G)(a) = D_v F(a) + D_v G(a).$$

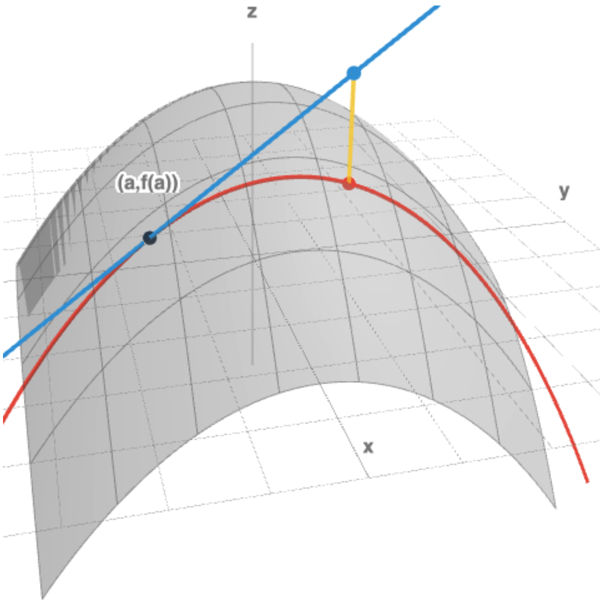
Applications and beyond

3.3.14 You are a data scientist who is designing a machine learning algorithm. At each step of the algorithm, you need to choose n variables $x_1, \ldots, x_n$ which minimize a certain error function $E : \mathbb{R}^n \to [0, \infty)$. Unfortunately, since n is extremely large and E is quite complicated, it is hopeless to solve this optimization problem algebraically. You attempt to numerically solve this problem using the directional derivative.

(a) You have made an initial guess for a point $p \in \mathbb{R}^n$ which minimizes E. You want to improve your guess by choosing a direction to move from p towards a new point. How should you choose this direction? Explain your answer in terms of the directional derivative.

(b) Once you have chosen a direction $u \in \mathbb{R}^n$, you want to move from p in that direction by a distance $h > 0$ to a new point $q \in \mathbb{R}^n$. Give a formula for q in terms of u, p, and h.

3.3.15 You can apply the four viewpoints to the directional derivatives of a **real-valued** function. Again, real-valued functions have special interpretations since partial derivatives are precisely like differentiation in single variable calculus. Let $f : \mathbb{R}^n \to \mathbb{R}$ and assume directional derivative of f exists at $a \in \mathbb{R}^n$ in the direction $v \in \mathbb{R}^n$.



Math3D demo

Use the above figure for parts (b) and (c).

(a) The **algebraic** viewpoint is:

$D_v f(a)$ is the instantaneous rate of change of f at a in the v direction.

Write this down precisely using a limit.

(b) The **analytic** viewpoint is:

$D_v f(a)$ defines a linear approximation of f at a in the v direction with small error.

First, label the two points as $(a+hv, f(a+hv))$ and $(a+hv, f(a)+D_v f(a)h)$. Second, label the "exact value", "approximate value", and "error" in the figure.

(c) The **geometric** viewpoint (for a *real-valued* function $f : \mathbb{R}^n \to \mathbb{R}$) is:

If $||v|| = 1$, then $D_v f(a)$ is the slope of the tangent line of $f(a+tv)$ at $t = 0$.

How does the same picture of the graph of f relate? Include a plot in $\mathbb{R}^2$ to explain.

(d) The **physical/applied** viewpoint is:

The marginal change $D_v f(a)$ approximates the change in f from a to $a+v$.

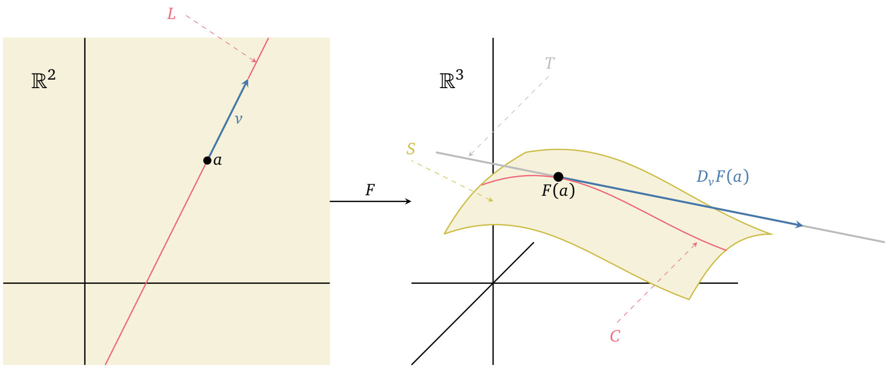
Let $R(p,q)$ be a company's revenue in dollars generated by selling PayStations at price p dollars per unit and Sintendos at price q dollars per unit. These are competing products so their prices affect each other's sales. Use plain language to describe what $D_{(-10,5)}R(500,200)$ represents in practical terms.

3.3.16 Directional derivatives can be interpreted in a much more general perspective using vectors tangent to a set in parametric form.

(a) For the **geometric** viewpoint, a function $F : \mathbb{R}^n \to \mathbb{R}^m$ with $a, v \in \mathbb{R}^n$ should satisfy:

The vector $D_vF(a)$, translated to $F(a)$, is tangent to the image of F at the point $F(a)$.

In other words, $D_vF(a)$ is a tangent vector to surface defined as the image of F. The figure below illustrates this idea with $n = 2$ and $m = 3$.



The labeling of the figure is incomplete. Match the line L, the surface S, the curve C, and the tangent line T to their corresponding expression.

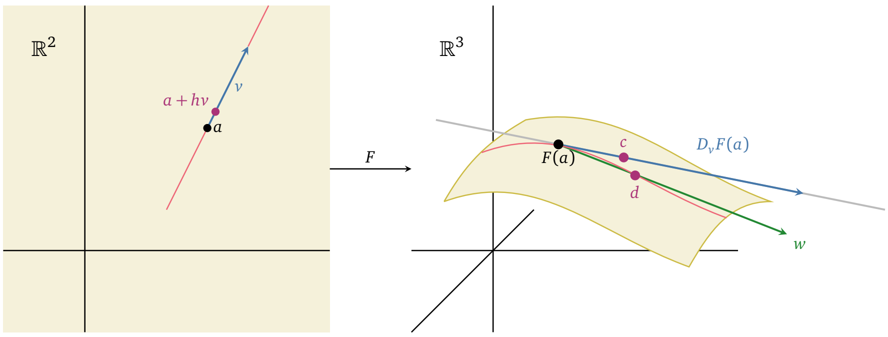
I) $\{tv : t \in \mathbb{R}\}$
II) $\{a + tv : t \in \mathbb{R}\}$
III) $\{F(a + tv) : t \in \mathbb{R}\}$
IV) $\{F(tv) : t \in \mathbb{R}\}$

I) $\{F(x) : x \in \mathbb{R}^n\}$
II) $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y = F(x)\}$
III) $\{F(a) + D_vF(a)t : t \in \mathbb{R}\}$
IV) $\{F(a) + D_aF(v)t : t \in \mathbb{R}\}$

(b) For the **analytic** viewpoint, a function $F : \mathbb{R}^n \to \mathbb{R}^m$ with $a, v \in \mathbb{R}^n$ should satisfy:

$F(a + hv) \approx F(a) + D_vF(a)h$ for small values of $h \in \mathbb{R}$.

In other words, the directional derivative allows you to linearly approximate the function along a line. The figure below illustrates this idea with $n = 2$ and $m = 3$ and small positive $h > 0$.



The labeling of the figure is incomplete. Match the point c, the point d, and the vector w to their corresponding expression. Afterwards, identify which of these expressions represent the "error in the approximation".

I) $F(a + hv)$
II) $F(a) + F(v)h$
III) $D_vF(a)h$
IV) $F(a) + D_vF(a)h$

I) $F(a + hv) - F(a)$
II) $F(a + hv) - F(a) - D_vF(a)h$
III) $\dfrac{F(a + hv) - F(a)}{h}$
IV) $\dfrac{F(a + hv) - F(a) - D_vF(a)h}{h}$

3.4. Gradients

Before pursuing the generalized definition of derivatives for any map $\mathbb{R}^n \to \mathbb{R}^m$, it is worthwhile to more deeply investigate real-valued maps $\mathbb{R}^n \to \mathbb{R}$ and your newfound discoveries with directional derivatives. The critical observation is Theorem 3.3.10.

If $f : \mathbb{R}^n \to \mathbb{R}$ is "differentiable" at $a \in \mathbb{R}^n$ then for any vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$,

$$D_v f(a) = \sum_{j=1}^{n} v_j \partial_j f(a) = (\partial_1 f(a), \ldots, \partial_n f(a)) \cdot (v_1, \ldots, v_n).$$

Informally speaking, the vector of partials $(\partial_1 f(a), \ldots, \partial_n f(a))$ captures all the necessary information about the change of f in *any* direction[8]. This warrants its own definition.

Definition 3.4.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. Let a be an interior point of A. The **gradient of f at a** is denoted[9] $\nabla f(a)$ and given by the n-dimensional vector

$$\nabla f(a) = (\partial_1 f(a), \ldots, \partial_n f(a)),$$

provided the partial derivatives $\partial_1 f(a), \ldots, \partial_n f(a)$ all exist.

The gradient is ubiquitous in applications and will repeatedly arise in many contexts. Since the gradient is defined with partial derivatives, it is straightforward to compute.

Example 3.4.2 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xy^2 + 2x$. Then $\partial_1 f(x, y) = y^2 + 2$ and $\partial_2 f(x, y) = 2xy$, so $\nabla f(x, y) = (y^2 + 2, 2xy)$. For example, $\nabla f(-1, 2) = (6, -4)$.

This definition allows you to rewrite Theorem 3.3.10 in terms of the gradient.

Theorem 3.4.3 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. Let a be an interior point of A. If f is differentiable at a, then for all $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, $D_v f(a)$ exists and

$$D_v f(a) = \nabla f(a) \cdot v = \big(\nabla f(a)\big)^T v.$$

Remark 3.4.4 Remember that the dot product of two column vectors $u, v \in \mathbb{R}^n$ is written as $u \cdot v = u_1 v_1 + \cdots + u_n v_n$ but in terms of matrix multiplication, this is written as $u^T v$ since the transpose u^T is a $1 \times n$ matrix and v is a $n \times 1$ matrix.

For now, you can freely assume functions are differentiable and apply Theorem 3.4.3.

Example 3.4.5 Let $f(x, y) = xy^2 + 2x$ be as in Example 3.4.2. Assuming f is differentiable at $(-1, 2)$, it follows that for any $v \in \mathbb{R}^2$,

$$D_v f(-1, 2) = (6, -4) \cdot (v_1, v_2) = 6v_1 - 4v_2.$$

This allows you to quickly linearly approximate values of f near $(-1, 2)$. For instance, since the point $(-0.9, 1.8)$ is equal to $(-1, 2) + (0.1, -0.2)$,

$$f(-0.9, 1.8) \approx f(-1, 2) + D_{(0.1,-0.2)} f(-1, 2) = -6 + 0.6 + 0.8 = -4.6.$$

The gradient is not simply a notational convenience. It has several deep geometric meanings.

[8] Notice a similarity with linear algebra? This property may remind you of how $\{e_1, \ldots, e_n\}$ is a basis for $\mathbb{R}^n$.
[9] The symbol ∇ is referred to as *nabla*.

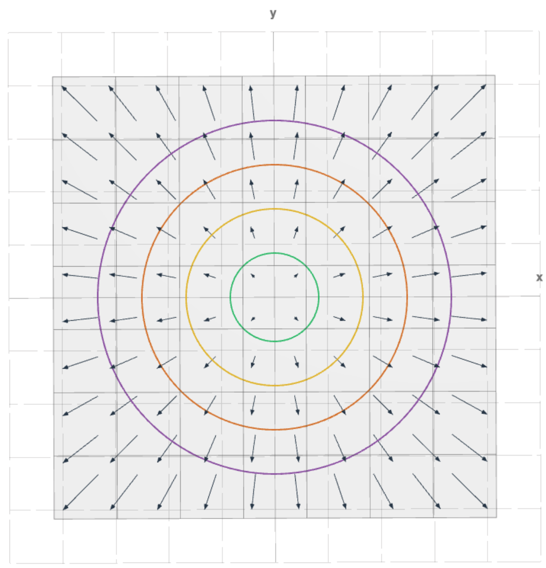
3.4.1 Gradient vector field

The gradient takes as input a (differentiable) function f and it outputs a vector field ∇f. This operation is formalized in the following definition.

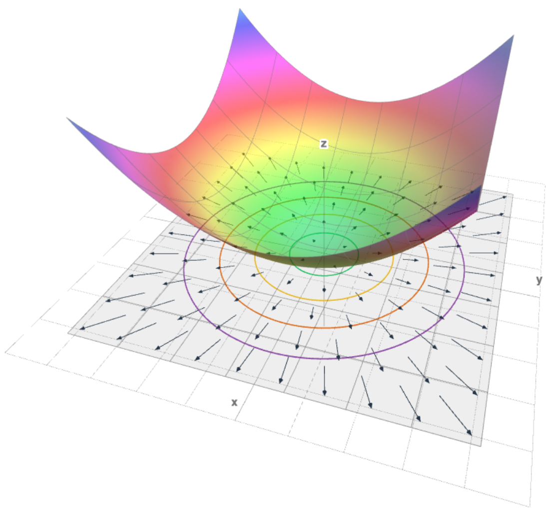
Definition 3.4.6 Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be a real-valued function. Assume all partial derivatives of f exist on all of U. The **gradient of** f (or **gradient vector field of** f) is the function $\nabla f : U \to \mathbb{R}^n$ given by $\nabla f(a) = (\partial_1 f(a), \ldots, \partial_n f(a))$ for all $a \in U$.

To discover the geometric meaning of the gradient, you can plot the gradient vector fields for functions of two variables along with their contours and their graphs.

Example 3.4.7 Let $f(x,y) = x^2 + y^2 + 1$, so $\nabla f(x,y) = (2x, 2y)$. Here is a picture in $\mathbb{R}^2$ of its gradient vector field along with a few level sets. View this Math3D demo for details.



Look at the gradient vectors on a given contour. They seem to point in the direction of the shortest path to the next larger contour. In other words, at a given point a, the gradient $\nabla f(a)$ appears to point in the direction of greatest change. This becomes more apparent when you plot the graph of f in $\mathbb{R}^3$ and embed the gradient vector field below the graph.

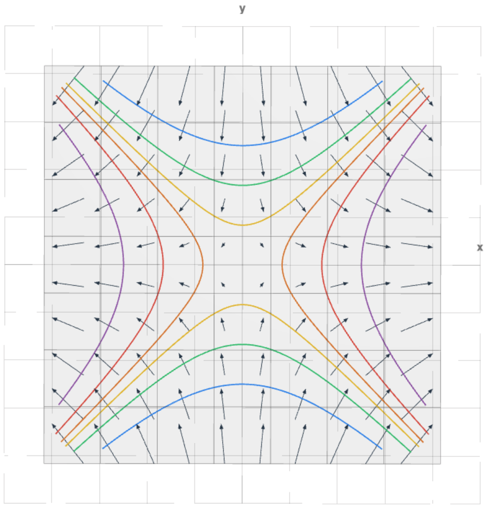


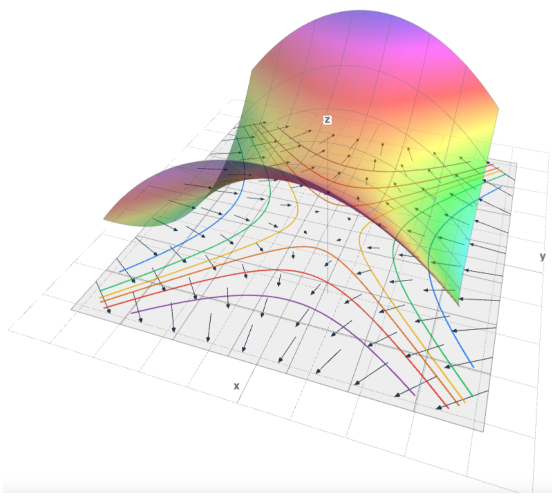
Returning to the 2D plot, notice that the vectors seem forced to cross the level sets "orthogonally". More precisely, at a given point p on the level set, the direction of the gradient $\nabla f(p)$ appears orthogonal to the tangent line of the circle at the point. You can check this with a computation for the level set

$$S = \{(x,y) \in \mathbb{R}^2 : f(x,y) = 3\} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 2\}$$

at the point $p = (1,1) \in S$. By single variable calculus, you can check that the tangent line at $(1,1)$ on this circle S is given by $y = -x + 2$ so $v = (1,-1)$ is a direction vector of the tangent line. On the other hand, $\nabla f(x,y) = (2x, 2y)$ so $\nabla f(p) = (2,2)$ which is orthogonal to $v = (1,-1)$. This confirms your observation!

Example 3.4.8 Let $g(x,y) = x^2 - y^2 + 4$, so $\nabla g(x,y) = (2x, -2y)$. On the left is a plot in $\mathbb{R}^2$ of its gradient vector field with some level sets, and on the right is its graph in $\mathbb{R}^3$ with the lefthand plot embedded in the $z = 0$ plane. View this Math3D demo for details.





You can notice the same features as before. First, the gradient vector field appears to point in the direction of steepest ascent. Second, the gradient vector field appears to "orthogonally" cross the level sets. Imagine you are an ant on top of the graph following the direction of the vector field. You will find yourself walking along the steepest path at every step!

These miraculous observations deserve some formalization.

3.4.2 Direction of steepest ascent

As suggested by these examples, the gradient points in the *direction of steepest ascent*.

Lemma 3.4.9 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. Let a be an interior point of A. Assume f is differentiable at a and $\nabla f(a) \neq 0$. Both of the following hold:

(a) The maximum of $D_u f(a)$ over all unit vectors u occurs when $u = +\frac{\nabla f(a)}{||\nabla f(a)||}$ and the maximum value is $||\nabla f(a)||$.

(b) The minimum of $D_u f(a)$ over all unit vectors u occurs when $u = -\frac{\nabla f(a)}{||\nabla f(a)||}$ and the minimum value is $-||\nabla f(a)||$.

Proof. By Theorem 3.4.3, $D_u f(a) = \nabla f(a) \cdot u$ for any unit vector $u \in S^{n-1}$. From linear algebra,

$$\nabla f(a) \cdot u = ||\nabla f(a)|| \cdot ||u|| \cos\theta = ||\nabla f(a)|| \cos\theta$$

where θ is the angle between $\nabla f(a)$ and u. This quantity is maximized when $\theta = 0$ (i.e. when u points in the same direction as $\nabla f(a)$ so $u = +\frac{\nabla f(a)}{||\nabla f(a)||}$) and minimized when $\theta = -\pi$ (i.e. when u points in the opposite direction as $\nabla f(a)$ so $u = -\frac{\nabla f(a)}{||\nabla f(a)||}$). Therefore, the maximum value is $||\nabla f(a)||$ and the minimum value is $-||\nabla f(a)||$. ■

Remark 3.4.10 The above argument is somewhat informal since the concept of "angle between vectors" is rarely explained in a rigorous way. A fully correct proof uses the Cauchy-Schwarz inequality which states that $|x \cdot y| \le ||x||||y||$ for any $x, y \in \mathbb{R}^n$ and equality holds if and only if x is a scalar multiple of y.

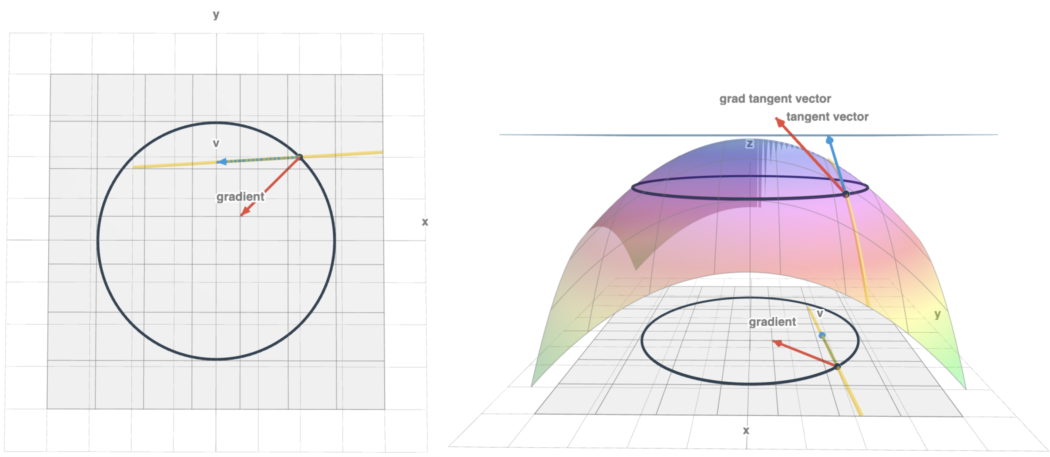
This lemma formalizes two notions.

First, the gradient $\nabla f(a)$ points in the direction of steepest ascent at a.

Second, the norm $||\nabla f(a)||$ is the rate of change of f in this direction.

Both of these fundamental properties are useful in applications. You can observe their geometric truth more carefully in an example.

Example 3.4.11 Consider the function $f(x, y) = 9 - x^2 - y^2$ whose gradient is $\nabla f(x, y) = (-2x, -2y)$. Since $f(1, 1) = 7$, the point $(1, 1)$ belongs to the 7-level set of f. See the lefthand figure below for this plot in $\mathbb{R}^2$ along with the gradient vector $\nabla f(1, 1) = (-2, -2)$.



To see that $\nabla f(1, 1) = (-2, -2)$ points in the direction of steepest ascent, you can compare the change in f in along a different direction $v \in \mathbb{R}^2$. This feature is illustrated by looking at the corresponding "tangent vectors" on the graph in the righthand Math3D demo above. Both vectors v and $\nabla f(1, 1)$ have the same magnitude in $\mathbb{R}^2$, but notice the tangent vector corresponding to $\nabla f(1, 1)$ rises higher compared to the tangent vector corresponding to v. That is precisely because the gradient is the direction of steepest ascent.

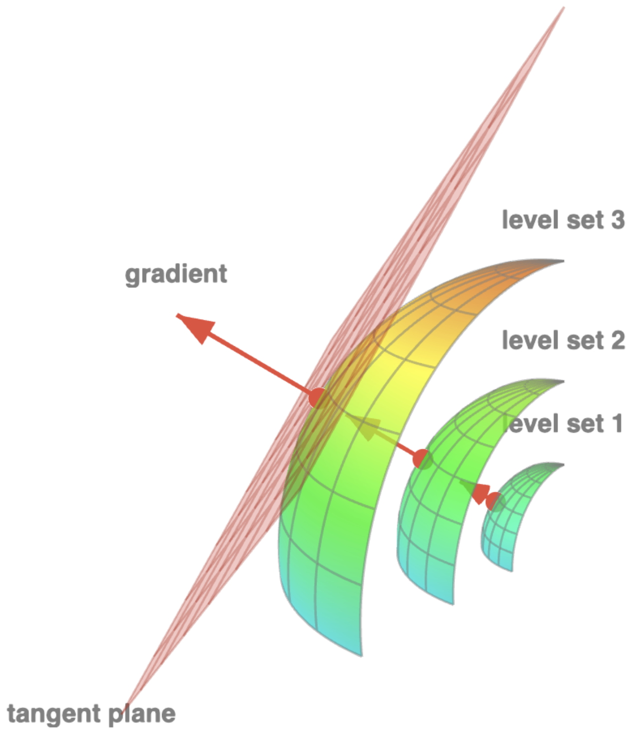
3.4.3 Orthogonality to level sets

Examples 3.4.7 and 3.4.8 both illustrate that the gradient of a real-valued function f appears to cross "orthogonally" to the level sets of f. For $c \in \mathbb{R}$ and $A \subseteq \mathbb{R}^n$, recall that the level set of a scalar-valued function $f : A \to \mathbb{R}$ is given by

$$S = \{x \in A : f(x) = c\} = f^{-1}(\{c\}).$$

Assuming f is continuous, this set is written in implicit form (see Definition 1.5.13). Based on your earlier observations, you may guess that the gradient of f at a point $p \in S$ is "orthogonal" to this set S. You have seen examples of this relationship for real-valued functions of two variables. Here is an example with three variables.

Example 3.4.12 Let $f(x, y, z) = x^2 + y^2 + z^2$. The level sets of f are spheres in $\mathbb{R}^3$. Below is a Math3D demo of some level sets and the gradient at a single point of each level set.



Again, since ∇f points in the direction of steepest ascent of f, the vectors seem forced to cross the level sets orthogonally. More precisely, at a given point p on the level set, the direction of the gradient $\nabla f(p)$ appears orthogonal to the "tangent plane" at that point. But what is the "tangent plane" at a point of a set? Consider the set

$$S = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 3\} = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 3\}$$

at the point $p = (1, 1, 1) \in S$. Notice $\nabla f(x, y, z) = (2x, 2y, 2z)$ so $\nabla f(p) = (2, 2, 2)$. How can you describe the "tangent plane" at p assuming $\nabla f(p)$ is normal to the plane? If (x, y, z) is a point on the plane, then the vector $(x, y, z) - (1, 1, 1)$ lies inside the plane so $\nabla f(p)$ should be orthogonal to it. That is, you may expect

$$\nabla f(p) \cdot (x - 1, y - 1, z - 1) = 0,$$

which should capture the idea of "orthogonally crossing level sets". Using the normal form for a plane in $\mathbb{R}^3$, you might therefore guess that the "tangent plane" of S at p should be

$$\{(x, y, z) \in \mathbb{R}^3 : 2(x - 1) + 2(y - 1) + 2(z - 1) = 0\}.$$

This conclusion is not rigorous at all; it is merely a well-founded conjecture for now.

Good mathematics leads to more questions and this property of the gradient is no exception. These empirical observations in Example 3.4.12 suggest you need to formalize a lot of concepts if you want to establish rigorous mathematical relationships between the gradient and sets in implicit form. You will need to answer many important questions.

> *What does it mean for a vector to be "tangent" to a set? How does this define a "tangent plane" to a set? Why does the gradient seem to define the tangent plane to a set written in implicit form?*

No matter how you choose to define these objects, they better satisfy the key properties of the gradient which you have already observed. Informally, these can be stated as:

> *Every tangent vector v at a point p of a set S written in implicit form is orthogonal to $\nabla f(p)$. Moreover, the gradient $\nabla f(p)$ is orthogonal to the tangent plane of S at p.*

It will take several chapters to build these foundations and properly resolve these questions. Once you have created the necessary tools, this property of the gradient will ultimately be confirmed by an incredibly general theorem, and that theorem will itself follow from one of the most grand theorems in this textbook: the implicit function theorem. Stay tuned!

Exercises for Section 3.4

Concepts and definitions

3.4.1 Let $A \subseteq \mathbb{R}^n$. Assume all the partial derivatives of $f : A \to \mathbb{R}$ exist at every interior point of A. Let a be an interior point of A. Fix $v \in \mathbb{R}^n$.

(a) For each quantity, identify the type of mathematical object that it is.

I) $\nabla f(a)$ II) ∇f III) $\nabla f(a) \cdot v$ IV) $(\nabla f(a))^T v$

(b) If $f : A \to \mathbb{R}$ is differentiable at a, then you have the following identity:

$$D_v f(a) = \nabla f(a) \cdot v = \sum_{j=1}^{n} \partial_j f(a) v_j$$

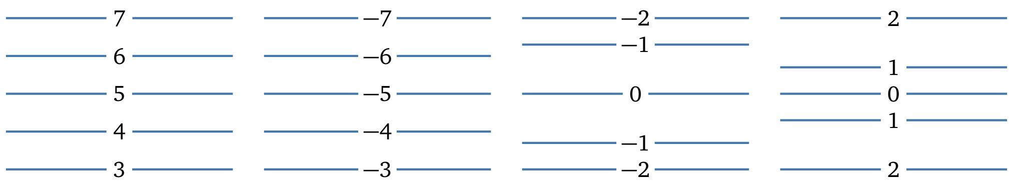
One equality holds by definition. One equality holds by a theorem. Identify which is which.

(c) Fill in the blank with a condition about partial derivatives so that the statement is always true.

The gradient ∇f is defined at a if and only if ________________________

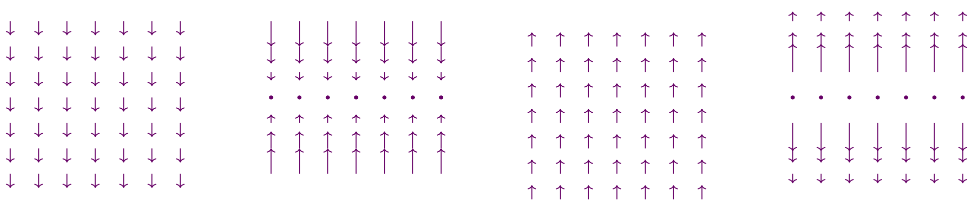
3.4.2 The contour plots of four functions F_1, F_2, F_3, F_4 from $\mathbb{R}^2$ to $\mathbb{R}$ are plotted below.

$z = F_1(x,y)$ $z = F_2(x,y)$ $z = F_3(x,y)$ $z = F_4(x,y)$

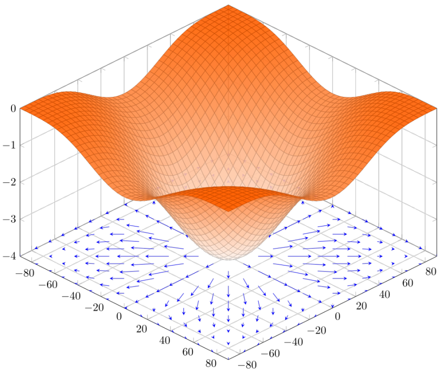


Match the graphs with their corresponding gradient vector field.

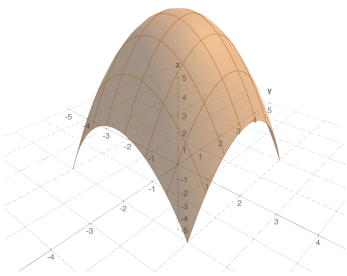
A B C D



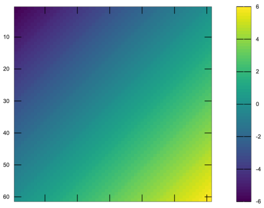
3.4.3 Recall that a real-valued function f in n-variables defines an n-dimensional gradient vector field ∇f. Your geometric intuition will be based on $n = 2$.



$z = f(x, y)$



$z = g(x, y)$



$z = h(x, y)$

For each part, you will use one of the 3 figures above[10].

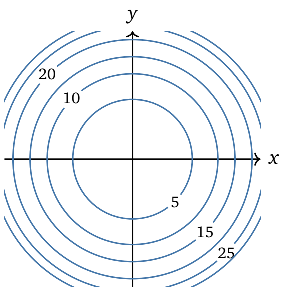
(a) Relate properties of the graph $z = f(x, y)$ from the leftmost figure to properties of gradient vector field $\nabla f(x, y)$. List as many as you can.

(b) Sketch the gradient vector field corresponding to the graph $z = g(x, y)$ from the centre figure.

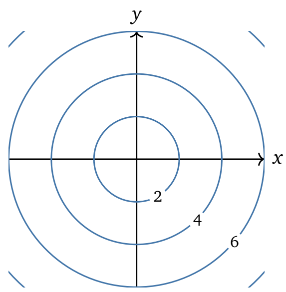
(c) Sketch the gradient vector field correspond to the heat map of $z = h(x, y)$.

3.4.4 For each differentiable map $\mathbb{R}^2 \to \mathbb{R}$, sketch its gradient vector field on the same contour plot.

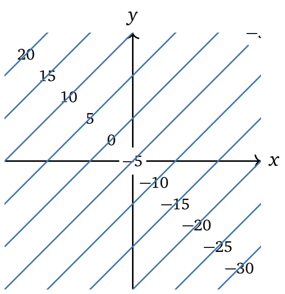
(a) $f : \mathbb{R}^2 \to \mathbb{R}$



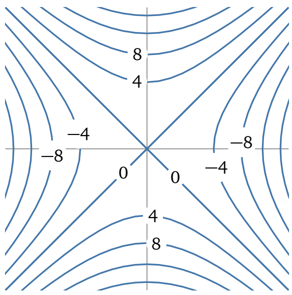
(b) $g : \mathbb{R}^2 \to \mathbb{R}$



(c) $F : \mathbb{R}^2 \to \mathbb{R}$



(d) $G : \mathbb{R}^2 \to \mathbb{R}$



(e) Abbigael and Amy notice something strange for ∇G near the origin $(0, 0)$.

Abbigael says: "*The gradient vector field ∇G near the origin is pointing both towards and away from the origin. That is a contradiction! So $\nabla G(0, 0)$ cannot be defined.*"

Amy replies: "*I think there is no contradiction. In fact, $\nabla G(0, 0)$ has only one possible value!*"

Who is correct? Briefly explain why.

[10]The graph of f is retrieved from Wikimedia Commons on 2024-07-23 licensed under CC0.

Computations

3.4.5 Compute the gradient ∇f for all of the following functions f.

(a) Fix $n \in \mathbb{N}^+$ and $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}$. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(A,B) = \frac{1}{n}\sum_{i=1}^{n}(y_i - Ax_i - B)^2$.

(b) Fix $v \in \mathbb{R}^n$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = x \cdot v$.

(c) Fix $a \in \mathbb{R}^n$. Define $f : \mathbb{R}^n \setminus \{a\} \to \mathbb{R}$ by $f(x) = \frac{1}{||x-a||}$.

3.4.6 Let $a \in \mathbb{R}^n$. For a fixed direction $v \in \mathbb{R}^n$, directional derivatives of functions $f : \mathbb{R}^n \to \mathbb{R}$ are basically the same as single-variable derivatives. They give linear approximations along a single line through a. The gradient is the glue which allows you to calculate directional derivatives in *all directions* from a. It gives a linear approximation in all directions.

(a) Fix $a \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$. Define the single-variable function $F_{a,v} : \mathbb{R} \to \mathbb{R}$ by $F_{a,v}(h) = f(a+hv)$. Express the linear approximation $L_{a,v} : \mathbb{R} \to \mathbb{R}$ for $F_{a,v}$ at 0 using $D_v f(a)$.

(b) Express $L_{a,v} : \mathbb{R} \to \mathbb{R}$ in terms of the gradient. State any assumptions that you use.

(c) Conjecture a linear approximation $L_a : \mathbb{R}^n \to \mathbb{R}$ of f at a.

(d) Let $f(x,y) = xe^{xy} + y$. Compute the conjectured linear approximation $L : \mathbb{R}^2 \to \mathbb{R}$ of f at $(0,0)$. Then estimate $f(0.1, 0.2)$ with L.

3.4.7 Let $C \subseteq \mathbb{R}^2$ be the set of points defined by the equation $y^2 + 2xy = x^3 - 4x + 3$.

(a) Express C as the level set of a function $f : \mathbb{R}^2 \to \mathbb{R}$. Use set builder notation.

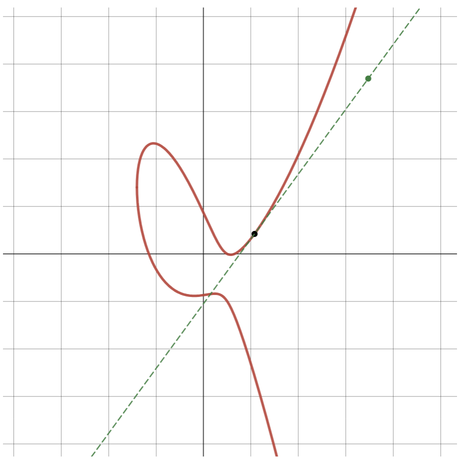
(b) Fix a point $(a,b) \in C$. Conjecturally, the gradient $\nabla f(a,b)$ should be *orthogonal* to the "tangent line" of C at (a,b). This suggests that the "tangent line" of C at (a,b) should be the line

$$L_{a,b} = \{(x,y) \in \mathbb{R}^2 : \nabla f(a,b) \cdot (x-a, y-b) = 0\}$$

This educated guess is a bit mysterious. To remember this expression, you will illustrate the key idea. For the diagram below:

- Label the set C, the line $L_{a,b}$, the point $(a,b) \in C$, and an arbitrary point $(x,y) \in L_{a,b}$.
- Draw a possible gradient vector $\nabla f(a,b)$ and label it.
- Draw the vector $(x-a, y-b)$ and label it.

Once you finish, reflect on how your diagram explains your educated guess.



(c) Compute the gradient $\nabla f(a,b)$ and express the line $L_{a,b}$ using this calculation.

3.4.8 Recall one of your key discoveries with the gradient.

The gradient is orthogonal to the level sets of a real-valued function.

Building off the previous question, you will heuristically explore the notion of a "tangent plane" in higher dimensions. This will not be rigorous, but it will build your spatial intuition.

(a) Let $S \subseteq \mathbb{R}^3$ be the 237-level set of a function $f : \mathbb{R}^3 \to \mathbb{R}$ so

$$S = \{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 237\}$$

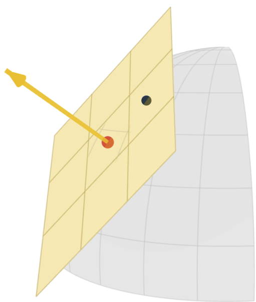
Fix $(a, b, c) \in S$. Your discovery suggests that the "tangent plane" of S at (a, b, c) should be given by

$$P_{a,b,c} = \{(x, y, z) \in \mathbb{R}^3 : \nabla f(a, b, c) \cdot (x - a, y - b, z - c) = 0\}.$$

For the Math3D figure below:

- Label the set S, the plane $P_{a,b,c}$, the point $(a, b, c) \in S$, and a point $(x, y, z) \in P_{a,b,c}$.
- Label the gradient vector $\nabla f(a, b, c)$.
- Draw the vector $(x - a, y - b, z - c)$ and label it.

Once you finish, reflect on how your diagram explains your educated guess.



(b) Let $S \subseteq \mathbb{R}^n$ be the 237-level set of a function $f : \mathbb{R}^n \to \mathbb{R}$. Fix a point $a \in S$. Let $P_a \subseteq \mathbb{R}^n$ be the "tangent hyperplane" of S at a. Guess an expression for the set P_a using the gradient. *Hint:* Redraw and relabel the above diagram with your new notation to guide your intuition.

3.4.9 Let $g(x, y, z) = 2x^2 + 3y^2 + 7xyz$. Guess the equation of the tangent plane to the surface $g(x, y, z) = -9$ at the point $(1, -1, 2)$.

Proofs

3.4.10 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. Let a be an interior point of A. The textbook provides a short proof of the key fact:

The gradient $\nabla f(a)$ points in the direction of steepest ascent on the graph of f at a.

This idea could be more precisely stated than Lemma 3.4.9.

(a) A **direction** is a unit vector in $\mathbb{R}^n$. In other words, it is an element of the $(n-1)$-dimensional unit sphere

$$S^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = 1\}.$$

Define a function $F_a : S^{n-1} \to \mathbb{R}$ which measures the steepness of the ascent of f in any given direction. State the *weakest* assumptions on f which ensures F_a is defined.

(b) Formulate a precise theorem statement using F_a which replaces Lemma 3.4.9.

3.4.11 Differentiation rules for the gradient are also straightforward since they are directly imported from partial derivatives with a bit of linear algebra. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Assume both ∇f and ∇g exist at $a \in \mathbb{R}^n$. Prove that $\nabla(fg)$ exists at a and

$$\nabla(fg)(a) = g(a)\nabla f(a) + f(a)\nabla g(a).$$

Applications and beyond

3.4.12 You will form a deeper understanding of the gradient by interpreting it in applications.

(a) For a given moment in time, let $T(\varphi, \theta)$ be the temperature in degrees Celsius at the point on the Earth with latitude $\varphi \in [-90, 90]$ and longitude $\theta \in [-180, 180]$ measured in degrees. Negative values of φ are south of the equator and negative values of θ are west of Greenwich. Explain what the temperature gradient $\nabla T(43.6, -79.3)$ represents in plain language.

(b) At the given moment in time, it is 9 am in Toronto. What can you likely say about $\nabla T(43.6, -79.3)$?

(c) TorontoQuills, a manufacturer of exquisite fountain pens, is able to produce $P(x, y)$ pens each week if they pay their employees a total of x thousand dollars and they use their pen production machines for a total of y thousand dollars. They currently spent \$40,000 on employees and \$8,000 on machine maintenance per week. Use plain language to explain what $\dfrac{\partial P}{\partial x}(40, 8) = 35$ and $\dfrac{\partial P}{\partial y}(40, 8) = 15$ means.

(d) If TorontoQuills has an extra \$2000 per week to spend on production and $\nabla P(40, 8) = (35, 15)$, how would you advise them to maximize their production? Give two different estimates on the number of additional pens they would produce based on your advice.

3.4.13 You are a data scientist who is continuing to build a machine learning algorithm (see C3 worksheet). At each step of the algorithm, you need to choose n variables $x_1, \ldots, x_n$ which minimize a certain error function $E : \mathbb{R}^n \to [0, \infty)$. Equipped with your knowledge of the gradient, you attempt to make progress on numerically solving this problem.

(a) You have made an initial guess for a point $p_0 \in \mathbb{R}^n$ which **minimizes** E. You want to improve your guess by choosing a direction to move from p_0 towards a new point. How should you choose this direction? Explain your answer in terms of the gradient.

(b) Once you have chosen a direction, you want to move from p_0 in that direction by a distance $h_0 > 0$ to a new point $p_1 \in \mathbb{R}^n$. Give a formula for p_1 in terms of p_0, h_0, and E.

(c) By repeating this process with a sequence of distances $\{h_k\}_k$ satisfying $h_k \to 0^+$, define a recursive sequence of approximations $\{p_k\}_k \subseteq \mathbb{R}^n$.

(d) Let $p \in \mathbb{R}^n$ and fix $h > 0$. Define $h_k = h||\nabla E(p_k)||$ for $k \geq 1$. Show that if $\{p_k\}_k$ converges to p and ∇E is continuous on $\mathbb{R}^n$ then $\nabla E(p) = 0$. *Hint:* Take limits in your recursion formula.

3.4.14 Let M be the mass of the Earth centred at the origin in $\mathbb{R}^3$. Let m be the mass of the moon. Newton's law of gravitation states that if $x \in \mathbb{R}^3$ is the moon's position relative to the Earth, then the gravitational pull from the Earth on the moon is given by

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

where G is the gravitational constant. Show that $F = \nabla\varphi$ where $\varphi(x) = \dfrac{GMm}{||x||}$.

3.5. Differentials and Jacobians

Generalizing the derivative to any map poses quite a few issues in higher dimensions. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a map and fix a point $a \in \mathbb{R}^n$. You might first guess:

$$\textit{Should } \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \textit{ or } \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \textit{ be the derivative of } f \textit{ at } a?$$

This idea stems from single variable calculus and Section 3.1, but there is an issue in higher dimensions. The quantities $x-a$ and h are vectors in $\mathbb{R}^n$. You cannot divide by a vector, so these expressions are nonsensical. Instead, you can take the norm of these quantities, and ask:

$$\textit{Then should } \lim_{x\to a} \frac{f(x)-f(a)}{||x-a||} \textit{ or } \lim_{h\to 0} \frac{f(a+h)-f(a)}{||h||} \textit{ be the derivative of } f \textit{ at } a?$$

These limits at least make some sense now. The limit variable h is a vector in $\mathbb{R}^n$ and the expression $\frac{1}{||h||}(f(a+h)-f(a))$ is a vector in $\mathbb{R}^m$. But this creates a different issue. The limit will rarely exist! This fails even for a single variable calculus example. For instance, if $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x$ and $a = 0$ then these limits are equal to $\lim_{h\to 0}(h/|h|)$ which does not exist. If this attempted definition fails for the simplest examples, then it has little hope in higher dimensions.

These outcomes are rather frustrating, so you may be exasperated and ask:

Can the map $v \mapsto D_v f(a)$ be the derivative of f at a? Is this good enough?

This seems reasonable since, after all, directional derivatives capture information about the change in f in every direction. However, Examples 3.3.8 and 3.3.9 demonstrate that these directional derivatives do not have any natural relationship with each other, even if you assume they all exist. After so many failed attempts, this may seem disheartening, but these failures are productive! Your explorations with functions of one variable, partial derivatives, and directional derivatives have illustrated great depth so, if you create a definition that addresses all of these issues, then you will be handsomely rewarded with many applications and a deeper insight.

3.5.1 Definitions

The key breakthrough is in Section 3.1, namely Theorem 3.1.9. Here is a simplified version.

Theorem. *A map $\gamma : \mathbb{R} \to \mathbb{R}^m$ is differentiable at the real number $a \in \mathbb{R}$ if and only if there exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that*

$$\lim_{h\to 0} \frac{\gamma(a+h)-\gamma(a)-L(h)}{h} = 0.$$

The *non-linear* map $\gamma(a+h)-\gamma(a)$ is approximated by the *linear* map $L(h)$ and, most crucially, the error $\gamma(a+h)-\gamma(a)-L(h)$ tends to the zero vector *faster* than h tends to the zero scalar. This beautiful mix of linear algebra and calculus generalizes almost seamlessly.

Definition 3.5.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. A function f is **differentiable at** a if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{f(a+h)-f(a)-L(h)}{||h||} = 0$$

The linear map L is the **differential of** f **at** a, denoted df_a.

Remark 3.5.2 Other sources refer to df_a as the "total derivative" of f at a.

Informally speaking, the viewpoints of algebra and analysis are blended to capture the same idea as before. The *non-linear* map $f(a+h)-f(a)$ is approximated by the *linear* map $L(h)$ and the error $f(a+h)-f(a)-L(h)$ tends to the zero vector *faster* than $||h||$ tends to the zero scalar. As a first example, you can illustrate this idea with a map $\mathbb{R}^2 \to \mathbb{R}^2$.

Example 3.5.3 Take the point $a = (1.5, 0.5)$ and define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$f(x,y) = \Big(\frac{x^2+y^2}{2}, xy\Big).$$

As you shall soon see, f is differentiable at a and the differential of f at a is given by $df_a : \mathbb{R}^2 \to \mathbb{R}^2$ where

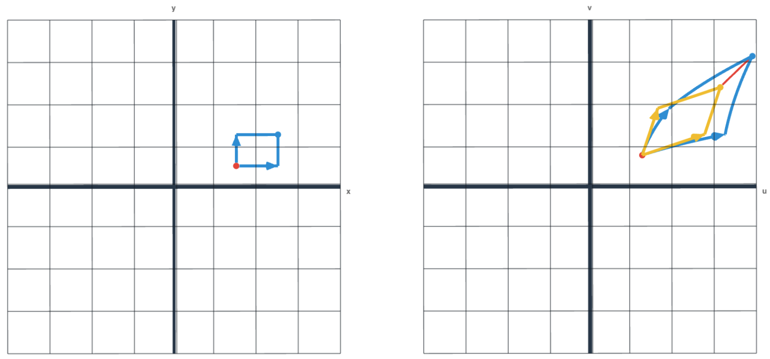
$$\forall h \in \mathbb{R}^2, \quad df_a(h) = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} h. \tag{3.5.1}$$

Thus, $df_a(h)$ approximates $f(a+h)-f(a)$ with error tending to zero faster than $||h||$ as $h \to 0$. What does this mean from a geometric perspective?

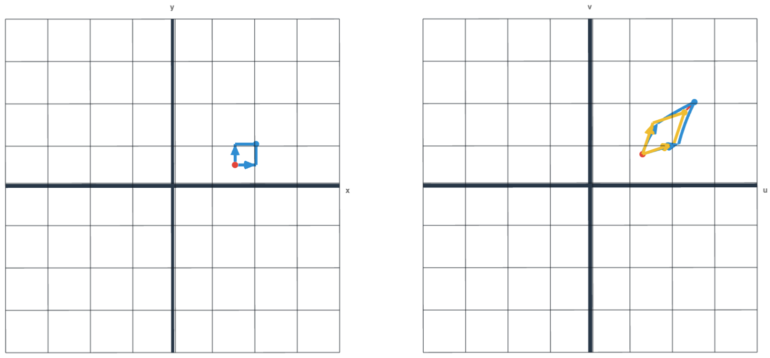
Write $h = (\Delta x, \Delta y)$ where Δx and Δy are small positive real numbers. The two points

$$a = (1.5, 0.5) \qquad \text{and} \qquad a+h = (1.5+\Delta x, 0.5+\Delta y)$$

form opposite corners of a rectangle. You can plot how this rectangle transforms under the non-linear map f compared to how it transforms under the linear map L.



The lefthand grid has a (in red) and $a+h$ on opposite corners of a rectangle. On the right, the blue bent square is the image under the non-linear map f whereas the yellow parallelogram is the image under the linear map L translated by $f(a)$. The top right corners are $f(a+h)$ and $f(a)+L(h)$ respectively. The above diagram illustrates the choice $\Delta x = 1$ and $\Delta y = 0.75$. If you shrink those values to say $\Delta x = \Delta y = 0.5$, the approximation visibly improves.



Play with this Math3D demo to watch this phenomenon; you can also vary the point a to see how the differential df_a changes.

This gives some intuition for what the differential df_a represents geometrically, but how do you prove it exists and also compute it? The goal is to choose a linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ such that $L(h)$ approximates $f(a+h) - f(a)$ with error tending to zero faster than $||h||$. Right now, you do not have a systematic method so you can try an ad hoc approach. Write $h = (\Delta x, \Delta y)$ where Δx and Δy are small positive real numbers. Notice that

$$\begin{aligned} f(a+h) - f(a) &= f(1.5+\Delta x, 0.5+\Delta y) - f(1.5, 0.5) \\ &= \begin{bmatrix} 1.5\Delta x + 0.5\Delta y + 0.5(\Delta x)^2 + 0.5(\Delta y)^2 \\ 0.5\Delta x + 1.5\Delta y + \Delta x \Delta y \end{bmatrix}. \end{aligned}$$

You can separate the linear terms (i.e. scalar multiples of Δx and Δy) from the quadratic terms (i.e. scalar multiples of $(\Delta x)^2, (\Delta y)^2$ or $\Delta x \Delta y$). As $h = (\Delta x, \Delta y) \to (0,0)$, the quadratic terms tend to zero faster than the linear terms. More formally, you can prove that

$$\begin{aligned} &\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{(\Delta x)^2}{||(\Delta x, \Delta y)||} = 0, \qquad \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{(\Delta y)^2}{||(\Delta x, \Delta y)||} = 0, \\ &\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{\Delta x \Delta y}{||(\Delta x, \Delta y)||} = 0. \end{aligned} \tag{3.5.2}$$

By separating these terms, you can see that

$$f(a+h) - f(a) = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} 0.5(\Delta x)^2 + 0.5(\Delta y)^2 \\ \Delta x \Delta y \end{bmatrix}.$$

A matrix with constant coefficients has appeared! This suggests defining $L : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$L(h) = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} h \quad \text{for all } h \in \mathbb{R}^2.$$

which, by no coincidence, is the same as the linear map in (3.5.1). From the calculations above, it follows by standard limit laws and (3.5.2) that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - L(h)}{||h||} = \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{1}{||(\Delta x, \Delta y)||} \begin{bmatrix} 0.5(\Delta x)^2 + 0.5(\Delta y)^2 \\ \Delta x \Delta y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This proves that L satisfies Definition 3.5.1, so f is differentiable at a and the differential df_a exists and equals L. Phew!

Note Definition 3.5.1 refers to the linear map as *the* differential of f at a. This statement implicitly claims that if such a linear map exists, then it is unique. This requires proof.

Lemma 3.5.4 (Uniqueness of differential) Let $A \subseteq \mathbb{R}^n$ and let a be an interior point of A. Let $f : A \to \mathbb{R}^m$. If there exists linear maps $L_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $L_2 : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - L_1(h)}{||h||} = 0 = \lim_{h\to 0} \frac{f(a+h) - f(a) - L_2(h)}{||h||}$$

then $L_1 = L_2$. In other words, the differential of f at a, if it exists, is unique.

Proof. By taking the difference of the two limits, it follows that

$$0 = \lim_{h\to 0} \frac{L_2(h) - L_1(h)}{||h||} = \lim_{h\to 0} \frac{T(h)}{||h||} \tag{3.5.3}$$

where $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by $T(v) = L_2(v) - L_1(v)$ for $v \in \mathbb{R}^n$. Since T is the difference of two linear maps, it is linear and so $T(0) = 0$. It therefore suffices to prove that $T(v) = 0$ for all non-zero vectors $v \in \mathbb{R}^n$. Fix $v \in \mathbb{R}^n \setminus \{0\}$. Define the sequence $\{x(k)\}_{k=1}^{\infty}$ by $x(k) = \frac{1}{k} v \in \mathbb{R}^n$ for $k \in \mathbb{N}^+$ so $||x(k)|| = \frac{1}{k}||v||$ and $x(k) \neq 0$. This implies that $||x(k)|| \to 0$ and hence $x(k) \to 0$ as $k \to \infty$. By the sequential definition of the limit, (3.5.3), and linearity of T, it follows that

$$0 = \lim_{k\to\infty} \frac{T(x(k))}{||x(k)||} = \lim_{k\to\infty} \frac{T(\frac{1}{k}v)}{\frac{1}{k}||v||} = \lim_{k\to\infty} \frac{\frac{1}{k}T(v)}{\frac{1}{k}||v||} = \frac{T(v)}{||v||}.$$

The last step uses a limit law for constant sequences. Thus, $T(v) = 0$ for $v \in \mathbb{R}^n$ as required. ∎

You will also need some terminology to discuss differntiability on subsets of a domain.

Definition 3.5.5 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. Let S be a subset of the interior A°. The function f is **differentiable on** S if f is differentiable at a for every $a \in S$. The function f is **differentiable** if f is differentiable on its domain A.

Remark 3.5.6 The set S is a subset of the interior of the domain, because a function can only be differentiable at interior points. This means the phrase "f is differentiable" only makes sense if its domain A is open.

This definition has some minor linguisitic subtleties.

Example 3.5.7 Define $A := (0, \infty)$ and $B := [0, \infty)$. Define $f : A \to \mathbb{R}$ by $f(x) = x^2$ and $g : B \to \mathbb{R}$ by $g(x) = x^2$. Both f and g are differentiable on $(0, \infty) = A^\circ = B^\circ$. Thus, the function f is differentiable, but g is not. Notice g cannot be differentiable at 0 because 0 is not an interior point of its domain B.

Also, you can assert differentiability on non-open sets. For example, since $[1, 5) \subseteq (0, \infty) = A^\circ = B^\circ$, the maps f and g are both differentiable on the non-open set $[1, 5)$.

This summarizes the core definitions for differentiability.

3.5.2 Properties

Example 3.5.3 indicates that calculating differentials will be an onerous task without more tools. At the very least, you can directly verify two special cases.

Example 3.5.8 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. For any $a \in \mathbb{R}^n$, you can prove that f is differentiable at a and that $df_a = f$.

Example 3.5.9 Let $\gamma : \mathbb{R} \to \mathbb{R}^m$ be a parametric curve. You can check that the definition of differentiability in Theorem 3.1.9 is equivalent to Definition 3.5.1.

Aside from these special cases, you will need more tools to compute differentials. Your first tool is the usual one. You can calculate differentials component-by-component.

Lemma 3.5.10 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. Then f is differentiable at a if and only if each of its component functions $f^1, f^2, \ldots, f^m$ is. If so,

$$df_a = (df_a^1, \ldots, df_a^m)$$

Proof. This is left as an exercise. Use Theorem 2.5.11 and Definition 3.5.1. ■

A second tool is a familiar fact from single variable calculus.

Lemma 3.5.11 Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is differentiable at a, then f is continuous at a.

Proof. This is left as an exercise. Explain why it suffices to show $\lim_{x \to a}[F(x) - F(a)] = 0$. To introduce the definition of differentiability, add zero and then multiply by one. ■

A third tool is a consequence of linearity because differentials are linear maps.

Lemma 3.5.12 Let $A \subseteq \mathbb{R}^n$ and let a be an interior point of A. Fix $\lambda \in \mathbb{R}$. If $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$ are differentiable at a, then the function $f + \lambda g$ is differentiable at a and

$$d(f + \lambda g)_a = df_a + \lambda dg_a.$$

Proof. This is left as an exercise. It follows quickly from the definition. ■

The fourth tool is a triumph. Differentiability implies the existence of directional derivatives!

Theorem 3.5.13 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is differentiable at a, then for all $v \in \mathbb{R}^n$, the directional derivative $D_v f(a)$ exists and

$$df_a(v) = D_v f(a).$$

Proof. Assume f is differentiable at a. Fix $v \in \mathbb{R}^n$. By linearity, $df_a(tv) = t df_a(v)$ for $t \in \mathbb{R}$. Thus,

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = \lim_{t \to 0} \left[\frac{f(a + tv) - f(a) - df_a(tv)}{t} + df_a(v) \right].$$

By the addition limit law (Theorem 2.5.14), it suffices to show that

$$\lim_{t \to 0} \frac{f(a + tv) - f(a) - df_a(tv)}{t} = 0.$$

For $t \in \mathbb{R}$, the identity $||tv|| = |t| \cdot ||v||$ implies that

$$\lim_{t \to 0} \frac{f(a + tv) - f(a) - df_a(tv)}{t} = \lim_{t \to 0} \left[\left(\frac{f(a + tv) - f(a) - df_a(tv)}{||tv||} \right) \cdot \frac{|t|}{t} \cdot ||v|| \right].$$

Since the quantity $\frac{|t|}{t} \cdot ||v||$ is bounded between $-||v||$ and $||v||$ as $t \to 0$, the result follows from the multivariable squeeze theorem (Theorem 2.5.15) provided

$$\lim_{t \to 0} \frac{f(a + tv) - f(a) - df_a(tv)}{||tv||} = 0. \tag{3.5.4}$$

Define $F : A \to \mathbb{R}^m$ by $F(a) = 0$ and, for $x \neq a$,

$$F(x) = \frac{f(x) - f(a) - df_a(x-a)}{||x-a||}.$$

Since f is differentiable at a, it follows that F is continuous at 0. As a is an interior point of A, there exists $\delta > 0$ such that $\{a + tv : t \in (-\delta, \delta)\} \subseteq A$. Define $\gamma : (-\delta, \delta) \to \mathbb{R}^m$ by $\gamma(t) = a + tv$ so γ is continuous and $\operatorname{im}(\gamma) \subseteq A$. Since the composition of continuous functions is continuous (Theorem 2.6.21), we have that

$$\lim_{t\to 0} \frac{f(a+tv) - f(a) - df_a(tv)}{||tv||} = \lim_{t\to 0} F(\gamma(t)) = F(\gamma(0)) = F(a) = 0.$$

as required by (3.5.4). ■

While Theorem 3.5.13 is a fantastic result, its converse is false.

Example 3.5.14 Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(0,0) = 0$ and $f(x,y) = \dfrac{2x^2y}{x^4 + y^2}$ for $(x,y) \neq (0,0)$. Let $v = (v_1, v_2) \in \mathbb{R}^2$. You can check that $D_{(v_1,0)} f(0,0) = 0$ and if $v_2 \neq 0$, then

$$\begin{aligned} D_v f(0,0) = \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} &= \lim_{t\to 0} \frac{1}{t} \frac{2v_1^2 v_2 t^3}{v_1^4 t^4 + v_2^2 t^2} \\ &= \lim_{t\to 0} \frac{2v_1^2 v_2}{v_1^4 t^2 + v_2^2} = \frac{2v_1^2}{v_2}, \end{aligned}$$

so every directional derivative exists. However, you can prove that f is not continuous at $(0,0)$ by considering the sequence $(x_k, y_k) = (1/k, 1/k^2)$ as $k \to \infty$. From Lemma 3.5.11, this implies f is not differentiable at $(0,0)$.

3.5.3 Matrix of the differential

For a map $f : \mathbb{R}^n \to \mathbb{R}^m$ and a point $a \in \mathbb{R}^n$, calculating the differential df_a is equivalent to finding the matrix defining this linear map. As you may recall, the matrix of a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined using the standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$. That is, the matrix of T is precisely the $m \times n$ matrix

$$\begin{bmatrix} | & & | \\ T(e_1) & \cdots & T(e_n) \\ | & & | \end{bmatrix} \tag{3.5.5}$$

Similar to the discussion with (3.3.1), each partial derivative $\partial_j f(a) \in \mathbb{R}^m$ mirrors each column vector $T(e_j) \in \mathbb{R}^m$. This suggests you to introduce a matrix of partial derivatives.

Definition 3.5.15 Let $A \subseteq \mathbb{R}^n$ and let $f = (f_1, \dots, f_m) : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. Assume all the partials of f exist at a. The **Jacobian of f at a** is the $m \times n$ matrix $Df(a)$ given by

$$Df(a) = \left[\partial_j f_i(a)\right]_{i,j} = \begin{bmatrix} | & & | \\ \partial_1 f(a) & \cdots & \partial_n f(a) \\ | & & | \end{bmatrix} = \begin{bmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_n f_m(a) \end{bmatrix}$$

Remark 3.5.16 The Jacobian $Df(a)$ is also called the **Jacobian matrix**. The notations

$$f'(a) \qquad Jf(a) \qquad J_f(a) \qquad \mathrm{Jac}_f(a).$$

are common equivalent conventions, but none of these will be used here. Some textbooks also use the phrase "derivative of f" or "total derivative of f" instead of Jacobian.

Since Jacobians are defined with partial derivatives, they are quick to compute.

Example 3.5.17 Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x,y) = \left(\frac{1}{2}(x^2+y^2), xy\right)$ as in Example 3.5.3. Since $\partial_1 f(x,y) = (x,y)$ and $\partial_2 f(x,y) = (y,x)$, it follows that if $a = (1.5, 0.5)$ then

$$Df(a) = \begin{bmatrix} | & | \\ \partial_1 f(a) & \partial_2 f(a) \\ | & | \end{bmatrix} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}.$$

It is no coincidence that this is the same as the matrix of the differential in (3.5.1).

Example 3.5.18 Define the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x,y,z) = (x^2y^2z, y+\sin z)$. To find its Jacobian at $(1,2,0)$, you can begin by computing all of its partials.

$$\begin{aligned} \partial_1 f_1(x,y,z) &= \frac{\partial}{\partial x}(x^2y^2z) = 2xy^2z & \qquad \partial_1 f_2(x,y,z) &= \frac{\partial}{\partial x}(y+\sin z) = 0 \\ \partial_2 f_1(x,y,z) &= \frac{\partial}{\partial y}(x^2y^2z) = 2x^2yz & \qquad \partial_2 f_2(x,y,z) &= \frac{\partial}{\partial y}(y+\sin z) = 1 \\ \partial_3 f_1(x,y,z) &= \frac{\partial}{\partial z}(x^2y^2z) = x^2y^2 & \qquad \partial_3 f_2(x,y,z) &= \frac{\partial}{\partial z}(y+\sin z) = \cos z \end{aligned}$$

At any $(x,y,z) \in \mathbb{R}^3$, this implies that

$$Df(x,y,z) = \begin{bmatrix} 2xy^2z & 2x^2yz & x^2y^2 \\ 0 & 1 & \cos z \end{bmatrix} \quad \text{so} \quad Df(1,2,0) = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 1 & 1 \end{bmatrix}.$$

Example 3.5.19 Let $\gamma : I \to \mathbb{R}^m$ be a parametric curve with domain $I \subseteq \mathbb{R}$. Since γ is a one-variable function, the partial derivative $\partial_1 \gamma$ is the same as the derivative γ' in Section 3.1. Thus, if γ is differentiable at $a \in I$, the Jacobian of γ at a is

$$D\gamma(a) = \begin{bmatrix} | \\ \partial_1 \gamma(a) \\ | \end{bmatrix} = \begin{bmatrix} \gamma_1'(a) \\ \vdots \\ \gamma_m'(a) \end{bmatrix} = \gamma'(a).$$

Example 3.5.20 Let $f : A \to \mathbb{R}$ be a real-valued function with $A \subseteq \mathbb{R}^n$. Let a be an interior point of A. If all the partials of f exist at a then the Jacobian of f at a is the $1 \times n$ matrix

$$Df(a) = \begin{bmatrix} \partial_1 f(a) & \cdots & \partial_n f(a) \end{bmatrix} = \nabla f(a)^T.$$

In other words, the Jacobian is the transpose of the gradient.

Example 3.5.21 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, so $f(x) = Ax$ for some $m \times n$ matrix A. For any $p \in \mathbb{R}^n$, you can verify by direct computation that $Df(p) = A$.

As Example 3.5.17 suggests, the Jacobian and the differential are intimately connected.

Theorem 3.5.22 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is differentiable at a then the matrix of the differential of f at a is the $m \times n$ Jacobian matrix of f at a. In other words,

$$\forall v \in \mathbb{R}^n, \quad df_a(v) = Df(a)v.$$

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. By Theorem 3.5.13 and Definition 3.5.15,

$$\forall j \in \{1, \ldots, n\}, \quad df_a(e_j) = D_{e_j} f(a) = \partial_j f(a) = Df(a)e_j.$$

The result now follows from linear algebra, namely (3.5.5). ■

Combined with Theorem 3.5.13, this theorem implies the ultimate chain of relationships.

If $f : A \to \mathbb{R}^m$ is differentiable at a and $A \subseteq \mathbb{R}^n$, then

$$df_a(v) = Df(a)v = D_v f(a) = \sum_{j=1}^{n} v_j \partial_j f(a). \tag{3.5.6}$$

This connects differentials, Jacobians, directional derivatives, and partial derivatives!

Exercise 3.5.23 Prove Theorem 3.3.10 using Theorems 3.5.13 and 3.5.22.

Assuming differentiability, you can now easily compute differentials using the Jacobian.

Example 3.5.24 From Example 3.5.18, if you assume that $f(x, y, z) = (x^2 y^2 z, y + \sin z)$ is differentiable at $(1, 2, 0)$ then by Theorem 3.5.22,

$$\forall v \in \mathbb{R}^3, \quad df_{(1,2,0)}(v) = Df(1, 2, 0)v = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 1 & 1 \end{bmatrix} v$$

This section was a great leap forward in the theory of differential calculus. You have finally defined differentiability for any map $\mathbb{R}^n \to \mathbb{R}^m$ and this definition connects to all prior derivatives, but one key hurdle remains.

How do you efficiently verify whether a function is differentiable?

Previous examples highlight some concerns. First, Example 3.5.3 demonstrates that your only existing approach is ad hoc and rather complicated. Second, Example 3.5.14 shows that the existence of all directional derivatives is not enough to guarantee differentiability. These obstacles suggest that you will need to establish a simple yet effective criterion that implies differentiability. Luckily, as you shall soon discover, there is a rather elegant sufficient condition.

Exercises for Section 3.5

Concepts and definitions

3.5.1 Let $A \subseteq \mathbb{R}^n$ and let a be an interior point of A. Let $F : A \to \mathbb{R}^m$ be differentiable at a. Fix $v \in \mathbb{R}^n$.

(a) For each quantity, identify the type of mathematical object and (if possible) its name.

i) dF_a
ii) $dF_a(v)$
iii) $DF(a)$
iv) $DF(a)v$
v) $(dF_a^1, \ldots, dF_a^m)$
vi) $\begin{bmatrix} \partial_1 F(a) & \cdots & \partial_n F(a) \end{bmatrix}$
vii) $\partial_j F^i(a)$ with fixed $i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}$
viii) $\left[\partial_j F^i(a)\right]_{i,j}$

(b) You have the following identity

$$D_v F(a) = dF_a(v) = DF(a)v = \sum_{j=1}^{n} \partial_j F(a) v_j = \sum_{j=1}^{n} \begin{bmatrix} \partial_j F^1(a) \\ \vdots \\ \partial_j F^m(a) \end{bmatrix} v_j$$

Two equalities follow by theorems. One equality follows by a lemma. One equality follows by definition. Identify which is which.

3.5.2 Each relationship listed below is either true or nonsense. Determine which is which. If true, decide whether it follows from some theorems/properties or it is simply a definition. Let $A \subseteq \mathbb{R}^n$ and let a be an interior point of A. Assume $F : A \to \mathbb{R}^m$ is differentiable at a. Fix $v \in \mathbb{R}^n$.

(a) $dF_a = DF(a)$
(b) $dF_a(v+1) = dF_a(v) + dF_a(1)$
(c) $dF_a^3(v) = (dF_a(v))^3$
(d) $dF_a(v) = DF(a)v$
(e) $DF(a) = \left[\partial_j F^i(a)\right]_{i,j}$
(f) $DF(a) = \left[\partial_i F^j(a)\right]_{i,j}$
(g) $DF_a(h) = \sum_{j=1}^{n} \partial_j F^i(a) v_j$

3.5.3 Determine which statements are true or false. If the statement is true, briefly justify by citing a theorem. If false, cite a counterexample or provide your own. Fix $a \in \mathbb{R}^n$. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a map. Let $\gamma : \mathbb{R} \to \mathbb{R}^m$ be a parametric curve. Let $f : \mathbb{R}^n \to \mathbb{R}$ be real-valued.

(a) If F is differentiable at a, then F is continuous at a.
(b) If F is differentiable at a, then all of the partials of F exist at a.
(c) If F is differentiable at a, then all of the directional derivatives of F exist at a.
(d) If F is differentiable at a, then the Jacobian of F at a exists.
(e) If F is differentiable at a, then the matrix of the differential of F at a is the Jacobian of F at a.
(f) If γ is differentiable at a, then $d\gamma_a = \gamma'(a)$.
(g) If f is differentiable at a, then $Df(a) = \nabla f(a)$.

Computations

3.5.4 Let $U = \{(x,y,z) \in \mathbb{R}^3 : z \neq 1\}$. Define $F : U \to \mathbb{R}^2$ by $F(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$.
Assume F is differentiable everywhere.

(a) Why do the partials of F exist? Explain in one sentence.
(b) Calculate the partials of F.

(c) Write the Jacobian of F at $(2,3,7)$.

(d) Give an explicit formula for the differential of F at $(2,3,7)$.

(e) Compute $D_{(-1,0,1)}F(2,3,7)$.

(f) Compute $dF_{(2,3,7)}((-1,0,1))$.

(g) Use the differential of F at $(2,3,7)$ to linearly approximate the value of $F(3,3,4)$.

3.5.5 Define $g(x,y)=(xy, x^2+y^2, x^2-y^2)$. Assume g is differentiable at $(1,2)$.

(a) Give an explicit formula for the differential of g at $(1,2)$.

(b) Linearly approximate $g(1.1,1.8)$ using $dg_{(1,2)}$.

3.5.6 Calculate the Jacobian of the polar coordinate transformation

$$f(r,\theta)=(r\cos\theta, r\sin\theta)$$

at any given point $(r,\theta)\in\mathbb{R}^2$. Assume f is differentiable everywhere.

3.5.7 Calculate the differential of the cylindrical coordinate transformation

$$f(r,\theta,z)=(r\cos\theta, r\sin\theta, z)$$

at any given point $(r,\theta,z)\in\mathbb{R}^3$. Assume f is differentiable everywhere.

3.5.8 Calculate the Jacobian of the spherical coordinate transformation

$$f(\rho,\theta,\phi)=(\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$$

at any given point $(\rho,\theta,\phi)\in\mathbb{R}^3$. Assume f is differentiable everywhere.

Proofs

3.5.9 Define $F:\mathbb{R}^2\to\mathbb{R}^2$ by $F(x,y)=(2+3y, 7x+xy)$. Use the limit definition of differentiability to prove that F is differentiable at $(4,6)$.

3.5.10 Define $G:\mathbb{R}^3\to\mathbb{R}^2$ by $G(x,y,z)=(y+2, 3x+7z^2)$. Use the limit definition of differentiability to prove that G is differentiable at $(4,1,6)$.

3.5.11 Let $f:\mathbb{R}^n\to\mathbb{R}^m$ be a linear map so there exists an $m\times n$ matrix M such that $f(x)=Mx$ for $x\in\mathbb{R}^n$.

(a) For any $a\in\mathbb{R}^n$, use the definition of the differential to prove that f is differentiable at a and show that $df_a=f$.

(b) For any $a\in\mathbb{R}^n$, show that $Df(a)=M$.

3.5.12 Let $A\subseteq\mathbb{R}^n$ and let $f:A\to\mathbb{R}^m$ be a function. Let a be an interior point of A.

(a) Prove that f is differentiable at a if and only if each of its component functions $f^1, f^2, \ldots, f^m$ is. If so,

$$df_a=(df_a^1,\ldots,df_a^m)$$

(b) Assume f is differentiable at a and write its component functions as $f = (f^1, \dots, f^m)$. Show by direct computation that

$$Df(a) = \begin{bmatrix} - & \nabla f^1(a)^T & - \\ & \vdots & \\ - & \nabla f^m(a)^T & - \end{bmatrix}.$$

3.5.13 Let $A \subseteq \mathbb{R}^n$ and let a be an interior point of A. Let $F : A \to \mathbb{R}^m$ and $G : A \to \mathbb{R}^m$ be differentiable at a. Fix $\lambda \in \mathbb{R}$. Use the limit definition of the differential to prove that $F + \lambda G$ is differentiable at a and

$$d(F + \lambda G)_a = dF_a + \lambda dG_a.$$

3.5.14 Beyond two and three dimensions, you have much less geometry or physics to guide your intuition. Instead, you will rely on your algebraic and analytic viewpoints. Consider the lemma:

> **Lemma.** *If $F : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$ then there exists $\varepsilon_a : \mathbb{R}^n \to \mathbb{R}^m$ such that*
> $$\forall h \in \mathbb{R}^n, \qquad F(a+h) = F(a) + DF(a)h + ||h||\varepsilon_a(h),$$
> *where $\varepsilon_a(h) \to 0$ as $h \to 0$.*

(a) What represents the linear approximation of F at a? And the error?

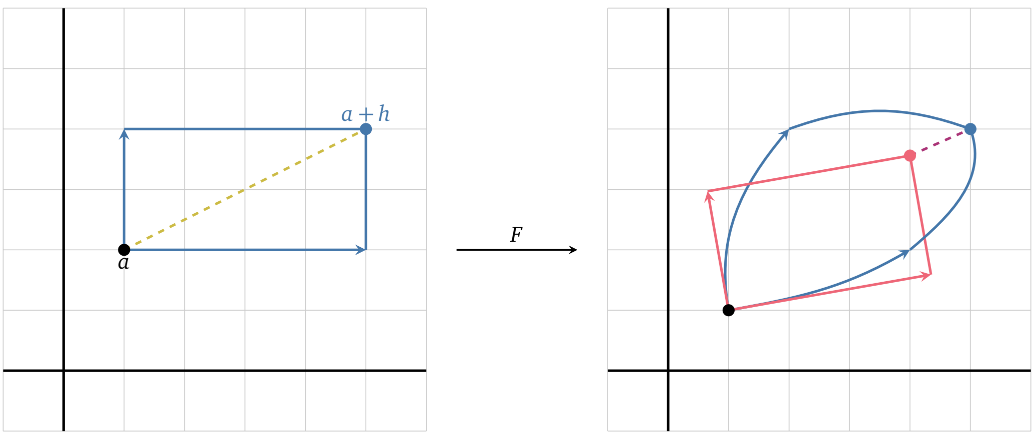
(b) Finish this summary of the lemma. The answer is more than 2 words.

"If F is differentiable at a then the error of its linear approximation at a tends ________"

(c) Prove the lemma. *Hint:* The choice of ε_a will appear circular but it is not. Why not?

Applications and beyond

3.5.15 Here is a picture for the geometry of a linear approximation for a map $F : \mathbb{R}^2 \to \mathbb{R}^2$.



Fix $a \in \mathbb{R}^2$ and $h \in \mathbb{R}^2$. Assume F is differentiable at a.

(a) The points $a \in \mathbb{R}^2$ and $a + h \in \mathbb{R}^2$ are labelled. On the righthand side, the curved rectangle is the image of F and the parallelogram is the image of the linear approximation to F. Label the three vertices on the righthand side using F, a, and h.

(b) Decide what objects are the "error", "exact value", and "approximate value".

(c) There are two dashed lines from one vertex to another. What is the length of each line?

(d) Since F is differentiable at a, it follows that

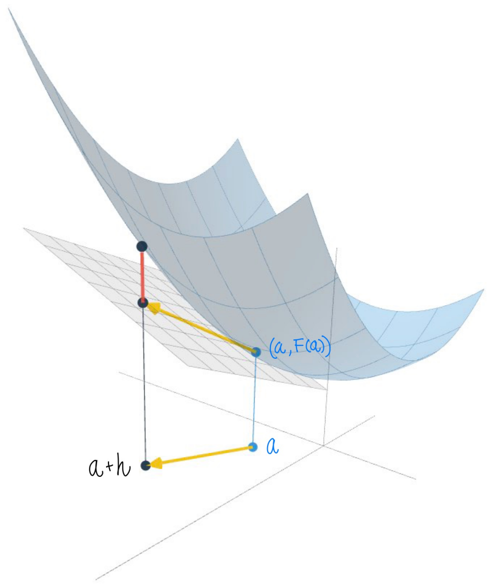
$$\lim_{h \to 0} \frac{F(a+h) - F(a) - dF_a(h)}{||h||} = 0.$$

How can you interpret this limit using the above diagram? Describe in plain language.
Hint: Use the previous part. Think of h as varying. Use this Math3D demo if it helps.

3.5.16 Here is a Math3D demo illustrating a linear approximation of $F : \mathbb{R}^n \to \mathbb{R}^m$. This gives you another geometric perspective of the differential by using the *the graph of* F, namely

$$S = \{(x, F(x)) : x \in \mathbb{R}^n\}.$$

You will interpret the analytic viewpoint using this figure.



(a) Identify the values of m and n that are used in the picture to illustrate the graph of $F : \mathbb{R}^n \to \mathbb{R}^m$.

(b) Label the "error", "exact point", and "approximate point" in the figure.

(c) State the coordinates of the exact point.

(d) State the coordinates of the approximate point.

(e) What is the difference in height between the exact point and approximate point?

(f) Describe the plane P in set builder notation.

3.5.17 Since linear maps are starting to appear in multivariable calculus, you might benefit from a refresher on some linear algebra; this exercise will help. You may return to this intuition later.

When studying maps from $\mathbb{R}^m$ to $\mathbb{R}^n$, you will want to distinguish between special and generic properties. This needs intuition for the non-rigorous idea of "generic" behaviour. You will explore a fundamental example of linear maps which rely on the heuristic principle:

> *If you randomly pick n vectors in $\mathbb{R}^n$ then they will form a basis for $\mathbb{R}^n$ with 100% likelihood.*

A property or object in linear algebra is considered "generic" if it respects this principle.

(a) Explain why a "generic" collection of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ should satisfy

$$\dim\big(\operatorname{span}\{v_1, \ldots, v_m\}\big) = \min\{m, n\}.$$

(b) Why should a "generic" $m \times n$ matrix A satisfy $\operatorname{rank}(A) = \min\{m, n\}$?

(c) Why should a "generic" linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ with $m \le n$ be injective with rank m?

(d) Why should a "generic" linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ with $m \ge n$ be surjective with nullity $m - n$?

(e) Why should a "generic" linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ be bijective?

3.6. Differentiability

You have finally constructed a definition of differentiability (Definition 3.5.1) that meets your wildest expectations. It applies to all non-linear maps of the form $\mathbb{R}^n \to \mathbb{R}^m$ and matches all previous special cases such as parametric curves or real-valued functions. It implies all directional derivatives exist and hence all partial derivatives exist. It merges linear algebra and calculus in a beautiful way. It agrees with your algebraic, analytic, geometric, and physical viewpoints both intuitively and rigorously. Unfortunately, one obstacle remains.

How do you verify that a function is differentiable?

Example 3.5.3 verifies differentiability by definition, but the approach is ad hoc and seems rather cumbersome. It would be preferable to have a simple criterion. Example 3.5.14 demonstrates that the existence of all directional derivatives (and hence the existence of partials) is not sufficient to ensure differentiability. That is disappointing since partial derivative calculations are quick and cheap. Remarkably, you only need to assume a tiny bit more to discover a sufficient condition. While the existence of partials is not enough, you shall see that the existence of *continuous* partials is sufficient! This section is dedicated to the proof and applications of this foundational theorem.

3.6.1 Continuously differentiable functions

Here is the critical new condition.

Definition 3.6.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Let a be an interior point of A. The function f is **continuously differentiable at** a (or C^1 **at** a or of **class** C^1 **at** a) if $\partial_1 f, \dots, \partial_n f$ are defined[11] on an open set containing a and are all continuous at a.

This definition is wonderful because it is relatively simple to check.

Example 3.6.2 Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x,y) = \left(\frac{x^2+y^2}{2}, xy\right)$$

from Example 3.5.3. Its partial derivatives are defined everywhere and given by $\partial_1 f(x,y) = (x,y)$ and $\partial_2 f(x,y) = (y,x)$. The functions $\pi_1 : \mathbb{R}^2 \to \mathbb{R}$ and $\pi_2 : \mathbb{R}^2 \to \mathbb{R}$ defined by $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$ are continuous since they are linear. These are the component functions of the partials of f, so f is continuously differentiable at every $a \in \mathbb{R}^2$.

It is convenient to introduce some language to describe being C^1 at every point in a set.

Definition 3.6.3 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Let S be a subset of the interior of A.
- The function f is **continuously differentiable on** S if f is C^1 at every point $a \in S$.
- The function f is **continuously differentiable** if f is C^1 on its domain.

Example 3.6.4 Define the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x,y,z) = (x^2y^2z, y + \sin z)$. From Example 3.5.18, you can verify that every partial of f is continuous on $\mathbb{R}^3$, so f is C^1.

As these examples illustrate, you can quickly verify whether a function is C^1 since partials are easy to compute. Properties of partial derivatives allow you to construct C^1 functions using other C^1 functions.

[11] If the partials are continuous at a, it may seem redundant to assume they are defined on an open set of a, but it is not. Unless you assume otherwise, it is possible that a is on the boundary of their domain, but these exotic scenarios are not typical. Although such strange functions exist, you will almost never encounter them.

Lemma 3.6.5 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. Let a be an interior point of A. The function f is C^1 at a if and only if each component function f_i is C^1 at a for all $i \in \{1, \ldots, m\}$.

Proof. This is left as an exercise. Combine Theorem 2.6.14 and Lemma 3.2.7. ■

Theorem 3.6.6 Let $A \subseteq \mathbb{R}^n$. Let f and g be $\mathbb{R}^m$-valued functions defined on A. Let ϕ and ψ be real-valued functions defined on A. Fix $\lambda \in \mathbb{R}$ and let a be an interior point of A.

(a) If f and g are C^1 at a then their linear combination $f + \lambda g$ is C^1 at a.
(b) If f and g are C^1 at a then their dot product $f \cdot g$ is C^1 at a.
(c) If f and ϕ are C^1 at a then the scalar product ϕf is C^1 at a.
(d) If ϕ and ψ are C^1 at a and $\psi(a) \neq 0$ then the quotient $\dfrac{\phi}{\psi}$ is C^1 at a.

Proof. This is left as an exercise. Use properties of partial derivatives (Theorem 3.2.8), properties of continuity (Theorem 2.6.18), and single variable differentiation rules. ■

There are two large families of functions which are shown to be C^1.

Lemma 3.6.7 All linear maps $\mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable.

Proof. This is left as an exercise. Directly compute the partials. ■

Lemma 3.6.8 All polynomials $\mathbb{R}^n \to \mathbb{R}$ are continuously differentiable.

Proof. This is left as an exercise. Use linearity to reduce to monomials. Then use induction on the degree of the monomial. ■

For now, you can directly compute partial derivatives and check whether a function is C^1 at a point. After this section, it will be enough to simply state whether a function is C^1 or not, because you can often accept this fact without explicitly doing the computation. This acceptance also occurs in single variable calculus, where most functions you encounter are differentiable wherever they are defined.

3.6.2 Differentiability criterion

This new easy-to-verify definition produces a machine for checking differentiability.

Theorem 3.6.9 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ be a function. Let a be an interior point of A. If f is continuously differentiable at a then f is differentiable at a.

Proof. Postponed to the end of this subsection. ■

First, you should take a moment to admire how useful it is. From Lemmas 3.6.7 and 3.6.8, it follows by Theorem 3.6.9 all linear maps and all polynomials are differentiable. That is quite a coup! You can also explicitly check examples without much trouble.

Example 3.6.10 The function $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f(x, y, z) = (x^2 y^2 z, y + \sin z)$ is C^1 by Example 3.6.4 so, by Theorem 3.6.9, f is differentiable.

Note, however, that the converse of Theorem 3.6.9 is false.

Example 3.6.11 Define $f : \mathbb{R} \to \mathbb{R}$ by $f(0) = 0$ and $f(x) = x^2 \sin(1/x)$ for $x \neq 0$. By direct computation, $f'(x) = 2x\sin(1/x) - \cos(1/x)$ for $x \neq 0$ and

$$f'(0) = \lim_{x\to 0} \frac{f(x)-f(0)}{x-0} = \lim_{x\to 0} x\sin(1/x) = 0,$$

so f is differentiable everywhere. However, $\partial_1 f = f'$ is not continuous at 0 since

$$\lim_{x\to 0} f'(x) = \lim_{x\to 0}(2x\sin(1/x) - \cos(1/x)) = 0 - \lim_{x\to 0}\cos(1/x)$$

and this last limit does not exist. This proves that the converse of Theorem 3.6.9 is false. View this function and its derivative on Desmos.

You can now connect many theorems together to compute any kind of derivative you like.

Example 3.6.12 By Example 3.6.2, the function $f(x,y) = (\frac{1}{2}x^2 + \frac{1}{2}y^2, xy)$ is C^1 everywhere so, by Theorem 3.6.9, f is differentiable everywhere. Thus, by Theorem 3.5.22,

$$df_{(a,b)}(v) = Df(a,b)v = \begin{bmatrix} \partial_1 f(a,b) & \partial_2 f(a,b) \end{bmatrix} v = \begin{bmatrix} a & b \\ b & a \end{bmatrix} v$$

for $(a,b) \in \mathbb{R}^2$ and $v \in \mathbb{R}^2$. Thus, by Theorem 3.5.13,

$$D_v f(a,b) = \begin{bmatrix} a & b \\ b & a \end{bmatrix} v.$$

The proof of Theorem 3.6.9 is a serious feat, so buckle up for a bumpy first read.

Proof of Theorem 3.6.9. By Lemmas 3.5.10 and 3.6.5, it suffices to prove Theorem 3.6.9 for a real-valued function $f : A \to \mathbb{R}$. Assume $f : A \to \mathbb{R}$ is C^1 at a. To prove that f is differentiable at a, it suffices to show that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - \nabla f(a)^T h}{||h||} = 0. \tag{3.6.1}$$

We shall proceed by the formal definition of the limit.

Fix $\varepsilon > 0$. By assumption, every partial of f is continuous at a and defined on an open ball containing a. Thus, for each $j \in \{1, \dots, n\}$, there exists $\delta_j > 0$ such that

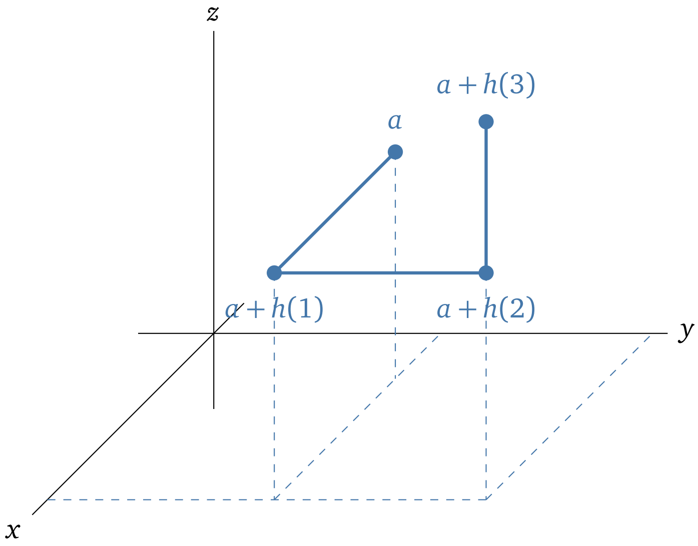
$$\forall x \in \mathbb{R}^n, ||x-a|| < \delta_j \implies \left|\partial_j f(x) - \partial_j f(a)\right| < \frac{\varepsilon}{n}. \tag{3.6.2}$$

Set $\delta = \frac{1}{2}\min\{\delta_1, \dots, \delta_n\}$. This choice ensures that the function f and its partials $\partial_1 f, \dots, \partial_n f$ are all defined on $B_{2\delta}(a)$.

Before continuing the proof, some notation must be introduced. Let $\{e_1, \dots, e_n\}$ be the standard basis in $\mathbb{R}^n$. For $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, define the point $h(0) = (0, 0, \dots, 0)$ and the points $h(1), \dots, h(n)$ by

$$h(k) = \sum_{j=1}^{k} h_j e_j = (h_1, h_2, \dots, h_k, 0, \dots, 0)$$

for $k \in \{1, \dots, n\}$. This construction of points is illustrated below for the case $n = 3$.



Now, continue with the proof. Fix $h \in \mathbb{R}^n$ such that $||h|| < \delta$. It follows that

$$f(a+h)-f(a)=\sum_{i=1}^{n}\big[f(a+h(i))-f(a+h(i-1))\big]. \tag{3.6.3}$$

Temporarily fix $i \in \{1,\dots,n\}$. Define $g_i : (a_i-\delta, a_i+\delta) \to \mathbb{R}$ by

$$g_i(t) = f(a_1+h_1,\dots,a_{i-1}+h_{i-1},t,a_{i+1},\dots,a_n)$$

The choice of δ implies that g_i is differentiable. By the single variable mean value theorem, it follows that

$$g_i(a_i+h_i)-g_i(a_i) = g_i'(c_i)h_i$$

for some $c_i \in \mathbb{R}$ between a_i and a_i+h_i. By definition of g_i, this implies that

$$\begin{aligned}
&f(a+h(i))-f(a+h(i-1))\\
&=f(a_1+h_1,\dots,a_{i-1}+h_{i-1},a_i+h_i,a_{i+1},\dots,a_n)-f(a_1+h_1,\dots,a_{i-1}+h_{i-1},a_i,a_{i+1},\dots,a_n)\\
&=\partial_i f(a_1+h_1,\dots,a_{i-1}+h_{i-1},c_i,a_{i+1},\dots,a_n)h_i.
\end{aligned}$$

Thus, for each $i \in \{1,\dots,n\}$, there exists a point $b_i \in B_\delta(a)$,

$$f(a+h(i))-f(a+h(i-1)) = h_i\partial_i f(b_i). \tag{3.6.4}$$

Consequently,

$$\begin{aligned}
\left|\frac{f(a+h)-f(a)-\nabla f(a)^T h}{||h||}-0\right| &= \frac{|f(a+h)-f(a)-\sum_{i=1}^{n}\partial_i f(a)h_i|}{||h||}\\
&=\frac{1}{||h||}\left|\sum_{i=1}^{n}\big[\partial_i f(b_i)-\partial_i f(a)\big]h_i\right|\\
&\le \sum_{i=1}^{n}\big|\partial_i f(b_i)-\partial_i f(a)\big|\frac{|h_i|}{||h||}\\
&\le \sum_{i=1}^{n}\big|\partial_i f(b_i)-\partial_i f(a)\big|\\
&< \sum_{i=1}^{n}\frac{\varepsilon}{n}=\varepsilon.
\end{aligned}$$

This completes the proof (3.6.6) and hence the theorem. ■

That was an absolute masterpiece of many ideas, so re-reading will certainly be necessary. Right now, take a moment to appreciate your long journey towards a grand definition.

You have navigated many definitions of a derivative: single variable, partial derivatives, directional derivatives, gradients, differentials, and Jacobians. Each development introduced a new idea and provided another insight on how to approximate change in maps $\mathbb{R}^n \to \mathbb{R}^m$. Every perspective – analytic, algebraic, geometric, and physical – contributed to a deeper understanding. The story climaxed with the perfect blend of linear algebra and calculus, namely the differential, which embodies this chapter's underlying philosophy.

Nonlinear maps are well approximated by linear maps.

The differential connected all of the prior definitions and solidified their theoretical foundations with a slew of core theorems (Theorems 3.5.13, 3.5.22 and 3.6.9). In the next chapter, you will leverage these tools to produce some fantastic applications in optimization and geometry.

Exercises for Section 3.6

Concepts and definitions

3.6.1 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in \mathbb{R}^n$. Which statements are equivalent to "F is differentiable at a"?

(a) There exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{F(a+h)-F(a)-L(h)}{||h||} = 0$.

(b) There exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{F(a+h)-F(a)-L(h)}{h} = 0$.

(c) There exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \left|\left|\dfrac{F(a+h)-F(a)-L(h)}{h}\right|\right| = 0$.

(d) There exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{||F(a+h)-F(a)-L(h)||}{||h||} = 0$.

(e) The limit $\lim\limits_{h\to 0} \dfrac{F(a+h)-F(a)}{h}$ exists.

3.6.2 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and fix $a \in \mathbb{R}^n$. Which of these statements are true or false? If true, briefly justify and cite any theorems used. If false, cite a counterexample or provide your own.

(a) If F is differentiable at a, then all directional derivatives of F at a exist.

(b) If all directional derivatives of F at a exist, then F is differentiable at a.

(c) F is differentiable at a if and only if all the partial derivatives of F exist.

(d) If F is differentiable at a, then the Jacobian of F at a is a matrix of partial derivatives.

(e) If F is continuously differentiable at a, then F is differentiable at a.

(f) If F is differentiable at a, then F is continuously differentiable at a.

3.6.3 A mind map of theorems, lemmas, and counterexamples is an effective strategy for organizing complex mathematical theories within your own mental framework. Fill in the mind map below for derivatives. Draw ($\Longrightarrow$) or ($\Longleftrightarrow$) or ($\not\Longrightarrow$) depending on whether it follows from a theorem, definition, or counterexample. If possible, cite a theorem or counterexample next to the arrow. Draw enough arrows, not all possible arrows. Note the phrasing is intended to be suggestive, not precise.

Continuous partials

Differential exists | Differentiable

All directional derivatives exist | directional derivative $= \sum$ partials

Jacobian exists | All partials exist | Continuous

Proofs

3.6.4 Let $F : \mathbb{R}^n \to \mathbb{R}^m$. Fix $a \in \mathbb{R}^n$. Prove that if F is differentiable at a then F is continuous at a. *Hint:* The proof is similar to the single variable proof but be careful to use multivariable definitions. Remember quotients of vectors do not make any sense.

3.6.5 Show that all linear maps $F : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable.

3.6.6 Show that all multivariable polynomials $f : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable.

3.6.7 The differential outputs the directional derivative. You will revisit its proof here.

> **Theorem A.** *Let $A \subseteq \mathbb{R}^n$ and let $a \in A^\circ$. If $f : A \to \mathbb{R}^m$ is differentiable at a, then the directional derivative $D_v f(a)$ exists for all $v \in \mathbb{R}^n$ and $df_a(v) = D_v f(a)$.*

(a) The first half of the proof is below.

1. *Assume f is differentiable at a. Fix $v \in \mathbb{R}^n$.*
2. *By linearity, $df_a(tv) = t\,df_a(v)$ for $t \in \mathbb{R}$.*
3. *Thus,* $D_v f(a) = \lim_{t\to 0} \dfrac{f(a+tv)-f(a)}{t} = \lim_{t\to 0}\left[\dfrac{f(a+tv)-f(a)-df_a(tv)}{t} + df_a(v)\right]$.
4. *By the addition limit law, it suffices to show* $\lim_{t\to 0} \dfrac{f(a+tv)-f(a)-df_a(tv)}{t} = 0$.
5. *For $t \in \mathbb{R}$, the identity $||tv|| = |t| \cdot ||v||$ implies that*
$$\lim_{t\to 0} \frac{f(a+tv)-f(a)-df_a(tv)}{t} = \lim_{t\to 0}\left[\left(\frac{f(a+tv)-f(a)-df_a(tv)}{||tv||}\right)\cdot \frac{|t|}{t}\cdot ||v||\right].$$
6. *Since the quantity $\frac{|t|}{t}\cdot ||v||$ is bounded between $-||v||$ and $||v||$ as $t \to 0$, the result follows from the multivariable squeeze theorem (Theorem 2.5.15) provided*
$$\lim_{t\to 0} \frac{f(a+tv)-f(a)-df_a(tv)}{||tv||} = 0. \tag{3.6.5}$$

Line 6 uses the multivariable squeeze theorem to conclude the claim in Line 4 holds, but the squeeze theorem only holds for real-valued functions. Justify this assertion with a proof.

(b) The rest of the proof establishes the limit in Line 6.

7. *Define $F : A \to \mathbb{R}^m$ by $F(a) = 0$ and, for $x \neq a$,* $F(x) = \dfrac{f(x)-f(a)-df_a(x-a)}{||x-a||}$.
8. *Since f is differentiable at a, it follows that F is continuous at 0.*
9. *As a is an interior point of A, there exists $\delta > 0$ such that $\{a+tv : t \in (-\delta,\delta)\} \subseteq A$.*
10. *Define $\gamma : (-\delta,\delta) \to \mathbb{R}^m$ by $\gamma(t) = a+tv$ so γ is continuous and $\operatorname{im}(\gamma) \subseteq A$.*
11. *Since the composition of continuous functions is continuous, we have that*
$$\lim_{t\to 0} \frac{f(a+tv)-f(a)-df_a(tv)}{||tv||} = \lim_{t\to 0} F(\gamma(t)) = F(\gamma(0)) = F(a) = 0.$$

Line 11 has four equalities. Identify which line(s) justify which equality.

(c) Line 9 requires a proof from topology. Justify this assertion.

3.6.8 The Jacobian is the matrix of the differential. Here is a simplified version the proof.

> **Theorem B.** *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a then the matrix of the differential of f at a is the $m \times n$ Jacobian matrix of f at a. In other words, $df_a(v) = Df(a)v$ for all $v \in \mathbb{R}^n$.*

1. *Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$.*
2. *By Theorem A and the definition of the Jacobian,*

$$\forall j \in \{1, \ldots, n\}, \quad df_a(e_j) = D_{e_j} f(a) = \partial_j f(a) = Df(a)e_j.$$

3. *The result now follows from linear algebra.*

(a) Line 2 has three equalities, each following for a different reason. Identify each justification.

(b) Line 3 is a bit vague. Precisely describe this result.

3.6.9 Finally, you will study the proof of the core theorem. Here is a special case.

> **Theorem C.** *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 at $a \in \mathbb{R}^n$ then f is differentiable at a.*

Its proof is not easy. You will study components of it by unravelling some line-by-line details.

(a) Here are two lemmas used in its proof.

> **Lemma I.** *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a if and only if each of its component functions $f^1, f^2, \ldots, f^m$ is. If so, $df_a = (df_a^1, \ldots, df_a^m)$.*

> **Lemma II.** *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 at a if and only if each of its component functions $f^1, f^2, \ldots, f^m$ is.*

The proof of Theorem C begins with this first line.

1. *By Lemmas I and II, it suffices to consider a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$.*

The phrase "*it suffices to consider*" suggests that if you prove another simplified version of Theorem C, then the lemmas and this other statement will imply Theorem C rather easily. Fill in the blank for this simplified version of Theorem C.

> **Theorem D.** *If* ____________________ *then f is differentiable at a.*

(b) Prove that Lemma I, Lemma II, and Theorem D imply Theorem C. It takes only a few lines.

(c) Theorem C continues with the next couple of lines.

2. *Assume $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 at a. To prove that f is differentiable at a, it suffices to show that*

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \nabla f(a)^T h}{||h||} = 0. \tag{3.6.6}$$

3. *We shall proceed by the formal definition of the limit.*

i) Why does $\nabla f(a)$ exist?

ii) Why does (3.6.6) prove that f is differentiable at a? Refer to the definition.

iii) State the formal definition of the limit in (3.6.6). This will guide the flow of the proof.

(d) The proof of (3.6.6) begins with a mysterious choice of δ. This choice will only make sense once you reach the end of the proof, but you can still verify that the steps are justified.

4. *Fix $\varepsilon > 0$.*
5. *Every partial of f is continuous at a and defined on an open ball containing a.*
6. *Thus, for each $j \in \{1,\dots,n\}$, there exists $\delta_j > 0$ such that*

$$\forall x \in \mathbb{R}^n, ||x-a|| < \delta_j \implies \left|\partial_j f(x) - \partial_j f(a)\right| < \frac{\varepsilon}{n}. \tag{3.6.7}$$

7. *Set $\delta = \frac{1}{2}\min\{\delta_1,\dots,\delta_n\}$.*
8. *This choice ensures that the function f and its partials $\partial_1 f,\dots,\partial_n f$ are all defined on $B_{2\delta}(a)$.*

How does Line 6 use that $\partial_j f$ is defined on an open ball containing a?

(e) The proof continues with an application of the single variable mean value theorem.

9. *Fix $h \in \mathbb{R}^n$ such that $||h|| < \delta$. Temporarily fix $i \in \{1,\dots,n\}$.*
10. *Recall $h(i) = (h_1,\dots,h_{i-1},h_i,0,\dots,0)$ and $h(i-1) = (h_1,\dots,h_{i-1},0,0,\dots,0) \in \mathbb{R}^n$.*
11. *Define $g_i : (a_i-\delta, a_i+\delta) \to \mathbb{R}$ by*

$$g_i(t) = f(a_1+h_1,\dots,a_{i-1}+h_{i-1},t,a_{i+1},\dots,a_n) \quad \text{for } t \in (a_i-\delta,a_i+\delta)$$

12. *The choice of δ implies that g_i is differentiable.*
13. *By the single variable mean value theorem, it follows that*

$$g_i(a_i+h_i) - g_i(a_i) = g_i'(c_i)h_i$$

for some $c_i \in \mathbb{R}$ between a_i and a_i+h_i.

i) Justify Line 12 using any of the previous 11 lines. This will unravel one aspect of the choice of δ.
ii) On what interval of $\mathbb{R}$ is the single variable mean value theorem applied? Explain why the assumptions are satisfied on this interval.

(f) The proof continues by translating from the single variable map g_i to the multivariable map f.

14. *By definition of g_i, this implies that*

$$\begin{aligned} &f(a+h(i)) - f(a+h(i-1)) \\ &= f(a_1+h_1,\dots,a_{i-1}+h_{i-1},a_i+h_i,a_{i+1},\dots,a_n) - f(a_1+h_1,\dots,a_{i-1}+h_{i-1},a_i,a_{i+1},\dots,a_n) \\ &= \partial_i f(a_1+h_1,\dots,a_{i-1}+h_{i-1},c_i,a_{i+1},\dots,a_n)h_i. \end{aligned}$$

15. *Thus, for each $i \in \{1,\dots,n\}$, there exists a point $b_i \in B_\delta(a)$,*

$$f(a+h(i)) - f(a+h(i-1)) = h_i \partial_i f(b_i). \tag{3.6.8}$$

i) By comparing limit definitions, verify that $g_i'(c_i) = \partial_i f(a_1+h_1,\dots,a_{i-1}+h_{i-1},c_i,a_{i+1},\dots,a_n)$.
ii) Express b_i in terms of the quantities in Line 14. Justify that $b_i \in B_\delta(a)$ using earlier lines.

(g) The conclusion lacks some detail. Add a short justification to the right of each step.

16. Consequently,

$$\begin{aligned}
\left|\frac{f(a+h)-f(a)-\nabla f(a)^T h}{||h||}-0\right| &= \frac{|f(a+h)-f(a)-\sum_{i=1}^n \partial_i f(a)h_i|}{||h||} \\
&= \frac{1}{||h||}\left|\sum_{i=1}^n \big[\partial_i f(b_i)-\partial_i f(a)\big]h_i\right| \\
&\le \sum_{i=1}^n \big|\partial_i f(b_i)-\partial_i f(a)\big|\frac{|h_i|}{||h||} \\
&\le \sum_{i=1}^n \big|\partial_i f(b_i)-\partial_i f(a)\big| \\
&< \sum_{i=1}^n \frac{\varepsilon}{n} = \varepsilon.
\end{aligned}$$

17. This completes the proof of (3.6.6) *and hence Theorem C.*

3.6.10 The big picture is always important. You will need to digest complicated proofs like Theorem C beyond a line-by-line reading. It is like understanding an algorithm beyond reading Python code line-by-line. You will try some general strategies here which will help you discover the big picture on your own.

(a) One strategy is to specialize to the simplest non-simple example. This helps cut notation and brings forward the ideas. Consider these sentences from the proof of Theorem C.

- *Given* $h=(h_1,\dots,h_n)$*, let* $h(0)=\mathbf{0}$ *and* $h(i)=(h_1,\dots,h_i,0,\dots,0)$ *for* $i\in\{1,\dots,n\}$.
- *Then* $f(a+h)-f(a)=\sum_{i=1}^n \big[f(a+h(i))-f(a+h(i-1))\big]$

Rewrite these sentences using $n=3$. Do not use Σ-notation or the phrase "for $i=1,2,3$".

(b) Explain how these rewritten sentences relate to the figure in the proof of Theorem 3.6.9.

(c) Another strategy is to summarize the entire proof in a few imprecise steps. Here is an attempt.

A) Reduce to the case of a scalar-valued function f.
B) Break down the total change in f from a to $a+h$ into component-by-component changes.
C) Estimate the ith component-by-component change using an ith partial at a point near a.
D) Since the partials are continuous, the error between the ith partial at a point near a compared to the ith partial at a is very small.
E) Combine everything to prove the total change in f from a to $a+h$ can be linearly approximated; hence, f is differentiable at a.

Below is a condensed proof of Theorem C. Highlight which step corresponds to which part of the proof. Sometimes it may be one line and sometimes it may be several lines.

1. *By Lemmas I and II, it suffices to consider a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$.*
2. *Fix $\varepsilon > 0$. For each $j \in \{1, \dots, n\}$, there exists $\delta_j > 0$ such that $\partial_j f$ is defined on $B_{\delta_j}(a)$ and*
$$\forall x \in \mathbb{R}^n, ||x - a|| < \delta_j \implies |\partial_j f(x) - \partial_j f(a)| < \varepsilon.$$
3. *Set $\delta = \frac{1}{2}\min\{\delta_1, \dots, \delta_n\}$ and let $\{e_1, \dots, e_n\}$ is the standard basis in $\mathbb{R}^n$.*
4. *For $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, define the point $h(0) = (0, 0, \dots, 0)$ and the points $h(1), \dots, h(n)$ by $h(k) = \sum_{j=1}^k h_j e_j = (h_1, h_2, \dots, h_k, 0, \dots, 0)$ for $k \in \{1, \dots, n\}$.*
5. *Let $h \in \mathbb{R}^n$ be such that $||h|| < \delta$. It follows that*
$$f(a+h) - f(a) = \sum_{i=1}^n \big[f(a + h(i)) - f(a + h(i-1))\big].$$
6. *Temporarily fix $i \in \{1, \dots, n\}$. By the single variable mean value theorem,*
$$\begin{aligned}&f(a + h(i)) - f(a + h(i-1))\\ &= f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i + h_i, a_{i+1}, \dots, a_n) - f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i, a_{i+1}, \dots, a_n)\\ &= \partial_i f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, c_i, a_{i+1}, \dots, a_n)\end{aligned}$$
for some $c_i \in \mathbb{R}$ between a_i and $a_i + h_i$, since $\partial_i f$ is the derivative of the single variable function
$$g_i(x) = f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, x, a_{i+1}, \dots, a_n).$$
7. *Thus, $f(a + h(i)) - f(a + h(i-1)) = h_i \partial_i f(b_i)$ for some point $b_i \in B_\delta(a)$.*
8. *Consequently,*
$$\begin{aligned}\frac{|f(a+h) - f(a) - \sum_{i=1}^n \partial_i f(a) h_i|}{||h||} &= \frac{1}{||h||}\left|\sum_{i=1}^n \big[\partial_i f(b_i) - \partial_i f(a)\big] h_i\right|\\ &\le \sum_{i=1}^n \big|\partial_i f(b_i) - \partial_i f(a)\big| \frac{|h_i|}{||h||}\\ &\le \sum_{i=1}^n \big|\partial_i f(b_i) - \partial_i f(a)\big| < \sum_{i=1}^n \frac{\varepsilon}{n} = \varepsilon.\end{aligned}$$
9. *This completes the proof.*

Applications and beyond

3.6.11 Recall our core philosophy for differentiation:

Nonlinear maps can be locally approximated by linear maps.

You will expand this intuition further. A property for a nonlinear map is usually considered to hold **locally** at a point p if there is a small enough open ball centered at p on which the property holds. This idea powers the study of nonlinear maps $F : \mathbb{R}^n \to \mathbb{R}^m$. Fix $a \in \mathbb{R}^n$. Guess a formal definition for each of the following statements.

(a) F is locally non-constant at a.

(b) F is locally injective at a.

3.6.12 By applying the idea of local properties from the previous question, you will now transfer the idea of "generic" objects from linear maps to nonlinear maps. Review Exercise 3.5.17 for details on this heuristic. This will bridge intuition between linear algebra and multivariable calculus.

(a) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be C^1 at $a \in \mathbb{R}^n$. Consider F to be "generic" at a if its differential dF_a is a "generic" linear map. What should be the rank of the Jacobian of F at a?

(b) If $n \leq m$ and F is "generic" at a then explain why F should be locally injective at a.

(c) If $n > m$ and F is "generic" at a then explain why F should not be locally injective at a.

4. Derivative applications

Equipped with a sophisticated theory of differentiation and a bounty of derivatives, you are ready to apply these notions to mathematical and scientific applications. Real world contexts are, however, much more intricate. Variables depend on each other in complex ways. A variable can depend on several variables, which themselves depend on more variables.

> *If my variables have several layers of dependencies, how does change in one multidimensional variable affect the change of another multidimensional variable?*

This question can be viewed as a generalization of related rates questions to higher dimensions. Resolving this problem will naturally prepare you for the star of the show: optimization.

Optimization is the pinnacle of your applications because of its incredible versatility across almost every subject imaginable, such as statistics, physics, computer science, economics, chemistry, and biology. With your substantial efforts in the previous chapter, you can tackle the next big questions.

> *How can you locate possible extrema? How do you actually ensure you have solved your optimization problem?*

By developing one more tool, you will be empowered to address a huge array of scenarios.

On the other hand, you will also encounter a surprising new obstacle.

> *What are the limitations of my optimization tools so far? Why does it seem so difficult to optimize over a "lower dimensional manifold"? What really is a "manifold"?*

These questions are surprisingly deep and spawn an entire subject of study.

> *What really is a "manifold" and what is its "dimension"? What does it mean for a plane to be "tangent" to a manifold?*

By the end of this chapter, you will investigate all of these fundamental geometric questions by rigorously defining tangent spaces, tangent planes, and manifolds. Derivatives created these problems, and derivatives will resolve them!

4.1. Chain rule

Variables in the real world depend on each other in complex unknown ways. This issue raises one of this chapter's motivating questions.

> *If my variables have several layers of dependencies, how does change in one multidimensional variable affect the change of another multidimensional variable?*

While the informal notion of "variables" is useful, you can formally rephrase this question in terms of maps and their differentials.

> *If a map is a composition of several or more maps, then how are the differentials of these maps related?*

The answer to this question is a beautiful theorem: the (multivariable) chain rule.

Theorem 4.1.1 (Chain rule) Let U and V be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. If the mappings $f : U \to V$ and $g : V \to \mathbb{R}^k$ are differentiable at $a \in U$ and $f(a) \in V$ respectively, then their composition $h = g \circ f$ is differentiable at a. Moreover, its differential is a composition of linear maps,

$$dh_a = dg_{f(a)} \circ df_a,$$

and its $k \times n$ Jacobian matrix is a product of a $k \times m$ Jacobian matrix with a $m \times n$ Jacobian matrix, namely

$$Dh(a) = Dg(f(a))Df(a).$$

Informally, the chain rule says that:

> *The differential of a composition is the composition of differentials!*

This elegant statement was also displayed at the end of Section 3.1 for single variable functions. You will practice with it in some explicit examples, interpret the statement geometrically, and reframe it with variables and Leibniz notation. Its proof is postponed to the end of this section.

4.1.1 Basic examples

The chain rule is algebraically elegant but it is quite messy to apply in practice.

Example 4.1.2 Here you consider a sequence of compositions $\mathbb{R}^2 \to \mathbb{R}^3 \to \mathbb{R}$. Namely, consider functions $f : \mathbb{R}^2 \to \mathbb{R}^3$ and $g : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x, y) = (x^2 y, x^2 + y^2, \cos y) \qquad g(x, y, z) = xz^3 + \sin y.$$

Is $g \circ f$ differentiable at the point $(1, 0)$ and, if so, what is its differential?

Both f and g are C^1 and hence differentiable by Theorem 3.6.9. By the chain rule, their composition $h = g \circ f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable. To calculate the Jacobian of h at $(1, 0)$, we first calculate the Jacobians of f at $(1, 0)$ and the Jacobian of g at $f(1, 0) = (0, 1, 1)$. Note

$$Df(x, y) = \begin{bmatrix} | & | \\ f_x & f_y \\ | & | \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 2x & 2y \\ 0 & -\sin y \end{bmatrix} \quad \text{so} \quad Df(1, 0) = \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

and similarly, as $f(1, 0) = (0, 1, 1)$,

$$Dg(x, y, z) = \begin{bmatrix} z^3 & \cos y & 3xz^2 \end{bmatrix} \quad \text{so} \quad Dg(f(1, 0)) = \begin{bmatrix} 1 & \cos 1 & 0 \end{bmatrix}.$$

By the chain rule, the 1×2 Jacobian of h at $(1,0)$ is given by

$$Dh(1,0) = Dg(f(1,0))Df(1,0) = \begin{bmatrix} 1 & \cos 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2\cos 1 & 1 \end{bmatrix}. \tag{4.1.1}$$

Thus, the differential satisfies $dh_{(1,0)}(v) = \begin{bmatrix} 2\cos 1 & 1 \end{bmatrix} v = 2(\cos 1)v_1 + v_2$ for $v \in \mathbb{R}^2$.

In the above calculation, there is nothing special about the point $(1,0)$. By the chain rule, the Jacobian of h at any $(x,y) \in \mathbb{R}^2$ is given by

$$\begin{aligned} Dh(x,y) &= Dg(f(x,y)) \cdot Df(x,y) \\ &= \begin{bmatrix} (\cos y)^3 & \cos(x^2+y^2) & 3x^2y(\cos y)^2 \end{bmatrix} \begin{bmatrix} 2xy & x^2 \\ 2x & 2y \\ 0 & -\sin y \end{bmatrix} \\ &= \begin{bmatrix} 2xy(\cos y)^3 + 2x\cos(x^2+y^2) & x^2(\cos y)^3 + 2y\cos(x^2+y^2) - 3x^2y(\sin y)(\cos y)^2 \end{bmatrix}. \end{aligned}$$

You can actually check this computation directly, because you can express $h = g \circ f$ as

$$h(x,y) = x^2 y \cos^3 y + \sin(x^2+y^2)$$

and again calculate the Jacobian $Dh(x,y) = \begin{bmatrix} h_x & h_y \end{bmatrix}$ directly. This method may be easier, but that does not matter. Computational efficiency is not the main value of the chain rule.

The previous example illustrates how a computation with the chain rule is executed. However, unlike single variable calculus, the purpose of the multivariable chain rule is *not* to perform computations for explicit choices of functions. Its primary goal is to express algebraic relationships between derivatives and it is often used in proofs or algorithms.

Example 4.1.3 The chain rule is commonly used to calculate derivatives of parametric curves, such as a composition of the form $\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}$. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\gamma : \mathbb{R} \to \mathbb{R}^n$. If γ is differentiable at $t \in \mathbb{R}$ and f is differentiable at $\gamma(t) \in \mathbb{R}^n$, then by the chain rule $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is differentiable at t and

$$\underbrace{D(f \circ \gamma)(t)}_{1\times 1 \text{ matrix}} = \underbrace{(Df)(\gamma(t))}_{1\times n \text{ matrix}} \underbrace{D\gamma(t)}_{n\times 1 \text{ matrix}}.$$

Since f is real-valued, the Jacobian of f is the transpose of the gradient, namely $Df(x) = \nabla f(x)^T$ if f is differentiable at $x \in \mathbb{R}^n$. Since $f \circ \gamma$ and γ are single-variable functions, their Jacobians are the $n \times 1$ derivative, namely $D\gamma(t) = \gamma'(t)$ and $D(f \circ \gamma)(t) = (f \circ \gamma)'(t)$. Thus, the above formula may be written as

$$(f \circ \gamma)'(t) = \nabla f(\gamma(t))^T \gamma'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t).$$

This identity will be critical when you later study implicit surfaces and tangent planes.

There is also a convenient consequence of the chain rule with C^1 functions.

Corollary 4.1.4 Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open sets. If the function $f : U \to V$ is C^1 and the function $g : V \to \mathbb{R}^k$ is C^1, then their composition $g \circ f : U \to \mathbb{R}^k$ is C^1.

Proof. This is left as an exercise. Use the chain rule and Theorem 3.6.9. ∎

Computationally speaking, the chain rule is extensively used to express partial derivatives in other coordinate systems. This poses a serious notational challenge to the current approach with Jacobians, so you must rely on another convenient notation.

4.1.2 Variables, Leibniz notation, and chain rule trees

The chain rule is extremely cumbersome if you attempt to write everything out carefully with formal notation for functions. Here is a simple illustration of this issue.

Example 4.1.5 Consider a composition of the form $\mathbb{R}^2 \to \mathbb{R}^3 \to \mathbb{R}$. So, assume $f : \mathbb{R}^2 \to \mathbb{R}^3$ and $g : \mathbb{R}^3 \to \mathbb{R}$ are both differentiable everywhere. By the chain rule, $h = g \circ f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and you want to construct a general formula for its Jacobian in terms of partial derivatives of f and g. How would you formally write this out?

The Jacobian of g at $f(a)$ and the Jacobian of $f = (f_1, f_2, f_3)$ at a are

$$Dg(f(a)) = \begin{bmatrix} \partial_1 g(f(a)) & \partial_2 g(f(a)) & \partial_3 g(f(a)) \end{bmatrix} \qquad Df(a) = \begin{bmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) \\ \partial_1 f_3(a) & \partial_2 f_3(a) \end{bmatrix}.$$

Multiplying these together, you find that

$$Dh(a) = \nabla h(a)^T = \begin{bmatrix} \partial_1 g(f(a))\partial_1 f_1(a) + \partial_2 g(f(a))\partial_1 f_2(a) + \partial_3 g(f(a))\partial_1 f_3(a) \\ \partial_1 g(f(a))\partial_2 f_1(a) + \partial_2 g(f(a))\partial_2 f_2(a) + \partial_3 g(f(a))\partial_2 f_3(a) \end{bmatrix}^T. \tag{4.1.2}$$

This is completely unambiguous and formally valid, but it is a rather gross expression and it is difficult to detect whether you have made an error.

This issue is alleviated by using a variable letter in place of a function, and by using the Leibniz notation for partial derivatives. For instance, if $g : \mathbb{R}^3 \to \mathbb{R}$ is differentiable and written as $g(x, y, z)$ then you can write

$$\partial_1 g \qquad \text{or} \qquad \frac{\partial g}{\partial x}$$

to indicate the partial derivative with respect to the 1st coordinate. Both of these are functions $\mathbb{R}^3 \to \mathbb{R}$. The former is completely unambiguous, but the latter relies on you understanding that x denotes the first coordinate. The second notation $\frac{\partial g}{\partial x}$ is referred to as *Leibniz notation*. Instead of writing the function $g(x, y, z)$, you may often instead write

$$u = g(x, y, z) \qquad \text{and} \qquad \frac{\partial u}{\partial x}.$$

The letter u is a "variable". This abuse of notation treats u as a function of x, y, and z and $\frac{\partial u}{\partial x}$ as the function $\partial_1 g$.

Example 4.1.6 Define $g : \mathbb{R}^3 \to \mathbb{R}$ by $g(x, y, z) = xz^3 + \sin y$ as in Example 4.1.2. If you write $u = g(x, y, z)$ then the identity $\frac{\partial u}{\partial x} = z^3$ is formally understood to mean $\partial_1 g(x, y, z) = z^3$.

A "variable" u is not a formal mathematical object; it can be formally interpreted as a function g with a domain and codomain. When using the chain rule, replacing functions with variables is helpful, because a variable can denote several different functions at once! Rather than formally explain this distinction, it may be easier to study an example.

Example 4.1.7 You can rewrite Example 4.1.5 with Leibniz notation. Define the variables

$$u = g(x, y, z) \qquad \text{and} \qquad (x, y, z) = f(s, t)$$

where $g : \mathbb{R}^3 \to \mathbb{R}$ and $f : \mathbb{R}^2 \to \mathbb{R}^3$ are differentiable functions. The second equation implies that $x = f_1(s,t)$, $y = f_2(s,t)$, and $z = f_3(s,t)$. Notice the letters x, y, z are being abused as both inputs of functions and as functions themselves. On one hand, you may write

$$\frac{\partial u}{\partial y} \quad \text{instead of} \quad \partial_2 g$$

so the letter y here denotes the second partial. On the other hand, you may write

$$\frac{\partial y}{\partial s} \quad \text{instead of} \quad \partial_1 f_2$$

since $y = f_2(s,t)$. This abuse of notation is acceptable because there is no other way to interpret these partial derivatives. Now, here is the magic. You can treat u as a function of x, y, z as well as a function of s and t! That is, you may think of u as satisfying

$$u = g \circ f(s,t).$$

Sometimes you will abuse notation even further and regard the above as equivalent to

$$u = g(x(s,t), y(s,t), z(s,t)).$$

This has no formal meaning but if you accept this equivalence, then you may write

$$\frac{\partial u}{\partial s} \qquad \text{instead of} \qquad \partial_1(g \circ f) = \partial_1 h.$$

Now, suppose you accept all this equivalent notation. Then you may rewrite (4.1.2) as

$$\begin{aligned} \frac{\partial u}{\partial s} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s} \\ \frac{\partial u}{\partial t} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial t} \end{aligned} \tag{4.1.3}$$

Now that is elegant and it is precisely what makes this notation worthwhile. The "fractions" appear to "cancel" so you can quickly see if you have written down a sensible identity.

Example 4.1.8 Consider Example 4.1.2 again in Leibniz notation. Then $u = xz^3 + \sin y$ so

$$\frac{\partial u}{\partial x} = z^3, \quad \frac{\partial u}{\partial y} = \cos y, \quad \frac{\partial u}{\partial z} = 3xz^2$$

and $(x, y, z) = (s^2 t, s^2 + t^2, \cos t)$ so

$$\begin{aligned} &\frac{\partial x}{\partial s} = 2st &&\frac{\partial y}{\partial s} = 2s &&\frac{\partial z}{\partial s} = 0 \\ &\frac{\partial x}{\partial t} = s^2 &&\frac{\partial y}{\partial t} = 2t &&\frac{\partial z}{\partial t} = -\sin t. \end{aligned}$$

Inserting these in (4.1.3), you find that

$$\frac{\partial u}{\partial s} = 2stz^3 + 2s\cos y, \qquad \frac{\partial u}{\partial t} = s^2 z^3 + 2t\cos y - 3xz^2 \sin t.$$

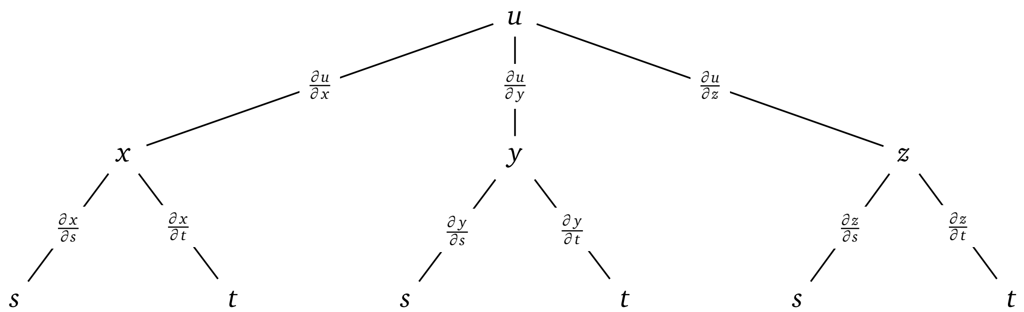
It may seem strange to have x, y, z remaining in these expressions but remember you can replace these with the variables s and t. For concreteness, you can check these relations a specific point. When $(s,t) = (1,0)$, you get that $(x,y,z) = f(1,0) = (0,1,-1)$ and so

$$\frac{\partial u}{\partial s} = 2\cos 1 \qquad \text{and} \qquad \frac{\partial u}{\partial t} = -1.$$

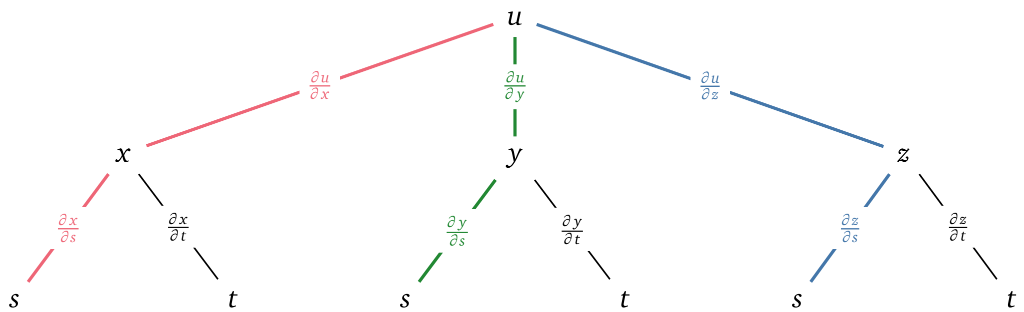
This matches (4.1.1) but was appreciably simpler to calculate.

You may already notice the ease of calculating with Leibniz notation. Another great advantage is that you can remember the chain rule using a mnemonic visual: trees!

Example 4.1.9 Consider again Example 4.1.7 so $u = g(x,y,z)$ and $(x,y,z) = f(s,t)$ where $g : \mathbb{R}^3 \to \mathbb{R}$ and $f : \mathbb{R}^2 \to \mathbb{R}^3$ are differentiable functions. The dependencies between the variables can be naturally expressed in a tree drawing.



If you want to calculate $\dfrac{\partial u}{\partial s}$, you collect the contribution from each branch ending with s.



Each branch contributes the product of the partials along each s-branch. Adding these contributions up gives

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s},$$

which is precisely the expression for $\frac{\partial u}{\partial s}$ in (4.1.3). You can do the same for $\frac{\partial u}{\partial t}$. This visual operation corresponds exactly to multiplying a row in one Jacobian matrix by a column in another Jacobian matrix.

Overall, you may find this abuse of notation to be quite confusing at first, but after enough practice you will find it to be remarkably convenient for applying the chain rule.

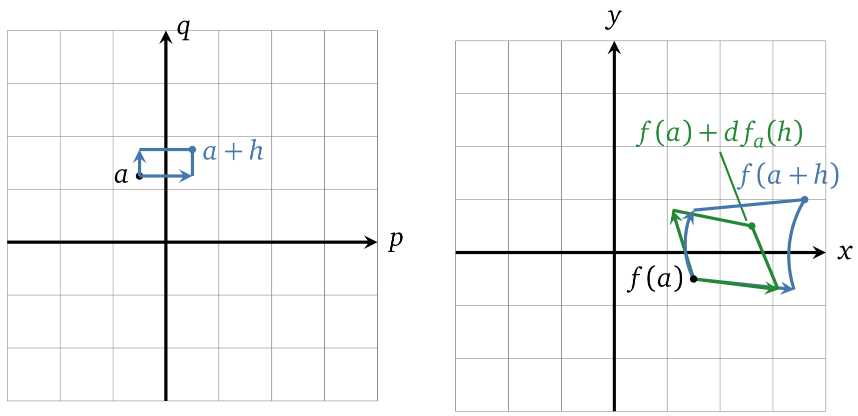
4.1.3 Proof of the chain rule

The formal proof of the chain rule is not easy, but the core ideas are natural.

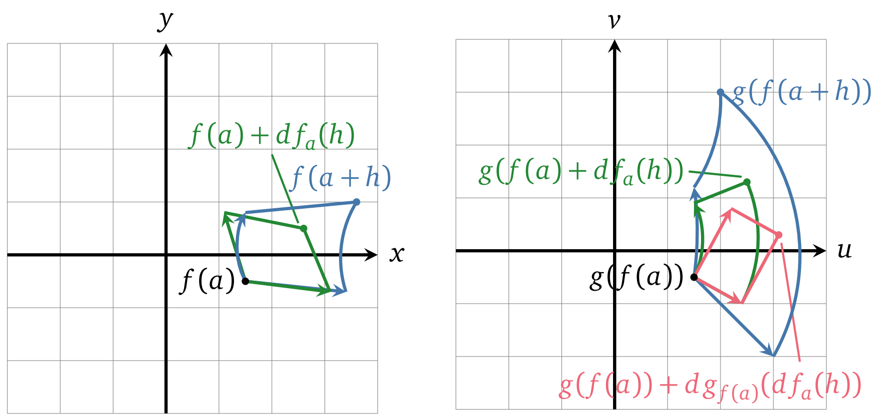
Example 4.1.10 Consider a composition of maps $\mathbb{R}^2 \to \mathbb{R}^2 \to \mathbb{R}^2$, namely

$$(x,y) = f(p,q) = (p^2+q^2, pq) \qquad (u,v) = g(x,y) = (x\cos y, x\sin y),$$

The functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^2$ are both C^1 and hence differentiable at any given $(p,q) \in \mathbb{R}^2$. What does it visually mean for $g \circ f$ to be differentiable at $a = (p,q)$? Fix $h \in \mathbb{R}^2$ small and consider how $a+h$ maps through the three planes: (p,q)-plane, (x,y)-plane, and (u,v)-plane. Start with f mapping from the (p,q)-plane to the (x,y)-plane.



The point $a+h$ is mapped *non-linearly* to $f(a+h)$ on the blue curved rectangle in the (x,y)-plane, and its *linear* approximation $f(a)+df_a(h)$ appears on the green parallelogram. Since f is differentiable at a, the distance between these two points should be small. Now, iterate this procedure with g from the (x,y)-plane to the (u,v)-plane.



Notice both the green and blue curves are mapped *non-linearly* by g to the (u,v)-plane. The point $f(a+h)$ is mapped non-linearly to $g(f(a+h))$ on the blue curved rectangle in the (u,v)-plane. Since $f(a+h)$ is close to its linear approximation $f(a)+df_a(h)$, the green point $g(f(a)+df_a(h))$ should be close to the blue point $g(f(a+h))$ because g is continuous at $f(a)$. On the other hand, if you use the *linear* approximation of g to map $f(a)+df_a(h)$, then you obtain the red point $g(f(a))+dg_{f(a)}\circ df_a(h)$. Since g is differentiable at $f(a)$, this red point should be close to the green point $g(f(a)+df_a(h))$.

Here is an informal summary of this reasoning with the points in the (u,v)-plane.

- The (non-linear non-linear) blue point is close to the (non-linear linear) green point by differentiability of f and continuity of g.

- The (non-linear linear) green point is close to the (linear linear) red point by differentiability of f and differentiability of g.

Both errors shrink faster than $||h||$ as $h \to 0$, so the overall error shrinks faster than $||h||$ as $h \to 0$. This suggests $d(g \circ f)_a = dg_{f(a)} \circ df_a$ as you would expect. The ideas in this heuristic argument will genuinely appear in the proof itself; in particular, the "green point" will act as an intermediate approximation. View the Math3D demo for further visuals.

Finally, here is the proof of the chain rule. It is quite clever.

Proof of Theorem 4.1.1. Since $a \in U$ and $f(a) \in V$ are interior points of U and V respectively, there exists $\delta_f > 0$ and $\delta_g > 0$ such that $B_{\delta_f}(a) \subseteq U$ and $B_{\delta_g}(f(a)) \subseteq V$. Since f is differentiable at a and g is differentiable at $f(a)$, the differentials df_a and $dg_{f(a)}$ both exist. Define the functions $\varepsilon_f : B_{\delta_f}(0) \to \mathbb{R}^m$ and $\varepsilon_g : B_{\delta_g}(0) \to \mathbb{R}^k$ by

$$\begin{aligned} &\forall v \neq 0, \quad \varepsilon_f(v) = \frac{f(a+v) - f(a) - df_a(v)}{||v||} \quad \text{and} \quad \varepsilon_f(0) = 0, \\ &\forall w \neq 0, \quad \varepsilon_g(w) = \frac{g(f(a)+w) - g(f(a)) - dg_{f(a)}(w)}{||w||} \quad \text{and} \quad \varepsilon_g(0) = 0. \end{aligned}$$

Since f is differentiable at a and g is differentiable at $f(a)$, it follows that ε_f and ε_g are continuous at 0; that is,

$$\lim_{v \to 0} \varepsilon_f(v) = 0 \quad \text{and} \quad \lim_{w \to 0} \varepsilon_g(w) = 0. \tag{4.1.4}$$

Now, you must show that

$$\lim_{v \to 0} \frac{(g \circ f)(a+v) - g(f(a)) - dg_{f(a)} \circ df_a(v)}{||v||} = 0. \tag{4.1.5}$$

By definition of ε_f, observe that

$$(g \circ f)(a+v) = g(f(a+v)) = g\big(f(a) + df_a(v) + ||v||\varepsilon_f(v)\big)$$

By definition of ε_g with $w = \eta(v) = df_a(v) + ||v||\varepsilon_f(v)$, it follows by linearity of $dg_{f(a)}$ that

$$\begin{aligned} (g \circ f)(a+v) &= g(f(a)) + dg_{f(a)}\big(\eta(v)\big) + \big|\big|\eta(v)\big|\big|\varepsilon_g\big(\eta(v)\big) \\ &= g(f(a)) + dg_{f(a)} \circ df_a(v) + ||v||dg_{f(a)}\big(\varepsilon_f(v)\big) + \big|\big|\eta(v)\big|\big|\varepsilon_g\big(\eta(v)\big) \end{aligned}$$

To deduce (4.1.5), it therefore suffices to prove that

$$\lim_{v \to 0} dg_{f(a)}(\varepsilon_f(v)) = 0 \quad \text{and} \quad \lim_{v \to 0} \frac{||\eta(v)||}{||v||}\varepsilon_g(\eta(v)) = 0.$$

The first limit follows from (4.1.4) and since $dg_{f(a)}$ is a linear map and hence continuous at 0 with $dg_{f(a)}(0) = 0$. For the second limit, notice $\eta(v) = df_a(v) + ||v||\varepsilon_f(v) \to 0$ as $v \to 0$ by (4.1.4) and continuity of the linear map df_a. Since ε_g is continuous at 0, it follows by Theorem 2.6.21 that

$$\lim_{v \to 0} \varepsilon_g(\eta(v)) = \varepsilon_g(\eta(0)) = \varepsilon_g(0) = 0.$$

Thus, by a version of the squeeze theorem, the second limit is equal to zero, provided the function $||\eta(v)||/||v||$ is bounded for v in a small punctured ball centered at 0.

Notice that

$$\frac{\eta(v)}{||v||} = \frac{1}{||v||} df_a(v) + \varepsilon_f(v) = df_a\Big(\frac{v}{||v||}\Big) + \varepsilon_f(v).$$

By (4.1.4), $\varepsilon_f(v)$ is bounded in $\mathbb{R}^m$ for v in a small open ball near 0. The unit sphere S^{n-1} in $\mathbb{R}^n$ is compact so its image under the (continuous) linear map $u \mapsto df_a(u)$ is compact and hence bounded in $\mathbb{R}^m$. Thus, the image of the map $df_a(v/||v||)$ for $v \neq 0$ is also bounded in $\mathbb{R}^m$. This implies that the image of $\eta(v)/||v||$ inside a small punctured ball centred at 0 is bounded so the scalar quantity $||\eta(v)||/||v||$ is bounded, as desired. This completes the proof. ■

Exercises for Section 4.1

Concepts and definitions

4.1.1 Let $F:\mathbb{R}^n\to\mathbb{R}^m$ and $G:\mathbb{R}^m\to\mathbb{R}^k$. Fix $a\in\mathbb{R}^n$. Assume F is differentiable at a and G is differentiable at $F(a)$. By the chain rule, $G\circ F$ is differentiable at a. Thus, for $v\in\mathbb{R}^n$, the following identity holds:

$$D(G\circ F)(a)v = d(G\circ F)_a(v) = (dG_{F(a)}\circ dF_a)(v) = DG(F(a))DF(a)v.$$

Two equalities follow from the same theorem, and one equality follows from another theorem. Identify which is which.

4.1.2 Let $F:\mathbb{R}^n\to\mathbb{R}^m$ and $G:\mathbb{R}^m\to\mathbb{R}^k$. Fix $a\in\mathbb{R}^n$. Which of the following are true?

(a) If F and G are differentiable, then $G\circ F$ is differentiable and $D(G\circ F)(a)=DG(a)DF(a)$.

(b) If F and G are differentiable, then $G\circ F$ is differentiable and $D(G\circ F)(a)=DG(F(a))DF(a)$.

(c) If F and G are C^1, then $G\circ F$ is C^1 and $D(G\circ F)(a)=DG(F(a))DF(a)$.

(d) If F and G are C^1, then $G\circ F$ is C^1 and $d(G\circ F)_a = dG_a\circ dF_a$.

(e) If F and G are C^1, then $G\circ F$ is differentiable and $d(G\circ F)_a = dG_{F(a)}\circ dF_a$.

Computations

4.1.3 Define $V=\{(x,y,z)\in\mathbb{R}^3 : z\neq 1\}$. Let $F:\mathbb{R}^2\to V$ and $G:V\to\mathbb{R}^2$ be given by

$$F(x,y)=\left(xy,\,2x-3y+7,\,y^2+2\right) \qquad\text{and}\qquad G(x,y,z)=\left(\frac{x}{1-z},\frac{y}{1-z}\right).$$

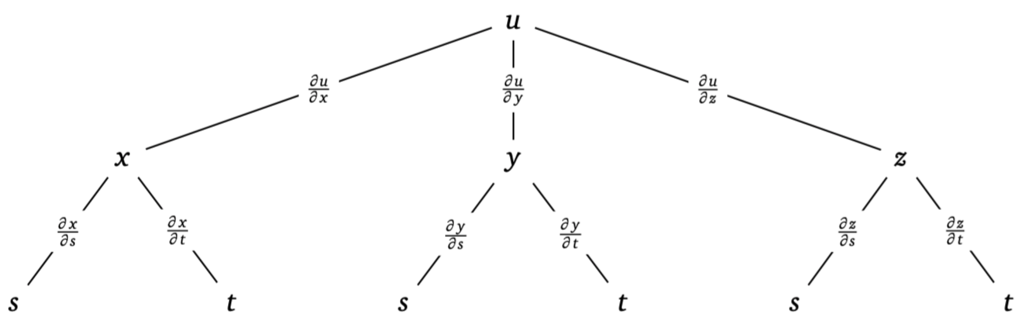
(a) Briefly explain why you can apply the chain rule to $G\circ F:\mathbb{R}^2\to\mathbb{R}^2$ at the point $(0,1)$.

(b) By a direct calculation, you can verify that

$$DF(x,y)=\begin{bmatrix} y & x\\ 2 & -3\\ 0 & 2y\end{bmatrix} \qquad DG(x,y,z)=\begin{bmatrix} \frac{1}{1-z} & 0 & \frac{x}{(1-z)^2}\\ 0 & \frac{1}{1-z} & \frac{y}{(1-z)^2}\end{bmatrix}$$

in their corresponding domains. Compute $d(G\circ F)_{(0,1)}(4,2)$ using the chain rule.

4.1.4 You can visualize the chain rule with "chain rule trees". Let $G:\mathbb{R}^3\to\mathbb{R}$ and $F:\mathbb{R}^2\to\mathbb{R}^3$. To distinguish your variables, write G as $G(x,y,z)$ and F as $F(s,t)$. Also write $u=(G\circ F)(s,t)$. Then you can compute $\dfrac{\partial u}{\partial s}$ and $\dfrac{\partial u}{\partial t}$ by using the following diagram:



(a) Use the above technique to write out a formula for $\dfrac{\partial u}{\partial s}$ and $\dfrac{\partial u}{\partial t}$.

(b) This method allows you to find the partial derivatives of $G \circ F$ when $G : \mathbb{R}^m \to \mathbb{R}$. How can you use this idea to find the partial derivatives of $G \circ F$ when $G : \mathbb{R}^m \to \mathbb{R}^k$ for any $k \in \mathbb{N}^+$?

(c) Let s, t be real variables so $s, t \in \mathbb{R}$. Define the variables u, v, x, y, z as follows:

$$(x, y, z) = \left(st, 2s - 3t + 7, t^2 + 2\right) \qquad \text{and} \qquad (u, v) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right).$$

i) Compute $\dfrac{\partial u}{\partial s}$ and $\dfrac{\partial u}{\partial t}$ when $(s, t) = (0, 1)$.

ii) Compute $\dfrac{\partial v}{\partial s}$ and $\dfrac{\partial v}{\partial t}$ when $(s, t) = (0, 1)$.

4.1.5 The chain rule has some conditions you need to verify before you can use it. But do you really need to check them? What can possibly go wrong if you do not? Let $F : \mathbb{R}^2 \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}^2$ be defined by

$$F(x, y) = \begin{cases} \dfrac{y^3}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases} \qquad \text{and} \qquad \gamma(t) = (t, t).$$

(a) Compute the $(F \circ \gamma)'(0)$ using the chain rule.

(b) Since $F \circ \gamma$ is a function from $\mathbb{R}$ to $\mathbb{R}$, you can compute its derivative directly. Find a formula for $(F \circ \gamma)(t)$ and use it to compute $(F \circ \gamma)'(0)$.

(c) Compare your calculations above. What happened? Explain why this might have occurred.

4.1.6 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinate transformation so $T(r, \theta) = (r\cos\theta, r\sin\theta)$ for all $(r, \theta) \in \mathbb{R}$. You will convert partial derivatives from Cartesian to polar in two equivalent ways.

(a) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable. Fix $(r, \theta) \in \mathbb{R}^2$. Using formal function notation, express the Jacobian of $f \circ T$ in terms of the partials of f, r, and θ.

(b) Equivalently, write $u = f(x, y)$ and $(x, y) = (r\cos\theta, r\sin\theta)$. Using Leibniz notation and chain rule trees, express $\dfrac{\partial u}{\partial r}$ and $\dfrac{\partial u}{\partial \theta}$ in terms of $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial y}$, r, and θ.

4.1.7 Define $f(x, y) = (3x + \sin y, x^2y^2 + x, e^{2x-y})$ and $g(x, y, z) = (xy - 2z, x - y + z)$.

(a) Calculate $D(f \circ g)(1, 0, -1)$.

(b) Calculate $D(g \circ f)(1, 0)$.

4.1.8 An ant is moving along a helical path $\gamma(t) = (4\cos t, 4\sin t, 3t)$. The temperature in space is given by $f(x, y, z) = x^2y^2 - z^2$ in Celsius.

(a) What is the velocity of the ant at $t = \pi$?

(b) At what rate does the ant detect the temperature to be changing at $t = 3\pi/2$?

4.1.9 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be differentiable. Write $u = f(x, y, z)$ and

$$(x, y, z) = (r\cos\theta, r\sin\theta, z).$$

Using Leibniz notation, express $\dfrac{\partial u}{\partial r}, \dfrac{\partial u}{\partial \theta}, \dfrac{\partial u}{\partial z}$ in terms of $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial y}, \dfrac{\partial u}{\partial z}$, r, θ, and z.

4.1.10 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be differentiable. Write $u = f(x, y, z)$ and

$$(x, y, z) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi).$$

Using Leibniz notation, express $\frac{\partial u}{\partial \rho}, \frac{\partial u}{\partial \theta}, \frac{\partial u}{\partial \phi}$ in terms of $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}$, ρ, θ, and ϕ.

Proofs

4.1.11 Remember that the gradient vector field of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ appears "orthogonal" to the level sets of f. The following lemma begins to formalize this idea.

> **Lemma.** *Let $S \subseteq \mathbb{R}^n$ be a level set of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Let $I \subseteq \mathbb{R}$ be an open interval. If $\gamma : I \to \mathbb{R}^n$ is differentiable at $t \in I$ and $\gamma(I) \subseteq S$, then*
>
> $$\nabla f(\gamma(t)) \cdot \gamma'(t) = 0.$$

Informally speaking, this says that

> *"The velocity of a particle on a level set of f is orthogonal to the gradient of f".*

(a) Sketch a picture illustrating the lemma using S as the unit sphere in $\mathbb{R}^3$.

(b) Prove the lemma using the chain rule.

4.1.12 Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open sets. Prove that if the function $f : U \to V$ is C^1 and the function $g : V \to \mathbb{R}^k$ is C^1, then their composition $g \circ f : U \to \mathbb{R}^k$ is C^1.

Applications and beyond

4.1.13 Recall derivatives can be viewed in four different contexts: algebraically, analytically, graphically, and physically. How can you interpret the chain rule in these different contexts?

(a) The algebraic viewpoint is:

The chain rule tells us how to linearly approximate $G \circ F$ at a using a composition of linear maps.

Write down these linear maps.

(b) The analytic viewpoint is:

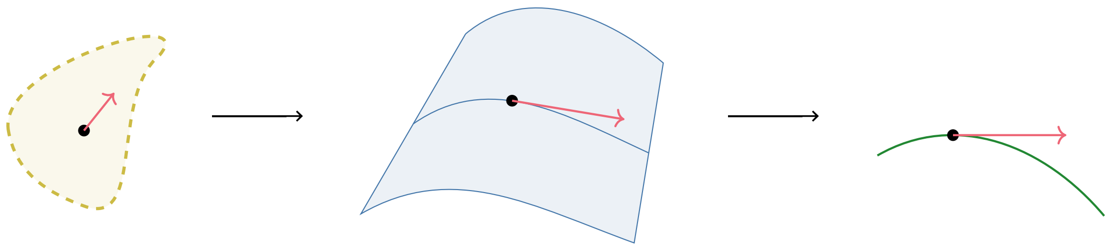
> *The chain rule tells us how to linearly approximate $G \circ F$ near a if we know how to approximate G near $F(a)$ and F near a.*

Write down the linear approximation of $G \circ F$ near a in terms of the differentials of G and F.

(c) The geometric interpretation is:

> *The chain rule tell us how composing functions transforms tangent vectors.*

The diagram below illustrates how tangent vectors are transformed via a composition of functions $G \circ F$. Label each quantity using the language of tangent vectors and differentials.



Hint: Use the labels $F, G, a, F(a), (G \circ F)(a), v, dF_a(v)$, and $d(G \circ F)_a(v)$

(d) A physical viewpoint is:

The chain rule tells us how our temperature changes as we move across the Earth's surface.

Write down the conclusion of the chain rule and give a physical interpretation for each quantity. *Hint:* Describe a path $\mathbb{R} \to \mathbb{R}^3$ and the temperature $\mathbb{R}^3 \to \mathbb{R}$.

4.2. Mean value theorem

The single-variable mean value theorem is a fundamental tool from single-variable calculus. It relates the derivative to the function itself, producing many fantastic results. While its generalization is not quite the same in multivariable calculus, it is a standard result for estimating the change in a function in terms of its derivative.

Theorem 4.2.1 (Mean value theorem) Let $U \subseteq \mathbb{R}^n$ be open and let $a, b \in U$. Let f be a differentiable real-valued function on U. If U contains the line segment L from a to b then there exists $c \in L$ such that

$$f(b) - f(a) = \nabla f(c) \cdot (b - a).$$

Its proof is an elegant application of the chain rule and your theory of derivatives.

Proof. Define $\gamma : [0,1] \to \mathbb{R}^n$ by $\gamma(t) = (1-t)a + tb$ for $0 \leq t \leq 1$, so $\gamma([0,1]) = L$ by definition. Since f is defined on U which contains L, we can define the single variable function $g : [0,1] \to \mathbb{R}$ by $g(t) = f(\gamma(t))$ for $0 \leq t \leq 1$.

Since γ is differentiable on $(0,1)$ and f is differentiable on U which contains L and hence $\gamma((0,1))$. It follows by the chain rule that g is differentiable on $(0,1)$. Since γ is continuous on $[0,1]$ and f is continuous on $L = \gamma([0,1])$, it follows that g is continuous on $[0,1]$.

By the single variable mean value theorem, there exists $s \in (0,1)$ such that

$$g'(s) = \frac{g(1) - g(0)}{1 - 0} = f(\gamma(1)) - f(\gamma(0)) = f(b) - f(a).$$

By the chain rule, $g'(s) = (f \circ \gamma)'(s) = \nabla f(\gamma(s)) \cdot \gamma'(s) = \nabla f(c) \cdot (b - a)$ where $c = \gamma(s) \in L$. ■

Here is a concrete example of the mean value theorem.

Example 4.2.2 Define the set $U = B_3(0,0)$ and the function $f : U \to \mathbb{R}$ as $f(x,y) = xy^2$ for all $(x,y) \in U$. Consider the points $(0,0)$ and $(1,2)$ contained in U connected by the line segment $L = \{(t, 2t) : t \in [0,1]\}$. First, since polynomials are always differentiable then f is differentiable. Further, the open ball U is open. Lastly, $L \subseteq U$ since

$$\sqrt{t^2 + (2t)^2} = \sqrt{5}|t| \leq \sqrt{5} < 3$$

for all $t \in [0,1]$. These observations imply by Theorem 4.2.1, there exists some $c \in L$ such that

$$f(1,2) - f(0,0) = \nabla f(c) \cdot ((1,2) - (0,0)).$$

It is not necessary but you can actually check that $c \in L$, as promised by Theorem 4.2.1.

Note, however, the mean value theorem does not hold for vector-valued functions.

Example 4.2.3 Define the function $f : \mathbb{R} \to \mathbb{R}^2$ by $f(t) = (\cos t, \sin t)$. Note that $f(0) = (1,0)$ and $f(\pi) = (-1,0)$ but there does *not* exist $c \in (0,\pi)$ such that

$$f(\pi) - f(0) = Df(c) \cdot (\pi - 0).$$

You can verify this directly since this equation states that $(-2,0) = (-\pi \sin c, \pi \cos c)$. If $\cos c = 0$ then $\sin c = \pm 1$ which shows no such c can satisfy this equation.

As with single variable calculus, the utility of the mean value theorem is almost entirely theoretical. This theorem (or the ideas behind its proof) can be used to prove a number of striking consequences. A couple are listed here.

Theorem 4.2.4 Let $U \subseteq \mathbb{R}^n$ be open and C^1 path-connected. Let $F : U \to \mathbb{R}^m$ be differentiable. The Jacobian $DF(x)$ is the $m \times n$ zero matrix for all $x \in U$ if and only if F is a constant map.

Proof. It suffices to prove this for each component of F, so you may assume F is real-valued. If F is constant, then it is immediate that $\nabla F(x) = 0 \in \mathbb{R}^n$ for all $x \in U$. Conversely, assume $\nabla F(x) = 0 \in \mathbb{R}^n$ for all $x \in U$. Fix $a, b \in U$ arbitrary. There exists a path $\gamma : [0,1] \to U$ such that γ is continuous on $[0,1]$, γ is C^1 on $(0,1)$, $\gamma(0) = a$, and $\gamma(1) = b$.

Define the map $f : [0,1] \to \mathbb{R}$ by $f(t) = F(\gamma(t))$ for all $t \in [0,1]$. Since γ and F are continuous, f is continuous on $[0,1]$. Since γ is differentiable on $(0,1)$ and F is differentiable, it follows that f is differentiable on $(0,1)$ by the chain rule. By the single variable mean value theorem, there exists $c \in (0,1)$ such that

$$f'(c) = f(1) - f(0) = F(\gamma(1)) - F(\gamma(0)) = F(b) - F(a).$$

On the other hand, by the chain rule,

$$f'(c) = Df(c) = D(F \circ \gamma)(c) = \nabla F(\gamma(c)) \cdot \gamma'(c) = 0,$$

so $F(b) = F(a)$. Since a, b were arbitrary, we conclude that F is constant. ■

Corollary 4.2.5 Let $U \subseteq \mathbb{R}^n$ be open and C^1 path-connected. Let $F : U \to \mathbb{R}^m$ and $G : U \to \mathbb{R}^m$ be differentiable. If $DF(x) = DG(x)$ for all $x \in U$, then there exists a constant $C \in \mathbb{R}^m$ such that $F(x) = G(x) + C$ for all $x \in U$.

Proof. This follows quickly from the previous result. ■

Comparing the proofs of Theorems 4.2.1 and 4.2.4, you may notice a recurring theme of ideas.

> *Create a path between points in the set. Produce a single variable function by composing the real-valued function with the path. Apply single variable calculus to this function.*

This nifty trick will be indispensable for many future applications of the derivative and beyond. Notice how the chain rule serves as a critical component for its success by translating change in one variable to change in many variables. The mean value theorem was your first application and you have many more to go.

Exercises for Section 4.2

Concepts and definitions

4.2.1 Fix $U \subseteq \mathbb{R}^n$ open. Let $a, b \in U$ and let L be the line segment between a and b. Let $f : U \to \mathbb{R}$ and $F : U \to \mathbb{R}^m$ be differentiable. Which of the following are true or false?

(a) There exists a $c \in U$ such that $f(b) - f(a) = \nabla f(c) \cdot (b - a)$.

(b) If U contains L then $\exists c \in L$ such that $f(b) - f(a) = \nabla f(c) \cdot (b - a)$.

(c) If U contains L then $\exists c \in L$ such that $f(b) = f(a) + Df(c)(b - a)$.

(d) If U contains L then $\exists c \in L$ such that $F(b) - F(a) = DF(c)(b - a)$.

Proofs

4.2.2 The mean value theorem is a tool for estimating the difference between two function values. Here you will use the mean value theorem to prepare lemmas for useful estimations. Let $U \subseteq \mathbb{R}^n$ be open and $a, b \in U$. Assume U contains the line segment L between a and b.

(a) Show if $f : U \to \mathbb{R}$ is differentiable then $\exists c \in L$ such that

$$|f(b) - f(a)| \leq ||\nabla f(c)|| \cdot ||b - a||.$$

Hint: Use the Cauchy-Schwarz inequality.

(b) Show if $f : U \to \mathbb{R}$ is C^1 then

$$|f(b) - f(a)| \leq \left(\max_{c \in L} ||\nabla f(c)|| \right) \cdot ||b - a||.$$

4.2.3 The mean value theorem does not generalize perfectly for vector-valued functions.

(a) If you were to guess a vector-valued MVT, then you might guess this:

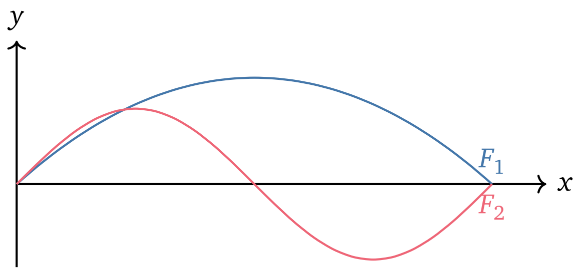
> **FALSE claim.** *Let $U \subseteq \mathbb{R}^n$ be open. If $F : U \to \mathbb{R}^m$ is a differentiable function, $a, b \in U$, and the line segment L between a and b lies in U, then $\exists c \in L$ such that*
>
> $$F(b) - F(a) = DF(c)(b - a).$$

Here is a sketch of a WRONG proof of this FALSE claim:

1. *Since F is differentiable, its real-valued component functions $F_1, \ldots, F_m$ are differentiable.*
2. *Fix $i \in \{1, \ldots, m\}$.*
3. *By the MVT on $F_i : U \to \mathbb{R}$, there exists $c \in L$ such that $F_i(b) - F_i(a) = \nabla F_i(c) \cdot (b - a)$.*
4. *Therefore, $F(b) - F(a) = DF(c)(b - a)$ as needed.*

Identify the fatal mistake in this proof.

(b) Below are the graphs of the two component functions of $F = (F_1, F_2) : \mathbb{R} \to \mathbb{R}^2$. Explain why this picture illustrates a counterexample to the above claim.



4.3. Local extrema and critical points

The story of optimization of real-valued functions continues in this chapter on derivatives. Thus far, you have established the extreme value theorem (Theorem 2.8.7) with topological ideas. It identifies the existence of global extrema and it is currently your only optimization tool. This addresses one fundamental question and now you must address a much harder one.

I know a solution to this optimization problem exists. How do I find possible solutions?

The answer will vary greatly with the precise nature of your optimization problem. Now that you possess a solid foundation in derivatives, you can make significant progress on this question for a very general class of such problems by developing your next optimization tool: the local extreme value theorem. This theorem will lead to a generalized notion of critical points from single variable calculus.

4.3.1 Local extreme value theorem

The notion of local extrema in single variable calculus generalizes nicely.

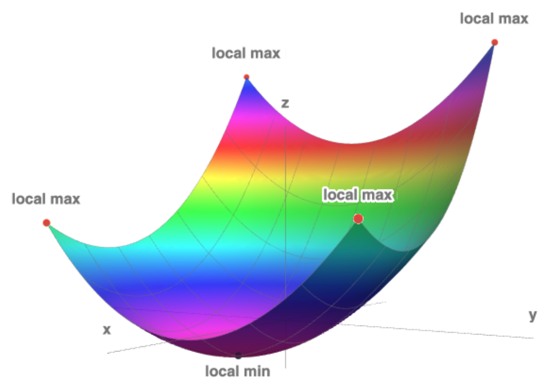
Definition 4.3.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. Let $a \in A$. The map f has a **local maximum** at a if there exists $\delta > 0$ such that $f(a) \geq f(x)$ for all $x \in A \cap B_\delta(a)$. If so, the value $f(a)$ is a **local maximum value** of f on A.

A **local minimum** and **local minimum value** are defined similarly. A **local extremum** is a local maximum or local minimum.

Remark 4.3.2 If you drop the adjective "local", then the meaning of the word is global. A maximum is always a global maximum. A minimum is always a global minimum. An extremum is always a global extremum.

Local extrema can exist on the interior or boundary of a domain. Moreover, every global extremum is a local extremum but the converse is not true. Here are a few examples.

Example 4.3.3 Consider the graph of the function $f(x, y) = x^2 - 2x + y^2$ on $A = [-2, 2]^2$. View this Math3D demo.

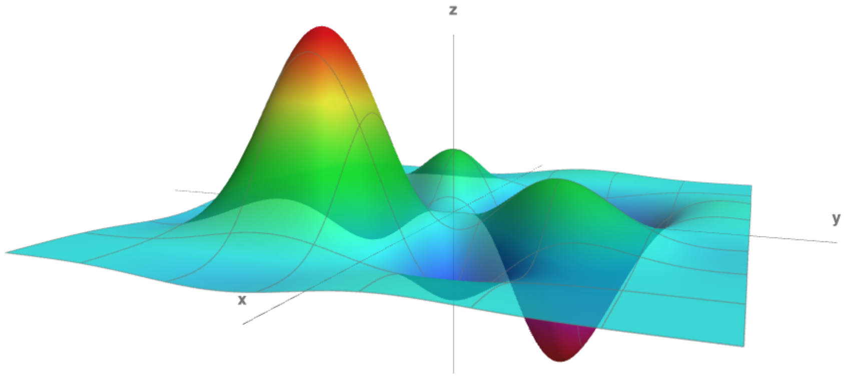


By carefully inspecting this graph, you might notice that f has a local minimum at $(1, 0)$ with value $f(1, 0) = -1$. This occurs on the interior of A and it is also the global minimum of f. On the other hand, you can guess that f attains a local maxima at $(2, 2), (-2, 2), (2, -2)$ and $(-2, -2)$ all of which are on the boundary of A. Notice $(-2, 2)$ and $(-2, -2)$ are global maxima of f whereas $(2, 2)$ and $(2, -2)$ are not.

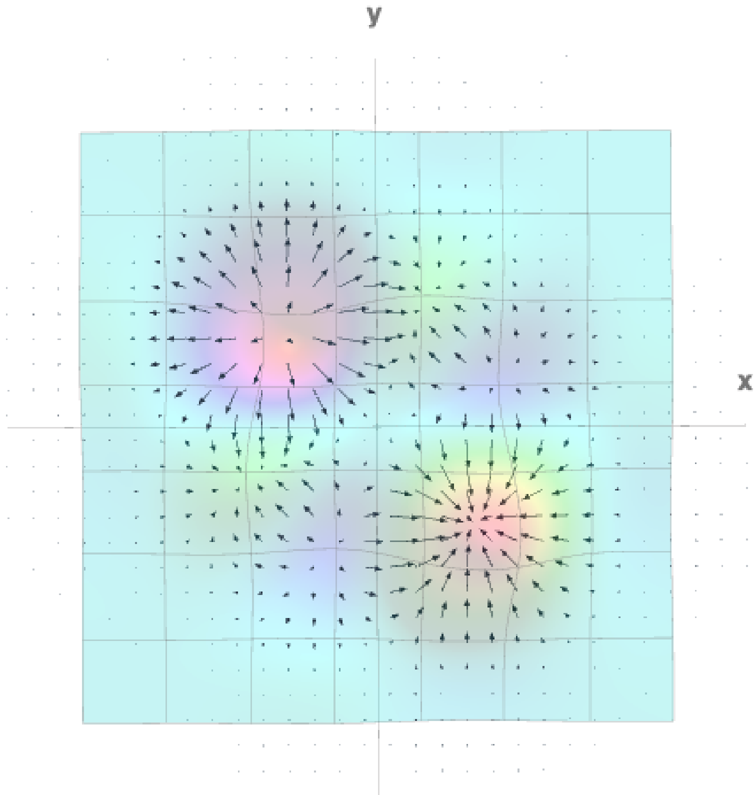
Example 4.3.4 Local extrema can be difficult to identify. For instance, consider the function

$$f(x,y) = \sin(x)\sin(y)e^{-x^2-y^2}(x+y).$$

View its on Math3D which is included below.

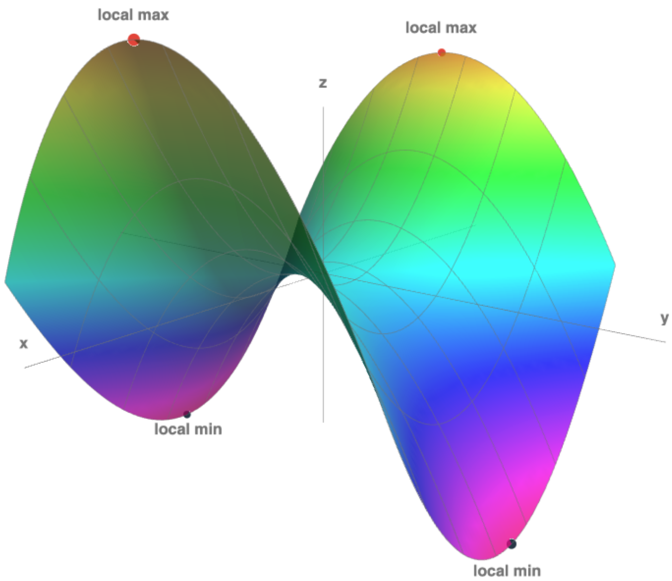


You can visibly see several local extrema in many spots, but actually finding them does not seem easy to do. The gradient vector field of f hints at how you might use derivatives to locate these extrema.

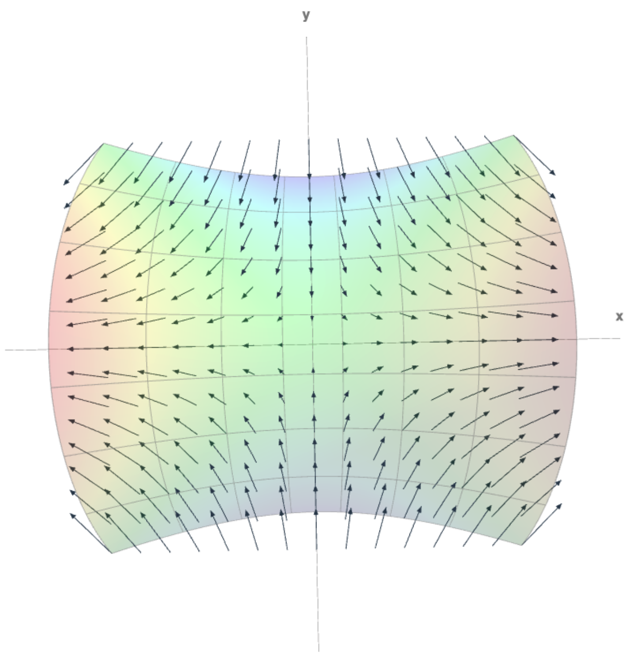


You can view the vector field alone on Math3D by toggling the surface. Notice the gradient appears to vanish at local extrema! This feature is reasonable since the gradient points in the direction of steepest ascent. However, if you look closely at the origin, the gradient vanishes there but it is not a local extrema.

Example 4.3.5 Consider the graph of the function $g(x,y) = x^2 - y^2$ on $A = [-2,2]^2$. View this Math3D demo.



Notice that g has no local extrema on the interior of A. Moreover, g attains a local maximum at $(2,0)$ and $(-2,0)$ and local minima at $(0,2)$ and $(0,-2)$. All of these occur on the boundary of A. Can you still observe these local extrema using the gradient vector field?



View this Math3D demo. Unlike the previous example, the gradient vector field does not give any information about these local extrema because they are not interior points. They lie on the boundary of the domain so the gradient does not need to vanish there. There is actually a worse issue. The gradient vanishes at the origin $(0,0)$, that is $\nabla f(0,0) = 0$, but this is *neither* a local maximum nor a local minimum! If you compare with the graph, the origin corresponds to what looks like a "saddle".

These examples lead to the discovery of a crucial result.

Theorem 4.3.6 (Local extreme value theorem) Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. If a is an interior point of A and f has a local extremum at a, then either $\nabla f(a) = 0$ or $\nabla f(a)$ does not exist.

Remark 4.3.7 Note that this gives no information on boundary points of A.

Proof. It suffices to prove that if $\nabla f(a)$ exists then $\nabla f(a) = 0$. Fix $j \in \{1, \ldots, n\}$ and let $\{e_1, \ldots, e_n\}$ denote the standard basis of $\mathbb{R}^n$. There exists $\varepsilon > 0$ such that the line segment $\{a + he_j : h \in (-\varepsilon, \varepsilon)\}$ lies inside A. Thus, we may define $g : (-\varepsilon, \varepsilon) \to \mathbb{R}$ by $g(h) = f(a + he_j)$. As $\nabla f(a)$ exists, the partial derivative $\partial_j f(a)$ exists so

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h} = \partial_j f(a)$$

and hence g is differentiable at 0. Since f has a local extremum at $a \in A$, it follows that g has a local extremum at $0 \in \mathbb{R}$. By the single variable local extreme value theorem, we have that $\partial_j f(a) = g'(0) = 0$. Since $j \in \{1, \ldots, n\}$ was arbitrary, we conclude that $\nabla f(a) = 0$. ■

Theorem 4.3.6 generalizes the single variable local extreme value theorem (see this MAT137 video). This constitutes a major achievement in optimization! You can narrow down your search for solutions to an optimization problem by identifying where the gradient vanishes.

4.3.2 Critical points

The local extreme value theorem suggests a generalization of critical points.

Definition 4.3.8 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. A point $a \in A$ is a **critical point** of f if a is an interior point of A and either $\nabla f(a) = 0$ or $\nabla f(a)$ does not exist.

Remark 4.3.9 Many texts define critical points to only be interior points a where $\nabla f(a) = 0$. Always check your source's conventions.

This definition and the local extreme value theorem yield an immediate consequence.

Lemma 4.3.10 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. If $a \in A$ is a local extremum of f, then either a is a boundary point of A or a is a critical point of f.

Proof. This is left as an exercise. It follows quickly from the local extreme value theorem. ■

Example 4.3.5 demonstrates that not every critical point is a local extremum, even when the gradient exists. These misleading points deserve a special name inspired by their shape.

Definition 4.3.11 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a real-valued function. An interior point $a \in A$ is a **saddle point** of f if[1] $\nabla f(a) = 0$ and f does not have a local extremum at a.

Similar to single variable calculus, critical points are "candidates for local extrema *at interior points*" but they are not always local extrema. This drawback is unavoidable. Derivatives possess local information for a function near a point, so they can measure "flatness" but not much more. Simply put, a derivative at a point can tell you if you are standing on a flat surface but it cannot tell you whether you are on a peak, a valley, or a saddle.

[1]Since $\nabla f(a)$ exists, it is necessary by Definition 3.2.2 that a is an interior point of A.

Example 4.3.12 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$. What are the critical points of f? First, you can calculate its partials:

$$\partial_1 f(x,y) = 6xy - 6x$$
$$\partial_2 f(x,y) = 3x^2 + 3y^2 - 6y.$$

Then, to find the critical points of f, you must find the points $(x,y) \in \mathbb{R}^2$ such that $\nabla f(x,y) = (0,0)$. In other words, you must solve for the system of (non-linear) equations

$$6xy - 6x = 0$$
$$3x^2 + 3y^2 - 6y = 0.$$

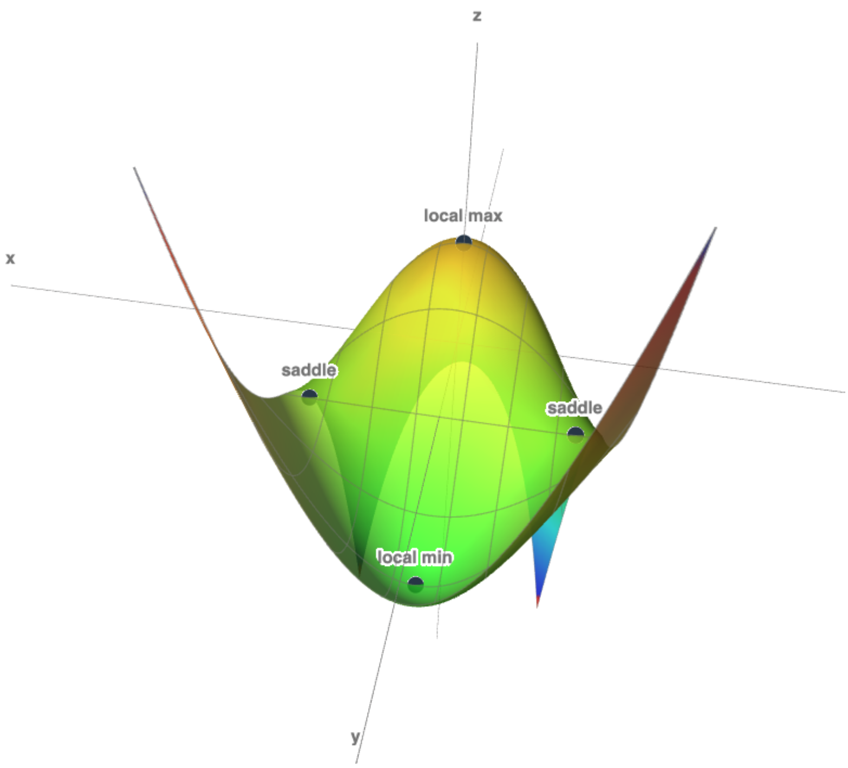
In general, solving non-linear systems of equations is extremely difficult but you can do so in simple examples like this situation. Here you can factor the first equation into $6x(y-1) = 0$, so $x = 0$ or $y = 1$. Consider these two cases separately. First, if $x = 0$, then the second equation implies that

$$3x^2 + 3y^2 - 6y = 0 \implies 3y(y-2) = 0 \implies y = 0 \text{ or } 2.$$

Otherwise, if $y = 1$, then the second equation implies that

$$3x^2 + 3y^2 - 6y = 0 \implies 3(x^2 - 1) = 0 \implies x = \pm 1.$$

Therefore, the critical points of f are $(0,0), (0,2), (1,1)$ and $(-1,1)$. You can view these on the graph of f in this Math3D demo.



From the graph, you can infer that $(0,0)$ is a local maximum, $(0,2)$ is a local minimum, and $(\pm 1,1)$ are saddle points. You do not yet have any techniques yet to classify critical points. Those results will be developed in a later chapter. For now, you can appreciate that you narrowed down the local extrema to only 4 possible points (and 2 of them are correct).

Remember a critical point may occur also when the gradient is not defined.

Example 4.3.13 Consider the following function $f : \mathbb{R}^n \to \mathbb{R}$:

$$f(x_1,\dots,x_n) = ||(x_1,\dots,x_n)|| = \sqrt{x_1^2+\cdots+x_n^2}.$$

To find its critical points, start by finding its partials. For $1 \le j \le n$ and $x \neq 0$,

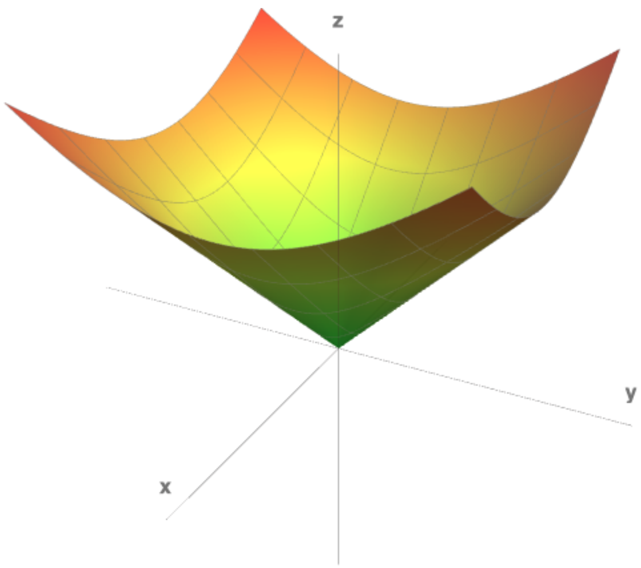
$$\partial_j f(x_1,\dots,x_n) = \frac{x_j}{\sqrt{x_1^2+\cdots+x_n^2}}$$

so the jth partial exists and is continuous on $\{x \in \mathbb{R}^n : x \neq 0\} = \mathbb{R}^n \setminus \{0\}$. Therefore, f is C^1 and hence differentiable on $\mathbb{R}^n \setminus \{0\}$. The calculations above imply that

$$\nabla f(x) = \frac{x}{||x||}.$$

Since the gradient is not defined at the origin, the point $x = 0$ is a critical point of f. Moreover, $\nabla f(x) \neq 0$ for any $x \in \mathbb{R}^n \setminus \{0\}$ so f does not have any critical points on the domain of ∇f. By the local extreme value theorem, f does not have any local extrema on $\mathbb{R}^n \setminus \{0\}$. The only possible local extrema of f is therefore at the critical point 0.

Indeed, the origin is in fact a local minimum of f. You can view this for the case $n = 2$ where the graph of f is a cone (Math3D).



These couple of examples mark the beginning of multivariable optimization. Equipped with the global and local extreme value theorems, you can begin solving unconstrained optimization problems. This will open up an entire world of applications.

Exercises for Section 4.3

Concepts and definitions

4.3.1 Let $A \subseteq \mathbb{R}^n$ be a set. Let $f : A \to \mathbb{R}$ and let $a \in A$.
Which of the following statements are equivalent to "f has a local maximum a"?

(a) $\forall x \in A,\ f(x) \le f(a)$.

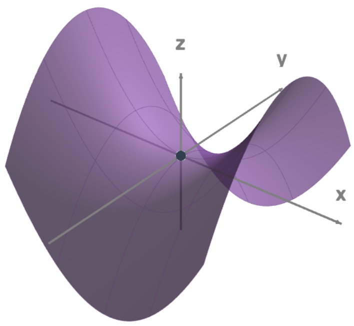
(b) $\exists \varepsilon > 0$ s.t. $\forall x \in B_\varepsilon(a),\ f(x) \le f(a)$.

(c) $\exists \varepsilon > 0$ s.t. $\forall x \in B_\varepsilon(a) \cap A,\ f(x) \le f(a)$.

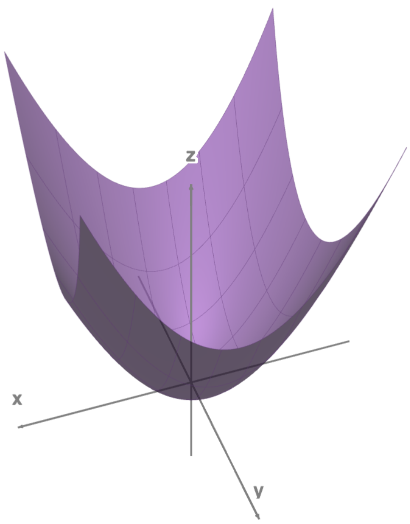
(d) $\exists \varepsilon > 0$ s.t. $\forall x \in B_\varepsilon(a) \cap A,\ |f(x)| \le |f(a)|$.

4.3.2 Below are contour plots or graphs of a C^1 function $\mathbb{R}^2 \to \mathbb{R}$. Identify whether at the origin the function has a local maximum, local minimum, saddle point, or none of these.

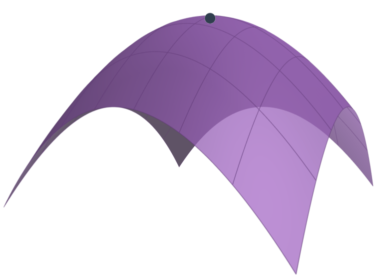
(a)



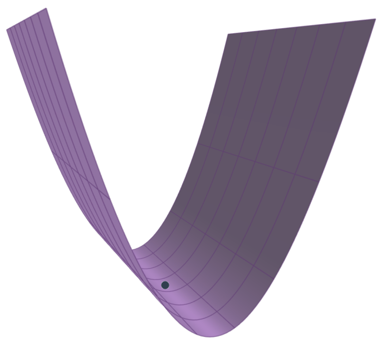
(b)



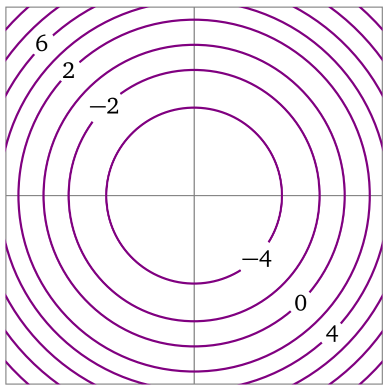
(c)



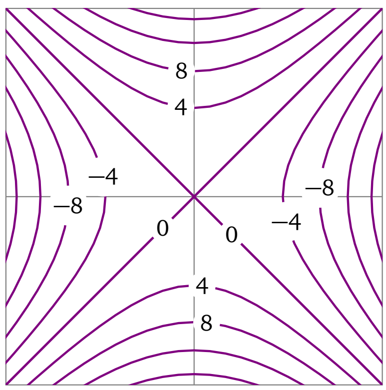
(d)



(e)

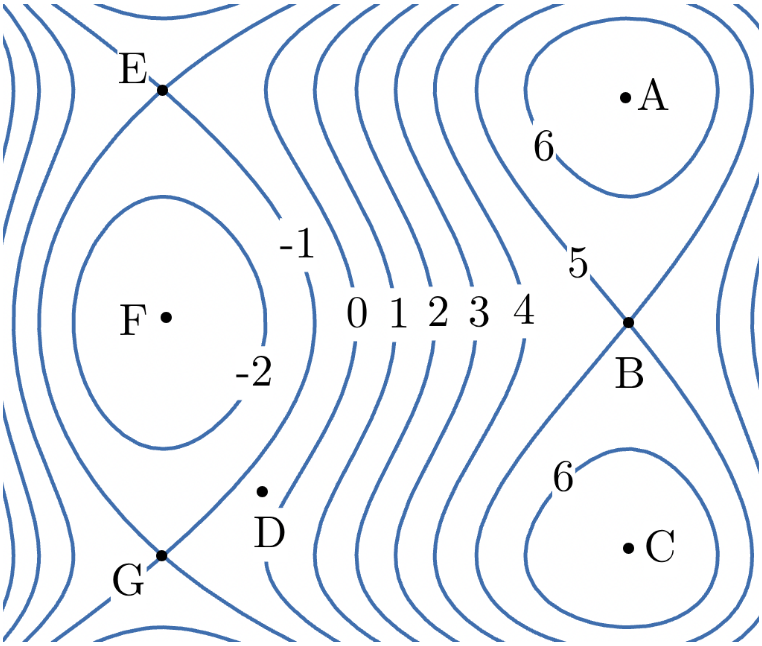


(f)

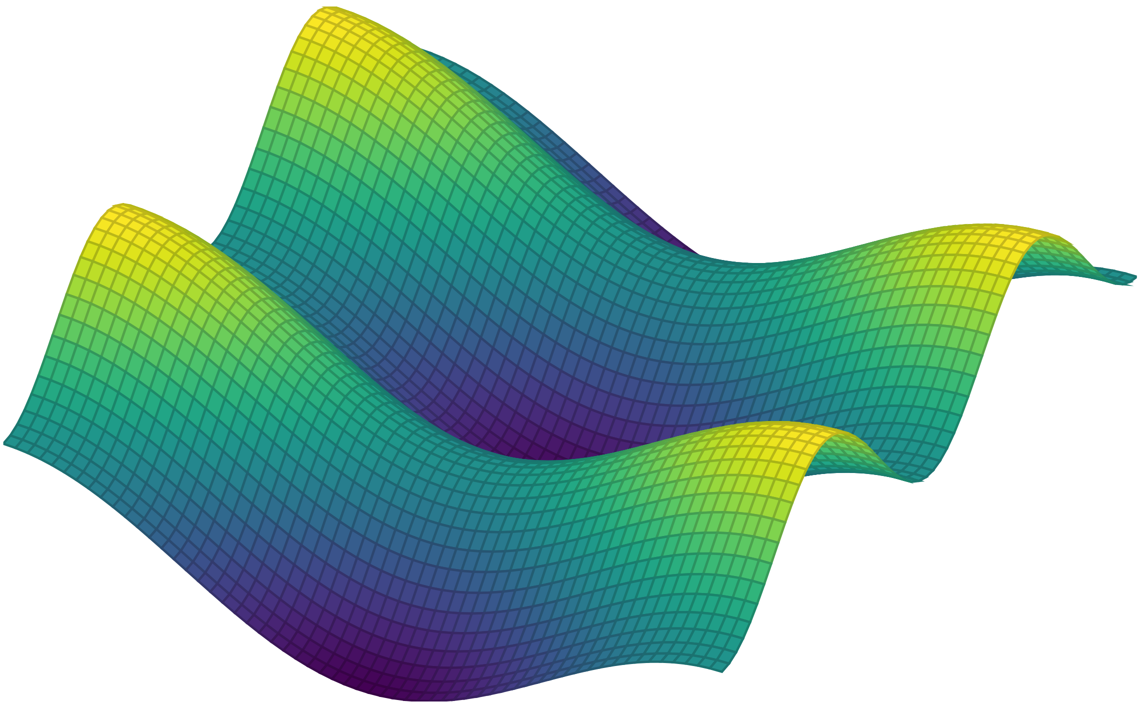


4.3.3

(a) Identify which points appear to be critical points, local maxima, local minima, or saddle points.



(b) Let D be a closed rectangle in $\mathbb{R}^2$. Consider the graph of $f : D \to \mathbb{R}$. Label all of the local minima, local maxima, saddle points, and critical points of f on D. (At least the ones you can see.)



4.3.4 Let $A \subseteq \mathbb{R}^n$ be open. Let $f : A \to \mathbb{R}$ be differentiable. Fix $a \in A$. Which statements are true?

(a) If f has a local extremum at a then $\nabla f(a) = 0$.

(b) If $\nabla f(a) = 0$ then a is a local extremum of f.

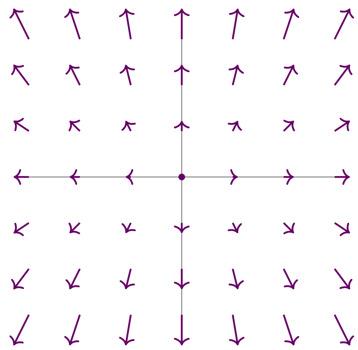
(c) If $\nabla f(a) = 0$ then a is either a local extremum of f or a saddle point of f.

(d) Every global extremum of f is a local extremum.

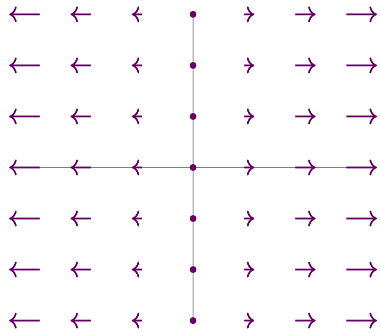
(e) Every saddle point of f is a critical point.

(f) Every local extremum of f is a critical point. *Hint:* The answer differs if A were not open.

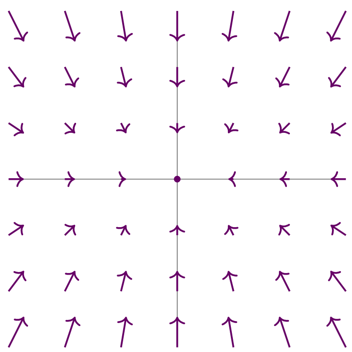
4.3.5 Below are six vector fields. Each vector field is the gradient of some C^1 function $f : \mathbb{R}^2 \to \mathbb{R}$. For each vector field, identify all critical points and classify whether they appear to be local maxima, local minima, saddle points, or none of these.



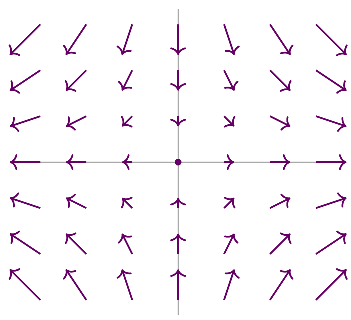
(a)



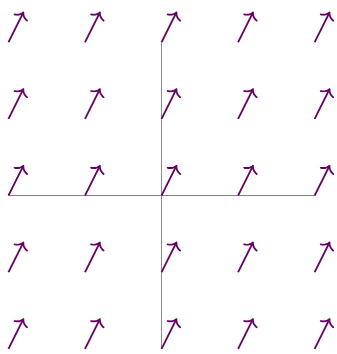
(d)



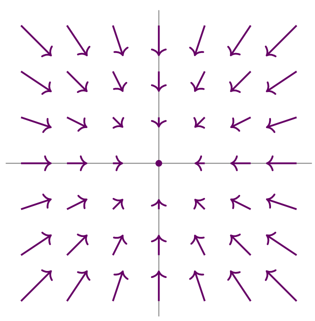
(b)



(e)



(c)



(f)

4.3.6 Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ with a saddle point at 0.

4.3.7 Give an example of a function $g : \mathbb{R}^2 \to \mathbb{R}$ with a saddle point at $(0,0)$.

Computations

4.3.8 Compute the critical points of each function on the given domain.

(a) Find the critical points of $f(x,y) = x^2 - 2xy + 3y^2 - 8y$.

(b) Find the critical points of $g : [-2,2]^2 \to \mathbb{R}$ given by $g(x,y) = x^2 - 2xy + 3y^2 - 8y$.

4.3.9 Find the critical points of $g(x,y) = 2xy + \dfrac{1}{x} + \dfrac{1}{y}$.

4.3.10 Find the critical points of $f(x,y,z) = (x^2+y^2+z^2)e^{3x^2+2y^2+z^2}$.

4.3.11 Find the critical points of $F : \mathbb{R}^n \to \mathbb{R}$ given by $F(x) = e^{-||x-a||^2}$ for some fixed $a \in \mathbb{R}^n$.

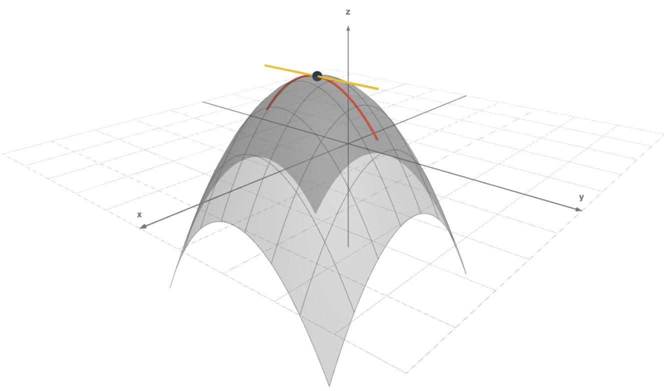
4.3.12 Find the critical points of $G : \mathbb{R}^n \to \mathbb{R}$ given by $G(x) = ||x-a||$ for some fixed $a \in \mathbb{R}^n$.

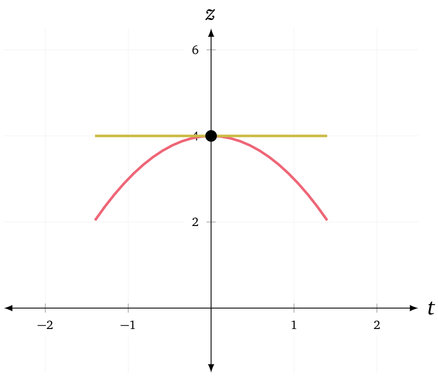
Proofs

4.3.13 The proof of the local extreme value theorem (Theorem 4.3.6) brushes over some technicalities. Here is a condensed proof. You will synthesize the main ideas and add important details.

1. *It suffices to prove that if $\nabla f(a)$ exists then $\nabla f(a) = 0$. Assume $\nabla f(a)$ exists.*
2. *Fix $j \in \{1,\dots,n\}$. Let $\{e_1,\dots,e_n\}$ be the standard basis for $\mathbb{R}^n$.*
3. *There exists $\varepsilon > 0$ such that $\{a + te_j : t \in (-\varepsilon,\varepsilon)\}$ is contained in A.*
4. *Thus, we can define the function $g : (-\varepsilon,\varepsilon) \to \mathbb{R}$ by $g(t) = f(a+te_j)$.*
5. *Notice $g'(0) = \lim\limits_{t\to 0} \dfrac{g(t)-g(0)}{t} = \lim\limits_{t\to 0} \dfrac{f(a+te_j)-f(a)}{t} = \partial_j f(a)$.*
6. *Since $\partial_j f(a)$ exists, this implies $g'(0)$ exists and hence g is differentiable at 0.*
7. *Since f has a local extremum at $a \in A$, it follows that g has a local extremum at $0 \in \mathbb{R}$.*
8. *By the single-variable local extreme value theorem, it follows that $\partial_j f(a) = g'(0) = 0$.*
9. *Since $j \in \{1,\dots,n\}$ was arbitrary, we conclude that $\nabla f(a) = 0$.*

(a) Below is a picture proof with $n = 2$ including a graph of f in $\mathbb{R}^3$ and a slice of the graph.





i) Label the lefthand Math3D demo using f and other quantities in the proof.

ii) Label the righthand figure using g and other quantities in the proof.

(b) Line 1 assumes $\nabla f(a)$ exists. Why can the author ignore when $\nabla f(a)$ does not exist?

(c) Line 3 uses an assumption without mention. Identify that assumption and justify Line 3.

(d) Line 6 claims $\partial_j f(a)$ exists. Briefly explain why.

(e) Line 7 is not justified. Justify it with a proof by definition of local extrema.

4.3.14 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable. Prove if the 237-level set of f is $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 4\}$ then f has at least one local extremum.

Applications and beyond

4.3.15 Simple linear regression is a statistical method for finding the most plausible linear relationship between one dependent variable $y \in \mathbb{R}$ and one independent variable $x \in \mathbb{R}$. If you have n data pairs $(x_1, y_1), \ldots, (x_n, y_n)$ in $\mathbb{R}^2$, you want to search for an optimal choice of parameters $\alpha, \beta \in \mathbb{R}$ such that

$$E(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

is minimized. Once you minimize it, the "line of best fit" for the data is therefore given by $y = \alpha + \beta x$. Show that if (α, β) is a critical point of E then

$$\begin{aligned} \Big(\sum_{i=1}^{n} 1\Big)\alpha + \Big(\sum_{i=1}^{n} x_i\Big)\beta &= \sum_{i=1}^{n} y_i \\ \Big(\sum_{i=1}^{n} x_i\Big)\alpha + \Big(\sum_{i=1}^{n} x_i^2\Big)\beta &= \sum_{i=1}^{n} x_i y_i \end{aligned}$$

4.4. Optimization

You have established two fundamental tools in multivariable optimization: the extreme value theorem and the local extreme value theorem. Amazingly, with these tools alone you can already solve many optimization problems with a wide range of applications. Due to the incalculable diversity of scenarios, there is no specific strategy that will always succeed. Nonetheless, the basic ingredients are simple.

> *Determine whether global extrema must exist. Identify the critical points on the interior of the domain. Check the boundary for possible extrema. Justify your conclusions.*

The first ingredient can vary in difficulty and may require a variant of the extreme value theorem. The second ingredient is a mechanical process, but that does not mean it is easy to do. The third and fourth step strongly depends on your function and its domain. The fourth step requires a careful synthesis of all the previous steps.

Some creativity is often required and the process may feel ad hoc. This section will describe the approach in some selected examples which illustrate the key ideas. You are encouraged to focus on these core principles and justifications, rather than trying to categorize the optimization problems. The first example begins with one of the most common scenarios: a differentiable function on a compact domain.

Example 4.4.1 You can optimize the function

$$f(x,y) = x^2 + y^2 + 4x - 6y,$$

on the closed disk $A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 16\}$. Since f is continuous and A is compact, the extreme value theorem implies that f attains a global maximum and global minimum on A. A global extremum of f will either belong to

$$A^{\circ} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 16\} \quad \text{or} \quad \partial A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 16\},$$

so you must divide into cases.

If p is a global extremum (and hence a local extremum) in the interior of A, then since f is differentiable on $\mathbb{R}^2$ it follows by the local extreme value theorem that $\nabla f(p) = (0,0)$. In other words, you must locate the critical points of f belonging to A. If a point $(x,y) \in \mathbb{R}^2$ satisfies $\nabla f(x,y) = (0,0)$, then

$$\begin{aligned} &(\partial_1 f(x,y), \partial_2 f(x,y)) = (0,0) \\ \Longrightarrow\ &(2x+4, 2y-6) = (0,0) \\ \Longrightarrow\ &x = -2, y = 3. \end{aligned}$$

Therefore the only critical point is $(-2,3)$ and $(-2)^2 + 3^2 < 16$ so this point lies inside A°.

Next, you must locate possible extrema of f on the boundary ∂A. This circle of radius 4 can be parametrized by

$$\gamma(t) = (4\cos t, 4\sin t), \qquad 0 \leq t \leq 2\pi,$$

so you must optimize the single variable function $g : [0, 2\pi] \to \mathbb{R}$ defined by

$$g(t) = f \circ \gamma(t) = (4\cos t)^2 + (4\sin t)^2 + 4(4\cos t) - 6(4\sin t) = 16 + 16\cos t - 24\sin t.$$

For $t \in (0, 2\pi)$, since $g'(t) = -16\sin t - 24\cos t$, you can see that $g'(t) = 0$ implies that

$$\begin{aligned} & -16\sin t - 24\cos t = 0 \\ \implies & \cos t = 0 \quad \text{or} \quad \tan t = -\frac{3}{2} \\ \implies & t = \frac{\pi}{2}, \frac{3\pi}{2} \quad \text{or} \quad t = \pi + \arctan(-\tfrac{3}{2}), 2\pi + \arctan(\tfrac{-3}{2}). \end{aligned}$$

It is also possible that the endpoints $t = 0$ and $t = 2\pi$ are extrema of g. By standard trigonometric identities, you can verify that all 6 of these values of t corresponds to the points $\gamma(t)$ given by

$$(0,4), \quad (0,-4), \quad \left(\frac{-8}{\sqrt{13}}, \frac{12}{\sqrt{13}}\right), \quad \left(\frac{8}{\sqrt{13}}, \frac{-12}{\sqrt{13}}\right), \quad (4,0).$$

These points correspond to all possible extrema for f on the ∂A.

All that remains is to compare the values of f at all possible extrema on the interior and the boundary of A. You can verify that

$$f(-2,3) = -13, \quad f(0,4) = -8, \quad f(0,-4) = 40$$

$$f(\tfrac{-8}{\sqrt{13}}, \tfrac{12}{\sqrt{13}}) \approx -12.8, \quad f(\tfrac{8}{\sqrt{13}}, \tfrac{-12}{\sqrt{13}}) \approx 44.8, \qquad f(4,0) = 32$$

This implies that the global minimum of f on A is $f(-2,3) = -13$, and the global maximum is $f(\frac{8}{\sqrt{13}}, \frac{-12}{\sqrt{13}}) \approx 44.8$. You can visually confirm this by plotting the graph of f.

View this Math3D demo for a better visual.

The previous example outlined a basic optimization strategy. Find the critical points on the interior. Parametrize the boundary and solve an optimization problem on a lower dimensional domain. Luckily, the boundary was fairly simple and could be parametrized with a single curve. The boundary can quickly get more complicated and you will need to divide-and-conquer.

Example 4.4.2 You can find the global extrema of

$$f(x,y) = x^2 - 2xy + 4y^2 - 4x - 2y + 24$$

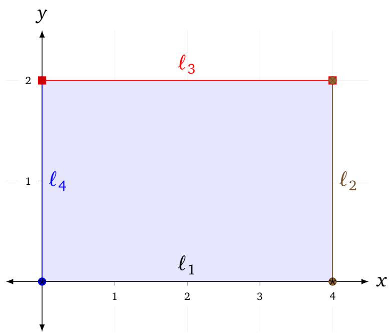
on the rectangle $A = [0,4] \times [0,2]$, namely on $0 \le x \le 4$, $0 \le y \le 2$. Since f is continuous and A is compact, the extreme value theorem implies that f attains a global maximum and global minimum. A global extremum of f will either belong to the interior or the boundary of A, so you must divide into cases.

If p is a global extremum (and hence a local extremum) in the interior of A, then since f is differentiable on $\mathbb{R}^2$ it follows by the local extreme value theorem that $\nabla f(p) = (0,0)$. In other words, you must locate the critical points of f belonging to A. If a point $(x,y) \in \mathbb{R}^2$ satisfies $\nabla f(x,y) = (0,0)$ then

$$\begin{aligned}
&\implies (\partial_1 f(x,y), \partial_2 f(x,y)) = (0,0) \\
&\implies (2x - 2y - 4, -2x + 8y - 2) = (0,0) \\
&\implies x = 3, y = 1.
\end{aligned}$$

Since $(3,1) \in A$, it follows that $(3,1)$ is a possible global extremum of f on A.

Next, assume $p \in \partial A$. We must determine the extrema of f on the boundary of its domain, which consists of 4 line segments labelled below:



The strategy is to parametrize each piece of the boundary and optimize f along these pieces.

- The line segment ℓ_1 can be parametrized as $\gamma_1(t) = (t,0)$ for $0 \le t \le 4$. You want to find the global extrema of the single-variable function

$$g_1(t) = f(\gamma_1(t)) = f(t,0) = t^2 - 4t + 24, \qquad 0 \le t \le 4.$$

  As $g_1'(t) = 2t - 4$, the only critical point is $t = 2$. Thus, the possible extrema of f on ℓ_1 occur at $\gamma_1(2) = (2,0)$ and the endpoints $\gamma_1(0) = (0,0)$ and $\gamma_1(4) = (4,0)$.
- The line segment ℓ_2 can be parametrized as $\gamma_2(t) = (4,t)$ for $0 \le t \le 2$. The function

$$g_2(t) = f \circ \gamma_2(t) = 4t^2 - 10t + 24, \qquad 0 \le t \le 2$$

  satisfies $g_2'(t) = 8t - 10$ so $t = 5/4$ is the only critical point. Thus, the possible extrema of f occur at $\gamma_2(\frac{5}{4}) = (4, \frac{5}{4})$ as well as the endpoints $(4,0)$ and $(4,2)$.

- The line segment ℓ_3 can be parametrized as $\gamma_3(t) = (t, 2)$ for $0 \leq t \leq 4$. The function

$$g_3(t) = f \circ \gamma_3(t) = t^2 - 8t + 36, \qquad 0 \leq t \leq 4$$

satisfies $g_3'(t) = 2t - 8$ so there are no critical points on $0 < t < 4$. Thus, the possible extrema of f occur at the endpoints $(0, 2)$ and $(4, 2)$.

- The line segment ℓ_4 can be parametrized as $\gamma_4(t) = (0, t)$ for $0 \leq t \leq 2$. The function

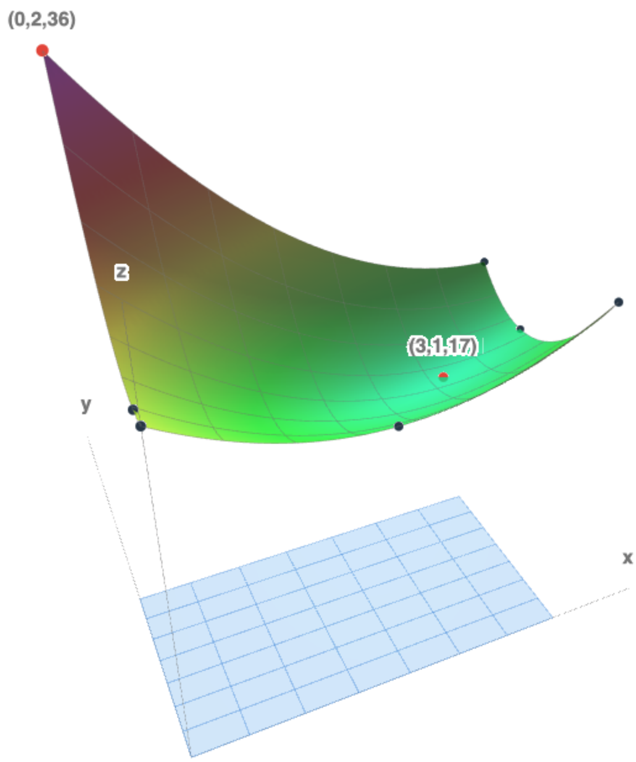
$$g_4(t) = f \circ \gamma_4(t) = 4t^2 - 2t + 24, \qquad 0 \leq t \leq 2$$

satisfies $g_4'(t) = 8t - 2$ so $t = 1/4$ is the only critical point. Thus, the possible extrema of f occur at $\gamma_4(\frac{1}{4}) = (0, \frac{1}{4})$ as well as the endpoints $(0, 0)$ and $(0, 2)$.

This completes the list of possible extrema of f on the boundary.

Compiling all of your data, all that remains is to compare the values of f at all possible extrema whether in the interior or on the boundary. This is calculated below.

a	$f(a)$
$(3,1)$	17
$(2,0)$	20
$(4,\frac{5}{4})$	17.75
$(4,2)$	20
$(0,\frac{1}{4})$	23.75
$(0,0)$	24
$(4,0)$	24
$(4,2)$	20
$(0,2)$	36



From the table, you can deduce that the global maximum of f on A is $f(0, 2) = 36$, and the global minimum is $f(3, 1) = 17$. The Math3D graph of f is included above to illustrate.

Example 4.4.3 You can also optimize the function

$$f(x, y) = x^2 + y^2 + 4x - 6y,$$

on the *unbounded* closed domain $A = \mathbb{R}^2$. It is not immediately obvious whether f even possesses any global extrema, so you must prove that it does before you can conclude anything. You need to understand the *limiting* behaviour of $f(x, y)$ as $||(x, y)|| \to \infty$. If you are lucky, you can evaluate this limit. Notice that

$$\begin{aligned}\lim_{||(x,y)||\to\infty} f(x,y) &= \lim_{||(x,y)||\to\infty}\left(x^2+y^2+4x-6y\right)\\ &= \lim_{||(x,y)||\to\infty} ||(x,y)||^2\left(1+\frac{4x}{||(x,y)||^2}-\frac{6y}{||(x,y)||^2}\right).\end{aligned}$$

You can verify that $\frac{x}{||(x,y)||^2}\to 0$ and $\frac{y}{||(x,y)||^2}\to 0$ as $||(x,y)||\to\infty$. By a product limit law for infinite limits, the above is equal to ∞. This automatically implies that f does *not* have a global maximum. Moreover, once you prove a lemma similar to Lemma 2.8.14, you can conclude that f must have a global minimum on A. It is left as an exercise to state and prove such a lemma.

Now, you know that f has a global minimum, you must find it. The domain $A=\mathbb{R}^2$ has empty boundary, so the global minimum must occur on the interior. Let $p\in\mathbb{R}^2$ be a minimum point of f. Note it may not be unique, but you have at least verified it exists. Since p is a global minimum point lying inside the interior of $A=\mathbb{R}^2$ and f is differentiable on $\mathbb{R}^2$, it follows that $\nabla f(p)=0$ by the local extreme value theorem. The calculation in Example 4.4.1 implies that the only critical point of f on $\mathbb{R}^2$ is the point $(-2,3)$. Since this critical point is unique and $\nabla f(p)=0$, the point $(-2,3)$ is necessarily the global minimum point p! That is, $f(-2,3)=-13$ is the global minimum of f on $\mathbb{R}^2$.

Now that you are equipped with the basic multivariable optimization techniques, you can begin applying them. This requires some additional care with explaining your setup, defining your variables, and modelling the scenario. You may need to make some assumptions or choices, so you should always be clear what you are assuming.

Example 4.4.4 A company sells two products in separate markets where it has a monopoly. The prices, p and q (in dollars), and their quantities demanded, x and y, are related by

$$p = 800-0.4x \quad \text{ and } q = 700-0.3y.$$

Such that price and quantity demanded are inversely proportional to each other. The company's total production cost is given by

$$C = 18+1.6x+1.6y+0.4xy$$

in dollars. How should they set prices and quantities to maximize profit? What is this maximum profit?

First, the total revenue R (in dollars) is the sum of the revenues px and qy from each market, so using the relations between the prices and quantities, you find that

$$R = px+qy = (800-0.4x)x+(700-0.3y)y.$$

The profit P (in dollars) is the revenue minus the cost so

$$P = R-C = -18+798.4x-0.4x^2+698.4y-0.3y^2-0.4xy.$$

Quantites x and y cannot be negative, so you must maximize the profit function $P(x,y)$ on the closed first quadrant

$$D = \{(x,y)\in\mathbb{R}^2 : x\geq 0, y\geq 0\}.$$

Similar to the previous example, you can verify that

$$P(x,y) \to -\infty \quad \text{as} \quad ||(x,y)|| \to \infty$$

inside the first quadrant. Since P is continuous on the closed set D, it follows from Lemma 2.8.14 that P has a global maximum on D. This maximum must lie on the interior or the boundary of D.

If the maximum is on the interior, then since P is differentiable everywhere, the local extreme value theorem implies that this maximum must be a critical point where ∇P vanishes. Thus, you must solve for $(x,y) \in D^\circ$ such that $\nabla P(x,y) = (0,0)$. This implies that

$$\frac{\partial P}{\partial x} = 798.4 - 0.8x - 0.4y = 0$$
$$\frac{\partial P}{\partial y} = 698.4 - 0.6y - 0.4x = 0.$$

Solving this system of linear equations, you will find that $x = 624$ and $y = 748$. Thus, there is exactly one critical point $(624,748)$ lying in the interior of D.

The boundary of D is the union of the positive y-axis and positive x-axis, so you must maximize $P(x,0)$ on $x \geq 0$ and $P(0,y)$ on $y \geq 0$.

- Note $g(x) = P(x,0) = -18 + 798.4x - 0.4x^2$ satisfies $g'(x) = 798.4 - 0.8x$ and $\lim_{x\to\infty} g(x) = -\infty$, so $g(x)$ is maximized on $x \geq 0$ either at the critical point $x = 998$ or the endpoint $x = 0$. These correspond to the points $(998,0)$ and $(0,0)$ in D.
- Note $h(y) = P(0,y) = -18 + 698.4y - 0.3y^2$ satisfies $h'(y) = 698.4 - 0.6y$ and $\lim_{y\to\infty} h(y) = -\infty$, so $h(y)$ is maximized on $y \geq 0$ either at the critical point $y = 1164$ or $y = 0$. These correspond to the points $(0,1164)$ and $(0,0)$ in D.

This gives $(0,0),(998,0)$ and $(0,1164)$ as possible maximum points on ∂D for the profit.

Overall, by combining these calculations, you can conclude that the maximum occurs at one of the four points $(0,0),(998,0),(0,1164)$, or $(624,748)$ lying inside D. By evaluating the profit at these values, you notice that

$$P(0,0) = -18 \qquad P(998,0) = 398{,}383.6,$$
$$P(0,1164) = 406{,}450.8 \qquad P(624,748) = 510{,}284.4.$$

Thus, the company can make a maximum profit \$510,284.40 when they set the price to be $p(624) = \$550.40$ per unit and $q(748) = \$475.60$ per unit.

Now, despite the power of your current tools, there are many optimization problems which you cannot immediately solve. Here is a simple example.

Example 4.4.5 Suppose you want to find the extrema of the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x_1^2 - x_2^2 + \cdots + (-1)^{n+1} x_n^2$$

on the closed unit ball $B = \overline{B_1}(0)$. Since f is continuous and B is compact, you can proceed by the strategies from before. Note

$$\nabla f(x) = (2x_1, -2x_2, \ldots, (-1)^{n+1} 2x_n)$$

so $\nabla f(x) = 0$ if and only if $x = 0$. As f is differentiable, the origin is its only critical point.

Now, you must find the possible extrema of f on the boundary

$$\partial B = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}, \tag{4.4.1}$$

the unit sphere in $\mathbb{R}^n$. In previous examples, you could parametrize the boundary and start a new "lower-dimensional" optimization problem, but how do you do that here? The set does not seem easy to parametrize at all! This is a major issue, so you are stuck for the moment.

Despite all the progress you have made with multivariable optimization, this final example illustrates that there are still serious hurdles. The basic strategy quickly unravels if you cannot *parametrize* your boundary. On the other hand, this boundary in Example 4.4.5 is a set written in implicit form and appears to be a "lower dimensional manifold" based on the ideas in Section 1.5 (see Example 1.5.17). This situation occurs often, presenting a major new problem.

> *How can you optimize a real-valued function over a "lower dimensional manifold" written in implicit form?*

This question is begging you to establish a rigorous definition for a manifold. Once you construct these definitions at the conclusion of this chapter, you will embark on a new journey to optimize on "lower dimensional manifolds" written in implicit form.

Exercises for Section 4.4

Concepts and definitions

4.4.1 Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$. Let $a \in S$. Which of the following are true or false?

(a) f has a global maximum at a if and only if $\forall x \in S, f(x) \le f(a)$.

(b) If f has a global extremum at a then f has a local extremum at a.

(c) If f has a global extremum at a and f is differentiable on the interior S^o then $\nabla f(a) = 0$.

(d) If f is continuous on S then the global extrema of f occur either on the boundary ∂S or the interior S^o.

(e) If f is continuous on S, f is differentiable on the interior S^o, and S is compact, then the global extrema of f occur either on the boundary ∂S or at a critical point p where $\nabla f(p) = 0$.

4.4.2 The basic optimization problem is often described as:

Let f be a real-valued function on a set $S \subseteq \mathbb{R}^n$. Find the global extrema of f on S.

There is not enough information here. Before you begin optimizing, what would you ask about the problem? List your questions about any aspect.

4.4.3

(a) You want to optimize a function $g : \mathbb{R}^2 \to \mathbb{R}$ on the closed disk $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 4\}$. After some computations, you have determined that:

- g is C^1 on $\mathbb{R}^2$.
- For $(x, y) \in D^o$, $\nabla g(x, y) = (0, 0)$ if and only if $(x, y) = (1, 1)$ or $(0, 0)$.
- $g(1, 1) = -\pi$ and $g(0, 0) = 6$

What can you conclude about g on the closed disk D? Briefly explain.

(b) Using the same g as above, you have also defined the function $\phi : [0, 2\pi] \to \mathbb{R}$ by $\phi(t) = g(2\cos t, 2\sin t)$ and determined that:

- For $t \in (0, 2\pi)$, $\phi'(t) = 0 \iff t \in \{\pi/4, \pi\}$.
- $\phi(0) = \phi(2\pi) = -1$, $\phi(\pi) = -2$, and $\phi(\pi/4) = 8$.

What more can you conclude about g on the closed disk D? Briefly explain.

(c) You want to optimize a function $f : S \to \mathbb{R}$ on a closed square $S = [0, 1]^2$. After some computations, you have determined that:

- f is continuous on S and C^1 on the interior of S.
- f has exactly one critical point and it is a saddle point.
- $f(0, 0) = 137$, $f(1, 0) = 223$, $f(0, 1) = 237$, $f(1, 1) = 334$

What can you conclude about f on the closed square S? Briefly explain.

Computations

4.4.4 Searching for global extrema of a continuously differentiable function on a compact set has a natural strategy. Define $S = [0, 4] \times [-2, 0]$ and $g(x, y) = \frac{1}{3}x^3 + y^2 + 2xy - 6x - 3y + 4$.

(a) Draw a picture of the situation and label it. Briefly summarize the strategy.

(b) First, search the interior of S for possible global extrema.

(c) Second, parametrize the boundary of S and search for global extrema.

(d) Determine the global extrema of g on S.

4.4.5 Find all critical points of $f(x,y) = \sin(x)\sin(y)$ on $S = (0,2\pi)\times(0,2\pi)$. Classify as many as you can.

4.4.6 What is the shortest distance from the surface $xy + 2x + z^2 = 9$ to the origin?

4.4.7 Find the point on the surface $z = x^2 + y^2 + 10$ nearest the plane $x + 2y - z = 0$. Identify the point on the plane. *Hint:* Sketch a well-labelled picture to help set up your equations.

4.4.8 Maximize $f(x,y) = x + y$ on the closed disk $S = \{(x,y)\in\mathbb{R}^2 : x^2 + y^2 \le 1\}$.

4.4.9 Maximize $f(x,y,z) = x + y + z$ on the closed ball $S = \{(x,y,z)\in\mathbb{R}^3 : x^2 + y^2 + z^2 \le 1\}$.

Proofs

4.4.10 You want to maximize $f(x,y) = x^2+y^2+4y$ on the closed disk $A = \{(x,y)\in\mathbb{R}^2 : x^2+y^2 \le 16\}$. Below are several bad attempts, but all derivatives and algebraic manipulations are correct.

(a) Asif scribbles down the following attempt.

> 1. $\nabla f(x,y) = (2x, 2y+4)$
> 2. $\nabla f(x,y) = (0,0) \implies (x,y) = (0,-2)$.
> 3. *So the maximum of* f *occurs at* $(0,-2)$ *and is equal to* $f(0,-2) = -4$.

In addition to being very poorly written, Asif's solution has a critical flaw. Identify the flaw.

(b) Thad tries to improve Asif's attempt.

> 1. *Since* $\nabla f(x,y) = (2x,2y+4)$*, the critical points of* f *occur when* $(2x,2y+4) = (0,0)$.
> 2. *Hence,* $(x,y) = (0,-2)$ *is the only critical point of* f *on* A.
> 3. *Define* $g(t) = f(4\cos t, 4\sin t) = 16 + 16\sin t$ *for* $0 \le t \le 2\pi$.
> 4. *Notice* $g'(t) = 16\cos t$ *so* $g'(t) = 0$ *on* $0 < t < 2\pi$ *if and only if* $t = \pi/2$ *or* $3\pi/2$.
> 5. *The values* $t = \frac{\pi}{2}, \frac{3\pi}{2}$ *correspond to* $(4\cos\frac{\pi}{2}, 4\sin\frac{\pi}{2}) = (0,4)$ *and* $(4\cos\frac{3\pi}{2}, 4\sin\frac{3\pi}{2}) = (0,-4)$.
> 6. *Overall,* $f(0,-2) = -4, f(0,-4) = 0$ *and* $f(0,4) = 32$.
> 7. *The maximum of* f *on* A *is therefore 32.*

In addition to missing justifications, Thad's argument has a flaw. Identify the flaw.

(c) Caleb tries to improve Thad's attempt.

> 1. *Since* $\nabla f(x,y) = (2x,2y+4)$*, the critical points of* f *occur when* $(2x,2y+4) = (0,0)$.
> 2. *Hence,* $(x,y) = (0,-2)$ *is the only critical point of* f *on* A.
> 3. *The boundary of* A *is the circle* $\partial A = \{(x,y)\in\mathbb{R}^2 : x^2+y^2 = 16\}$.
> 4. *Parametrize the boundary of* A *with* $\gamma(t) = (4\cos t, 4\sin t)$ *for* $0 \le t \le 2\pi$.
> 5. *Define* $g(t) = f\circ\gamma(t) = f(4\cos t, 4\sin t) = 16 + 16\sin t$ *for* $0 \le t \le 2\pi$.
> 6. *Notice* $g'(t) = 16\cos t$ *on* $0 < t < 2\pi$*, so* $g'(t) = 0$ *if and only if* $t = \pi/2$ *or* $3\pi/2$.
> 7. *The critical points of* g *are* $t = \frac{\pi}{2}, \frac{3\pi}{2}$ *and the endpoints of its domain are* $t = 0, 2\pi$.
> 8. *The maximum of* g *must occur at one of these values.*
> 9. *These values of* t *correspond to* $\gamma(0) = \gamma(2\pi) = (4,0), \gamma(\frac{\pi}{2}) = (0,4)$*, and* $\gamma(\frac{3\pi}{2}) = (0,-4)$.
> 10. *Overall, the maximum of* f *must occur at one of* $(0,-2),(4,0),(0,4)$ *or* $(0,-4)$.
> 11. *By direct computation,* $f(0,-2) = -4, f(4,0) = 16, f(0,4) = 32$ *and* $f(0,-4) = 0$,
> 12. *The maximum of* f *on* A *is therefore 32.*

Caleb's argument does not have a serious flaw, but it is missing many key justifications.

(a) Line 1 is missing a justification. Add the necessary reasoning.
(b) Identify which line(s) apply the global extreme value theorem.
(c) Identify which line(s) apply the local extreme value theorem.
(d) Justify line 10. You may use previous lines.

4.4.11 Optimization on non-compact sets requires some new ideas and careful reasoning. Here are examples of *terribly* written arguments: one is an unbounded set and the other is not closed.

(a) You attempt to find the maximum of $f(x,y) = e^{-x^2-y^2}$ on $S = \mathbb{R}^2$.

> 1. *Then* $\nabla f(x,y) = (-2xe^{-x^2-y^2}, -2ye^{-x^2-y^2})$.
> 2. $\nabla f(x,y) = (0,0) \iff (-2xe^{-x^2-y^2}, -2ye^{-x^2-y^2}) = (0,0) \iff (x,y) = (0,0)$ *critical point*
> 3. *Thus, the maximum occurs at* $(0,0)$ *and is equal to* $f(0,0) = 1$.

The calculations and conclusion are correct, but the reasoning is flawed. Identify the errors.

(b) You attempt to find the minimum of $g(x,y) = \frac{x}{x^2+y^2}$ on $S = \{(x,y) \in \mathbb{R}^2 : 0 < x^2+y^2 \leq 1\}$.

> 1. *By direct computation,* $\nabla g(x,y) = \left(\dfrac{y^2-x^2}{(x^2+y^2)^2}, \dfrac{-2xy}{(x^2+y^2)^2}\right)$.
> 2. *Then* $\nabla g(x,y) = (0,0)$ *if and only if* $x^2 = y^2$ *and* $xy = 0$ *which occurs if and only if* $(x,y) = (0,0)$.
> 3. *Since* $(0,0) \notin S$, *the function* g *does not have a global minimum.*

The calculations and conclusion are correct, but the reasoning is flawed. Identify the errors.

4.4.12 The most common optimization scenario uses the following lemma. Prove it.

> **Lemma.** *Let* $S \subseteq \mathbb{R}^n$ *and* $f : S \to \mathbb{R}$. *If* S *is compact,* f *is continuous on* S, *and* f *is differentiable on the interior* S^o, *then the global extrema of* f *occur either on the boundary* ∂S *or at a critical point* $p \in S^o$ *where* $\nabla f(p) = 0$.

4.4.13 Unique critical points are special because they can often produce a global extremum. These lemmas are particularly useful in optimization problems.

(a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Prove that if f has a unique critical point p and $\lim_{||x|| \to \infty} f(x) = \infty$ then f attains a global minimum at p.

(b) Let $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Prove that if g has a unique critical point p and $\lim_{||x|| \to \infty} g(x) = -\infty$ then g attains a global maximum at p. *Hint:* Use the previous result for a quick proof.

Applications and beyond

4.4.14 Applied optimization problems require you to model the situation, make reasonable choices, and explain a bit more. Here is an example.

> *A maths professor wants your help designing a container to hold their supply of chalk. Their requirements are for the width, height, and length to be less than or equal to 4500 mm. Find the dimensions of this container that would maximize its volume.*

(a) Setup the optimization problem. Precisely define a function f and region S. Explain choices.

(b) Search for extrema lying in the interior of S.

(c) Search for extrema lying on the boundary of S.

(d) Conclude your solution with proper reasoning and a full sentence.

4.4.15 Find the line of best fit for the data points $(0,5), (1,2)$, and $(2,0)$ by minimizing the sum of the vertical square errors. *Hint:* See question 4.3.15.

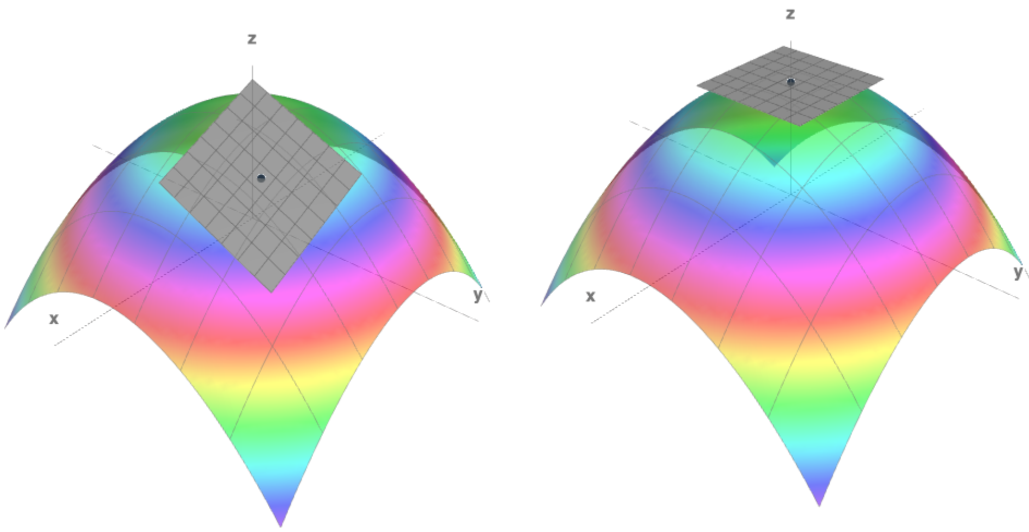
4.4.16 A rocket ship needs to deliver a combined 2400 cubic metres of various payloads to astronauts aboard a space station. You are tasked with designing a rectangular cargo bay that can carry these items, potentially across multiple trips, for the rocket. The total cost of this operation is the cost of the cargo bay plus a fixed \$128 million per back-and-forth trip. The cargo bay must have height 8 metres, but can have varying width and length.
Due to the layout of existing systems and infrastructure inside the rocket, engineers estimate the cost of the cargo bay to be \$50 million/m^2 for the front and back, \$25 million/m^2 for the sides, and \$200 million/m^2 for the top and bottom. What dimensions should the container have to minimize the total cost?

4.4.17 Maximize $f(x) = x_1 + \cdots + x_n$ on the closed unit ball $S = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$.

4.5. Tangent spaces

The geometric viewpoint of derivatives has been intimately connected to the idea of a "manifold", whether the set is written in explicit form, implicit form, or parametric form (Section 1.5). Directional derivatives can be interpreted as vectors “tangent” to a set in parametric form (Section 3.3.3). Gradient vector fields “orthogonally” cross sets in implicit form (Section 3.4.3). Local extrema occur when the graph of a real-valued function is “flat” (Section 4.3.1). Optimization requires you to analyze the boundary of a domain (Section 4.4), which is often a set in implicit form. All of these observations are now compelling you to formalize these notions.

Before proceeding with any of these investigations, you must begin by formalizing the idea of the “tangent plane to a set”. Pictures can build some intuition; see this Math3D demo.



A tangent plane should barely “touch the surface". For instance, at a local extrema, you may expect your tangent plane to be “flat”. This intuition will be formalized in smaller steps.

> *When is a vector tangent to a set? What is the space of all tangent vectors? How do these vectors define a tangent plane?*

This section is dedicated to resolving these questions.

4.5.1 Tangent vectors, spaces, and planes

The notion of tangency originates with parametric curves in Section 3.1. Recall if $\gamma : \mathbb{R} \to \mathbb{R}^m$ is differentiable at $t \in \mathbb{R}$ then its derivative $\gamma'(t) \in \mathbb{R}^m$ can be interpreted in several ways. Geometrically speaking, $\gamma'(t)$ is a direction vector defining the tangent line of γ at t. That is,

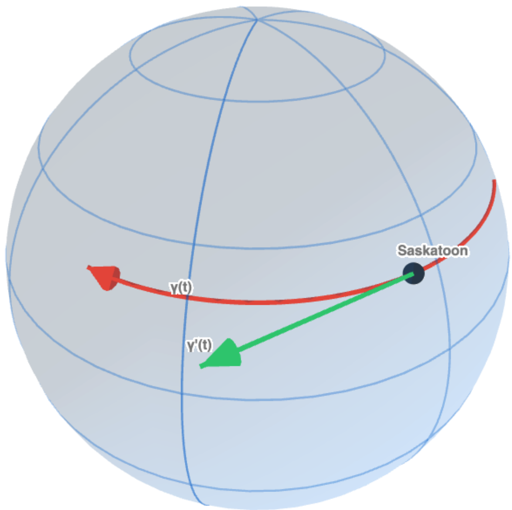
> *The vector $v = \gamma'(t)$ is **tangent** to the curve defined by γ at the point $p = \gamma(t)$.*

Physically speaking, $\gamma'(t)$ is the velocity of the particle moving along γ at time t. That is,

> *The vector $v = \gamma'(t)$ is the **velocity** of a particle moving along γ at position $p = \gamma(t)$.*

You can apply these perspectives to discover the definition of tangent vectors.

Example 4.5.1 You are in a vehicle traveling for 12 hours on the Earth’s surface, which is modeled as a sphere $S \subseteq \mathbb{R}^3$. Your position is given by a parametric curve $\gamma : (0, 12) \to \mathbb{R}^3$. Halfway through the trip, you are passing through Saskatoon. Your position is $\gamma(6) \in S$ and your velocity is $\gamma'(6) \in \mathbb{R}^3$. What are all possible velocities of people traveling by car through Saskatoon? Can it be any possible vector in $\mathbb{R}^3$? For instance, can your velocity point radially away from the Earth? Explore the diagram below on Math3D.



That is absurd since that would mean your car can leave the Earth's surface. If you are traveling on the Earth's surface, it must be true that $\gamma(t) \in S$ for all $t \in (0,12)$ or equivalently $\gamma((0,12)) \subseteq S$. This condition is the key to defining tangent vectors because it restricts the possible velocities! Intuitively, you cannot defy gravity in your car but you can drive in at least two different directions through Saskatoon, e.g. north-south and east-west. You may therefore guess that the set of possible velocities is a 2-dimensional subspace of $\mathbb{R}^3$ spanned by these two directions.

This informal exploration leads to a natural definition.

Definition 4.5.2 Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. A vector $v \in \mathbb{R}^n$ is a **tangent vector of** S **at** p if there exists an open interval $I \subseteq \mathbb{R}$ containing 0 and a map $\gamma : I \to \mathbb{R}^n$ such that[2] γ is C^1, $\gamma(I) \subseteq S$, $\gamma(0) = p$, and $\gamma'(0) = v$.

You can interpret this definition from a physical perspective.

There exists a particle moving along S through p with velocity v.

It is worth studying a running example for this entire section.

Example 4.5.3 Define the set

$$S = \{(x, y, 9 - x^2 - y^2) : x, y \in \mathbb{R}\}$$

and the point $p = (1,2,4) \in S$. What are some tangent vectors of S at p? There are an infinitude of differentiable curves on S all of which pass through p. Here is a few of them.

$$\begin{array}{ll}
\gamma_1 : \mathbb{R} \to \mathbb{R}^3 & \gamma_1(t) = (1+t, 2-3t, 9-(1+t)^2-(2-3t)^2), \\
\gamma_2 : (-6, 0.3) \to \mathbb{R}^3 & \gamma_2(t) = (\cos t, 2e^t, 9-(\cos t)^2-(2e^t)^2), \\
\gamma_3 : (-1.5, 1.5) \to \mathbb{R}^3 & \gamma_3(t) = (1+t^2, 2+t^3, 9-(1+t^2)^2-(2+t^3)^2), \\
\gamma_4 : (-1, 2) \to \mathbb{R}^3 & \gamma_4(t) = (1-2t, 2-t, 9-(1-2t)^2-(2-t)^2).
\end{array}$$

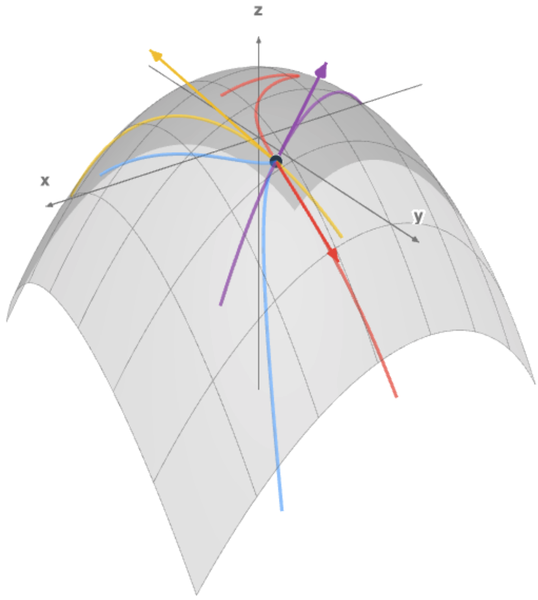
You can verify that $\gamma_1, \gamma_2, \gamma_3$, and γ_4 are all C^1 and their range is a subset of the set S. Notice

[2]You may wonder why γ is required to be C^1 instead of differentiable. This choice was made for simplicity but will hardly make a noticeable difference for your purposes. Some proofs might be a little easier with the C^1 assumption, but that is about it. Most mathematical texts actually require more derivatives, often C^∞ instead of C^1. It all comes down to context; you ask for as many derivatives as you may need.

also that $\gamma_1(0) = \gamma_2(0) = \gamma_3(0) = \gamma_4(0) = (1,2,4) = p$ and, by direct calculation,

$$\gamma_1'(0) = (1,-3,10), \qquad \gamma_2'(0) = (0,2,-8), \qquad \gamma_3'(0) = (0,0,0), \qquad \gamma_4'(0) = (-2,-1,8)$$

Thus, the vectors $v_1 = (1,-3,10), v_2 = (0,2,-4), v_3 = (0,0,0)$ and $v_4 = (-2,-1,8)$ are all tangent vectors of S at $p = (1,2,4)$ by definition. A Math3D demo is included below.



Note $\gamma_1, \gamma_2, \gamma_3$, and γ_4 are respectively the yellow, red, blue, and purple curves.

Now, what are all possible tangent vectors? This set of vectors warrants its own definition.

Definition 4.5.4 Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. The **tangent space of S at p**, denoted T_pS, is the set of tangent vectors to S at p. In other words,

$$T_pS = \{v \in \mathbb{R}^n : v \text{ is a tangent vector of } S \text{ at } p\}.$$

You can again interpret this definition from a physical perspective.

T_pS is the set of all possible velocities for a particle moving along S through p.

Does this set form a subspace of $\mathbb{R}^n$? This question will be central to the future study of manifolds. For now, you can prove that the zero vector always belongs to the tangent space.

Example 4.5.5 For any set $S \subseteq \mathbb{R}^n$ any point $p \in S$, you can verify that $0 \in T_pS$ by definition. This is left as an exercise. A good choice of curve is one that goes nowhere.

You can attempt to guess the tangent space in an explicit example.

Example 4.5.6 Continue with Example 4.5.3, so $S = \{(x,y,9-x^2-y^2) : x,y \in \mathbb{R}\}$ and $p = (1,2,4)$. From before, you have verified by definition that

$$v_1 = (1,-3,10), \quad v_2 = (0,2,-8), \quad v_3 = (0,0,0), \quad v_4 = (-2,-1,8)$$

are all tangent vectors of S at p and hence, they are elements of T_pS. Producing all tangent vectors seems like an impossible task since you would have to consider all possible parametric curves passing through p. Luckily, S is a rather special set: it is a graph!

Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = 9 - x^2 - y^2$, so indeed

$$S = \{(x,y,f(x,y)) : x,y \in \mathbb{R}\}$$

This allowed the tangent vectors v_1 and v_4 from Example 4.5.3 to be created with straight lines passing through $(1,2) \in \mathbb{R}^2$ in the domain of f. The line $(1+t, 2-3t)$ in $\mathbb{R}^2$ corresponded to the curve $\gamma_1(t)$ on the graph in $\mathbb{R}^3$ and the line $(1-2t, 2-t)$ in $\mathbb{R}^2$ corresponding to the curve $\gamma_4(t)$ on the graph in $\mathbb{R}^3$. These are inspired by the directional derivative of f. In particular, you can check that

$$\gamma_1'(0) = \big(1, -3, D_{(1,-3)}f(1,2)\big), \quad \text{and} \quad \gamma_4'(0) = \big(-2, -1, D_{(-2,-1)}f(1,2)\big).$$

This relation holds much more generally.

Fix $w = (w_1, w_2) \in \mathbb{R}^2$ and $a = (1,2)$ so $p = (a, f(a)) = (1,2,4)$. Define $g : \mathbb{R} \to \mathbb{R}^2$ to be the straight line defined by $g(t) = a + tw$. Then define $\gamma : \mathbb{R} \to \mathbb{R}^3$ by

$$\begin{aligned} \gamma(t) = (g(t), f \circ g(t)) &= (a + tw, f(a+tw)) \\ &= \big(1 + tw_1, 2 + tw_2, f(1+tw_1, 2+tw_2)\big), \end{aligned}$$

so $\gamma(0) = (a, f(a)) = p$ and $\gamma(t) \in S$ for all $t \in \mathbb{R}$. In other words, $\gamma(\mathbb{R}) \subseteq S$. You can verify that both g and f are C^1 and hence differentiable on their domains. Thus, by Lemma 3.1.3 and the chain rule (Corollary 4.1.4), γ is C^1 and

$$\gamma'(t) = \big(g'(t), \nabla f(g(t)) \cdot g'(t)\big).$$

Overall, this implies that

$$\gamma'(0) = (g'(0), \nabla f(g(0)) \cdot g'(0)) = (w, \nabla f(a) \cdot w)$$

is a tangent vector of S at $p = \gamma(0) = (1,2,4)$. By direct calculation, you can check that $\nabla f(1,2) = (-2,-4)$ so the above becomes

$$\gamma'(0) = (w_1, w_2, -2w_1 - 4w_2) = w_1 \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} + w_2 \begin{bmatrix} 0 \\ 1 \\ -4 \end{bmatrix}.$$

Thus, any linear combination of $(1,0,-2)$ and $(0,1,-4)$ is a tangent vector of S at p.

This argument proves that

$$\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -4 \end{bmatrix} \right\} \subseteq T_pS$$

The lefthand set is a 2-dimensional subspace of $\mathbb{R}^3$, so you may wonder: are these sets actually equal? After all, you considered all curves on the set S produced by straight lines in the domain of f. However, as demonstrated in Example 4.5.3 with the curves γ_2 and γ_3, there are curves that do not arise this way so the answer remains unclear.

The tangent space describes all possible tangent vectors to a point on a set. By translating the tangent space to the point of tangency, you can finally define the tangent plane.

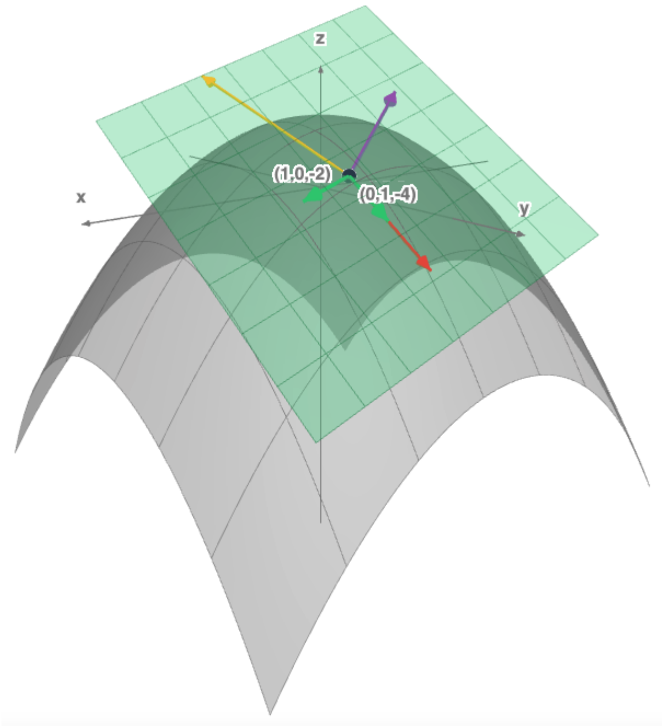
Definition 4.5.7 Let $S \subseteq \mathbb{R}^n$ be a set and $p \in S$. The **tangent plane of** S **at** p, denoted $p + T_pS$, is the tangent space translated to p. That is,

$$p + T_pS = \{p + v : v \in T_pS\}.$$

Example 4.5.8 From Example 4.5.6, the subspace $V = \text{span}\{(1,0,-2),(0,1,-4)\}$ is conjecturally equal to the tangent space T_pS and therefore the tangent plane of S at $p = (1,2,4)$ is conjecturally given by

$$\begin{aligned} p + T_pS &= \{(1,2,4) + v : v \in V\} \\ &= \{(1,2,4) + v_1(1,0,-2) + v_2(0,1,-4) : v_1, v_2 \in \mathbb{R}\} \\ &= \{(1+v_1, 2+v_2, 4-2v_1-4v_2) : v_1, v_2 \in \mathbb{R}\}. \end{aligned}$$

This expression gives a 2-dimensional plane in parametric form. This tangent plane is illustrated below along with the tangent vectors v_1, v_2, and v_4 (translated to p) from Example 4.5.3. View this Math3D demo for a better visual.



You can also rewrite this plane in implicit form with variables $x = 1 + v_1$, $y = 2 + v_2$, and $z = 4 - 2v_1 - 4v_2$. Then

$$(x,y,z) \in p + T_pS \iff z = 4 - 2(x-1) - 4(y-2) \iff 2(x-1) + 4(y-2) + (z-4) = 0.$$

Therefore, you can also guess that the tangent plane is given by

$$p + T_pS = \{(x,y,z) \in \mathbb{R}^3 : 2(x-1) + 4(y-2) + (z-4) = 0\}.$$

Either expression for the tangent plane is valid.

4.5.2 Tangent space of a graph

Unfortunately, there is a major drawback to the definition of the tangent space.

How do you explicitly describe all differentiable parametric curves ***lying inside*** $S \subseteq \mathbb{R}^n$*?*

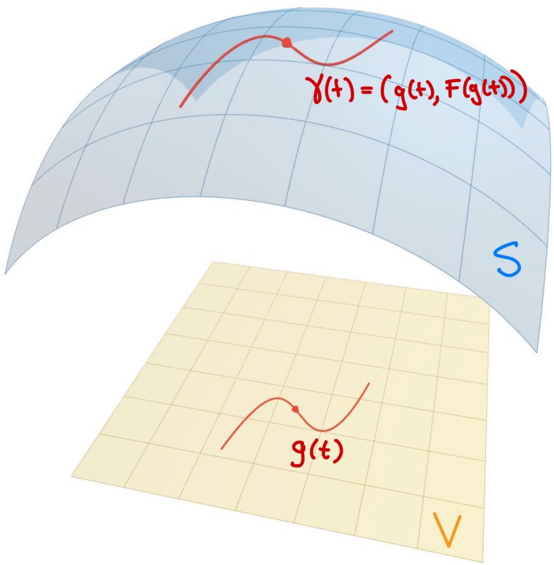
This issue already prevented you from finding the tangent space in Example 4.5.6. While resolving this problem for sets written in implicit form or parametric form will be much harder, you can luckily resolve it for graphs!

Lemma 4.5.9 Let $V \subseteq \mathbb{R}^k$ be open. Let $F : V \to \mathbb{R}^{n-k}$ be C^1. Let $S \subseteq \mathbb{R}^n$ be the graph of F so

$$S = \{(x, F(x)) : x \in V\}.$$

Let $I \subseteq \mathbb{R}$ be open and let $\gamma : I \to \mathbb{R}^n$. The function γ is C^1 and $\gamma(I) \subseteq S$ if and only if there exists a C^1 function $g : I \to \mathbb{R}^k$ such that $g(I) \subseteq V$ and $\gamma(t) = \big(g(t), F(g(t))\big)$ for $t \in I$.

The basic idea of the proof is illustrated in the Math 3D demo below.



There is a bijective correspondence between the points in $V \subseteq \mathbb{R}^k$ and the points in $S \subseteq \mathbb{R}^n$. This observation is informally justified by the following two heuristic reasons.

- For any point $(x_1, \ldots, x_k) \in V$, the corresponding point is $(x_1, \ldots, x_k, F(x_1, \ldots, x_k)) \in S$.
- For any point $(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \in S$, you can express the $n-k$ variables $x_{k+1}, \ldots, x_n$ as an *explicit* differentiable function of the first k variables $x_1, \ldots, x_k$, namely

$$(x_{k+1}, \ldots, x_n) = F(x_1, \ldots, x_k).$$

  This ensures that the correspondence is bijective, because the $n-k$ variables $x_{k+1}, \ldots, x_n$ are uniquely determined by the first k variables $x_1, \ldots, x_k$.

These ideas allow you to go back and forth between g and γ in the lemma.

Proof of Lemma 4.5.9. By definition, $S = \{(x, F(x)) : x \in V\}$. Begin with the "if" direction. Assume $g : I \to \mathbb{R}^k$ is C^1 such that $g(I) \subseteq V$ and $\gamma(t) = \big(g(t), F \circ g(t)\big)$ for all $t \in I$. By definition of S, it follows that $\gamma(t) = \big(g(t), F(g(t))\big) \in S$ for all $t \in I$, so $\gamma(I) \subseteq S$. Moreover, since F and g are C^1, γ is C^1 by Lemma 3.1.3 and the chain rule (Corollary 4.1.4).

Conversely, assume γ is C^1 and $\gamma(I) \subseteq S$. Let $\pi : \mathbb{R}^n \to \mathbb{R}^k$ be the standard projection onto the first k coordinates so $\pi(x_1, \ldots, x_n) = (x_1, \ldots, x_k)$. By definition of π and the set

S, every $p \in S$ satisfies $p = \big(\pi(p), F(\pi(p))\big)$. As $\gamma(I) \subseteq S$, this implies for $t \in I$ that $\gamma(t) = (\pi \circ \gamma(t), F(\pi \circ \gamma(t))) \in S$, so $\pi \circ \gamma(t) \in V$ by definition of S.

Therefore, you may define $g : I \to \mathbb{R}^k$ by $g = \pi \circ \gamma$ and hence $g(I) \subseteq V$ and $\gamma(t) = \big(g(t), F(g(t))\big)$ for all $t \in I$. Since π is a linear map (and hence C^1) and γ is C^1 by assumption, the chain rule implies that g is C^1 as required. ■

Example 4.5.10 Consider again Example 4.5.3. So, you want to find the tangent space of the set $S = \{(x, y, 9 - x^2 - y^2) : x, y \in \mathbb{R}^2\}$ at the point $p = (1, 2, 4)$. Again, S is the graph of the function $f(x, y) = 9 - x^2 - y^2$. Let $\gamma : I \to \mathbb{R}^3$ be a differentiable curve such that $\gamma(I) \subseteq S$ with $I \subseteq \mathbb{R}$ open. By Lemma 4.5.9, if $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ then for all $t \in I$,

$$\gamma_3(t) = f(\gamma_1(t), \gamma_2(t)) = 9 - \big(\gamma_1(t)\big)^2 - \big(\gamma_2(t)\big)^2.$$

In other words, the 3rd component of γ is expressed in terms of the 1st and 2nd components.

Now, equipped with Lemma 4.5.9, you can determine the tangent space of a graph.

Theorem 4.5.11 (Tangent space of graphs) Let $V \subseteq \mathbb{R}^k$ be open. Let $F : V \to \mathbb{R}^{n-k}$ be C^1. Let $S \subseteq \mathbb{R}^n$ be the graph of F so

$$S = \big\{(x, F(x)) : x \in V\big\}.$$

For $a \in V$ and $p = (a, F(a))$, all of the following hold:

(a) $T_pS = \big\{(w, dF_a(w)) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : w \in \mathbb{R}^k\big\}$

(b) $T_pS = \operatorname{span}\big\{\big(e_1, \partial_1 F(a)\big), \ldots, \big(e_k, \partial_k F(a)\big)\big\}$, where $\{e_1, \ldots, e_k\}$ is the standard $\mathbb{R}^k$ basis.

(c) T_pS is a k-dimensional subspace of $\mathbb{R}^n$.

Proof. (a) By an argument similar to Example 4.5.6, you can prove that

$$\big\{(w, dF_a(w)) : w \in \mathbb{R}^k\big\} \subseteq T_pS.$$

This is left as an exercise. For the reverse containment, let $v \in T_pS$ be arbitrary. By definition of the tangent space, there exists a C^1 map $\gamma : I \to \mathbb{R}^n$ with $I \subseteq \mathbb{R}$ open and $0 \in I$ such that $\gamma(I) \subseteq S, \gamma(0) = p$, and $\gamma'(0) = v$. From Lemma 4.5.9, there exists a C^1 map $g : I \to \mathbb{R}^k$ such that $g(I) \subseteq V$ and $\gamma(t) = (g(t), (F \circ g)(t))$ for all $t \in I$. This implies that

$$(a, F(a)) = p = \gamma(0) = (g(0), F(g(0)))$$

and hence $a = g(0)$. Take $w = g'(0)$ so by Lemma 3.1.3, it follows that

$$v = \gamma'(0) = (g'(0), (F \circ g)'(0)) = (w, (F \circ g)'(0)).$$

The definition of the Jacobian and the chain rule implies that

$$(F \circ g)'(0) = D(F \circ g)(0) = DF(g(0))Dg(0) = DF(a)w$$

since $g(0) = a$ and $Dg(0) = g'(0) = w$. By Theorem 3.5.22, the Jacobian is the matrix of the differential so $DF(a)w = dF_a(w)$, so we may conclude that $v = \gamma'(0) = (w, dF_a(w))$ as required.

(b) This is left as an exercise in linear algebra. You must prove that the finite set of k vectors given by $\{(e_1, dF_a(e_1)), \ldots, (e_k, dF_a(e_k))\}$ is a basis for the set defined in (a).

(c) This follows from (b) and standard linear algebra arguments. ■

Theorem 4.5.11 allows you to quickly find the tangent plane to a graph.

Example 4.5.12 You can finally determine the tangent space of the set

$$S = \{(x, y, 9 - x^2 - y^2) : x, y \in \mathbb{R}\}$$

at the point $p = (1, 2, 4)$. Recall S is the graph of the C^1 function $f(x, y) = 9 - x^2 - y^2$ and $p = (a, f(a))$ where $a = (1, 2)$. By Theorem 4.5.11,

$$T_pS = \{(w, df_a(w)) : w \in \mathbb{R}^2\}.$$

By Theorem 3.5.22, it follows that

$$df_a(w) = \nabla f(a) \cdot w = \nabla f(1, 2) \cdot w = (-2, -4) \cdot w = -2w_1 - 4w_2,$$

so, as conjectured in Example 4.5.6, you may conclude that

$$T_pS = \{(w_1, w_2, -2w_1 - 4w_2) : w_1, w_2 \in \mathbb{R}\} = \operatorname{span}\{(1, 0, -2), (0, 1, -4)\}.$$

This concludes a major achievement towards the theory of manifolds! You have formalized the idea of tangent planes. For sets in explicit form, you can compute them and they are naturally the translation of subspace in $\mathbb{R}^n$. However, many sets in $\mathbb{R}^n$ cannot be described explicitly, such as the unit sphere.

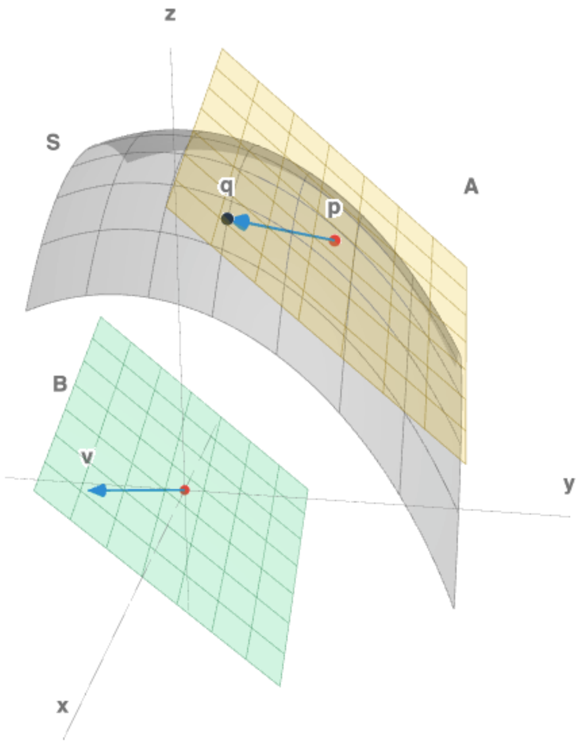
> *How do you find the tangent planes of sets in implicit form or parametric form? Are they always subspaces of $\mathbb{R}^n$?*

These investigations are postponed to future chapters as they are much deeper. You will finish this chapter in the next section with the miraculous definition of a manifold.

Exercises for Section 4.5

Concepts and definitions

4.5.1 The Math3D demo below shows a set S, a plane A, a plane B, a point p, and a vector v.



(a) Which set is the tangent space of S at p?

(b) Which set is the tangent plane of S at p?

(c) The same vector v is drawn twice in the figure. Express q in terms of p and v.

4.5.2 Let $S \subseteq \mathbb{R}^3$ be the upper hemisphere of radius 1 centred at the origin. Fix $p = (0,0,1) \in S$.

(a) Which of the following are tangent vectors of S at p? Select all that apply.

A. $(2,0)$
B. $(2,3,0)$
C. $(0,0,7)$
D. $(2,3,7)$
E. $(0,0,0)$

(b) Which of the following is equal to the tangent space of S at p? Select all that apply.

A. $\mathbb{R}^2$
B. $\mathbb{R}^3$
C. $\{(1,0,0),(0,1,0)\}$
D. $\{(u,v,0) : u,v \in \mathbb{R}\}$
E. $\operatorname{span}\{(1,0,0),(0,1,0)\}$

(c) What is the tangent plane of S at p? Select all that apply.

A. $\mathbb{R}^2$
B. $\{(u,v,0) : u,v \in \mathbb{R}\}$
C. $\{(u,v,1) : u,v \in \mathbb{R}\}$
D. $\operatorname{span}\{(1,0,1),(0,1,1)\}$
E. $\{(x,y,z) \in \mathbb{R}^3 : z = 1\}$

4.5.3 Which statements are equivalent to "$v \in \mathbb{R}^n$ is a tangent vector of $S \subseteq \mathbb{R}^n$ at $p \in S$"?

(a) $v = \gamma'(0)$ and $p = \gamma(0)$ for some C^1 map $\gamma : I \to S$ with $I \subseteq \mathbb{R}$ open and $0 \in I$.

(b) $p + v$ is an element of S.

(c) v is an element of the tangent plane of S at p.

(d) v is an element of the tangent space of S at p.

(e) v is the velocity of a particle passing through p.

4.5.4 Let $S = \{(x, y, z) \in \mathbb{R}^3 : z = xy\} \subseteq \mathbb{R}^3$ and let $p = (2, 3, 6)$, so $p \in S$.

(a) The vector $v = (4, -1, 10)$ is a tangent vector of S at p if and only if there is a C^1 map $\gamma : I \to \mathbb{R}^3$ such that

i) I is ________________________________.

ii) $\gamma(0)$ is ________________________________.

iii) $\gamma'(0)$ is ________________________________.

iv) $\operatorname{im}(\gamma)$ is ________________________________.

(b) Finding a γ satisfying these four properties is tricky for general sets but your set S is luckily a *graph*! Writing $\gamma(t) = (x(t), y(t), z(t))$ in components, rewrite the conditions (b), (c), and (d).

i) I is ________________________________.

ii) ________________________________.

iii) ________________________________.

iv) ________________________________.

(c) Guess a choice for $\gamma : I \to \mathbb{R}^3$ that will conjecturally satisfy (a)–(d). *Hint:* Straight line.

(d) Use your γ to show that $v = (4, -1, 10)$ is a tangent vector of S at p. Is this choice of γ unique?

(e) Use Theorem 4.5.11 to calculate T_pS in this example. State your choices of n, k, V, F, and a.

(f) Give an explicit formula for the differential $dF_a(w)$ at $w \in \mathbb{R}^k$.

(g) Use the theorem to express T_pS as the span of a set of k vectors in $\mathbb{R}^n$.

(h) Express the tangent plane $p + T_pS$ as a set.

Computations

4.5.5 Let $S = \{(x, y, z, x^2 + z^2, xyz) : x, y \in \mathbb{R}\}$ and let $p = (1, 2, -2, 5, -4) \in S$.

(a) Find the tangent space of S at p.

(b) Find the tangent plane of S at p.

4.5.6 Let S be the graph of a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ with $F(1, 1) = (2, -1)$. Assume

$$T_{(1,1,2,-1)}S = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -\pi \\ 4 \end{bmatrix} \right\}.$$

(a) What is the Jacobian of F at $(1,1)$?

(b) Estimate the value of $F(1.2, 0.8)$.

(c) You want to linearly approximate a point q on the set S nearby $(1,1,2,-1) \in S$. After some calculations, you find $(2,0,5+\pi,-7)$ approximates q. What is the point q?

Proofs

4.5.7 Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. Prove that $0 \in \mathbb{R}^n$ belongs to T_pS.

4.5.8 Let $S = \{(x,y) \in \mathbb{R}^2 : y = f(x)\}$ for some C^1 function $f : \mathbb{R} \to \mathbb{R}$. Let $p = (a, f(a)) \in S$ be arbitrary. Prove that $T_pS = \operatorname{span}\{(1, f'(a))\}$ and the tangent plane of S at p is the usual tangent line of f at a.

4.5.9 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be C^1. Let $S = \{(x,y,z) \in \mathbb{R}^3 : z = f(x,y)\} \subseteq \mathbb{R}^3$ be its graph.

(a) Use Theorem 4.5.11 to show that the tangent space at $p = (a,b,f(a,b)) \in S$ is given by

$$T_pS = \operatorname{span}\{(1,0,\partial_1 f(a,b)), (0,1,\partial_2 f(a,b))\}.$$

(b) Conclude that T_pS is a 2-dimensional subspace of $\mathbb{R}^3$ using the previous part.

(c) Show that the tangent plane at $p = (a,b,f(a,b)) \in S$ is given by:

$$\left\{(x,y,z) \in \mathbb{R}^3 : z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b)\right\}$$

(d) Find a normal to the tangent plane of S at p.

(e) Let $q = (s,t,f(s,t))$ be a point nearby $p = (a,b,f(a,b))$ on the set S. What point on the tangent plane of S at p approximates q?

4.5.10 Let $F : V \to \mathbb{R}^{n-k}$ be a C^1 function where the set $V \subseteq \mathbb{R}^k$ is open. Fix $a \in V$.

(a) Show that the set $W = \left\{(w, dF_a(w)) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : w \in \mathbb{R}^k\right\}$ is a subspace of $\mathbb{R}^n$.

(b) Show that $\{(e_1, \partial_1 F(a)), \ldots, (e_k, \partial_k F(a))\}$ is a basis for W.

(c) Conclude that if $S = \{(x,y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : y = F(x)\}$ and $p = (a, F(a))$, then the tangent space T_pS is a k-dimensional subspace of $\mathbb{R}^n$ and

$$T_pS = \operatorname{span}\{(e_1, \partial_1 F(a)), \ldots, (e_k, \partial_k F(a))\}.$$

Applications and beyond

4.5.11 The notion of tangent space still makes perfect sense when the objects are low dimensional. Define the set $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 2\}$. Let $p = (1,1) \in S$.

(a) Draw a figure with the point p, the set S, the tangent space T_pS, and the tangent plane $p + T_pS$.

(b) Guess an expression for the tangent space T_pS as a span of a finite set of vectors.

(c) Guess an expression for the tangent plane $p + T_pS$ of the form $\{(x,y) \in \mathbb{R}^2 : \underline{\qquad\qquad}\}$.

4.5.12 Here is a simplified version of the key lemma for tangent spaces of graphs.

Lemma. *Let $S \subseteq \mathbb{R}^n$ be the graph of a C^1 function $F : \mathbb{R}^k \to \mathbb{R}^{n-k}$. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a parametric curve. Then γ is C^1 and $\gamma(\mathbb{R}) \subseteq S$ if and only if $\gamma(t) = \big(g(t), F(g(t))\big)$ for some C^1 function $g : \mathbb{R} \to \mathbb{R}^k$.*

This describes all possible C^1 parametric curves on a graph. The proof crucially uses the projection map $\pi : \mathbb{R}^n \to \mathbb{R}^k$ given by

$$\pi(x_1, \ldots, x_k, \ldots, x_n) = (x_1, \ldots, x_k).$$

You ultimately argue that a restriction of π is invertible. Normally, projections are not invertible so this property is quite special to graphs! You will formally establish this central property here.

(a) Use S as defined in the lemma. Write S and $\pi(S)$ in set builder notation.

(b) Prove that the restriction $\pi|_S : S \to \pi(S)$ is invertible. *Hint:* What is its inverse?

4.6. Smooth manifolds

Recall some fundamental questions that were sparked long ago in Section 1.5.

> *If a set is described by nonlinear equations, then what is its "dimension"? How do you even define "dimension"?*

This inquiry generated the heuristic idea of a manifold and the above was rephrased as a motivating question for this chapter.

> *What really is a "manifold" and what is its "dimension"?*

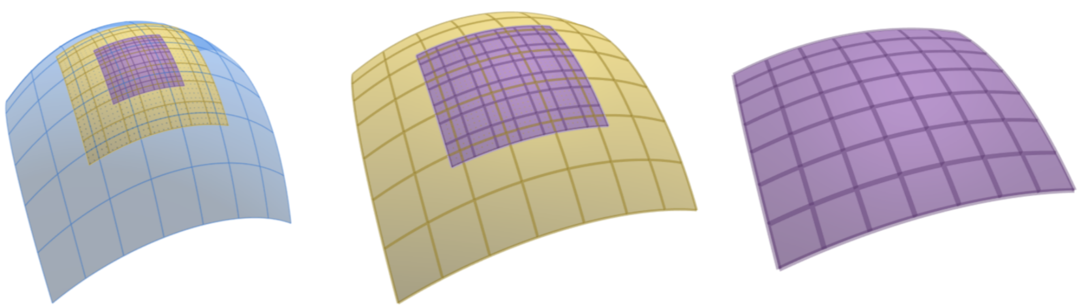
After some substantial theory building, you are prepared to formulate an ingenious solution.

A good definition for whether a set $S \subseteq \mathbb{R}^n$ is a manifold should have three key features. First, it should ensure S is "lower dimensional"; an open ball should not be a manifold, but the unit sphere and a plane in $\mathbb{R}^3$ should presumably be 2-dimensional manifolds. Second, the definition should characterize if S is "smooth"; a sphere should be considered smooth, but a cone should not be. Third, it should not depend on how the set S is written; a hemisphere should be a considered a manifold, regardless of whether you describe it in explicit form, implicit form, or parametric form.

Your definitions of the tangent space and tangent plane are the perfect first step towards these features. First, a tangent plane can have a formal notion of dimension from linear algebra (if the tangent space is indeed a subspace). Second, a tangent plane is "flat" which is as smooth as you can expect. Third, the tangent space does not depend on how the set is described. Given all of this fortune, can you potentially use tangent planes to define a manifold?

The first hint comes from the overarching theme of differentiability: non-linear maps can be *locally* well-approximated by linear maps. You can extend this algebraic philosophy to try to define smooth manifolds[3] . Presumably, a smooth manifold should be *locally* well-approximated by a plane as illustrated by the Math3D demo below. This suggests your first informal guess.

> *A set S is a k-dimensional smooth manifold if you zoom in close enough and it always looks like a k-dimensional plane in $\mathbb{R}^n$.*



More formally, you might phrase this idea as:

> *A set S is a k-dimensional smooth manifold if the tangent space T_pS is a k-dimensional subspace of $\mathbb{R}^n$ at every point p in S.*

[3]Our definition of smooth manifold (Definition 4.6.4) is more commonly referred to as a smooth *embedded* manifold in mathematical literature. The traditional definition of a smooth manifold is inspired by general relativity and is better suited for a course in differential geometry like MAT367. Ultimately, this distinction does not really matter due to the Whitney embedding theorem but that is far beyond the scope of this text.

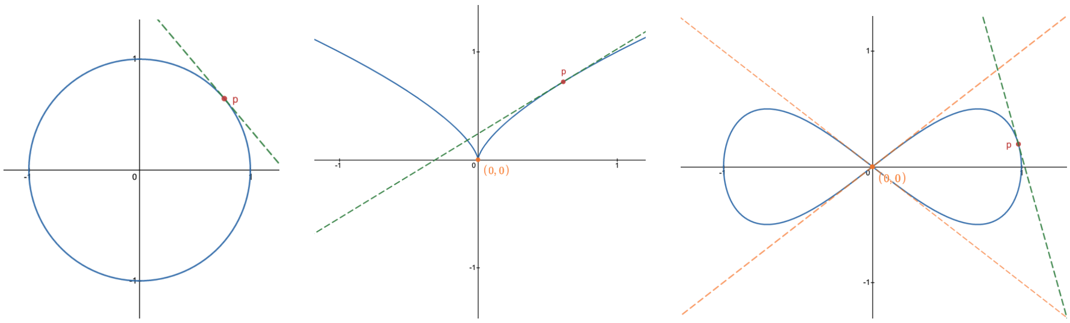
This idea is fantastic! It suggests a beautiful connection between geometry and algebra. Unfortunately, this attempted definition has a serious drawback. Tangent spaces are incredibly difficult to compute from the definition, so it seems too hard to verify this in any example. Thus far, you only have one solid theorem (Theorem 4.5.11) on tangent spaces when S is the *graph* of a differentiable function. Otherwise, for sets defined parametrically or implicitly, you do not yet have any tools to determine their tangent spaces.

Your goal in this section is therefore to construct a good definition for smooth manifolds and confirm that their tangent space is always a subspace of the same dimension. Miraculously, the key ingredient will be the special case you have already considered: *graphs*.

4.6.1 Issues with defining manifolds

Before diving into formal definitions, it is instructive to visually explore some examples. The phrases "smooth manifold" and "dimension" will be used non-rigorously but with the aim of motivating the desired idea. You will begin with examples and non-examples of 1-dimensional smooth manifolds, also known as *smooth curves*.

Example 4.6.1 Three examples of curves in $\mathbb{R}^2$ include a circle given by $x^2+y^2=1$, a cusp given by $x^2=y^3$, and a figure eight given by $x^4=x^2-y^2$.



Play with each Desmos activity (circle, cusp, figure 8) for a better visual.

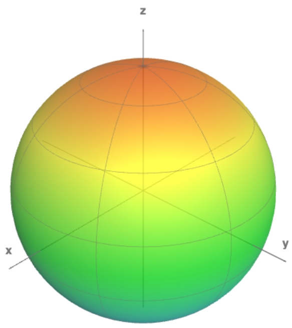
For almost every point p on each curve $S \subseteq \mathbb{R}^2$, you can see that the tangent space T_pS is usually a 1-dimensional subspace of $\mathbb{R}^2$. For the circle, every point satisfies this property, so it should presumably be a smooth curve. The cusp and figure 8, however, should not be.

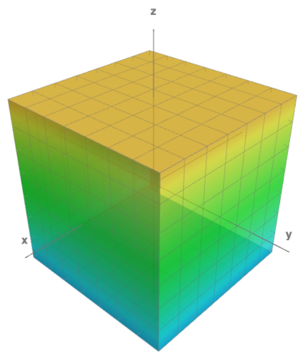
For the cusp, there is an issue at the origin $p=(0,0)$. If a particle on the cusp $S \subseteq \mathbb{R}^2$ moves through the origin, then it must always "stop" and turn around which suggests the tangent space at the origin is only the zero vector $T_{(0,0)}S = \{(0,0)\}$. The dimension of the tangent space appears to drop from 1 to 0 here, and that seems problematic.

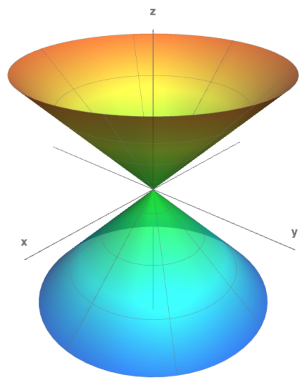
For the figure eight $S \subseteq \mathbb{R}^2$, there is a worse issue at the origin. The tangent space appears to be the union of two subspaces, namely $T_{(0,0)}S = \{(u,v) \in \mathbb{R}^2 : u = \pm v\}$. This is *not* a subspace of $\mathbb{R}^2$. That is terrible!

This example illustrates that the circle, cusp, and figure eight look like 1-dimensional smooth manifolds everywhere except possibly at some troublesome points. You will want to avoid these troublesome points. Next, you can consider some classic examples and non-examples of 2-dimensional smooth manifolds in $\mathbb{R}^3$, which are also known as *smooth surfaces*.

Example 4.6.2 Three examples of surfaces in $\mathbb{R}^3$ include a sphere given by $x^2+y^2+z^2=16$, a cube given by $\partial[-3,3]^3$, and a two-sided cone given by $z^2=x^2+y^2$.







View these on Math3D. Again, at almost every point p on one of these sets $S \subseteq \mathbb{R}^3$, you can see that the tangent space T_pS is usually a 2-dimensional subspace of $\mathbb{R}^3$. For the sphere, every point satisfies this property, so the sphere should presumably be a smooth surface. The cone and cube, however, should not be smooth surfaces.

For the cube, the edges and corners appear to have an issue; the tangent space at those points does not appear to be a 2-dimensional subspace of $\mathbb{R}^3$. They are possibly either 1-dimensional or 0-dimensional, so the dimension drops again.

For the cone, the origin $(0,0,0)$ has a more serious issue. A particle can pass straight through the origin along the cone, so every straight line lying on the cone should belong to the tangent plane. This means the tangent space of the cone S at the origin should be entire cone itself! That is, you may strangely expect $T_{(0,0,0)}S = S$. A cone looks nothing like a subspace of $\mathbb{R}^3$, so this is a rather alarming outcome.

You will again want to identify the troublesome points in these examples. Unsurprisingly, there are also more issues in higher dimensions.

Example 4.6.3 Consider the three sets defined by three different equations

$$\begin{aligned}
A &= \{(x_1,\dots,x_n) \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\},\\
B &= \{(xy, y^2+z, xyz, ye^{xz}) : x,y,z \in \mathbb{R}\},\\
C &= \{(x_1,\dots,x_n) \in \mathbb{R}^n : x_1 \cdots x_n = 137, x_1 + x_2^2 + x_3^3 + \cdots + x_n^n = 237\}.
\end{aligned}$$

These sets are not graphs of a function and they are now impossible to adequately visualize. Which of these should be smooth manifolds and what is their dimension? You only have algebraic descriptions, so you will need some serious tools to handle these mysterious situations. Remember high dimensional smooth manifolds are not just a mathematician's playground; they naturally arise in many applications such as general relativity in physics, data sets in statistics, cost constraints in economics, and models in machine learning.

These pose serious challenges to defining a smooth manifold using tangent spaces. The first informal guess was flawed because tangent spaces for many reasonable sets in $\mathbb{R}^n$ do not need to have constant dimension (when they are a subspace) and, even worse, they do not need to look anything like subspaces of $\mathbb{R}^n$. However, if the set S is a *graph* of a differentiable function, then Theorem 4.5.11 shows that the tangent space at every point is always a subspace of $\mathbb{R}^n$ with the same dimension. The ingenious idea is to instead define smooth manifolds *using graphs*, so you can exploit this theorem to make conclusions about tangent spaces.

4.6.2 Definition of a smooth manifold

Now, you can return to the goal of rigorously defining a "smooth manifold". The first informal guess was that a smooth manifold should be locally well-approximated by a *plane* but this is too hard to verify directly. The second informal guess is that a smooth manifold should be locally well-approximated by a *graph*. Here is an equivalent informal phrasing.

> *A set S in $\mathbb{R}^n$ is a k-dimensional smooth manifold if you zoom in close enough and it always looks like the graph of a function with k-variables in $\mathbb{R}^n$.*

This idea is magical and finally produces a good definition.

Definition 4.6.4 Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. The set S is a **(k-dimensional) smooth manifold**[4] **at** p if there exists an open set $U \subseteq \mathbb{R}^n$ containing p such that $S \cap U$ is a graph of a C^1 function $f : V \to \mathbb{R}^{n-k}$ where $V \subseteq \mathbb{R}^k$ is open.

Remark 4.6.5 There are several subtleties. First, the function $f : V \to \mathbb{R}^{n-k}$ may depend on the point p. Second, the choice of f is *not* necessarily unique. Third, the set $S \cap U$ is *a* graph of f, and not *the* graph of f; see Definition 1.5.6 for details. These comments probably do not make much sense now, but you will see some of these through examples.

One aspect of this definition is perhaps not immediately intuitive.

> *How does the set $S \cap U$ for an open set U containing p correspond to "zooming into" the set S near p? Why not replace U with a small open ball instead?*

This natural suggestion is actually equivalent.

Lemma 4.6.6 Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. The set S is a k-dimensional smooth manifold at p if and only if there exists $\varepsilon > 0$ such that $S \cap B_\varepsilon(p)$ is a graph of a C^1 function $f : V \to \mathbb{R}^{n-k}$ where $V \subseteq \mathbb{R}^k$ is open.

Proof. (Sketch) The "if" direction is immediate since an open ball is an open set. The "only if" direction requires more work, but we shall sketch the details. Assume S is a k-dimensional smooth manifold at p. By definition, there exists an open set $U \subseteq \mathbb{R}^n$ containing p such that $S \cap U$ is a graph of C^1 function $f : V \to \mathbb{R}^{n-k}$ where $V \subseteq \mathbb{R}^k$ is open. Without loss of generality, you can permute the coordinates in $\mathbb{R}^n$ and assume that $S \cap U$ is *the* graph of $f : V \to \mathbb{R}^{n-k}$ so

$$S \cap U = \{(x, f(x)) : x \in V\}$$

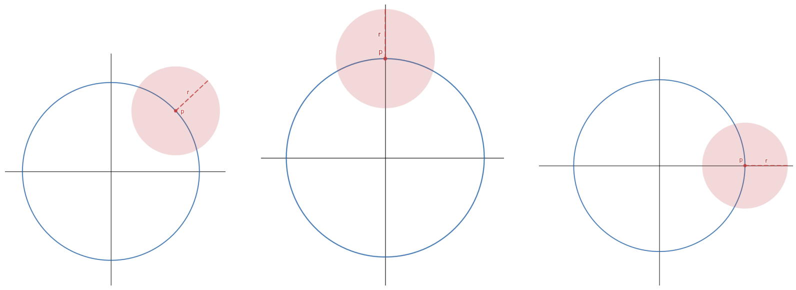
Since U is open and contains p, there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. Let $\pi : \mathbb{R}^n \to \mathbb{R}^k$ be the projection map onto the first k coordinates. Define $V' = \pi(B_\varepsilon(p)) \cap V$ and the restriction $g := f|_{V'} : V' \to \mathbb{R}^{n-k}$ so $g(x) = f(x)$ for all $x \in V'$. With some additional effort, you can prove that V' is open and $S \cap B_\varepsilon(p)$ is the graph of $g : V' \to \mathbb{R}^{n-k}$. ■

This equivalent definition of smooth manifolds with small open balls (Lemma 4.6.6) is helpful for building your geometric intuition.

Example 4.6.7 Consider the circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$ at three different points p lying in

$$\{(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}), (0,1), (1,0)\}$$

with a ball $B_\varepsilon(p)$ for a small fixed radius $\varepsilon > 0$. Play with this Desmos graph.



The radius $\varepsilon > 0$ is small enough so that $B_\varepsilon(p) \cap S$ is a graph of a C^1 function for each p.

- At $p = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, the set $B_\varepsilon(p) \cap S$ is a piece of the *upper* half of the circle. Hence, for some open interval I with $\frac{1}{\sqrt{2}} \in I$, this is a graph $y = f(x)$ where $f : I \to \mathbb{R}$ is C^1 given by $f(x) = \sqrt{1-x^2}$. Also, $B_\varepsilon(p) \cap S$ is a piece of the *right* half of the circle. Hence, for some open interval I' with $\frac{1}{\sqrt{2}} \in I'$, this is also a graph $x = g(y)$ where $g : I' \to \mathbb{R}$ is C^1 given by $g(y) = \sqrt{1-y^2}$. This shows the choice of graph is not unique.
- At $p = (0,1)$, the set $B_\varepsilon(p) \cap S$ is a piece of the upper half of the circle, so for some open interval I with $0 \in I$, it is a graph $y = \sqrt{1-x^2}$ for $x \in I$. However, no matter how small you choose ε, you can prove that the set $B_\varepsilon(p) \cap S$ *cannot* be written as a graph of the form $x = g(y)$. Informally, it always fails the horizontal line test.
- At $p = (1,0)$, the set $B_\varepsilon(p) \cap S$ is the graph $x = \sqrt{1-y^2}$ for $y \in I$ and some open interval I. Again, for any $\varepsilon > 0$, this *cannot* be written as a graph of the form $y = f(x)$.

Overall, the circle is a 1-dimensional smooth manifold at each of these three points. You can prove these observations formally but the focus is currently on intuition.

Lemma 4.6.6 provides intuition. However, it is cumbersome to use in a formal proof because it can be difficult to precisely determine the intersection of the set with an open ball. Definition 4.6.4 is much more flexible in a formal proof.

Example 4.6.8 Here is a formal proof that the circle $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is a 1-dimensional smooth manifold at $p = (1,0)$.

Proof. Let $U = \{(x,y) \in \mathbb{R}^2 : x > 0\}$ so U is open and $p = (1,0) \in U$. By definition,

$$S \cap U = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1, x > 0\}.$$

Define the open set $V = (-1,1) \subseteq \mathbb{R}$ and $f : V \to \mathbb{R}$ by

$$f(y) = \sqrt{1-y^2} \quad \text{for } y \in V.$$

Note f is C^1 on its domain. It follows that for $(x,y) \in \mathbb{R}^2$,

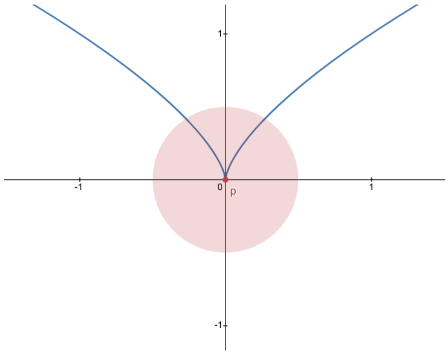
$$\begin{aligned}(x,y) \in S \cap U &\iff x^2 + y^2 = 1, x > 0 \\ &\iff x = \sqrt{1-y^2} \text{ and } -1 < y < 1 \\ &\iff x = f(y) \text{ and } y \in V\end{aligned}$$

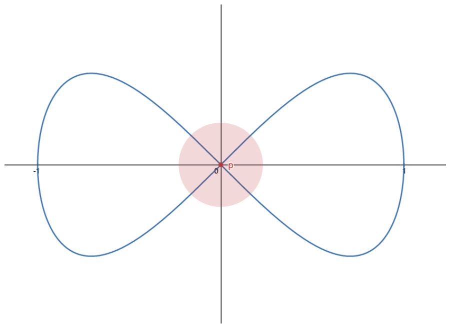
Thus, $S \cap U = \{(f(y), y) : y \in V\}$ is a graph of the 1-variable C^1 real-valued function f. By Definition 4.6.4, S is a 1-dimensional smooth manifold at $p = (1,0)$. ■

Remark 4.6.9 Notice the set U was cleverly chosen to make the equivalences easy to verify.

Next, you can revisit some non-examples of 1-dimensional smooth manifolds.

Example 4.6.10 Consider the cusp $x^2 = y^3$ and figure eight $x^4 = x^2 - y^2$ in $\mathbb{R}^2$. Neither of them are a 1-dimensional smooth manifold at the origin $p = (0,0)$.





If S is the cusp then, for any $\varepsilon > 0$, the set $B_\varepsilon(p) \cap S$ *cannot* be written as a C^1 graph of the form $y = f(x)$ because $f'(0)$ cannot exist, and it *cannot* be written as a C^1 graph of the form $x = g(y)$ because it will fail the horizontal line test. Play with the Desmos cusp.

If S is the figure eight then, for any $\varepsilon > 0$, the set $B_\varepsilon(p) \cap S$ *cannot* be written as C^1 graph of either form $y = f(x)$ or $x = g(y)$ because it will always fail the vertical and horizontal line tests. Play with the Desmos figure eight.

Again, the claims in Example 4.6.10 can be proved formally, but it takes much more effort as Definition 4.6.4 is quite delicate.

Example 4.6.11 Let $S = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3\}$ so S is the cusp. Here you will prove that S is not a 1-dimensional smooth manifold at the origin $p = (0,0)$.

Proof. Suppose, for a contradiction, that S is a 1-dimensional smooth manifold at $(0,0)$. There exists an open set $U \subseteq \mathbb{R}^2$ containing $(0,0)$ such that $S \cap U$ is a graph of a C^1 function $f : V \to \mathbb{R}$ where $V \subseteq \mathbb{R}$ is open. This implies that one of the following holds:

$$\forall (x,y) \in \mathbb{R}^2, (x,y) \in S \cap U \iff y = f(x), x \in V \tag{4.6.1}$$
$$\forall (x,y) \in \mathbb{R}^2, (x,y) \in S \cap U \iff x = f(y), y \in V \tag{4.6.2}$$

Assume the first statement holds. We claim it suffices to show that there exists $\delta > 0$ such that $f(x) = x^{2/3}$ for $x \in (-\delta, \delta)$. Assuming the claim, it follows that

$$f'(0) = \lim_{x\to 0} \frac{x^{2/3} - 0}{x - 0} = \lim_{x \to 0} x^{-1/3}$$

does not exist, so f is not C^1, a contradiction. It remains to prove the claim. Since U is open and contains $(0,0)$, there exists $\varepsilon > 0$ such that $B_{2\varepsilon}((0,0)) \subseteq U$. Take $\delta = \min\{\varepsilon, \varepsilon^{3/2}\}$ and let $x \in (-\delta, \delta)$. Note that $(x, x^{2/3}) \in S$ by definition of S. As $|x| < \delta$,

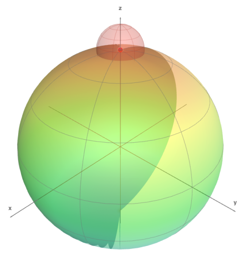
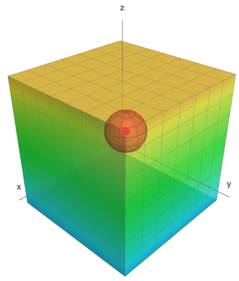
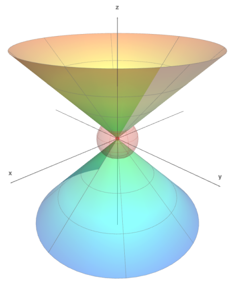
$$||(x, x^{2/3})|| = \sqrt{x^2 + x^{4/3}} \leq \sqrt{\varepsilon^2 + \varepsilon^2} < 2\varepsilon$$

in which case $(x, x^{2/3}) \in U$. By Equation (4.6.1), it follows that $f(x) = x^{2/3}$ as required.

Assume the second statement holds. Since U is open and contains $(0,0)$, there exists $0 < \varepsilon < 1$ such that $B_{2\varepsilon}(0,0) \subseteq U$. The points $(\pm\varepsilon^3, \varepsilon^2)$ belong to the set S by definition and, since $0 < \varepsilon < 1$, you can verify that $||(\pm\varepsilon^3, \varepsilon^2)|| = \sqrt{2\varepsilon^6} \leq 2\varepsilon$. This means that the two points $(\pm\varepsilon^3, \varepsilon^2)$ belong to $S \cap U$. By (4.6.2), it follows that $f(\varepsilon^2) = \varepsilon^3$ and $f(\varepsilon^2) = -\varepsilon^3$. This contradicts that f is a function. ∎

These outcomes (both informal and formal) for 1-dimensional smooth manifolds in $\mathbb{R}^2$ demonstrate that Definition 4.6.4 (or equivalently Lemma 4.6.6) is effective, because it is precisely identifying the troublesome points. The same benefits carry over to sets in $\mathbb{R}^3$.

Example 4.6.12 Again, consider the sphere, cube, and cone from Example 4.6.2.

Play with these Math3D demos (sphere, cube, cone).

If $S \subseteq \mathbb{R}^3$ is the sphere and $p = (0,0,1)$, then the set $B_\varepsilon(p) \cap S$ in the picture is a piece of the upper hemisphere. Thus, it can be written as a C^1 graph $z = f(x,y)$ for $f : U \to \mathbb{R}$ given by $f(x,y) = \sqrt{16 - x^2 - y^2}$ and $U = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < \delta^2\}$ for some $\delta > 0$. This implies the sphere is a 2-dimensional smooth manifold at p. Note, however, the set $B_\varepsilon(p) \cap S$ cannot be written as a C^1 graph of the form $y = g(x,z)$ or $x = h(y,z)$.

If S is the cube and $p = (1,1,1)$ is the corner, then the set $B_\varepsilon(p) \cap S$ cannot be written as a C^1 graph of the form $z = f(x,y)$, $y = g(x,z)$ or $x = h(y,z)$. That is because all three sides of the cube meet at the corner. Hence, the cube is not a 2-dimensional smooth manifold at the corner $p = (1,1,1)$. The same is true for any point along the edge.

If S is the cone and $p = (0,0,0)$, then the set $B_\varepsilon(p) \cap S$ cannot be written as a C^1 graph of the form $z = f(x,y)$, $y = g(x,z)$ or $x = h(y,z)$. That is because any line parallel to any axis will intersect the cone at two points (except the axes themselves). Hence, the cone is not a 2-dimensional smooth manifold at the origin $p = (0,0,0)$.

As usual, all of these claims can be formally proved with Definition 4.6.4 by arguments similar to Examples 4.6.8 and 4.6.11.

By comparing the corresponding examples, you may notice that a set is not a smooth manifold at a point precisely when the tangent space fails to be a subspace with the expected dimension. This suggests you have discovered a good definition for a smooth manifold.

Definition 4.6.13 Fix $k, n \in \mathbb{N}^+$ with $k < n$. A set $S \subseteq \mathbb{R}^n$ is a **(k-dimensional) smooth manifold** if S is a (k-dimensional) smooth manifold at every point in S.

Remark 4.6.14 A 2-dimensional smooth manifold $S \subseteq \mathbb{R}^3$ is also called a **smooth surface** and a 1-dimensional smooth manifold $C \subseteq \mathbb{R}^n$ is also called a **smooth curve**.

Example 4.6.15 You can verify that the circle is a smooth curve. From the observations in Example 4.6.10, you can prove that the cusp and figure eight are not smooth curves since they are not smooth curves at the origin $(0,0)$.

Example 4.6.16 You can verify that the sphere is a smooth surface. From the observations in Example 4.6.12, you can prove that the cube and cone are not smooth surfaces since they are not smooth surfaces at the corner $(1,1,1)$ and the origin $(0,0,0)$ respectively.

After some significant labour, you have finally created a definition for a smooth manifold using graphs. It is therefore sensible that a graph of a C^1 function is itself a smooth manifold.

Lemma 4.6.17 Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let $S \subseteq \mathbb{R}^n$ be a graph of a C^1 function $F : V \to \mathbb{R}^{n-k}$ with $V \subseteq \mathbb{R}^k$ open. The set S is a k-dimensional smooth manifold.

Proof. By permuting coordinates in $\mathbb{R}^n$, you may assume without loss of generality that S is *the* graph of F, so $S = \{(x, F(x)) : x \in V\}$ by definition. The rest of the proof is left as an exercise. Once you choose an appropriate set U in Definition 4.6.4, the proof is very short. ■

4.6.3 Tangent space of a smooth manifold

By definition, a set is a k-dimensional smooth manifold at a point p if it is locally a k-variable graph at p. How are the tangent spaces at p of the smooth manifold and its local graph related?

Theorem 4.6.18 (Tangent spaces are local) Let S and S' be subsets of $\mathbb{R}^n$ with $p \in S \cap S'$. If there exists an open set $U \subseteq \mathbb{R}^n$ containing p such that $S \cap U = S' \cap U$, then $T_pS = T_pS'$.

Informally speaking, this theorem states:

If two sets are locally the same near p, then their tangent spaces at p are the same.

Proof. By symmetry, it is enough to show $T_pS \subseteq T_pS'$. Fix a tangent vector $v \in T_pS$ so there exists a differentiable $\gamma : I \to \mathbb{R}^n$ with $I \subseteq \mathbb{R}$ open, $\gamma(I) \subseteq S$, $\gamma(0) = p$, and $\gamma'(0) = v$. Since γ is differentiable and hence continuous at 0, I is open, U is open and contains p, and $\gamma(0) = p$, there exists $\delta > 0$ such that $I' = (-\delta, \delta) \subseteq I$ and $\gamma((-\delta, \delta)) \subseteq U$.

Set $g = \gamma|_{I'}$ so $g : I' \to \mathbb{R}^n$ is given by $g(t) = \gamma(t)$ for $t \in I'$. We will show v is a tangent vector of the set S' at p using the parametric curve g. By definition,

$$g(0) = \gamma(0) = p, \qquad g'(0) = \gamma'(0) = v, \qquad \text{and} \qquad g(I') = \gamma(I') \subseteq \gamma(I) \subseteq S.$$

Also, $g(I') = \gamma(I') \subseteq U$ so overall $g(I') \subseteq S \cap U = S' \cap U \subseteq S'$. Thus, by definition, $v \in T_pS'$. ■

This creates the final bridge you need to reward your monumental efforts with smooth manifolds.

Corollary 4.6.19 Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let $S \subseteq \mathbb{R}^n$ be a set and let p be a point in S. If S is a k-dimensional smooth manifold at p, then the tangent space T_pS is a k-dimensional subspace of $\mathbb{R}^n$.

Proof. Since S is a k-dimensional smooth manifold at p, there exists an open set $U \subseteq \mathbb{R}^n$ such that $S \cap U$ is a graph S' of a C^1 function $F : V \to \mathbb{R}^{n-k}$ with $V \subseteq \mathbb{R}^k$ open. Since $S' = S' \cap U$ it follows that $S \cap U = S' \cap U$ so, by Theorem 4.6.18, $T_pS = T_pS'$. By Theorem 4.5.11, T_pS' is a k-dimensional subspace of $\mathbb{R}^n$. ■

In other words, at every point of a smooth manifold, the tangent space is always a subspace of the same dimension! That is exactly the original idea at the start of this section.

This corollary is a great accomplishment and capstone to this chapter on applications of derivatives. Nonetheless, you must ask: how do you check if a set is a smooth manifold? If the set is a graph, then the answer is immediate by Lemma 4.6.17. If the set is written in implicit form, then the examples with the circle and sphere show that you can directly find the appropriate graph. For instance, you can solve the implicit equation

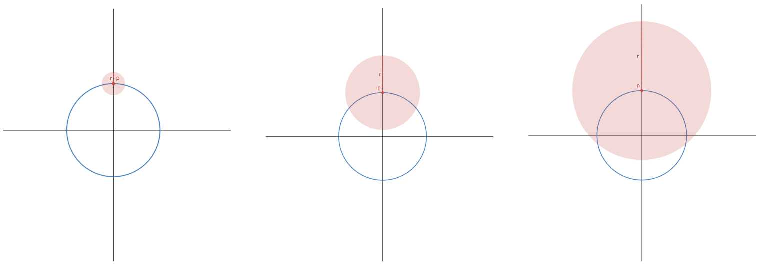
$$x^2 + y^2 = 1$$

for either x or y, depending on what is needed. However, as Example 4.6.3 illustrates, solving for variables in terms of others is tremendously difficult and sometimes impossible to do. So how do you verify whether a set in implicit form is a smooth manifold *without* actually solving for variables? This grand mystery launches you into the next chapter, where you will study implicit functions and ultimately find a solution to optimizing over sets in implicit form.

Exercises for Section 4.6

Concepts and definitions

4.6.1 Let $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ be the unit circle. Fix the point $p = (0, 1) \in S$. The Desmos plots below illustrate the sets $B_\varepsilon(p) \cap S$ for three different values $\varepsilon \in \{0.1, 0.7, 1.6\}$.

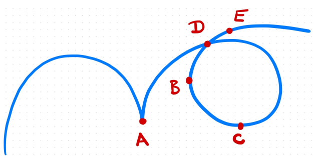


(a) For which of the 3 values of ε is $B_\varepsilon(p) \cap S$ the graph of a C^1 function of the form $y = f(x)$?

(b) For which of the 3 values of ε is $B_\varepsilon(p) \cap S$ the graph of a C^1 function of the form $x = f(y)$?

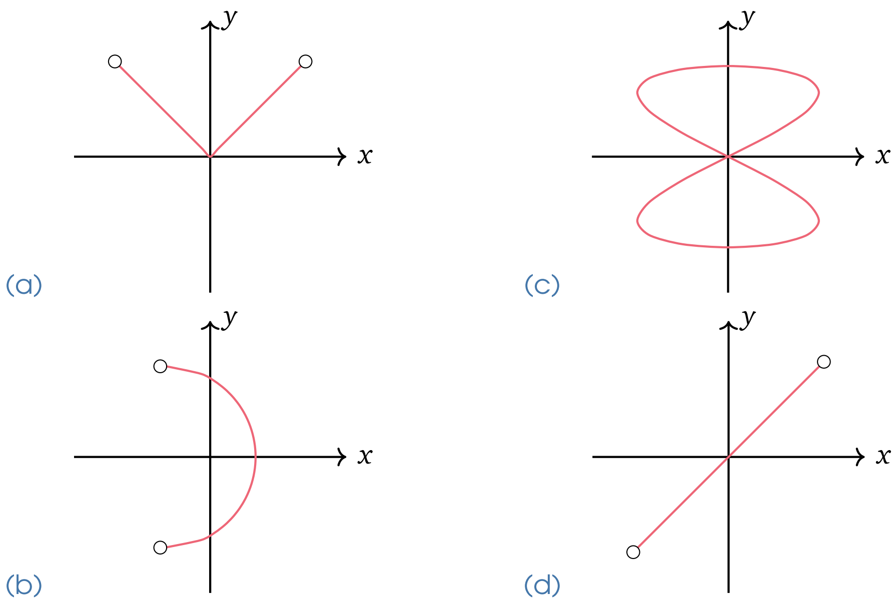
(c) For which of the 3 values of ε is $B_\varepsilon(p) \cap S$ a graph of a C^1 function?

4.6.2 Below is a plot of a curve $S \subseteq \mathbb{R}^2$. Five points A, B, C, D, E are listed on the curve.



At which points is S a 1-dimensional smooth manifold?

4.6.3 Identify which of the sets are 1-dimensional smooth manifolds. If not, identify where the set fails to be a 1-dimensional smooth manifold.

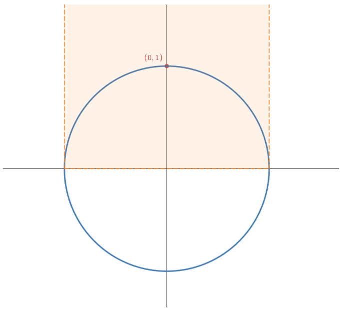


4.6.4 Define the set $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ and fix $p = (0, 1)$. Recall:

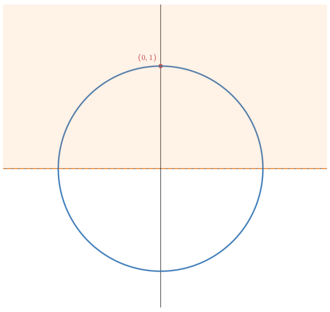
> *The set S is a 1-dimensional smooth manifold at p if and only if there exists an open set $U \subseteq \mathbb{R}^2$ containing p such that $S \cap U$ is a graph of a C^1 map $f : V \to \mathbb{R}$ where $V \subseteq \mathbb{R}$ is open.*

Which choice of $U \subseteq \mathbb{R}^2$ can successfully lead to a correct proof? If so, define $f : V \to \mathbb{R}$. Play with this Desmos graph for better visuals.

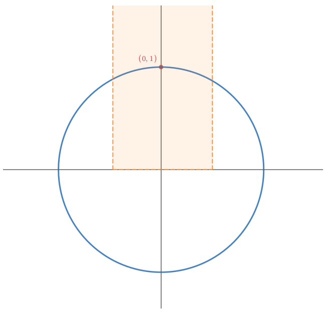
(a) $U = (-1, 1) \times (0, \infty)$



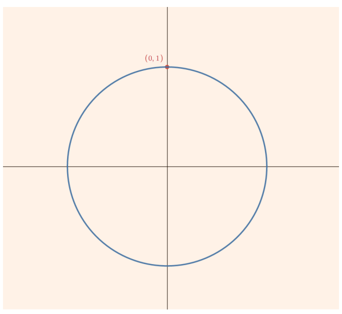
(b) $U = \mathbb{R} \times (0, \infty)$



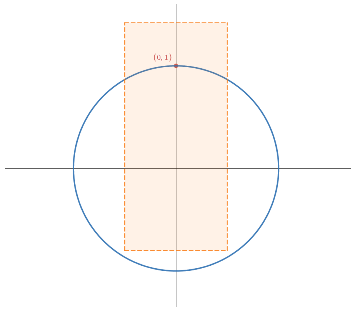
(c) $U = (-0.5, 0.5) \times (0, \infty)$



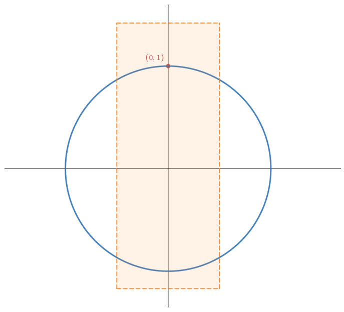
(d) $U = \mathbb{R}^2$



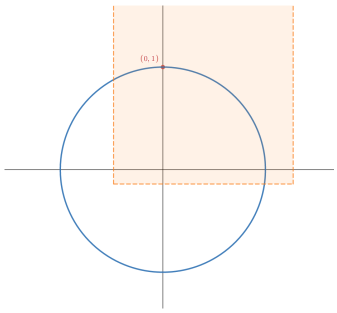
(e) $U = (-0.5, 0.5) \times (-0.8, 1.5)$



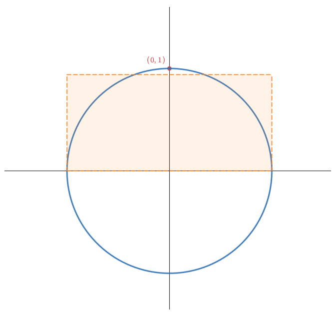
(f) $U = (-0.5, 0.5) \times (-1.2, 1.5)$



(g) $U = (-0.5, 1.2) \times (-0.1, \infty)$

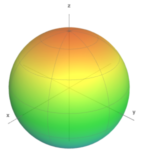


(h) $U = (-1, 1) \times (0, 0.9)$

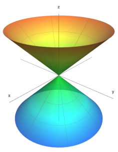


4.6.5 Identify which of the sets are 2-dimensional smooth manifolds. If a set is not a smooth manifold, identify at which points the definition fails. View this Math3D demo for better visuals.

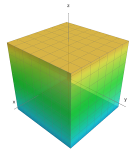
(a) A sphere



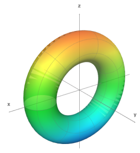
(b) A double cone



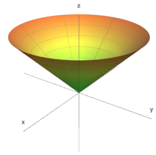
(c) A cube



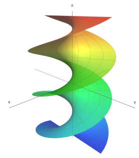
(d) A torus



(e) A single cone



(f) A helicoid

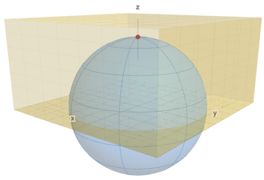


4.6.6 Define the set $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 16\}$ and fix $p = (0,0,4) \in S$. Recall:

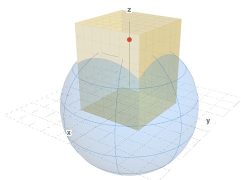
> *The set S is a 2-dimensional smooth manifold at p if and only if there exists an open set $U \subseteq \mathbb{R}^3$ containing p such that $S \cap U$ is a graph of a C^1 map $f : V \to \mathbb{R}$ where $V \subseteq \mathbb{R}^2$ is open.*

Which choice of $U \subseteq \mathbb{R}^2$ can successfully lead to a correct proof? If so, define $f : V \to \mathbb{R}$. Play with this Math3D box demo and this Math3D cylinder demo for better visuals.

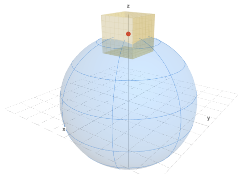
(a) $U = (-5,5) \times (-5,5) \times (0,\infty)$



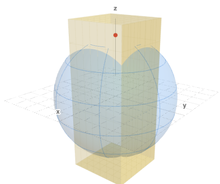
(b) $U = (-2,2) \times (-2,2) \times (0,\infty)$



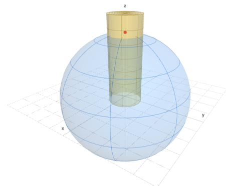
(c) $U = (-1,1) \times (-1,1) \times (3,5)$



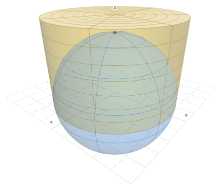
(d) $U = (-2,2) \times (-2,2) \times (-\infty,\infty)$



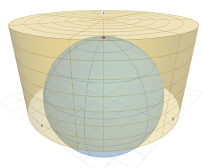
(e) $U = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 < 1, z > 0\}$



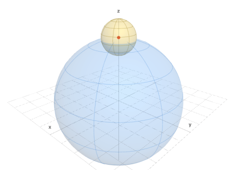
(f) $U = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 < 16, z > 0\}$



(g) $U = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 < 25, z > 0\}$



(h) $U = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+(z-4)^2 < 1\}$



Computations

4.6.7 Let $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ be the unit circle. Fix $p = (0,1) \in S$. No justification is necessary for any part below.

(a) For which values of $\varepsilon > 0$ is $B_\varepsilon(p) \cap S$ a graph of the form $y = f(x)$ for a C^1 function f?

(b) For which values of $\varepsilon > 0$ is $B_\varepsilon(p) \cap S$ a graph of the form $x = g(y)$ for a C^1 function g?

4.6.8 Let $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ be the unit sphere. Fix $p = (0,0,1) \in S$. No justification is necessary for any part below.

(a) For which values of $\varepsilon > 0$ is $B_\varepsilon(p) \cap S$ a graph of the form $z = f(x,y)$ for a C^1 function f?

(b) For which values of $\varepsilon > 0$ is $B_\varepsilon(p) \cap S$ a graph of the form $y = g(x,z)$ for a C^1 function g?

(c) For which values of $\varepsilon > 0$ is $B_\varepsilon(p) \cap S$ a graph of the form $x = h(y,z)$ for a C^1 function h?

4.6.9 Since smooth manifolds are locally graphs and you can compute the tangent space of graphs, you can compute the tangent space of a smooth manifold. You will implement this strategy with the 1-dimensional smooth manifold $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 2\}$ at the point $p = (1,1) \in S$.

(a) Exhibit without proof an open set $U \subseteq \mathbb{R}^2$ containing p and a C^1 function $f : V \to \mathbb{R}$ with $V \subseteq \mathbb{R}$ open such that $S \cap U$ is a graph of f.

(b) Define the set $S' = S \cap U$. Compute the tangent space $T_p(S')$ of the graph S' at the point p.

(c) Compute the tangent space T_pS of the smooth manifold S at the point p.

4.6.10 Define smooth manifold $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 3\}$ and fix $p = (1,1,1) \in S$.

(a) Exhibit without proof an open set $U \subseteq \mathbb{R}^3$ containing p and a C^1 function $f : V \to \mathbb{R}$ with $V \subseteq \mathbb{R}^2$ open such that $S \cap U$ is a graph of f.

(b) Define the set $S' = S \cap U$. Compute the tangent space $T_p(S')$ of the graph S' at the point p.

(c) Compute the tangent space T_pS of the smooth manifold S at the point p.

Proofs

4.6.11 The unit circle

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$

is the model example for studying smooth manifolds. If you thoroughly investigate this single example, then you will gain a strong grasp of the key ideas.

(a) Andy describes a strategy to prove that S is a 1-dimensional smooth manifold at $p = (0,1)$.

> 1. *The upper half of the circle is the graph* $y = \sqrt{1-x^2}$ *for* $-1 < x < 1$.
> 2. *The point* $p = (0,1)$ *lies on this graph.*
> 3. *I can use this graph to prove that* S *is a 1-dimensional smooth manifold at* $p = (0,1)$.

Andy has the key idea. To begin converting this outline into a formal proof, fill in the blanks.

- *Define the set* $U =$ ____________________ *so* U *is open and* $p = (0,1) \in U$.
- *Define* $f : V \to \mathbb{R}$ *by* $V =$ ____________ *and* $f(t) =$ ____________ *for* $t \in V$.
- *It suffices to prove the following equality of sets:*

$$\underline{\hspace{3cm}} = \underline{\hspace{10cm}}.$$

Now, finish the proof.

(b) Cameron describes a strategy to prove that S is a smooth curve.

> 1. *The upper half of the circle is the graph* $y = \sqrt{1-x^2}$ *for* $-1 \le x \le 1$.
> 2. *The lower half of the circle is the graph* $y = -\sqrt{1-x^2}$ *for* $-1 \le x \le 1$.
> 3. *Since* S *is the union of these two graphs, I can use these two graphs to show that* S *is a smooth curve.*

Cameron has some good ideas, but needs more. Why are these two graphs not enough?

(c) Jinan describes a strategy to prove that S is a smooth curve.

> 1. *The upper half of the circle is the graph* $y = \sqrt{1-x^2}$ *for* $-1 < x < 1$.
> 2. *The lower half of the circle is the graph* $y = -\sqrt{1-x^2}$ *for* $-1 < x < 1$.
> 3. *The right half of the circle is the graph* $x = \sqrt{1-y^2}$ *for* $-1 < y < 1$.
> 4. *The left half of the circle is the graph* $x = -\sqrt{1-y^2}$ *for* $-1 < y < 1$.
> 5. *Since* S *is the union of these four graphs, the set* S *is a 1-dimensional smooth manifold.*

Jinan's suggestion has all the key ideas, but there is a serious logical error. Identify it.

4.6.12 Define the cusp $S = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3\}$. The proof that S is not a 1-dimensional smooth manifold at $(0,0)$ is quite delicate, so you will further analyze it here.

(a) The first few steps are below. In Line 3, why are (4.6.3) and (4.6.4) possible scenarios instead of just (4.6.3)?

> 1. *Suppose, for a contradiction, that* S *is a 1-dimensional smooth manifold at* $(0,0)$.
> 2. *There exists an open set* $U \subseteq \mathbb{R}^2$ *containing* $(0,0)$ *such that* $S \cap U$ *is a graph of a* C^1 *function* $f : V \to \mathbb{R}$ *where* $V \subseteq \mathbb{R}$ *is open.*
> 3. *This implies that one of the following holds:*
>
> $$\forall (x,y) \in \mathbb{R}^2, (x,y) \in S \cap U \iff y = f(x) \text{ and } x \in V \tag{4.6.3}$$
> $$\forall (x,y) \in \mathbb{R}^2, (x,y) \in S \cap U \iff x = f(y) \text{ and } y \in V \tag{4.6.4}$$

(b) The next few steps continue. How exactly is (4.6.3) applied? Add one more line of explanation.

> 4. *Assume the first statement holds.*
> 5. *We claim it suffices to show that there exists* $\delta > 0$ *such that* $f(x) = x^{2/3}$ *for* $x \in (-\delta, \delta)$.
> 6. *Assuming the claim, it follows that* $f'(0) = \lim_{x \to 0} \frac{x^{2/3} - 0}{x - 0} = \lim_{x \to 0} x^{-1/3}$ *does not exist.*
> 7. *Hence,* f *is not* C^1*, a contradiction. It remains to prove the claim.*
> 8. *Since* U *is open and contains* $(0,0)$*, there exists* $\varepsilon > 0$ *such that* $B_{2\varepsilon}((0,0)) \subseteq U$.
> 9. *Take* $\delta = \min\{\varepsilon, \varepsilon^{3/2}\}$ *and let* $x \in (-\delta, \delta)$.
> 10. *Note that* $(x, x^{2/3}) \in S$ *by definition of* S.
> 11. *As* $|x| < \delta$, $\|(x, x^{2/3})\| = \sqrt{x^2 + x^{4/3}} \le \sqrt{\varepsilon^2 + \varepsilon^2} \le 2\varepsilon$ *in which case* $(x, x^{2/3}) \in U$.
> 12. *By* (4.6.3)*, it follows that* $f(x) = x^{2/3}$ *as required.*

(c) Sketch a picture proof of the last few steps below. Label all relevant quantities.

13. *Assume the second statement holds.*
14. *Since U is open and contains $(0,0)$, there exists $0 < \varepsilon < 1$ such that $B_{2\varepsilon}(0,0) \subseteq U$.*
15. *The points $(\pm\varepsilon^3, \varepsilon^2)$ belong to the set S by definition.*
16. *Since $0 < \varepsilon < 1$, you can verify that $||(\pm\varepsilon^3, \varepsilon^2)|| = \sqrt{2\varepsilon^6} \le 2\varepsilon$.*
17. *This means that the two points $(\pm\varepsilon^3, \varepsilon^2)$ belong to $S \cap U$.*
18. *By (4.6.4), it follows that $f(\varepsilon^2) = \varepsilon^3$ and $f(\varepsilon^2) = -\varepsilon^2$. This contradicts that f is a function.*

4.6.13 Below is the proof of Theorem 4.6.18, which asserts that tangent spaces are local. In other words, if two sets are locally the same near p, then their tangent spaces at p are the same.

1. *By symmetry, it is enough to show $T_pS \subseteq T_pS'$.*
2. *Fix a tangent vector $v \in T_pS$.*
3. *There exists a differentiable $\gamma : I \to \mathbb{R}^n$ with $I \subseteq \mathbb{R}$ open, $\gamma(I) \subseteq S$, $\gamma(0) = p$, and $\gamma'(0) = v$.*
4. *Since γ is continuous at $0 \in I$, I is open, and U is open, there exists $\delta > 0$ such that $(-\delta, \delta) \subseteq \gamma^{-1}(U)$.*
5. *Set $I' = (-\delta, \delta)$, so this implies $\gamma(I') \subseteq U$.*
6. *Define the restriction $g = \gamma|_{I'}$ so $g : I' \to \mathbb{R}^n$ is given by $g(t) = \gamma(t)$ for $t \in I'$.*
7. *Then $g(0) = \gamma(0) = p$ and $g'(0) = \gamma'(0) = v$.*
8. *Moreover, $g(I') = \gamma(I') \subseteq \gamma(I) \subseteq S$ and also $g(I') = \gamma(I') \subseteq U$.*
9. *Overall, this implies that $g(I') \subseteq S \cap U = S' \cap U \subseteq S'$.*
10. *Thus, $v \in T_pS'$ as desired.*

This proof is correct but you will fill in a few details.

(a) Line 1 mentions "by symmetry". Explain what this means and why this implies the conclusion.

(b) Line 4 does not include all the details. Further justify Line 4 with more details.

4.6.14 Let $S = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3\}$ and fix $(a,b) \in S$ with $(a,b) \neq (0,0)$. Prove that S is a 1-dimensional smooth manifold at (a,b).

4.6.15 Prove that the unit circle $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is a smooth curve.

4.6.16 Prove that the unit sphere $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ is a smooth surface.

4.6.17 Show that the square $S = \partial[-1,1]^2$ is not a 1-dimensional smooth manifold at $p = (1,1)$.

4.6.18 Show that the cube $S = \partial[-1,1]^3$ is not a 2-dimensional smooth manifold at $p = (1,1,1)$.

4.6.19 Let $S \subseteq \mathbb{R}^n$ be the graph of a C^1 function $F : U \to \mathbb{R}^{n-k}$ with $U \subseteq \mathbb{R}^k$ open. Prove that the set S is a k-dimensional smooth manifold.

4.6.20 Show that the set $S = \{(q,q) : q \in \mathbb{Q}\} \subseteq \mathbb{R}^2$ is not a 1-dimensional smooth manifold.

5. Inverse and implicit functions

To solve any optimization problem, you need to locate possible extrema on the boundary of a function's domain. These boundaries are often described as sets in implicit form. This poses an intriguing question.

How do you optimize over a set written in implicit form?

For instance, you may want to find the extrema of a real-valued function $f : \mathbb{R}^2 \to \mathbb{R}$ on the set S in implicit form given by

$$S = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$$

for some nonlinear map $g : \mathbb{R}^2 \to \mathbb{R}$. The equation $g(x, y) = 0$ is a *constraint*. Many real-world optimization problems have constraints. If you can parametrize S, then you can continue trying to solve this problem with your usual strategy. However, parametrizing a set in implicit form is not always possible or, at best, it can be incredibly difficult.

You will therefore need to understand how a function changes along a set in implicit form without parametrizing it. As a first step, you have defined smooth manifolds and you successfully determined their tangent planes, so differentiation on smooth manifolds should be possible. This leads to a deep question:

How can you determine whether a set S written in implicit form is a smooth manifold?

For instance, can the *nonlinear* equation $g(x, y) = 0$ be written locally as a graph $y = f(x)$? In other words, is y an *implicit function* of x? This can be stated more generally.

When does a nonlinear system of equations have an explicit set of solutions?

This is a fundamental problem in calculus and geometry that has many ramifications beyond optimization, such as to inverse functions and coordinate systems. In this chapter, you will study the most powerful and intricate tools in your multivariable calculus arsenal: the inverse function theorem and the implicit function theorem. Your reward will be a newfound strategy for optimizing with constraints.

5.1. Diffeomorphisms

Finding the inverse of a function is actually an example of solving a system of equations. Informally speaking, given a nonlinear map $F : \mathbb{R}^n \to \mathbb{R}^m$, you have a system of m nonlinear equations with n variables given by

$$y = F(x) \quad \Longleftrightarrow \quad \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} F_1(x) \\ \vdots \\ F_m(x) \end{bmatrix} \quad \Longleftrightarrow \quad \begin{array}{ccc} y_1 & = & F_1(x_1, \dots, x_n) \\ & \vdots & \\ y_m & = & F_m(x_1, \dots, x_n) \end{array}. \tag{5.1.1}$$

Solving this system means you want to express $x \in \mathbb{R}^n$ as a function of $y \in \mathbb{R}^m$. That is, you want to find a function $G : \mathbb{R}^m \to \mathbb{R}^n$ such that for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$,

$$y = F(x) \quad \Longleftrightarrow \quad x = G(y).$$

Equivalently, $F \circ G : \mathbb{R}^m \to \mathbb{R}^m$ and $G \circ F : \mathbb{R}^n \to \mathbb{R}^n$ should both be the identity maps.

Based on your experience with linear algebra, the dimensions m and n should not be arbitrary. Consider the special case of a linear map. A linear map $\mathbb{R}^n \to \mathbb{R}^m$ cannot be surjective when $n < m$ and it cannot be injective when $n > m$. The only reasonable scenario for a linear map to be invertible is therefore when $m = n$. Since your theory of differential calculus relies so heavily on linear algebra, this suggests that you should restrict your attention to finding inverse functions of *nonlinear* maps when

$$m = n.$$

This equality will always be assumed for your study of inverse functions. In this section, you will formally explain what it means to *globally* or *locally* solve the nonlinear system (5.1.1) when $m = n$. The resulting definitions will introduce invertible maps that preserve useful topological and geometric properties of sets, which you can interpret with coordinate systems.

5.1.1 Global diffeomorphisms

First, you will begin by defining a global solution to (5.1.1) when $m = n$.

Definition 5.1.1 Let U and V be subsets of $\mathbb{R}^n$. The **(global) inverse of** $F : U \to V$ is a map $G : V \to U$ satisfying $G \circ F(x) = x$ for all $x \in U$ and $F \circ G(y) = y$ for all $y \in V$. Equivalently,

$$\text{for all } x \in U \text{ and } y \in V, \qquad y = F(x) \quad \Longleftrightarrow \quad x = G(y).$$

The inverse of F is unique and denoted F^{-1}.

Remark 5.1.2 The abuse of notation of F^{-1} with the preimage is acceptable because these agree whenever F has an inverse.

Recall F is invertible if and only if F is bijective. Thus, assuming F is bijective, the inverse function F^{-1} produces a solution $x = F^{-1}(y)$ to (5.1.1) but it does not allow you to utilize differential calculus. As y varies, you may want to understand how x will vary and vice versa; that is, you will want both F and its inverse to be differentiable to understand how solutions vary. Is it enough to assume that F is C^1? Surprisingly, it is not.

Example 5.1.3 Consider the function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = x^3$. Notice F is C^1 and F is bijective with inverse given by $G : \mathbb{R} \to \mathbb{R}$ with $G(x) = x^{1/3}$. However, G is *not* differentiable at 0! This example illustrates that a bijective C^1 function may not have a C^1 inverse.

This motivates you to introduce a definition of inverse maps that requires differentiability.

Definition 5.1.4 Let U and V be open subsets of $\mathbb{R}^n$. A function $F : U \to V$ is a **(global) diffeomorphism**[1] if F is bijective, F is C^1, and its inverse function $F^{-1} : V \to U$ is C^1.

Remark 5.1.5 Many other texts will define diffeomorphisms to require stronger differentiability assumptions than C^1, or to use domains and codomains other than subsets of $\mathbb{R}^n$, e.g. MAT363 Curves and Surfaces or MAT367 Differential Geometry. The definition highly depends on the context, so pay attention to your source's conventions.

Example 5.1.6 Consider the function

$$F : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}, \qquad F(x) = \tan x.$$

Its derivative $F'(x) = \sec^2 x$ is continuous on its domain, so F is C^1. You can verify that the function F is bijective, and its inverse function of F is given by

$$G : \mathbb{R} \longrightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \qquad G(x) = \arctan x.$$

Moreover, $G'(x) = \frac{1}{1+x^2}$ is continuous on $\mathbb{R}$ so $G = F^{-1}$ is C^1. Thus, F is a diffeomorphism from $(-\frac{\pi}{2}, \frac{\pi}{2})$ to $\mathbb{R}$.

A map usually fails to be a diffeomorphism because it is not bijective.

Example 5.1.7 The function $F : \mathbb{R} \to \mathbb{R}$ given by $F(x) = x^2$ is not a diffeomorphism because it is not bijective. For instance, $F(-1) = 1 = F(1)$ so F is not injective. On the other hand, the function $G : (0,2) \to (0,4)$ given by $G(x) = x^2$ is bijective and C^1 with inverse $H : (0,4) \to (0,2)$ given by $H(x) = x^{1/2}$. Note $H'(x) = x^{-1/2}$, so the inverse $H = G^{-1}$ is C^1 on its domain. Thus, G is a diffeomorphism.

You have already seen that a C^1 bijective map can fail to be a diffeomorphism.

Example 5.1.8 Revisit Example 5.1.3. The function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = x^3$ is C^1 and bijective with inverse given by $G : \mathbb{R} \to \mathbb{R}$ with $G(x) = x^{1/3}$. However, F is *not* a diffeomorphism, because its inverse $G = F^{-1}$ is not differentiable at 0 and hence not C^1.

Now, finding inverses of nonlinear maps $\mathbb{R}^n \to \mathbb{R}^n$ has important geometric consequences and applications to coordinate systems (Section 1.4). If you want to translate between two coordinate systems, say between polar and rectangular in $\mathbb{R}^2$, then you must find the inverse of the coordinate transformation.

Example 5.1.9 Consider the polar coordinate transformation

$$F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad F(r,\theta) = (r\cos\theta, r\sin\theta).$$

Note F is not a diffeomorphism because F is not bijective. For example, $F(1,0) = (1,0) = F(1,2\pi)$, so F is not injective. If you instead define the open subsets

$$U = (0,\infty) \times (-\tfrac{\pi}{2}, \tfrac{\pi}{2}), \qquad V = (0,\infty) \times \mathbb{R}$$

[1] The suffix *-morphism* is used in mathematics to refer to maps that preserve structure, whether it is algebraic, topological, geometric, or whatever may be your interest. The prefix *diffeo-* is used to refer to objects that possess some smoothness, such as differentiable, C^1, infinitely differentiable, or whatever the context requires. The combined phrase *diffeomorphism* roughly means "a smooth map that preserves smooth structure".

of $\mathbb{R}^2$, then by an argument similar to Lemma 1.4.10 you can show that F maps U bijectively to V. In other words, the map given by

$$G : U \longrightarrow V, \qquad G(r,\theta) = (r\cos\theta, r\sin\theta)$$

is bijective. Moreover, you can verify that $G^{-1} : V \to U$ is given by

$$G^{-1}(x,y) = \left(\sqrt{x^2+y^2}, \arctan(y/x)\right)$$

Since G and G^{-1} are both C^1 on their domains, it follows that G is a diffeomorphism.

Linear maps are one of the few families for which you can classify all diffeomorphisms.

Example 5.1.10 Let A be a $n \times n$ matrix. Define $F : \mathbb{R}^n \to \mathbb{R}^n$ by $F(x) = Ax$. You can verify that F is a diffeomorphism if and only if A is invertible. The proof is left as a short exercise.

5.1.2 Properties of diffeomorphisms

Global diffeomorphisms possess some basic natural properties.

Lemma 5.1.11 (Symmetry) Let U and V be open subsets of $\mathbb{R}^n$. Assume the map $F : U \to V$ is bijective. The map F is a diffeomorphism if and only if its inverse F^{-1} is a diffeomorphism.

Proof. This is left as an exercise. Note the inverse of the inverse is the function itself. ■

Lemma 5.1.12 (Transitivity) Let U, V, and W be open subsets of $\mathbb{R}^n$. If $F : U \to V$ and $G : V \to W$ are diffeomorphisms, then $G \circ F : U \to W$ is a diffeomorphism.

Proof. This is left as an exercise. You will need to use the chain rule twice. ■

Diffeomorphisms are wonderful because they preserve so much structure. For instance, since a diffeomorphism *and its inverse* are both continuous, they preserve almost every topological property of sets that you have studied so far.

Lemma 5.1.13 (Diffeomorphisms preserve topology) Let U and V be open subsets of $\mathbb{R}^n$. Let $F : U \to V$ be a diffeomorphism. For every subset S of U, all of the following hold:

(a) S is open if and only if $F(S)$ is open.
(b) S is closed if and only if $F(S)$ is closed.
(c) S is compact if and only if $F(S)$ is compact.
(d) S is path-connected if and only if $F(S)$ is path-connected.

Proof. For (c), if S is compact then since F is C^1 (and hence continuous) it follows that $F(S)$ is compact. Conversely, if $F(S)$ is compact then since F^{-1} exists and is C^1 (and hence continuous), the set $F^{-1}(F(S)) = S$ is compact. Items (a), (b), and (d) are left as exercises. For (a), follow the ideas in the proof of Theorem 2.6.27 but be careful that the domain is not all of $\mathbb{R}^n$. For (b), use sequences. For (d), use Theorem 2.7.8 twice. ■

Moreover, diffeomorphisms also preserve smoothness properties such as tangency.

Lemma 5.1.14 (Diffeomorphisms preserve tangency) Let U and V be open subsets of $\mathbb{R}^n$. Let $F : U \to V$ be a diffeomorphism. Let $S \subseteq U$ be a set. Fix $p \in S$. A vector $v \in \mathbb{R}^n$ is a tangent vector of S at p if and only if $dF_p(v) \in \mathbb{R}^n$ is a tangent vector of $F(S)$ at $F(p)$.

Proof. This is left as an exercise. Proceed by definition, and use symmetry. ■

This preservation of structure is especially helpful if you view your diffeomorphism as a coordinate transformation. Lemmas 5.1.13 and 5.1.14 imply that when you transform a set from one coordinate system to another via a diffeomorphism, then your set retains all of its original topological and smoothness properties! In other words, switching coordinate systems is mostly a matter of relabelling and it does not change the underlying characteristics of a set.

5.1.3 Local diffeomorphisms

Finding global inverses for nonlinear maps $F:\mathbb{R}^n\to\mathbb{R}^n$ is not always tractable. In other words, there may be no global solution to (5.1.1) when $m=n$ because, as illustrated by previous examples, F may not actually be bijective. There is a compromise that you can make by restricting the domain.

Example 5.1.15 Define $F:\mathbb{R}\to\mathbb{R}$ given by $F(x)=x^2$. Example 5.1.7 showed that F is not a diffeomorphism since it is not bijective. However, the same example also noted that if you restrict F to the domain $U=(0,2)$ so $F(U)=(0,4)$, then the map $G=F|_U:U\to F(U)$ is a diffeomorphism.

Instead of searching for a *global* solution, you should therefore search for *local* solutions.

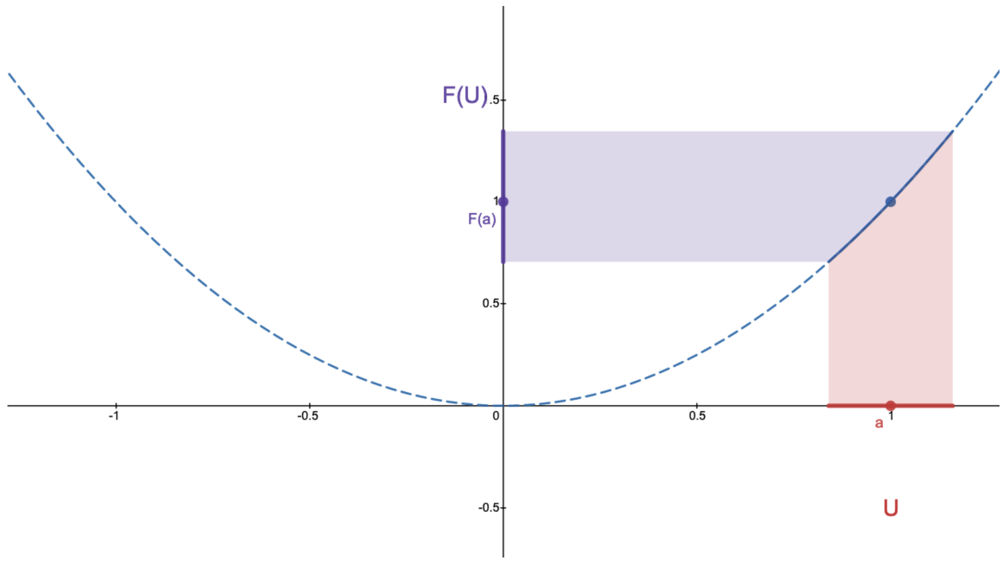
Definition 5.1.16 Let A and B be open subsets of $\mathbb{R}^n$. Fix a point $a\in A$. A function $F:A\to B$ is a **local diffeomorphism at** a if there exists an open subset $U\subseteq A$ containing a such that $F(U)$ is open and the restriction

$$F|_U:U\to F(U)$$

is a diffeomorphism. The inverse function $G=F|_U^{-1}:F(U)\to U$ is **a local inverse of** F **at** a.

A basic example for a map $\mathbb{R}\to\mathbb{R}$ illustrates the key technical points.

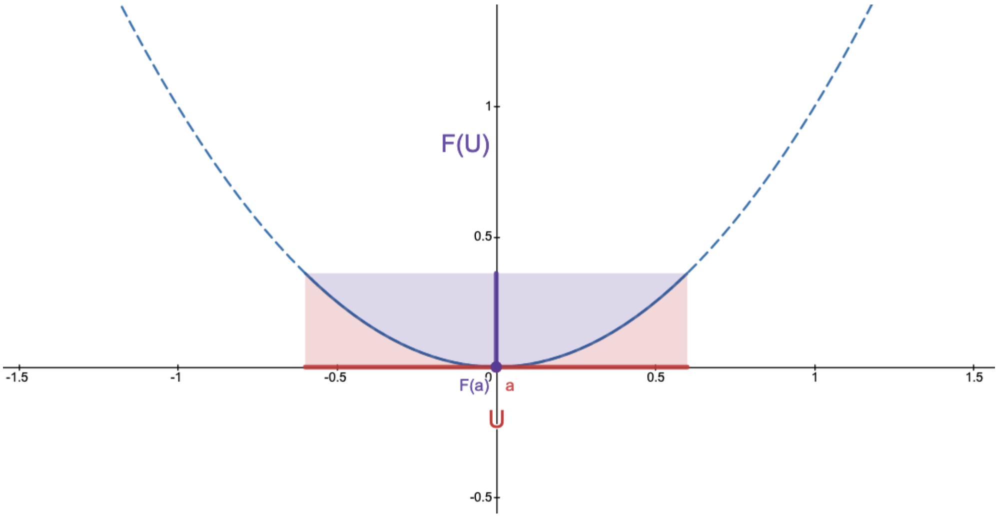
Example 5.1.17 Consider again $F:\mathbb{R}\to\mathbb{R}$ given by $F(x)=x^2$. Although F is not bijective, the Desmos plot below illustrates that F is local diffeomorphism at $a=1$.



By restricting F to the open set $U=(0.8,1.2)$, the function $G=F|_U:U\to F(U)$ is now bijective with inverse $G^{-1}(x)=\sqrt{x}$. Since $G'(x)=2x$ is continuous on U and $(G^{-1})'(x)=$

$\frac{1}{2}x^{-1/2}$ is continuous on $F(U) = (0.64, 1.44)$, the maps G and G^{-1} are both C^1. Hence, F is a local diffeomorphism at $a = 1$.

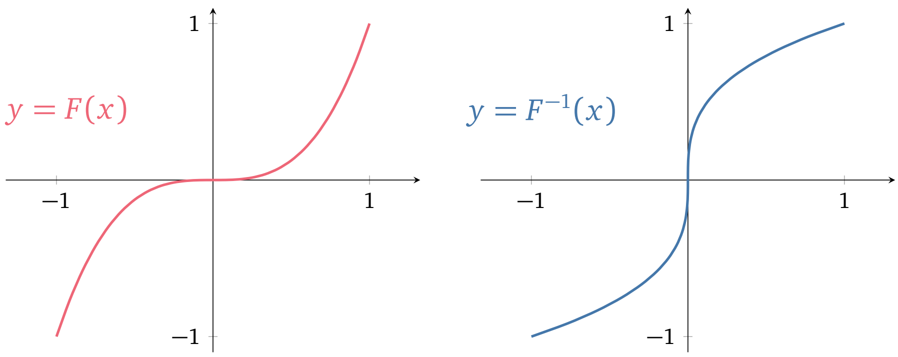
On the other hand, F is not a local diffeomorphism at $a = 0$, because its restriction on any open set $U \subseteq \mathbb{R}$ containing 0 always fails to be bijective. Here is a proof. Let $U \subseteq \mathbb{R}$ be an open set containing 0 and define $G = F|_U : U \to F(U)$. There exists $\varepsilon > 0$ such that $(-2\varepsilon, 2\varepsilon) \subseteq U$. Thus, $G(-\varepsilon) = \varepsilon^2 = G(\varepsilon)$, so G is not injective and hence not bijective. Since U was arbitrary, this proves by definition that F is not a local diffeomorphism at $a = 0$.



You can informally confirm this observation with the Desmos plot above.

A function can fail to be a local diffeomorphism for reasons other than bijectivity.

Example 5.1.18 The function $F : \mathbb{R} \to \mathbb{R}$ given by $F(x) = x^3$ is a local diffeomorphism at each $x \in \mathbb{R}$ except for at $x = 0$. Notice that F is invertible with inverse $F^{-1}(x) = x^{1/3}$.



The vertical tangent at $x = 0$ implies that $F^{-1}(x)$ is not differentiable at $x = 0$, so $F^{-1}(x)$ is not C^1 on any open interval containing $x = 0$.

You can prove that F is a local diffeomorphism at every $a \in \mathbb{R}$ with $a \neq 0$. Here is the argument for $a = 1$. Take $U = (0, 2)$ so $1 \in U$ and $F(U) = (0, 8)$. Define the restriction $G = F|_U : U \to F(U)$ so $G : (0, 2) \to (0, 8)$ and $G(x) = x^3$. Notice G is C^1 on its domain as $G'(x) = 3x^2$ is continuous on $(0, 2)$. Since F is globally invertible with $F^{-1}(x) = x^{1/3}$, it follows that G is invertible with $G^{-1} : (0, 8) \to (0, 2)$ and $G^{-1}(x) = x^{1/3}$. Notice that G^{-1} is also C^1, since $(G^{-1})'(x) = \frac{1}{3}x^{-2/3}$ is continuous on $(0, 8)$. Hence, G is C^1 and invertible on $U = (0, 2)$ with C^1 inverse, so $G = F|_U$ is a diffeomorphism by definition. As $1 \in U$, it follows that F is a local diffeomorphism at $a = 1$.

Note that the choice of interval $U = (0,2)$ is not unique. There are infinitely many valid choices of U in this situation, each of which gives a different local inverse.

There is a simple relationship between global and local diffeomorphisms.

Lemma 5.1.19 Let A and B be open subsets of $\mathbb{R}^n$. If $F : A \to B$ is a global diffeomorphism, then F is a local diffeomorphism at every $a \in A$.

Proof. This follows quickly from Definition 5.1.16 by taking $U = A$. It is left as an exercise. ■

Somewhat surprisingly, the converse of Lemma 5.1.19 is false. You will uncover a counterexample in the next section. Now, it is much more difficult to construct inverses for maps $\mathbb{R}^n \to \mathbb{R}^n$ in higher dimensions. You have already seen this with the polar coordinate transformation in Section 1.4 which you will continue to analyze here.

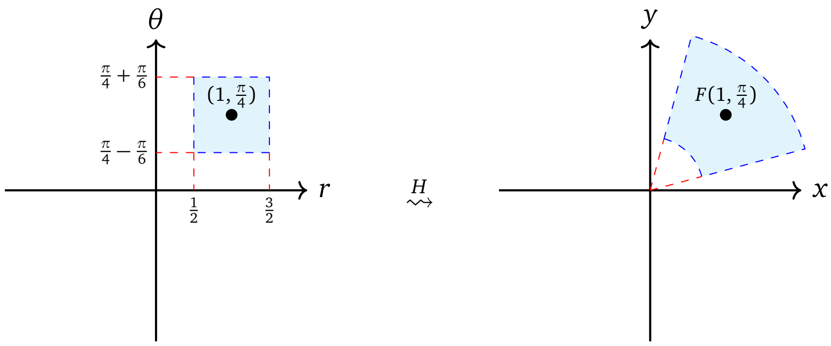
Example 5.1.20 Define the polar coordinate transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ by $F(r,\theta) = (r\cos\theta, r\sin\theta)$. As you have seen in Example 5.1.9, F is not bijective and therefore is not a diffeomorphism. Example 5.1.9 also explained that the restriction $G = F|_U : U \to F(U)$ with

$$U = (0,\infty) \times (-\tfrac{\pi}{2}, \tfrac{\pi}{2}), \qquad F(U) = V = (0,\infty) \times \mathbb{R}$$

is a diffeomorphism. This implies, for example, that F is a local diffeomorphism at $(1, \frac{\pi}{4})$ since $(1, \frac{\pi}{4}) \in U$. Again, to prove F is a local diffeomorphism at $(1, \frac{\pi}{4})$, the choice of restriction is not unique. You could instead choose the open set

$$W = (\tfrac{1}{2}, \tfrac{3}{2}) \times (\tfrac{\pi}{4} - \tfrac{\pi}{6}, \tfrac{\pi}{4} + \tfrac{\pi}{6})$$

and prove that $H = F|_W : W \to F(W)$ is a diffeomorphism. The map H is illustrated below.



Since $(1, \frac{\pi}{4}) \in W$, this also shows that F is a local diffeomorphism at $(1, \frac{\pi}{4})$.

In fact, F is a local diffeomorphism at every point $(r,\theta) \in \mathbb{R}^2$ with $r \neq 0$, but F is not a local diffeomorphism at any point $(0,\theta)$ for $\theta \in \mathbb{R}$. You can play with the Math3D demo to informally confirm these observations.

Consider again the nonlinear system (5.1.1) when $m = n$. Finding a global solution to $y = F(x)$ for all $x \in \mathbb{R}^n$ is often impossible; most maps $F : \mathbb{R}^n \to \mathbb{R}^n$ are not diffeomorphisms. However, you can *locally* solve $y = F(x)$ for the variable x, provided x is near a and F is a local diffeomorphism at $a \in \mathbb{R}^n$. This reduces the problem to a new challenge.

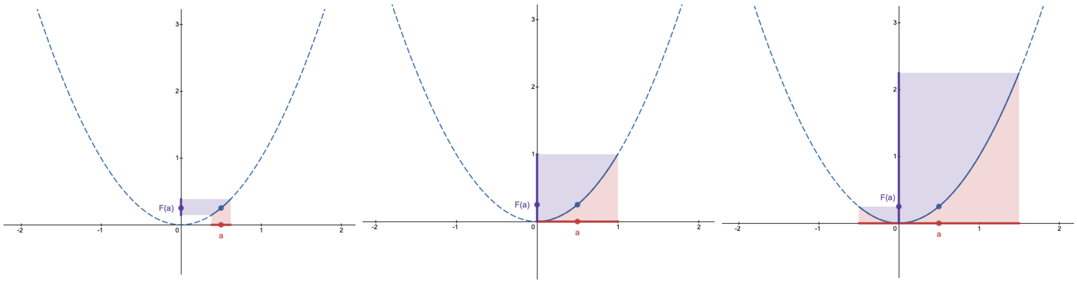
How do you verify whether a map is a local diffeomorphism at a point?

That will be the focus of the next section.

Exercises for Section 5.1

Concepts and definitions

5.1.1 Define $F : \mathbb{R} \to \mathbb{R}$ by $F(x) = x^2$. Consider three different open sets $U, V, W \subseteq \mathbb{R}$.



$U = (0.4, 0.6)$ $\qquad$ $V = (0, 1)$ $\qquad$ $W = (-0.5, 1.5)$

Each set contains the point $a = 0.5$. Use the above Desmos figures for each part below.

(a) Consider the maps

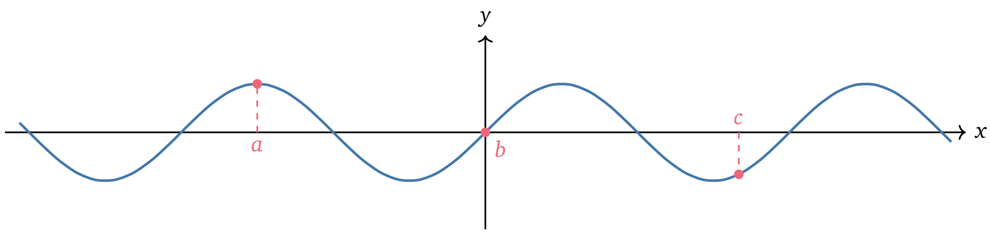
$$F|_U : U \to F(U) \qquad F|_V : V \to F(V) \qquad F|_W : W \to F(W)$$

Which of these maps is a diffeomorphism?

(b) Is F a local diffeomorphism at $a = 0.5$?

(c) State the largest open set $S \subseteq \mathbb{R}$ such that $a \in S$ and $F|_S : S \to F(S)$ is a diffeomorphism.

5.1.2 Consider the graph of $F(x) = \sin x$ and three real numbers $a, b, c \in \mathbb{R}$.

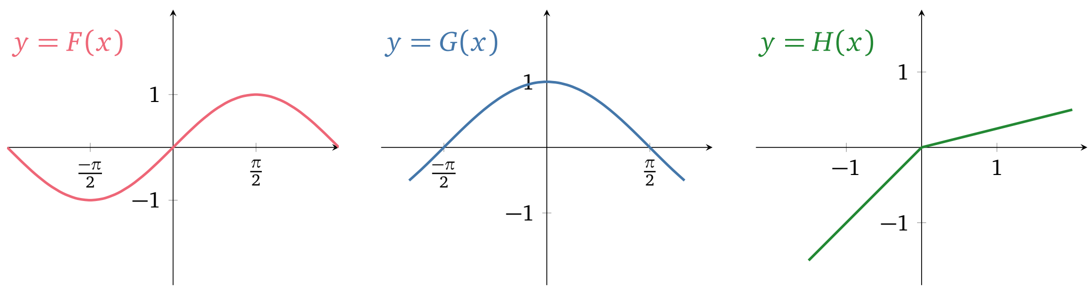


(a) Is F a local diffeomorphism at a? If so, define two different local inverses at a. If not, explain.

(b) Is F a local diffeomorphism at b? If so, define two different local inverses at b. If not, explain.

(c) Is F a local diffeomorphism at c? If so, define two different local inverses at c. If not, explain.

5.1.3 Consider the three functions F, G, H.



(a) Which of F, G, H appear to be a diffeomorphism?

(b) Which of F, G, H appear to be a local diffeomorphism at 0?

(c) Which of F, G, H appear to be a local diffeomorphism at 1?

5.1.4 Is the following statement true or false?

If $F : \mathbb{R}^n \to \mathbb{R}^n$ is C^1 and bijective, then F is a diffeomorphism.

If true, briefly explain. If false, give a counterexample.

5.1.5 Diffeomorphisms are fantastic since they preserve the properties of many sets. Let

$$F : \mathbb{R} \to \mathbb{R}, \qquad G : \mathbb{R}^2 \to \mathbb{R}^2, \qquad H : \mathbb{R}^3 \to \mathbb{R}^3$$

be diffeomorphisms. Which statements below are necessarily false? Explain why.

(a) $F([0,1]) = (137, 237)$

(b) $F((137, 237)) = [0,1]$.

(c) $G(\{(x,y) \in \mathbb{R}^2 : y \neq 0\}) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$

(d) $H([0,1]^3) = \{(x,y,z) \in \mathbb{R}^3 : x \geq 0, y \geq 0, z \geq 0\}$

Proofs

5.1.6 Let A be a $n \times n$ matrix. Define $F : \mathbb{R}^n \to \mathbb{R}^n$ by $F(x) = Ax$. Prove that F is a diffeomorphism if and only if A is invertible.

5.1.7 Let A and B be open subsets of $\mathbb{R}^n$. Prove that if $F : A \to B$ is a global diffeomorphism, then F is a local diffeomorphism at every $a \in A$.

5.1.8 Let $F : \mathbb{R} \to \mathbb{R}$ be $F(x) = \sin x$. Fardin and Amy try to prove F is a local diffeomorphism at π.

(a) Fardin writes two short lines.

1. *Note* $\sin x$ *is* C^1 *and its inverse* $\arcsin x$ *is* C^1.
2. *Thus,* F *is a local diffeomorphism at* π.

There is a false claim in one of the lines. Identify it and explain what is the false claim.

(b) Amy does better, but also makes a serious mistake in one line.

1. *Note* $F(x) = \sin x$ *is* C^1.
2. *Also,* $F(x) = \sin x$ *bijectively maps* $U = (\frac{\pi}{2}, \frac{3\pi}{2})$ *to* $V = (-1, 1)$.
3. *The inverse map is* $G(x) = \arcsin x$.
4. *Note* G *is* C^1, *since* $G'(x) = \frac{1}{\sqrt{1-x^2}}$ *is continuous for* $-1 < x < 1$.
5. *Thus,* F *is a local diffeomorphism at* π.

Identify the flawed line and briefly explain what is wrong.

(c) Improve upon the arguments of Fardin and Amy. Prove that F is a local diffeomorphism at π.

5.1.9 Prove that $F : \mathbb{R} \to \mathbb{R}$ given by $F(x) = \sin x$ is not a local diffeomorphism at $\frac{\pi}{2}$.

5.1.10 Recall two standard properties of diffeomorphisms.

> **Lemma A.** *Let U and V be open subsets of $\mathbb{R}$. Assume the map $F : U \to V$ is bijective. The map F is a diffeomorphism if and only if its inverse F^{-1} is a diffeomorphism.*

> **Lemma B.** *Let $U, V,$ and W be open subsets of $\mathbb{R}$. If $F : U \to V$ and $G : V \to W$ are diffeomorphisms, then $G \circ F : U \to W$ is a diffeomorphism.*

(a) Here is a bad proof of Lemma A.

1. *Assume F is a diffeomorphism. By definition, F is C^1, bijective, and its inverse F^{-1} is C^1.*
2. *Note $F : U \to V$ is bijective implies that $F^{-1} : V \to U$ is bijective by definition.*
3. *Hence, F^{-1} is C^1 and bijective.*
4. *Thus, F^{-1} is a diffeomorphism.*
5. *The converse follows by a similar argument.*

This argument is missing a key justification. Fix it by adding one line.

(b) Here is a bad proof of Lemma B.

1. *Assume $F : U \to V$ and $G : V \to W$ are diffeomorphisms.*
2. *By definition, F and G are C^1 and bijective with C^1 inverses.*
3. *A composition of bijective maps is bijective. Since F and G are bijective, then so is $G \circ F$.*
4. *Since F and G are C^1, it follows that $G \circ F$ is C^1 by a consequence of the chain rule.*
5. *Since F^{-1} and G^{-1} are C^1, it follows that $G^{-1} \circ F^{-1}$ is C^1 by a consequence of the chain rule.*
6. *Thus, $G \circ F$ is a diffeomorphism.*

One line has a fatal flaw. Identify the flaw and fix it.

5.1.11 Let U and V be open subsets of $\mathbb{R}^n$. Let $F : U \to V$ be a global diffeomorphism. Let S be a subset of U.

(a) Prove that S is path-connected if and only if $F(S)$ is path-connected.

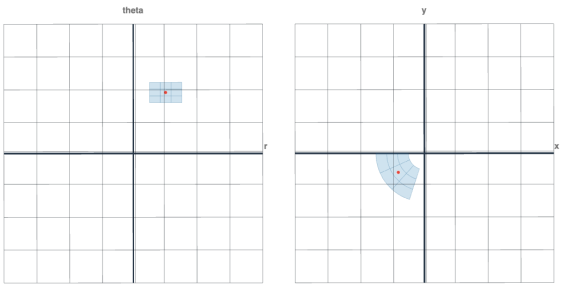
(b) Prove that S is closed if and only if $F(S)$ is closed. *Hint:* Use sequences.

Applications and beyond

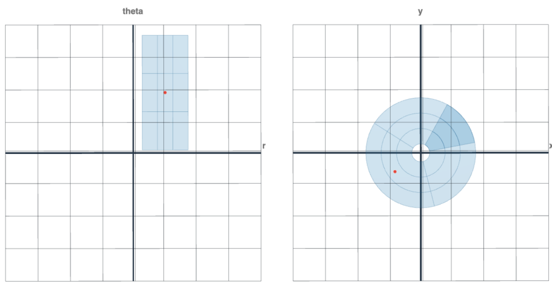
5.1.12 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinates map, so $F(r, \theta) = (r\cos\theta, r\sin\theta)$.

(a) Fix a point $p = (r_0, \theta_0) \in \mathbb{R}^2$. Consider four different open sets $A, B, C, D \subseteq \mathbb{R}^2$ containing p.

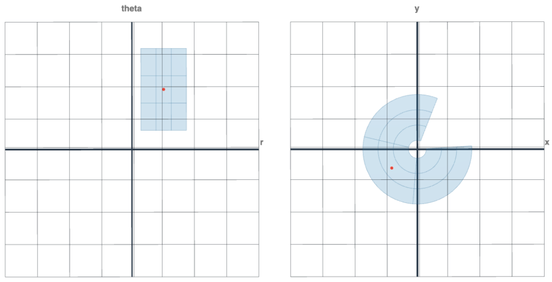
i) $F|_A : A \to F(A)$



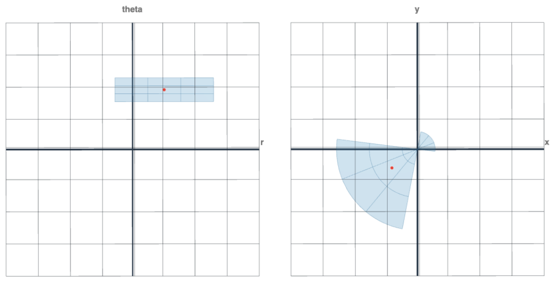
ii) $F|_B : B \to F(B)$



iii) $F|_C : C \to F(C)$



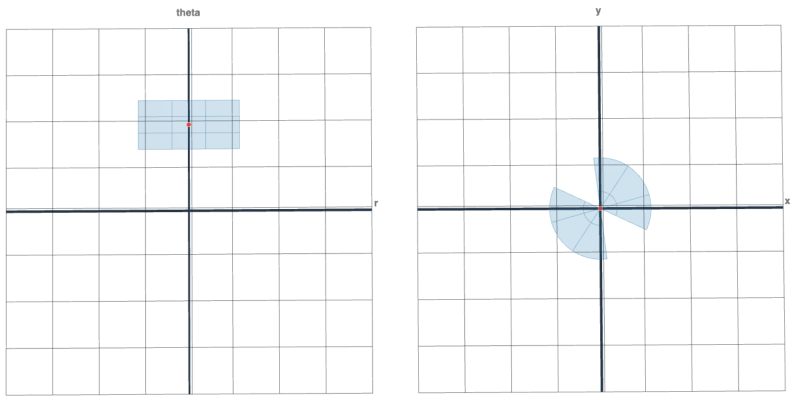
iv) $F|_D : D \to F(D)$



Which maps $F|_A, F|_B, F|_C, F|_D$ appear to be a diffeomorphism? Briefly explain each case. (The transformation of these sets under F can be illustrated with this Math3D demo.)

(b) Is F a local diffeomorphism at p? Identify which of the maps above support your assertion.

(c) Fix a point $q = (0, \theta_0) \in \mathbb{R}^2$. The polar coordinates map F is not a local diffeomorphism at q. This can be informally observed with the illustration below of $F|_U : U \to F(U)$ for some open set U containing q.



How does this illustration show that $F|_U$ is not a diffeomorphism? Explain in one sentence.

(d) Prove that F is not a local diffeomorphism at $(0, \theta_0) \in \mathbb{R}^2$.

5.1.13 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinates map, so $F(r, \theta) = (r\cos\theta, r\sin\theta)$. How can you prove by definition that F is a local diffeomorphism at every $(r, \theta) \in \mathbb{R}^2$ with $r \neq 0$? This exercise will illustrate the key ingredients for a full proof. You will not verify every detail, but you will make enough progress to hopefully convince yourself. Set

$$S = (0, \infty) \times (-\tfrac{\pi}{2}, \tfrac{\pi}{2}), \qquad \text{so} \qquad F(S) = (0, \infty) \times \mathbb{R}.$$

The restriction $F|_S : S \to F(S)$ is a diffeomorphism with inverse $(F|_S)^{-1} : F(S) \to S$ given by

$$(F|_S)^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, \arctan(y/x)\right).$$

You will assume many maps are diffeomorphisms without proof.

(a) The restriction $F|_S$ proves that F is a local diffeomorphism at which points of $\mathbb{R}^2$?

(b) Set $U = (-\infty, 0) \times (-\frac{\pi}{2}, \frac{\pi}{2})$. Sketch $F|_U : U \to F(U)$ as a transformation and describe the set $F(U)$. Then guess an explicit formula for the inverse map $(F|_U)^{-1}$.

(c) If you prove that $F|_S$ and $F|_U$ are diffeomorphisms, then at which points of $\mathbb{R}^2$ can you conclude that F is a local diffeomorphism?

(d) Set $V = (0,\infty) \times (\frac{\pi}{2}, \frac{3\pi}{2})$. Sketch $F|_V : V \to F(V)$ as a transformation and describe the set $F(V)$. Guess an explicit formula for the inverse map $(F|_V)^{-1}$.

(e) Set $W = (-\infty, 0) \times (0, \pi)$. Sketch $F|_W : W \to F(W)$ as a transformation and describe the set $F(W)$. Guess an explicit formula for the inverse map $(F|_W)^{-1}$.

(f) If you prove that $F|_S, F|_U, F|_V, F|_W$ are diffeomorphisms, then at which points of $\mathbb{R}^2$ can you conclude that F is a local diffeomorphism?

5.1.14 Define the cylindrical coordinate map $F : \mathbb{R}^3 \to \mathbb{R}^3$ by $F(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$. Fix $\theta_0, z_0 \in \mathbb{R}$. Prove that F is not a local diffeomorphism at $(0, \theta_0, z_0)$.

5.1.15 Define the spherical coordinate map $F : \mathbb{R}^3 \to \mathbb{R}^3$ by $F(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$.

(a) Fix $\theta_0, \phi_0 \in \mathbb{R}$. Prove that F is not a local diffeomorphism at $(0, \theta_0, \phi_0)$.

(b) Fix $\rho_0, \theta_0 \in \mathbb{R}$. Prove that F is not a local diffeomorphism at (ρ_0, θ_0, π).

5.2. Inverse function theorem

Given a nonlinear C^1 map $F : \mathbb{R}^n \to \mathbb{R}^n$, you have the system of equations

$$\begin{array}{rcl} y_1 & = & F_1(x_1, \dots, x_n) \\ & \vdots & \\ y_n & = & F_n(x_1, \dots, x_n) \end{array} \tag{5.2.1}$$

which, as you have seen, may be impossible to solve globally. Your investigations thus far show that if F is a *local* diffeomorphism at $a \in \mathbb{R}^n$, then you can *locally* solve this system $y = F(x)$ for x near a. That is, you can express $x = G(y)$ where G is a *local* C^1 inverse of F at a. This outcome would be a fantastic achievement since (5.2.1) can be an incredibly nasty system of equations. Now, how can you verify that the C^1 map F is a local diffeomorphism at a point?

Your first hint of an answer will come from differentiating diffeomorphisms, which will lead to a complete resolution and a grand result: the inverse function theorem! This amazing theorem will allow you to easily check the existence of local solutions to (5.2.1). There are, however, some drawbacks that you will uncover.

5.2.1 Derivatives of diffeomorphisms

You begin by learning how to calculate the derivatives of the inverse of a diffeomorphism.

Theorem 5.2.1 Let U and V be open subsets of $\mathbb{R}^n$. Assume the function $F : U \to V$ is a diffeomorphism. For every $x \in U$, the Jacobian $DF(x)$ is an invertible $n \times n$ matrix and the Jacobian of the inverse function $G = F^{-1} : V \to U$ satisfies

$$DG(y) = [DF(x)]^{-1} \qquad \text{for every } x \in U \text{ and } y = F(x).$$

This theorem is an immediate consequence of the chain rule.

Proof. Since F and G are inverses, they satisfy the equation

$$\forall x \in U, \quad G \circ F(x) = x = \mathrm{id}(x),$$

where $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map. This implies their Jacobians are equal so that

$$\forall x \in U, \quad D(G \circ F)(x) = D\,\mathrm{id}(x). \tag{5.2.2}$$

The Jacobian of id is the $n \times n$ identity matrix I_n, that is, $D\,\mathrm{id}(x) = I_n$ for all $x \in \mathbb{R}^n$. As F is a diffeomorphism, both F and G are C^1 and hence differentiable, so the chain rule (Theorem 4.1.1) implies that $D(G \circ F)(x) = DG(F(x))DF(x)$ for $x \in U$. Combining these observations with (5.2.2) and writing $y = F(x)$ gives that

$$\forall x \in U, \quad DG(y)DF(x) = I_n.$$

Hence, for $x \in U$, the $n \times n$ matrix $DF(x)$ is invertible and its inverse is $DG(y)$. ∎

Given a diffeomorphism, this theorem is relatively straightforward to apply.

Example 5.2.2 Recall the function $F : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ given by $F(x) = \tan x$ is a diffeomorphism from Example 5.1.6. Let $G : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ be its inverse, so $G = F^{-1}$. By Theorem 5.2.1, for every $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ and $y = F(x) = \tan x$,

$$G'(y) = [F'(x)]^{-1} = \frac{1}{F'(x)} = \frac{1}{\sec^2 x} = \frac{1}{1 + \tan^2 x} = \frac{1}{1 + y^2}$$

by the trigonometric identity $\sec^2 x = 1+\tan^2 x$. Notice this calculation did not actually require the explicit formula for the inverse! That is really useful when it is difficult to compute the inverse (as it often is). In this case, you happen to know that $G(y) = \arctan y$ so you can safely double check that $G'(y) = (1+y^2)^{-1}$.

Example 5.2.3 The polar coordinates map $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $F(r,\theta) = (r\cos\theta, r\sin\theta)$ is not a diffeomorphism, but its restriction $H = F|_U : U \to F(U)$ with

$$U = (0,\infty)\times(-\tfrac{\pi}{2},\tfrac{\pi}{2}) \quad \text{and} \quad F(U) = (0,\infty)\times\mathbb{R}$$

is a diffeomorphism by Example 5.1.9. Let $G = H^{-1} : F(U) \to U$ be the C^1 inverse of H. By Theorem 5.2.1, it follows that for every $(r,\theta) \in U$ and $(x,y) = H(r,\theta) = (r\cos\theta, r\sin\theta)$,

$$DG(x,y) = [DH(r,\theta)]^{-1} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}^{-1} = \frac{1}{r}\begin{bmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

by the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$. This can be expressed in terms of x and y. Since $(r,\theta) \in U$, it follows that $r = \sqrt{x^2+y^2}$, $\cos\theta = \frac{x}{\sqrt{x^2+y^2}}$, and $\sin\theta = \frac{y}{\sqrt{x^2+y^2}}$ so the above may also be written as

$$DG(x,y) = \begin{bmatrix} x/\sqrt{x^2+y^2} & y/\sqrt{x^2+y^2} \\ -y/(x^2+y^2) & x/(x^2+y^2) \end{bmatrix}.$$

Again, notice you did not require the explicit formula for G from Example 5.1.9 to perform this calculation. You can double check the result with this alternate approach, if you wish.

At first glance, Theorem 5.2.1 may seem unhelpful for your overall goal, since it requires a diffeomorphism as an assumption and that is hard to verify. Its key asset is the creation of a quick-and-easy necessary condition for a local diffeomorphism.

Corollary 5.2.4 Let A and B be open subsets of $\mathbb{R}^n$. Fix $a \in A$. Let $F : A \to B$ be a C^1 function. If F is a local diffeomorphism at a, then the Jacobian $DF(a)$ is an invertible $n \times n$ matrix.

Proof. Since F is a local diffeomorphism at a, there exists an open set $U \subseteq A$ containing a such that $H = F|_U : U \to F(U)$ is a diffeomorphism. By Theorem 5.2.1, it follows that $DH(x) = DF(x)$ is invertible for $x \in U$. Since $a \in U$, this implies that $DF(a)$ is invertible. ■

This corollary can prove that a map is *not* a local diffeomorphism.

Example 5.2.5 Let $F : \mathbb{R} \to \mathbb{R}$ be given by $F(x) = x^3$. The Jacobian of F is $DF(x) = \begin{bmatrix} 3x^2 \end{bmatrix}$, so $DF(0)$ is not invertible. The contrapositive of Corollary 5.2.4 implies that F is not a local diffeomorphism at $0 \in \mathbb{R}$.

Example 5.2.6 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinate map so $F(r,\theta) = (r\cos\theta, r\sin\theta)$. For $(r,\theta) \in \mathbb{R}^2$,

$$DF(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix},$$

so its determinant is $\det DF(r,\theta) = r\cos^2\theta + r\sin^2\theta = r$. Thus, $DF(r,\theta)$ is not invertible whenever $r = 0$. By Corollary 5.2.4, F is not a local diffeomorphism at $(0,\theta)$ for $\theta \in \mathbb{R}$.

Example 5.2.7 Let $F : \mathbb{R}^4 \to \mathbb{R}^4$ be given by

$$F(x_1, x_2, x_3, x_4) = (2x_1 + x_4^2, x_2 x_3 - 2x_4, x_1^2 - x_3, x_2)$$

It is not at all apparent where this function may or may not be a local diffeomorphism. By standard calculations, its Jacobian is

$$DF(x_1, x_2, x_3, x_4) = \begin{bmatrix} 2 & 0 & 0 & 2x_4 \\ 0 & x_3 & x_2 & -2 \\ 2x_1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

so $\det DF(x_1, x_2, x_3, x_4) = -4 - 4x_1 x_2 x_4$. Hence, Corollary 5.2.4 implies that F is not a local diffeomorphism whenever $x_1 x_2 x_4 = -1$.

Corollary 5.2.4 continues the philosophy of differentiation that nonlinear maps are approximated by linear maps. Informally, it says that

> *If a nonlinear C^1 map is locally invertible, then its linear approximation is globally invertible.*

This gives a necessary condition for a map to be a local diffeomorphism.

5.2.2 Inverse function theorem

Miraculously, this necessary condition is also *sufficient*! That is the remarkable surprise of the inverse function theorem.

Theorem 5.2.8 (Inverse function theorem) Let A and B be open subsets of $\mathbb{R}^n$. Fix $a \in A$. Let $F : A \to B$ be a C^1 function. If the Jacobian $DF(a)$ is an invertible $n \times n$ matrix, then the map F is a local diffeomorphism at a.

Remark 5.2.9 In particular, by Theorem 5.2.1 and Definition 5.1.16, there exists an open set $U \subseteq A$ containing a such that the restriction $F|_U : U \to F(U)$ is a diffeomorphism and the inverse map $G = F|_U^{-1}$ satisfies $DG(y) = \big[DF(x)\big]^{-1}$ for all $x \in U$ and $y = F(x)$.

Note, however, the inverse function theorem does *not* explicitly give you the open set U, nor does it provide an explicit formula for a local inverse G. You can only conclude that they exist and nothing more.

Proof. The proof is far beyond the scope of this text. Solving nonlinear systems of equations requires much deeper analytic machinery, such as contraction mappings and the Banach fixed-point theorem[2]. See, for example, [6, Chapter III] or [18, Chapter 6]. ∎

Informally speaking, the inverse function theorem and Corollary 5.2.4 can be rephrased as:

> *A nonlinear C^1 map is locally invertible if and only if its linear approximation is globally invertible.*

You can heuristically explain its truth as follows. Suppose your goal is to locally solve a *nonlinear* system of equations

$$y = F(x) \quad \text{when } x \approx a$$

[2]See MAT337 Introduction to Real Analysis or MAT357 Analysis III for details on this beautiful theorem.

for x in terms of y. Linearly approximating F near a, you can replace this nonlinear system with the *linear* system of equations

$$y \approx F(a) + DF(a)(x-a) \quad \text{when } x \approx a$$

and attempt to solve for x in terms of y. This linear system has a unique solution *if and only if* the $n \times n$ Jacobian matrix $DF(a)$ is invertible! If so, then

$$x \approx a + DF(a)^{-1}(y - F(a)) \quad \text{when } x \approx a.$$

Notice $x \approx a$ is equivalent to $y \approx F(a)$. The righthand side should presumably linearly approximate a local inverse of F at a. In other words, you must pray that there exists a local inverse G of F at a such that

$$x = G(y) \qquad \text{when } y \approx F(a)$$

and the linear approximation of $G(y)$ near $y = F(a)$ is precisely $a + DF(a)^{-1}(y - F(a))$. This last idea is very wishful thinking that requires some deep machinery to justify. Nonetheless, this heuristic explanation is hopefully convincing enough for you.

Now, equipped with this beautiful theorem, you can accomplish the original goal of this section. Namely, you can easily identify at which points a map is a local diffeomorphism and therefore find local solutions to the nonlinear system (5.2.1).

Example 5.2.10 Let $F : \mathbb{R} \to \mathbb{R}$ be given by $F(x) = x^3$. The Jacobian of F is $DF(x) = \begin{bmatrix} 3x^2 \end{bmatrix}$ so $DF(x)$ is invertible for $x \neq 0$. The inverse function theorem therefore implies that if $x \in \mathbb{R} \setminus \{0\}$ then F is a local diffeomorphism at x. Combined with the application of Corollary 5.2.4 in Example 5.2.5, you can finally conclude that F is a local diffeomorphism at x if and only if $x \neq 0$. This confirms precisely what you manually verified in Example 5.1.3, except now you can enjoy the luxury of the inverse function theorem.

Example 5.2.11 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinate map so $F(r, \theta) = (r\cos\theta, r\sin\theta)$. From Example 5.2.6, the determinant of its Jacobian is $\det DF(r,\theta) = r$, so $DF(r,\theta)$ is invertible if and only if $r \neq 0$. This implies by the inverse function theorem that if $r \neq 0$ then F is a local diffeomorphism at $(r,\theta) \in \mathbb{R}^2$. Combined with the application of Corollary 5.2.4 in Example 5.2.6 you can therefore conclude that F is a local diffeomorphism at $(r,\theta) \in \mathbb{R}$ if and only if $r \neq 0$. You should appreciate that this argument was nearly effortless. You never needed to compute a local inverse, yet you showed precisely when they exist!

You can now verify that the converse of Lemma 5.1.19 is false.

Example 5.2.12 Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by $F(x,y) = (e^x \cos y, e^x \sin y)$. For $(x,y) \in \mathbb{R}^2$, its Jacobian is given by

$$DF(x,y) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}$$

which has determinant $\det DF(x,y) = e^{2x}$. Since the exponential is never zero, the Jacobian $DF(x,y)$ is always invertible, so the inverse function theorem implies that F is a local diffeomorphism at every $(x,y) \in \mathbb{R}^2$. However, F is *not* a global diffeomorphism! For instance, $F(x,y) = F(x, y+2\pi)$ for any $(x,y) \in \mathbb{R}^2$ so F is not injective and hence not a diffeomorphism.

The next example shows that the inverse function theorem genuinely requires that F is C^1 to find a local inverse.

Example 5.2.13 Let $F : \mathbb{R} \to \mathbb{R}$ be given by

$$F(x) = \begin{cases} x + 2x^2 \sin\left(\frac{1}{x}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Notice F is differentiable everywhere. First, $F'(x) = 1 + 4x\sin(1/x) - 2\cos(1/x)$ for $x \neq 0$. Second, computing $F'(0)$ directly from the limit definition gives $F'(0) = 1$, so the Jacobian $DF(0) = F'(0) = 1$ is invertible. Notice that $\lim_{x\to 0} F'(x)$ does not exist yet $F'(0) = 1$. Hence, all of the assumptions of the inverse function theorem are satisfied, except F is not C^1 at 0.

Nonetheless, F is *not* invertible on any open interval I centered at 0. You can prove this by verifying that $F'(x) = 1 + 4x\sin(1/x) - 2\cos(1/x)$ oscillates between being positive and negative values infinitely often as you approach 0. This feature violates the conclusion of the following lemma: if $\phi : (a,b) \to \mathbb{R}$ is differentiable and injective, then either $\phi'(x) \geq 0$ for all $x \in (a,b)$ or $\phi'(x) \leq 0$ for all $x \in (a,b)$. Since F is differentiable everywhere, this implies F cannot be injective on any open interval containing 0.

The inverse function theorem is truly amazing, but it has drawbacks as noted in Remark 5.2.9. Keeping with the notation of the theorem, the key limitations include:

1. The theorem does not give an explicit formula for a local inverse G, so you cannot compute $G(y)$ for any specific y other than $y = F(a)$
2. The theorem does not give an explicit domain U for which $F|_U$ is a diffeomorphism, so you cannot explicitly determine how close x must be to a in order for a local inverse G to take x as input.

These drawbacks are more apparent in a more complicated example.

Example 5.2.14 Let $F : \mathbb{R}^4 \to \mathbb{R}^4$ be given by

$$F(x_1, x_2, x_3, x_4) = (x_1, x_1x_2, x_1x_2x_3, x_1x_2x_3 + x_4).$$

You can prove that F is a local diffeomorphism at $a = (1,1,1,1)$. Note F is C^1 and

$$DF(x) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ x_2 & x_1 & 0 & 0 \\ x_2x_3 & x_1x_3 & x_1x_2 & 0 \\ x_2x_3 & x_1x_3 & x_1x_2 & 1 \end{bmatrix} \quad \text{so} \quad DF(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The matrix $DF(a)$ has determinant 1 so it is invertible. By the inverse function theorem, F is a local diffeomorphism at a.

Thus, there exists a local inverse of F at a, that is, there exists some open set $U \subseteq \mathbb{R}^4$ containing a such that $F|_U : U \to F(U)$ is a diffeomorphism with C^1 inverse $G = (F|_U)^{-1}$. You have no formula for G but you know it exists and you also know that

$$DG(y) = [DF(x)]^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\dfrac{x_2}{x_1} & \dfrac{1}{x_1} & 0 & 0 \\ 0 & -\dfrac{x_3}{x_1x_2} & \dfrac{1}{x_1x_2} & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix} \text{for } x = (x_1, x_2, x_3, x_4) \in U \text{ and } y = F(x).$$

The inverse can be calculated by routine linear algebra calculations; this is often performed with computer algebra software but you should be able to do small examples by hand. In particular, at $y = F(a) = F(1,1,1,1) = (1,1,1,2)$,

$$DG(1,1,1,2) = [DF(1,1,1,1)]^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Although you do not have an explicit formula for G, you can *linearly approximate* G near $F(a)$ using this Jacobian. That is, for $y \in F(U)$,

$$\begin{aligned} G(y) &\approx G(1,1,1,2) + DG(1,1,1,2)(y-(1,1,1,2)) \\ &= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 - 1 \\ y_2 - 1 \\ y_3 - 1 \\ y_4 - 2 \end{bmatrix} = \begin{bmatrix} y_1 \\ -y_1 + y_2 + 1 \\ -y_2 + y_3 + 1 \\ -y_3 + y_4 \end{bmatrix}. \end{aligned}$$

Unfortunately, the inverse function theorem does not give you any information about U and you require that $y \in F(U)$ for this approximation to be valid. Thus, you have no information about how close y must be to $F(a) = (1,1,1,2)$ in order for this approximation to be reasonable. Despite these setbacks, it is still an awesome theorem.

Overall, the inverse function theorem constitutes yet another triumph for differential calculus, because information about a derivative (i.e. invertible Jacobian) implies a property about the map itself (i.e. local diffeomorphism). For instance, in single variable calculus, the mean value theorem shows that a positive derivative implies a function is increasing. This trend will continue throughout this chapter and the next. You have locally solved the nonlinear system of equations for inverse functions given by (5.2.1), and your next step is to expand these ideas to *any* nonlinear system of equations.

Exercises for Section 5.2

Concepts and definitions

5.2.1 Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be the linear map given by $L(x) = Ax$. In which cases is L invertible?

(a) $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 4 \\ -2 & 0 & -2 & 0 \\ 5 & 6 & 7 & 8 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$

(d) A has eigenvalues $-237, 0, 237$

(e) L is a surjective map

(f) L is an injective map

5.2.2 Determine whether the function is a local diffeomorphism at the given point. Briefly justify.

(a) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be C^1. If $DF(0,0) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, is F a local diffeomorphism at $(0,0)$?

(b) Let $G : \mathbb{R}^3 \to \mathbb{R}^3$ be C^1. If $DG(2,3,7) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$, is G a local diffeomorphism at $(2,3,7)$?

(c) Let $H : \mathbb{R}^{237} \to \mathbb{R}^{237}$ be C^1 and let $a \in \mathbb{R}^{237}$. If the differential dH_a is injective, then is H a local diffeomorphism at a?

(d) Let $\phi : \mathbb{R}^{237} \to \mathbb{R}^{237}$ be C^1 and let $a \in \mathbb{R}^{237}$. If the differential $d\phi_a$ is surjective, then is ϕ a local diffeomorphism at a?

5.2.3 Let $U, V \subseteq \mathbb{R}^n$ be open sets and let $F : U \to V$ be a C^1 function. Which of the following are true or false? If it is true, cite a theorem. If it is false, state a counterexample without proof.

(a) If F is a diffeomorphism, then F is a local diffeomorphism at every $a \in U$.

(b) If F is local diffeomorphism at every $a \in U$, then F is a diffeomorphism.

Computations

5.2.4 Let $p = (\frac{\pi}{4}, 0)$ and let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $F(x,y) = \begin{bmatrix} xy \\ \sin x + \cos y \end{bmatrix}$

(a) Show F is a local diffeomorphism at p.

(b) Let G be a local C^1 inverse of F at p. Compute the Jacobian of G by hand.

5.2.5 Let $p = (0,0)$ and let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $F(x,y) = \begin{bmatrix} e^x(y^2 - 3x + 1) \\ x\log(y^2+1) + y \end{bmatrix}$

(a) Show F is a local diffeomorphism at p.

(b) Let G be a local C^1 inverse of F at p. Compute the Jacobian of G at $F(p)$ only.

5.2.6 Let $p = (\pi, 0, \pi, 0)$ and let $F : \mathbb{R}^4 \to \mathbb{R}^4$ be given by

$$F(w,x,y,z) = \begin{bmatrix} x^2 + y^2 + z^2 \\ wx + xy + yz - zw \\ x \sin y + w \cos z \\ 2x^2 + 4wxyz - 2z^2 + z + 1 \end{bmatrix}$$

(a) Show F is a local diffeomorphism at p.

(b) Let G be a local C^1 inverse of F at p. Compute the Jacobian of G with a computer.

5.2.7 Show that $F(x,y) = (x^2+y^2, xy)$ is a local diffeomorphism at $(a,b) \in \mathbb{R}^2$ if and only if $a \neq \pm b$.

5.2.8 Define $F(x,y,z) = \big(x+y+z, xy+xz+yz, xyz\big)$. Let a, b, c be three *distinct* real numbers.

(a) Show that the function F is a local diffeomorphism at $(a,b,c) \in \mathbb{R}^3$.

(b) Let G be a local inverse of F at (a,b,c). Compute $DG(F(x,y,z))$ using a computer.

Proofs

5.2.9 The inverse function theorem (IFT) is amazing, but it has limitations. Here are several flawed arguments using the inverse function theorem beyond what it permits.

(a) Cameron gets too excited about IFT and writes an incorrect argument. Explain what is wrong.

1. *Define* $F(x,y) = (x^2+y^2, xy)$ *so* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *is* C^1.
2. *Note* $DF(x,y) = \begin{bmatrix} 2x & 2y \\ y & x \end{bmatrix}$, *so* $DF(1,0) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ *has determinant 2 and hence invertible.*
3. *By the inverse function theorem,* $F|_U : U \to F(U)$ *is invertible for any open set* U *with* $(1,0)$.

(b) Alisa is overjoyed by the IFT, but also makes a mistake. Explain what is wrong.

1. *The function* $F : \mathbb{R} \to \mathbb{R}$ *given by* $F(x) = \sin x$ *is* C^1.
2. *Notice* $F'(x) = \cos x$ *so* $F'(0) = \cos 0 = 1$ *is invertible.*
3. *By the inverse function theorem,* $G(x) = \arcsin x$ *for* $-1 < x < 1$ *is a local inverse of* F *at* 0.
4. *Thus,* F *is a local diffeomorphism at* 0.

(c) Abbigael is blinded by her love of IFT by concluding too much. Explain what is wrong.

1. *The function* $F : \mathbb{R} \to \mathbb{R}$ *given by* $F(x) = x^5$ *is* C^1.
2. *Note* $F'(x) = 4x^4$ *so* $F'(0) = 0$ *is not invertible.*
3. *By the inverse function theorem,* F *does not have a global inverse.*

(d) Sarah is a diehard IFT fan, but got lost in its magic. Explain what is wrong.

1. *The function* $F : \mathbb{R} \to \mathbb{R}$ *given by* $F(x) = xe^x$ *is* C^1.
2. *Note* $F'(x) = (x+1)e^x$ *so* $F'(1) = 2e^2$.
3. *By the inverse function theorem,* F *has a local* C^1 *inverse* G *at* 1.
4. *The linear approximation of* G *at* $F(1) = e$ *is given by* $\ell(y) = 1 + (2e^2)^{-1}(y - e)$.
5. *Thus,* $F(x) = 3$ *when* $x = G(3) \approx \ell(3) = 1 + (2e^2)^{-1}(3 - e) \approx 1.019$.

5.2.10 Prove that $F : \mathbb{R} \to \mathbb{R}$ given by $F(x) = \sin x$ is a local diffeomorphism at $a \in \mathbb{R}$ if and only if $a \neq \frac{\pi}{2} + k\pi$ for some $k \in \mathbb{Z}$.

5.2.11 Define the cylindrical coordinate map $F : \mathbb{R}^3 \to \mathbb{R}^3$ by $F(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$. Fix $r_0, \theta_0, z_0 \in \mathbb{R}$. Prove that F is a local diffeomorphism at (r_0, θ_0, z_0) if and only if $r_0 \neq 0$.

5.2.12 Define the spherical coordinate map $F : \mathbb{R}^3 \to \mathbb{R}^3$ by $F(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$. Fix $\rho_0, \theta_0, \phi_0 \in \mathbb{R}$. Prove that F is a local diffeomorphism at $(\rho_0, \theta_0, \phi_0)$ if and only if $\rho_0 \neq 0$ and ϕ_0 is not an integer multiple of π.

Applications and beyond

5.2.13 From an **algebraic** viewpoint, the inverse function theorem can be reformulated in several ways. Let $A, B \subseteq \mathbb{R}^n$ be open subsets of $\mathbb{R}^n$. Fix $a \in A$. Let $F : A \to B$ be a C^1 function. Fill in the blank so that the statement is equivalent to "*F is a local diffeomorphism at a.*"

(a) ________________________________ is an invertible $n \times n$ matrix.

(b) ________________________________ is an invertible linear map.

(c) ________________________________ are a set of n linearly independent vectors in $\mathbb{R}^n$.

5.2.14 From an **analytic** viewpoint, the inverse function theorem informally states that:

> *If the linear approximation of F at a is globally invertible, then a local inverse of F at a exists and you can linearly approximate its value.*

Let $A, B \subseteq \mathbb{R}^n$ be open subsets of $\mathbb{R}^n$. Fix $a \in A$. Let $F : A \to B$ be a C^1 function. The linear approximation of F at a is the map $\ell : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\ell(x) = F(a) + DF(a)(x - a).$$

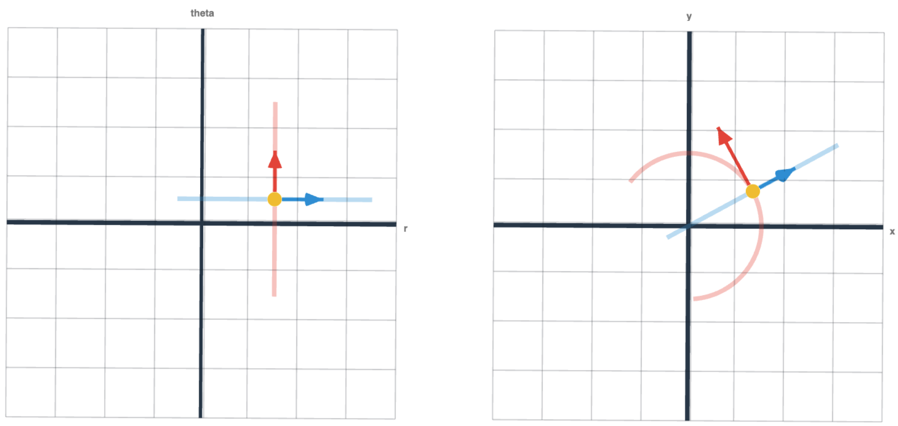
Assume the Jacobian matrix $DF(a)$ is invertible. Let G be a local C^1 inverse of F at a.

(a) Define the linear approximation $L : \mathbb{R}^n \to \mathbb{R}^n$ of G at $F(a)$.

(b) Show that L is the inverse of ℓ, and hence ℓ is globally invertible.

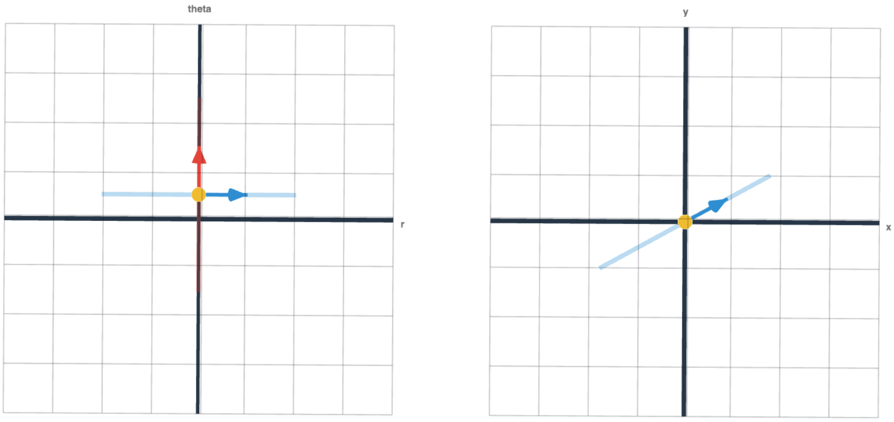
5.2.15 A **geometric/physical** viewpoint of the inverse function theorem refers to "directions of motion".

(a) Define the polar coordinates map $F : \mathbb{R}^2 \to \mathbb{R}^2$ by $F(r,\theta) = (r\cos\theta, r\sin\theta)$. Fix a point $(a,b) \in \mathbb{R}^2$ with $a \neq 0$. Label the points (a,b), $F(a,b)$, and vectors in the plot below of F.



Notice F is a local diffeomorphism at (a, b).

(b) The Math3D demo below illustrates the polar coordinates map F. Fix a point $(0, b) \in \mathbb{R}^2$. Label the points $(0, b), F(0, b)$, and the vectors.



Notice F is *not* a local diffeomorphism at $(0, b)$.

(c) Let $A, B \subseteq \mathbb{R}^n$ be open. Let $F : A \to B$ be C^1 and fix $a \in A$. Complete the geometric viewpoint.

F is a local diffeomorphism at a if and only if the n directions of motion ____________________

transformed to ____________________ *remain* ____________________.

5.3. Nonlinear systems

Solving *linear* systems of equations was the cornerstone of linear algebra. Extending those ideas to *nonlinear* systems of equations will be a remarkable achievement for multivariable calculus with applications to analysis, algebra , and geometry. Many scientific phenomena also involve such systems. For instance, the pressure, volume, and temperature of an ideal gas are related by a nonlinear equation in thermodynamics. Economists' models of global supply chains can be described with nonlinear systems of equations. Data scientists try to describe their high dimensional data in lower dimensions via smooth manifolds defined by nonlinear maps. The applications beyond mathematics are aplenty.

5.3.1 Setup of a nonlinear system

The basic problem is reminiscent of linear algebra and informally described here. Given a nonlinear map $F : \mathbb{R}^N \to \mathbb{R}^k$, you have a system of k nonlinear equations with N variables $x_1, \dots, x_N$ given by

$$F(x) = 0 \quad \Longleftrightarrow \quad \begin{array}{rcl} F_1(x_1,\dots,x_N) & = & 0 \\ & \vdots & \\ F_k(x_1,\dots,x_N) & = & 0 \end{array}$$

Note the zero vector $0 \in \mathbb{R}^k$ is chosen by convention[3] to define this system of equations.

Example 5.3.1 The nonlinear system of equations given by

$$\begin{array}{rcl} x_1x_2 + x_2x_3 + x_1x_3 & = & 0 \\ x_1^2 - x_2^2 + x_3^2 & = & 0 \end{array}$$

corresponds to $F(x_1, x_2, x_3) = (x_1x_2 + x_2x_3 + x_1x_3, x_1^2 - x_2^2 + x_3^2)$.

There are several basic questions you must address. Here is the first.

1. *Does there exist a single solution to this system of nonlinear equations?*

This question is notoriously difficult in general; it is an active area of research! In most scenarios, your nonlinear system will have at least one solution that is "obvious" because of how you constructed the system. For instance, for any $k \times N$ matrix A, the linear system $Ax = 0$ always has the trivial solution $x = 0 \in \mathbb{R}^N$. It is better to make an assumption.

1. *Assume at least one (easy-to-find) solution exists.*

This handles the first question by simply ignoring it.

Example 5.3.2 The nonlinear system in Example 5.3.1 has at least one easy-to-find solution, namely $(x_1, x_2, x_3) = (0, 0, 0)$.

Here is the second fundamental question.

2. *Do infinitely many solutions exist for this system of nonlinear equations?*

You will usually want the answer to be 'yes', but this second question is still quite difficult. You can gain some insight from linear algebra. For a $k \times N$ matrix A, the linear system $Ax = 0$ is over-determined if $N < k$, or uniquely determined if $N = k$. In other words, if $N \leq k$, then

[3] You can replace $0 \in \mathbb{R}^k$ with another constant vector, say $c \in \mathbb{R}^k$, but the problems are of equivalent difficulty. If you define $G(x) = F(x) - c$ then $G(x) = 0$ if and only if $F(x) = c$. In an effort to reduce notation, you will keep the convention of having all equations equal 0 in all definitions and theorems.

there are at least as many equations as variables, so you would usually expect that no other solutions exist for the linear system $Ax = 0$ aside from $x = 0$ for most[4] matrices A.

On the other hand, if $N > k$, then there are more variables than equations, so a $k \times N$ linear system $Ax = 0$ is under-determined and there *must* be infinitely many solutions. Again, since your theory of differential calculus relies so heavily on linear algebra, this suggests that you should restrict your attention to finding solutions of nonlinear systems when

$$N = n + k > k$$

for some $n \in \mathbb{N}^+$. Unlike for linear systems, this assumption will not guarantee a nonlinear system has infinitely many solutions, but it gives you a much better chance.

Example 5.3.3 The system in Example 5.3.1 has $N = 3$ and $k = 2$ so $n = 1$ and $N > k$.

How can you show infinitely many solutions exist? The simplest idea is to explicitly solve for some variables in terms of the others.

Example 5.3.4 The nonlinear equation $x^2 + y^2 - 1 = 0$ in 2 variables x and y can be partially solved. If $y = \sqrt{1-x^2}$ and $-1 < x < 1$, then the pair $(x, y) \in \mathbb{R}^2$ satisfies the equation. By expressing y as a function of x, each $x \in (-1, 1)$ produces an *explicit* solution $(x, \sqrt{1-x^2})$ to the *implicit* equation $x^2 + y^2 - 1 = 0$. This gives infinitely many solutions!

The second fundamental question therefore leads to the third.

3. Can you express some variables as a function of the others in a nonlinear system?

If the answer is yes, then you will have produced infinitely many solutions. You may want to know how changes in some variables affects others, so it is reasonable to require these variables depend smoothly on the others. This refines the third question.

3. Can you express some variables as a C^1 function of the others?

However, for a system with $n + k$ variables and k equations, how many variables do you expect to depend on the others? It will help to specify this in advance. Linear algebra again provides some insight. For most matrices A with k rows and $n+k$ columns, you expect the set of solutions $\{x \in \mathbb{R}^{n+k} : Ax = 0\}$ to form a subspace of dimension $n = (n+k)-k$ by the rank-nullity theorem. Informally speaking, $n + k$ variables with k equations will usually have n *independent variables* and k *dependent variables*. This leads to the ultimate question.

4. For a nonlinear system with $n + k$ variables and k equations, can you express k variables as a C^1 function of the other n variables?

This question motivates the following setup, which will be the centre of your study.

Given an open set $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ and a C^1 map

$$F : U \longrightarrow \mathbb{R}^k,$$

you want to solve the system of k nonlinear equations with $n + k$ variables given by

$$F(x, y) = 0 \quad \Longleftrightarrow \quad \begin{array}{rcl} F_1(x_1, \dots, x_n, y_1, \dots, y_k) & = & 0 \\ & \vdots & \\ F_k(x_1, \dots, x_n, y_1, \dots, y_k) & = & 0 \end{array} \tag{5.3.1}$$

Thus, your ultimate goal is to express the k dependent variables $y_1, \dots, y_k$ in terms of the n independent variables $x_1, \dots, x_n$.

[4] By "most matrices", it is better to think of a matrix with randomly-chosen entries. Some matrices will have infinitely many solutions but you can consider this highly unlikely.

Remark 5.3.5 There are a few conventions to clarify. First, note $(x, y) \in \mathbb{R}^n \times \mathbb{R}^k$ means $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^k$. Second, the choice of letters x and y does not matter, but it is often used to remind you which variables are expressed in terms of the others. Third, the order of the variables as inputs of F does not matter (e.g. $F(y_1, x_2, x_1, y_3, \dots)$ is valid) but the ordering $F(x_1, \dots, x_n, y_1, \dots, y_k)$ is often written for the sake of consistency. Any theorem or definition will be written with this variable ordering, but it applies to any ordering with suitable minor modifications.

That completes the informal setup! Now, your current objective is to formally describe what it means to solve the nonlinear system (5.3.1).

5.3.2 Single nonlinear equation

To keep things simple, focus on the case of $k = 1$ nonlinear equation in (5.3.1) with n variables. The unit circle in $\mathbb{R}^2$ is the quintessential example.

Example 5.3.6 Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2 - 1$. The unit circle in $\mathbb{R}^2$ is described by the implicit equation

$$f(x, y) = 0 \iff x^2 + y^2 - 1 = 0.$$

Can you solve for x as a C^1 function of y? Or for y as a C^1 function of x? On one hand, the answer is 'No' because the unit circle is not the graph of any function; it does not pass the vertical line test nor the horizontal line test. On the other hand, the answer is 'Yes' because you can *locally* express the circle as at least one of the four following graphs of C^1 functions:

$$\begin{aligned} y &= \sqrt{1 - x^2} \quad -1 < x < 1; \qquad & y &= -\sqrt{1 - x^2} \quad -1 < x < 1; \\ x &= \sqrt{1 - y^2} \quad -1 < y < 1; \qquad & x &= -\sqrt{1 - y^2} \quad -1 < y < 1. \end{aligned}$$

This idea is akin to smooth manifolds from Section 4.6, except now you will care about *which* of these four graphs you are using.

This example illustrates that finding a global solution to (5.3.1) will often be impossible. As with inverse functions and smooth manifolds, you must settle for *local* solutions.

Definition 5.3.7 Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be an open set. Let $f : U \to \mathbb{R}$ be a real-valued C^1 function. Let $(a, b) \in U$ so $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Assume $f(a, b) = 0$. The equation

$$f(x_1, \dots, x_n, y) = 0$$

locally defines y **as a** C^1 **function of** $x = (x_1, \dots, x_n)$ **near** (a, b) if there exists an open set $V \subseteq \mathbb{R}^n$ containing a, an open set $W \subseteq \mathbb{R}$ containing b, and a C^1 function $\phi : V \to W$ such that $V \times W \subseteq U$ and

$$\forall (x, y) \in V \times W, \qquad f(x, y) = 0 \iff y = \phi(x).$$

In other words, $\{(x, y) \in V \times W : f(x, y) = 0\} = \{(x, \phi(x)) : x \in V\}$.

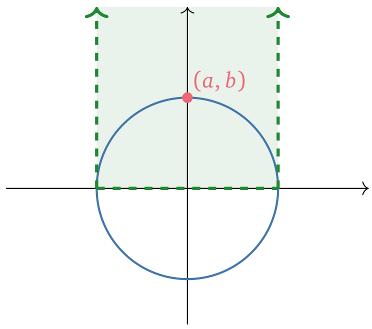
This definition is rather overwhelming, so use the unit circle for context.

Example 5.3.8 Using the notation of Definition 5.3.7, take $U = \mathbb{R} \times \mathbb{R}$ so $n = 1$ and define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2 - 1$. Let $(a, b) = (0, 1) \in U$ so $f(a, b) = 0^2 + 1^2 - 1 = 0$.

Does the equation

$$f(x,y) = 0 \iff x^2 + y^2 - 1 = 0$$

locally define y as a C^1 function of x near $(a,b) = (0,1)$? The diagram below suggests yes.



Take $V = (-1,1)$ and $W = (0,\infty)$, so V is an open set containing $a = 0$ and W is an open set containing $b = 1$. Define the C^1 function $\phi : V \to W$ by

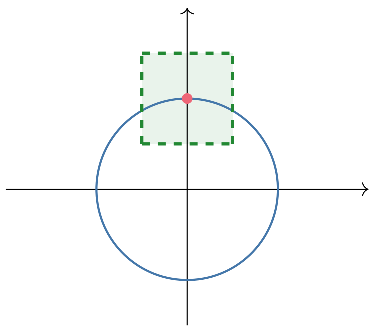
$$\forall x \in (-1,1), \quad \phi(x) = \sqrt{1-x^2}.$$

Since $V \times W \subseteq U = \mathbb{R}^2$, it suffices to prove for all $(x,y) \in V \times W$ that $f(x,y) = 0$ if and only if $y = \phi(x)$. Fix $(x,y) \in V \times W$ so $x \in (-1,1)$ and $y \in (0,\infty)$. Then

$$\begin{aligned} x^2 + y^2 - 1 = 0 &\iff y^2 = 1 - x^2 \\ &\iff y = \sqrt{1-x^2} && \text{as } y > 0. \end{aligned}$$

This proves the equation $x^2 + y^2 - 1 = 0$ locally defines y as C^1 function of x near $(0,1)$.

Now, does $x^2 + y^2 - 1 = 0$ locally define x as C^1 function of y near $(0,1)$? Informally, the circle cannot pass the horizontal line test near $(0,1)$ no matter how small you shrink $V \times W$. Hence, the answer must be 'No'. This is illustrated by the plot below.



A formal proof must swap the roles of x and y in Definition 5.3.7. Let V be any open subset containing 1 and let W be any open subset containing 0. Suppose, for a contradiction, that there exists a C^1 function $\phi : V \to W$ such that $f(x,y) = 0$ if and only if $x = \phi(y)$ for all $y \in V$ and $x \in W$. Since $1 \in V$, there exists $\varepsilon > 0$ such that $(1-2\varepsilon, 1+2\varepsilon) \subseteq V$ so $\phi(1+\varepsilon)$ must be defined. This implies that

$$f(\phi(1+\varepsilon), 1+\varepsilon) = 0 \iff \phi(1+\varepsilon)^2 + (1+\varepsilon)^2 - 1 = 0 \iff \phi(1+\varepsilon)^2 = -2\varepsilon - \varepsilon^2,$$

so $\phi(1+\varepsilon)^2 < 0$ which is a contradiction. This proves $f(x,y) = 0$ does not locally define x as a C^1 function of y near $(0,1)$.

Remark 5.3.9 To obtain a contradiction, you can instead show that there are two values $\pm\delta \in W$ where $f(\pm\delta, \sqrt{1-\delta^2}) = 0$ and $\sqrt{1-\delta^2} \in V$ which would imply $\phi(\sqrt{1-\delta^2})$ outputs two different values $\pm\delta$. That is a contradiction since ϕ is a function.

As you can see, it is quite tricky to verify Definition 5.3.7 even in simple examples. Similar equations can suddenly be impossible to solve.

Example 5.3.10 The equations $x^2 - y^3 = 0$ has a solution at the origin $(0,0)$. Does this define y locally as a C^1 function of x, or vice versa? The equation itself seems like a harmless polynomial at first glance, but it is the cusp from Example 4.6.10. You cannot locally solve for either variable near the origin!

The equation $x^4 - x^2 + y^2 = 0$ in $\mathbb{R}^2$ has a solution at the origin $(0,0)$. Does this define y locally as a C^1 function of x, or vice versa? Again, it seems innocuous but it is the figure eight from Example 4.6.10. You cannot locally solve for either variable near the origin!

It should be frightening that such simple equations can have such terrible behaviour. How would you detect this bad behaviour with more variables? For instance, the 4-variable implicit equation

$$x^2 + y^2 - z^2 + xyz - wz^3 + y^3 = 0$$

is impossible to plot. Can you solve for z in terms of the other variables or not? It is hard to tell. Despite this disheartening situation for a nonlinear equation, you can at least fully characterize when you can solve a *linear* equation and this leads to a critical insight.

Example 5.3.11 Fix constants $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $\beta \in \mathbb{R}$. When does the linear equation

$$\alpha_1 x_1 + \cdots + \alpha_n x_n + \beta y = 0$$

locally define y as a C^1 function of $x_1, \ldots, x_n$? This answer is straightforward, since you can actually *globally* solve for y. Namely, you can express

$$y = -\frac{1}{\beta}(\alpha_1 x_1 + \cdots + \alpha_n x_n) \quad \text{if and only if} \quad \beta \neq 0.$$

The simple criterion $\beta \neq 0$ fully determines whether you can solve the linear equation. If you reformulate everything in terms of Definition 5.3.7, you can discover an important feature. Namely, define $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ by

$$f(x_1, \ldots, x_n, y) = \alpha_1 x_1 + \cdots + \alpha_n x_n + \beta y$$

so the linear equation $f(x_1, \ldots, x_n, y) = 0$ *globally* defines y as C^1 function of x provided

$$\frac{\partial f}{\partial y} = \beta \neq 0.$$

The partial derivative detects whether the linear equation is solvable! This discovery will be incredibly helpful in the next section. Note the $y = \phi(x)$ where $\phi : \mathbb{R}^n \to \mathbb{R}$ is the linear function given by $\phi(x) = -\frac{1}{\beta}(\alpha_1 x_1 + \cdots + \alpha_n x_n)$.

5.3.3 Many nonlinear equations

Now, you can consider the nonlinear system (5.3.1) in full generality, so the number of equations $k \in \mathbb{N}^+$ and the number of variables $n \in \mathbb{N}^+$ are both arbitrary. On the bright side, the definition of locally solving such systems generalizes Definition 5.3.7 without much difficulty.

Definition 5.3.12 Let $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be an open set. Let $F : U \to \mathbb{R}^k$ be a $\mathbb{R}^k$-valued C^1 function. Let $(a,b) \in U$ so $a = (a_1,\ldots,a_n) \in \mathbb{R}^n$ and $b = (b_1,\ldots,b_k) \in \mathbb{R}^k$. Assume $F(a,b) = 0$. The equation

$$F(x_1,\ldots,x_n,y_1,\ldots,y_k) = 0$$

locally defines y **as a** C^1 **function of** x **near** (a,b) if there exists an open set $V \subseteq \mathbb{R}^n$ containing a, an open set $W \subseteq \mathbb{R}^k$ containing b, and a C^1 function $\phi : V \to W$ such that $V \times W \subseteq U$ and for all $(x,y) \in V \times W$,

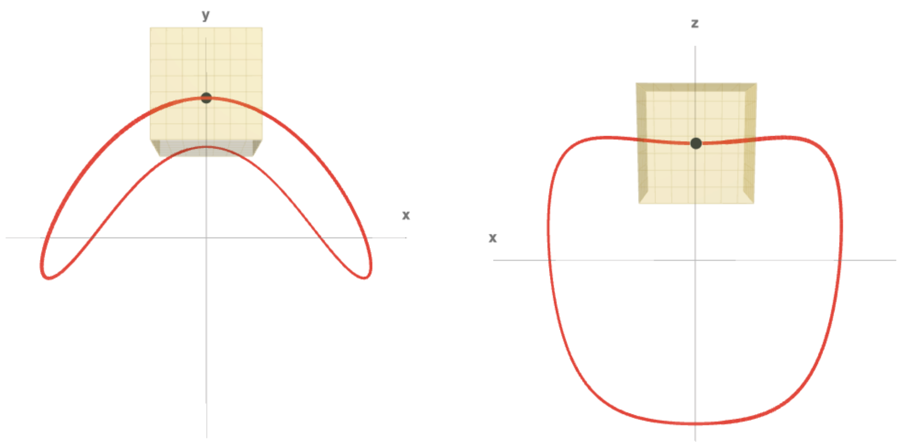
$$F(x,y) = 0 \iff y = \phi(x).$$

On the dark side, many nonlinear equations are a nightmare to solve, usually impossible to visualize, and the notation can get out of control. There is exactly one scenario with multiple equations that you can visualize: 3 variables and 2 equations in $\mathbb{R}^3$. Here is an example.

Example 5.3.13 Consider the system of nonlinear equations

$$x^2 + y^2 + z^2 = 16$$
$$(y-2)^2 + z^2 = 9$$

Set $a = 0$ and $b = (\frac{11}{4}, \frac{3\sqrt{15}}{4})$. You can check that $(a,b) = (0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$ is a solution. Can you locally define (y,z) as a C^1 function of x near (a,b)? Even for such simple equations, the answer is not obvious but you can explore this geometrically with this Math3D demo. It turns out the solutions to this system form a 1-dimensional regular curve.



Fix $\varepsilon > 0$. Set $V = (-\varepsilon, \varepsilon)$ and $W = (\frac{11}{4} - \varepsilon, \frac{11}{4} + \varepsilon) \times (\frac{3\sqrt{15}}{4} - \varepsilon, \frac{3\sqrt{15}}{4} + \varepsilon)$ so $a \in V$ and $b \in W$. The cube $V \times W$ is illustrated in the plots above with $\varepsilon = 1$. The plot on the left illustrates that inside the cube, y appears to be a function of x. The plot on the right illustrates that inside the cube, z appears to be a function of x. Thus, the variables (y,z) appear to be locally a C^1 function of x near (a,b). Confirming this formally will need some serious theorems.

Luckily, you can still fully characterize solutions to any system of *linear* equations.

Example 5.3.14 Consider the system of linear equations

$$x_1 + 2x_2 + x_3 - x_4 = 0$$
$$x_1 + 0x_2 + x_3 + x_4 = 0$$

Can you express x_1 and x_2 as C^1 functions of x_3 and x_4? By row-reducing the corresponding matrix A, you find that

$$A = \begin{bmatrix} 1 & 2 & 1 & -1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Treating x_3 and x_4 as the free variables, this implies that

$$x_1 = -x_3 - x_4, \quad x_2 = x_4,$$

so you can *globally* express x_1 and x_2 as C^1 functions of x_3 and x_4. Can you express x_1 and x_3 as C^1 functions of x_2 and x_4? Unfortunately from the reduced echelon form of A, the only equation you have with x_1 or x_3 is $x_1 + x_3 = -x_4$. It is therefore impossible to express x_1 (or x_3) in terms of only x_2 and x_4.

The difference between these two situations was the submatrices of the corresponding columns. Namely, for rref(A), the columns of x_1 and x_2 correspond to the 2×2 matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. This matrix is invertible, so the variables x_1 and x_2 are independent and you can solve for them in terms of x_3 and x_4. On the other hand, the columns of x_1 and x_3 correspond to the 2×2 matrix $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. This matrix is not invertible so the variables x_1 and x_3 are not independent so you cannot solve for them in terms of the others.

This idea for linear systems holds more generally.

Lemma 5.3.15 Let A be an $k \times n$ matrix and let B be a $k \times k$ matrix. The matrix B is invertible if and only if the system of k linear equations with $n + k$ variables

$$\begin{bmatrix} A & | & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

globally defines $y = (y_1, \dots, y_k) \in \mathbb{R}^k$ as a C^1 function of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

This is a standard consequence of linear algebra and generalizes Example 5.3.11.

Proof. (Sketch) If B is invertible, you can row reduce the $k \times (n + k)$ matrices as follows:

$$\begin{bmatrix} A & | & B \end{bmatrix} \sim \cdots \sim \begin{bmatrix} R & | & I \end{bmatrix}$$

where I is the $k \times k$ identity matrix and R is similar to A. Since I is the identity matrix, you can directly solve for y in terms of x with the equivalent system

$$\begin{bmatrix} R & | & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

because the jth row gives a linear equation for y_j. If B is not invertible, then a similar linear algebra argument shows you cannot solve for y in terms of x because there will be more than n free variables after row reduction. ∎

Thus far, you have formalized what it means to locally solve the nonlinear system (5.3.1), but these systems of equations are so complicated that it seems to be nearly impossible to actually verify this definition. Your only glimmer of hope comes from Lemma 5.3.15, which demonstrates that you can *globally* solve linear systems with a straightforward criterion. Since derivatives can approximate nonlinear quantities by linear quantities, will differential calculus come to your rescue with solving nonlinear systems? That is the next episode.

Exercises for Section 5.3

Concepts and definitions

5.3.1 Linear systems inform your intuition for nonlinear systems, so you will need a solid foundation. Suppose you are given a (homogenous) *linear* system of equations with coefficient matrix A.

(a) There is always at least one solution $x \in \mathbb{R}^N$ to the equation $Ax = 0$. Identify that solution.

(b) The set of solutions is given by $\operatorname{null}(A) = \{x \in \mathbb{R}^N : Ax = 0\}$, the null space of A. It is a subspace of $\mathbb{R}^N$. Assume the entries of your matrix are *randomly chosen*. For each case below, what would you expect to be the value of the dimension of $\operatorname{null}(A)$?

i) The system has 7 linear equations and 23 variables, so A is a 7×23 matrix.
ii) The system has 23 linear equations and 7 variables, so A is a 23×7 matrix.
iii) The system has 23 linear equations and 23 variables, so A is a 23×23 matrix.

5.3.2 Consider the *linear* map $L : \mathbb{R}^4 \to \mathbb{R}^2$ defined by $L(v) = Av$ for $v \in \mathbb{R}^4$ where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 12 \end{bmatrix} \quad \text{and} \quad \operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

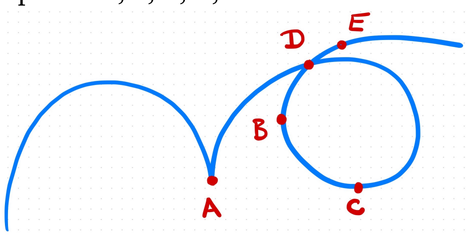
Almost no calculations are needed for any part below. Use 2×2 submatrices to decide.

(a) Does the *linear* system $L(x_1, x_2, x_3, x_4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ globally define (x_1, x_2) as a C^1 map f of (x_3, x_4)?

(b) Does the *linear* system $L(x_1, x_2, x_3, x_4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ globally define (x_1, x_3) as a C^1 map g of (x_2, x_4)?

(c) Does the *linear* system $L(x_1, x_2, x_3, x_4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ globally define (x_2, x_4) as a C^1 map h of (x_1, x_3)?

5.3.3 Below a curve in $\mathbb{R}^2$ defined by $F(x, y) = 0$. Five points A, B, C, D, E are listed on the curve.



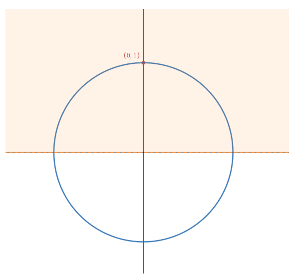
(a) At which points does the equation $F(x, y) = 0$ locally define y as a C^1 function of x?

(b) At which points does the equation $F(x, y) = 0$ locally define x as a C^1 function of y?

5.3.4 The equation $x^2 + y^2 = 1$ locally defines y as a C^1 map $\phi : V \to W$ of x near $(0, 1)$, i.e.

$$\forall (x, y) \in V \times W, \quad x^2 + y^2 - 1 = 0 \iff y = \phi(x).$$

Victoria needs to choose sets V and W and a map ϕ to write a successful proof. She starts by choosing the sets $V = (-\infty, \infty)$ and $W = (0, \infty)$, but still needs to choose a function.



(a) Victoria defines $\phi_A(x) = \sqrt{1 - x^2}$ for $|x| < 1$. Can her choice $\phi = \phi_A$ succeed?

(b) Victoria defines $\phi_B(x) = \sqrt{1 - x^2}$ for $|x| < 1$ and $\phi_B(x) = 0$ for $|x| \geq 1$. Can her choice $\phi = \phi_B$ succeed?

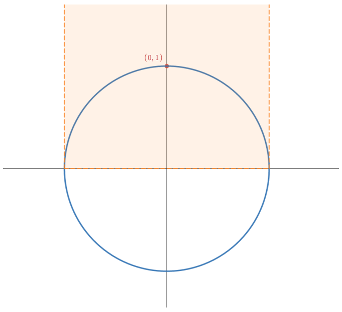
(c) What are all possible choices of ϕ that can succeed? Explain.

5.3.5 The equation $x^2 + y^2 = 1$ locally defines y as a C^1 map $\phi : V \to W$ of x near $(0, 1)$, i.e.

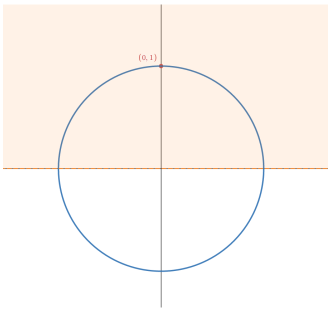
$$\forall (x, y) \in V \times W, \quad x^2 + y^2 - 1 = 0 \iff y = \phi(x).$$

Which choice of $V \subseteq \mathbb{R}$ and $W \subseteq \mathbb{R}$ can successfully lead to a correct proof? If so, define ϕ. Play with this Desmos graph for better visuals.

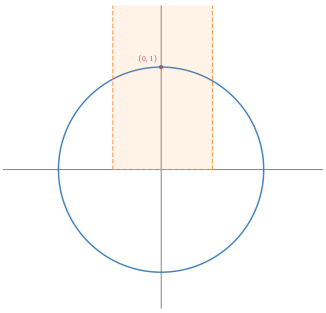
(a) $V = (-1, 1)$ and $W = (0, \infty)$



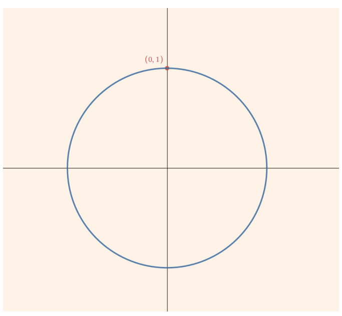
(b) $V = \mathbb{R}$ and $W = (0, \infty)$



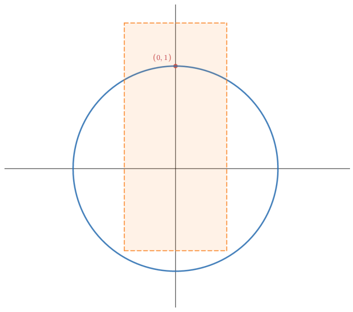
(c) $V = (-0.5, 0.5)$ and $W = (0, \infty)$



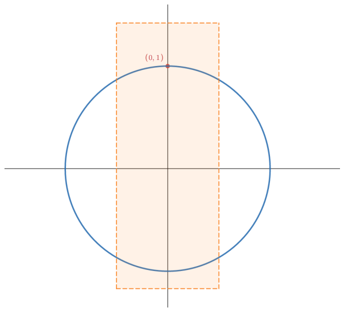
(d) $V = \mathbb{R}$ and $W = \mathbb{R}$



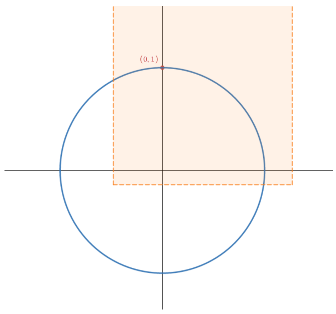
(e) $V = (-0.5, 0.5)$ and $W = (-0.8, 1.5)$



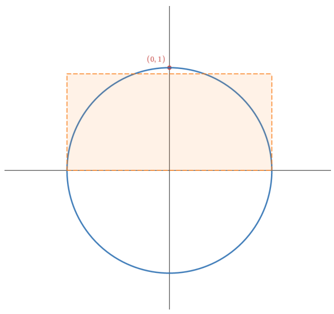
(f) $V = (-0.5, 0.5)$ and $W = (-1.2, 1.5)$



(g) $V = (-0.5, 1.2)$ and $W = (-0.1, \infty)$



(h) $V = (-1, 1)$ and $W = (0, 0.9)$

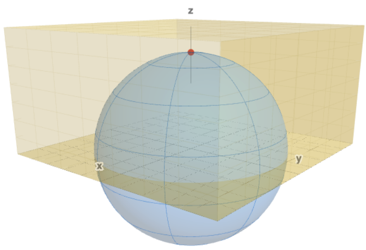


5.3.6 Note $x^2 + y^2 + z^2 = 16$ locally defines z as a C^1 map $\phi : V \to W$ of (x, y) near $(0, 0, 4)$, i.e.

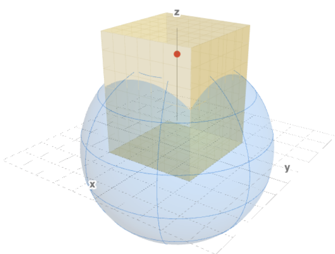
$$\forall (x, y, z) \in V \times W, \quad x^2 + y^2 + z^2 = 16 \iff z = \phi(x, y).$$

Which choice of $V \subseteq \mathbb{R}^2$ and $W \subseteq \mathbb{R}$ can successfully lead to a correct proof? If so, define ϕ. Play with this Math3D box demo and this Math3D cylinder demo for better visuals.

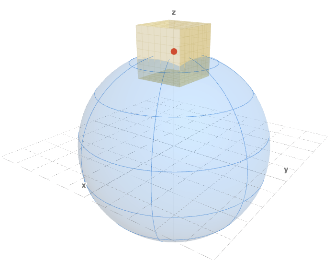
(a) $V = (-5, 5)^2$
$W = (0, \infty)$



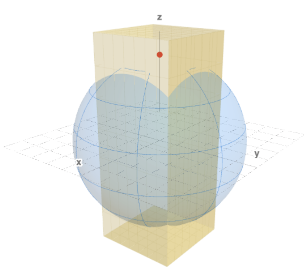
(b) $V = (-2, 2)^2$
$W = (0, \infty)$



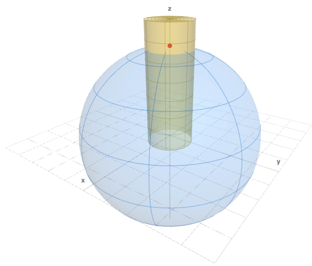
(c) $V = (-1, 1)^2$
$W = (3, 5)$



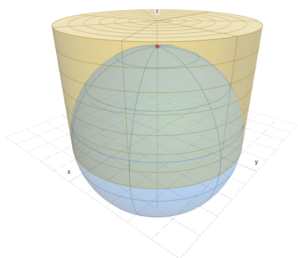
(d) $V = (-2, 2)^2$
$W = (-\infty, \infty)$



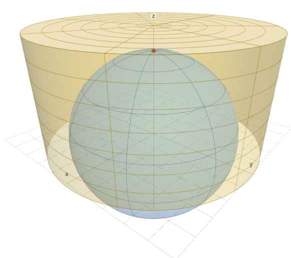
(e) $V = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$
$W = (0, \infty)$



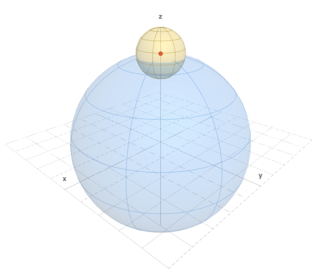
(f) $V = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 16\}$
$W = (0, \infty)$



(g) $V = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 25\}$
$W = (0, \infty)$



(h) $V \times W = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + (z-4)^2 < 1\}$



Computations

5.3.7 Consider the *linear* maps

$$F(w,x,y,z) = A_1w + B_1x + C_1y + D_1z \qquad G(w,x,y,z) = A_2w + B_2x + C_2y + D_2z$$

for fixed constants $A_1, A_2, B_1, B_2, C_1, C_2, D_1, D_2 \in \mathbb{R}$.

(a) What condition on the constants $A_1, B_1, C_1, D_1 \in \mathbb{R}$ guarantees that you can solve the equation

$$F(w,x,y,z) = 0$$

for y in terms of w, x, and z?

(b) What condition on the constants $A_1, A_2, B_1, B_2, C_1, C_2, D_1, D_2 \in \mathbb{R}$ guarantees that you can solve the system of equations

$$F(w,x,y,z) = 0, \qquad G(w,x,y,z) = 0$$

for y, z in terms of w, x?

5.3.8 Consider the linear equation $Ax = 0$ for $x \in \mathbb{R}^5$, where

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 4 \\ 1 & 2 & 1 & 1 & 6 \\ 0 & 1 & 1 & 1 & 3 \end{bmatrix} \quad \text{so} \quad \operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Almost no calculations are needed for any part below.

(a) Does the linear system $Ax = 0$ globally define (x_1, x_2, x_3) as a C^1 function of (x_4, x_5)?

(b) Does the linear system $Ax = 0$ globally define (x_1, x_2, x_4) as a C^1 function of (x_3, x_5)?

(c) Does the linear system $Ax = 0$ globally define (x_2, x_3, x_4) as a C^1 function of (x_1, x_5)?

5.3.9 Fix $k, n \in \mathbb{N}^+$. Let A be a matrix with $k + n$ columns and k rows.

(a) A matrix has **full rank** if its rank is as large as possible. If A has full rank, what is rank(A)?

(b) The linear system $Ax = 0$ gives a relationship for the $n + k$ variables $x_1, \ldots, x_{n+k}$. If A has full rank, then explain why you can choose k of these variables so that they are *globally* defined in terms of the other n variables.

5.3.10 The equation $x^2 + y^2 = 1$ locally defines (at least) 1 variable in terms of the other 1 variable near every point. The proof only needs exactly four functions $\phi : V \to W$ with $V \subseteq \mathbb{R}, W \subseteq \mathbb{R}$.

- $\phi_1 : V_1 \to W_1$ and $V_1 \times W_1 \subseteq \mathbb{R}^2$
- $\phi_2 : V_2 \to W_2$ and $V_2 \times W_2 \subseteq \mathbb{R}^2$
- $\phi_3 : V_3 \to W_3$ and $W_3 \times V_3 \subseteq \mathbb{R}^2$
- $\phi_4 : V_4 \to W_4$ and $W_4 \times V_4 \subseteq \mathbb{R}^2$

What are these four functions? Choose ones that yield a simple argument. No proof required.

5.3.11 The equation $x^2+y^2+z^2 = 1$ locally defines (at least) 1 variable in terms of the other 2 variables near every point. The proof only needs exactly six functions $\phi : V \to W$ with $V \subseteq \mathbb{R}^2, W \subseteq \mathbb{R}$. What are these six functions? Choose the ones that yield a simple argument. No proof required.

Proofs

5.3.12 Formal proofs that an equation can be locally solved (or not) are quite tricky since there are many technical details. Here are attempts related to the curve $x^2 = y^3$ plotted in Desmos.

(a) Maryam argues that $x^2 - y^3 = 0$ locally defines x as a C^1 function of y near $(1,1)$.

> 1. *Let* $V = (0,\infty)$ *and* $W = (0,\infty)$.
> 2. *Define* $\phi : V \to W$ *by* $\phi(y) = y^{3/2}$.
> 3. *Notice* $x^2 - y^3 = 0$ *if and only if* $x = y^{3/2} = \phi(y)$.
> 4. *Thus,* $x^2 - y^3 = 0$ *locally defines* x *as a* C^1 *function of* y *near* $(1,1)$.

There is no fatal error, but each of the first 3 lines could use a bit more detail. Add those details.

(b) Fabian argues that $x^2 - y^3 = 0$ does *not* locally define y as a C^1 function of x near $(0,0)$.

> 1. *Suppose for a contradiction that there exists a* C^1 *function* $\phi : \mathbb{R} \to \mathbb{R}$ *such that*
>
> $$\{(x,y) \in \mathbb{R}^2 : x^2 - y^3 = 0\} = \{(x,\phi(x)) : x \in \mathbb{R}\}.$$
>
> 2. *This implies for* $x \in \mathbb{R}$ *that* $x^2 - [\phi(x)]^3 = 0$ *so* $[\phi(x)]^3 = x^2$.
> 3. *Then* $\phi(x) = x^{2/3}$ *but* ϕ *is not differentiable at* 0.
> 4. *Thus,* ϕ *is not* C^1*, a contradiction.*

The proof has a flaw. Identify the flaw and suggest how to fix it.

(c) Semeon argues that $x^2 - y^3 = 0$ does *not* locally define x as a C^1 function of y near $(0,0)$.

> 1. *Let* $V,W \subseteq \mathbb{R}$ *be open subsets containing* 0*, so* $W \times V$ *contains* $(0,0)$.
> 2. *Suppose for a contradiction that there exists a* C^1 *function* $\phi : V \to W$ *such that*
>
> $$\{(x,y) \in W \times V : x^2 - y^3 = 0\} = \{(\phi(y),y) : y \in V\}.$$
>
> 3. *This implies for* $y \in V$ *that* $[\phi(y)]^2 - y^3 = 0$ *so* $\phi(y) = y^{3/2}$.
> 4. *Then* $\phi(y) = y^{3/2}$ *but* ϕ *is not differentiable at* 0.
> 5. *Thus,* ϕ *is not* C^1*, a contradiction.*

His proof has a serious flaw. Identify it and fix his proof.

5.3.13 Fix $(a,b) \in \mathbb{R}^2$ lying on the unit circle. You want to prove by definition that the equation $x^2 + y^2 = 1$ locally defines y as C^1 function of x near $(a,b) \in \mathbb{R}^2$.

(a) If $b > 0$ then what choice of sets $V,W \subseteq \mathbb{R}$ and C^1 map $\phi : V \to W$ will be successful and easy? *Hint:* Your choice should not depend on (a,b) aside from $b > 0$.

(b) If $b < 0$ then what choice of sets $V,W \subseteq \mathbb{R}$ and C^1 map $\phi : V \to W$ will be successful and easy?

(c) Prove the case $b < 0$ with your choices above.

5.3.14 The equation $x^2 - 4y^2 = 0$ is satisfied by the points $(2,-1)$ and $(0,0)$.

(a) Prove that $x^2 - 4y^2 = 0$ locally defines x as a C^1 function of y near $(2,-1)$.

(b) Prove that $x^2 - 4y^2 = 0$ does not locally define y as a C^1 function of x near $(0,0)$.

(c) Prove that $x^2 - 4y^2 = 0$ does not locally define x as a C^1 function of y near $(0,0)$.

5.3.15 Prove that $x^2 + y^2 + z^2 = 16$ locally defines z as a C^1 function of (x,y) near $(0,0,4)$.

5.4. Implicit function theorem

This section will answer one of this chapter's fundamental questions.

When does a nonlinear system of equations have an explicit set of (local) solutions?

You have successfully formalized the concept of locally solving the nonlinear system (5.3.1) with Definition 5.3.12. This matches the geometric intuition illustrated by some basic examples, but it is not easy to verify. In addition to the definition's intricacy, finding local solutions to a nonlinear system is genuinely hard. You already assume at least one solution exists, but this critical assumption does not seemingly ease the task. As you shall see, the implicit function theorem overcomes this monumental challenge albeit with some limitations.

The first hint comes from asking another question in differential calculus.

Assume local solutions exist. If some variables change in a nonlinear system, how do the other variables change?

That is, you want to **implicitly differentiate** variables in a nonlinear system. Resolving this question suggests a sufficient condition for local solutions to exist and miraculously produces the implicit function theorem.

5.4.1 Implicit differentiation for one variable

To keep things simple, assume your nonlinear system has 1 equation and $n+1$ variables. Assuming some local solutions exist, if you change n variables, how will the remaining 1 variable change? You can explore some examples.

Example 5.4.1 The unit circle in $\mathbb{R}^2$ is defined by

$$x^2 + y^2 = 1,$$

which has 2 variables and 1 equation. Let (a,b) be a point on the unit circle. How does y change with x near (a,b)? This question does not make formal sense unless y can be treated as a C^1 function of x. In other words, you need that local solutions exist near (a,b).

Luckily, you can directly check Definition 5.3.7 to show that the equation $x^2+y^2=1$ locally defines y as a C^1 function of x near (a,b) provided $(a,b) \neq (\pm 1, 0)$. That is, you can confirm that local solutions exist near (a,b) for $(a,b) \neq (\pm 1, 0)$. This was proved in Example 5.3.8 with $(a,b) = (0,1)$.

Assume $(a,b) \neq (\pm 1, 0)$. Thus, there exists an open set $V \subseteq \mathbb{R}$ containing a, an open set $W \subseteq \mathbb{R}$ containing b, and a C^1 function $\phi : V \to W$ satisfying

$$\forall (x,y) \in V \times W, \qquad x^2 + y^2 = 1 \iff y = \phi(x).$$

Hence, for all $x \in V$,

$$x^2 + [\phi(x)]^2 = 1.$$

Differentiating this with single variable chain rule, you find that

$$2x + 2\phi(x)\phi'(x) = 0.$$

Ideally, you want to isolate for $\phi'(x)$ which requires dividing by $\phi(x)$. Is this quantity non-zero? Notice $\phi(a) = b$ by definition and $b \neq 0$ by assumption. Since ϕ is continuous, this implies that $\phi(x) \neq 0$ for $x \in I$ where $I \subseteq V$ is an open interval centered at a. Thus, you can conclude that

$$\phi'(x) = -\frac{x}{\phi(x)} \quad \text{for } x \in I.$$

Notice in particular that $\phi'(a) = -\frac{a}{\phi(a)} = -\frac{a}{b}$ will be defined if and only if $(a,b) \neq (\pm 1, 0)$. These are exactly the places where y can be written as a function of x near (a,b).

The calculation in Example 5.4.1 is the formal equivalent of *implicit differentiation*, which you have likely observed in single variable calculus.

Example 5.4.2 Here you will repeat Example 5.4.1 with variables and Leibniz notation. This will not be as formal and lack details, but it is quite convenient for calculations.

Assume $(a,b) \neq (\pm 1, 0)$. Since $x^2 + y^2 = 1$ locally defines y as a function of x near (a,b), you can implicitly differentiate y with respect to x. Abusing notation further, you may sometimes write $y = y(x)$. By the chain rule,

$$\frac{d}{dx}(x^2+y^2) = \frac{d}{dx}(1) \quad \Longrightarrow \quad 2x + 2y\frac{dy}{dx} = 0.$$

Since $b \neq 0$, this implies $y \neq 0$ for (x,y) near (a,b). It follows that

$$\frac{dy}{dx} = -\frac{x}{y} \quad \text{for } (x,y) \text{ near } (a,b).$$

Comparing with Example 5.4.1, observe that each informal step with variables corresponds to a rigorous formal explanation with functions.

Example 5.4.3 Consider the unit sphere in $\mathbb{R}^3$ defined by $x^2+y^2+z^2 = 1$. Fix $(a,b,c) \in \mathbb{R}^3$ and assume $c \neq 0$. You can verify that $x^2+y^2+z^2 = 1$ locally defines $z = z(x,y)$ as a C^1 function of (x,y) near (a,b,c). By the chain rule,

$$\begin{aligned}
\frac{\partial}{\partial x}(x^2+y^2+z^2) = \frac{\partial}{\partial x}(1) \quad &\Longrightarrow \quad 2x + 2z\frac{\partial z}{\partial x} = 0 \\
\frac{\partial}{\partial y}(x^2+y^2+z^2) = \frac{\partial}{\partial y}(1) \quad &\Longrightarrow \quad 2y + 2z\frac{\partial z}{\partial y} = 0
\end{aligned}$$

Since $c \neq 0$, this implies $z \neq 0$ for (x,y,z) near (a,b,c). It follows that

$$\frac{\partial z}{\partial x} = -\frac{x}{z}, \quad \frac{\partial z}{\partial y} = -\frac{y}{z}, \quad \text{for } (x,y,z) \text{ near } (a,b,c).$$

For instance, if $(a,b,c) = (0,0,1)$ then

$$\left.\frac{\partial z}{\partial x}\right|_{(x,y)=(0,0)} = 0, \quad \left.\frac{\partial z}{\partial y}\right|_{(x,y)=(0,0)} = 0.$$

as you might expect since the tangent plane is flat at the top of the sphere.

Implicit differentiation is almost always written with variables instead of functions because you can always convert the less formal calculation to a formal one. You will need to interpret the variable notation in terms of functions and, if necessary, translate for the sake of clarity. This technique of implicit differentiation is described in the following lemma.

Lemma 5.4.4 Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be an open set. Let $f : U \to \mathbb{R}$ be a real-valued C^1 function. Let $(a,b) \in U$ so $a = (a_1,\ldots,a_n) \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Assume $f(a,b) = 0$ and f is not constant. If the equation

$$f(x_1,\ldots,x_n,y) = 0$$

locally defines y as a C^1 function $\phi : V \to W$ of $x = (x_1,\ldots,x_n)$ near (a,b) then for every point $v = (v_1,\ldots,v_n) \in V$ with $w = \phi(v)$ and every $j \in \{1,2,\ldots,n\}$,

$$\frac{\partial f}{\partial x_j}(v,w) + \frac{\partial f}{\partial y}(v,w)\frac{\partial \phi}{\partial x_j}(v) = 0.$$

Remark 5.4.5 In particular, if $\frac{\partial f}{\partial y}(v,w) \neq 0$ then

$$\frac{\partial \phi}{\partial x_j}(v) = -\left(\frac{\partial f}{\partial y}(v,w)\right)^{-1} \frac{\partial f}{\partial x_j}(v,w). \tag{5.4.1}$$

Using variables and Leibniz notation, this can be informally written as

$$\frac{\partial y}{\partial x_j} = -\left(\frac{\partial f}{\partial y}\right)^{-1} \frac{\partial f}{\partial x_j}.$$

Proof. Define $g : V \to V \times W$ by $g(x) = (x,\phi(x)) = (x_1,\ldots,x_n,\phi(x_1,\ldots,x_n))$. For $x \in V$, it follows that

$$(f \circ g)(x) = f(x_1,\ldots,x_n,\phi(x_1,\ldots,x_n)) = 0$$

Fix $v \in V$ and $w = \phi(v)$. Thus, the Jacobian of $f \circ g : V \to \mathbb{R}$ at v is a $1 \times n$ zero matrix, so

$$D(f \circ g)(v) = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}.$$

From $g(v) = (v,\phi(v)) = (v,w)$ and the chain rule, it follows that $D(f \circ g)(v) = Df(v,w)Dg(v)$, which is equal to

$$\underbrace{\begin{bmatrix} \partial_1 f(v,w) & \partial_2 f(v,w) & \cdots & \partial_n f(v,w) & \partial_{n+1} f(v,w) \end{bmatrix}}_{1\times(n+1)} \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \partial_1\phi(v) & \partial_2\phi(v) & \cdots & \partial_n\phi(v) \end{bmatrix}}_{(n+1)\times n}.$$

The result follows by multiplying these and taking the jth component of the resulting vector. ∎

It is not worth memorizing Lemma 5.4.4, but digesting the ideas of the proof can be helpful. You may notice the proof is a slight generalization of the argument in Example 5.4.1. Remark 5.4.5 provides the crucial insight: assuming local solutions exist, if $\frac{\partial f}{\partial y} \neq 0$ then you can calculate how the 1 variable y changes with the n variables $x = (x_1,\ldots,x_n)$, that is, $\frac{\partial y}{\partial x_1},\ldots,\frac{\partial y}{\partial x_n}$ make formal sense. This condition for expressing derivatives of *local* solutions of a *nonlinear* equation matches the criterion in Example 5.3.11 for the existence of *global* solutions to a *linear* equation. This apparent coincidence is actually a breakthrough.

5.4.2 Implicit function theorem for one variable

These investigations have unearthed a spectacular theorem. By checking a simple partial derivative condition, you can locally solve any nonlinear equation given at least one solution.

Theorem 5.4.6 (Implicit function theorem) Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be open and let f be a real-valued C^1 function on U. Let $(a,b) \in U$ so $a \in \mathbb{R}^n, b \in \mathbb{R}$. If $f(a,b) = 0$ and $\dfrac{\partial f}{\partial y}(a,b) \neq 0$, then the equation $f(x,y) = 0$ defines y locally as a C^1 function ϕ of x near (a,b).

Remark 5.4.7 Notice the theorem does not give you an explicit function ϕ and it does not tell you how near you must be to (a,b) for local solutions to exist.

Proof. The proof is far beyond the scope of this course. See Shifrin [18, Chapter 6] or Krantz and Parks [15] for a detailed proof. ■

Informally speaking, the implicit function theorem states:

> *Assume a solution exists to your nonlinear equation. If you can globally solve an approximate linear equation, then you can locally solve the nonlinear equation.*

Information about a derivative gives information about the map itself. Yet another triumph for differential calculus! This theorem is also the rigorous justification for implicit differentiation in single variable calculus and (5.4.1) tells you how to perform implicit differentiation. You can apply it in examples.

Example 5.4.8 The equation $x^2 + \sin(x+y+z) = 1$ has a solution $(x,y,z) = (1,-4,3)$. Can you express x locally as a C^1 function of y and z? Define $F(x,y,z) = x^2 + \sin(x+y+z) - 1$. Notice $F : \mathbb{R}^3 \to \mathbb{R}$ is C^1 and $F(1,-4,3) = 0$. Moreover,

$$\frac{\partial F}{\partial x} = 2x + \cos(x+y+z) \implies \frac{\partial F}{\partial x}(1,-4,3) = 2 + \cos 0 = 3 \neq 0.$$

By the implicit function theorem, $F(x,y,z) = 0$ defines x locally as a C^1 function of y and z near $(1,-4,3)$. The ease of the argument is remarkable, since the equation itself seems impossible to explicitly solve. Also, compared to Example 5.3.8 with the unit circle and Definition 5.3.7, the simplicity of this argument is quite impressive.

While the implicit function theorem is spectacular, there are drawbacks Keeping with the notation of the theorem, the key limitations include:

1. The theorem requires you have at least one solution (a,b), which can be hard to find.
2. The theorem does not give an explicit formula for ϕ, so you cannot compute any solution (x,y) aside from (a,b).
3. The theorem does not give an explicit open set containing local solutions, so you cannot explicitly determine how close you must be to (a,b) in order for other solutions to exist.

Nonetheless, it is still a fantastic achievement and it will yield many important consequences.

5.4.3 Implicit function theorem for many variables

With a bit of linear algebra, implicit differentiation and the implicit function theorem can be generalized to any nonlinear systems with many equations. For this section, you will consider arbitrary nonlinear systems with k equations and $n+k$ variables as in (5.3.1).

Example 5.4.9 Recall Example 5.3.13 and the system of nonlinear equations $F(x,y,z) = (0,0)$ where $F : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$F(x,y,z) = \begin{bmatrix} x^2+y^2+z^2-16 \\ (y-2)^2+z^2-9 \end{bmatrix}.$$

Set $a = 0$ and $b = (\frac{11}{4}, \frac{3\sqrt{15}}{4})$ so $(a,b) = (0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$ is a solution to $F(x,y,z) = (0,0)$.

Assume $F(x,y,z) = (0,0)$ locally defines (y,z) as a C^1 function of x near (a,b). This is supported by the illustrations from Example 5.3.13. Thus, there exists an open set $V \subseteq \mathbb{R}$ containing 0, an open set $W \subseteq \mathbb{R}^2$ containing b, and a C^1 function $\phi : V \to W$ such that

$$\forall (x,y,z) \in V \times W, \quad F(x,y,z) = (0,0) \iff (y,z) = (\phi_1(x), \phi_2(x)).$$

Hence, for $x \in V$,

$$F(x, \phi_1(x), \phi_2(x)) = (0,0) \quad \implies \quad \begin{bmatrix} x^2+\phi_1(x)^2+\phi_2(x)^2-16 \\ (\phi_1(x)-2)^2+\phi_2(x)^2-9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Differentiating and applying the chain rule, this implies that

$$\begin{aligned} &\begin{bmatrix} 2x+2\phi_1(x)\phi_1'(x)+2\phi_2(x)\phi_2'(x) \\ 2(\phi_1(x)-2)\phi_1'(x)+2\phi_2(x)\phi_2'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \implies &\begin{bmatrix} 2x \\ 0 \end{bmatrix} + \begin{bmatrix} 2\phi_1(x) & 2\phi_2(x) \\ 2(\phi_1(x)-2) & 2\phi_2(x) \end{bmatrix} \begin{bmatrix} \phi_1'(x) \\ \phi_2'(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Thus, you can solve for $\phi'(x) = (\phi_1'(x), \phi_2'(x))$ provided the 2×2 matrix is invertible. Notice its determinant is given by

$$4(\phi_1(x)\phi_2(x) - (\phi_1(x)-2)\phi_2(x)) = 8\phi_2(x).$$

The determinant is non-zero and hence the matrix is invertible exactly when $\phi_2(x) \neq 0$. Note $\phi_2(a) = \frac{3\sqrt{15}}{4} \neq 0$ so $\phi_2(x) \neq 0$ for $x \in I$ where $I \subseteq \mathbb{R}$ is some open interval centred at $\frac{3\sqrt{15}}{4}$. Overall, this implies that for $x \in I$,

$$\begin{bmatrix} \phi_1'(x) \\ \phi_2'(x) \end{bmatrix} = \begin{bmatrix} 2\phi_1(x) & 2\phi_2(x) \\ 2(\phi_1(x)-2) & 2\phi_2(x) \end{bmatrix}^{-1} \begin{bmatrix} -2x \\ 0 \end{bmatrix} = \frac{1}{8\phi_2(x)} \begin{bmatrix} -4x\phi_2(x) \\ 4x\phi_1(x)-8x \end{bmatrix} = \begin{bmatrix} -x/2 \\ x(y-2)/2z \end{bmatrix}.$$

This example can be rewritten more concisely but less formally with variables.

Example 5.4.10 Assume the nonlinear equations

$$\begin{aligned} x^2+y^2+z^2 &= 16 \\ (y-2)^2+z^2 &= 9 \end{aligned}$$

locally define (y,z) as a C^1 function of x near $(a,b) = (0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$. Abusing notation, you may write $y = y(x)$ and $z = z(x)$. Taking partials of both equations with respect to x,

$$\begin{aligned} 2x + 2y\frac{\partial y}{\partial x} + 2z\frac{\partial z}{\partial x} &= 0 \\ 2(y-2)\frac{\partial y}{\partial x} + 2z\frac{\partial z}{\partial x} &= 0 \end{aligned}$$

This implies that

$$\begin{bmatrix}2x\\0\end{bmatrix}+\begin{bmatrix}2y & 2z\\2(y-2) & 2z\end{bmatrix}\begin{bmatrix}\frac{\partial y}{\partial x}\\\frac{\partial z}{\partial x}\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}.$$

The 2×2 determinant $4yz-4(y-2)z=8z$ is non-zero when $(x,y,z)=(0,\frac{11}{4},\frac{3\sqrt{15}}{4})$. This implies that

$$\begin{bmatrix}\frac{\partial y}{\partial x}\\\frac{\partial z}{\partial x}\end{bmatrix}=\begin{bmatrix}2y & 2z\\2(y-2) & 2z\end{bmatrix}^{-1}\begin{bmatrix}-2x\\0\end{bmatrix}=\begin{bmatrix}-x/2\\x(y-2)/2z\end{bmatrix}.$$

Implicit differentiation can be formally written as a lemma.

Lemma 5.4.11 Let $U\subseteq\mathbb{R}^n\times\mathbb{R}^k$ be an open set. Let $F:U\to\mathbb{R}^k$ be a $\mathbb{R}^k$-valued C^1 function. Let $(a,b)\in U$ so $a=(a_1,\dots,a_n)\in\mathbb{R}^n$ and $b=(b_1,\dots,b_k)\in\mathbb{R}^k$. Assume $F(a,b)=0$. If the equation

$$F(x_1,\dots,x_n,y_1,\dots,y_k)=0$$

locally defines y as a C^1 function $\phi:V\to W$ of x near (a,b), then for every $v\in V$ with $w=\phi(v)$, the Jacobian $D\phi(v)$ is a $k\times n$ matrix satisfying

$$\frac{\partial F}{\partial x}(v,w)+\frac{\partial F}{\partial y}(v,w)D\phi(v)=0.$$

Here $\dfrac{\partial F}{\partial x}=\dfrac{\partial(F_1,\dots,F_k)}{\partial(x_1,\dots,x_n)}:=\left[\dfrac{\partial F_i}{\partial x_j}\right]_{i,j}$ is a $k\times n$ matrix and $\dfrac{\partial F}{\partial y}=\dfrac{\partial(F_1,\dots,F_k)}{\partial(y_1,\dots,y_k)}:=\left[\dfrac{\partial F_i}{\partial y_j}\right]_{i,j}$ is a $k\times k$ matrix.

Remark 5.4.12 In particular, if the $k\times k$ matrix $\frac{\partial F}{\partial y}$ is invertible, then

$$D\phi(v)=-\Big(\frac{\partial F}{\partial y}(v,w)\Big)^{-1}\frac{\partial F}{\partial x}(v,w). \tag{5.4.2}$$

In Leibniz notation, this can be informally written as

$$D\phi=-\Big(\frac{\partial F}{\partial y}\Big)^{-1}\frac{\partial F}{\partial x}.$$

Note $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ are submatrices of the $k\times(n+k)$ Jacobian $DF=\left[\frac{\partial F}{\partial x}\;\middle|\;\frac{\partial F}{\partial y}\right]$.

The proof is another careful application of chain rule.

Proof. Note $V\subseteq\mathbb{R}^n$ and $W\subseteq\mathbb{R}^k$ are open sets containing a and b respectively. Define $G:V\to V\times W$ by $G(x)=(x,\phi(x))$. For $x\in V$, it follows that $F\circ G(x)=0$. Fix $v\in V$ and $w=\phi(v)$. The Jacobian of $F\circ G:V\to\mathbb{R}^k$ at v must be the $k\times n$ zero matrix $0_{k\times n}$, so

$$D(F\circ G)(v)=0_{k\times n}$$

From $G(v)=(v,\phi(v))=(v,w)$ and the chain rule, it follows that $D(F\circ G)(v)=DF(v,w)DG(v)$. By direct calculation, you can verify that this expression is equal to

$$\left[\frac{\partial F}{\partial x}(v,w)\;\middle|\;\frac{\partial F}{\partial y}(v,w)\right]\begin{bmatrix}I_{n\times n}\\ D\phi(v)\end{bmatrix}=\frac{\partial F}{\partial x}(v,w)+\frac{\partial F}{\partial y}(v,w)D\phi(v)$$

where $I_{n\times n}$ is the $n\times n$ identity matrix. ∎

Remark 5.4.12 provides the same vital insight: assuming local solutions exist, if $\frac{\partial F}{\partial y}$ is invertible, then you can calculate how the k variables $y = (y_1, \ldots, y_k)$ changes with the n variables $x = (x_1, \ldots, x_n)$. This spawns the full fledged version of the implicit function theorem in all its spectacular glory.

Theorem 5.4.13 (Implicit function theorem) Let $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be open. Let $F : U \to \mathbb{R}^k$ be a C^1 map. Let $(a, b) \in U$ so $a \in \mathbb{R}^n, b \in \mathbb{R}^k$. If $F(a, b) = 0$ and the $k \times k$ matrix

$$\frac{\partial F}{\partial y}(a,b) = \frac{\partial(F_1, \ldots, F_k)}{\partial(y_1, \ldots, y_k)}(a,b) := \left[\frac{\partial F_i}{\partial y_j}(a,b)\right]_{i,j}$$

is invertible, then the equation $F(x, y) = 0$ locally defines $y = (y_1, \ldots, y_k)$ as a $\mathbb{R}^k$-valued C^1 function ϕ of $x = (x_1, \ldots, x_n)$ near (a, b).

By (globally) solving a *linear* system of equations from a derivative condition, you can (locally) solve ***any*** *non-linear* system of equations given at least one solution. Incredible!

Example 5.4.14 Return to the nonlinear system $F(x, y, z) = (0, 0)$ of Example 5.4.9 where $F : \mathbb{R}^3 \to \mathbb{R}^2$ is the C^1 map given by

$$F(x,y,z) = \begin{bmatrix} x^2 + y^2 + z^2 - 16 \\ (y-2)^2 + z^2 - 9 \end{bmatrix}.$$

You can now prove that the system $F(x, y, z) = (0, 0)$ locally defines (y, z) as a C^1 function of x near $(0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$. Note $F(0, \frac{11}{4}, \frac{3\sqrt{15}}{4}) = (0, 0)$ and the 2×2 matrix

$$\frac{\partial(F_1, F_2)}{\partial(y, z)} = \begin{bmatrix} 2y & 2z \\ 2(y-2) & 2z \end{bmatrix}$$

has determinant $8z$, so it is invertible when $(x, y, z) = (0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$. By the implicit function theorem, $F(x, y, z) = (0, 0)$ locally defines (y, z) as C^1 function of x near $(0, \frac{11}{4}, \frac{3\sqrt{15}}{4})$.

The implicit function theorem (Theorem 5.4.13) has the same limitations as Theorem 5.4.6, but this does not detract from its accomplishments. Its impact is tremendous in multivariable calculus from any of the algebraic, analytic, and geometric viewpoints. The next section will describe a major consequence for identifying when sets written in implicit form are smooth manifolds and, if so, how to calculate their tangent spaces.

Exercises for Section 5.4

Concepts and definitions

5.4.1 Let $F(x,y,z) = x^3 + 3xyz + z^2y - 4y - 6z + 5$. Note $F(1,1,1) = 0$ and $\nabla F(1,1,1) = (6,0,-1)$.

(a) Does the equation $F(x,y,z) = 0$ locally define z as a C^1 function of (x,y) near $(1,1,1)$?

(b) Does the equation $F(x,y,z) = 0$ locally define y as a C^1 function of (x,z) near $(1,1,1)$?

(c) Does the equation $F(x,y,z) = 0$ locally define x as a C^1 function of (y,z) near $(1,1,1)$?

5.4.2 The implicit function theorem is spectacular because it reduces a non-linear problem to a linear one. Let $F : \mathbb{R}^4 \to \mathbb{R}^2$ be a nonlinear C^1 map. Let $p \in \mathbb{R}^4$ sastisfy $F(p) = (0,0)$. You compute

$$DF(p) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 12 \end{bmatrix} \quad \text{and} \quad \operatorname{rref}(DF(p)) = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

Since the map is not necessarily linear, you need the implicit function theorem to find nearby solutions to the *nonlinear* equation $F(x_1,x_2,x_3,x_4) = (0,0)$.

(a) Does this equation locally define (x_1,x_2) as a $\mathbb{R}^2$-valued C^1 function f of (x_3,x_4) near p?
Hint: Look at the 2×2 submatrix $\dfrac{\partial(F_1,F_2)}{\partial(x_1,x_2)}$

(b) Does this equation locally define (x_1,x_3) as a $\mathbb{R}^2$-valued C^1 function g of (x_2,x_4) near p?
Hint: Look at the 2×2 submatrix $\dfrac{\partial(F_1,F_2)}{\partial(x_1,x_3)}$

(c) Does this equation locally define (x_2,x_4) as a $\mathbb{R}^2$-valued C^1 function h of (x_1,x_3) near p?
Hint: Look at the 2×2 submatrix $\dfrac{\partial(F_1,F_2)}{\partial(x_2,x_4)}$

(d) If you had an explicit formula for F, for which of f, g, or h can you find an explicit formula?

5.4.3 The implicit function theorem is a mouthful to correctly state. Most mathematicians do not try to memorize the precise wording of such complicated theorems. Instead, you can rely on informal interpretations that possess the essential assumptions and conclusions. You will practice this skill by relating some helpful interpretations to the precise statement of the theorem.

(a) The implicit function theorem can be interpreted using intuition from linear algebra.

$$\textit{If you have } \underbrace{\boxed{k \textit{ non-linear equations}}}_{(A)} \textit{ between } \underbrace{\boxed{n+k \textit{ variables}}}_{(B)} \textit{ that are } \underbrace{\boxed{\textit{locally linearly independent}}}_{(C)}$$

$$\textit{then you have } \underbrace{\boxed{n = (n+k) - k \textit{ free variables}}}_{(D)} \textit{ and the } \underbrace{\boxed{\textit{remaining } k \textit{ variables}}}_{(E)} \textit{ depend}$$

on them.

For (A)–(E), give matching quantities or statements that correspond to the formal version.

(b) The theorem can also be interpreted with intuition about "generic" and "degrees of freedom".

$$\textit{Each of the } \underbrace{\boxed{n+k \textit{ variables}}}_{(A)} \textit{ is a degree of freedom. If } \underbrace{\boxed{\textit{the non-linear equations are generic}}}_{(B)},$$

each equation should cut down 1 degree of freedom. Since there are $\underbrace{\boxed{k \text{ non-linear equations}}}_{(C)}$,

this leaves $\underbrace{\boxed{n=(n+k)-k \text{ degrees of freedom}}}_{(D)}$.

For (A)–(D), give matching quantities or statements that correspond to the formal version.

5.4.4 Let $F:\mathbb{R}^4\to\mathbb{R}$ and $G:\mathbb{R}^4\to\mathbb{R}$ be real-valued maps. Fix a point $p\in\mathbb{R}^4$ such that $F(p)=G(p)=0$. Assume that

$$\nabla F(p)=(A_1,B_1,C_1,D_1) \qquad \nabla G(p)=(A_2,B_2,C_2,D_2)$$

for fixed constants $A_1,A_2,B_1,B_2,C_1,C_2,D_1,D_2\in\mathbb{R}$.

(a) What condition on the constants $A_1,B_1,C_1,D_1\in\mathbb{R}$ guarantees that you can locally solve

$$F(w,x,y,z)=0$$

for y in terms of w,x, and z?

(b) What condition on the constants $A_1,B_1,C_1,D_1\in\mathbb{R}$ guarantees that you can locally solve

$$F(w,x,y,z)=0$$

for some choice of 1 variable in terms of the other 3 variables?

(c) What condition on the constants $A_1,A_2,B_1,B_2,C_1,C_2,D_1,D_2\in\mathbb{R}$ guarantees that you can locally solve the system of equations

$$F(w,x,y,z)=0, \qquad G(w,x,y,z)=0$$

for y,z in terms of w,x?

(d) What condition on the constants $A_1,A_2,B_1,B_2,C_1,C_2,D_1,D_2\in\mathbb{R}$ guarantees that you can locally solve the system of equations

$$F(w,x,y,z)=0, \qquad G(w,x,y,z)=0$$

for some choice of 2 variables in terms of the other 2 variables?

Computations

5.4.5 Let $F:\mathbb{R}^5\to\mathbb{R}^3$ be a C^1 map. Fix a point $p\in\mathbb{R}^5$. You have calculated that

$$F(p)=\begin{bmatrix}0\\0\\0\end{bmatrix}, \quad DF(p)=\begin{bmatrix}1&1&0&1&4\\1&2&1&1&6\\0&1&1&1&3\end{bmatrix} \quad \text{and} \quad \operatorname{rref}(DF(p))=\begin{bmatrix}1&0&-1&0&1\\0&1&1&0&2\\0&0&0&1&1\end{bmatrix}.$$

Almost no calculations are needed for any part below.

(a) Does the nonlinear system $F(x)=0$ locally define (x_1,x_2,x_3) as a C^1 function of (x_4,x_5)?

(b) Does the nonlinear system $F(x)=0$ locally define (x_1,x_2,x_4) as a C^1 function of (x_3,x_5)?

(c) Does the nonlinear system $F(x)=0$ locally define (x_2,x_3,x_4) as a C^1 function of (x_1,x_5)?

5.4.6 Let $F(x,y)=x^2-y^2$. Notice $F(2,2)=0$.

(a) Show that the equation $F(x,y)=0$ locally defines y as a C^1 function ϕ of x near $(2,2)$.

(b) Express $\phi'(x)$ in terms of x and $\phi(x)$. Remember to specify the domain of ϕ.

5.4.7 Define $F(x,y)=x^2+y^2-2\cos(\pi(x+y))$ and notice $F(-1,-1)=0$.

(a) Show that $F(x,y)=0$ defines y locally as a C^1 function $\phi(x)$ near $(-1,-1)$.

(b) Calculate $\phi'(x)$. Remember to specify the domain of ϕ.

5.4.8 Define $F(x,y,z)=e^{2xyz}-(\arcsin x)/z^2-1+\frac{\pi}{6}$ and notice $F(1/2,0,1)=0$.

(a) Show that $F(x,y,z)=0$ defines z locally as a C^1 function $\phi(x,y)$ near $(1/2,0,1)$.

(b) Calculate $\nabla\phi(x,y)$. Remember to specify the domain of ϕ.

5.4.9 Fix $p=(1,0,1,-1,1)\in\mathbb{R}^5$. Let $F:\mathbb{R}^5\to\mathbb{R}^3$ be given by

$$F(v,w,x,y,z)=\begin{bmatrix} vw^2+vy+x^2z^2-yz^3-1 \\ wxy^2+w^2x^2-xy^2z+1 \\ vxy-w^3z+1 \end{bmatrix}.$$

(a) Show that $F(v,w,x,y,z)=(0,0,0)$ defines (w,y,z) implicitly as a C^1 map ϕ of (v,x) near p.

(b) Compute the Jacobian $D\phi(v,w)$. Remember to specify the domain of ϕ.

5.4.10 Let $F:\mathbb{R}^4\to\mathbb{R}^2$ be given by

$$F(w,x,y,z)=\begin{bmatrix} 3w^2yz-x^3+2y^2+z-3 \\ w^3-x^2y^2+3yz^4-27 \end{bmatrix}.$$

(a) Show that $F(w,x,y,z)=(0,0)$ defines (x,y) implicitly as a C^1 map ϕ of (w,z) near $(3,-1,0,2)$.

(b) Compute the Jacobian $D\phi(w,z)$. Remember to specify the domain of ϕ.

Proofs

5.4.11 The implicit function theorem (IFT) is amazing but it has limitations. Here are several flawed arguments using the implicit function theorem beyond what it permits.

(a) Adele is singing songs about the IFT, but makes a mistake during one of her proofs.

> 1. *Define $f(x,y)=x^2+y^2-1$ and note $f(0,1)=0$.*
> 2. *Note $\partial_1 f(x,y)=2x$ so $\partial_1 f(0,1)=0$.*
> 3. *By the implicit function theorem, the equation $f(x,y)=0$ does not locally define x as a C^1 function of y near $(0,1)$.*

Identify the flaw and briefly explain.

(b) After dancing with excitement over the IFT, Beyoncé accidentally writes a flawed argument.

> 1. *Consider $F(x,y,u,v)=\begin{bmatrix}0\\0\end{bmatrix}$ where $F(x,y,u,v)=\begin{bmatrix} x^2-y^2-u^3+v^2+4 \\ 2xy+y^2-2u^2+3v^4+8\end{bmatrix}$.*
> 2. *Notice $\dfrac{\partial(F_1,F_2)}{\partial(u,v)}=\begin{bmatrix}-3u^2 & 2v \\ -4u & 12v^3\end{bmatrix}$ so $\dfrac{\partial(F_1,F_2)}{\partial(u,v)}(2,0,1,-1)=\begin{bmatrix}-3 & -2\\ -4 & -12\end{bmatrix}$ is invertible.*
> 3. *By the implicit function theorem, the equation $F(x,y,u,v)=(0,0)$ locally defines (u,v) as a C^1 function of (x,y) near $(2,0,1,-1)$.*

Identify the flaw and briefly explain.

(c) Cardi is writing poetry about the IFT and goes a bit too fast.

> 1. *The equation $x^2 + y^2 = 2$ defines a circle of radius $\sqrt{2}$ and $(1,1)$ is on this circle.*
> 2. *Define $F(x,y) = x^2 + y^2 - 2$ which is C^1 on $\mathbb{R}^2$.*
> 3. *Since $\frac{\partial F}{\partial y}(x,y) = 2y$, notice $\frac{\partial F}{\partial y}(1,1) = 2 \neq 0$.*
> 4. *By the implicit function theorem, $y = \sqrt{2-x^2}$ for $-\sqrt{2} < x < \sqrt{2}$.*
> 5. *Thus, the equation $F(x,y) = 0$ defines y locally as a C^1 function of x near $(1,1)$.*

She has made several errors on 1 line. Identify the line and explain its flaws.

5.4.12 Prove that the equation $x^2 + y^2 + z^2 = 1$ is locally defines 1 variable as a C^1 function of the other 2 variables at the north pole $(0,0,1)$. State which variable can be expressed in terms of the others.

5.4.13 Show that the equation $x^2 + y^2 = 1$ locally defines (at least) 1 variable in terms of the other 1 variable near every point.

5.4.14 Prove that the equation $x^2 + y^2 + z^2 = 1$ locally defines (at least) 1 variable in terms of the other 2 variables near every point.

5.4.15 Show that $S^{n-1} = \{x \in \mathbb{R}^n : ||x|| = 1\}$ is locally the graph of a C^1 function at every point.

Applications and beyond

5.4.16 The implicit function theorem (IFT) for real-valued maps answers natural questions. You will explore such questions from each of the four viewpoints and how the implicit function theorem applies. Assume the functions f, g, h are $\mathbb{R}$-valued and C^1 everywhere.

(a) From an **algebraic** viewpoint, you might ask:

> *Does the non-linear equation $f(x_1, \ldots, x_n, y) = 0$ implicitly define the variable y as a function of the remaining variables $x_1, \ldots, x_n$?*

Briefly explain how the IFT addresses this question. Include extra assumptions, if necessary.

(b) From an **analytic** viewpoint, you might ask:

> *Given a solution $(x_1, \ldots, x_n) = (a_1, \ldots, a_n)$ to a non-linear equation $g(x_1, \ldots, x_n) = 0$, can you find other solutions nearby?*

Briefly explain how the IFT addresses this question. Include extra assumptions, if necessary.

(c) From a **physical** viewpoint, you might ask:

> *Under certain physical conditions, the pressure $x_1 \in \mathbb{R}$ and volume $x_2 \in \mathbb{R}$ of a gas are implicitly related to the temperature y of the gas by some equation $h(x_1, x_2, y) = 0$. How does the temperature change with respect to the volume if you hold pressure fixed?*

Briefly explain how the IFT applies to this question. Include extra assumptions, if necessary.

(d) Now, repeat these questions for $\mathbb{R}^k$-valued maps. Assume F, G, H, are $\mathbb{R}^k$-valued and C^1. From an **algebraic** viewpoint, you might ask:

> *Does the non-linear system of k equations $F(x_1, \dots, x_n, y_1, \dots, y_k) = 0$ implicitly define the variables $y = (y_1, \dots, y_k)$ as a function of the remaining variables $x = (x_1, \dots, x_n)$?*

Briefly explain how the IFT addresses this question. Include extra assumptions, if necessary.

(e) From an **analytic** viewpoint, you might ask:

> *Given a single solution $(x_1, \dots, x_n) = (a_1, \dots, a_n)$ to the non-linear system of k equations $H(x_1, \dots, x_n) = 0$, can you find other solutions nearby?*

Briefly explain how the IFT addresses this question. Include extra assumptions, if necessary.

5.4.17 Prove the implicit function theorem for $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ follows from the inverse function theorem. *Hint:* Define $G : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n \times \mathbb{R}^k$ by $G(x, y) = (x, F(x, y))$.

5.5. Smooth manifolds and implicit form

Thus far, the implicit function theorem has been described from the algebraic and analytic perspectives. It also has deep geometric consequences to sets written implicit form, namely sets

$$S = \{x \in \mathbb{R}^n : F(x) = 0\}$$

where $F : \mathbb{R}^n \to \mathbb{R}^k$ is a C^1 map with $k < n$. In this section, you will revisit a core question that has lingered from your study of smooth manifolds (Section 4.6).

How can you determine if a set $S \subseteq \mathbb{R}^n$ written in implicit form is a smooth manifold?

Once you can identify smoothness, the tangent planes on S are translations of a subspace of $\mathbb{R}^n$ (see Corollary 4.6.19); that is, they really are "planes" in the linear algebra sense. This puts you in the perfect position to return to a magical observation about the gradient (Section 3.4.3).

Is the gradient orthogonal to the tangent plane of a level set?

You will finally prove powerful theorems answering these questions and generalize this property of the gradient to any set in implicit form. The implicit function theorem will be your hammer.

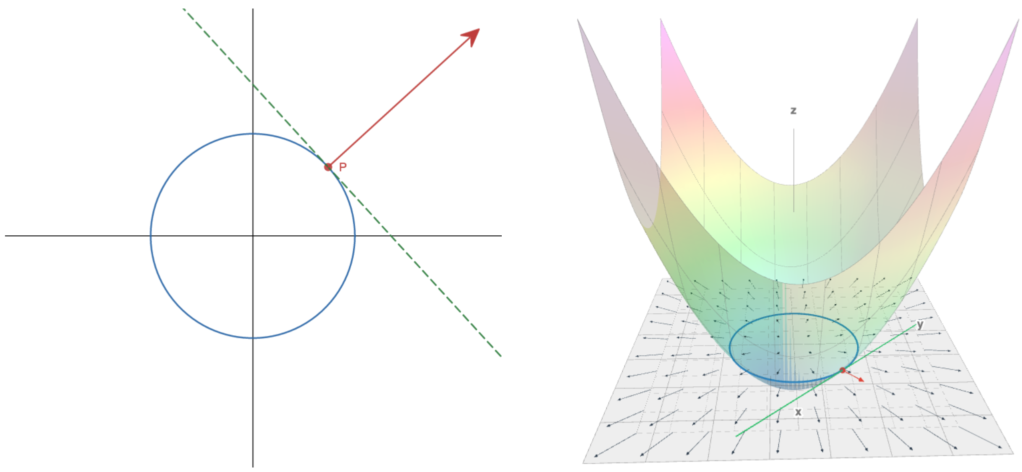
5.5.1 Level sets and gradients

Recall a level set of a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is a set written in implicit form, namely

$$S = \{x \in \mathbb{R}^n : f(x) = 0\} = f^{-1}(\{0\}).$$

Notice S is defined by 1 nonlinear equation $f(x_1, \dots, x_n) = 0$ with n variables $x_1, \dots, x_n$. This suggests S should be an $(n-1)$-dimensional smooth manifold, but how can you detect this feature? The gradient of f demonstrates the key insight

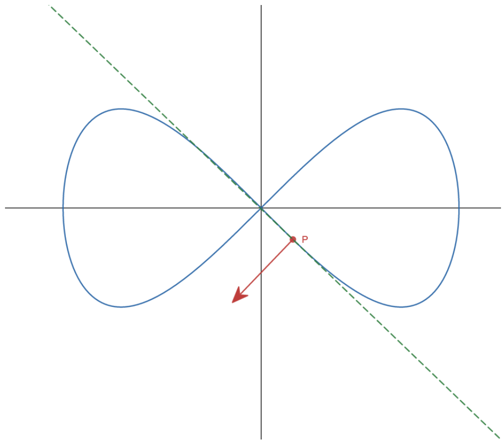
Example 5.5.1 Let $f(x,y) = x^2 + y^2 - 1$ so the set $S = \{(x,y) \in \mathbb{R}^2 : f(x,y) = 0\}$ is the 0-level set of f. Is this a 1-dimensional smooth manifold? Intuitively speaking, for any $p = (a,b) \in S$, the tangent plane $p + T_pS$ is a line and the gradient $\nabla f(a,b) = (2a, 2b)$ appears orthogonal to this line. The lefthand Desmos graph illustrates this phenomenon.

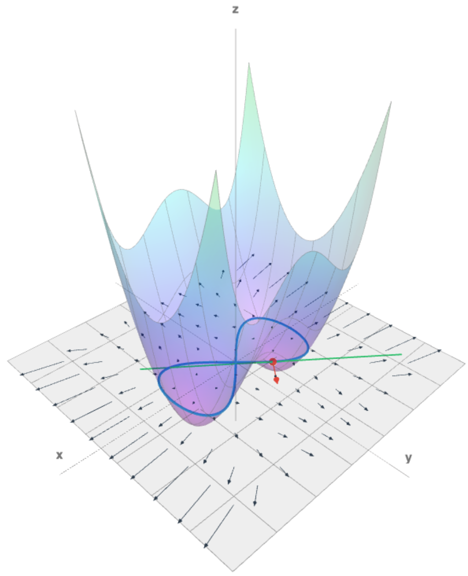


Hence, the gradient $\nabla f(a,b)$ actually defines the tangent line of the unit circle, a 1-dimensional smooth manifold. The righthand Math3D demo displays the 2-dimensional graph of the paraboloid $z = f(x,y) = x^2 + y^2 - 1$, the 1-dimensional level set, and the gradient all together.

The tangent line may not always be defined for every implicit curve.

Example 5.5.2 Let $g(x,y) = x^4 - x^2 + y^2$ so the set $S = \{(x,y) \in \mathbb{R}^2 : g(x,y) = 0\}$ is the 0-level set of g. Is it a 1-dimensional smooth manifold? Intuitively speaking, for any $p = (a,b) \in S$, the tangent plane $p + T_pS$ is again a line and the gradient $\nabla g(a,b) = (4a^3 - 2a, 2b)$ appears orthogonal to this line, *except* when $(a,b) = (0,0)$. Play with the lefthand Desmos graph to view this issue. What is happening at the origin?





Notice the level set S is *not* a 1-dimensional smooth manifold at $(0,0)$, but this occurs precisely when the gradient vanishes! That is, $\nabla g(0,0) = (0,0)$ and this happens exactly when the tangent space $T_{(0,0)}S$ is not a subspace of $\mathbb{R}^2$ (see also Example 4.6.1). View the righthand Math3D demo and notice that the graph $z = g(x,y)$ has a saddle point at $(0,0)$.

Algebraically speaking, the gradient ∇g vanishes at the origin so you cannot use the implicit function theorem to locally solve the equation $g(x,y) = 0$ for (x,y) near $(0,0)$. If $\partial_1 g(0,0) = 0$ then you cannot use the implicit function theorem (Theorem 5.4.6) to show x is locally a C^1 function of y near $(0,0)$. Similarly, if $\partial_2 g(0,0) = 0$ then you cannot use the theorem to show y is locally a C^1 function of x near $(0,0)$. These are the only two options. This does not prove anything about S at the origin, but this example does demonstrate that, as long as $\nabla g(a,b) \neq 0$ and $(a,b) \in S$, the level set S should be a smooth manifold at (a,b) and the tangent space $T_{(a,b)}S$ should be a 1-dimensional subspace of $\mathbb{R}^2$.

These empirical observations are confirmed by a wonderfully beautiful theorem.

Theorem 5.5.3 Let $U \subseteq \mathbb{R}^n$ be an open set. Let $f : U \to \mathbb{R}$ be a C^1 function. Assume the set

$$S = \{x \in U : f(x) = 0\} = f^{-1}(\{0\})$$

is non-empty. Fix $p \in S$. If $\nabla f(p) \neq 0$, then S is a $(n-1)$-dimensional smooth manifold at p. Moreover, a vector $v \in \mathbb{R}^n$ is a tangent vector of S at p if and only if $\nabla f(p) \cdot v = 0$. That is,

$$T_pS = \{v \in \mathbb{R}^n : \nabla f(p) \cdot v = 0\}.$$

Therefore, the tangent plane of S at p is given by

$$p + T_pS = \{x \in \mathbb{R}^n : \nabla f(p) \cdot (x - p) = 0\}.$$

This theorem simultaneously answers both of your longstanding questions.

> *A level set in $\mathbb{R}^n$ is a $(n-1)$-dimensional smooth manifold wherever the gradient does not vanish, and the gradient is orthogonal to the $(n-1)$-dimensional tangent plane.*

Its proof is a stunning display of the implicit function theorem, implicit differentiation (or the chain rule), the local property of tangent spaces, and tangent spaces of graphs.

Proof. Since $\nabla f(p) \neq 0$, assume without loss of generality that $\partial_n f(p) \neq 0$. Since f is C^1, $f(p) = 0$, and $\partial_n f(p) \neq 0$, the implicit function theorem (Theorem 5.4.6) implies that the equation $f(x_1, \ldots, x_n) = 0$ locally defines x_n as a C^1 function of $x_1, \ldots, x_{n-1}$ near p. That is, there exists open sets $V \subseteq \mathbb{R}^{n-1}$ and $W \subseteq \mathbb{R}$ containing $(p_1, \ldots, p_{n-1}) \in \mathbb{R}^{n-1}$ and $p_n \in \mathbb{R}$ respectively and a C^1 function $\phi : V \to W$ such that

$$\forall (x_1, \ldots, x_{n-1}, x_n) \in V \times W, \quad f(x_1, \ldots, x_n) = 0 \iff x_n = \phi(x_1, \ldots, x_{n-1}).$$

Since $V \times W$ is an open set containing p, the set $S \cap (V \times W)$ is the graph of $\phi : V \to W$, so

$$S \cap (V \times W) = \{(x_1, \ldots, x_{n-1}, x_n) \in \mathbb{R}^n : x_n = \phi(x_1, \ldots, x_{n-1})\}. \tag{5.5.1}$$

Hence, S is an $(n-1)$-dimensional smooth manifold by Definition 4.6.4.

Continuing with this setup, it remains to show that $v \in T_p S$ if and only if $\nabla f(p) \cdot v = 0$. Theorem 4.6.18 shows that S and $S' = S \cap (V \times W)$ have the same tangent space at p, because $p \in S \cap S'$ and $V \times W$ is an open set containing p. Set $a = (p_1, \ldots, p_{n-1}) \in \mathbb{R}^{n-1}$ so $(a, \phi(a)) = p$. Since S' is the graph of ϕ given by (5.5.1), it follow by Theorem 4.5.11 that

$$T_p S = T_p S' = \{(w, d\phi_a(w)) : w \in \mathbb{R}^{n-1}\}.$$

From Theorem 3.5.22, we have for $w \in \mathbb{R}^{n-1}$ that

$$(w, d\phi_a(w)) = (w, \nabla \phi(a) \cdot w) = \big(w_1, \ldots, w_{n-1}, \partial_1 \phi(a) w_1 + \cdots + \partial_{n-1} \phi(a) w_{n-1}\big)$$

Partials of ϕ and f are related by implicit differentiation. From Lemma 5.4.4, for $1 \leq j \leq n-1$,

$$\begin{aligned}
&\frac{\partial f}{\partial x_j}(a, \phi(a)) + \frac{\partial f}{\partial x_n}(a, \phi(a)) \frac{\partial \phi}{\partial x_j}(a) = 0 \\
\implies{}& \partial_j f(p) + \partial_n f(p) \partial_j \phi(a) = 0 \\
\implies{}& \partial_j \phi(a) = -\frac{\partial_j f(p)}{\partial_n f(p)}
\end{aligned}$$

as $(a, \phi(a)) = p$ and $\partial_n f(p) \neq 0$. Combining these observations, it follows that $v \in T_p S$ if and only if there exists $w \in \mathbb{R}^{n-1}$ such that

$$\begin{aligned}
v = (w, d\phi_a(w)) &\iff (v_1, \ldots, v_{n-1}, v_n) = (w_1, \ldots, w_{n-1}, \partial_1 \phi(a) w_1 + \cdots + \partial_{n-1} \phi(a) w_{n-1}) \\
&\iff v_n = \partial_1 \phi(a) v_1 + \cdots + \partial_{n-1} \phi(a) v_{n-1} \\
&\iff v_n = -\frac{\partial_1 f(p)}{\partial_n f(p)} v_1 - \cdots - \frac{\partial_{n-1} f(p)}{\partial_n f(p)} v_{n-1} \\
&\iff \partial_1 f(p) v_1 + \cdots + \partial_{n-1} f(p) v_{n-1} + \partial_n f(p) v_n = 0.
\end{aligned}$$

Thus, $v \in T_p S$ if and only if $\nabla f(p) \cdot v = 0$ as required. The expression with the tangent space $T_p S$ and the tangent plane $p + T_p S$ follow quickly from the definitions. ■

Theorem 5.5.3 is an outstanding accomplishment of multivariable calculus. It solidifies your geometric intuition with a rigorous foundation. It is also brilliant for computing.

Example 5.5.4 The set $S = \{x \in \mathbb{R}^{2k} : x_1 + x_2^2 + x_3^3 + \cdots + x_{2k}^{2k} = 0\}$ is the 0-level set of the C^1 function
$$f(x) = x_1 + x_2^2 + x_3^3 + \cdots + x_{2k}^{2k}.$$
Notice $p = (-1, 1, \ldots, -1, 1) \in S$ since $f(p) = -1 + 1 - 1 + \cdots - 1 + 1 = 0$ as $2k$ is even. Observe
$$\nabla f(x) = (1, 2x_2, 3x_3^2, \ldots, 2kx_{2k}^{2k-1})$$
so $\nabla f(p) = (1, 2, 3, \ldots, 2k-1, 2k)$. Hence, for $x \in \mathbb{R}^{2k}$,
$$\begin{aligned}\nabla f(p) \cdot (x - p) &= (1, 2, 3, \ldots, 2k-1, 2k) \cdot (x_1 + 1, x_2 - 1, \ldots, x_{2k-1} + 1, x_{2k} - 1) \\ &= (x_1 + 1) + 2(x_2 - 1) + 3(x_3 + 1) + \cdots + 2k(x_{2k} - 1).\end{aligned}$$
By Theorem 5.5.3, the tangent plane of S at p is the $(2k-1)$-dimensional plane given by
$$\{x \in \mathbb{R}^{2k} : (x_1 + 1) + 2(x_2 - 1) + 3(x_3 + 1) + \cdots + 2k(x_{2k} - 1) = 0\}.$$

Note Theorem 5.5.3 cannot be used to prove a set is not a smooth manifold.

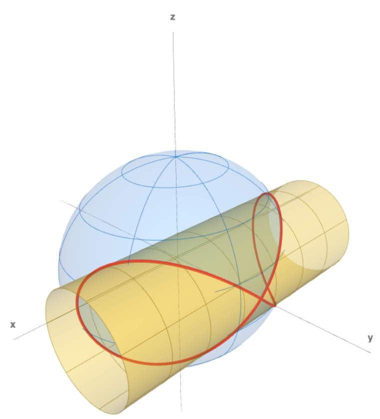
Example 5.5.5 The 0-level set of the function $g(x, y) = x^4 - x^2 + y^2$ is the figure eight curve. The gradient $\nabla g(x, y) = (4x^3 - 2x, 2y)$ vanishes at $(0, 0)$ but Theorem 5.5.3 does not apply in this situation. You must prove the set is not a smooth manifold at $(0, 0)$ by definition.

On the other hand, the 0-level set of the function $h(x, y) = (y - x^2)^2$ is the parabola $y = x^2$, which is a graph everywhere. However, the gradient $\nabla h(x, y) = (4x(y - x^2), 2(y - x^2))$ vanishes on the entire level set! This shows that the converse of Theorem 5.5.3 does not hold. In other words, the gradient may vanish even if the set is a smooth manifold.

5.5.2 Implicit form and kernels

This completes the study of tangent planes for sets written in implicit form defined by a single nonlinear equation. You proved that the gradient is orthogonal to a point on a level set, provided that the function was C^1 and the gradient did not vanish at the point. This theorem generalizes to any set written in implicit form defined by many nonlinear equations, but the precise statement requires a bit more intuition from linear algebra and geometry.

Example 5.5.6 The single equation $x^2 + y^2 + z^2 = 16$ is a sphere (a 2-dimensional smooth manifold) and the single equation $(y - 2)^2 + z^2 = 4$ is a cylinder (a 2-dimensional smooth manifold). Solutions to both nonlinear equations correspond to the *intersection* of these two 2-dimensional smooth manifolds. Play with this Math3D demo and toggle the manifolds.



The intersection appears to be a 1-dimensional smooth manifold, except possibly at the crossing point $(0,4,0)$. This phenomenon corresponds to how 2-dimensional planes in $\mathbb{R}^3$ usually intersect to form a 1-dimensional line. This geometric connection with linear algebra suggests there may be a deep theorem lurking in the background.

First, you can prove that the sphere and the cylinder are 2-dimensional smooth manifolds with Theorem 5.5.3. Define the C^1 real-valued functions

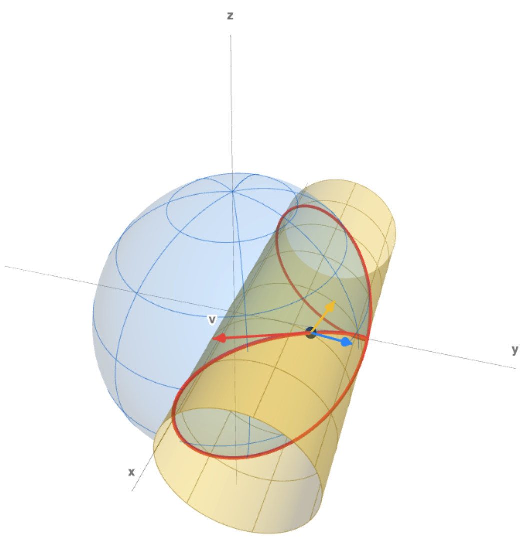
$$f(x,y,z) = x^2 + y^2 + z^2 - 16, \qquad g(x,y,z) = (y-2)^2 + z^2 - 4$$

so $\nabla f(x,y,z) = (2x, 2y, 2z)$ and $\nabla g(x,y,z) = (0, 2y-4, 2z)$. This implies ∇f vanishes only at $(0,0,0)$, but this does not lie on the sphere, i.e. it does not satisfy $f(x,y,z) = 0$. Similarly, ∇g vanishes only at $(0,2,0)$, but this does not lie on the cylinder, i.e. it does not satisfy $g(x,y,z) = 0$. Hence, Theorem 5.5.3 implies that

$$S_1 = \{(x,y,z) \in \mathbb{R}^3 : f(x,y,z) = 0\} \qquad S_2 = \{(x,y,z) \in \mathbb{R}^3 : g(x,y,z) = 0\}$$

are both 2-dimensional smooth manifolds in $\mathbb{R}^3$. What about their intersection $S_1 \cap S_2$?

Fix a point $p \in S_1 \cap S_2$. Since p lies on the sphere, $\nabla f(p)$ is orthogonal to the tangent plane of the sphere by Theorem 5.5.3. Since p lies on the cylinder, $\nabla g(p)$ is orthogonal to the tangent plane of the cylinder by Theorem 5.5.3. This suggests that any tangent vector $v \in \mathbb{R}^3$ of $S_1 \cap S_2$ at p must be orthogonal to *both* of $\nabla f(p)$ and $\nabla g(p)$. Play with the Math3D demo for a visual.



Algebraically speaking, how can you interpret this observation for $S = S_1 \cap S_2$? Combine the nonlinear equations. Define $F : \mathbb{R}^3 \to \mathbb{R}^2$ by $F = (f, g)$ so

$$F(x,y,z) = (x^2 + y^2 + z^2 - 16, (y-2)^2 + z^2 - 4)$$

and hence $S = \{(x,y,z) \in \mathbb{R}^3 : F(x,y,z) = (0,0)\}$ by definition of intersection. Notice that

$$DF(p) = \begin{bmatrix} \partial_1 f(p) & \partial_2 f(p) & \partial_3 f(p) \\ \partial_1 g(p) & \partial_2 g(p) & \partial_3 g(p) \end{bmatrix} = \begin{bmatrix} - & \nabla f(p)^T & - \\ - & \nabla g(p)^T & - \end{bmatrix}. \tag{5.5.2}$$

Hence for $v \in \mathbb{R}^3$, the simultaneous conditions $\nabla f(p) \cdot v = 0 \in \mathbb{R}$ and $\nabla g(p) \cdot v = 0 \in \mathbb{R}$ is equivalent to $DF(p)v = 0 \in \mathbb{R}^2$. Equivalently, $dF_p(v) = 0 \in \mathbb{R}^2$ so $v \in \ker dF_p$.

Finally, how can you detect when the intersection S is a 1-dimensional smooth manifold at a point? The insight comes from looking at the only point where S appears to be *not* a 1-dimensional smooth manifold, namely at $(0,4,0) \in S$. Notice that

$$\nabla f(0,4,0) = (0,8,0), \qquad \nabla g(0,4,0) = (0,4,0).$$

so the vectors $\nabla f(0,4,0)$ and $\nabla g(0,4,0)$ are *not* linearly independent. You can confirm this with earlier Math3D demo. Otherwise, at all other points $p \in S$, it appears that $\{\nabla f(p), \nabla g(p)\}$ is linearly independent, so the 2 conditions $\nabla f(p) \cdot v = 0$ and $\nabla g(p) \cdot v = 0$ force $v \in \mathbb{R}^3$ to lie in a subspace of dimension $3 - 2 = 1$. From (5.5.2), you can reformulate this condition:

If $DF(p)$ has rank 2, then S is a 1-dimensional smooth manifold at p.

Since $DF(p)$ is a 2×3 matrix, the rank 2 condition is equivalent to saying the Jacobian matrix $DF(p)$ has full rank (or the differential dF_p has full rank).

This extended example leads to the higher dimensional generalization of Theorem 5.5.3.

Theorem 5.5.7 Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let $U \subseteq \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^k$ be a C^1 map. Assume the set $S = F^{-1}(\{0\})$ is non-empty. Fix $p \in S$. If the differential dF_p has full rank (that is, $\operatorname{rank} DF(p) = k$), then S is a $(n-k)$-dimensional smooth manifold at p. Moreover, the tangent space to S at p is a $(n-k)$-dimensional subspace of $\mathbb{R}^n$ given by

$$T_pS = \ker dF_p = \{v \in \mathbb{R}^n : DF(p)v = 0\}.$$

Remark 5.5.8 Using some linear algebra and definition of the Jacobian, you can prove that the rank of the matrix $DF(p) = k$ occurs if and only if the gradient vectors of the components $\nabla F_1(p), \ldots, \nabla F_k(p)$ are k linearly independent vectors in $\mathbb{R}^n$.

Proof. The argument is very similar to Theorem 5.5.3. You must apply the more general implicit function theorem (Theorem 5.4.13 instead of Theorem 5.4.6) and the more general implicit differentiation (Lemma 5.4.11 instead of Lemma 5.4.4). The linear algebra gets a bit more complicated since you must manipulate matrices rather then vectors. Otherwise the proof is almost identical, so this is left as a challenging exercise. ■

Now, you can apply this theorem in examples to verify when a set written in implicit form is a smooth manifold.

Example 5.5.9 Revisit $S = \{(x,y,z) \in \mathbb{R}^3 : F(x,y,z) = (0,0)\}$ in Example 5.5.6, where

$$F(x,y,z) = (x^2 + y^2 + z^2 - 16, (y-2)^2 + z^2 - 4).$$

By direct calculation, the C^1 map F satisfies

$$DF(x,y,z) = \begin{bmatrix} 2x & 2y & 2z \\ 0 & 2y-4 & 2z \end{bmatrix}.$$

For which $(x,y,z) \in S$ does the 3×2 matrix $DF(x,y,z)$ have rank 2? This is the critical assumption in Theorem 5.5.7. Notice that $DF(x,y,z)$ has rank 2 if a pair of columns form an invertible 2×2 matrix. The first and second column form a 2×2 matrix with determinant

$4x(y-2)$, so this is not invertible whenever $x(y-2)=0$. The first and third column form a 2×2 matrix with determinant $4xz$ so this is not invertible whenever $xz=0$. The second and third column for a 2×2 matrix with determinant $4yz-(2y-4)2z=8z$ so this is not invertible whenever $z=0$.

Overall, these calculations imply that $DF(x,y,z)$ has rank 2 unless

$$x(y-2)=0, \quad xz=0, \quad \text{and} \quad z=0.$$

This forces either $x=z=0$ or $y-2=z=0$. If $x=z=0$, then

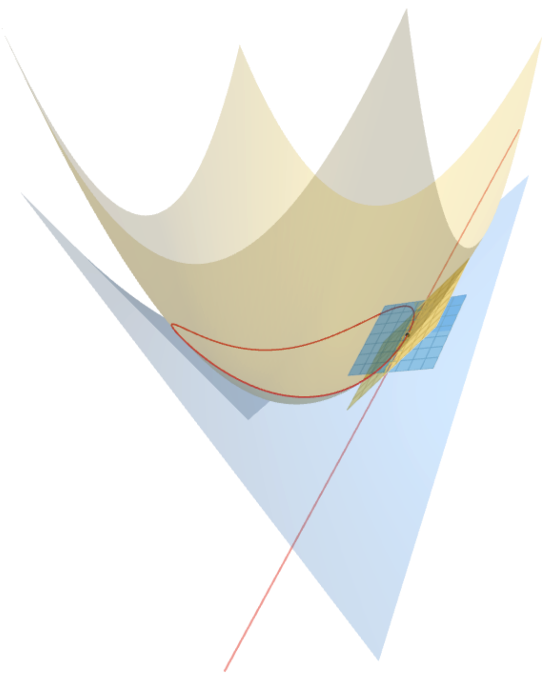
$$F(0,y,0)=(0,0) \iff y^2=16, \quad (y-2)^2=4 \iff y=4,$$

so $(0,4,0)\in S$ and $DF(0,4,0)$ has rank <2. In fact, $DF(0,4,0)$ has rank 1. Otherwise, if $y-2=z=0$ then

$$F(x,2,0)=(0,0) \iff x^2+4=16, \quad -4=0$$

and this has no solutions. Thus, by Theorem 5.5.7, S is a 1-dimensional smooth manifold at every point except possibly $(0,4,0)$. Note Theorem 5.5.7 cannot conclude anything at the point $(0,4,0)$.

This concludes some major achievements in the study of manifolds and sets written in implicit form. You have reduced the verification of smoothness and calculation of tangent planes to routine linear algebra computations! The geometry of intersecting smooth manifolds also matches your intuition from linear algebra with intersecting planes, which you can view in this Math3D demo. These consequences demonstrate the spectacular power of the implicit function theorem.



As a big reward, you will develop a brand new optimization technique using this smooth manifold theory for sets in implicit form and their tangent planes.

Exercises for Section 5.5

Concepts and definitions

5.5.1 An $m \times n$ matrix A has **full rank** if $\operatorname{rank} A = \min\{m, n\}$. Which matrices are full rank?

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & 3 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$

5.5.2 Theorem 5.5.7 is quite powerful as it consider any set defined in implicit form.

(a) Fill in the blank.

Let S be a level set of a C^1 function $F : \mathbb{R}^{237} \to \mathbb{R}^{137}$. Fix $p \in S$.

If $\operatorname{rank} dF_p =$__________, *then T_pS is a subspace of $\mathbb{R}^{237}$ of dimension* __________.

(b) Keep the notation of Theorem 5.5.7. The core assumption is that the differential dF_p has full rank. Which of the following are equivalent to this assumption?

i) The matrix $DF(p)$ has rank k.
ii) The linear map dF_p is injective.
iii) The linear map dF_p is surjective.
iv) The vectors $\{\nabla F_1(p), \dots, \nabla F_k(p)\}$ are linearly independent.
v) The vectors $\{\partial_1 F(p), \dots, \partial_k F(p)\}$ are linearly independent.

5.5.3 Which statements are true or false? If true, briefly explain. If false, state a counterexample.

(a) Every set written in explicit form can also be written in implicit form.

(b) Every set written in implicit form can also be written in explicit form.

(c) Every set written in explicit form using a C^1 function is a smooth manifold.

(d) Every set written in implicit form using a C^1 function is a smooth manifold.

5.5.4 Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^1. Assume $S = \{x \in \mathbb{R}^n : f(x) = 0\}$ is non-empty.
Fix $p \in S$ and $v \in \mathbb{R}^n$. Which statements are true or false?

(a) The set $\{x \in \mathbb{R}^n : \nabla f(p) \cdot (x - p) = 0\}$ is the tangent plane of S at p.

(b) If $\nabla f(p) \neq 0$, then the tangent space T_pS is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$.

(c) If v is a tangent vector of S at p, then $\nabla f(p) \cdot v = 0$.

(d) If $\nabla f(p) \cdot v = 0$, then v is a tangent vector of S at p.

5.5.5 Let $F : \mathbb{R}^5 \to \mathbb{R}^3$ be a C^1 map. Fix a point $p \in \mathbb{R}^5$. You have calculated that

$$F(p) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad DF(p) = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 2 & 3 & 3 & 0 & 2 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \operatorname{rref}(DF(p)) = \begin{bmatrix} 1 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -2 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Is the set $S = F^{-1}(\{(0,0,0)\})$ a smooth manifold at p? If so, of what dimension?

Computations

5.5.6 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be C^1. Define $S = f^{-1}(\{0\})$ and assume

$$f(2,3,7) = 0 \qquad \nabla f(2,3,7) = (-1,0,4).$$

(a) Is S a 2-dimensional smooth manifold at $(2,3,7)$?

(b) Compute T_pS. You may express it as a set written in implicit form with 1 linear equation.

(c) Compute the tangent plane of S at p.

5.5.7 Let $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 14\}$ be the sphere of radius $\sqrt{14}$ centred at the origin.

(a) Show that S is a smooth surface.

(b) Compute the tangent plane of S at $(1,2,3)$.

5.5.8 Let $F : \mathbb{R}^4 \to \mathbb{R}^2$ be a C^1 map. Let $p \in \mathbb{R}^4$ be such that $F(p) = (0,0)$. You compute

$$DF(p) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \quad \text{and} \quad \operatorname{rref}(DF(p)) = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

Define the level set $S = \{x \in \mathbb{R}^4 : F(x) = 0\} = F^{-1}(\{0\})$.

(a) Is the set S a smooth manifold at p? If so, what is its dimension?

(b) If possible, express the tangent space T_pS as the span of 2 vectors.

(c) Express the tangent plane $p + T_pS$ using set builder notation.

5.5.9 Let $S = \{(w,x,y,z) \in \mathbb{R}^4 : xz + w = 0 \text{ and } x^2 + wy + z^2 = 1\}$ be a subset of $\mathbb{R}^4$.

(a) Express S as a set written in implicit form using a single C^1 function.

(b) Compute the differential of your function at $p = (1,0,0,1) \in S$.

(c) Show S is a smooth manifold at p and determine its dimension.

(d) Give a basis for the tangent space T_pS.

5.5.10 Let $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + 2y^3 - 3xyz = 0\}$.

(a) Show that S is a 2-dimensional smooth manifold at $(1,1,1)$.

(b) Compute the tangent plane at $(1,1,1)$.

5.5.11 If possible, use Theorem 5.5.3 or Corollary A to determine which sets are smooth manifolds.

(a) S is the zero level set of the function $F(x,y) = x^2 + 2xy + xy^3 + 1$.

(b) S is the zero level set of the function $G(x,y,z) = x^2 + y^2 - z^2$.

(c) S is the zero level set of the function $H(x,y,z) = x^2 + 4xyz + xy^5 + 1$.

(d) S is the explicit surface defined by $z = \cos(x)\sin(y)$.

5.5.12 Let S^{n-1} be the $(n-1)$-dimensional unit sphere in $\mathbb{R}^n$.

(a) Show that S^{n-1} is a $(n-1)$-dimensional smooth manifold.

(b) Compute the tangent plane of S^{n-1} at any given point $p \in S^{n-1}$.

Proofs

5.5.13 Theorem 5.5.3 has a nice corollary about smooth manifolds which you can use in many cases.

(a) Fill in the blank.

Corollary A. *Let $U \subseteq \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}$ be a C^1 function. Assume $S = f^{-1}(\{0\}) \neq \varnothing$. If ________________, then S is a $(n-1)$-dimensional smooth manifold.*

(b) Show that any graph $y = \phi(x)$ of a C^1 function $\phi : \mathbb{R} \to \mathbb{R}$ is a smooth curve in $\mathbb{R}^2$.

(c) Show that the hyperbola $x^2 - y^2 = 1$ is a smooth curve in $\mathbb{R}^2$.

(d) The union of lines $x^2 - y^2 = 0$ is not a smooth curve, but the given argument is incorrect.

1. *Define $f(x,y) = x^2 - y^2$ and the set*
$$S = \{(x,y) \in \mathbb{R}^2 : x^2 - y^2 = 0\} = f^{-1}(\{0\}).$$
2. *Notice f is C^1 and $\nabla f(x,y) = (2x, -2y)$.*
3. *The point $(0,0)$ satisfies $f(0,0) = 0$ so $(0,0) \in S$.*
4. *However, $\nabla f(0,0)$ vanishes so S is not a smooth curve by Corollary A.*

Identify the false claim and briefly explain.

5.5.14 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a C^1 function. Let $(a,b) \in \mathbb{R}^2$. Define

$$S = \{(x,y,z) : z = f(x,y)\} \subseteq \mathbb{R}^3$$

to be the graph of f.

(a) Use Theorem 5.5.3 to give another proof that S is a 2-dimensional smooth manifold at (a,b). *Hint:* Define a function to write the set S in implicit form.

(b) Use Theorem 5.5.3 to give another proof that the tangent plane at $p = (a,b,f(a,b)) \in S$ is

$$\left\{(x,y,z) \in \mathbb{R}^3 : z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b)\right\}$$

5.5.15 Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a C^1 map. Define its graph

$$S = \{(x, \phi(x)) : x \in \mathbb{R}^n\}$$

in $\mathbb{R}^{n+1}$. You already have results from Sections 4.5 and 4.6 which prove S is a n-dimensional smooth manifold and allow you to compute its tangent plane.

(a) Use Theorem 5.5.3 to reprove that S is a n-dimensional smooth manifold in $\mathbb{R}^{n+1}$.

(b) Fix $a \in \mathbb{R}^n$. Use Theorem 5.5.3 to recompute the tangent plane of S at $p = (a, \phi(a)) \in S$.

5.5.16 Use Theorem 5.5.7 to show that the set

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 16,\ (x-1)^2 + (y-2)^2 + (z-3)^2 = 16\}$$

is a 1-dimensional smooth manifold.

5.5.17 Prove that Theorem 5.5.7 with $k = 1$ implies Theorem 5.5.3.

Applications and beyond

5.5.18 Big proofs can make it difficult to see the big ideas. It is helpful to loosely summarize the main steps. Each summarized step should highlight a key idea. You will practice this with a simplified version of Theorem 5.5.3.

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function. Assume*

$$S = \{x \in \mathbb{R}^n : f(x) = 0\} = f^{-1}(\{0\})$$

is non-empty. Fix $p \in S$. If $\nabla f(p) \neq 0$, then S is a $(n-1)$-dimensional smooth manifold at p. Moreover, a vector $v \in \mathbb{R}^n$ is a tangent vector of S at p if and only if $\nabla f(p) \cdot v = 0$.

(a) Here is the first stage of the proof.

1. *Since $\nabla f(p) \neq 0$, assume without loss of generality that $\partial_n f(p) \neq 0$.*
2. *Since f is C^1, $f(p) = 0$, and $\partial_n f(p) \neq 0$, the implicit function theorem implies that the equation*
$$f(x_1, \ldots, x_n) = 0$$
locally defines x_n as a C^1 function of $x_1, \ldots, x_{n-1}$ near p.
3. *That is, there exists open sets $V \subseteq \mathbb{R}^{n-1}$ and $W \subseteq \mathbb{R}$ containing $(p_1, \ldots, p_{n-1}) \in \mathbb{R}^{n-1}$ and $p_n \in \mathbb{R}$ respectively and a C^1 function $\phi : V \to W$ such that*
$$\forall (x_1, \ldots, x_{n-1}, x_n) \in V \times W, \quad f(x_1, \ldots, x_n) = 0 \iff x_n = \phi(x_1, \ldots, x_{n-1}).$$
4. *Since $V \times W$ is an open set containing p, the set $S \cap (V \times W)$ is the graph of $\phi : V \to W$, so*
$$S \cap (V \times W) = \{(x_1, \ldots, x_{n-1}, x_n) \in \mathbb{R}^n : x_n = \phi(x_1, \ldots, x_{n-1})\}. \tag{5.5.3}$$
5. *Hence, S is an $(n-1)$-dimensional smooth manifold.*

Summarize this step with 1 full sentence. It will be coarse and leave out many details, but that is expected.

(b) Here is the second stage of the proof.

1. *It remains to show that $v \in T_p S$ if and only if $\nabla f(p) \cdot v = 0$.*
2. *Since tangent spaces are local, the sets S and $S' = S \cap (V \times W)$ have the same tangent space at p, because $p \in S \cap S'$ and $V \times W$ is an open set containing p.*
3. *Set $a = (p_1, \ldots, p_{n-1}) \in \mathbb{R}^{n-1}$ so $(a, \phi(a)) = p$.*
4. *Since S' is the graph of ϕ given by (5.5.3), it follows by another theorem that*
$$T_p S = T_p S' = \{(w, d\phi_a(w)) : w \in \mathbb{R}^{n-1}\}.$$

Summarize this step with 1 full sentence. It will be coarse and leave out many details, but that is expected.

(c) Here is the third stage of the proof.

1. *For $w \in \mathbb{R}^{n-1}$, we have that*

$$(w, d\phi_a(w)) = (w, \nabla\phi(a) \cdot w) = \big(w_1, \dots, w_{n-1}, \partial_1\phi(a)w_1 + \cdots + \partial_{n-1}\phi(a)w_{n-1}\big)$$

2. *By implicit differentiation, for $1 \le j \le n-1$,*

$$\begin{aligned} \frac{\partial f}{\partial x_j}(a, \phi(a)) + \frac{\partial f}{\partial x_n}(a, \phi(a))\frac{\partial \phi}{\partial x_j}(a) = 0 \\ \implies \partial_j f(p) + \partial_n f(p)\partial_j\phi(a) = 0 \\ \implies \partial_j\phi(a) = -\frac{\partial_j f(p)}{\partial_n f(p)} \end{aligned}$$

as $(a, \phi(a)) = p$ and $\partial_n f(p) \neq 0$.

3. *Combining these observations, it follows that $v \in T_pS$ if and only if there exists $w \in \mathbb{R}^{n-1}$ such that*

$$\begin{aligned} v = (w, d\phi_a(w)) &\iff (v_1, \dots, v_{n-1}, v_n) = (w_1, \dots, w_{n-1}, \partial_1\phi(a)w_1 + \cdots + \partial_{n-1}\phi(a)w_{n-1}) \\ &\iff v_n = \partial_1\phi(a)v_1 + \cdots + \partial_{n-1}\phi(a)v_{n-1} \\ &\iff v_n = -\frac{\partial_1 f(p)}{\partial_n f(p)}v_1 - \cdots - \frac{\partial_{n-1} f(p)}{\partial_n f(p)}v_{n-1} \\ &\iff \partial_1 f(p)v_1 + \cdots + \partial_{n-1} f(p)v_{n-1} + \partial_n f(p)v_n = 0. \end{aligned}$$

4. *Thus, $v \in T_pS$ if and only if $\nabla f(p) \cdot v = 0$ as required.*

Summarize this step with 1 full sentence. It will be coarse and leave out many details, but that is expected.

(d) Reread all three sentences. Cite the key theorems used in each step. After you finish, take a moment to appreciate how much beautiful theory was developed to achieve this proof.

5.6. Lagrange multipliers

It has been a long journey. You are finally ready to return to this chapter's motivating question.

How can you optimize a real-valued function f over a set S written in implicit form?

Remember that your existing optimization techniques are insufficient to adequately handle this situation. The local extreme value theorem can be used to search for extrema on the interior of a closed region, but it cannot provide any information about the boundary. If you can parametrize the boundary, then you can re-apply the local extreme value theorem on a lower dimensional optimization problem. However, if your boundary is a set written in *implicit* form, then it can be unreasonably complicated (or genuinely impossible) to parametrize.

This conundrum motivates the original question above and creates an important class of optimization problems, namely *constrained* optimization.

Let $f, g_1, \ldots, g_k$ be real-valued functions with domain $A \subseteq \mathbb{R}^n$. Fix $c_1, \ldots, c_k \in \mathbb{R}$. Optimize the objective function $f(x)$ subject to the constraints that

$$g_1(x) = c_1, \cdots, g_k(x) = c_k.$$

In other words, you want to optimize f on a set S written in implicit form by $g_1, \cdots, g_k$. The terminology "objective" refers to the function you want to optimize and the word "constraint" refers to the fact that the n variables $x_1, \ldots, x_n$ are constrained by the k nonlinear equations defined by $g_1, \ldots, g_k$. After all of your tremendous efforts with nonlinear systems, implicit functions, and smooth manifolds, you will reap the ultimate reward of a new technique designed specifically for constrained optimization problems: *the method of Lagrange multipliers*.

5.6.1 Extrema on subsets

Before diving into the new technique, you must precisely answer a basic question.

What exactly are (local or global) solutions of a constrained optimization problem?

Based on your prior experience with extrema (Sections 2.8 and 4.3.1), the formal description has a natural choice.

Definition 5.6.1 Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}$ be real-valued. Let S be a subset of A. The function f has a **global maximum on the set S (at the point a)** if

$$\forall x \in S, \ f(x) \leq f(a).$$

The function f has a **local maximum on the set S (at the point a)** if

$$\exists \varepsilon > 0 \text{ s.t. } \forall x \in S \cap B_\varepsilon(a), \ f(x) \leq f(a).$$

Global and local minima on S are defined similarly.

Remark 5.6.2 Equivalently, f has a local (or global) maximum of S at a if and only if the restriction $f|_S : S \to \mathbb{R}$ has a local (or global) maximum at a.

At first glance, this definition may appear indistinguishable from the earlier definitions (Definitions 2.8.1 and 4.3.1), but notice this refers to *subsets* of the domain. Otherwise, extrema of a function refer to the entire domain. These notions do not have to agree at all.

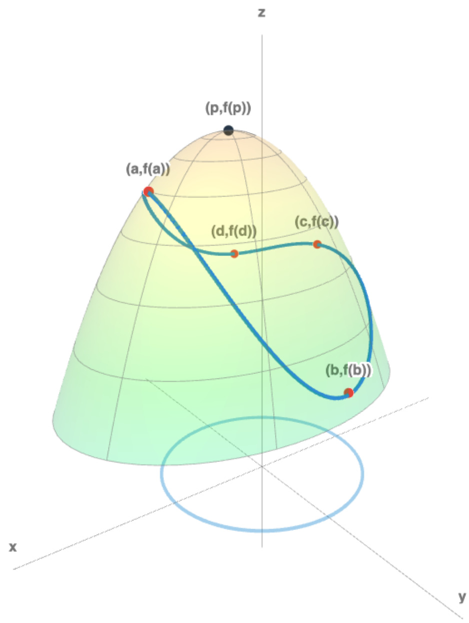
Example 5.6.3 Define the objective function $f:\mathbb{R}^2\to\mathbb{R}$ by $f(x,y)=16-x^2-2y^2+x-y$. Here is a classic constrained optimization problem.

Optimize $f(x,y)$ subject to the constraint $x^2+y^2=4$.

The constraint function is therefore $g:\mathbb{R}^2\to\mathbb{R}$ given by $g(x,y)=x^2+y^2$. More formally, you want to find the extrema of f on the set

$$S=\{(x,y)\in\mathbb{R}^2 : g(x,y)=4\}=\{(x,y)\in\mathbb{R}^2 : x^2+y^2=4\}.$$

The extrema of f on S are completely different than the extrema of f on its domain $\mathbb{R}^2$. This is illustrated with the Math3D demo below. The graph $z=f(x,y)$ is a 2-dimensional smooth manifold in $\mathbb{R}^3$ and the restricted graph $z=f|_S(x,y)$ is a 1-dimensional smooth manifold in $\mathbb{R}^3$.



The point $p\in\mathbb{R}^2$ corresponds to the only local extremum of f on its domain $\mathbb{R}^2$. In fact, p is the global maximum of f. Notice p is not a local extremum of f on S since $g(p)\neq 4$ implying $p\notin S$. By the local extreme value theorem, $\nabla f(p)=(0,0)$.

On the other hand, the points $a,b,c,d\in S$ correspond to the local extrema of f on S. None of them are local extrema of f and none of them are critical points of f. Note f has a local minimum on S at b,d and f has a local maximum on S at a,c. Moreover, f has a global maximum on S at a and a global minimum on S at b.

These definitions are actually the same if the point is an interior point of S.

Lemma 5.6.4 Let $A\subseteq\mathbb{R}^n$ and $f:A\to\mathbb{R}$ be a real-valued function. Let S be a subset of A. Assume a is an interior point of S. The function f has a local maximum on S at a if and only if f has a local maximum at a. The same equivalence holds for local minima.

Proof. Since a is an interior point of S, there exists $\delta>0$ such that $B_\delta(a)\subseteq S\subseteq A$. Thus, for $0<\varepsilon<\delta$, we have that $B_\varepsilon(a)\cap S=B_\varepsilon(a)=B_\varepsilon(a)\cap A$. Assume f has a local maximum on S at a. Then there exists $0<\varepsilon_1<\delta$ such that $\forall x\in S\cap B_{\varepsilon_1}(a), f(x)\le f(a)$. Since $B_{\varepsilon_1}(a)\cap S=B_{\varepsilon_1}(a)\cap A$, the same ε_1 shows that f has a local maximum at a (on its domain). The reverse direction follows similarly. ■

The local extreme value theorem (Theorem 4.3.6) is your primary tool for finding extrema on the *interior* of your function's domain. On the other hand, you want to find extrema on a *smooth manifold*. Unfortunately, the local extreme value theorem does not apply here.

Lemma 5.6.5 Fix $k, n \in \mathbb{N}^+$ with $k < n$. If S is a k-dimensional smooth manifold in $\mathbb{R}^n$, then S has empty interior.

Proof. Let $p \in S$ and let $\delta > 0$ be arbitrary. We wish to show that $B_\delta(p) \cap S^c$ is non-empty. By assumption, there exists an open set $U \subseteq \mathbb{R}^n$ such that $S \cap U$ is a graph of a function $\phi : V \to W$ for some open sets $V \subseteq \mathbb{R}^k$ and $W \subseteq \mathbb{R}^{n-k}$ with $V \times W \subseteq U$. Without loss of generality, assume it is *the* graph of ϕ so

$$S \cap U = \{(x, \phi(x)) \in V \times W : x \in V\}$$

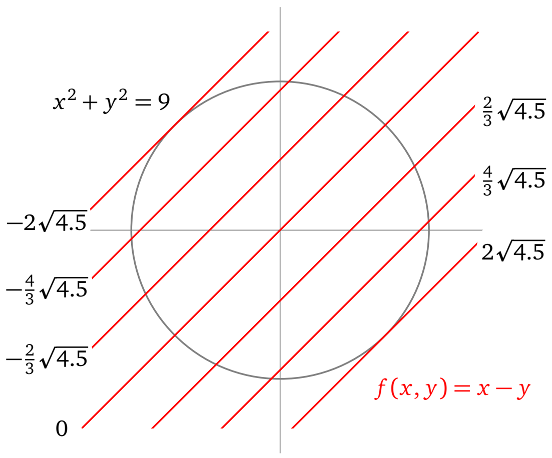
and $p = (a, \phi(a))$ for some $a \in V$. Since $p = (a, \phi(a)) \in U$ and U is open, there exists $0 < \varepsilon < \delta$ such that $B_\delta(p) \subseteq U$. The line $L = \{(a, \phi(a) + t) : t \in \mathbb{R}^{n-k}, ||t|| < \varepsilon\}$ lies inside $B_\varepsilon(p)$ but $S \cap L = \{(a, \phi(a))\}$. Therefore, there is a point (on the line) lying outside S but inside $B_\varepsilon(p) \subseteq B_\delta(p)$, as desired. ∎

These observations collectively illustrate that you really need to develop a new version of the local extreme value theorem that applies to smooth manifolds written in implicit form. How can you narrow down the possible locations for local extrema on a smooth manifold written in implicit form? The method of Lagrange multipliers is a flexible tool for such problems.

5.6.2 Lagrange multipliers with one constraint

For simplicity, assume there is only one constraint. The key ideas can be visualized with a 2-variable optimization problem.

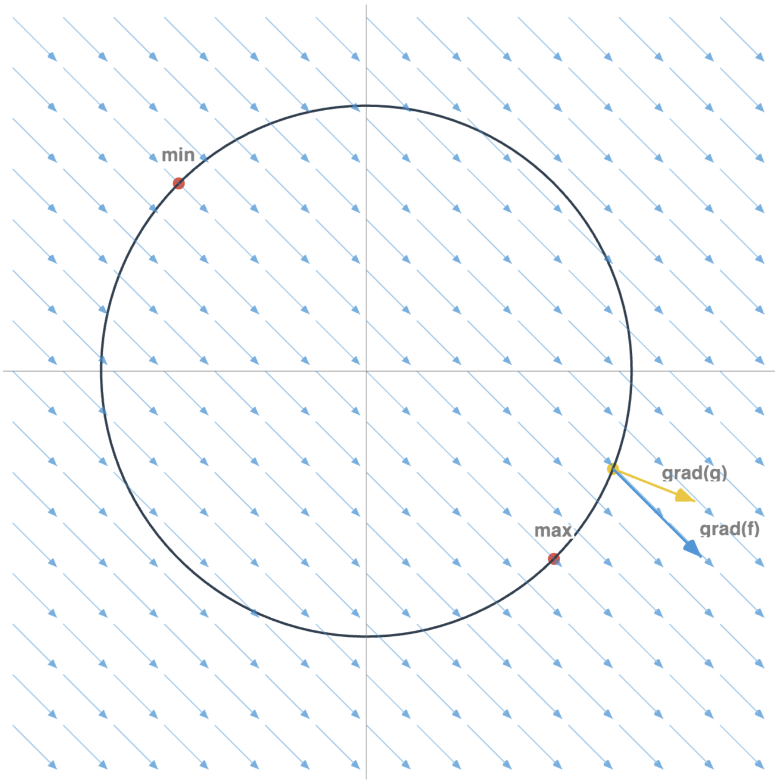
Example 5.6.6 Suppose you want to find the extrema of $f(x, y) = x - y$ subject to the constraint $g(x, y) = x^2 + y^2 = 9$. You can plot, the set $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 9\}$ along with some contours of f. Play with the Desmos graph of the same plot.



Here is the crucial observation: the extrema of f on S occur with value k when the curve S is *tangent* to the contour $f(x, y) = k$. For instance, the contour $f(x, y) = 2\sqrt{4.5}$ is tangent to the circle $x^2 + y^2 = 9$ at the point $(\sqrt{4.5}, \sqrt{4.5})$. More formally, if $S' = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 2\sqrt{4.5}\}$ and $p = (\sqrt{4.5}, \sqrt{4.5})$, then the tangent lines are the same, that is,

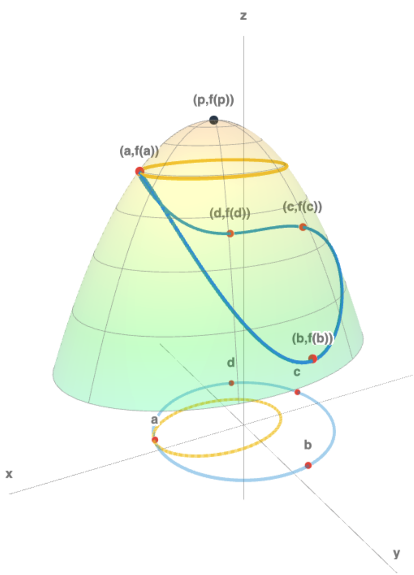
$$p + T_p S' = p + T_p S.$$

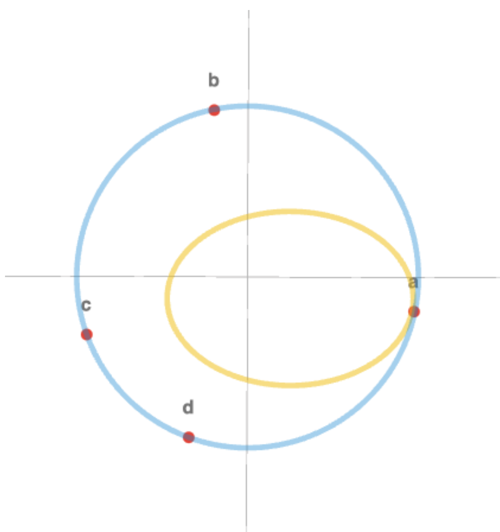
Due to the relationship between tangent spaces and gradients (Theorem 5.5.3), this observation can be rephrased in terms of gradient vectors. Since $\nabla f(p)$ is a normal vector defining the tangent line $p + T_pS'$ and $\nabla g(p)$ is a normal vector defining the tangent line $p + T_pS$, the gradient vectors $\nabla f(p)$ and $\nabla g(p)$ should be *parallel* if and only if $p + T_pS' = p + T_pS$.



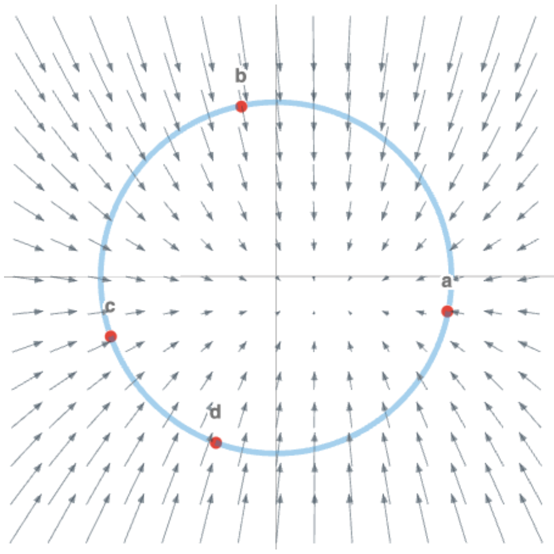
Play with the Math3D demo above to better visualize this phenomenon.

Example 5.6.7 The same features of Example 5.6.6 can be viewed in Example 5.6.3. At local extrema of f on the set $S = \{(x, y) : g(x, y) = 4\}$, the contours $f(x, y) = k$ are again tangent to the constraint $g(x, y) = x^2 + y^2 = 4$. Play with this Math3D contour demo.





Similarly, at local extrema of f on S, the gradient ∇f is parallel to the gradient ∇g. This is illustrated below. Play with this Math3D gradient demo for better visuals.



Notice the gradient vector field ∇f crosses the level set S orthogonally at the local extrema $a, b, c, d \in S$. Everywhere else, it does not cross orthogonally.

These illustrations inspire the method of Lagrange multipliers.

Theorem 5.6.8 (Lagrange multipliers with one constraint) Let U be an open subset of $\mathbb{R}^n$. Let $g : U \to \mathbb{R}$ and $f : U \to \mathbb{R}$ be C^1 real-valued functions. Fix $c \in \mathbb{R}$. Define

$$S = \{x \in U : g(x) = c\}.$$

Assume $\nabla g(p) \neq 0$ for any $p \in S$. If f has a local extremum on S at the point a, then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(a) = \lambda \nabla g(a).$$

The quantity λ is the **Lagrange multiplier**.

Remark 5.6.9 This theorem plays a role similar to the local extreme value theorem. Solutions $a \in U, \lambda \in \mathbb{R}$ to the **Lagrange system**

$$\nabla f(a) = \lambda \nabla g(a); \qquad g(a) = c$$

are *candidates* for local extrema of f on S. They are not necessarily local extrema.

Its proof is the crowning achievement for your theory of smooth manifolds and sets written in implicit form. The key ingredients include the definition of tangent vectors, linear algebra of orthogonal projections, multivariable chain rule, and local extreme value theorem.

Proof. Without loss of generality, assume that f has a local maximum on S at a. By Theorem 5.5.3, since $\nabla g(a) \neq 0$,

$$T_a S = \{v \in \mathbb{R}^n : \nabla g(a) \cdot v = 0\}.$$

and this tangent space is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. Thus, if a vector w is orthogonal to $T_a S$, then $w = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$. It therefore suffices to prove that $\nabla f(a)$ is orthogonal to any tangent vector of S at a.

Fix a tangent vector $v \in T_aS$. By Definition 4.5.2, there exists an open interval $I \subseteq \mathbb{R}$ containing 0, a C^1 parametric curve $\gamma : I \to \mathbb{R}^n$ with $\gamma(0) = a, \gamma'(0) = v$, and $\gamma(I) \subseteq S$. Since f has a local maximum on S at a, there exists $\varepsilon > 0$ such that $\forall x \in S \cap B_\varepsilon(a), f(x) \leq f(a)$. As γ is continuous at 0 and $\gamma(0) = a$, there exists $\delta > 0$ such that $\gamma((-\delta, \delta)) \subseteq B_\varepsilon(a)$.

Now, define the single variable function $h : (-\delta, \delta) \to \mathbb{R}$ by $h(t) = f(\gamma(t))$. As f and γ are differentiable, it follows that h is differentiable. We claim h has a maximum at 0. Let $t \in (-\delta, \delta)$, so $\gamma(t) \in S \cap B_\varepsilon(a)$ by choice of δ and definition of γ. The definition of ε implies that $h(t) = f(\gamma(t)) \leq f(a) = f(\gamma(0)) = h(0)$. This proves the claim. Finally, by the single variable local extreme value theorem, $h'(0) = 0$, so the multivariable chain rule implies that

$$0 = h'(0) = (f \circ \gamma)'(0) = \nabla f(\gamma(0)) \cdot \gamma'(0) = \nabla f(a) \cdot v$$

as desired. This completes the proof. ■

The method of Lagrange multipliers is a flexible tool that lends itself well to computations.

Example 5.6.10 Return to Example 5.6.3. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = 16 - x^2 - 2y^2 + x - y$ and $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(x, y) = x^2 + y^2$. Suppose you want to maximize $f(x, y)$ subject to the constraint $g(x, y) = 4$. In other words, you want to maximize f on the set $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 4\}$. Since S is compact and f is continuous, the function f attains a maximum value on S at a point $p \in S$ by the global extreme value theorem. It follows that f also has a local extrema on S at p.

By the method of Lagrange multipliers (Theorem 5.6.8), it follows that the point p must satisfy the Lagrange system:

$$\begin{aligned} & \nabla f(x, y) = \lambda \nabla g(x, y); \quad g(x, y) = 4 \\ \iff & \begin{bmatrix} -2x + 1 \\ -4y - 1 \end{bmatrix} = \begin{bmatrix} 2\lambda x \\ 2\lambda y \end{bmatrix}; \quad x^2 + y^2 = 4 \end{aligned}$$

for $\lambda \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^2$. The goal is to solve this system for all possible solutions. In general, this can be quite hard to solve and you must do careful casework.

If $x = 0$ then the system has no solution since $-2x + 1 = 2\lambda x$ implies $1 = 0$. If $y = 0$ then again the system has no solution since $-4y - 1 = 2\lambda y$ implies $-1 = 0$. Thus, both x, y are non-zero. The system is therefore equivalent to

$$\begin{bmatrix} -2 + x^{-1} \\ -4 - y^{-1} \end{bmatrix} = \begin{bmatrix} 2\lambda \\ 2\lambda \end{bmatrix}; \quad x^2 + y^2 = 4$$

Hence, $-2 + x^{-1} = -4 - y^{-1}$ implies that $-2xy + y = -4xy - x$ so either $x = -1/2$ or $y = \frac{-x}{1+2x}$. If $x = -1/2$ then $\lambda = -2 + x^{-1} = -4$, but then $-4 - y^{-1} = 2\lambda = -4$ has no solution for y. Otherwise, substituting $y = \frac{-x}{1+2x}$ into the equation $x^2 + y^2 = 4$ gives a quartic equation in x so you can find the 4 solutions

$$x = \frac{1}{4}\big(-1 - \sqrt{17} \pm \sqrt{14 - 2\sqrt{17}}\big), \frac{1}{4}\big(-1 + \sqrt{17} \pm \sqrt{14 + 2\sqrt{17}}\big)$$

after using a computer algebra system. From the relation $y = \frac{-x}{1+2x}$, these give corresponding y-values and thus produce 4 points $a, b, c, d \in S$. (These four points are exactly those appearing in Example 5.6.3 and they all happen to be local extrema of f on $x^2 + y^2 = 4$.)

As argued above, the maximum of f on the curve S must be one of $a, b, c, d \in S$. By evaluating f at each of these points, you can find that f attains a maximum value of

$f(a) = \frac{1}{4}(40 + \sqrt{142 + 34\sqrt{17}}) \approx 14.199$ on the curve S at the point

$$a = \left(\frac{1}{4}\left(-1 + \sqrt{17} + \sqrt{14 - 2\sqrt{17}}\right), \frac{1}{4}\left(-1 + \sqrt{17} - \sqrt{14 - 2\sqrt{17}}\right)\right) \approx (1.960, -0.398).$$

That was a really messy example, but that can happen with any optimization problem. The above situation was also somewhat silly since it can be solved much more easily by parametrizing the circle in $\mathbb{R}^2$. The next examples showcases why Lagrange multipliers is so much more powerful for optimizing on sets written in implicit form.

Example 5.6.11 Suppose you want to maximize $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = x_1 + \cdots + x_n$ on the unit sphere $x_1^2 + \cdots + x_n^2 = 1$. Since the unit sphere S is compact and f is continuous, a maximum of f on S must exist by the global extreme value theorem. By the method of Lagrange multipliers, this maximum must satisfy the Lagrange system:

$$\nabla f(x) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 2\lambda x_1 \\ \vdots \\ 2\lambda x_n \end{bmatrix} = \lambda \nabla g(x); \qquad x_1^2 + \cdots + x_n^2 = 1.$$

for some $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$. Note $\lambda = 0$ is impossible so it must be that $x_i = (2\lambda)^{-1}$ for $1 \leq i \leq n$. Evaluating this in the constraint equation, it follows that

$$(1/2\lambda)^2 + \cdots + (1/2\lambda)^2 = 1 \iff n(1/2\lambda)^2 = 1 \iff 2\lambda = \pm\sqrt{n}.$$

This gives two possible solutions to the system, namely $a = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$ or $b = (-1/\sqrt{n}, \ldots, -1/\sqrt{n})$. As argued above, the maximum of f on S must be one of these two points and you can verify that f attains a maximum on S at a with value $f(a) = 1/\sqrt{n} + \cdots + 1/\sqrt{n} = n/\sqrt{n} = \sqrt{n}$. It would have been much worse to try and optimize by parametrizing the unit sphere in $\mathbb{R}^n$.

Remember that the method of Lagrange multipliers only allows you to detect where local extrema *may* occur on a set written in implicit form. It narrows your options but it does not directly prove that any of them are local extrema on the set.

5.6.3 Lagrange multipliers with many constraints

You can also optimize with several constraints using a similar technique and some linear algebra.

Theorem 5.6.12 (Lagrange multipliers with many constraints) Let $U \subseteq \mathbb{R}^n$ be open. Let $g_1 : U \to \mathbb{R}, \ldots, g_k : U \to \mathbb{R}$ and $f : U \to \mathbb{R}$ be C^1 real-valued functions. Fix $c_1, \ldots, c_k \in \mathbb{R}$. Define the set

$$S = \{x \in U : g_1(x) = c_1, \ldots, g_k(x) = c_k\}.$$

Assume for every $p \in S$ that the set $\{\nabla g_1(p), \ldots, \nabla g_k(p)\}$ is linearly independent. If a is a local extremum of f on S, then there exists $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that

$$\nabla f(a) = \sum_{i=1}^{k} \lambda_i \nabla g_i(a).$$

Remark 5.6.13 Comparing with Theorem 5.5.7, it is hopefully unsurprising that you require the k gradients to be linearly independent. This ensures that the set S is a smooth manifold and allows you to compute its tangent space.

Proof. This is left as a challenging exercise. The argument is similar to the proof of Theorem 5.6.8 with some additional linear algebra. First, show that the gradients $\nabla g_1(a), \ldots, \nabla g_k(a)$ form a basis for the subspace $V \subseteq \mathbb{R}^n$ of vectors which are orthogonal to $T_a S$. Second, show that $\nabla f(a)$ belongs to V. ■

Theorem 5.6.12 is also useful in applications but the computations become drastically more difficult with every extra constraint.

Example 5.6.14 Suppose you want to minimize $f(x,y,z) = x^2 + y^2 + z^2$ subject to the constraints

$$0 = g(x,y,z) = x^2 + y^2 - z^2, \qquad 0 = h(x,y,z) = x - 2z - 3.$$

Let S be the implicit set defined by g and h. You can verify that S is compact (though it is not easy to see) and f is continuous, so f has a minimum on S by the global extreme value theorem. By Theorem 5.6.12, this minimum on S must satisfy the Lagrange system given by:

$$\begin{aligned} &\nabla f(x,y,z) = \lambda \nabla g(x,y,z) + \mu \nabla h(x,y,z); \qquad g(x,y,z) = 0; \qquad h(x,y,z) = 0 \\ \iff \quad & \begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix} = \begin{bmatrix} 2\lambda x + \mu \\ 2\lambda y \\ -2\lambda z - 2\mu \end{bmatrix}; \qquad x^2 + y^2 - z^2 = 0; \quad x - 2z - 3 = 0. \end{aligned}$$

for $x, y, z, \lambda, \mu \in \mathbb{R}$. This is a system of 5 nonlinear equations with 5 variables. What a nightmare! At this point, you may willingly surrender to your nearest computer algebra system to solve such equations.

You will obtain one solution of $(x,y,z) = (1,0,-1)$ with $(\lambda,\mu) = (-\frac{1}{3}, \frac{-4}{3})$ and another solution of $(x,y,z) = (-3,0,-3)$ with $(\lambda,\mu) = (-3,-12)$. (The values of λ,μ do not affect the final minimum but you will need to solve for them anyways.) The minimum of f on S must occur at one of these two points $(-3,0,-3)$ or $(1,0,-1)$. Evaluating f, you can find that $f(-3,0,-3) = 18$ and $f(1,0,-1) = 2$ so the minimum of f on S is 2.

Although this technique may appear rather complicated in specific numerical examples, if you really need to compute something, then you can use computer algebra software to solve systems of equations; there is no need to feel intimidated by the calculations. The method of Lagrange multipliers is a celebrated optimization technique with a rich mathematical theory behind it. More importantly, it is a fundamental tool for solving constrained optimization problems. This produces a vast array of applications across many disciplines including economics, statistics, probability, physics, data science, mathematics, and much more. Next, you will triumphantly conclude this chapter by revisiting more advanced optimization problems and feeling the full extent of your newfound powers.

Exercises for Section 5.6

Concepts and definitions

5.6.1 Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are C^1 functions and $S = \{x \in \mathbb{R}^n : g(x) = 237\}$ is a level set of g with $\nabla g(p) \neq 0$ for all $p \in S$. Which statements are true or false?

(a) If f has a local extremum on S at a, then $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$.

(b) If $a \in S$ and $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$, then f has a local extremum on S at a.

(c) If f has a maximum on S at a, then $\nabla f(a) = 0$.

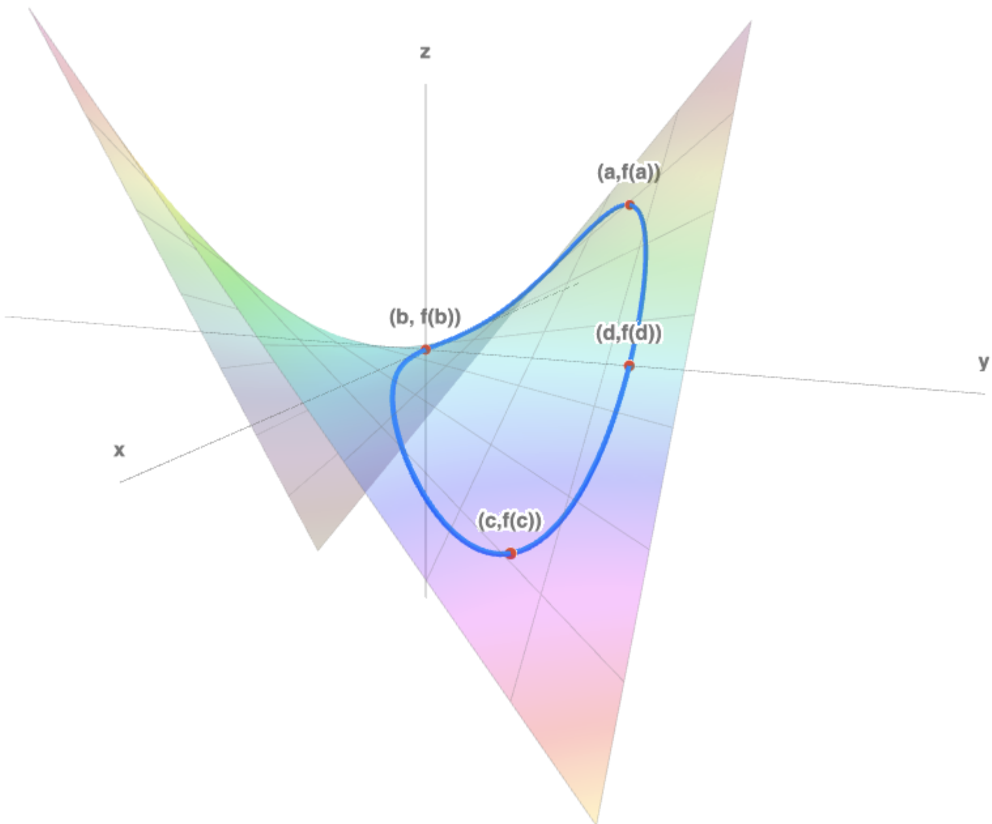
(d) If f has a local maximum on S at a, then f has a local maximum at a.

(e) If f has a maximum on S at a, then f has a local maximum at a.

5.6.2 Define $f(x, y) = -xy$ and $g(x, y) = x^2/4 + (y-1)^2$. Define the set

$$S = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 1\}.$$

The graph of $z = f(x, y)$ and the graph of $z = f|_S(x, y)$ are shown below.



View Math3D to see the points on the graph corresponding to $a, b, c, d \in S$. **Assume** $\nabla f(b) = 0$.

(a) Which of $a, b, c, d \in S$ appear to be local extrema of f?

(b) Which of $a, b, c, d \in S$ appear to be critical points of f?

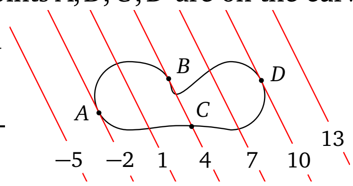
(c) Which of $a, b, c, d \in S$ appear to be local extrema of f on S?

(d) Which of $a, b, c, d \in S$ appear to be solutions to the Lagrange multipliers system for f on S?

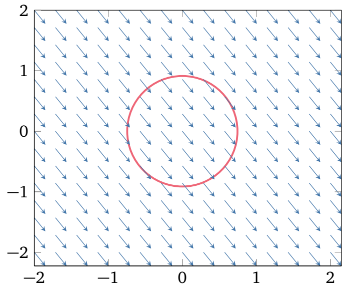
5.6.3 Let $S \subseteq \mathbb{R}^2$ be the level set of a real-valued C^1 map g. The points A, B, C, D are on the curve S.

(a) Decide which of these points are local maxima on S, local minima on S, or neither.

(b) Decide which of these points appear to satisfy the corresponding Lagrange system of equations.



5.6.4 The graph below shows a vector field $\nabla f(x,y)$ and a constraint $S = \{(x,y) \in \mathbb{R}^2 : g(x,y) = 0\}$.



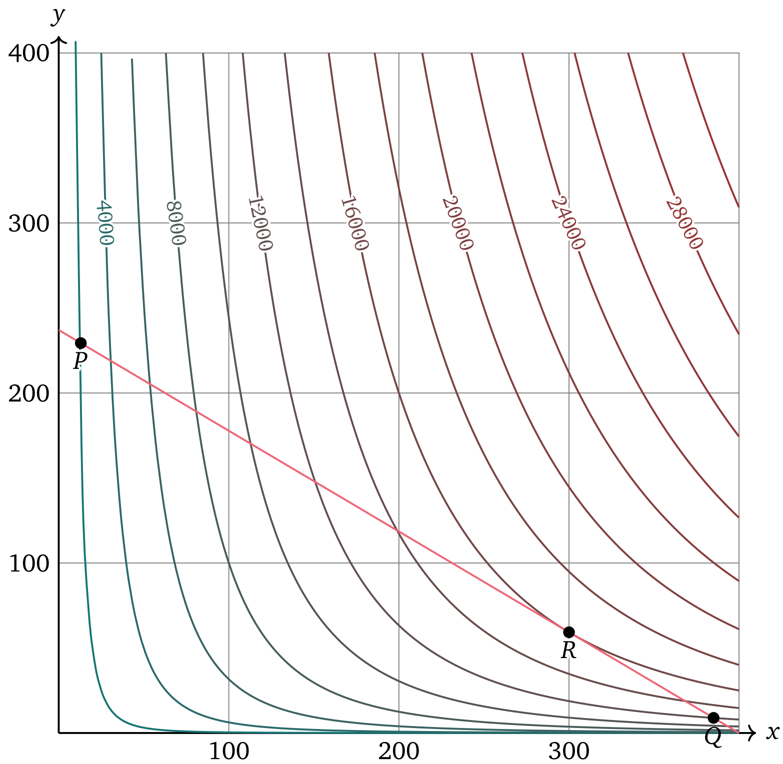
(a) Label point(s) at which f possibly has a local extremum on S.

(b) At each of those point(s), sketch the gradient ∇g and decide whether the corresponding Lagrange multiplier λ is positive or negative. *Hint:* There are two sets of valid answers.

5.6.5 Define

$$f(x,y) = 80x^{3/4}y^{1/4} \qquad g(x,y) = 16x + 27y$$

We want to maximize $f(x,y)$ subject to the constraint $g(x,y) = 6400$.[5]



(a) What is the maximum of f subject to the constraint?

(b) At points P, Q and R, roughly sketch the directions of ∇f and ∇g. (Ignore the lengths.)

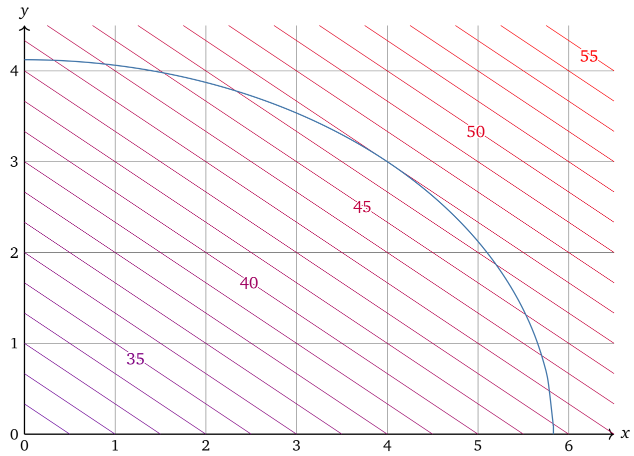
(c) If you solve the Lagrange multiplier system, which point(s) will you find?

[5] Used with permission of Cindy Blois from MAT133 2022-23.

5.6.6 An ant walks on a flat stove. The temperature in $^\circ C$ at a location x cm east and y cm north from the southwest corner of the stove is approximately

$$T(x, y) = 3y + 2x + 30$$

She walks along the path $x^2 + 2y^2 = 34$, with $x \geq 0$ and $y \geq 0$[6] .



(a) What are the warmest and coldest temperatures she experiences?

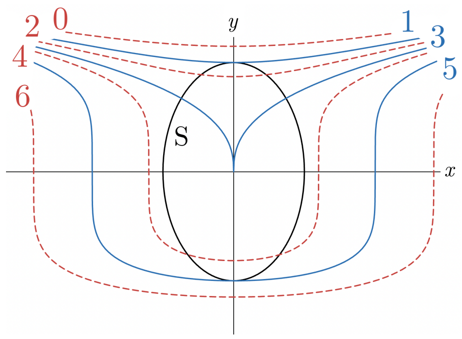
(b) At approximately which points does she experience the warmest and coldest temperatures?

(c) If you solve the Lagrange multiplier system, which point(s) will you find? Explain why.

5.6.7 Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be C^1 functions. Define

$$S = \{x \in \mathbb{R}^2 : g(x) = 0\}$$

to be the constraint curve. This curve S and level sets of f are plotted below.



Assume f attains its maximum on S at $a \in S$, and its minimum on S at $b \in S$.

(a) Label a and b on the contour plot. Identify $f(a)$ and $f(b)$.

(b) Draw the tangent line to the level sets $g(x) = 0$ and $f(x) = f(a)$ at the point a.

(c) Draw the tangent line to the level sets $g(x) = 0$ and $f(x) = f(b)$ at the point b.

(d) Draw the gradients $\nabla f(a)$ and $\nabla f(b)$. Do not worry about length.

(e) Draw the gradients $\nabla g(a)$ and $\nabla g(b)$. Do not worry about length.

[6] Used with permission of Cindy Blois from MAT133 2022-23.

Computations

For all optimization problems below, remember to justify your calculations.

5.6.8 Find the maximum and minimum values of $f(x,y) = 2y - 2\sqrt{3}x$ subject to the constraint $x^2 + y^2 = 16$. Justify that you have found the maximum.

5.6.9 Minimize $f(x,y) = x^2 + 2y^2$ on the hyperbola $xy = -2$.

5.6.10 Find the extrema of $f(x,y,z) = x + 2y + 4z$ on $x^2 + y^2 + z^2 = 4$.

5.6.11 Fix $a \in \mathbb{R}^n$ non-zero. Find the extrema of $f(x) = a_1 x_1 + \cdots + a_n x_n$ on the unit sphere S^{n-1}.

5.6.12 Find the extrema of $f(x,y,z) = -x^2 - y^2 - z^2$ subject to the constraints $x^2 = y^2 + z^2$ and $y + z - x + 1 = 0$.

5.6.13 Find the maximum value of $f(x,y,z) = x + y^2 - z^2$ on the intersection of the surfaces $x^2 + y^2 + z^2 = 32$ and $x^2 - y^2 + z^2 = 4$. Use the method of Lagrange multipliers with many constraints.

Proofs

5.6.14 Lagrange multipliers is often presented as a straightforward procedure but justifications are easy to miss. If you are not careful, these can lead to genuinely serious errors.

(a) Arthur is trying to maximize $f(x,y) = x^2 + 4y^2 - 2x + 8y$ subject to the constraint $x + 2y = 7$.

1. *Define* $S = \{(x,y) \in \mathbb{R}^2 : x + 2y = 7\}$ *and* $g(x,y) = x + 2y$.
2. *By the method of Lagrange multipliers, the maximum of* f *on* S *satisfies the system of equations*
$$\nabla f(x,y) = \lambda \nabla g(x,y), \qquad g(x,y) = 7$$
for some $x, y, \lambda \in \mathbb{R}$.
3. *This implies that*
$$2x - 2 = \lambda, \qquad 8y + 8 = 2\lambda, \qquad x + 2y = 7.$$
4. *By the almighty WolframAlpha, the solutions are* $x = 5, y = 1$, *and* $\lambda = 8$.
5. *Thus, the maximum of* f *on* S *occurs at* $(x,y) = (5,1)$ *with value* $f(5,1) = 27$.

Arthur has committed many sins, but one of his flaws is fatal. Identify the flaw. Can he fix it?

(b) Alyssa is searching for the minimum of $f(x,y) = x + y$ subject to the constraint $x^2 + y^2 = 4$.

> 1. *Define* $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 4\}$ *and* $g(x, y) = x^2 + y^2$.
> 2. *By the method of Lagrange multipliers, the extrema of* f *on* S *satisfies the system of equations*
> $$\nabla f(x, y) = \lambda \nabla g(x, y), \qquad g(x, y) = 4$$
> *for some* $x, y, \lambda \in \mathbb{R}$.
> 3. *This implies that*
> $$1 = 2\lambda x, \qquad 1 = 2\lambda y, \qquad x^2 + y^2 = 4.$$
> 4. *Using my fountain pen, I found the solutions are* $(x, y) = (\sqrt{2}, \sqrt{2})$ *and* $(-\sqrt{2}, -\sqrt{2})$.
> 5. *As* $f(\sqrt{2}, \sqrt{2}) = 2\sqrt{2}$ *and* $f(-\sqrt{2}, -\sqrt{2}) = -2\sqrt{2}$*, the minimum value of* f *on* S *is* $-2\sqrt{2}$.

Alyssa has not made a serious error, but she is missing two crucial justifications. Identify the missing justifications and suggest how to fix it.

(c) Federico wants the minimum of $f(x, y) = x + y$ when $x^2 + y^2 = 4$ and $y \geq 0$.

> 1. *Define* $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 4, y \geq 0\}$ *and* $g(x, y) = x^2 + y^2$.
> 2. *By the method of Lagrange multipliers, the extrema of* f *on* S *satisfies the system of equations*
> $$\nabla f(x, y) = \lambda \nabla g(x, y), \qquad g(x, y) = 4$$
> *for some* $x, y, \lambda \in \mathbb{R}$.
> 3. *This implies that*
> $$1 = 2\lambda x, \qquad 1 = 2\lambda y, \qquad x^2 + y^2 = 4.$$
> 4. *I have computed that the solutions are* $(x, y) = (\sqrt{2}, \sqrt{2})$ *and* $(-\sqrt{2}, -\sqrt{2})$.
> 5. *Since* $(-\sqrt{2}, -\sqrt{2}) \notin S$*, the minimum of* f *on* S *is* $f(\sqrt{2}, \sqrt{2}) = 2\sqrt{2}$.

Aside from missing justifications, Federico messed up. Identify his mistake and how to fix it.

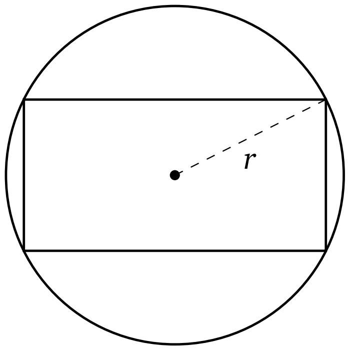
(d) Sirui wants to minimize $f(x, y) = x + y$ subject to the constraint $2x + 3y = 7$ and $x \geq 0, y \geq 0$.

> 1. *Define* $S = \{(x, y) \in \mathbb{R}^2 : 2x + 3y = 7, x \geq 0, y \geq 0\}$ *and* $g(x, y) = 2x + 3y$.
> 2. *By the method of Lagrange multipliers, the extrema of* f *on* S *satisfies the system of equations*
> $$\nabla f(x, y) = \lambda \nabla g(x, y), \qquad g(x, y) = 7$$
> *for some* $x, y, \lambda \in \mathbb{R}$.
> 3. *This implies that*
> $$1 = 2\lambda, \qquad 1 = 3\lambda, \qquad 2x + 3y = 7.$$
> 4. *There are no solutions to this system.*
> 5. *Thus,* f *does not attain a minimum on* S.

Aside from missing justifications, Sirui made an error. Identify the error and how to fix it.

Applications and beyond

5.6.15 A polygon is **inscribed** on a circle if all of its vertices are on the circle. You will find the maximum perimeter of a rectangle that you can inscribe in a circle of radius r.



(a) Label the rectangle above with two variables x, y and write the perimeter as a function $P(x, y)$.

(b) Determine the constraints on x and y so that the vertices of the rectangle are on the circle.

(c) Find the values of x and y that maximuize the perimeter of the rectangle and the corresponding maximum perimeter. Justify that you have found the maximum.

5.6.16 A company CHAYR must produce 20,000 chairs. The number of chairs they can produce is given by $P(K, L) = 5K^{0.2}L^{0.8}$, where K is their capital expenditure and L is their labour costs. They want to produce those 20,000 chairs while minimizing their total costs.

(a) Identify the function you wish to optimize and the domain over which you are optimizing.

(b) Find the choice of K and L that solves the Lagrange multiplier system.

(c) Solving the Lagrange multiplier system does not guarantee you have a global minimum. Draw a "picture proof" which illustrates why you have found the global minimum cost. Consider how you might make this a formal proof. *Hint:* Draw curves in the (K, L)-plane using a computer.

5.6.17 You want to maximize consumer satisfaction. You invest in products at quantities x and y with prices p_1 and p_2 hundred thousands of dollars per unit respectively. From market research, the function representing consumer satisfaction is given by $F(x, y) = x^{1/4}y^{3/4}$. What is the maximum satisfaction that can be obtained with a budget of c hundred thousand dollars? Your answer may be expressed in terms of p_1, p_2, and c.

5.6.18 To distill the big ideas in a long proof, you can divide it into stages and draw a picture for each stage. By drawing each step, you can look at the series of pictures to visualize the proof. You will practice this with the theorem on Lagrange multipliers (Theorem 5.6.8).

(a) Here is the first stage of the proof.

> 1. *Without loss of generality, assume that f has a local maximum on S at a.*
> 2. *By a theorem on smooth manifolds written in implicit form, since $\nabla g(a) \neq 0$,*
>
> $$T_aS = \{v \in \mathbb{R}^n : \nabla g(a) \cdot v = 0\}$$
>
> *and this tangent space is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$.*
> 3. *Thus, if a vector w is orthogonal to T_aS, then $w = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$.*
> 4. *It therefore suffices to prove that $\nabla f(a)$ is orthogonal to any tangent vector of S at a.*

Draw a plot in $\mathbb{R}^2$ of the constraint $g(x, y) = x^2 + y^2 = 4$, a point a, the gradient vector $\nabla g(a)$, and the tangent plane of S at a.

(b) Here is the second stage of the proof.

> 1. *Fix a tangent vector $v \in T_aS$.*
> 2. *There exists an open interval $I \subseteq \mathbb{R}$ containing 0, a differentiable function $\gamma : I \to \mathbb{R}^n$ with $\gamma(0) = a, \gamma'(0) = v$, and $\gamma(I) \subseteq S$.*
> 3. *Since f has a local maximum on S at a, there exists $\varepsilon > 0$ such that $\forall x \in S \cap B_\varepsilon(a), f(x) \leq f(a)$.*
> 4. *As γ is continuous at 0 and $\gamma(0) = a$, there exists $\delta > 0$ such that $\gamma((-\delta, \delta)) \subseteq B_\varepsilon(a)$.*
> 5. *Define the single variable function $h : (-\delta, \delta) \to \mathbb{R}$ by $h(t) = f(\gamma(t))$.*

Add γ and v to your diagram above.

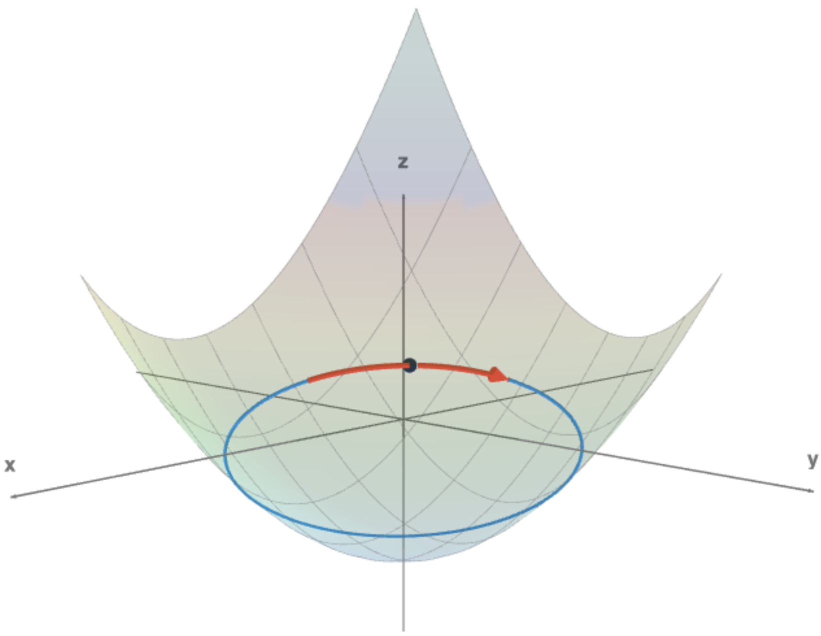
(c) Here is the third stage of the proof.

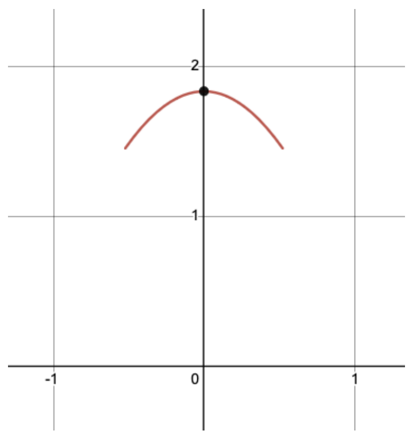
1. *As f and γ are differentiable, it follows by the chain rule that h is differentiable.*
2. *We claim h has a maximum at 0.*
3. *Let $t \in (-\delta, \delta)$, so $\gamma(t) \in S \cap B_\varepsilon(a)$ by choice of δ and definition of γ.*
4. *The definition of ε implies that $h(t) = f(\gamma(t)) \leq f(a) = f(\gamma(0)) = h(0)$.*
5. *This proves the claim.*
6. *Finally, by the single variable local extreme value theorem, $h'(0) = 0$.*
7. *The multivariable chain rule implies that*

$$0 = h'(0) = (f \circ \gamma)'(0) = \nabla f(\gamma(0)) \cdot \gamma'(0) = \nabla f(a) \cdot v$$

as desired. This completes the proof.

Choosing $f(x, y) = (x+1)^2 + (y+1)^2 - 4$ gives the Math3D plot in $\mathbb{R}^3$ of the graphs $z = f(x, y)$ and $z = f|_S(x, y)$ below. We also have the plot in $\mathbb{R}^2$ of the graph of $h : (-\delta, \delta) \to \mathbb{R}$.





Label the plots above with corresponding quantities in the proof:

- The graph of $z = f(x, y)$ in $\mathbb{R}^3$
- The graph of $z = f|_S(x, y)$ in $\mathbb{R}^3$
- The point $(a, f(a))$ in $\mathbb{R}^3$.
- The curve $(\gamma(t), f(\gamma(t))$ in $\mathbb{R}^3$.
- The graph of $z = h(t)$ in $\mathbb{R}^2$.
- The point $(0, h(0))$ in $\mathbb{R}^2$.

Once finished, your sequence of pictures will be complete! You have distilled the big ideas.

5.7. Optimization with constraints

Real world problems almost always have constraints, so these problems are pervasive in economics, statistics, data science, physics, computer science and mathematics. The list is endless. Thus far, you have built three core tools in multivariable optimization: the extreme value theorem, the local extreme value theorem, and the method of Lagrange multipliers.

The method of Lagrange multipliers opens a new strategy for solving more advanced optimization problems, namely *constrained* optimization problems. This strategy is usually in combination with your other tools, so this section will display some examples that blend all of these together. The basic ideas from Section 4.4 remain the same.

Determine whether global extrema must exist. Identify the critical points on the interior. Check the boundary for extrema, either by parametrizing or by Lagrange. Justify.

Notice you are now able to handle more boundary scenarios. As before, there is no fixed procedure to solve every kind of optimization problem. Each scenario can require a different approach, so remember to be flexible and think creatively. There are limitations and utilities for each method that you will discover through practice. This section will display some selected examples which illustrate the key ideas. The first example begins with a classic constrained optimization problem.

Example 5.7.1 A company CATS4EVER wants to maximize production of luxury kitty nail polish Artemis Azure under a budget constraint. They purchase x bins of polymer and y bins of pigments at prices 2 and 3 thousand dollars per unit respectively. They will produce f thousand bottles of Artemis Azure with these raw materials and they estimate that their production function is given by

$$f(x,y) = x^{2/3}y^{1/3}$$

Assuming they spend their entire budget of 6 million dollars, what is their maximum production at this budget? The budget constraint implies that

$$2000x + 3000y = 6000000 \iff 2x + 3y = 6000.$$

Since x and y are quantities of raw materials, these variables must be non-negative. Thus, the company wants to maximize the function $f(x,y) = x^{2/3}y^{1/3}$ over the domain

$$L = \{(x,y) \in \mathbb{R}^2 : x, y \geq 0, 2x + 3y = 6000\}.$$

Note L is a closed line segment in the first quadrant, so L is compact. Since f is continuous, the global extreme value theorem implies that the function p has a maximum value on L.

Define $U = \{(x,y) \in \mathbb{R}^2 : x > 0, y > 0\}$ and $g : U \to \mathbb{R}$ by $g(x,y) = 2x + 3y$ so g is C^1 on U and $\nabla g(x,y) = (2,3)$ is non-zero on U. Define the line segment without endpoints

$$S = \{(x,y) \in U : g(x,y) = 6000\} = \{(x,y) \in \mathbb{R}^2 : x, y > 0, 2x + 3y = 6000\},$$

so $L = S \cup \{(3000,0),(0,2000)\}$. Then the maximum value of f on L occurs either at $(3000,0),(0,2000)$, or on the set S.

By the method of Lagrange multipliers (Theorem 5.6.8), a local maximum of f on S must satisfy the system of equations given by

$$\begin{aligned} &\nabla f(x,y) = \lambda \nabla g(x,y); \qquad g(x,y) = 6000 \\ \iff &\frac{2}{3}x^{-1/3}y^{1/3} = 2\lambda; \quad \frac{1}{3}x^{2/3}y^{-2/3} = 3\lambda; \quad 2x + 3y = 6000. \end{aligned} \tag{5.7.1}$$

for some $(x,y) \in U$ and $\lambda \in \mathbb{R}$. These equations imply that

$$\frac{1}{3}x^{-1/3}y^{1/3} = \frac{1}{9}x^{2/3}y^{-2/3} \implies x = 3y$$

in which case $2(3y)+3y = 6000$ gives that $y = 2000/3 \approx 666$ and hence $x = 2000$. Thus, the only solution to the Lagrange multipliers system has $(x,y) = (2000, 2000/3)$.

Overall, the maximum value of f on the closed line segment L occurs at either $(3000,0)$, $(0,2000)$, or $(2000,2000/3)$. Since $f(3000,0) = f(0,2000) = 0$ and $f(2000,2000/3) \approx 1387$, the maximum value is 1387. Assuming the CATS4EVER company spends their entire budget of 6 million dollars, they should buy 667 bins of polymer and 2000 bins of pigments to produce approximately $1{,}387{,}000$ bottles of luxury Artemis Azure kitty nail polish.

After solving constrained optimization problems with the method of Lagrange multipliers, you may notice that the value of the Lagrange multiplier λ for the optimal value is never used in the final solution. What is the meaning of λ at the optimal value? A formal investigation leads to the amazing envelope theorem. This is far beyond the scope of this text and better reserved for a course with variational calculus. Instead, an informal interpretation will suffice.

The value of λ is the rate of change of the optimum value of f as c increases, where $g(x) = c$ is the constraint.

This can be expressed a bit more formally with language from the implicit function theorem.

For $c \in \mathbb{R}$, let $x(c) \in \mathbb{R}^n$ be the optimum point and $\lambda(c) \in \mathbb{R}$ be the Lagrange multiplier corresponding to the optimum value of $f(x)$ subject to the constraint $g(x) = c$. If the multiplier $\lambda(c)$ and the point $x(c)$ are locally defined as C^1 functions of c, then

$$\lambda(c) = \frac{d}{dc} f(x(c)).$$

This interpretation of λ can be quite useful in applied contexts, especially economics.

Example 5.7.2 How can you interpret the Lagrange multiplier in Example 5.7.1? The company CATS4EVER found that $x = 2000$, $y = \frac{2000}{3}$ yielded the optimal value for production $f(x,y)$ with constraint $g(x,y) = 6000$. From (5.7.1), you can solve for λ to find that $\lambda \approx 0.231$. Therefore, if you increase from budget $c = 6000$ to $c+1 = 6001$, then the *optimal* production f subject to this increased budget constraint should increase by approximately $\lambda \approx 0.231$. In plain language, if the company CATS4EVER increases their budget by \$1000, then they will be able to produce approximately 231 more bottles of kitty nail polish.

By combining all of your techniques (global extreme value theorem, local extreme value theorem, and Lagrange multipliers), you can optimize over more complicated regions, such as those defined by inequalities.

Example 5.7.3 Continue with CATS4EVER and the description in Example 5.7.1. Their CEO, Meowth Moneybags, is having second thoughts over her decision to spend the entire budget. What is the maximum production if CATS4EVER spends *at most* 6 million dollars? By following the same arguments as before, the company now wants to maximize the function $f(x,y) = x^{2/3}y^{1/3}$ over the domain

$$A = \{(x,y) \in \mathbb{R}^2 : x, y \geq 0, 2x+3y \leq 6000\}.$$

This domain A is a closed triangular region in $\mathbb{R}^2$. Since A is compact and f is continuous on A, it follows by the global extreme value theorem that f has a maximum on A. This maximum point $p \in A$ occurs either on the interior or the boundary.

The interior is given by

$$A^{\circ} = \{(x, y) \in \mathbb{R}^2 : x, y > 0, 2x + 3y < 6000\}.$$

and $f(x, y) = x^{2/3} y^{1/3}$ is C^1 on this open set. If the maximum of f occurs at (x, y) inside A°, then by the local extreme value theorem, it must be satisfy $\nabla f(x, y) = (0, 0)$ so

$$\frac{2}{3} x^{-1/3} y^{1/3} = 0, \quad \frac{1}{3} x^{2/3} y^{-2/3} = 0,$$

This implies $x = y = 0$ but $(0, 0)$ does not belong to A°. Hence, the maximum of f on A cannot occur inside A°.

The boundary of A can be written as a union of three line segments $L_1 \cup L_2 \cup L_3$, where

$$\begin{aligned} L_1 &= \{(x, 0) : 0 \le x \le 3000\}, \\ L_2 &= \{(0, y) : 0 \le y \le 2000\}, \\ L_3 &= \{(x, y) \in \mathbb{R}^2 : x, y \ge 0, 2x + 3y = 6000\}. \end{aligned}$$

You must optimize f on each closed line segment. You can parametrize them or apply the method Lagrange multipliers. In this case, the calculations are straightforward for any of them. For L_1 and L_2, notice $f(x, y) = 0$ for all $(x, y) \in L_1 \cup L_2$, so the maximum value of f on these two line segments is 0. For L_3, notice $L_3 = L$ from Example 5.7.1 so it follows that the maximum value of f on L_3 is ≈ 1387.

Combining all of these observations of the interior and boundary of A, you can conclude that indeed the maximum production occurs when CATS4EVER spends their entire budget of 6 million dollars and they should spend exactly as suggested at the end of Example 5.7.1 to produce 1,387,000 bottles of Artemis Azure.

Constraints may not necessarily define compact regions, so you may need some limiting information to solve such optimization problems.

Example 5.7.4 Suppose you want to find the point on the set $x^2 - 2xy + y^2 - x + y = 0$ with minimum distance to the point $(1, 2, -3)$. This problem is equivalent to minimizing the square of the distance function

$$f(x, y, z) = (x - 1)^2 + (y - 2)^2 + (z + 3)^2$$

on the set $S = \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) = 0\}$ where

$$g(x, y, z) = x^2 - 2xy + y^2 - x + y.$$

Since $g : \mathbb{R}^3 \to \mathbb{R}$ is continuous, the set $S = g^{-1}(\{0\})$ is closed. Notice S is unbounded since, for example, $g(0, 0, z) = 0$ for all $z \in \mathbb{R}$. (This means you cannot directly apply the extreme value theorem to show the existence of a minimum.) On the other hand, if $||(x, y, z)|| \to \infty$, then one of $|x|, |y|$, or $|z|$ must tend to ∞ which implies that $f(x, y, z) \to \infty$. Thus,

$$f(x, y, z) \to \infty \qquad \text{as } ||(x, y, z)|| \to \infty \text{ with } (x, y, z) \in S.$$

Therefore, since S is closed and unbounded and f is continuous, this implies by a lemma similar to Lemma 2.8.14 that f attains a minimum on S.

Now, you can verify that the method of Lagrange multipliers (Theorem 5.6.8) applies in this situation. Thus, the minimum of f on S must satisfy the Lagrange multipliers system given by

$$\begin{bmatrix} 2(x-1) \\ 2(y-2) \\ 2(z+3) \end{bmatrix} = \lambda \begin{bmatrix} 2x-2y-1 \\ -2x+2y+1 \\ 0 \end{bmatrix} \quad ; \quad x^2-2xy+y^2-x+y=0$$

for some $x, y, z \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. Although it may be tedious, you can solve this system by hand and find that the only solutions are $(x, y, z) = (\frac{3}{2}, \frac{3}{2}, -3)$ with $\lambda = -1$, and $(x, y, z) = (2, 1, -3)$ with $\lambda = 2$.

Combining all of these observations, the minimum of f on S occurs either at $(\frac{3}{2}, \frac{3}{2}, -3)$ or at $(2, 1, -3)$. Note $f(\frac{3}{2}, \frac{3}{2}, -3) = 0.5$ and $f(2, 1, -3) = 2$, so the minimum of f on S occurs at $(\frac{3}{2}, \frac{3}{2}, -3)$. In other words, the point on the surface S closest to $(1, 2, -3)$ is $(\frac{3}{2}, \frac{3}{2}, -3)$.

These examples on multivariable optimization serve as a landmark for how much you have achieved over several chapters. You developed a substantial theory of differential calculus for analyzing a large class of optimization problems. Namely, you can show the existence of an optimal solution (Chapter 2), locate possible extrema on the interior of a domain or a parametrized boundary (Chapters 3 and 4), and search for solutions on an implicitly defined boundary (Chapter 5). While there are many more optimization techniques beyond this text, this concludes our story on *globally* optimal solutions.

There remains one lingering question on *locally* optimal solutions. Locally optimal solutions can be valuable. For example, if you have no hope of identifying the global extrema, then you may be satisfied if you at least know your solution is locally optimal. Now, from the local extreme value theorem, you can identify possible extrema on the interior of a function's domain. If the function is differentiable, then these critical points must have vanishing gradient. It is possible that the critical point corresponds to a local minimum, a local maximum, or a saddle point. This creates a new question for multivariable optimization.

> *Given a critical point of the form $\nabla f(p) = 0$, can you classify whether p corresponds to a local minimum, a local maximum, or a saddle?*

The gradient (and hence a linear approximation) alone cannot distinguish between local minima, local maxima, or saddles. You will need higher order derivatives (and hence higher order approximations) to classify local extrema. This desire for higher order approximations launches you into the next chapter.

Exercises for Section 5.7

Concepts and definitions

5.7.1 The basic optimization problem is often described as:

> *Let f be a real-valued function on a set $S \subseteq \mathbb{R}^n$. Find the global extrema of f on S.*

There is not enough information about f or S, but you can still ask yourself some key questions:

> *Does a solution exist? Where might a solution exist? What tools can I use to find possible solutions? How do I conclude that I have found the optimum?*

Use these guiding questions to formulate a general strategy for what you need to do. (Do not worry about the details or specific assumptions; just record some possible ideas.)

5.7.2 You want to maximize a function $f : \mathbb{R}^2 \to \mathbb{R}$ on the closed disk $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$.

(a) First, you find that f is C^1 on $\mathbb{R}^2$. What have you learned about the maximum of f on D?

(b) Second, you find that $\nabla f(x, y) = (0, 0)$ if and only if $(x, y) = (1, 1)$ or $(0, 0)$. What have you learned about the maximum of f on D?

(c) Third, you find that the only solutions to the system

$$\nabla f(x, y) = \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix} \qquad x^2 + y^2 = 4$$

are $(x, y) = (\sqrt{2}, \sqrt{2})$ with $\lambda = 3$, and $(x, y) = (0, 2)$ with $\lambda = 2$. What have you learned about the maximum of f on D?

(d) Fourth, you compute that $f(0, 0) = 137$, $f(\sqrt{2}, \sqrt{2}) = 223$, $f(1, 1) = 237$, $f(0, 2) = 334$. What have you learned about the maximum of f on D?

5.7.3 Arthur and Cameron want to optimize a C^1 function $f : \mathbb{R}^3 \to \mathbb{R}$ over the sphere

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

By the extreme value theorem, f has a maximum and a minimum on S. However, S has no interior points so the local extreme value theorem is not immediately useful.

(a) Arthur is obsessed with parametrizing. Outline his strategy along with parametrizations needed. Draw pictures to support your explanation.

(b) Cameron is obsessed with Lagrange multipliers. Outline his strategy along with the constraint and domain choices. Draw pictures to support your explanation.

(c) Whose strategy would you rather execute? Briefly explain why.

5.7.4 Amy and Victoria want to optimize a C^1 function $f : \mathbb{R}^3 \to \mathbb{R}$ over the hemisphere

$$H = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \geq 0\}.$$

By the extreme value theorem, f has a maximum and a minimum on H. However, H has no interior points so the local extreme value theorem is not immediately useful.

(a) Amy is obsessed with parametrizing. Outline her strategy along with parametrizations needed. Draw pictures to support your explanation.

(b) Victoria is obsessed with Lagrange multipliers. Outline her strategy along with the constraint and domain choices. Draw pictures to support your explanation.

(c) Whose strategy would you rather execute? Briefly explain why.

5.7.5 Assume you have a function $f : \mathbb{R}^3 \to \mathbb{R}$ that is C^1. You want to find the maximum of f on

$$R = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1, z \geq 0\}.$$

You can verify that R is compact. As f is C^1 and hence continuous, the extreme value theorem implies f has a maximum on R.

(a) To optimize on R, you must divide R into at least four subsets. Sketch R and define the subsets.

(b) For each subset, briefly describe possible strategies needed to find the maximum. If possible, provide more than one option.

Computations

5.7.6 Define $f(x, y) = xy + 2y$.

(a) Use two distinct approaches to find the maximum of f on

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

(b) Find the maximum of f on the set

$$B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

5.7.7 A chocolate company has a unique design for their chocolates, which come in a box shaped like a triangular prism. To appeal to environmentally concerned stakeholders, they want to develop new packaging that uses less material. If they keep an equilateral triangle face, and their package must contain 250cm^3 of volume, what dimensions should their box have to minimize the amount of material used?

5.7.8 A company sells phones and laptops. It costs them

$$C(x, y) = 100x^{0.6} + 60y^{0.5} + 2,000,000$$

thousand dollars to produce x phones and y computers. They sell each device for \$1,000. Currently, they are spending \$12,000,000 thousand (\$12 billion) on producing these two products.

(a) By solving a Lagrange multiplier problem, they calculated a multiplier λ and chose values x, y accordingly to maximize their profit. Set up the system and, using a computer algebra system, determine their choices x, y, and λ.

(b) If they increase their production budget to \$13 billion, estimate how much their maximum possible profit increases using the solution from the previous part. Do not solve another Lagrange multiplier system. *Hint:* Interpret the Lagrange multiplier.

(c) How much does the maximum profit actually increase? Compare with your estimate above.

5.7.9 A company has a production function with three inputs x, y, and z given by

$$f(x, y, z) = 32x^{1/4}y^{1/2}z^{1/4}.$$

The total budget is \$36,000 but they do not necessarily want to spend all of it. The company can buy x, y, and z at \$9, \$30, and \$18 per unit, respectively. What combination of inputs will maximize production?

5.7.10 Find the minimum value of $\sum_{i=1}^{n} x_i^2$ subject to the constraint that $\sum_{i=1}^{n} x_i = 1$.

Proofs

5.7.11 You want to maximize $f(x,y) = x^2+y^2+4y$ on the closed disk $A = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 16\}$. Semeon writes an argument. All of his derivatives and algebraic manipulations are correct.

1. *Since f is C^1 and $\nabla f(x,y) = (2x, 2y+4)$, the critical points of f occur precisely when $(2x, 2y+4) = (0,0)$.*
2. *Hence, $(x,y) = (0,-2)$ is the only critical point of f on A.*
3. *The boundary of A is the circle $\partial A = \{(x,y) \in \mathbb{R}^2 : g(x,y) = 16\}$ where $g(x,y) = x^2 + y^2$.*
4. *The system of equations*

$$\nabla f(x,y) = \lambda \nabla g(x,y) \qquad g(x,y) = 16$$

with $x, y, \lambda \in \mathbb{R}$ is equivalent to

$$2x = 2\lambda x, \qquad 2y + 4 = 2\lambda y, \qquad x^2 + y^2 = 16.$$

5. *The first equation implies $2x(\lambda - 1) = 0$ so $\lambda = 1$ or $x = 0$.*
6. *If $\lambda = 1$, then the second equation implies $4 = 0$ so there are no solutions in this case.*
7. *If $x = 0$, then the third equation implies $y^2 = 16$ so $y = \pm 4$.*
8. *Thus, the only solutions to the system are $(x,y) = (0, \pm 4)$ with corresponding λ values.*
9. *Thus, the maximum of f on ∂A must occur at one of $(0,4)$ or $(0,-4)$.*
10. *Overall, the maximum of f on A must occur at one of $(0,-2), (0,4)$ or $(0,-4)$.*
11. *By direct computation, $f(0,-2) = -4, f(0,4) = 32$ and $f(0,-4) = 0$,*
12. *The maximum of f on A is therefore 32.*

Semeon's argument does not have a serious logical error, but it is missing important justifications.

(a) Identify which line(s) require the global extreme value theorem.

(b) Identify which line(s) require the local extreme value theorem.

(c) Identify which line(s) require the method of Lagrange multipliers. What key assumption was not verified?

(d) Justify Line 10 more carefully. Use theorems and previous lines if needed.

5.7.12 Let A be a $m \times n$ matrix. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = ||Ax||^2$.

(a) Justify that $f(x)$ has a maximum subject to the constraint $||x|| = 1$.

(b) Show that $\nabla f(x) = 2A^T Ax$. *Hint:* Use chain rule.

(c) Define $g : \mathbb{R}^n \to \mathbb{R}$ by $g(x) = ||x||^2$. Prove that $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ is a solution to the system

$$\nabla f(x) = \lambda \nabla g(x) \qquad g(x) = 1$$

if and only if x is a normalized eigenvector of $A^T A$ with eigenvalue λ.

(d) Conclude that the maximum of $f(x)$ subject to the constraint $||x|| = 1$ is equal to the maximum eigenvalue of $A^T A$.

Applications and beyond

5.7.13 There are many other optimization methods that stem from ideas you have already seen. You will investigate **linear programming** – an area of mathematics that focuses on optimizing linear functions.

(a) The set S which consists of the triangle in $\mathbb{R}^2$ with vertices $(0,0)$, $(1,0)$, and $(0,1)$ and all points inside the triangle can be described as the intersection of the regions $\{(x,y) : x \geq 0\}$, $\{(x,y) : y \geq 0\}$, and $\{(x,y) : x + y \leq 1\}$. Fix $m, n \in \mathbb{R}$. Show that maximum of any linear function $f(x,y) = mx + ny$ on S must occur at one of the vertices of the triangle.

(b) Let $S \subseteq \mathbb{R}^2$ be the pentagon with vertices $(2,0), (1,1), (0.5,-1), (-0.5,-1)$, and $(-2,0)$, including the points inside the pentagon. Describe S in terms of regions of the form $\{(x,y) \in \mathbb{R}^2 : ax + by \leq c\}$ for some $a, b, c \in \mathbb{R}$. Then, show that any linear function f on S must have its maximum occur at one of the vertices of the pentagon.

(c) Let S be any convex polygon in $\mathbb{R}^2$ and f a linear function on S. Come up with a conjecture about where the maximum and minimum values of f occur on S.

(d) The general problem can be stated in a very similar way: *find the maximum value of a linear function f on a convex polytope in $\mathbb{R}^n$.* George Dantzig devised the **simplex method** for solving this type of problem. This method gives an algorithm for searching the vertices of a polytope to find the optimal solution instead of just checking each vertex. Why is it better to have an algorithm which searches the vertices, instead of simply computing that value of f on each vertex? *Hint:* Think of n as large.

6. Approximations

Linear approximations have been a smashing success for optimizing real-valued functions $\mathbb{R}^n \to \mathbb{R}$. From the local extrema value theorem, you can identify possible local extrema on the interior of a function's domain. If the function is differentiable, then these critical points must have vanishing gradient. In other words, the linear approximation is a *constant* function at critical points. Critical points can correspond to a local minimum, a local maximum, or a saddle point yet the linear approximation does not distinguish between them! This leads to this chapter's core question.

> *Given a critical point of the form $\nabla f(p) = 0$, can you classify whether p corresponds to a local extrema? If so, of what type?*

The gradient (and hence a linear approximation) alone cannot distinguish between local minima, local maxima, or saddles. You will need higher order derivatives (and hence higher order approximations) to better approximate your function and classify local extrema.

To approximate nonlinear functions, you will first need to answer a fundamental question.

> *What functions are good to use for higher order approximations?*

From a computational perspective, multivariable polynomials are a natural candidate. You can calculate them rapidly with basic addition and multiplication. Polynomials are easy to use but it will take a bit of notation to describe them fully. Once you have become comfortable with these functions, you will need to answer another fundamental question.

> *How can I determine the quality of my approximation?*

Limits and single variable calculus concepts will be revived to handle this question. Once these two questions are addressed, you will have one remaining concern.

> *How do I choose the best polynomial approximation?*

This final question will ultimately lead to multivariable Taylor's theorem and consequently, a resolution for classifying critical points.

6.1. Second order derivatives and the Hessian

First order partial derivatives contain information on the rate of change of a function. These are used to linearly approximate a function. If you want more refined information, you will want to know the rate of change of the *partial derivatives* themselves. In other words, you want a *partial derivative of a partial derivative*. This will presumably give you a better picture of the function itself and motivates the focus of this section.

6.1.1 Definition

Second order derivatives capture the idea of "partial derivative of a partial derivative".

Definition 6.1.1 Let $U \subseteq \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}^m$ be C^1. Fix $i, j \in \{1, \dots, n\}$ and $a \in U$. The **second order partial derivative** $\partial_i \partial_j f$ at a is defined by $\partial_i \partial_j f(a) := \partial_i(\partial_j f)(a)$, provided i^{th} partial of $\partial_j f$ exists at a. If $i \neq j$ then the partial is **mixed** and if $i = j$ then the partial is **pure**.

Remark 6.1.2 For pure partials, you can write ∂_i^2 instead of $\partial_i \partial_i$. For mixed partials, the order may matter, so $\partial_i \partial_j$ is not necessarily the same as $\partial_j \partial_i$. Replacing x_i with x and x_j with y, here are other pieces of equivalent notation

$$f_{xx} = (f_x)_x = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} \qquad f_{yy} = (f_y)_y = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2}$$

$$f_{xy} = (f_x)_y = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} \qquad f_{yx} = (f_y)_x = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y}$$

Pay close attention to the parentheses for mixed partials.

For concreteness, this definition can be fully expanded in terms of limits.

Example 6.1.3 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be some C^1 function. Fix $(a, b) \in \mathbb{R}^2$. Set $g = \partial_2 f : \mathbb{R}^2 \to R$. Then $\partial_1 g(a) = \partial_1 \partial_2 f(a)$ is given by

$$\partial_1 \partial_2 f(a) = \lim_{h \to 0} \frac{\partial_2 f(a+h, b) - \partial_2 f(a, b)}{h}.$$

Computing second order partial derivatives is usually quite straightforward.

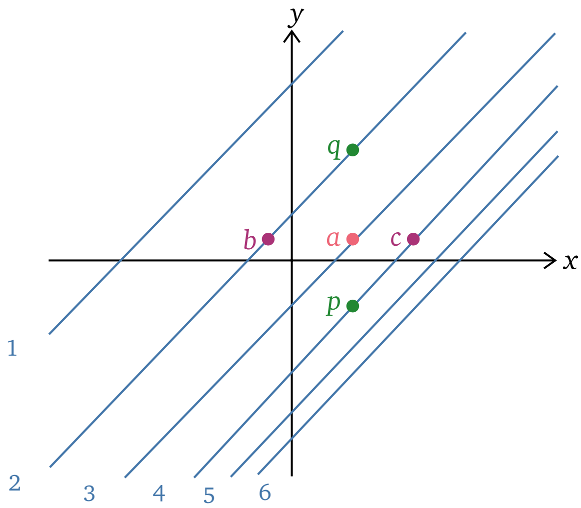
Example 6.1.4 Define $g : \mathbb{R}^2 \to \mathbb{R}$ as $g(x, y) = 3x^3 + 4x^2 y + 7xy + y^2$ for all $(x, y) \in \mathbb{R}^2$. You can compute $\partial_i \partial_j g$ for $i, j \in \{1, 2\}$. Remember that when taking partial derivatives, you keep all other variables constant except the differentiating variable. This gives

$$\begin{aligned}
\partial_1^2 g(x,y) &= \frac{\partial^2}{\partial x^2}\left(3x^3 + 4x^2 y + 7xy + y^2\right) = \frac{\partial}{\partial x}\left(9x^2 + 8xy + 7y\right) = 18x + 8y, \\
\partial_1 \partial_2 g(x,y) &= \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}\left(3x^3 + 4x^2 y + 7xy + y^2\right)\right] = \frac{\partial}{\partial x}\left[4x^2 + 7x + 2y\right] = 8x + 7, \\
\partial_2 \partial_1 g(x,y) &= \frac{\partial}{\partial y}\left[\frac{\partial}{\partial x}\left(3x^3 + 4x^2 y + 7xy + y^2\right)\right] = \frac{\partial}{\partial y}\left[9x^2 + 8xy + 7y\right] = 8x + 7, \\
\partial_2^2 g(x,y) &= \frac{\partial^2}{\partial y^2}\left(3x^3 + 4x^2 y + 7xy + y^2\right) = \frac{\partial}{\partial y}\left(4x^2 + 7x + 2y\right) = 2.
\end{aligned}$$

Notice that $\partial_1 \partial_2 g = \partial_2 \partial_1 g$ but this is not obvious a priori.

The geometry of second order partial derivatives is much more subtle due to mixed partials.

Example 6.1.5 Below is a contour map of a C^2 function $f : \mathbb{R}^2 \to \mathbb{R}$ and a fixed point $a \in \mathbb{R}^2$. Four points nearby a are also included, namely $b, c, p, q \in \mathbb{R}^2$.



The signs of the partial derivatives of a can be guessed from this contour map.

- What are the signs of $f_x(a)$ and $f_y(a)$? For $f_x(a)$, you can see as you move from left to right through a that contours increase so $f_x(a) > 0$. Similarly, as you move from bottom to top through a that the contours decrease so $f_y(a) < 0$.
- What are the signs of $f_{xx}(a)$ and $f_{yy}(a)$? For $f_{xx}(a)$, as you move from left to right through a, the contours are getting closer and closer while rising in height; in other words, $f_x(b) < f_x(a) < f_x(c)$. This suggests that $f_{xx}(a) > 0$. Similarly, as you move from bottom to top through a, the contours are getting farther apart while lowering in height; in other words, $f_y(p) < f_y(a) < f_y(q)$. This suggests that $f_{yy}(a) > 0$.
- What are the signs of $f_{xy}(a)$ and $f_{yx}(a)$? Mixed partials are always much trickier.

  First consider $f_{yx}(a)$. The quantity $f_{yx} = (f_y)_x$ measures how the y-slope of f (namely f_y) changes when you move in the x-direction. That is,

$$f_{yx}(a) = \lim_{h\to 0} \frac{f_y(a + he_1) - f_y(a)}{h}.$$

  The diagram can be used to observe how f_y changes for points nearby a in the x-direction, namely points b and c. The diagram suggests that $f_y(b) > f_y(a) > f_y(c)$ so the y-slope of f is decreasing as you move in the x-direction, or equivalently

$$\frac{f_y(a + he_1) - f_y(a)}{h} < 0$$

  for small values of $h \approx 0$. Overall, you can reasonably infer that $f_{yx}(a) < 0$.

  Second, consider $f_{xy}(a)$. The quantity $f_{xy} = (f_x)_y$ measures how the x-slope of f (namely f_x) changes when you move in the y-direction. That is,

$$f_{xy}(a) = \lim_{h\to 0} \frac{f_x(a + he_2) - f_x(a)}{h}.$$

The diagram can be used to observe how f_x changes for points nearby a in the y-direction, namely points p and q. The diagram suggests that $f_x(p) > f_x(a) > f_x(q)$, so the x-slope of f is decreasing as you move along in the y-direction, or equivalently

$$\frac{f_x(a+he_2)-f_x(a)}{h} < 0$$

for small values of $h \approx 0$. Overall, you can reasonably infer that $f_{xy}(a) < 0$.

The intricacy of this geometry might make you believe that mixed partials have no relationship whatsoever. As you shall surprisingly see, mixed partials are usually equal!

6.1.2 Clairaut's theorem

In general, mixed partials are not necessarily related.

Example 6.1.6 It sometimes may be the case that a function has second order partial derivatives at every point, but its mixed partials do not equal. Define

$$f(x,y) = \begin{cases} \frac{x^3y}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}.$$

You can directly compute the rest of its first order partials and find that

$$\partial_1 f(x,y) = \begin{cases} \frac{x^4y+3x^2y^3}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases} \qquad \partial_2 f(x,y) = \begin{cases} \frac{x^5-x^3y^2}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

Note the partial derivatives at $(0,0)$ were computed using the definition of the partial derivative. For example,

$$\partial_1 f(0,0) = \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{h^3 \cdot 0}{h^3} = 0$$

and likewise for $\partial_2 f(0,0) = 0$. Now, you can compute $\partial_1\partial_2 f(0,0)$ and $\partial_2\partial_1 f(0,0)$ by the limit definition. Namely,

$$\partial_1\partial_2 f(0,0) = \lim_{h\to 0} \frac{\partial_2 f(0+h,0)-\partial_2 f(0,0)}{h} = \frac{h^5-h^3\cdot 0}{h^5} = \lim_{h\to 0} 1 = 1.$$

and similarly you can verify that $\partial_2\partial_1 f(0,0) = 0$. Therefore,

$$\partial_2\partial_1 f(0,0) = 0 \neq 1 = \partial_1\partial_2 f(0,0).$$

Play with this Math3D demo to visually observe this phenomenon.

You can miraculously avoid this dire situation with a mild assumption.

Definition 6.1.7 Let $U \subseteq \mathbb{R}^n$ be an open set. A function $f : U \to \mathbb{R}^m$ is **twice continuously differentiable** (or C^2) provided for all $i,j \in \{1,\ldots,n\}$, the second partials $\partial_i\partial_j f$ exist and are continuous everywhere in U.

Theorem 6.1.8 (Clairaut) Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \to \mathbb{R}^m$. If f is C^2, then

$$\forall i, j \in \{1, \dots, n\}, \qquad \partial_i \partial_j f = \partial_j \partial_i f.$$

Proof. The proof is postponed to Section 6.1.4. ■

Informally speaking, Clairaut's theorem encapsulates a natural property.

Mixed partials commute[1].

Clairaut's theorem is brilliant because it is so simple to check that a function is C^2.

Example 6.1.9 Consider again $g(x, y) = 3x^3+4x^2y+7xy+y^2$. From Example 6.1.4, you can see that all 4 second order partial derivatives $\partial_i \partial_j g$ for $i, j \in \{1, 2\}$ exist and are continuous on all of $\mathbb{R}^2$. Thus, g is C^2 Clairaut's theorem proves that $\partial_1 \partial_2 g = \partial_2 \partial_1 g$. You checked this manually already but this is a good confirmation.

Similar to C^1 functions, C^2 functions possess basic properties similar to those in Theorem 3.6.6. There are also standard examples of C^2 functions.

Example 6.1.10 Every polynomial is C^2 because every polynomial C^1 and their partial derivatives are themselves polynomials. Hence, their partials are C^1 so every polynomial is C^2.

Example 6.1.11 Every linear map is C^2 since the partials of a linear map are constant maps and constant maps are C^1. Hence, their partials are C^1 so every linear map is C^2.

Clairaut's theorem can be used to find elegant identities for iterated directional derivatives.

Lemma 6.1.12 Let $U \subseteq \mathbb{R}^n$ be open. If $f : U \to \mathbb{R}^m$ is C^2, then for $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ and $p \in U$,

$$D_h^2 f(p) := D_h(D_h f)(p) = \sum_{i=1}^{n} h_i^2 (\partial_i^2 f)(p) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2h_i h_j (\partial_i \partial_j f)(p).$$

Proof. This is left as an exercise. Use Clairaut's theorem and Theorem 3.3.10. ■

This identity will be surprisingly useful in the proof of Taylor's theorem.

6.1.3 Hessian matrix

Now, to later construct a second order approximation, it will be convenient to introduce a matrix of the second partials.

Definition 6.1.13 Let f be a real-valued function which is C^2 at $a \in \mathbb{R}^n$.
The **Hessian of f at a** is the $n \times n$ matrix $Hf(a)$ defined by

$$Hf(a) = \left[\partial_i \partial_j f(a)\right]_{i,j}.$$

Remark 6.1.14 By Clairaut's theorem, the Hessian is a symmetric matrix.

Example 6.1.15 Define $f : \mathbb{R}^3 \to \mathbb{R}$ as $f(x, y, z) = ze^{xy}$ for all $(x, y, z) \in \mathbb{R}^3$. By 6.1.13, the

[1] The word "commute" in mathematics means that the order of operations does not affect the result.

Hessian of f is given by

$$Hf = \begin{bmatrix} \partial_1^2 f & \partial_1\partial_2 f & \partial_1\partial_3 f \\ \partial_2\partial_1 f & \partial_2^2 f & \partial_2\partial_3 f \\ \partial_3\partial_1 f & \partial_3\partial_2 f & \partial_3^2 f \end{bmatrix}.$$

Since f is C^2 then $\partial_i\partial_j f = \partial_j\partial_i f$ for $i,j \in \{1,2,3\}$ by by 6.1.8. This means that you only need to compute six second order partial derivatives instead of nine. You can verify that

$$Hf(x,y,z) = \begin{bmatrix} y^2ze^{xy} & ze^{xy}+xyze^{xy} & ye^{xy} \\ ze^{xy}+xyze^{xy} & x^2ze^{xy} & xe^{xy} \\ ye^{xy} & xe^{xy} & 0 \end{bmatrix}.$$

Admittedly, the Hessian is not well motivated right now but you can at least familiarize yourself with its defintiion. You will later see that the Hessian describes quadratic approximations similar to how the Jacobian matrix describes linear approximations. Before proceeding to these approximations, you will generalize partial derivatives to any order.

6.1.4 Proof of Clairaut's theorem

To conclude the section, we provide the proof of Clairaut's theorem for the ambitious reader. It is a sleek combination of the single variable mean value theorem and symmetry. Initially it may seem mysterious but it is more natural than it appears. As you read, you are encouraged to consider the special case in $\mathbb{R}^2$ and draw a square with a bunch of labelled points corresponding to the proof's quantities.

Proof. Without loss, assume $f : U \to \mathbb{R}$ is real-valued. Fix $a \in U$. It suffices to show

$$\forall \varepsilon > 0, |\partial_i\partial_j f(a) - \partial_j\partial_i f(a)| < \varepsilon.$$

Fix $\varepsilon > 0$. Since f is C^2, there exists $\delta_{ij}, \delta_{ji} > 0$ such that

$$\begin{aligned} &\forall x \in \mathbb{R}^n, ||x-a|| < 2\delta_{ij} \implies |\partial_i\partial_j f(x) - \partial_i\partial_j f(a)| < \frac{\varepsilon}{2}, \\ &\forall x \in \mathbb{R}^n, ||x-a|| < 2\delta_{ji} \implies |\partial_j\partial_i f(x) - \partial_j\partial_i f(a)| < \frac{\varepsilon}{2}. \end{aligned} \tag{6.1.1}$$

Take $\delta := \min\{\delta_{ij}, \delta_{ji}\}$. Fix $h,k \in (0,\delta)$ and define the single-variable map $\phi : (-\delta,\delta) \to \mathbb{R}$ by

$$\phi(t) = f(a + te_i + ke_j) - f(a + te_i).$$

Since f is C^2 (and hence C^1), we may apply the single-variable mean value theorem to ϕ on the interval $[0,h]$. Thus, there exists $h^* \in (0,h)$ such that

$$\partial_i f(a+h^*e_i+ke_j) - \partial_i f(a+h^*e_i) = \frac{f(a+he_i+ke_j) - f(a+he_i) - f(a+ke_j) + f(a)}{h} \tag{6.1.2}$$

Define the single-variable map $\psi : (-\delta,\delta) \to \mathbb{R}$ by

$$\psi(t) = \partial_i f(a + h^*e_i + te_j) - \partial_i f(a + h^*e_i).$$

Since f is C^2, we may apply the single-variable mean value theorem to ψ on the interval $[0,k]$. Thus, there exists $k^* \in (0,k)$ such that

$$\begin{aligned} \partial_j\partial_i f(a+h^*e_i+k^*e_j) &= \frac{\partial_i f(a+h^*e_i+ke_j) - \partial_i f(a+h^*e_i)}{k} \\ &= \frac{f(a+he_i+ke_j) - f(a+he_i) - f(a+ke_j) + f(a)}{hk} \quad \text{by (6.1.2).} \end{aligned}$$

By symmetry, there exists $h^{**} \in (0,h)$ and $k^{**} \in (0,k)$ such that

$$\partial_i \partial_j f(a + h^{**}e_i + k^{**}e_j) = \frac{f(a+he_i+ke_j) - f(a+ke_j) - f(a+he_i) + f(a)}{hk}.$$

Setting $x = a + h^*e_i + k^*e_j$ and $y = a + h^{**}e_i + k^{**}e_j$, this implies that

$$\partial_i \partial_j f(x) = \partial_i \partial_j f(y).$$

Note x and y satisfy $||x-a|| < 2\delta \le 2\delta_{ij}$ and $||y-a|| < 2\delta \le 2\delta_{ji}$. It follows by the triangle inequality and (6.1.1) that

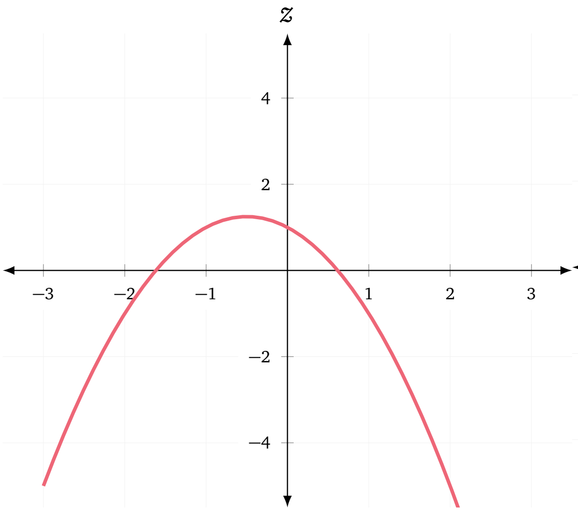
$$\begin{aligned}
|\partial_i\partial_j f(a) - \partial_j\partial_i f(a)| &\le |\partial_i\partial_j f(a) - \partial_i\partial_j f(x)| + |\partial_i\partial_j f(x) - \partial_j\partial_i f(y)| + |\partial_j\partial_i f(y) - \partial_j\partial_i f(a)| \\
&< \frac{\varepsilon}{2} + 0 + \frac{\varepsilon}{2} \\
&= \varepsilon
\end{aligned}$$

as required. ■

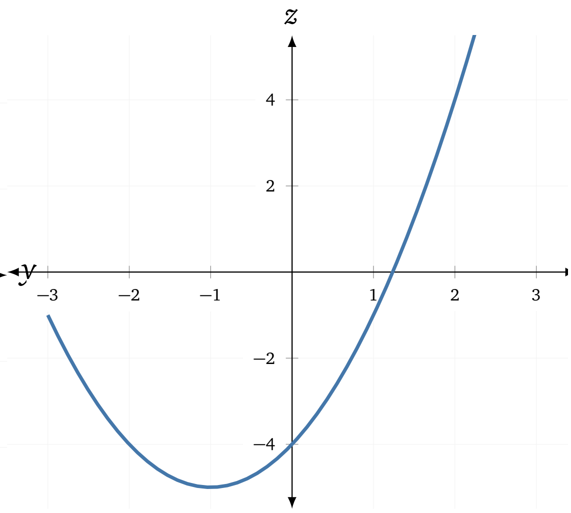
Exercises for Section 6.1

Concepts and definitions

6.1.1 Below are slices of the graph of a C^2 function $f : \mathbb{R}^2 \to \mathbb{R}$.



x-slice at $x = -1$



y-slice at $y = 2$

(a) If possible, determine whether $f(-1,2)$ is positive, negative, or zero.

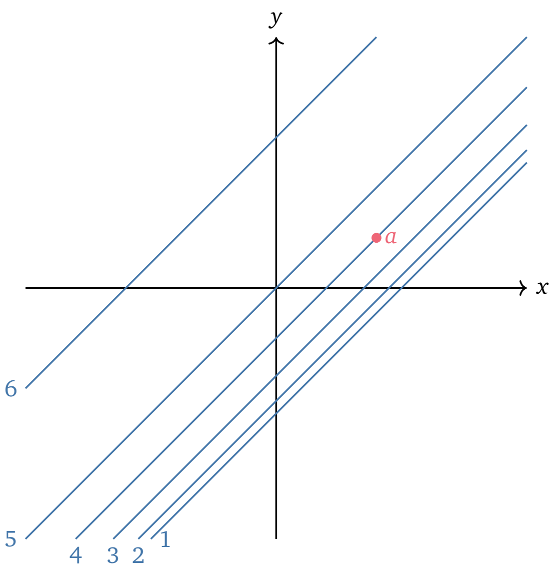
(b) If possible, determine whether each first order derivative is positive, negative, or zero.

i) $\dfrac{\partial f}{\partial x}(-1,2)$ ii) $\dfrac{\partial f}{\partial y}(-1,2)$

(c) If possible, determine whether each second order derivative is positive, negative, or zero.

i) $\dfrac{\partial^2 f}{\partial x^2}(-1,2)$ ii) $\dfrac{\partial^2 f}{\partial y^2}(-1,2)$ iii) $\dfrac{\partial^2 f}{\partial x \partial y}(-1,2)$ iv) $\dfrac{\partial^2 f}{\partial y \partial x}(-1,2)$

6.1.2 Below is a contour plot of the C^2 function $f : \mathbb{R}^2 \to \mathbb{R}$ and a point $a \in \mathbb{R}^2$.



(a) If possible, determine whether $f(a)$ is positive, negative, or zero.

(b) If possible, determine whether each first order derivative is positive, negative, or zero.

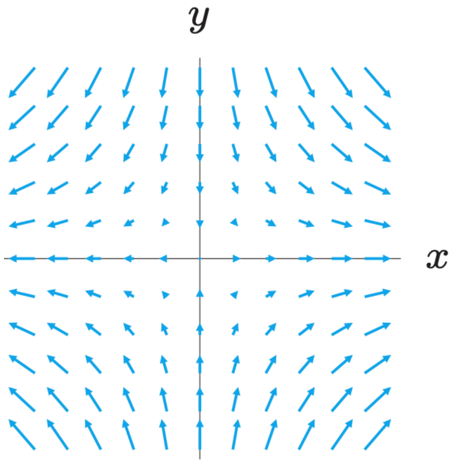
i) $\dfrac{\partial f}{\partial x}(a)$ ii) $\dfrac{\partial f}{\partial y}(a)$

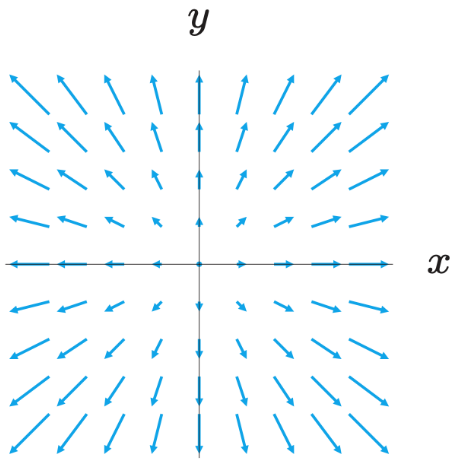
(c) If possible, determine whether each second order derivative is positive, negative, or zero.

i) $\dfrac{\partial^2 f}{\partial x^2}(a)$ ii) $\dfrac{\partial^2 f}{\partial y^2}(a)$

iii) $\dfrac{\partial^2 f}{\partial x \partial y}(a)$ iv) $\dfrac{\partial^2 f}{\partial y \partial x}(a)$

6.1.3 Below are gradient vector fields of $f : \mathbb{R}^2 \to \mathbb{R}$ (on the left) and $g : \mathbb{R}^2 \to \mathbb{R}$ (on the right). Both functions are C^2.

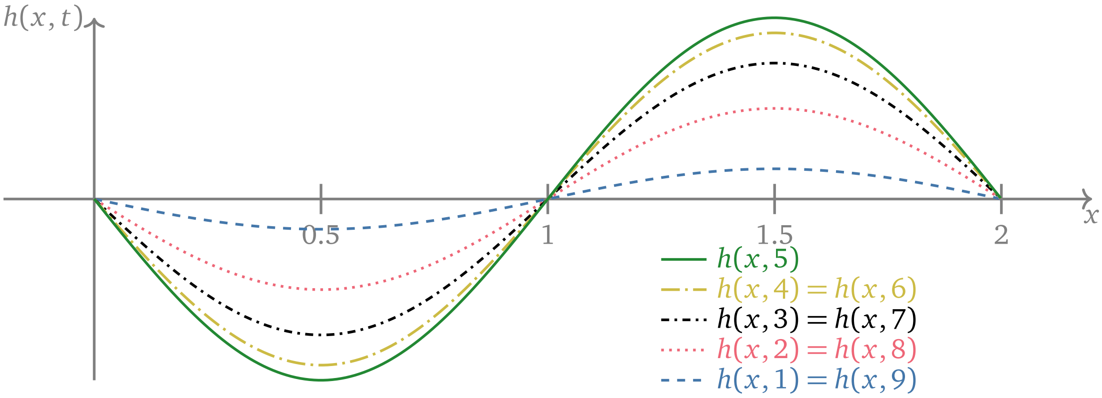




If possible, determine whether these quantities are positive, negative, or zero.

- $f_x(0,0)$
- $f_{xx}(0,0)$
- $f_y(0,0)$
- $f_{yy}(0,0)$
- $g_x(0,0)$
- $g_{xx}(0,0)$
- $g_y(0,0)$
- $g_{yy}(0,0)$

6.1.4 Below is a plot of a vibrating string $h : \mathbb{R}^2 \to \mathbb{R}$ as a C^2 function $h(x,t)$ for several times t.



(a) Determine whether each value is positive, negative, or zero.

i) $h(0.5,5)$ ii) $h(1,3)$

(b) Determine whether each first order partial is positive, negative, or zero.

i) $\dfrac{\partial h}{\partial x}(0.5,5)$ ii) $\dfrac{\partial h}{\partial t}(0.5,5)$ iii) $\dfrac{\partial h}{\partial x}(1,3)$ iv) $\dfrac{\partial h}{\partial t}(1,3)$

(c) Determine whether each pure second order partial is positive, negative, or zero.

i) $\dfrac{\partial^2 h}{\partial x^2}(0.5,5)$ ii) $\dfrac{\partial^2 h}{\partial t^2}(0.5,5)$ iii) $\dfrac{\partial^2 h}{\partial x^2}(1,3)$ iv) $\dfrac{\partial^2 h}{\partial t^2}(1,3)$

(d) Determine whether each mixed second order partial is positive, negative, or zero.

i) $\dfrac{\partial^2 h}{\partial x \partial t}(0.5,5)$ ii) $\dfrac{\partial^2 h}{\partial t \partial x}(0.5,5)$ iii) $\dfrac{\partial^2 h}{\partial x \partial t}(1,3)$ iv) $\dfrac{\partial^2 h}{\partial t \partial x}(1,3)$

Computations

6.1.5 Find all second order derivatives of $f(x,y)=x^2+xy^2+y^3+1$ and write the Hessian $Hf(x,y)$.

6.1.6 Find all second order derivatives of $g(x,y,z)=xye^{xz}$ and write the Hessian matrix $Hg(x,y,z)$.

6.1.7 Let $A,B,C\in\mathbb{R}$ and $q(x,y)=Ax^2+Bxy+Cy^2$. Compute the Hessian Hq and its determinant.

6.1.8 Let A be a $n\times n$ symmetric matrix. Let $q:\mathbb{R}^n\to\mathbb{R}$ be the quadratic form given by

$$q(x)=\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}x_ix_j.$$

Compute the Hessian Hq.

6.1.9 Define $f:\mathbb{R}^2\to\mathbb{R}$ by

$$f(x)=\begin{cases}\dfrac{xy(x^2-y^2)}{x^2+y^2} & (x,y)\neq 0,\\ 0 & (x,y)=0.\end{cases}$$

(a) Compute $\partial_1 f(x,y)$ and $\partial_2 f(x,y)$ for $(x,y)\neq(0,0)$.

(b) Compute $\partial_1 f(0,0)$ and $\partial_2 f(0,0)$ for $(x,y)=(0,0)$ from the definition.

(c) Compute $\partial_2\partial_1 f(0,0)$ and $\partial_1\partial_2 f(0,0)$ from the definition.

(d) How do these computations relate to Clairaut's theorem?

Proofs

6.1.10 Fix a point $p\in\mathbb{R}^2$. If $f:\mathbb{R}^2\to\mathbb{R}$ is C^2, then show for $h=(h_1,h_2)\in\mathbb{R}^2$ that

$$D_h^2 f(p):=D_h(D_hf)(p)=h_1^2(\partial_1^2 f)(p)+2h_1h_2(\partial_1\partial_2 f)(p)+h_2^2(\partial_2^2 f)(p).$$

6.1.11 Fix a point $p\in\mathbb{R}^3$. If $f:\mathbb{R}^3\to\mathbb{R}$ is C^2, then show for $h=(h_1,h_2,h_3)\in\mathbb{R}^3$ that

$$D_h^2 f(p):=D_h(D_hf)(p)=\sum_{i=1}^{3}h_i^2(\partial_i^2 f)(p)+\sum_{i=1}^{3}\sum_{j=i+1}^{3}2h_ih_j(\partial_i\partial_j f)(p).$$

Applications and beyond

6.1.12 Second order derivatives can be used to construct *quadratic* polynomial approximations at a given point. The key idea originates from Taylor polynomials in one variable.

> *A quadratic polynomial approximation $P(x)$ at a point $x=a$ should match the 0th, 1st, and 2nd derivatives of the function $f(x)$ at $x=a$.*

You will apply this idea to two-variable functions and discover the quadratic polynomial approximation. This motivates the definition of the Hessian. Let $f:\mathbb{R}^2\to\mathbb{R}$ be a C^2 function.

(a) A degree 0 polynomial in 2 variables is given by $P(x,y) = A$ for some constant $A \in \mathbb{R}$. If $P(0,0) = f(0,0)$, express A in terms of f.

(b) A degree 1 polynomial in 2 variables is given by $P(x,y) = A + Bx + Cy$ for some constants $A, B, C \in \mathbb{R}$. If

$$P(0,0) = f(0,0), \quad \partial_1 P(0,0) = \partial_1 f(0,0), \quad \partial_2 P(0,0) = \partial_2 f(0,0) \tag{6.1.3}$$

express A, B, C in terms of f.

(c) Prove that if P is a degree 1 polynomial satisfying (6.1.3) then $\displaystyle\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P(x,y)}{||(x,y)||} = 0.$

(d) A degree 2 polynomial in 2 variables is given by

$$P(x,y) = A + Bx + Cy + Dx^2 + Exy + Fy^2$$

for some constants $A, B, C, D, E, F \in \mathbb{R}$. If the zeroth, first, and second partials match f, i.e.

$$\begin{aligned} &P(0,0) = f(0,0), && \partial_1 P(0,0) = \partial_1 f(0,0), && \partial_2 P(0,0) = \partial_2 f(0,0), \\ &\partial_1^2 P(0,0) = \partial_1^2 f(0,0), && \partial_1\partial_2 P(0,0) = \partial_1\partial_2 f(0,0), && \partial_2^2 P(0,0) = \partial_2^2 f(0,0), \end{aligned} \tag{6.1.4}$$

then express A, B, C, D, E, F in terms of f.

(e) If P is the degree 2 polynomial satisfying (6.1.4), then conjecture a limit relationship with f similar to 6.1.12.3.

(f) If P is the degree 2 polynomial satisfying (6.1.4), then verify by direct computation that

$$P(x,y) = f(0,0) + \nabla f(0,0) \cdot (x,y) + \frac{1}{2}(x,y)^T Hf(0,0)(x,y).$$

The polynomial P is a *quadratic approximation* of f at $(0,0)$. The definition of the Hessian $Hf(0,0)$ is motivated by the above relationship with P.

6.1.13 Consider the vibrating string $h : \mathbb{R}^2 \to \mathbb{R}$ from Question 6.1.4. Explain what Clairaut's theorem says about the vibrating string using a full sentence with plain language only. Do not use any technical terms or mathematical symbols.

6.2. Higher order derivatives

Second order derivatives will help you construct quadratic approximations of a nonlinear map, which are better than linear approximations. To construct better and better approximations, you will need higher order derivatives. These will take some effort to calculate but the computations are mostly tedious computations. The primary challenge in this section is notational. By introducing the concise yet informative *multi-index notation*, you will be able to conveniently keep track of calculations. More importantly, you will become familiar with multivariable polynomials and seamlessly generalize many ideas from single variable calculus.

6.2.1 Generalized Clairaut's theorem

Motivated by C^1 and C^2 functions, you can introduce a definition that requires the existence and continuity of more partial derivatives.

Definition 6.2.1 Let $k \in \mathbb{N}^+$. Let $U \subseteq \mathbb{R}^n$ be an open set. A function $f : U \to \mathbb{R}^m$ is **k-times continuously differentiable** (or C^k) provided all of its k^{th} order partials exist and are continuous everywhere in U. That is, for all $i_1, \dots, i_k \in \{1, \dots, n\}$, the k^{th} **order partial derivative** $\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f$ exists and is continuous everywhere in U.

Remark 6.2.2 A function f is **smooth** (or C^∞) if it is C^k for every $k \in \mathbb{N}^+$.

Most functions that you encounter will be smooth on their domain. For instance, your favourite family of functions is continuously differentiable to any order.

Lemma 6.2.3 Every polynomial is C^∞.

Proof. Fix $n \geq 1$. We prove polynomials in n variables are C^k by induction on $k \in \mathbb{N}^+$. The base case $k = 1$ is Lemma 3.6.8. For $k \geq 2$, every polynomial is C^{k-1} by assumption. Any partial derivative of a polynomial p is also a polynomial, so for any $i_k \in \{1, \dots, n\}$ the partial derivative $\partial_{i_k} p$ is also C^{k-1}. Hence, you can take any $k-1$ additional partial derivatives $\partial_{i_1}\cdots\partial_{i_{k-1}}\partial_{i_k} p$ and this will be continuous by the induction assumption. This proves that p is C^k. ■

You can naturally extend Clairaut's theorem (Theorem 6.1.8) to hold for C^k functions, so mixed partials of order k commute if a function is C^k.

Theorem 6.2.4 (Generalized Clairaut) Let $U \subseteq \mathbb{R}^n$ be open. If $f : U \to \mathbb{R}^m$ is C^k, then the mixed partial derivatives of f up to order k commute. That is, for any integers $1 \leq i_1, i_2, \dots, i_k \leq n$ and any reordering $j_1, \dots, j_k$ of $i_1, \dots, i_k$,

$$\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f = \partial_{j_1}\partial_{j_2}\cdots\partial_{j_k} f.$$

A fully justified proof requires either an annoyingly careful induction or some group theory[2], so instead a sketch is provided.

Proof. (Sketch) The case $k = 1$ is trivial and $k = 2$ is Clairaut's theorem, so assume $k \geq 3$. The ordering $i_1, \dots, i_k$ can be rearranged into any other ordering $j_1, \dots, j_k$ by swapping pairs of consecutive indices enough times[3]. It therefore suffices to show that you can swap any pair of consecutive partials in an ordering of indices.

[2] See an introductory course on group theory.
[3] This algorithm is referred to as bubble sort in computer science.

For any $\ell \in \{2, \ldots, k-1\}$, the function $g = \partial_{i_{\ell+1}} \cdots \partial_{i_k} f$ is a partial derivative of f of order $k-\ell$. Hence, f is C^k implies g is C^ℓ and hence C^2. By Clairaut's theorem (Theorem 6.1.8), it follows that

$$\partial_{i_{\ell-1}} \partial_{i_\ell} (\partial_{i_{\ell+1}} \cdots \partial_{i_k} f) = \partial_{i_\ell} \partial_{i_{\ell-1}} (\partial_{i_{\ell+1}} \cdots \partial_{i_k} f)$$

Differentiating both sides by $\ell - 2$ partials $\partial_{i_1} \cdots \partial_{i_{\ell-2}}$ shows that you can swap any pair of consecutive partials in an ordering, as desired. ■

Higher order partial derivatives are straightforward albeit tedious to calculate.

Example 6.2.5 Define $g : \mathbb{R}^3 \to \mathbb{R}$ as $g(x,y,z) = x^3yz^2 + x^2y\cos(z) + zy^3 + xy + x$ for all $(x,y,z) \in \mathbb{R}^3$. Notice that g is a composition of C^∞ functions so g is C^∞ as well. By Theorem 6.2.4, all of the mixed partial derivatives of g of any order commute. You can verify this directly. For instance, you can check $\partial_1\partial_2\partial_3 g = \partial_3\partial_1\partial_2 g$. Notice that

$$\begin{aligned}
\partial_1\partial_2\partial_3 g(x,y,z) &= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left[\frac{\partial}{\partial z}\left(x^3yz^2 + x^2y\cos(z) + zy^3 + xy + x\right)\right]\right) \\
&= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left[2x^3yz - x^2y\sin(z) + y^3\right]\right) \\
&= \frac{\partial}{\partial x}\left(2x^3z - x^2\sin(z) + 3y^2\right) \\
&= 6x^2z - 2x\sin(z)
\end{aligned}$$

and moreover,

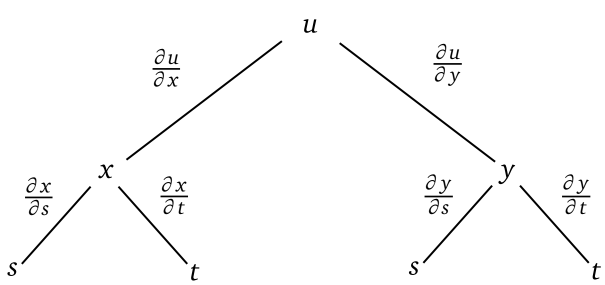
$$\begin{aligned}
\partial_3\partial_1\partial_2 g(x,y,z) &= \frac{\partial}{\partial z}\left(\frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}\left(x^3yz^2 + x^2y\cos(z) + zy^3 + xy + x\right)\right]\right) \\
&= \frac{\partial}{\partial z}\left(\frac{\partial}{\partial x}\left[x^3z^2 + x^2\cos(z) + 3zy^2 + x\right]\right) \\
&= \frac{\partial}{\partial z}\left(3x^2z^2 + 2x\cos(z) + 1\right) \\
&= 6x^2z - 2x\sin(z).
\end{aligned}$$

Thus, by direct calculation, $\partial_1\partial_2\partial_3 g = \partial_3\partial_1\partial_2 g$ as expected by generalized Clairaut.

6.2.2 Chain rule with higher derivatives

Applying the chain rule with higher derivatives is a mechanical and iterative process, but it quickly becomes quite a nightmare. Using variables instead of functions somewhat helps simplify the process.

Example 6.2.6 Let $u = u(x,y)$, $x = x(s,t)$ and $y = y(s,t)$. Assume all variables are C^2. You can derive a formula for $\frac{\partial^2 u}{\partial s^2}$ using the chain rule. Recall the chain rule tree.



By totaling the contribution of each branch from u to s, you can calculate $\frac{\partial u}{\partial s}$ as usual to obtain

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} \tag{6.2.1}$$

You can repeat this process with $\frac{\partial u}{\partial s}$ instead of u. In other words, to find $\frac{\partial^2 u}{\partial s^2}$, you use the chain rule tree again to apply $\frac{\partial}{\partial s}$ to the quantity $\frac{\partial u}{\partial s}$. There is, however, a new phenomenon! Each partial derivative term is itself a function, so you must also apply the product rule wherever required.

$$\begin{aligned}\frac{\partial^2 u}{\partial s^2} = \frac{\partial}{\partial s}\frac{\partial u}{\partial s} &= \frac{\partial}{\partial s}\left(\frac{\partial u}{\partial x}\frac{\partial x}{\partial s}\right) + \frac{\partial}{\partial s}\left(\frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right)\\ &= \frac{\partial^2 u}{\partial s\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial^2 u}{\partial s\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}\end{aligned}$$

Since u is C^2, you can write $\frac{\partial^2 u}{\partial s\partial x} = \frac{\partial}{\partial x}(\frac{\partial u}{\partial s})$ and $\frac{\partial^2 u}{\partial s\partial y} = \frac{\partial}{\partial y}(\frac{\partial u}{\partial s})$. You can again apply the chain rule from (6.2.1). By more applications of the product rule, this implies that

$$\begin{aligned}\frac{\partial^2 u}{\partial s^2} = \frac{\partial}{\partial x}&\left(\frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right)\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2}\\ &+ \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right)\frac{\partial y}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}.\end{aligned}$$

Notice that you will need to use the product rule on terms like $\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\frac{\partial x}{\partial s}\right)$ and $\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right)$. Since $\frac{\partial^2 x}{\partial x\partial s} = \frac{\partial}{\partial s}\frac{\partial x}{\partial x} = \frac{\partial}{\partial s}(1) = 0$ and $\frac{\partial^2 y}{\partial x\partial s} = \frac{\partial}{\partial s}\frac{\partial y}{\partial x} = \frac{\partial}{\partial s}(0) = 0$, some of the terms will disappear. Overall, after a few more small calculations, you will conclude that

$$\begin{aligned}\frac{\partial^2 u}{\partial s^2} = &\left(\frac{\partial^2 u}{\partial^2 x}\frac{\partial x}{\partial s} + \frac{\partial^2 u}{\partial x\partial y}\frac{\partial y}{\partial s}\right)\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2}\\ &+ \left(\frac{\partial^2 u}{\partial x\partial y}\frac{\partial x}{\partial s} + \frac{\partial^2 u}{\partial y^2}\frac{\partial y}{\partial s}\right)\frac{\partial y}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}.\end{aligned}$$

A similar formula holds for $\frac{\partial^2 u}{\partial t^2}$ and you can similarly derive another formula for $\frac{\partial^2 u}{\partial s\partial t}$. Although this was quite complicated, the goal was to demonstrate relationships between second order partials of u with respect to x, y compared to partials of u with respect to s, t.

Example 6.2.6 demonstrates that you can change the variables by which you express higher order partial derivatives. This ability can be quite handy in many situations, such as solving partial differential equations by changing coordinates systems from Cartesian to polar.

Example 6.2.7 If $u = f(x, y)$ for some C^2 function $f : \mathbb{R}^2 \to R$ and $(x, y) = (r\cos\theta, r\sin\theta)$, then you can verify by direct calculation that

$$\begin{aligned}\frac{\partial^2 u}{\partial r^2} &= \cos^2\theta\frac{\partial^2 u}{\partial^2 x} + 2\cos\theta\sin\theta\frac{\partial^2 u}{\partial x\partial y} + \sin^2\theta\frac{\partial^2 u}{\partial y^2},\\ \frac{\partial^2 u}{\partial\theta^2} &= r^2\sin^2\theta\frac{\partial^2 u}{\partial^2 x} - 2r^2\cos\theta\sin\theta\frac{\partial^2 u}{\partial x\partial y} + r^2\cos^2\theta\frac{\partial^2 u}{\partial y^2} - r\cos\theta\frac{\partial u}{\partial x} - r\sin\theta\frac{\partial u}{\partial y}.\end{aligned}$$

6.2.3 Multi-index notation and polynomials

Notation for higher order partials is already getting very messy. Clairaut's theorem for C^k functions provides a valuable convention for writing partials of C^k functions. Namely, for any integers $1 \leq i_1, \dots, i_k \leq n$ and C^k function f, you can always reorder the partials so that

$$\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f = \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_n^{\alpha_n} f,$$

where $\alpha_\ell \geq 0$ denotes the number of times that ∂_ℓ appears on the lefthand side. This implies that $\alpha_1+\cdots+\alpha_n = k$. Notice the righthand side does *not* depend on the ordering of $i_1, \dots, i_k$. The non-negative exponents $\alpha_1, \dots, \alpha_n$ can be packaged in a helpful piece of notation.

Definition 6.2.8 An element $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ is a **multi-index**. The **degree of** α is the non-negative integer $|\alpha| = \alpha_1 + \cdots + \alpha_n$. The **factorial** $\alpha!$ is the positive integer $\alpha_1!\alpha_2!\cdots\alpha_n!$.

Example 6.2.9 The element $\alpha = (2,0,5) \in \mathbb{N}^3$ is a multi-index with degree $|\alpha| = 2+0+5 = 7$ and $\alpha! = 2!0!5! = 240$.

The prior observations lead to an elegant rewriting of higher order derivatives.

Definition 6.2.10 Let $U \subseteq \mathbb{R}^n$ be an open set. Let $f : U \to \mathbb{R}^m$ be a C^k function For any multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ with degree $|\alpha| \leq k$, define the **α-partial derivative** by

$$\partial^\alpha f = \partial_1^{\alpha_1}\cdots\partial_n^{\alpha_n} f.$$

Calculations are easier to track and describe with this notation.

Example 6.2.11 Suppose you want to compute all partials of order up to 3 of $f(x,y) = x^2e^{xy}$. In other words, you want to compute $\partial^\alpha f$ for all multi-indices $\alpha \in \mathbb{N}^2$ with degree $|\alpha| \leq 3$. All possible multi-indexes of degree ≤ 3 are $(0,3),(1,2),(2,1),(3,0),(0,2),(1,1),(2,0),(1,0),(0,1)$ and $(0,0)$. For instance, if $\alpha = (0,3)$, then $\partial^\alpha f = \partial^{(0,3)} f$, so

$$\partial^{(0,3)} f(x,y) = \partial_2^3 f(x,y) = \frac{\partial^3}{\partial y^3} x^2 e^{xy} = x^5 e^{xy}.$$

Similarly,

$$\begin{aligned}
\partial^{(1,2)} f(x,y) &= \tfrac{\partial}{\partial x}\tfrac{\partial^2}{\partial y^2} x^2e^{xy} = 4x^3e^{xy} + x^4ye^{xy} \\
\partial^{(2,1)} f(x,y) &= \tfrac{\partial^2}{\partial x^2}\tfrac{\partial}{\partial y} x^2e^{xy} = e^{xy}(6x + 3x^2y + 3x^2y + x^3y^2) \\
\partial^{(3,0)} f(x,y) &= \tfrac{\partial^3}{\partial x^3} x^2e^{xy} = e^{xy}(6y + 6xy^2 + x^2y^3) \\
\partial^{(0,2)} f(x,y) &= \tfrac{\partial^2}{\partial y^2} x^2e^{xy} = x^4e^{xy} \\
\partial^{(1,1)} f(x,y) &= \tfrac{\partial^2}{\partial x\partial y} x^2e^{xy} = e^{xy}(3x^2 + x^3y) \\
\partial^{(2,0)} f(x,y) &= \tfrac{\partial^2}{\partial x^2} x^2e^{xy} = e^{xy}(2 + 4xy + x^2y^2) \\
\partial^{(1,0)} f(x,y) &= \tfrac{\partial}{\partial x} x^2e^{xy} = 2xe^{xy} + x^2ye^{xy} \\
\partial^{(0,1)} f(x,y) &= \tfrac{\partial}{\partial y} x^2e^{xy} = x^3e^{xy} \\
\partial^{(0,0)} f(x,y) &= x^2e^{xy}
\end{aligned}$$

Notice that $\partial^{(1,2)}$ and $\partial^{(2,1)}$ are not a reordering of derivatives; they are two completely different derivatives. That is, you can usually expect $\partial^{(2,1)} f(x,y) \neq \partial^{(1,2)} f(x,y)$.

Multi-indices are also the ideal notation for monomials.

Definition 6.2.12 Let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ be a multi-index and let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. The **monomial** x^α in the variables $x_1, \ldots, x_n$ is defined by

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

The monomial x^α has **degree** equal to $|\alpha| = \alpha_1 + \cdots + \alpha_n$ in the variables $x_1, \ldots, x_n$.

Example 6.2.13 Set $\alpha = (2, 0, 5) \in \mathbb{N}^3$ so $(x, y, z)^\alpha = x^2 y^0 z^5 = x^2 z^5$ has degree $|\alpha| = 7$ in the variables x, y, and z. Similarly $(x-1, y-2, z-3)^\alpha = (x-1)^2(z-3)^5$ has degree 7 in the variables $x-1, y-2$, and $z-3$.

The language of polynomials is perfectly suited for these new definitions.

Definition 6.2.14 A **polynomial** is a finite linear combination of monomials. The **degree of a polynomial** is the maximum of the degree of its monomials (with non-zero coefficients).

Example 6.2.15 The set of degree 3 monomials in 2 variables x, y is given by

$$\{(x, y)^\alpha : \alpha \in \mathbb{N}^2, |\alpha| = 3\} = \{x^3, x^2 y, x y^2, y^3\}.$$

The set of degree ≤ 3 monomials in 2 variables x, y is given by

$$\{(x, y)^\alpha : \alpha \in \mathbb{N}^2, |\alpha| \leq 3\} = \{1, x, y, x^2, xy, y^2, x^3, x^2 y, x y^2, y^3\}.$$

An arbitrary polynomial $P(x, y)$ of degree ≤ 3 in 2 variables x, y can be written as

$$\begin{aligned} P(x, y) = \sum_{\substack{\alpha \in \mathbb{N}^2 \\ |\alpha| \leq 3}} C_\alpha (x, y)^\alpha &= C_{(0,0)} + C_{(1,0)} x + C_{(0,1)} y + C_{(2,0)} x^2 + C_{(1,1)} xy + C_{(0,2)} y^2 \\ &\quad + C_{(3,0)} x^3 + C_{(2,1)} x^2 y + C_{(1,2)} x y^2 + C_{(0,3)} y^3 \end{aligned}$$

for some constants $C_\alpha \in \mathbb{R}$ with $\alpha \in \mathbb{N}^2$ and $|\alpha| \leq 3$. For instance, the polynomial $1 + x^2 + 2xy - 7xy^2$ has degree 3 and the polynomial $1 + x + y + xy$ has degree 2.

Monomials and polynomials are intimately connected with higher order partial derivatives. Excellent notation like multi-indices can illuminate these patterns much more clearly.

Lemma 6.2.16 Let $\alpha, \beta \in \mathbb{N}^n$ be any multi-indices. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = x^\beta$. Then

(a) If $\alpha = \beta$ then $\partial^\alpha f(x) = \alpha!$ for $x \in \mathbb{R}^n$.
(b) If $|\alpha| > |\beta|$ then $\partial^\alpha f(x) = 0$ for $x \in \mathbb{R}^n$.
(c) If $\alpha \neq \beta$ then $\partial^\alpha f(0) = 0$.

Proof. For (a), we proceed by induction on the degree $|\alpha|$. If $|\alpha| = 1$, then $\partial^\alpha = \partial_j$ and $x^\alpha = x_j$ for some $j \in \{1, \ldots, n\}$, so $\partial^\alpha f(x) = \partial_j(x_j) = 1 = \alpha!$. Let $|\alpha| \geq 2$ and assume the result holds for lower degrees. Without loss of generality, assume $\alpha_1 \neq 0$ so $(0, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n$ has degree $< |\alpha|$. By the induction hypothesis, it follows that

$$\begin{aligned} \partial^\alpha f(x) = \partial_1^{\alpha_1} \partial^{(0,\alpha_2,\ldots,\alpha_n)} (x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}) &= \partial_1^{\alpha_1}\big(x_1^{\alpha_1} \cdot \partial^{(0,\alpha_2,\ldots,\alpha_n)}(x_2^{\alpha_2} \cdots x_n^{\alpha_n})\big) \\ &= \partial_1^{\alpha_1}\big(x_1^{\alpha_1} \cdot \alpha_2! \cdots \alpha_n!\big) \\ &= \alpha_2! \cdots \alpha_n! \partial_1^{\alpha_1}(x_1^{\alpha_1}) = \alpha_1! \alpha_2! \cdots \alpha_n! = \alpha! \end{aligned}$$

as required. Claims (b) and (c) are left as exercises; both use a similar induction argument. ■

You can verify this lemma in an explicit example.

Example 6.2.17 You can check 6.2.16 for some explicit choices of α and β with $n=3$. Fix multi-indices $\alpha=(0,1,2)$ and $\beta=(1,1,0)$, so $|\beta|=2<3=|\alpha|$. Notice that

$$\partial^\alpha(x,y,z)^\alpha = \partial^{(0,1,2)}(x,y,z)^{(0,1,2)} = \frac{\partial}{\partial y}\frac{\partial^2}{\partial z^2}yz^2 = 2 = 0!\cdot 1!\cdot 2! = (0,1,2)! = \alpha!$$

$$\partial^\alpha(x,y,z)^\beta = \partial^{(0,1,2)}(x,y,z)^{(1,1,0)} = \frac{\partial}{\partial y}\frac{\partial^2}{\partial z^2}xy = 0$$

as promised by Lemma 6.2.16. Fix also $\beta'=(1,3,2)$ so $\beta'\neq\alpha$. Notice that

$$\partial^\alpha(x,y,z)^{\beta'} = \partial^{(0,1,2)}(x,y,z)^{(1,3,2)} = \frac{\partial}{\partial y}\frac{\partial^2}{\partial z^2}(xy^3z^2) = 6xy^2$$

so evaluating at $(x,y,z)=(0,0,0)$ gives 0 as expected.

Informally speaking, Lemma 6.2.16 gives a natural correspondence between the multi-index of a partial derivative and the multi-index defining a monomial. This can be fully exploited to detect coefficients of a polynomial exactly as you would in single variable calculus.

Lemma 6.2.18 Let $P:\mathbb{R}^n\to\mathbb{R}$ be a polynomial in n variables of degree $\leq k$, so

$$P(x) = \sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|\leq k}} C_\alpha x^\alpha$$

for some constants $C_\alpha\in\mathbb{R}$ with $\alpha\in\mathbb{N}^n$ and $|\alpha|\leq k$. Then $C_\alpha = \dfrac{\partial^\alpha P(0)}{\alpha!}$ for every such α.

Proof. Let $\alpha\in\mathbb{N}^n$ with $|\alpha|\leq k$. By linearity of partials,

$$\partial^\alpha P(x) = \sum_{\substack{\beta\in\mathbb{N}^n\\|\beta|\leq k}} C_\beta\,\partial^\alpha(x^\beta)$$

Evaluating this identity at $x=0$ and using Lemma 6.2.16(c), notice that every term with $\beta\neq\alpha$ vanishes. The only remaining term corresponds to $\beta=\alpha$ by Lemma 6.2.16(a) giving $\partial^\alpha P(0) = C_\alpha\cdot\alpha!$ as desired. ■

This lemma generalizes a key property of single variable polynomials. Namely, for a single variable polynomial $P:\mathbb{R}\to\mathbb{R}$ of degree $k\in\mathbb{N}^+$,

$$P(x) = \sum_{n=0}^{k}\frac{P^{(n)}(0)}{n!}x^n.$$

This identity motivates the definition of single variable Taylor polynomials. Lemma 6.2.18 therefore provides the critical insight for defining higher order approximations, which will be the goal of the next section.

Exercises for Section 6.2

Concepts and definitions

6.2.1 Let $f:\mathbb{R}^3\to\mathbb{R}$ and let $\phi:\mathbb{R}^n\to\mathbb{R}$. Which statements are true or false?

(a) If f is C^5 then $\partial^{(2,1,2)}f = \dfrac{\partial^5 f}{\partial x\partial z^2\partial x\partial y}$

(b) If f is C^5 then $\partial^{(2,2,1)}f = \partial^{(1,2,2)}f$.

(c) If f is C^7 then $\partial^{(4,3)}f$ is a partial derivative of order 7.

(d) If ϕ is C^k then $\partial^\alpha\phi$ is defined for any multi-index $\alpha\in\mathbb{N}^n$ up to order k.

(e) If $\phi:\mathbb{R}^n\to\mathbb{R}$ is C^2 and $a\in\mathbb{R}^n$ then the Hessian $H\phi(a)=\left[\partial_i\partial_j\phi(a)\right]_{i,j}$ at a is an $n\times n$ symmetric matrix of real numbers.

(f) All multivariable polynomials are smooth.

6.2.2 Let $f:\mathbb{R}^3\to\mathbb{R}$ be C^3. Write $\dfrac{\partial^3 f}{\partial x_3\partial x_1\partial x_2}$ as a limit in terms of a second order partial derivative.

Computations

6.2.3

(a) Find all partial derivatives of $g(x,y)=x^2-xy+3y^2+1$.

(b) Find all partial derivatives of $f(x,y,z)=ye^{xz}$ up to order 3.

6.2.4 Multi-index notation simplifies the expressions for higher order partial derivatives.

Let $f:\mathbb{R}^n\to\mathbb{R}^m$ be C^k function and $\alpha=(\alpha_1,\ldots,\alpha_n)\in\mathbb{Z}^n$. Recall

$$\partial^\alpha f = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} f$$

and the order of the derivative $\partial^\alpha f$ is $|\alpha|=\alpha_1+\ldots+\alpha_n$

(a) Let $f(x,y)=e^{xy^2}+\sin(x)$. Compute $\partial^{(1,1)}f$. What is the order of this derivative?

(b) Let $f(x,y,z)=xye^{xz^2}$. Compute $\partial^{(2,2,2)}f$. What is the order of this derivative?

6.2.5 Here you will practice quick computations with multi-index notation. Sometimes you will take advantage of a nice property or pattern. Consider what makes them simple.

(a) Compute $\partial^{(1,2,3)}(2x^3+3y+5z^2+7xyz+11x^2y^2z^2+13yz^3)$.

(b) Compute $\partial^{(1,2,3)}(xy^2z^3)$.

(c) Compute $\partial^{(1,2,3)}(x^3y^2z+x^2yz^3+xy^3z^2+xy^2z^3)$.

6.2.6 The chain rule gets messy with higher order derivatives. It is mostly just tedious and will require a little more practice. Let $f:\mathbb{R}^2\to\mathbb{R}^3$ and $g:\mathbb{R}^3\to\mathbb{R}$ be given by

$$f(s,t)=\big(st,2s+3t+7,s^2+(t-1)^2\big) \qquad \text{and} \qquad g(x,y,z)=3x+z.$$

Abusing notation, denote $u=g(x,y,z)=3x+z$ and write

$$x=f_1(s,t)=st, \qquad y=f_2(s,t)=2s+3t+7, \qquad z=f_3(s,t)=s^2+(t-1)^2.$$

Thus, $u = g \circ f(s,t)$. Recall by the chain rule,

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s}.$$

(a) Express $\frac{\partial^2 u}{\partial s^2}$ in terms of other "intermediate" partial derivatives and evaluate it at $(s,t) = (0,0)$.

(b) Express $\frac{\partial^2 u}{\partial t^2}$, $\frac{\partial^2 u}{\partial s \partial t}$ and $\frac{\partial^2 u}{\partial t \partial s}$ in terms of other "intermediate" partials and evaluate at $(0,0)$.

Proofs

6.2.7 The readings sketches the proof of generalized Clairaut's theorem from the readings. The basic idea is to use induction and the standard Clairaut's theorem for C^2 functions, but a fully justified argument is fairly subtle. To get a flavour for the key ideas, you can prove a special case with no induction. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be C^3. Prove that

$$\partial_1\partial_2\partial_3 f = \partial_2\partial_3\partial_1 f$$

using Clairaut's theorem for C^2 functions.

6.2.8 Assume $u = f(x,y)$ for some C^2 function $f : \mathbb{R}^2 \to \mathbb{R}$ and $(x,y) = (r\cos\theta, r\sin\theta)$. Show that

$$\frac{\partial^2 u}{\partial r^2} = \cos^2\theta\frac{\partial^2 u}{\partial^2 x} + 2\cos\theta\sin\theta\frac{\partial^2 u}{\partial x\partial y} + \sin^2\theta\frac{\partial^2 u}{\partial y^2},$$
$$\frac{\partial^2 u}{\partial \theta^2} = r^2\sin^2\theta\frac{\partial^2 u}{\partial^2 x} - 2r^2\cos\theta\sin\theta\frac{\partial^2 u}{\partial x\partial y} + r^2\cos^2\theta\frac{\partial^2 u}{\partial y^2} - r\cos\theta\frac{\partial u}{\partial x} - r\sin\theta\frac{\partial u}{\partial y}.$$

Applications and beyond

6.2.9 Multivariable polynomials are central to approximations, so you can review some facts.

(a) Write out a general degree 2 polynomial in two variables.

(b) How many degree 2 terms does the above polynomial have? List them.

(c) How many degree k terms will a polynomial of degree $N \geq k$ with 2 variables have? List them.

(d) Write out a general degree 2 polynomial in three variables.

(e) How many degree 2 terms does the above polynomial have? List them.

(f) How many degree k terms will a polynomial of degree $N \geq k$ with 3 variables have?

6.2.10 Recall an equivalent definition of 1-variable Taylor polynomials:

> *The Nth Taylor polynomial of a C^N function $f : \mathbb{R} \to \mathbb{R}$ at $a \in \mathbb{R}$ is the polynomial P_N of smallest possible degree such that its zeroth and first N derivatives agree with the zeroth and first N derivatives of f at a. That is,*
>
> $$f(a) = P_N(a),\ f'(a) = P_N'(a),\ \cdots,\ f^{(N)}(a) = P_N^{(N)}(a).$$

Using higher order derivatives, you will generalize this equivalent definition to two variables.

(a) Define the 2nd Taylor polynomial for the function $f : \mathbb{R}^2 \to \mathbb{R}$ at $(a, b) \in \mathbb{R}^2$.
Hint: You will need six equations.

(b) Define the N^{th} Taylor polynomial for $f : \mathbb{R}^2 \to \mathbb{R}$ at $(a, b) \in \mathbb{R}^2$ without multi-index notation.

(c) Define the N^{th} Taylor polynomial for $f : \mathbb{R}^2 \to \mathbb{R}$ at $(a, b) \in \mathbb{R}^2$ with multi-index notation.

(d) Compute the 2nd Taylor polynomial P_2 for $f(x, y) = \cos(x)\sin(y)$ at $(0, 0)$.

6.2.11 Recall another equivalent definition of 1-variable Taylor polynomials:

> *The Nth Taylor polynomial P_N of a C^N function $f : \mathbb{R} \to \mathbb{R}$ at $a \in \mathbb{R}$ is given by an explicit formula involving the first N derivatives of f at a. Namely,*
>
> $$P_N(x) = \sum_{k=0}^{N} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

Unfortunately, the formula for multivariable functions is a bit messy with so many partial derivatives. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be C^N. The N^{th} Taylor polynomial P_N of f at $(a, b) \in \mathbb{R}^2$ is given by the explicit formula:

$$\begin{aligned} P_N(x, y) &= \sum_{k=0}^{N} \sum_{j=0}^{k} \frac{1}{(k-j)!\,j!} \frac{\partial^k f}{\partial x^j \partial y^{k-j}}(a, b)(x-a)^{k-j}(y-b)^j \\ &= \sum_{k=0}^{N} \sum_{i+j=k} \frac{1}{i!\,j!} \frac{\partial^k f}{\partial x^i \partial y^j}(a, b)(x-a)^i(y-b)^j. \end{aligned}$$

You will practice using this formula and rewriting it using different notation.

(a) Use multi-index notation to rewrite the formula for $P_N(x, y)$.

(b) For a multi-index $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$ and $z = (z_1, z_2) \in \mathbb{R}^2$, define the integer $\alpha! = \alpha_1!\alpha_2!$ and the real number $z^\alpha = z_1^{\alpha_1} z_2^{\alpha_2}$. Use these definitions and multi-index notation to rewrite the formula for $P_N(x, y)$.

(c) Let P_2 be the second Taylor polynomial of $f : \mathbb{R}^2 \to \mathbb{R}$ at $(0, 0)$. Use a formula above to explicitly write out $P_2(h_1, h_2)$ for $h = (h_1, h_2) \in \mathbb{R}^2$. Do not use sigma notation.

(d) Let $h = (h_1, h_2) \in \mathbb{R}^2$. You can use powers of the directional derivative operator

$$D_h = h_1 \frac{\partial}{\partial x} + h_2 \frac{\partial}{\partial y}$$

to write an equivalent explicit formula for Taylor polynomials. By direct computation, show that the second Taylor polynomial P_2 for $f : \mathbb{R}^2 \to \mathbb{R}$ at $(0, 0)$ is also given by

$$P_2(h_1, h_2) = \sum_{k=0}^{2} \frac{D_h^k f(0, 0)}{k!}.$$

6.3. Taylor polynomials

Now, with a firm grip on higher order derivatives, you are prepared to define higher order approximations of nonlinear functions, namely *(multivariable) Taylor polynomials*. The core motivation comes from three equivalent definitions of single variable Taylor polynomials; the Nth Taylor polynomial of $f : \mathbb{R} \to \mathbb{R}$ at $a \in \mathbb{R}$ can be described as:

the polynomial given by an explicit formula with the first N derivatives of f at a;

or the polynomial of degree $\leq N$ such that its zeroth and first N derivatives agree with the zeroth and first N derivatives of f at a;

or the polynomial of degree $\leq N$ which is an Nth order approximation of f at a.

In this section, you will use higher order derivatives and multi-index notation to generalize these definitions to multivariable Taylor polynomials. Fix $N \in \mathbb{N}$ and $n \in \mathbb{N}^+$ throughout the section. Assume all functions are real-valued with domains lying in $\mathbb{R}^n$.

6.3.1 Explicit formula

The first definition of Taylor polynomials uses an explicit formula which can be elegantly written using multi-index notation. Let $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ be a multi-index and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Recall that the positive integer $\alpha!$ and real number x^α are given by

$$\alpha! = \alpha_1! \cdots \alpha_n! \qquad \text{and} \qquad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

Note x^α is a monomial in the variables $x_1, \dots, x_n$ of degree $|\alpha| = \alpha_1 + \cdots + \alpha_n$. Moreover, the set $\{x^\alpha : \alpha \in \mathbb{N}^n, |\alpha| = N\}$ is the set of all monomials of degree N.

Definition 6.3.1 Let f be a real-valued function of class C^N on an open ball centred at $a \in \mathbb{R}^n$. The **Nth Taylor polynomial of f at a** is

$$P_N(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq N}} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha.$$

More explicitly,

$$P_N(x) = \sum_{k=0}^{N} \sum_{\substack{\alpha \in \mathbb{N}^n \\ \alpha_1 + \cdots + \alpha_n = k}} \frac{1}{\alpha_1! \cdots \alpha_n!} \frac{\partial^k f(a)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} (x_1 - a_1)^{\alpha_1} \cdots (x_n - a_n)^{\alpha_n}.$$

Remark 6.3.2 Technically, you should write $P_{N,f,a}$ since the polynomial depends on a and f. However, usually the point $a \in \mathbb{R}^n$ and function f are understood, so you can safely suppress them and write P_N.

This first definition probably seems unmotivated and that is a fair criticism. You are seeing the top of the mountain without going on the journey. Rest assured, this formula is natural. You will see that it satisfies all of the desirable equivalent definitions including those about higher order approximations. The real purpose of this definition is to possess a computable and explicit description of Taylor polynomials. You can practice with it in an example. The calculations are very tedious.

Example 6.3.3 Let $f(x,y,z) = ze^{-x-y^2}$ so f is C^∞ on $\mathbb{R}^3$ and hence on an open ball at $a = (0,0,1)$. You can compute the second Taylor polynomial P_2 directly. It is better to organize your computations in a table.

α	$\alpha!$	$(x,y,z-1)^\alpha$	$\partial^\alpha f$	$\partial^\alpha f(a)$
$(0,0,0)$	1	1	ze^{-x-y^2}	1
$(1,0,0)$	1	x	$-ze^{-x-y^2}$	-1
$(0,1,0)$	1	y	$-2yze^{-x-y^2}$	0
$(0,0,1)$	1	$z-1$	e^{-x-y^2}	1
$(2,0,0)$	2	x^2	ze^{-x-y^2}	1
$(1,1,0)$	1	xy	$2yze^{-x-y^2}$	0
$(1,0,1)$	1	$x(z-1)$	$-e^{-x-y^2}$	-1
$(0,2,0)$	2	y^2	$-2ze^{-x-y^2}+4y^2ze^{-x-y^2}$	-2
$(0,1,1)$	1	$y(z-1)$	$-2ye^{-x-y^2}$	0
$(0,0,2)$	2	$(z-1)^2$	0	0

Therefore, by definition of the second Taylor polynomial,

$$\begin{aligned}
P_2(x,y,z) &= \frac{f(a)}{1} + \frac{\partial^{(1,0,0)}f(a)}{1}x + \frac{\partial^{(0,1,0)}f(a)}{1}y + \frac{\partial^{(0,0,1)}f(a)}{1}(z-1) \\
&\quad + \frac{\partial^{(2,0,0)}f(a)}{2}x^2 + \frac{\partial^{(1,1,0)}f(a)}{1}xy + \frac{\partial^{(1,0,1)}f(a)}{1}x(z-1) \\
&\quad + \frac{\partial^{(0,2,0)}f(a)}{2}y^2 + \frac{\partial^{(0,1,1)}f(a)}{1}y(z-1) + \frac{\partial^{(0,0,2)}f(a)}{2}(z-1)^2 \\
&= 1 - x + (z-1) + \frac{x^2}{2} - x(z-1) - y^2.
\end{aligned}$$

This is nothing exciting but it is satisfying to have a straightforward algorithm to compute it.

The first few Taylor polynomials have simple formulas.

Lemma 6.3.4 Let f be of class C^2 on an open ball centred at $a \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$,

$$\begin{aligned}
P_0(x) &= f(a), \\
P_1(x) &= f(a) + \nabla f(a)\cdot(x-a), \\
P_2(x) &= f(a) + \nabla f(a)\cdot(x-a) + \frac{1}{2}(x-a)^T Hf(a)(x-a).
\end{aligned}$$

Proof. This is left as an exercise. It is a direct computation with (6.3.1); the notation is a bit simpler if you write $x = a + h$. ∎

Example 6.3.5 Define the function $h : \mathbb{R}^2 \to \mathbb{R}$ by as $h(x,y) = \sin(x)\cos(y)$. You can use Lemma 6.3.4 to compute the 1st and 2nd Taylor polynomial of h at $a = \left(\frac{\pi}{4}, 0\right) \in \mathbb{R}^2$. Note

$$\nabla h(x,y) = (\cos(x)\cos(y), -\sin(x)\sin(y)) \quad \text{so} \quad \nabla h\left(\frac{\pi}{4}, 0\right) = \left(\frac{1}{\sqrt{2}}, 0\right).$$

Lemma 6.3.4 implies that the 1st Taylor polynomial P_1 of h at $\left(\frac{\pi}{4},0\right)$ is given by

$$\begin{aligned}P_1(x,y) &= h\left(\tfrac{\pi}{4},0\right)+\nabla h\left(\tfrac{\pi}{4},0\right)\cdot\left(x-\tfrac{\pi}{4},y-0\right)\\ &= \tfrac{1}{\sqrt{2}}+\left(\tfrac{1}{\sqrt{2}},0\right)\cdot\left(x-\tfrac{\pi}{4},y\right)\\ &= \tfrac{1}{\sqrt{2}}x-\tfrac{\pi}{4\sqrt{2}}+\tfrac{1}{\sqrt{2}}.\end{aligned}$$

Next, you can compute the Hessian $Hh\left(\frac{\pi}{4},0\right)$. You can verify that

$$Hh(x,y)=\begin{bmatrix}-\sin(x)\cos(y) & -\cos(x)\sin(y)\\ -\cos(x)\sin(y) & -\sin(x)\cos(y)\end{bmatrix} \quad so \quad Hh\left(\frac{\pi}{4},0\right)=\begin{bmatrix}-\frac{1}{\sqrt{2}} & 0\\ 0 & -\frac{1}{\sqrt{2}}\end{bmatrix}.$$

Thus, Lemma 6.3.4 the 2nd Taylor polynomial P_2 of h at $\left(\frac{\pi}{4},0\right)$ is given by

$$\begin{aligned}P_2(x,y) &= P_1(x,y)+\frac{1}{2}\left(x-\tfrac{\pi}{4},y\right)^T Hh\left(\tfrac{\pi}{4},0\right)\left(x-\tfrac{\pi}{4},y\right)\\ &= \tfrac{1}{\sqrt{2}}x-\tfrac{\pi}{4\sqrt{2}}+\tfrac{1}{\sqrt{2}}+\tfrac{1}{2}\begin{bmatrix}x-\frac{\pi}{4} & y\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}} & 0\\ 0 & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x-\frac{\pi}{4}\\ y\end{bmatrix}\\ &= \tfrac{1}{\sqrt{2}}\left(1-\tfrac{\pi}{4}-\tfrac{\pi^2}{32}+x\left(1+\tfrac{\pi}{4}\right)-\tfrac{1}{2}x^2-\tfrac{1}{2}y^2\right).\end{aligned}$$

6.3.2 Matching derivatives

Taylor polynomials are equivalently defined using matching higher order derivatives.

Lemma 6.3.6 Let f be a real-valued function of class C^N on an open ball at $a\in\mathbb{R}^n$. Then a polynomial P is the Nth Taylor polynomial of f at a if and only if P is a polynomial of degree $\leq N$ such that for all multi-indices $\alpha\in\mathbb{N}^n$ with $|\alpha|\leq N$,

$$\partial^\alpha f(a)=\partial^\alpha P(a).$$

Proof. For the "only if" direction, assume $P=P_N$ from Definition 6.3.1. Fix $\alpha\in\mathbb{N}^n$ with $|\alpha|\leq N$. By linearity of partials,

$$\partial^\alpha P_N(x)=\sum_{\substack{\beta\in\mathbb{N}^n\\ |\beta|\leq N}}\frac{\partial^\beta f(a)}{\beta!}\partial^\alpha(x-a)^\beta.$$

Evaluating this identity at $x=a$ and using Lemma 6.2.16(c), notice that every term with $\beta\neq\alpha$ vanishes. The only remaining term corresponds to $\beta=\alpha$ by Lemma 6.2.16 giving $\partial^\alpha P_N(a)=\frac{\partial^\alpha f(a)}{\alpha!}\cdot\alpha!=\partial^\alpha f(a)$ as required.

For the "if" direction, assume P is a polynomial of degree $\leq N$ satisfying $\partial^\alpha f(a)=\partial^\alpha P(a)$ for all $\alpha\in\mathbb{N}^n$ with $|\alpha|\leq N$. Since P is a polynomial of degree $\leq N$, you may write

$$P(x)=\sum_{\substack{\alpha\in\mathbb{N}^n\\ |\alpha|\leq N}}C_\alpha(x-a)^\alpha.$$

for some coefficients $C_\alpha\in\mathbb{R}$ with $\alpha\in\mathbb{N}^n$ and $|\alpha|\leq N$. By Lemma 6.2.18 with $Q(x)=P(x+a)$, it follows that

$$C_\alpha=\frac{\partial^\alpha Q(0)}{\alpha!}=\frac{\partial^\alpha P(a)}{\alpha!}=\frac{\partial^\alpha f(a)}{\alpha!}.$$

Thus, $P=P_N$ by Definition 6.3.1. ∎

This proof shows how Lemma 6.2.18 motivates the explicit formula in Definition 6.3.1. Namely, the coefficients of a polynomial are defined by the value of its derivatives. Lemma 6.3.6 allows you to translate derivative information between Taylor polynomials and f. It is quick and convenient once you have this observation on coefficients of polynomials.

Example 6.3.7 Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is C^∞ and its third Taylor polynomial at $a = (0,2,0)$ is

$$P(x,y,z) = 1-2x+3(y-2)-7z+x^2-x(y-2)-3xz+(y-2)^2-6(y-2)z+2z^2+3x^3+7xz^2.$$

Can you determine the gradient $\nabla f(a)$ and the Hessian $Hf(a)$? This is built right into the polynomial. To see the pattern, you need only compute some derivatives.

α	$\alpha!$	α-coefficient of P	$\partial^\alpha P$	$\partial^\alpha P(a)$
$(0,0,0)$	1	1	$P(x,y,z)$	1
$(1,0,0)$	1	-2	$-2+2x-(y-2)-3z+9x^2+7z^2$	-2
$(0,1,0)$	1	3	$3-x+2(y-2)-6z$	3
$(0,0,1)$	1	-7	$-7-3x-6(y-2)+4z+14xz$	-7
$(2,0,0)$	2	1	$2+18x$	2
$(1,1,0)$	1	-1	-1	-1
$(1,0,1)$	1	-3	$-3+14z$	-3
$(0,2,0)$	2	1	2	2
$(0,1,1)$	1	-6	-6	-6
$(0,0,2)$	2	2	$4+14x$	4

By Lemma 6.3.6, this implies

$$\nabla f(a) = \begin{bmatrix} -2 \\ 3 \\ -7 \end{bmatrix} \quad \text{and} \quad Hf(a) = \begin{bmatrix} 2 & -1 & -3 \\ -1 & 2 & -6 \\ -3 & -6 & 4 \end{bmatrix}.$$

You could have figured this out much more quickly by reading the coefficients of P because:

$\partial^\alpha P(a)$ is the α-coefficient of P multiplied by $\alpha!$.

This relationship is precisely Lemma 6.2.18 which can be seen in the above table. It is not that surprising if you consider how repeated differentiation of polynomials creates factorials.

6.3.3 Higher order approximations

The true purpose of Taylor polynomials is their equivalent definition as higher order approximations. The order of an approximation measures the rate at which the error of an approximation tends to zero by using a "standard family of functions". Powers of the norm

$$1, ||x||, ||x||^2, ||x||^3, \ldots$$

are a natural choice for this standard family. If you want to approximate functions by polynomials, this suggests you need to understand the rate at which n-variable monomials tend to zero compared to powers of the norm.

Lemma 6.3.8 Let $\alpha \in \mathbb{N}^n$. If $|\alpha| \geq N+1$ then $\lim\limits_{x\to 0} \dfrac{x^\alpha}{||x||^N} = 0$.

Proof. This is left as an exercise of limits. Use the formal definition. ■

With this basic understanding, you can generalize higher order approximations.

Definition 6.3.9 Let f and g be real-valued functions defined on an open ball centred at $a \in \mathbb{R}^n$. The function g is an N**th order approximation of** f **at** $a \in \mathbb{R}^n$ if

$$\lim_{x \to a} \frac{f(x) - g(x)}{||x-a||^N} = 0.$$

The crowning achievement for Taylor polynomials is this equivalent definition with higher order approximations. It is famously known as (multivariable) Taylor's theorem[4].

Theorem 6.3.10 (Taylor) Let f be a real-valued function of class C^N on an open ball at $a \in \mathbb{R}^n$. A polynomial P is the Nth Taylor polynomial of f at a if and only if P is the unique degree $\leq N$ polynomial which is an Nth order approximation of f at a.

Proof. The proof is postponed to Section 6.5. It is a big one. ■

Taylor's theorem has a marvelous analytic and geometric interpretations.

Example 6.3.11 Define the C^∞ function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = (y-2) \cdot e^{-x^2}$. By direct calculation, the first few Taylor polynomials of f at $(0,0)$ are

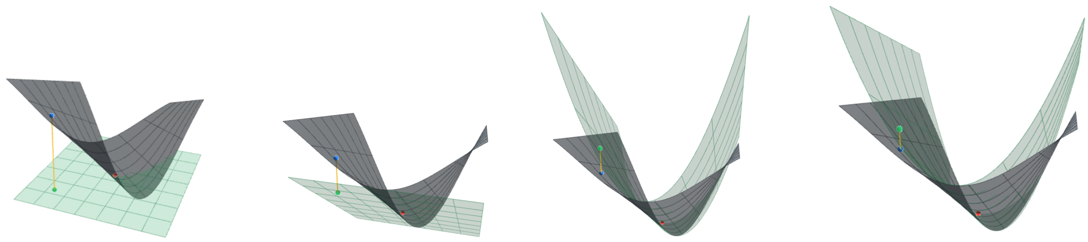
$$P_0(x,y) = -2, \quad P_1(x,y) = -2+y, \quad P_2(x,y) = -2+y+2x^2, \quad P_3(x,y) = -2+y+2x^2-x^2y.$$

By Taylor's theorem, you know for $N \in \mathbb{N}^+$ that

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P_N(x,y)}{||(x,y)||^N} = 0.$$

How can you interpret this analytic statement from a geometric viewpoint? Below are the graphs of $z = f(x,y)$ with each Taylor polynomial $z = P_N(x,y)$ with $N = 1,2,3$.

$z = P_0(x,y)$ $z = P_1(x,y)$ $z = P_2(x,y)$ $z = P_3(x,y)$



Given a point $(x,y) \in \mathbb{R}^2$ near $(0,0)$, each plot illustrates the approximation between $f(x,y)$ and $P_N(x,y)$. Taylor's theorem shows that as $(x,y) \to (0,0)$, the error $f(x,y) - P_N(x,y)$ tends to zero faster than $||(x,y)||^N$ tends to zero. For larger values of N, the error tends to zero faster. View this phenomenon with the Math3D demo by toggling the surfaces.

Taylor polynomials can be computed with both multi- and single-variable Taylor's theorem.

[4] In earlier versions of this text, Taylor's theorem was stated with the stronger assumption that f was C^{N+1}. This extra smoothness was unnecessary and chosen only for the convenience of a more elegant proof. Inspired by a proof shared by Hanfu G. (a former student) on the second derivative test, I decided to upgrade the proof of Taylor's theorem with the weaker assumption of C^N. While the proof is more complicated, the theorem statement brings inner peace.

Example 6.3.12 You can compute the second Taylor polynomial of $f(x,y) = e^{x+y}$ at $a = (0,0)$. Since e^t is analytic everywhere and hence C^3, single variable Taylor's theorem implies that

$$\lim_{t\to 0} \frac{e^t - (1 + t + \frac{t^2}{2} + \frac{t^3}{6})}{t^3} = 0$$

That is, there exists a function R defined on an open interval of $0 \in \mathbb{R}$ such that

$$e^t = 1 + t + \frac{t^2}{2} + \frac{t^3}{6} + R(t), \qquad \lim_{t\to 0} \frac{R(t)}{t^3} = 0. \tag{6.3.1}$$

Then

$$f(x,y) = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2} + \frac{(x+y)^3}{6} + R(x+y).$$

By collecting lower order terms, we claim $P(x,y) = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2}$ is the second Taylor polynomial of f at $(0,0)$. By multivariable Taylor's theorem, it suffices to show that

$$\lim_{(x,y)\to(0,0)} \frac{\frac{(x+y)^3}{6} + R(x+y)}{||(x,y)||^2} = 0.$$

Define $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = R(t)/t^3$ for $t \neq 0$ and $g(0) = 0$ so, by (6.3.2), g is continuous at 0. Moreover, $s(x,y) = x + y$ is continuous at $(0,0)$. Therefore, by Theorem 2.6.21,

$$\lim_{(x,y)\to(0,0)} \frac{R(x+y)}{(x+y)^3} = \lim_{(x,y)\to(0,0)} g(s(x,y)) = g(s(0,0)) = g(0) = 0.$$

Applying Lemma 6.3.8 four times and basic limit laws,

$$\lim_{(x,y)\to(0,0)} \frac{(x+y)^3}{x^2+y^2} = \lim_{(x,y)\to(0,0)} \frac{x^3 + 3x^2y + 3xy^2 + y^3}{||(x,y)||^2} = 0$$

since each monomial x^3, x^2y, xy^2, y^3 is degree 3. By the sum and product limit law,

$$\lim_{(x,y)\to(0,0)} \frac{\frac{(x+y)^3}{6} + R(x+y)}{||(x,y)||^2} = 0 + \lim_{(x,y)\to(0,0)} \frac{R(x+y)}{(x+y)^3} \cdot \frac{(x+y)^3}{x^2+y^2} = 0 + 0 \cdot 0 = 0.$$

Thus, $P(x,y) = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2}$ is the 2nd Taylor polynomial of f at $(0,0)$.

These three definitions of Taylor polynomials are a spectacular achievement for multivariable differential calculus. You can construct better and better approximations to non-linear functions using multivariable polynomials of higher and higher degrees! In other words, smooth non-linear maps can be treated like high degree polynomials without losing too much information. Polynomials are your friends, so this is a really fantastic outcome.

Remark 6.3.13 Other sources may define Taylor polynomials so that P_N is centred at 0 instead of a, in which case $f(x) \approx P_N(x-a)$ for x near a. If so, the shifted polynomial $P_N(x-a)$ is an Nth order approximation of $f(x)$ at a,

$$\forall h \in \mathbb{R}^n,\ P_N(h) = \sum_{\substack{\alpha\in\mathbb{N}^n \\ |\alpha|\le N}} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha, \qquad \text{and} \qquad \forall \alpha \in \mathbb{N}^n \text{ with } |\alpha| \le N,\ \partial^\alpha f(a) = \partial^\alpha P(0).$$

Exercises for Section 6.3

Concepts and definitions

6.3.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^∞ function. Let $a \in \mathbb{R}^n$ and $N \in \mathbb{N}^+$. Which statements are true or false?

(a) The Nth Taylor polynomial of f at a exists.

(b) If P is the Nth Taylor polynomial of f at a then $\displaystyle\lim_{x\to a} \frac{f(x)-P(x)}{||x-a||^N} = 0$.

(c) If a polynomial $P : \mathbb{R}^n \to \mathbb{R}$ satisfies $\displaystyle\lim_{x\to a} \frac{f(x)-P(x)}{||x-a||^N} = 0$ then P is the N^{th} Taylor polynomial of f at a.

(d) If P is the 10th Taylor polynomial of f at a then $\displaystyle\lim_{x\to a} \frac{f(x)-P(x)}{||x-a||^7} = 0$.

(e) If P is the second Taylor polynomial of f at a then $P(a) = f(a)$, $\nabla f(a) = \nabla P(a)$, and $Hf(a) = HP(a)$.

(f) If P_1 is the first Taylor polynomial of f at a then $f(a) + df_a(h) = P_1(a+h)$ for all $h \in \mathbb{R}^n$.

6.3.2 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be C^3 and suppose its 2nd Taylor polynomial at $(0,0,0)$ is given by

$$P_2(x,y,z) = x + 2y + 3z + x^2 - 3xy + 4y^2 - 7yz.$$

Let P_0, P_1 and P_3 be the 0th, 1st, and 3rd Taylor polynomials of f at $(0,0,0)$.

(a) Estimate the value $f(0.1, 0, 0.3)$ with a quadratic approximation of f at $(0,0,0)$.

(b) If possible, determine each of the following quantities:
$f(0,0,0)$ $\quad$ $P_2(0,-1,0)$ $\quad$ $P_1(0,-1,0)$ $\quad$ $f(0,-1,0)$

(c) If possible, determine each of the following quantities:
$\dfrac{\partial f}{\partial x}(0,0,0)$ $\quad$ $\dfrac{\partial^2 f}{\partial y^2}(0,0,0)$ $\quad$ $\dfrac{\partial^3 f}{\partial x \partial y \partial z}(0,0,0)$

(d) If possible, determine each of the following quantities:
$\nabla f(0,0,0)$ $\quad$ $Hf(0,0,0)$

(e) Which of the following could possibly be equal to P_3? Select all possible options.

i) $9xyz - 4x^2y - 8z^3$

ii) $9x^2yz - 4x^2y^2 - 8z^3$

iii) $x + 2y + 3z + x^2 - 3xy + 4y^2 - 7yz + 9xyz - 4x^2y - 8z^3$

iv) $x + 2y + 3z + x^2 - 3xy + 4y^2 - 7yz + 9x^2yz - 4x^2y^2 - 8z^3$

v) $x + 2y + 3z + x^2 - 3xy + 4y^2 - 7yz$

6.3.3 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be C^∞. Let P_1 and P_2 be its 1st and 2nd Taylor polynomials at $(0,0)$. Assume

$$P_2(x,y) = -1 - 3x - 7y + 2x^2 + 3xy + 7y^2$$

For each statement, determine whether it is must be true, whether it must be false, or whether there is not enough information to decide.

(a) $f(0,0) = P_1(0,0)$

(b) $P_1(x,y) = -1 - 3x - 7y$

(c) $\lim_{(x,y)\to(0,0)} \left| f(x,y) - P_1(x,y) \right| = 0$

(d) $\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P_1(x,y)}{||(x,y)||} = 0$

(e) $\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P_2(x,y)}{||(x,y)||} = 0$

(f) $\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P_2(x,y)}{||(x,y)||^2} = 0$

(g) $\lim_{(x,y)\to(0,0)} \frac{f(x,y) - P_2(x,y)}{||(x,y)||^3} = 0$

(h) $\lim_{(x,y)\to(0,0)} \frac{P_1(x,y) - P_2(x,y)}{||(x,y)||^2} = 0$

Computations

6.3.4 Recall the Nth Taylor polynomial of f at $a \in \mathbb{R}^n$ is given by the explicit formula

$$P_N(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \le N}} \frac{\partial^\alpha f(a)}{\alpha!} (x-a)^\alpha.$$

For both parts below, use this explicit formula for your calculations.

(a) Compute the 2nd Taylor polynomial of $f(x,y,z) = x\sin(y+z)$ at $(x,y,z) = (1,0,0)$. Use the table template below to help organize your calculations.

α	$\alpha!$	$(x-1,y,z)^\alpha$	$\partial^\alpha f$	$\partial^\alpha f(a)$

(b) Compute the 3rd Taylor polynomial of $g(x,y) = e^{-xy^2}$ at $(x,y) = (0,0)$.

Proofs

6.3.5 Fix $M, N \in \mathbb{N}^+$. Show that if g is an Mth order approximation of f at $a \in \mathbb{R}^n$ and $M \ge N$, then g is an Nth order approximation of f at a.

6.3.6 Computing Taylor polynomials using the explicit formula is quite tedious. Another option is to use single variable Taylor's theorem. This may seem counterintuitive at first but it can be much more efficient. You only need a bit of care to properly justify your work.

(a) Federico computes the 2nd Taylor polynomial of $f(x,y) = e^{x+y}$ at $(0,0)$.

1. *Since e^t is analytic everywhere and hence C^2 everywhere, single variable Taylor's theorem implies that there exists a real-valued function R defined on an open interval of 0 such that*

$$e^t = 1 + t + \frac{t^2}{2} + R(t), \qquad \lim_{t\to 0} \frac{R(t)}{t^2} = 0. \tag{6.3.2}$$

2. *Then $f(x,y) = P(x,y) + R(x+y)$ where $P(x,y) = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2}$.*
3. *By* (6.3.2),

$$\lim_{(x,y)\to(0,0)} \frac{R(x+y)}{||(x,y)||^2} = \lim_{(x,y)\to(0,0)} \frac{R(x+y)}{(x+y)^2} \cdot \frac{(x+y)^2}{||(x,y)||^2} = 0.$$

4. *Therefore, Taylor's theorem implies that P is the 2nd Taylor polynomial of f at $(0,0)$.*

He has the main ideas but one line is missing some key justifications. Identify this line and what needs to be justified. Then fill in the details to justify this line.

(b) Determine the 2nd Taylor polynomial of $f(x,y,z) = x\sin(y+z)$ at $(x,y,z) = (1,0,0)$ by using single variable Taylor's theorem. Remember you must justify your calculations.

6.3.7 Second order approximations are the most common form of higher order approximations and they have many equivalent formulations. Starting with the explicit formula for Taylor polynomials, you should be able to translate between these formulations. Let f be C^2 on an open ball at $a \in \mathbb{R}^n$.

(a) Show that $\displaystyle\sum_{\substack{\alpha\in\mathbb{N}^n \\ |\alpha|=2}} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i-a_i)(x_j-a_j)$.

(b) Conclude that $\displaystyle P_2(x) = f(a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a)(x_i-a_i) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i-a_i)(x_j-a_j)$.

(c) Show that $\displaystyle\sum_{\substack{\alpha\in\mathbb{N}^n \\ |\alpha|=2}} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha = \frac{1}{2}(x-a)^T Hf(a)(x-a)$. *Hint:* Expand and use 6.3.7(a)

(d) Conclude that $P_2(x) = f(a) + \nabla f(a)\cdot(x-a) + \frac{1}{2}(x-a)^T Hf(a)(x-a)$. *Hint:* Use 6.3.7(b,c).

Applications and beyond

6.3.8 As usual, you cannot really visualize higher order approximations in higher dimensions so you need to rely on your algebraic and analytic understanding of Taylor polynomials.

> **Lemma.** *Let f be C^N on an open ball $B_\delta(a)$ of radius $\delta > 0$ centered at $a \in \mathbb{R}^n$. Let P_N be the Nth Taylor polynomial of f at a. There exists $\varepsilon_{a,N} : B_\delta(0) \to \mathbb{R}$ such that*
>
> $$\forall h \in B_\delta(0), \quad f(a+h) = P_N(a+h) + ||h||^N \varepsilon_{a,N}(h)$$
>
> *and $\varepsilon_{a,N}(h) \to 0$ as $h \to 0$.*

(a) What represents the Nth order approximation of f at a? And the error?

(b) Finish this informal summary of the lemma. The answer is more than 2 words.

> *"If f is C^N at a then the error of its Nth Taylor approximation at a tends ____________."*

(c) Prove the lemma. *Hint:* The choice of $\varepsilon_{a,N}$ will appear circular.

6.3.9 Let $f : \mathbb{R}^n \to R$ be C^N at $a \in \mathbb{R}^n$. Taylor's theorem states that P_N is the unique degree $\leq N$ polynomial which is an Nth order approximation of f at a. You can interpret this statement from the key viewpoints.

(a) The **algebraic** viewpoint is:

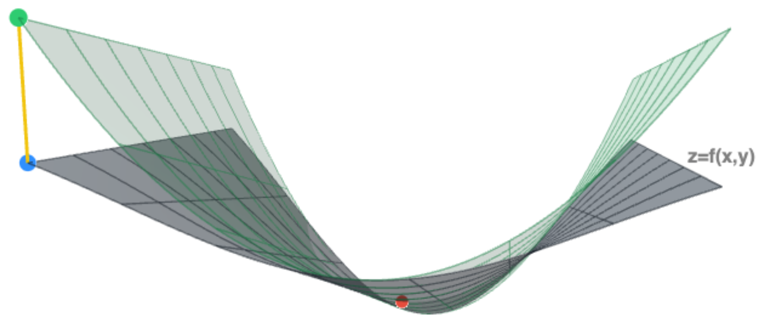
> *The Nth Taylor polynomial P_N of f at a is an Nth order approximation to f at a.*

Write this down precisely using a limit.

(b) The **analytic** viewpoint is:

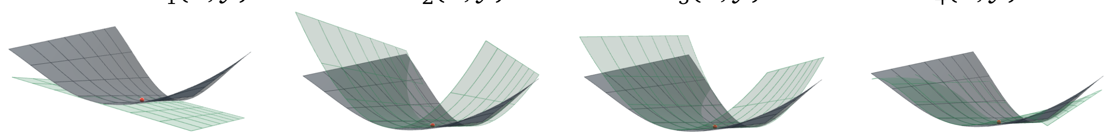
> *For x near a, $f(x) \approx P_N(x)$ with error smaller than $||x-a||^N$.*

How does this picture of an approximation of $f : \mathbb{R}^2 \to \mathbb{R}$ at $a = (a_1, a_2) \in \mathbb{R}^2$ relate to this viewpoint? Label each surface, line, and point using this context.



(c) The **geometric** viewpoint of Taylor's theorem is illustrated with these pictures.

$z = P_1(x, y)$ $z = P_2(x, y)$ $z = P_3(x, y)$ $z = P_4(x, y)$



Select the sentence below which best describes Taylor's theorem from a geometric viewpoint.

i) *P_N defines a n-dimensional polynomial surface which is close to the graph of f near a.*

ii) *P_N defines a n-dimensional polynomial surface which approaches the graph of f near a as N grows.*

iii) *P_N defines a n-dimensional polynomial surface which fits the graph of f near a up with order N error.*

iv) *P_N defines the n-dimensional polynomial surface of degree $\leq N$ which best fits the graph of f near a.*

(d) View this Math3D demo by toggling surfaces. Describe how it illustrates Taylor's theorem using one or two sentences. Use precise language from one or more of the above viewpoints.

6.4. Classification of critical points

You are now prepared to address this chapter's motivating question.

Given a critical point, can you classify if it is a local extrema? If so, of what type?

Let f be a real-valued function that is differentiable at $a \in \mathbb{R}^n$. If a is a critical point of f, then $\nabla f(a) = 0$ so the first Taylor polynomial of f at a is given by

$$P_1(x) = f(a)$$

from Lemma 6.3.4. Since P_1 is constant, you cannot tell whether f has a local extrema at a.

To obtain deeper information, you can now use a higher order approximation, namely a *quadratic* approximation. Assuming f is C^2 on an open ball containing $a \in \mathbb{R}^n$, you can apply Taylor's theorem to approximate f up to the *second* order. Thus, if a is a critical point of f then $\nabla f(a) = 0$ so

$$P_2(x) = f(a) + \tfrac{1}{2}(x-a)^T Hf(a)(x-a)$$

by Lemma 6.3.4. The behaviour of f near a is therefore dictated by the second order term $\frac{1}{2}(x-a)^T Hf(a)(x-a)$ with the $n \times n$ Hessian matrix $Hf(a)$. This object will be the focus of this section, so it warrants its own definition.

Definition 6.4.1 Let f be a real-valued function that is C^2 on an open ball containing $a \in \mathbb{R}^n$. The **quadratic form of f at a** is the function $q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\forall v \in \mathbb{R}^n, \qquad q(v) = v^T Hf(a)v.$$

If a is a critical point of f and q is the quadratic form of f at a, then

$$P_2(x) = f(a) + \tfrac{1}{2}q(x-a).$$

Properties of $q : \mathbb{R}^n \to \mathbb{R}$ will translate to properties of the polynomial P_2, such as:

- If q is always positive on $\mathbb{R}^n \setminus \{0\}$, then P_2 will have a global minimum at a.
- If q is always negative on $\mathbb{R}^n \setminus \{0\}$, then P_2 will have a global maximum at a.
- If q is negative and positive on $\mathbb{R}^n \setminus \{0\}$, then P_2 will not have a global extremum at a.

Since the quadratic polynomial P_2 *locally* approximates f, can you translate these *global* conclusions about P_2 at a into *local* conclusions about f at a itself? Classifying the critical point will hinge upon a careful analysis of the quadratic form q.

6.4.1 Quadratic forms

A Hessian matrix $Hf(a)$ is always symmetric for a C^2 function f. It will be helpful to temporarily ignore the calculus and study quadratic forms in a pure linear algebra context. This means you will replace $Hf(a)$ with any $n \times n$ symmetric matrix A and study its quadratic form.

Definition 6.4.2 Let A be a real $n \times n$ symmetric matrix. The **quadratic form associated to A** is the function $q : \mathbb{R}^n \to \mathbb{R}$ given by $q(v) = v^T Av$ for $v \in \mathbb{R}^n$.

Notice $q(0) = 0$ for any quadratic form. These forms can be explicitly computed.

Example 6.4.3 The symmetric matrix

$$A = \begin{bmatrix} 3 & 5 & 1 \\ 5 & 2 & 8 \\ 1 & 8 & 4 \end{bmatrix}$$

defines a quadratic form $q : \mathbb{R}^3 \to \mathbb{R}$ by $q(v) = v^T Av$ for all $v \in \mathbb{R}^3$. Explicitly,

$$\begin{aligned} q(x, y, z) &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 3 & 5 & 1 \\ 5 & 2 & 8 \\ 1 & 8 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= 3x^2 + 10xy + 2xz + 2y^2 + 4z^2 + 16yz \end{aligned}$$

is the quadratic form associated to A.

How does the symmetric matrix A relate to the positivity (or negativity) of its associated quadratic form $q(v) = v^T Av$? This question is at the heart of classifying critical points. The key observation is elementary but insightful.

Lemma 6.4.4 Let A be a $n \times n$ real symmetric matrix. Let $q : \mathbb{R}^n \to \mathbb{R}$ be its associated quadratic form. If $v \in \mathbb{R}^n$ is an eigenvector of A with eigenvalue $\lambda \in \mathbb{R}$, then

$$q(v) = \lambda ||v||^2.$$

Proof. By definition, $q(v) = v^T Av = v^T (\lambda v) = \lambda v^T v = \lambda ||v||^2$. ■

This lemma illustrates that at an *eigenvector*, the quadratic form will be positive if the eigenvalue is positive, negative if the eigenvalue is negative, and zero if the eigenvalue is zero. This is fantastic progress already. Wishful thinking leads to a reasonable question.

Can the eigenvectors of A tell you everything about the quadratic form q?

Miraculously, the answer is yes! Real symmetric matrices possess a special structure.

Theorem 6.4.5 (Spectral theorem) Every $n \times n$ real symmetric matrix has an orthonormal[5] basis of eigenvectors with real eigenvalues.

Proof. Omitted. This proof is usually reserved for a course in abstract linear algebra. ■

The spectral theorem and Lemma 6.4.4 give you the perfect opportunity for analyzing quadratic forms using their eigenvectors and eigenvalues.

Example 6.4.6 As claimed by the spectral theorem, the 2×2 symmetric matrix

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

has an orthonormal basis of eigenvectors given by

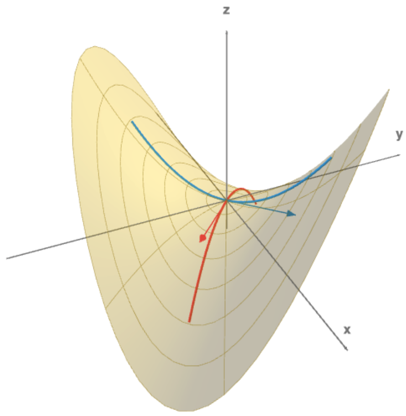
$$\left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}.$$

The eigenvalues are respectively 1 and -1. You can compute this directly by routine linear algebra. The quadratic form $q : \mathbb{R}^2 \to \mathbb{R}$ associated to A is given by

$$q(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2xy.$$

[5] An *orthonormal basis* is a basis $\{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ such that $||v_j||^2 = 1$ and $v_j \cdot v_k = 0$ for $1 \le j, k \le n$ with $j \ne k$.

The graph of $z = q(x,y)$ and slices in the eigenvector directions can be viewed with Math3D.



Notice the eigenvector $v_1 = \frac{1}{\sqrt{2}}(1,1)$ with positive eigenvalue $+1$ corresponds to the upward parabola slice, and the eigenvector $v_2 = \frac{1}{\sqrt{2}}(1,-1)$ with negative eigenvalue -1 corresponds to the downward parabola slice. This matches the observation in Lemma 6.4.4.

Example 6.4.7 There are four classic examples of quadratic forms in $\mathbb{R}^2$ and these illustrate all the necessary phenomena. Define the matrices

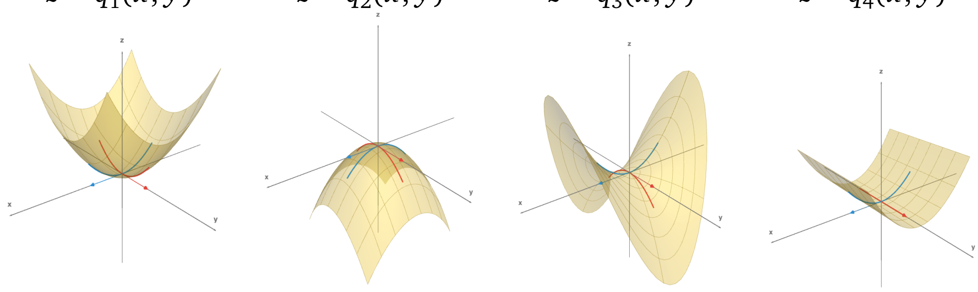
$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The associated quadratic forms are therefore given by

$$q_1(x,y) = x^2 + y^2, \qquad q_2(x,y) = -x^2 - y^2, \qquad q_3(x,y) = x^2 - y^2, \qquad q_4(x,y) = x^2$$

The plots of these quadratic forms illustrate their behaviour; see this Math3D demo.

$z = q_1(x,y)$ $\quad z = q_2(x,y)$ $\quad z = q_3(x,y)$ $\quad z = q_4(x,y)$



You can verify that the set $\{(1,0),(0,1)\}$ in $\mathbb{R}^2$ is an orthonormal basis for all four matrices. Moreover, you can check that A_1 has both positive eigenvalues $\{1,1\}$ and q_1 is always positive away from the origin. Also, A_2 has both negative eigenvalues $\{-1,-1\}$ and q_2 is always negative away from the origin. Also, A_3 has a positive and a negative eigenvalue $\{1,-1\}$, and q_3 is both positive and negative. Finally, A_4 has a positive and a zero eigenvalue $\{1,0\}$, and q_4 is non-negative everywhere but equal to zero along a line.

This last example suggests that the maximum and minimum eigenvalues of your symmetric matrix A dictate the behaviour of the associated quadratic form $q(v) = v^T A v$. Since $q(tv) = t^2 q(v)$ for $t \in \mathbb{R}$ and $v \in \mathbb{R}^n$, it suffices to analyze the extrema of q on the unit sphere S^{n-1}.

Theorem 6.4.8 Let A be a $n \times n$ real symmetric matrix. Let $q : \mathbb{R}^n \to \mathbb{R}$ be its quadratic form defined by $q(v) = v^T A v$ for $v \in \mathbb{R}^n$. The maximum and minimum values of q on the unit sphere S^{n-1} are respectively equal to the maximum and minimum eigenvalue of A.

The proof is a beautiful display of the method of Lagrange multipliers.

Proof. Since q is continuous and S^{n-1} is compact, the extreme value theorem implies that q has a maximum and a minimum on S^{n-1}. Define $g : \mathbb{R}^n \to \mathbb{R}$ by $g(x) = ||x||^2$. By the method of Lagrange multipliers, the extrema of q on S^{n-1} must satisfy the Lagrange system

$$\nabla q(x) = \lambda \nabla g(x), \qquad g(x) = 1,$$

where $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. You can verify by direct calculation that $\nabla q(x) = 2Ax$ and $\nabla g(x) = 2x$, so the Lagrange system is equivalent to

$$Ax = \lambda x, \qquad ||x||^2 = 1.$$

In other words, x must be a normalized eigenvector of A with eigenvalue λ and moreover, $q(x) = \lambda$ by Lemma 6.4.4. By the spectral theorem, A has an orthonormal basis of eigenvectors with real eigenvalues, so there exists solutions to the Lagrange system corresponding to the maximum and minimum eigenvalue of A. ■

Theorem 6.4.8 is the last vital ingredient for classifying critical points.

6.4.2 Second derivative test

You are finally ready to classify critical points using an imperfect but elegant criterion.

Theorem 6.4.9 (Second derivative test) Let f be a real-valued function of class C^2 on an open ball at $a \in \mathbb{R}^n$. If a is a critical point of f, then f has
(a) a local minimum if all the eigenvalues of the Hessian $Hf(a)$ are positive.
(b) a local maximum if all the eigenvalues of the Hessian $Hf(a)$ are negative.
(c) a saddle point if the Hessian $Hf(a)$ has a positive and a negative eigenvalue.

Remark 6.4.10 Notice the second derivative test is *inconclusive* in exactly two situations:
- The Hessian has some zero eigenvalues and the other eigenvalues are all positive.
- The Hessian has some zero eigenvalues and the other eigenvalues are all negative.

Taylor's theorem allows you to approximate f by the second Taylor polynomial, and Theorem 6.4.8 allows you to determine the behaviour of the quadratic form of f at a. The proof of the second derivative test is a delicate combination of these tools.

Proof. Since a is a critical point of f, the second Taylor polynomial of f at a is given by

$$P_2(x) = f(a) + \tfrac{1}{2}(x-a)^T Hf(a)(x-a) = f(a) + \tfrac{1}{2}q(x-a),$$

where $q(v) = v^T Hf(a)v$ is the quadratic form of f at a. By Taylor's theorem,

$$\lim_{x \to a} \frac{f(x) - P_2(x)}{||x-a||^2} = 0. \qquad (6.4.1)$$

First we prove (a). Assume all the eigenvalues of the Hessian $Hf(a)$ are positive. Let λ be the minimum eigenvalue of $Hf(a)$ so $\lambda > 0$. By (6.4.3), there exists $\delta > 0$ such that

$$0 < ||x-a|| < \delta \implies |f(x) - f(a) - \tfrac{1}{2}q(x-a)| < \frac{\lambda}{4}||x-a||^2.$$

Fix $x \in \mathbb{R}^n$ with $0 < ||x-a|| < \delta$. It follows that

$$f(x) > f(a) + \frac{1}{2}q(x-a) - \frac{\lambda}{4}||x-a||^2.$$

From Theorem 6.4.8, it follows that $q(x-a) = ||x-a||^2 q(\frac{x-a}{||x-a||}) \geq \lambda ||x-a||^2$, so

$$f(x) > f(a) + \frac{\lambda}{2}||x-a||^2 - \frac{\lambda}{4}||x-a||^2 > f(a).$$

This proves f has a local minimum at a, as required. The proof of (b) is similar.

It remains to prove (c). In this case, the Hessian $Hf(a)$ has a minimum eigenvalue $\lambda_1 < 0$ and a maximum eigenvalue $\lambda_2 > 0$. Let v_1 and v_2 be respective normalized eigenvectors so $||v_1|| = ||v_2|| = 1$. By (6.4.3), there exists $\delta > 0$ such that

$$0 < ||x-a|| < \delta \implies |f(x) - f(a) - \tfrac{1}{2}q(x-a)| < \frac{\min\{|\lambda_1|, |\lambda_2|\}}{4}||x-a||^2. \tag{6.4.2}$$

Fix $0 < \varepsilon < \delta/2$. It suffices to show there exists $x_1, x_2 \in B_{2\varepsilon}(a)$ such that

$$f(x_1) < f(a) < f(x_2).$$

Take $x_1 = a + \varepsilon v_1$ and $x_2 = a + \varepsilon v_2$ so $||x_1 - a|| = ||x_2 - a|| = \varepsilon$ implying $x_1, x_2 \in B_{2\varepsilon}(a)$. From Lemma 6.4.4, it follows that

$$q(x_1 - a) = q(\varepsilon v_1) = \varepsilon^2 \lambda_1 \quad \text{and} \quad q(x_2 - a) = q(\varepsilon v_2) = \varepsilon^2 \lambda_2.$$

By (6.4.2), it follows that

$$\begin{aligned} f(x_1) &\leq f(a) + \tfrac{1}{2}q(x_1 - a) + \frac{\min\{|\lambda_1|, |\lambda_2|\}}{4}||x_1 - a||^2 \\ &\leq f(a) + \frac{\varepsilon^2 \lambda_1}{2} + \frac{\varepsilon^2 |\lambda_1|}{4} \\ &\leq f(a) \end{aligned}$$

as $\lambda_1 < 0$. You can also verify that $f(x_2) > f(a)$. This proves (c). ∎

From a computational perspective, the second derivative test can often (but not always) identify local extrema and saddle points.

Example 6.4.11 You can try to classify the critical points of $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = x^2 y + x^2 y^2 + xy^2 + xy^3$$

using the second derivative test. Note f is C^∞ and

$$\nabla f(x,y) = \big(y(2x+y)(y+1), x(3y^2 + 2y(x+1) + x)\big).$$

After some tedious but straightforward calculations, you can verify that $\nabla f(x,y) = (0,0)$ if and only if $(x,y) = (0,0), (\frac{3}{8}, \frac{-3}{4}), (0,-1)$, or $(1,-1)$. To classify these critical points, you can compute the Hessian matrix

$$Hf(x,y) = \begin{bmatrix} 2y^2 + 2y & 4xy + 2x + 3y^2 + 2y \\ 4xy + 2x + 3y^2 + 2y & 2x^2 + 6xy + 2x \end{bmatrix}.$$

Plugging in each critical point, you can check that

- $Hf(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ has eigenvalues $\{0,0\}$.

- $Hf\left(\frac{3}{8},-\frac{3}{4}\right)=\begin{bmatrix}-\frac{3}{8} & -\frac{3}{16}\\ -\frac{3}{16} & -\frac{21}{32}\end{bmatrix}$ has eigenvalues $-\frac{9}{32}$ and $-\frac{3}{4}$.
- $Hf(0,-1)=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ has eigenvalues $\{1,-1\}$.
- $Hf(1,-1)=\begin{bmatrix}0 & -1\\ -1 & -2\end{bmatrix}$ has eigenvalues $\{-1-\sqrt{2},\sqrt{2}-1\}$.

By the second derivative test, f has a local maximum at $(\frac{3}{8},-\frac{3}{4})$ and saddle points at $(0,-1)$ and $(1,-1)$. Notice, however, that the second derivative test is inconclusive at $(0,0)$; this could be a saddle point, a local maximum, or a local minimum. View this Math3D demo to compare the surface $z=f(x,y)$ and the (shifted) quadratic form at each critical point.

The second derivative test has many equivalent formulations other than Theorem 6.4.9. The translation between these formulations is merely some linear algebra about real symmetric matrices. For instance, here is a common rephrasing for two variable functions.

Corollary 6.4.12 Let f be a two-variable real-valued C^2 function on an open ball at $p\in\mathbb{R}^2$. If p is a critical point of f, then f has
(a) a local minimum if $f_{xx}(p)f_{yy}(p)-(f_{xy}(p))^2>0$ and $f_{xx}(p)>0$.
(b) a local maximum if $f_{xx}(p)f_{yy}(p)-(f_{xy}(p))^2>0$ and $f_{xx}(p)<0$.
(c) a saddle point if $f_{xx}(p)f_{yy}(p)-(f_{xy}(p))^2<0$.

Proof. The 2×2 symmetric matrix $Hf=\begin{bmatrix}f_{xx} & f_{xy}\\ f_{xy} & f_{yy}\end{bmatrix}$ satisfies

$$f_{xx}(p)f_{yy}(p)-(f_{xy}(p))^2=\det Hf(p)=\lambda_1\lambda_2,$$

where λ_1 and λ_2 are its two real eigenvalues. By performing some casework, you can verify that the corollary follows from the above identity and Theorem 6.4.9. ∎

You can apply this formulation just as easily as Theorem 6.4.9. However, you must be careful to respect that there are strict limitations on the second derivative test and what you can conclude. It is only designed to classify *local* extrema.

Example 6.4.13 Consider the function $f(x,y)=e^{-(x^2+y^2+y^3)}$, so

$$\nabla f(x,y)=\left(-2xe^{-(x^2+y^2+y^3)},-(2y+3y^2)e^{-(x^2+y^2+y^3)}\right).$$

You can verify that $\nabla f(x,y)=(0,0)$ if and only if $(x,y)=(0,0)$ or $\left(0,-\frac{2}{3}\right)$. To classify these critical points, you can compute the Hessian matrix

$$Hf(x,y)=e^{-(x^2+y^2+y^3)}\begin{bmatrix}4x^2-2 & -2x(-3y^2-2y)\\ -2x(3y^2-2y) & (-3y^2-2y)^2+(-6y-2)\end{bmatrix}.$$

Plugging in these critical points, you obtain that

$$Hf\left(0,-\frac{2}{3}\right)=\begin{bmatrix}-2e^{\frac{4}{27}} & 0\\ 0 & 2e^{\frac{4}{27}}\end{bmatrix}\quad\text{and}\quad Hf(0,0)=\begin{bmatrix}-2 & 0\\ 0 & -2\end{bmatrix}$$

Note $\det Hf\left(0,-\frac{2}{3}\right)=-4e^{8/27}<0$, so the point $(0,\frac{-2}{3})$ is a saddle point of f by the second derivative test, namely Corollary 6.4.12. Similarly, notice that $\det Hf(0,0)=4>0$ and $f_{xx}(0,0)=-2<0$, so the point $(0,0)$ is a local maximum of f by the second derivative test.

You may be tempted to conclude that:

As f has no more critical points, the point $(0,0)$ must be the global maximum of f.

But this is actually false! The second derivative test cannot make any conclusions about global extrema. It has no effective capacity in global optimization problems. You must use other tools (e.g. global extreme value theorem, local extreme value theorem, Lagrange multipliers) if you wish to solve global optimization problems. Here you have not even confirmed whether a global maximum of f exists. In fact, notice that

$$\lim_{y\to-\infty} f(0,y) = \lim_{y\to-\infty} e^{-y^2-y^3} = \lim_{y\to-\infty} e^{-y^3(\frac{1}{y}+1)} = \infty,$$

which you can visually confirm in this Math3D demo. Thus, f does not have a global maximum. This reminder demonstrates why the second derivative test is incompatible with global optimization problems.

This concludes your long and storied journey in optimization. You have constructed powerful tools to globally optimize and to locally optimize. You can show an optimal solution exists (global extreme value theorem). You can locate possible extrema (local extreme value theorem, Lagrange multipliers). With the help of higher order approximations and Taylor's theorem, you can now classify local extrema (second derivative test). None of these techniques are perfect but, with a bit of creativity, you can apply them in potent combination to investigate many optimization problems. Now, go forth and conquer!

Exercises for Section 6.4

Concepts and definitions

6.4.1 Let f be a real-valued function. Assume f is C^2 on an open ball of the critical point $a \in \mathbb{R}^n$ so $\nabla f(a) = 0$. Which statements are true or false?

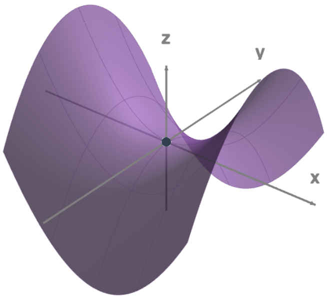
(a) If $Hf(a)$ has a zero eigenvalue, then you cannot classify a by the second derivative test.

(b) If $Hf(a)$ has a zero eigenvalue and the rest of its eigenvalues are non-negative, then you cannot classify the critical point a by the second derivative test.

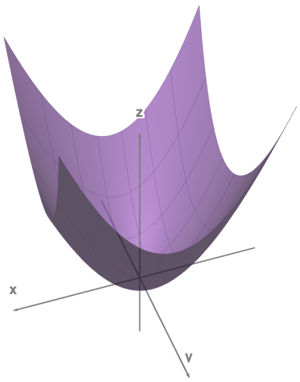
(c) If $Hf(a)$ has a zero eigenvalue and the rest of its eigenvalues are non-negative, then the critical point a may be something other than a local maximum, a local minimum, or a saddle point.

6.4.2 Each quadratic form $q : \mathbb{R}^2 \to \mathbb{R}$ has an associated 2×2 symmetric matrix A. Use the graph of q to determine how many eigenvalues of A are positive, negative, or zero.

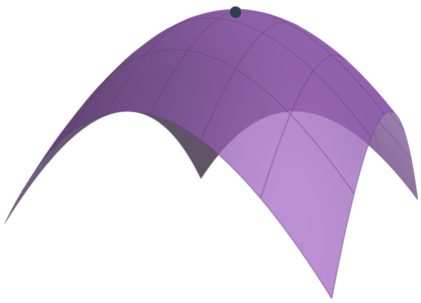
(a)



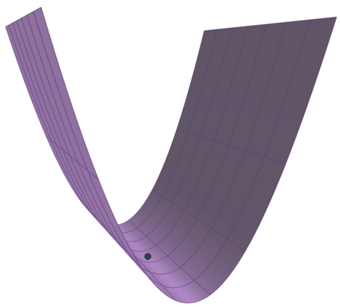
(b)



(c)



(d)



6.4.3 The illustrations above suggest a formal connection between extrema of quadratic forms and the eigenvalues of their matrices. Conjecture this formal connection.

Theorem B. *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form with associated symmetric matrix A. Then*

i) q has a unique global minimum at the origin if ____________________.

ii) q has a global minimum at the origin if ____________________.

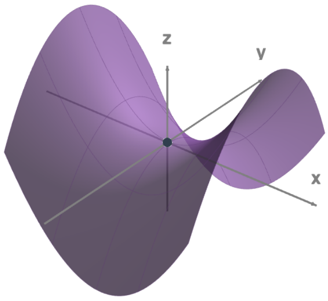
iii) q has a unique global maximum at the origin if ____________________.

iv) q has a global maximum at the origin if ____________________.

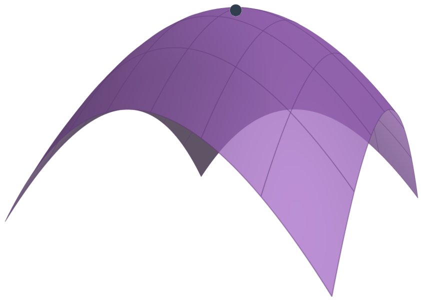
v) q does not have a global extremum at the origin if ____________________.

6.4.4 Each graph below is the graph of the quadratic form for a C^2 real-valued function $\mathbb{R}^2 \to \mathbb{R}$ at one of its critical points $a, b, c, d \in \mathbb{R}^2$. If possible, apply the second derivative test to classify the critical point.

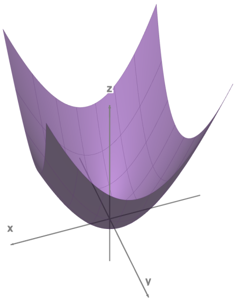
(a) quadratic form of f at a.



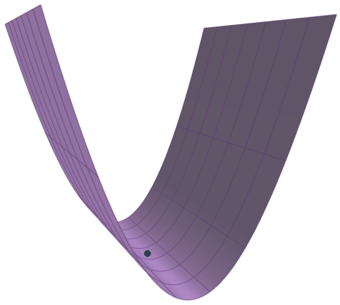
(c) quadratic form of f at c.



(b) quadratic form of f at b.



(d) quadratic form of f at d.



6.4.5 The second derivative test also applies to single-variable functions but, since the dimension is small, something different happens.

(a) List all possible quadratic forms $q : \mathbb{R} \to \mathbb{R}$.

(b) Exhibit a C^2 function $f : \mathbb{R} \to \mathbb{R}$ with a saddle point at 0. Write its quadratic form at 0.

(c) True or false?

The second derivative test cannot be used to classify saddle points for single-variable functions.

Computations

6.4.6 Suppose $g : \mathbb{R}^3 \to \mathbb{R}$ is C^2 and has critical point $(2, 3, 7)$. If possible, classify this critical point given that

$$Hg(2,3,7) = \begin{bmatrix} -524 & -87 & 263 \\ -87 & -20 & 27 \\ 263 & 27 & -182 \end{bmatrix}.$$

Use WolframAlpha for any linear algebra calculations.

6.4.7 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be C^2. You find it has exactly three critical points $a, b, c \in \mathbb{R}^3$ and compute the Hessian matrices at each point to be

$$Hf(a) = \begin{bmatrix} 378 & 374 & 173 \\ 374 & 372 & 169 \\ 173 & 169 & 83 \end{bmatrix}, \quad Hf(b) = \begin{bmatrix} -77 & -107 & -109 \\ -107 & -149 & -151 \\ -109 & -151 & -155 \end{bmatrix}, \quad Hf(c) = \begin{bmatrix} 335 & 339 & 139 \\ 339 & 341 & 143 \\ 139 & 143 & 55 \end{bmatrix}.$$

Use WolframAlpha to help you classify the critical points a, b, c, if possible.

6.4.8 The case of two variables is special. Recall if A is a 2 x 2 matrix with eigenvalues λ_1, λ_2 then

$$\det A = \lambda_1 \lambda_2.$$

For two variable functions, the determinant of the Hessian at a critical point tells you a lot.

(a) Use the second derivative test to write the corresponding conclusion.

Corollary A. *Let f be real-valued and C^2 on an open ball of the critical point $a \in \mathbb{R}^2$ so $\nabla f(a) = 0$.*

i) If $\det Hf(a) > 0$, then ______________________________ .

ii) If $\det Hf(a) < 0$, then ______________________________ .

iii) If $\det Hf(a) = 0$, then ______________________________ .

(b) The point $(0,0)$ is a critical point of $h(x,y) = e^{xy}$. Compute its quadratic form and, if possible, classify the critical point.

6.4.9 Here is another version of the second derivative test with determinants and submatrices.

> **Corollary C.** *Let f be real-valued and C^2 on an open ball of the critical point $a \in \mathbb{R}^n$ so $\nabla f(a) = 0$. For $k \in \{1, \ldots, n\}$, let Δ_k be the determinant of the upper lefthand $k \times k$ submatrix of the Hessian $Hf(a)$.*
>
> *i) If $\Delta_k > 0$ for all $k = 1, \ldots, n$ then a is a local minimum of f.*
> *ii) If $(-1)^k \Delta_k > 0$ for all $k = 1, \ldots, n$ then a is a local maximum of f.*
> *iii) If $\det Hf(a) \neq 0$ and neither i) nor ii) is satisfied then a is a saddle point of f.*

This avoids an eigenvalue computation which can be convenient in some contexts.

(a) Conjecture the linear algebra theorem combining with Theorem A to prove Corollary C.

Theorem C. *Let A be* ______________________________.
For $k \in \{1, \ldots, n\}$, let Δ_k be the determinant of the upper lefthand $k \times k$ submatrix of A.

i) If $\Delta_k > 0$ for all $k = 1, \ldots, n$ then ______________________________.

ii) If $(-1)^k \Delta_k > 0$ for all $k = 1, \ldots, n$ then ______________________________.

iii) If $\det(A) \neq 0$ and neither i) nor ii) holds then ______________________________.

iv) If $\det(A) = 0$ then ______________________________.

(b) Specialize Corollary C to functions of two variables. This common version has easy calculations. Write expressions in $f_{xx}(a), f_{yy}(a)$, and $f_{xy}(a)$ for all of your answers.

Corollary D. *Let f be real-valued and C^2 on an open ball of the critical point $a \in \mathbb{R}^2$ so $\nabla f(a) = 0$.*

i) If __________ > 0 *and* ______________ > 0 *then a is a local minimum of f.*

ii) If __________ < 0 *and* ______________ > 0 *then a is a local maximum of f.*

iii) If ______________ < 0 *then a is a saddle point of f.*

iv) If ______________ $= 0$ *then the test is inconclusive.*

(c) Classify the critical points of $g(x,y) = \frac{1}{3}x^3 + y^2 + 2xy - 6x - 3y + 4$ using Corollary D.

6.4.10 Classify the critical points of $f(x,y) = x^4 + y^4 - 8x^2 + 4y$, if possible.

6.4.11 Classify the critical points of $f(x,y,z) = x^2 + xz + y^4 + z^2$, if possible.

6.4.12 Let $a \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = e^{-\|x-a\|^2}$. Classify the critical points of f.

Proofs

6.4.13 The second derivative test has some very common pitfalls and incorrect applications.

(a) Spongebob tries to find the global extrema of $f(x,y) = e^{-x^2-y^2}$ on $\mathbb{R}^2$.

1. $\nabla f(x,y) = (-2xe^{-x^2-y^2}, -2ye^{-x^2-y^2})$
2. $\nabla f(x,y) = (0,0) \iff (-2xe^{-x^2-y^2}, -2ye^{-x^2-y^2}) = (0,0) \iff (x,y) = (0,0)$ *critical point*
3. $Hf(x,y) = \begin{bmatrix} -2+4x^2 & 4xy \\ 4xy & -2+4y^2 \end{bmatrix} e^{-x^2-y^2} \implies Hf(0,0) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$.
4. $\therefore \det Hf(0,0) = 4 > 0$ *and* $f_{xx}(0,0) = -2 < 0$.
5. *By second derivative test,* $(0,0)$ *is the global maximum.*

In addition to Spongebob's atrocious writing, what is wrong with his reasoning?

(b) Mr. Crab tries to find classify the critical points of $g(x,y) = x^4 + y^4$ on $\mathbb{R}^2$.

1. *By direct computation,* $\nabla g(x,y) = (4x^3, 4y^3)$.
2. *Then* $\nabla g(x,y) = (0,0)$ *if and only if* $4x^3 = 0$ *and* $4y^3 = 0$ *which occurs if and only if* $(x,y) = (0,0)$.
3. *After some calculations,* $Hg(0,0)$ *is the zero matrix, so* $(0,0)$ *is not a local extrema.*

In addition to Mr. Crab skipping many details, what is wrong with his argument?

6.4.14 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be C^2. Assume

$$P(x,y) = 4 - 3x^2 + xy - y^2 + 2x^3 + 3xy^2 + 7y^2x + y^3$$

is the 3rd Taylor polynomial of f at $(0,0)$. Prove that f has a local maximum at $(0,0)$.

Applications and beyond

6.4.15 You will summarize the big ideas for the proof of the second derivative test (Theorem 6.4.9) when all the eigenvalues of the Hessian are positive, i.e. the critical point is a local minimum.

1. *Assume all the eigenvalues of the Hessian $Hf(a)$ are positive.*
2. *Since a is a critical point of f, the second Taylor polynomial of f at a is given by*

$$P_2(x) = f(a) + \tfrac{1}{2}(x-a)^T Hf(a)(x-a) = f(a) + \tfrac{1}{2}q(x-a),$$

   where $q(v) = v^T Hf(a)v$ is the quadratic form of f at a.
3. *By Taylor's theorem,*

$$\lim_{x\to a} \frac{f(x) - P_2(x)}{||x-a||^2} = 0. \tag{6.4.3}$$

4. *Let λ be the minimum eigenvalue of $Hf(a)$ so $\lambda > 0$.*
5. *By* (6.4.3)*, there exists $\delta > 0$ such that*

$$0 < ||x-a|| < \delta \implies |f(x) - f(a) - \tfrac{1}{2}q(x-a)| < \frac{\lambda}{4}||x-a||^2.$$

6. *Fix $x \in \mathbb{R}^n$ with $0 < ||x-a|| < \delta$. It follows that*

$$f(x) > f(a) + \frac{1}{2}q(x-a) - \frac{\lambda}{4}||x-a||^2.$$

7. *It follows that $q(x-a) = ||x-a||^2 q(\frac{x-a}{||x-a||}) \geq \lambda||x-a||^2$.*
8. *Thus,*

$$f(x) > f(a) + \frac{\lambda}{2}||x-a||^2 - \frac{\lambda}{4}||x-a||^2 > f(a).$$

9. *This proves f has a local minimum at a, as required.*

(a) Identify which lines correspond to this crude sketch of the proof.

I) *Since a is a critical point, f is well approximated by its quadratic form q at a.*
II) *The Hessian has only positive eigenvalues, so q has a global minimum at the origin.*
III) *Since q near 0 approximates f near a, this implies f has a local minimum at a.*

(b) There are 3 lines of the proof that directly or indirectly use that all the Hessian eigenvalues are positive. Identify all 3 of these lines.

6.5. Proof of Taylor's theorem

Taylor's theorem is a landmark achievement in differential calculus, illustrating that nonlinear functions can be arbitrarily approximated by high degree polynomials. You will conclude this chapter with its proof. It is a genuinely intricate and beautiful multi-stage argument with several critical ideas, yet the underlying principle is a familiar one.

> *Reduce everything to single variable calculus.*

This mantra continues throughout multivariable calculus, and the proof of Taylor's theorem will be one of its greatest successes. Before embarking any further, it is worthwhile to set the stage for the entire section.

Fix $N \in \mathbb{N}^+$. Let f be C^N on an open ball centred at $a \in \mathbb{R}^n$. Recall the Nth Taylor polynomial of f at a is defined by the explicit formula

$$P_N(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq N}} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha. \tag{6.5.1}$$

Recall if $\alpha \in \mathbb{N}^n$ then $\alpha! = \alpha_1! \cdots \alpha_n! \in \mathbb{N}^+$ and $(x-a)^\alpha = (x_1 - a_1)^{\alpha_1} \cdots (x_n - a_n)^{\alpha_n} \in \mathbb{R}$. The goal is to prove multivariable Taylor's theorem (Theorem 6.3.10), namely

> *A polynomial P is the Nth Taylor polynomial of f at a if and only if P is the unique polynomial of degree $\leq N$ which is an Nth order approximation of f at a.*

The basic proof outline is simple. First, you will prove P_N is an Nth order approximation. Second, you will prove that P_N is the unique degree $\leq N$ polynomial satisfying this property. The second part is a relatively straightforward consequence of polynomials. The first part is the core challenge and will take several steps. The first step is to introduce another explicit formula for P_N, which connects to the single variable Taylor's theorem. The second step is to prove multivariable Lagrange's remainder theorem. The final step is to deduce Taylor's theorem from these two tools. You are ready to begin!

6.5.1 Another explicit formula

For $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$, by Theorem 3.3.10, the directional derivative satisfies

$$D_h f = h_1 \frac{\partial f}{\partial x_1} + \cdots + h_n \frac{\partial f}{\partial x_n} = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}$$

since f is C^N (and hence differentiable) on an open ball centred at $a \in \mathbb{R}^n$. This implies $D_h f$ is C^N on an open ball centred at a. Repeating this, $D_h(D_h f)$ is C^{N-1} and so on. This allows you to define the iterated directional derivative.

Definition 6.5.1 Let f be a real-valued function of class C^k. The **kth iterated directional derivative of** f is a map defined by

$$D_h^k f = \underbrace{D_h(D_h(\cdots(D_h}_{k \text{ times}} f))).$$

You can express this in terms of partials in some small examples.

Example 6.5.2 Let $f:\mathbb{R}^2\to\mathbb{R}$ be C^2. Let $h=(h_1,h_2)\in\mathbb{R}^2$. By Theorem 3.3.10 many times,

$$\begin{aligned} D_h^2 f &= D_h\left(h_1\frac{\partial f}{\partial x}+h_2\frac{\partial f}{\partial y}\right)\\ &= h_1 D_h\left(\frac{\partial f}{\partial x}\right)+h_2 D_h\left(\frac{\partial f}{\partial y}\right)\\ &= h_1\left(h_1\frac{\partial^2 f}{\partial x^2}+h_2\frac{\partial^2 f}{\partial y\partial x}\right)+h_2\left(h_1\frac{\partial^2 f}{\partial x\partial y}+h_2\frac{\partial^2 f}{\partial y^2}\right)\end{aligned}$$

Since f is C^2, Clairaut's theorem implies that $\frac{\partial^2 f}{\partial x\partial y}=\frac{\partial^2 f}{\partial y\partial x}$ so you may expand these expressions, collect the middle terms, and see that

$$D_h^2 f = h_1^2\frac{\partial^2 f}{\partial x^2}+2h_1h_2\frac{\partial^2 f}{\partial x\partial y}+h_2^2\frac{\partial^2 f}{\partial y^2}.$$

If f is C^3, you can use $D_h^2 f$ to calculate $D_h^3 f = D_h(D_h^2 f)$ and find that

$$D_h^3 f = h_1^3\frac{\partial^3 f}{\partial x^3}+3h_1^2h_2\frac{\partial^3 f}{\partial x^2\partial y}+3h_1h_2^2\frac{\partial^3 f}{\partial x\partial y^2}+h_2^3\frac{\partial^3 f}{\partial y^3}.$$

Example 6.5.3 Let f be a real-valued function of class C^k on an open disk at $(a,b)\in\mathbb{R}^2$. You can prove by induction on $k\in\mathbb{N}^+$ that for all $h=(h_1,h_2)\in\mathbb{R}^2$,

$$D_h^k f(a,b)=\sum_{j=0}^{k}\binom{k}{j}h_1^j h_2^{k-j}\frac{\partial^k f}{\partial x^j\partial y^{k-j}}(a,b).$$

With some clever combinatorics, you can extend this pattern to any dimension.

Lemma 6.5.4 *Let f be of class C^k on an open ball at $a\in\mathbb{R}^n$. For all $h\in\mathbb{R}^n$,*

$$\frac{D_h^k f(a)}{k!}=\sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=k}}\frac{\partial^\alpha f(a)}{\alpha!}h^\alpha.$$

Proof. Omitted. A full proof of the theorem will require fancy combinatorics along with either a messy induction or some advanced algebra. To at least give you some idea, here is a pseudo-argument. The elegant combinatorial identity

$$(z_1+\cdots+z_n)^k=\sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=k}}\frac{k!}{\alpha!}z_1^{\alpha_1}\cdots z_n^{\alpha_n}$$

implies that

$$\begin{aligned} D_h^k f(a) = \left(h_1\frac{\partial}{\partial x_1}+\cdots+h_n\frac{\partial}{\partial x_n}\right)^k f(a) &= \Big(\sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=k}}\frac{k!}{\alpha!}h_1^{\alpha_1}\cdots h_n^{\alpha_n}\frac{\partial^k}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}\Big)f(a)\\ &= k!\sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=k}}\frac{\partial^\alpha f(a)}{\alpha!}h^\alpha.\end{aligned}$$

This may be somewhat convincing but you should be skeptical. It skips a lot of steps and pushes symbols without any justification. ■

Assume this lemma and do not worry about its full proof. You can convince yourself with the earlier examples and exercise. Now, Lemma 6.5.4 gives another explicit formula for (6.3.1).

Corollary 6.5.5 Let f be of class C^N on an open ball at $a \in \mathbb{R}^n$. If P_N is the Nth Taylor polynomial of f at a given by (6.3.1) then for $h \in \mathbb{R}^n$,

$$P_N(a+h) = \sum_{k=0}^{N} \frac{D_h^k f(a)}{k!}.$$

Corollary 6.5.5 creates an avenue to apply single variable Taylor's theorem because it expresses a Taylor polynomial in terms of directional derivatives. Remember a directional derivative is a derivative *along a line*. Since a line is a 1-dimensional object, this connection opens up tools from single variable calculus.

6.5.2 Lagrange's remainder theorem

Approximations only hold value if you can estimate their error. For multivariable Taylor polynomials, you can use multivariable Lagrange's remainder theorem to explicitly estimate this error between f and its Taylor polynomial. This theorem is also your hammer to prove Taylor's theorem.

Theorem 6.5.6 (Lagrange's remainder theorem) Let $N \in \mathbb{N}$. If f is of class C^N on an open set D containing the line segment L from $a \in \mathbb{R}^n$ to $a+h \in \mathbb{R}^n$, then there exists a point[6] $\xi \in L$ such that

$$f(a+h) = P_{N-1}(a+h) + \frac{D_h^N f(\xi)}{N!}.$$

This proof exposition mostly follows [6, Section II.7 Theorem 7.1].

Proof. Since L is contained in D, there exists an open interval I containing $[0,1]$ such that

$$L = \{a+th : t \in [0,1]\} \subseteq \{a+th : t \in I\} \subseteq D.$$

Define the C^∞ function $\varphi : I \to \mathbb{R}^n$ by

$$\varphi(t) = a+th$$

so $L \subseteq \varphi(I) \subseteq D$ by definition of the interval I. Define $g = f \circ \varphi : I \to \mathbb{R}$. We claim that

$$\forall 1 \le k \le N, \quad g^{(k)}(t) = D_h^k f(a+th). \tag{6.5.2}$$

Assume the claim. Since f is C^N, it follows that g is C^N. Since $g(0) = f(a)$, $g(1) = f(a+h)$, and g is C^N on an open interval containing $[0,1]$, Lagrange's remainder theorem in one variable (see e.g. this MAT137 video or [6, II.6 Theorem 6.1]) implies that

$$g(1) = \sum_{k=0}^{N-1} \frac{g^{(k)}(0)}{k!} + \frac{g^{(N)}(c)}{N!}$$

[6]The symbol ξ is the Greek letter 'xi' and pronounced *ksaai* or *ksee*. It is notoriously difficult to legibly write by hand, especially if you want to distinguish it from the symbol ζ which is the Greek letter 'zeta'. When I started, I found it helpful to handwrite ξ as "triple squiggle" and ζ as "double squiggle".

for some $c \in (0,1)$. Define $\xi = \varphi(c) \in L$. From the claim (6.5.6), it follows that

$$\begin{aligned} f(a+h) &= \sum_{k=0}^{N-1} \frac{D_h^k f(a)}{k!} + \frac{D_h^N f(\varphi(c))}{N!} \\ &= P_{N-1}(a+h) + \frac{D_h^N f(\xi)}{N!} && \text{by Corollary 6.5.5.} \end{aligned}$$

This completes the proof assuming the claim (6.5.6).

It suffices to prove (6.5.6) by induction on k. Since φ and f are C^N on their domains, it follows by the chain rule that for $t \in I$,

$$\begin{aligned} g'(t) &= \nabla f(\varphi(t)) \cdot \varphi'(t) \\ &= \nabla f(a+th) \cdot h \\ &= D_h f(a+th). \end{aligned}$$

This proves g is C^1 and $g'(0) = D_h f(a)$. Assume (6.5.6) holds for some $1 \leq k \leq N-1$. Define $F = D_h^k f$ so $g^{(k)} = F \circ \varphi$. As F is C^{N-k} (and hence differentiable) and φ is C^∞, it follows by the chain rule that

$$\begin{aligned} g^{(k+1)}(t) = \frac{d}{dt} g^{(k)}(t) &= \frac{d}{dt}\big(F \circ \varphi(t)\big) \\ &= \nabla F(\varphi(t)) \cdot \varphi'(t) \\ &= \nabla F(a+th) \cdot h \\ &= D_h F(a+th) \\ &= D_h^{k+1} f(a+th) && \text{by definition of } F. \end{aligned}$$

This proves (6.5.6) and hence completes the proof. ■

Lagrange's remainder theorem is your hammer and now it is time to wield it.

6.5.3 Completing the proof

You need one last ingredient on distinguishing polynomials.

Lemma 6.5.7 Let Q be a polynomial in n variables of degree $\leq N$. Then Q is the zero polynomial if and only if

$$\lim_{x \to 0} \frac{Q(x)}{||x||^N} = 0. \tag{6.5.3}$$

Heuristically speaking, if Q is not the zero polynomial, then Q should contain a non-zero monomial x^α with degree $|\alpha| \leq N$. The limit

$$\lim_{x \to 0} \frac{x^\alpha}{||x||^N}$$

does not exist, so (6.5.3) cannot hold. The formal proof requires some attention to detail.

Proof. The "only if" direction is immediate. For the "if" direction, assume $Q(x)/||x||^N \to 0$ as $x \to 0$ and suppose, for a contradiction, that Q is not the zero polynomial. By Lemma 6.2.18,

$$Q(x) = \sum_{k=0}^{N} Q_k(x), \qquad \text{where} \qquad Q_k(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = k}} \frac{\partial^\alpha Q(0)}{\alpha!} x^\alpha.$$

In other words, Q_k has all the degree k terms for $k \in \{0, 1, \ldots, N\}$. Since Q is not the zero polynomials, there exists a minimum $\ell \in \{0, 1, \ldots, N\}$ such that Q_ℓ is non-zero in which case $Q = \sum_{k=\ell}^{N} Q_k$. Fix a point $b \in \mathbb{R}^n \setminus \{0\}$ with $Q_\ell(b) \neq 0$. For $t > 0$ and any multi-index $\alpha \in \mathbb{N}^n$, notice that $(tb)^\alpha = (tb_1)^{\alpha_1} \cdots (tb_n)^{\alpha_n} = t^{\alpha_1 + \cdots + \alpha_n} b_1^{\alpha_1} \cdots b_n^{\alpha_n} = t^{|\alpha|} b^\alpha$, so

$$Q_k(tb) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = k}} \frac{\partial^\alpha Q(0)}{\alpha!} t^{|\alpha|} b^\alpha = t^k Q_k(b).$$

Since $Q(x)/||x||^N \to 0$ as $x \to 0$ implies that $Q(x)/||x||^\ell \to 0$ as $x \to 0$, it follows that

$$0 = \lim_{t \to 0^+} \frac{Q(tb)}{||tb||^\ell} = \sum_{k=\ell}^{N} \lim_{t \to 0^+} \frac{Q_k(tb)}{||tb||^\ell} = \sum_{k=\ell}^{N} \lim_{t \to 0^+} \frac{t^k Q_k(b)}{t^\ell ||b||^\ell} = \frac{Q_\ell(b)}{||b||^\ell}.$$

The righthand side is non-zero, which is a contradiction. ■

You are finally ready to assemble a masterpiece.

Proof of Taylor's theorem. Let f be of class C^N on an open ball at $a \in \mathbb{R}^n$. Let P_N be the Nth Taylor polynomial of f at a. Fix $\varepsilon > 0$. For each multi-index $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq N$, as f is C^N, there exists $\delta_\alpha > 0$ such that

$$\forall x \in \mathbb{R}^n, ||x - a|| < \delta_\alpha \implies |\partial^\alpha f(x) - \partial^\alpha f(a)| < \frac{\varepsilon}{(N+1)^n}. \tag{6.5.4}$$

Take $\delta := \min\{\delta_\alpha : \alpha \in \mathbb{N}^n, |\alpha| = N\} > 0$. To prove P_N is an Nth order approximation of f, it suffices to prove that

$$\forall x \in \mathbb{R}^n, ||x - a|| < \delta \implies \left| \frac{f(x) - P_N(x)}{||x - a||^N} \right| < \varepsilon. \tag{6.5.5}$$

Fix $x \in \mathbb{R}^n$ satisfying $||x - a|| < \delta$. By Lagrange's remainder theorem (Theorem 6.5.6), there exists a point ξ on the line segement from a to $x = a + h$ such that

$$f(a+h) = P_{N-1}(a+h) + \frac{D_h^N f(\xi)}{N!}.$$

Note $||\xi - a|| \leq ||x - a|| < \delta$ by assumption. By Corollary 6.5.5,

$$P_N(a+h) = P_{N-1}(a+h) + \frac{D_h^N f(a)}{N!}.$$

Taking the difference of these equalities, it follows by Lemma 6.5.4 that

$$\begin{aligned} \frac{f(x) - P_N(x)}{||x - a||^N} &= \frac{f(a+h) - P_N(a+h)}{||h||^N} \\ &= \frac{D_h^N f(\xi) - D_h^N f(a)}{N! ||h||^N} \\ &= \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = N}} \frac{\partial^\alpha f(\xi) - \partial^\alpha f(a)}{\alpha!} \frac{h^\alpha}{||h||^N}. \end{aligned}$$

Applying absolute values and the triangle inequality, we have that

$$\begin{aligned}
\left|\frac{f(x)-P_N(x)}{||x-a||^N}\right| &\leq \sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=N}} \frac{|\partial^\alpha f(\xi)-\partial^\alpha f(a)|}{\alpha!}\frac{|h^\alpha|}{||h||^N} \\
&\leq \sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=N}} |\partial^\alpha f(\xi)-\partial^\alpha f(a)| && \text{since } |h^\alpha| \leq ||h||^{|\alpha|} \text{ and } \alpha! \geq 1, \\
&< \sum_{\substack{\alpha\in\mathbb{N}^n\\|\alpha|=N}} \frac{\varepsilon}{(N+1)^n} && \text{by (6.5.4) and } ||\xi - a|| < \delta, \\
&\leq \varepsilon && \text{as } |\{\alpha\in\mathbb{N}^n : |\alpha| = N\}| \leq (N+1)^n.
\end{aligned}$$

In the second last step, we used that $\delta \leq \delta_\alpha$ for every $\alpha\in\mathbb{N}^n$ with $|\alpha| = N$. This proves (6.5.5).

It remains to show P_N is the unique degree $\leq N$ polynomial with this property. Assume P is another degree $\leq N$ polynomial which is an Nth order approximation of f at a. Hence,

$$\lim_{h\to 0}\frac{P_N(a+h)-P(a+h)}{||h||^N} = \lim_{h\to 0}\frac{P_N(a+h)-f(a+h)}{||h||^N} + \lim_{h\to 0}\frac{f(a+h)-P(a+h)}{||h||^N} = 0+0,$$

since P_N and P are both Nth order approximations of f at a. Applying Lemma 6.5.7 with the degree $\leq N$ polynomial $Q = P_N - P$ implies that $P_N = P$, as desired. ■

Congratulations on reaching the summit of multivariable differential calculus! You have created a diverse and rich world of derivatives modeling the real world, established a bevy of global and local optimization techniques, extracted a great definition for surfaces, solved arbitrary nonlinear systems, and constructed polynomial approximations to any nonlinear map. That is an impressive list of mathematical accomplishments, so take a moment to appreciate it. Your multivariable calculus odyssey shall continue with the wonderful world of integrals.

Exercises for Section 6.5

Concepts and definitions

6.5.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Let $a \in \mathbb{R}^n$ and $N \in \mathbb{N}^+$. If it exists, let P_N be the Nth Taylor polynomial of f at a and let R_N be the Nth remainder of f at a. Which are true or false?

(a) Fix $i \in \{1, \dots, n\}$. If f is C^k then $\dfrac{\partial f}{\partial x_i}$ is C^{k-1}.

(b) Fix $h \in \mathbb{R}^n$. If f is C^N then $D_h f$ is C^{N-1}.

(c) Fix $h \in \mathbb{R}^n$ and $k \in \{1, \dots, N\}$. If f is C^N then $D_h^k f$ is C^{N-k}.

(d) If $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 and $\varphi : \mathbb{R} \to \mathbb{R}^n$ is C^1 then $f \circ \varphi : \mathbb{R} \to \mathbb{R}$ is C^1.

(e) If f is C^N on an open ball of a then $\displaystyle\lim_{x\to a} \frac{R_N(x)}{||x-a||^N} = 0$.

(f) If Q is a polynomial in n variables of degree $\le N$ such that $\displaystyle\lim_{x\to 0} \frac{Q(x)}{||x||^N} = 0$, then Q is the zero polynomial.

Proofs

6.5.2 Let f be a real-valued function of class C^k on an open disk at $(a,b) \in \mathbb{R}^2$. Prove by induction on $k \in \mathbb{N}^+$ that for all $h = (h_1, h_2) \in \mathbb{R}^2$,

$$D_h^k f(a,b) = \sum_{j=0}^{k} \binom{k}{j} h_1^j h_2^{k-j} \frac{\partial^k f}{\partial x^j \partial y^{k-j}}(a,b).$$

6.5.3 You will analyze the proof of Lagrange's remainder theorem by filling in some details and drawing informative pictures. The proof is tough so you will analyze it in stages.

Lagrange's remainder theorem. *Let $N \in \mathbb{N}^+$. If f is C^N on an open set D containing the line segment L from $a \in \mathbb{R}^n$ to $a + h \in \mathbb{R}^n$ then there exists a point $\xi \in L$ such that*

$$f(a+h) = P_{N-1}(a+h) + \frac{D_h^N f(\xi)}{N!}.$$

1. *Since L is contained in D, there exists an open interval I containing $[0,1]$ such that*

$$L = \{a + th : t \in [0,1]\} \subseteq \{a + th : t \in I\} \subseteq D.$$

2. *Define the C^∞ function $\varphi : I \to \mathbb{R}^n$ by $\varphi(t) = a + th$.*
3. *So $L \subseteq \varphi(I) \subseteq D$ by definition of the interval I.*

(a) Illustrate this stage with a picture of D. Label your picture with everything that you can.

(b) Explain why such an open interval must exist. Use your diagram for insights.

4. *Recall $g = f \circ \varphi : I \to \mathbb{R}$. We claim that*

$$\forall 1 \le k \le N, \quad g^{(k)}(t) = D_h^k f(a + th).$$

5. *Assume the claim. Since f is C^N, it follows that g is C^N.*

(c) Explain in a bit more detail why line 5 follows from line 4. You will need a line from 1 to 3.

(d) Draw a new picture based on 6.5.3.1 using $g = f \circ \varphi$. Illustrate a two-step transformation going from the open interval $I \subseteq \mathbb{R}$ then to the open set $D \subseteq \mathbb{R}^n$ then to $\mathbb{R}$.

6. *Since $g(0) = f(a), g(1) = f(a+h)$, and g is C^N on an open interval containing $[0,1]$, Lagrange's remainder theorem in one variable implies for some $c \in (0,1)$ that*

$$g(1) = \sum_{k=0}^{N-1} \frac{g^{(k)}(0)}{k!} + \frac{g^{(N)}(c)}{N!}.$$

7. *If we define $\xi = \varphi(c) \in L$ then from the claim (6.5.6) it follows that*

$$f(a+h) = \sum_{k=0}^{N-1} \frac{D_h^k f(a)}{k!} + \frac{D_h^N f(\varphi(c))}{N!} = P_N(a+h) + \frac{D_h^N f(\xi)}{N!} \qquad \textit{by Corollary 1.}$$

8. *By definition of the Nth remainder R_N, this completes the proof assuming the claim (6.5.6).*

(e) A version of Lagrange's remainder theorem in one variable states that

> *Let $J \subseteq \mathbb{R}$ be an open interval and $a, x \in J$. If h is C^N on J then $\exists c$ between a and x such that*
> $$h(x) = \sum_{k=0}^{N-1} \frac{h^{(k)}(a)}{k!}(x-a)^k + \frac{h^{(N)}(c)}{N!}(x-a)^N.$$

Explain how the author applies this theorem in line 6 by identifying the choices for h, J, a, x.

(f) Review lines 1 to 8. Why did you need the open interval I from line 1? Identify the key line.

(g) Add $\xi = \varphi(c)$ and any other related points to your picture in 6.5.3.4.

A technical claim remains: *Define $\varphi : I \to \mathbb{R}$ by $\varphi(t) = a + th$ and $g = f \circ \varphi : I \to \mathbb{R}$. Then*

$$\forall 1 \le k \le N, \quad g^{(k)}(t) = (D_h^k f)(a+th). \tag{6.5.6}$$

Recall I is an open interval of $\mathbb{R}$, f is C^N on the open set D, and $\varphi(I) \subseteq D$ by assumption.

9. *It suffices to prove (6.5.6) by induction on k.*
10. *Since φ and f are C^N on their domains, it follows by the chain rule that for $t \in I$,*

$$g'(t) = \nabla f(\varphi(t)) \cdot \varphi'(t) = \nabla f(a+th) \cdot h = (D_h f)(a+th).$$

11. *This proves g is C^1 and $g'(0) = D_h f(a)$.*
12. *Assume (6.5.6) holds for some $1 \le k \le N-1$. Define $F = D_h^k f$ so $g^{(k)} = F \circ \varphi$.*
13. *As F is C^{N-k} (and hence differentiable) and φ is C^∞, it follows by the chain rule that*

$$\begin{aligned} g^{(k+1)}(t) = \frac{d}{dt} g^{(k)}(t) = \frac{d}{dt}\big(F \circ \varphi(t)\big) &= \nabla F(\varphi(t)) \cdot \varphi'(t) \\ &= \nabla F(a+th) \cdot h \\ &= (D_h F)(a+th) \\ &= (D_h^{k+1} f)(a+th) \qquad \textit{by definition of } F. \end{aligned}$$

14. *This proves (6.5.6) and hence completes the proof.*

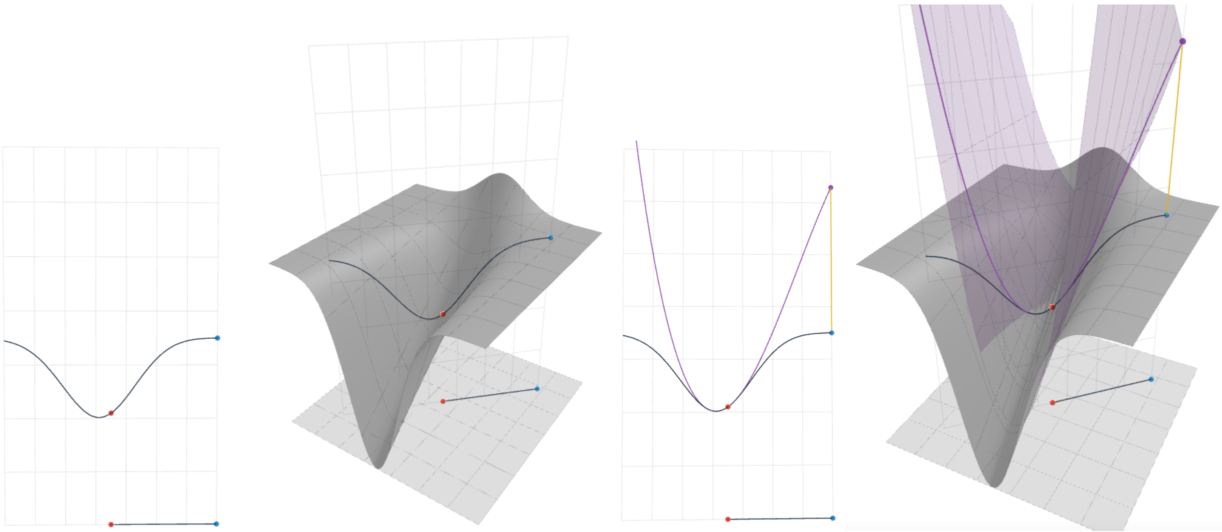
(h) Identify which lines corresponds to the base case and which correspond to the induction step. Pinpoint precisely where the inductive hypothesis is applied.

(i) Lines 10 and 13 could use more justifications next to the centered equations. Add them with specific references to theorems or definitions when possible.

6.5.4 The proof of Lagrange's remainder theorem appears complicated because there are many technical pieces to handle. However, the entire idea for this vast proof is simple:

Multivariable Lagrange follows by single variable Lagrange along a line segment.

You should process this big picture after finishing Question 6.5.3. Here you will label visuals for $f : \mathbb{R}^2 \to \mathbb{R}$ with notation from the proof to clarify how each of the technical pieces fit together.



You will use the four graphs above for parts (a) and (b) below.

(a) Use lines 1 to 3 from Question 6.5.3 to label the Math3D graphs included below.

- Label the leftmost graph using $g = f \circ \varphi$ and $[0, 1]$.
- Label the middle left graph using f, D, L, φ, a, and $a + h$.

(b) Use lines 6 to 8 from Question 6.5.3 to label the Math3D graphs included below.

- Label the middle right graph with quantities involving g and c.
- Label the rightmost graph using quantities involving f, φ, ξ, a, and $a + h$.

Do not repeat the same labels as the previous part.

(c) Finally, watch this Math3D proof of single variable Taylor's theorem. Toggle the surfaces.

6.5.5 You will now analyze the proof of Taylor's theorem.

Taylor's theorem. *Let f be of class C^N on an open ball at $a \in \mathbb{R}^n$. Then P_N is the Nth Taylor polynomial of f at a if and only if P_N is the unique degree $\leq N$ polynomial which is an Nth order approximation of f at a.*

This proof has two main steps. First, P_N is an Nth order approximation of f at a.

1. *Let P_N be the Nth Taylor polynomial of f at a. Fix $\varepsilon > 0$.*
2. *For each multi-index $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq N$, there exists $\delta_\alpha > 0$ such that*
$$\forall x \in \mathbb{R}^n, ||x-a|| < \delta_\alpha \implies |\partial^\alpha f(x) - \partial^\alpha f(a)| < \frac{\varepsilon}{(N+1)^n}.$$
3. *Take $\delta := \min\{\delta_\alpha : \alpha \in \mathbb{N}^n, |\alpha| = N\} > 0$.*
4. *Fix $x \in \mathbb{R}^n$ satisfying $||x-a|| < \delta$. By Lagrange's remainder theorem (Theorem 6.5.6), there exists a point ξ on the line segement from a to $x = a+h$ such that*
$$f(a+h) = P_{N-1}(a+h) + \frac{D_h^N f(\xi)}{N!}.$$
5. *By Corollary 6.5.5, $P_N(a+h) = P_{N-1}(a+h) + \dfrac{D_h^N f(a)}{N!}$.*
6. *It follows by Lemma 6.5.4 that*
$$\frac{f(x) - P_N(x)}{||x-a||^N} = \frac{D_h^N f(\xi) - D_h^N f(a)}{N!||h||^N} = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = N}} \frac{\partial^\alpha f(\xi) - \partial^\alpha f(a)}{\alpha!} \frac{h^\alpha}{||h||^N}.$$
7. *Applying absolute values and the triangle inequality, we have that*
$$\begin{aligned}\left|\frac{f(x) - P_N(x)}{||x-a||^N}\right| \leq \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = N}} \frac{|\partial^\alpha f(\xi) - \partial^\alpha f(a)|}{\alpha!} \frac{|h^\alpha|}{||h||^N} &\leq \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = N}} |\partial^\alpha f(\xi) - \partial^\alpha f(a)| \\ &< \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = N}} \frac{\varepsilon}{(N+1)^n} \leq \varepsilon.\end{aligned}$$

(a) Line 2 uses that f is C^N. Explain how.

(b) Line 3 choose δ. At what exact moment is this choice applied?

(c) Line 4 selects ξ on the line segment from a to x. Formally prove that $||\xi - a|| \leq ||x-a||$.

(d) Line 7 uses that $|h^\alpha| \leq ||h||^{|\alpha|}$ and $|\{\alpha \in \mathbb{N}^n : |\alpha| = N\}| \leq N^n$. Explain both inequalities.

Second, P_N is the unique degree $\leq N$ polynomial which is an Nth order approximation.

8. *This proves P_N is an Nth order approximation of f at a.*
9. *It remains to show P_N is the unique degree $\leq N$ polynomial with this property.*
10. *Assume P is another degree $\leq N$ polynomial which is an Nth order approximation of f at a.*
11. *By the sum limit law,*
$$\lim_{h\to 0} \frac{P_N(a+h) - P(a+h)}{||h||^N} = \lim_{h\to 0} \frac{P_N(a+h) - f(a+h)}{||h||^N} + \lim_{h\to 0} \frac{f(a+h) - P(a+h)}{||h||^N} = 0 + 0,$$
since P_N and P are both Nth order approximations of f at a.
12. *Applying Lemma 6.5.7 with the degree $\leq N$ polynomial $Q = P_N - P$ implies that $P_N = P$.*

(e) Modify lines 9 and 10 to establish a slightly stronger version of Taylor's theorem.

> *Let f be C^N on an open ball at $a \in \mathbb{R}^n$. The Nth Taylor polynomial of f at a is the smallest degree polynomial which is an Nth order approximation of f at a.*

III

Integral calculus

7. Integrals

Differential calculus describes rates of change. Integral calculus describes volumes, net totals, and averages. This chapter builds the foundation for the definition of the integral. To qualify as a good definition, the integral must address three fundamental questions.

> *What is the volume of a set S in $\mathbb{R}^n$?*
> *What is the total mass of an object S in $\mathbb{R}^n$ with variable density $f : S \to [0, \infty)$?*
> *What is the average of a real-valued function f over a region S in $\mathbb{R}^n$?*

The multivariable integral should also effortlessly generalize the single variable integral. Building this definition will be complicated but the key idea is natural.

> *Chop the region $S \subseteq \mathbb{R}^n$ into little pieces. Estimate the value of f on each piece. Total your estimates to approximate the total value of f over the region S. Repeat this process for finer and finer pieces to approach the exact answer.*

An abbreviated version of this process forms a memorable mantra:

> *Chop. Estimate. Refine.*

The idea is great but will you always get the *same* answer? What if you chop in another way? What if you estimate each piece differently? What if you refine your chopping by some other approach? The goal of this chapter is to build a robust definition of the integral which always gives the same answer, permits many choices of functions or regions, and allows you to chop, estimate, or refine in many different ways. This will be a monumental task but, as you shall see, it is well worth the effort. In the current and subsequent chapters, you will wield this rigorous theory to address the original three questions above and explore other applications.

7.1. Partitions

Generalizing the integral to its full form in $\mathbb{R}^n$ will take multiple stages. The first stage is to define integration over the most basic shape: a rectangle. This will take several sections to carefully unfold and you will follow the natural process of chopping, estimating, and refining. In this section, you will focus on step one of this stage.

> *Chop the rectangle into little pieces.*

Chopping is the source of all the complications for integration in higher dimensions, because you have many more different ways to chop. You will need more notation, more definitions, and more subtle proofs, but this need not be overwhelming. Luckily, you can rely on intuition from $\mathbb{R}$ and pictures in $\mathbb{R}^2$ to guide your thinking.

7.1.1 Constructing partitions

To build the concepts slowly, you will define partitions in $\mathbb{R}$, then $\mathbb{R}^2$, and then $\mathbb{R}^n$. Chopping in $\mathbb{R}$ is much simpler.

Definition 7.1.1 A **rectangle** in $\mathbb{R}$ is a closed interval $[a,b]$ where $a,b \in \mathbb{R}$ and $a < b$. The **length of** $[a,b]$ is defined to be

$$\text{length}([a,b]) = b - a.$$

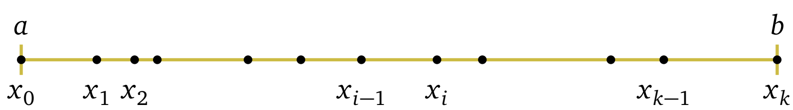
A **partition** P of the interval $[a,b]$ is a finite set such that $\{a,b\} \subseteq P \subseteq [a,b]$. Write $P = \{x_0, x_1, \ldots, x_k\}$ where $a = x_0 < x_1 < \cdots < x_{k-1} < x_k = b$.
- The **index set of** P is the set $\{1, 2, \ldots, k\}$.
- For $i \in \{1, \ldots, k\}$, the i**-subrectangle of** P is given by $R_i = [x_{i-1}, x_i]$.
- The **subrectangles of** P is the set of rectangles $\{R_1, \ldots, R_k\}$.

Remark 7.1.2 The above definition of a partition includes several additional terms, because this extra data will be helpful when generalizing to higher dimensions. Note that rectangles (resp. subrectangles) in $\mathbb{R}$ are also referred to as **intervals** (resp. **subintervals**).

Remark 7.1.3 When introducing a partition $P = \{x_0, x_1, \ldots, x_k\}$ of $[a,b]$, it will often be implicitly assumed that $a = x_0 < x_1 < \cdots < x_{k-1} < x_k = b$ without explicit mention.

Example 7.1.4 The interval $R = [1,4]$ has many partitions. For example, $P_1 = \{1,2,3,4\}$ and $P_2 = \{1, e, \pi, 4\}$ are both partitions of I. The partition $P_3 = \{1,4\}$ of R is the *trivial* partition. For $N \in \mathbb{N}^+$, the partition $P = \{x_0, \ldots, x_N\}$ of R defined by $x_i = 1 + \frac{3i}{N}$ for $i \in \{1, \ldots, N\}$ is a *regular* partition since each subinterval has the same length, namely $x_i - x_{i-1} = \frac{3}{N}$.

You can visualize chopping in $\mathbb{R}$ with a line segment.



Physically speaking, this looks like cutting a string with a knife at many points. This visual also suggests some basic properties and a formula. Namely, if $P = \{x_0, \ldots, x_k\}$ is a partition of the interval $[a,b]$, then

$$[a,b] = \bigcup_{i=1}^{k} [x_{i-1}, x_i], \tag{7.1.1}$$

$$\forall i,j \in \{1,\dots,k\},\ i \neq j \implies \operatorname{int}([x_{i-1},x_i] \cap [x_{j-1},x_j]) = \varnothing, \tag{7.1.2}$$

$$\operatorname{length}([a,b]) = \sum_{i=1}^{k} \operatorname{length}([x_{i-1},x_i]). \tag{7.1.3}$$

Recall $\operatorname{int}(S)$ is equivalent notation for the interior S^o of a set S. You can prove (7.1.1) and (7.1.2) using standard arguments with sets and topology. A proof of (7.1.3) involves standard manipulations with sigma notation.

Before generalizing to any dimension, you will focus on the two-dimensional setting so you can take advantage of visuals and temporarily keep notation to a minimum. Partitions in $\mathbb{R}^2$ are defined using partitions in $\mathbb{R}$.

Definition 7.1.5 A **rectangle** in $\mathbb{R}^2$ is a set R of the form $R = [a,b] \times [c,d]$ where $a,b,c,d \in \mathbb{R}$ and $a < b$ and $c < d$. The **area** of a rectangle is defined to be

$$\operatorname{area}(R) = (b-a)(d-c).$$

An ordered pair of sets $P = (P_1, P_2)$ is a **partition** of $R = [a,b] \times [c,d]$ if $P_1 = \{x_0, x_1, \dots, x_k\}$ is a partition of the interval $[a,b]$ and $P_2 = \{y_0, y_1, \dots, y_\ell\}$ is a partition of the interval $[c,d]$.

- The **index set of** P is the set $I = \{(i,j) \in \mathbb{N}^2 : 1 \le i \le k, 1 \le j \le \ell\}$.
- For $(i,j) \in I$, the (i,j)**-subrectangle of** P is given by

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j].$$

- The **subrectangles of** P is the set of rectangles $\{R_{ij} : 1 \le i \le k, 1 \le j \le \ell\}$

Remark 7.1.6 When introducing a partition P of a rectangle R, you often only need to mention the subrectangles of P. These are often written informally as $\{R_{ij}\}_{i,j}$ which avoids introducing all of the real numbers x_i and y_j if you do not need to refer to them.

On an unrelated note, the (i,j)-subrectangle of P should technically be denoted as $R_{(i,j)}$ but the equivalent notation R_{ij} is acceptable as there is usually no ambiguity.

Example 7.1.7 The rectangle $R = [0,1] \times [2,4]$ has $\operatorname{area}(R) = (1-0)(4-2) = 2$ and there are an infinite number of ways to partition R. For example,

$$P = (P_1, P_2) = \left(\{0, \tfrac{1}{2}, 1\}, \{2,3,4\}\right) \quad \text{and} \quad P' = (P_1', P_2') = \left(\{0, \tfrac{1}{4}, 1\}, \{2,4\}\right)$$

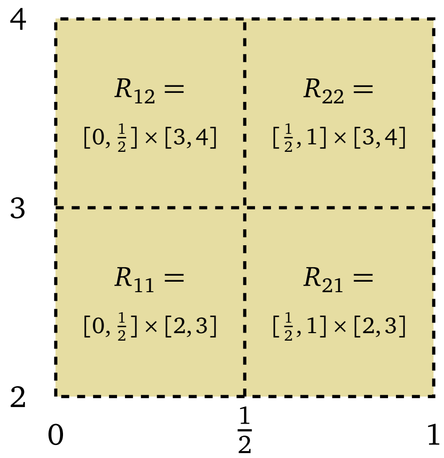
are partitions of R. The index set of P and subrectangles of P are respectively

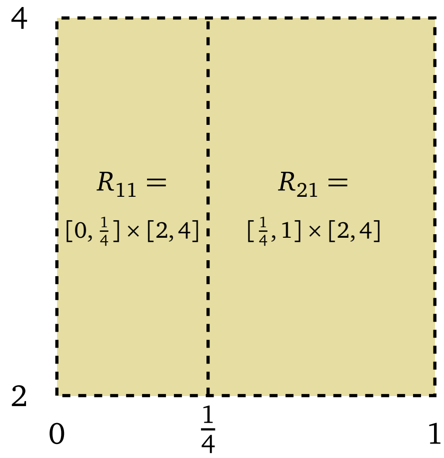
$$\{(1,1),(1,2),(2,1),(2,2)\}, \quad \{[0,\tfrac{1}{2}] \times [2,3], [0,\tfrac{1}{2}] \times [3,4], [\tfrac{1}{2},1] \times [2,3], [\tfrac{1}{2},1] \times [3,4]\}$$

The index set of P' and subrectangles of P' are respectively

$$\{(1,1),(2,1)\}, \qquad \{[0,\tfrac{1}{4}] \times [2,4], [\tfrac{1}{4},1] \times [2,4]\}.$$

Each partition is illustrated below with P on the left and P' on the right.





Notice P partitions R into subrectangles with identical dimensions, whereas P' partitions R into two subrectangles with different dimensions. Below are some other ways to partition R.

- $P'' = (\{0, 1-\varepsilon, 1\}, \{2,4\})$ for any fixed $\varepsilon \in (0,1)$.
- $P''' = (\{0,1\}, \{2,4\})$.

Notice that P''' is a partition that makes no "cuts". It simply creates the same rectangle. This trivially satisfies the definition of a partition and is called the *trivial* partition of R.

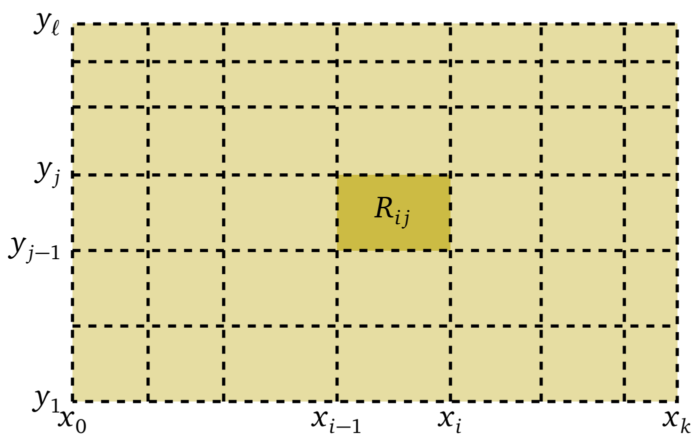
This example suggests properties paralleling the one-dimensional case. Namely, if $P = (P_1, P_2)$ is a partition of $[a, b] \times [c, d]$ with $P_1 = \{x_0, x_1, \ldots, x_k\}$ and $P_2 = \{y_0, y_1, \ldots, y_\ell\}$, then

$$[a,b] \times [c,d] = \bigcup_{i=1}^{k} \bigcup_{j=1}^{\ell} R_{ij} \tag{7.1.4}$$

$$\forall (i,j) \in I, \forall (i',j') \in I, \quad (i,j) \neq (i',j') \implies \operatorname{int}(R_{ij} \cap R_{i'j'}) = \varnothing \tag{7.1.5}$$

$$\operatorname{area}([a,b] \times [c,d]) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \operatorname{area}(R_{ij}), \tag{7.1.6}$$

where $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ for $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \ell\}$. These properties are geometrically natural. The partition P chops up R into subrectangles R_{ij} such that if you were to "combine" all of these subrectangles together, you would get back R. Adding the area of each subrectangle, you should get the total area of R. Below is an illustration.



Physically, this looks like cutting a chocolate bar with a knife while only making cuts parallel to its edges. Identity (7.1.6) states that the total amount of chocolate remains the same before and after cutting. You can formally prove this algebraically.

Finally, you can construct partitions in $\mathbb{R}^n$. Chopping in higher dimensions takes some notational finesse but nothing more.

Definition 7.1.8 A **rectangle** in $\mathbb{R}^n$ is a set R of the form $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ where $a_1, b_1, \ldots, a_n, b_n \in \mathbb{R}$ and $a_i < b_i$ for all $1 \le i \le n$. The **volume** of a rectangle is defined to be

$$\operatorname{vol}(R) = (b_1 - a_1)(b_2 - a_2)\cdots(b_n - a_n).$$

An ordered n-tuple of sets $P = (P_1, \ldots, P_n)$ is a **partition** of R if $P_j = \left\{x_0^{(j)}, x_1^{(j)}, \ldots, x_{k_j}^{(j)}\right\}$ is a partition of the closed interval $[a_j, b_j]$ for each $j \in \{1, \ldots, n\}$.

- The **index set of** P is the set $I = \{(i_1, \ldots, i_n) \in \mathbb{N}^n : 1 \le i_1 \le k_1, \ldots, 1 \le i_n \le k_n\}$.
- For $i = (i_1, \ldots, i_n) \in I$, the i**-subrectangle of** P is given by

$$R_i = R_{(i_1,\ldots,i_n)} := \left[x_{i_1-1}^{(1)}, x_{i_1}^{(1)}\right] \times \left[x_{i_2-1}^{(2)}, x_{i_2}^{(2)}\right] \times \cdots \times \left[x_{i_n-1}^{(n)}, x_{i_n}^{(n)}\right].$$

- The **subrectangles of** P is the set of rectangles $\{R_i : i \in I\}$.

Remark 7.1.9 You may also write $\{R_i\}_i$ or $\{R_i\}_{i\in I}$ for the subrectangles of P. Moreover, the word "volume" can be used for rectangles in $\mathbb{R}$ or $\mathbb{R}^2$ but you may also use "length" for $\mathbb{R}$ and "area" for $\mathbb{R}^2$.

All of the previous properties of partitions (and their proofs) can be generalized.

Theorem 7.1.10 Let R be a rectangle in $\mathbb{R}^n$. Let P be a partition of R. Let I be the index set of P and let $\{R_i : i \in I\}$ be the subrectangles of P.

(a) $R = \bigcup_{i\in I} R_i$

(b) $\forall i, j \in I,\ i \ne j \implies \operatorname{int}(R_i \cap R_j) = \varnothing$

(c) $\operatorname{vol}(R) = \sum_{i\in I} \operatorname{vol}(R_i)$

Proof. Write $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ and $P = (P_1, \ldots, P_n)$ where $P_j = \{x_0^{(j)}, \ldots, x_{k_j}^{(j)}\}$ is a partition of $[a_j, b_j]$ for each $j \in \{1, \ldots, n\}$.

(a) Fix $i \in I$ and $y = (y_1, \ldots, y_n) \in R_i$. For $j \in \{1, \ldots, n\}$, note by definition that $y_j \in [x_{i_j-1}^{(j)}, x_{i_j}^{(j)}]$ so, since $[x_{i_j-1}^{(j)}, x_{i_j}^{(j)}] \subseteq [a_j, b_j]$, it follows that $y_j \in [a_j, b_j]$. Hence, $y = (y_1, \ldots, y_n) \in [a_1, b_1] \times \cdots \times [a_n, b_n] = R$. This proves $R_i \subseteq R$. As $i \in I$ was arbitrary, $\cup_{i\in I} R_i \subseteq R$.

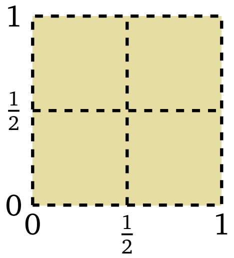
For the reverse containment, fix $z = (z_1, \ldots, z_n) \in R$. For $j \in \{1, \ldots, n\}$, since $z_j \in [a_j, b_j]$, it follows by definition of a partition in $\mathbb{R}$ that there exists $i_j \in \{1, \ldots, k_j\}$ such that $z_j \in [x_{i_j-1}^{(j)}, x_{i_j}^{(j)}]$. Therefore, $z = (z_1, \ldots, z_n) \in R_{(i_1,\ldots,i_n)}$ for some $(i_1, \ldots, i_n) \in I$, as required.

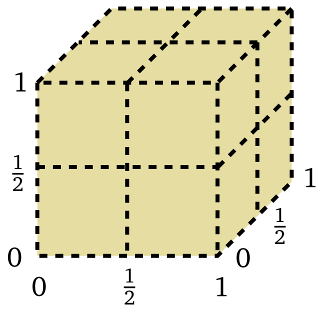
(b) Fix $i, j \in I$ and assume $i = (i_1, \ldots, i_n) \ne (j_1, \ldots, j_n) = j$. Without loss of generality, assume $i_1 \ne j_1$. In that case, by definition of a partition over $\mathbb{R}$, the subset of real numbers $A := [x_{i_1-1}^{(1)}, x_{i_1}^{(1)}] \cap [x_{j_1-1}^{(1)}, x_{j_1}^{(1)}]$ is either $\varnothing$ or equal to $\{x_{i_1}^{(1)}\}$ or equal to $\{x_{j_1}^{(1)}\}$. Since $R_i \cap R_j \subseteq A \times \mathbb{R}^{n-1}$, it suffices to prove that $A \times \mathbb{R}^{n-1}$ has empty interior; this is left as an exercise.

(c) This is left as an exercise. There will be n telescoping sums. ■

Writing explicit examples of partitions in $\mathbb{R}^n$ can be difficult to do. It is helpful to first try lower dimensional examples and then generalize those attempts.

Example 7.1.11 There are many ways to partition the unit cube $R = [0, 1]^n$ into 2^n subrectangles. Here you will construct a simple way to do so. You can start by investigating the lower dimensional cases $n = 2$ and $n = 3$. Note $[0,1]^2$ is a unit square in $\mathbb{R}^2$, which you want to partition into $2^2 = 4$ subrectangles. The simplest method is to chop each side in the same way. This produces the partition $P_2 = (\{0, \frac{1}{2}, 1\}, \{0, \frac{1}{2}, 1\})$ illustrated below on the left.





Similarly, $[0,1]^3$ is a unit cube in $\mathbb{R}^3$ which you want to partition into $2^3 = 8$ subrectangles. Using the same idea as $n = 2$ produces a partition P_3, which is illustrated above on the right.

Notice that you did not need need to partition $[0,1]^2$ or $[0,1]^3$ into identically-sized subrectangles, but this choice makes their algebraic descriptions much easier. The subrectangles of P_2 and the subrectangles of P_3 are respectively given by

$$\{R_{ij}\}_{i,j} = \left\{\left[\frac{i-1}{2}, \frac{i}{2}\right] \times \left[\frac{j-1}{2}, \frac{j}{2}\right] : i, j \in \{1,2\}\right\},$$
$$\{R_{ijk}\}_{i,j,k} = \left\{\left[\frac{i-1}{2}, \frac{i}{2}\right] \times \left[\frac{j-1}{2}, \frac{j}{2}\right] \times \left[\frac{k-1}{2}, \frac{k}{2}\right] : i, j, k \in \{1,2\}\right\}.$$

The pattern suggests how to generalize this partition of $[0,1]^n$ to any fixed dimension $n \in \mathbb{N}$. Namely, a partition P_n of $[0,1]^n$ into 2^n subrectangles can be given by

$$P_n = (\{0, \tfrac{1}{2}, 1\}, \ldots, \{0, \tfrac{1}{2}, 1\})$$

whose subrectangles are

$$\{R_i\}_i = \left\{\left[\frac{i_1 - 1}{2}, \frac{i_1}{2}\right] \times \cdots \times \left[\frac{i_n - 1}{2}, \frac{i_n}{2}\right] : i_1, \ldots, i_n \in \{1,2\}\right\}.$$

There is a certain family of partitions that have a nice explicit description.

Definition 7.1.12 A partition $P = (P_1, \ldots, P_n)$ of the rectangle $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$ is **regular** if, for each $j \in \{1, \ldots, n\}$, the set P_j is a regular partition of the interval $[a_j, b_j]$.

Remark 7.1.13 A regular partition of an interval $[a, b]$ is a partition $\{x_0, \ldots, x_k\}$ where every subinterval $[x_{i-1}, x_i]$ has the same length. Explicitly, $x_i = a + \frac{b-a}{k} i$ for $0 \le i \le k$.

Example 7.1.14 Consider the rectangle $R = [2,4] \times [1,3]$ and two partitions:

$$P = (\{2, 3.5, 4\}, \{1, 2, 3\}),$$
$$P' = (\{2, 2.5, 3, 3.5, 4\}, \{1, 2, 3\}).$$

Partition P' is regular whereas P is not. Notice P' is regular but does not have the same number of subintervals along each side.

Now that you can construct partitions, your next goal is to describe how to refine them.

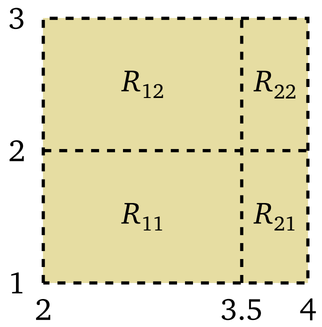
7.1.2 Refining partitions

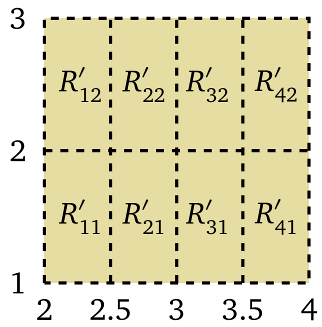
A refinement of partitions in $\mathbb{R}^n$ closely mirrors the traditional description in $\mathbb{R}$.

Definition 7.1.15 Let $P = (P_1, \dots, P_n)$ and $P' = (P'_1, \dots, P'_n)$ be partitions of a rectangle R. The partition P' is a **refinement of** P if $P_j \subseteq P'_j$ for every $j \in \{1, \dots, n\}$.

A simple example in $\mathbb{R}^2$ will demonstrate why this definition agrees with intuition.

Example 7.1.16 Consider the rectangle $R = [2,4] \times [1,3]$ and the two partitions P and P' from Example 7.1.14. These are illustrated below and their subrectangles are labelled.





You can verify by definition that P' is a refinement of P. More importantly, this implies valuable relationships between their subrectangles. For instance, $R'_{31} \subseteq R_{11}$ and $R'_{41} \subseteq R_{21}$. Now, since partitions are so rigid, the definition of a refinement also forces each subrectangle of P to be divided into subrectangles of P'. For instance,

$$R_{12} = R'_{12} \cup R'_{22} \cup R'_{32}.$$

The subrectangles $R'_{12}, R'_{22}, R'_{32}$ overlap but barely since, for example, $\operatorname{int}(R'_{22} \cap R'_{32}) = \varnothing$.

Relationships between subrectangles and refinements will be fundamental to your study.

Theorem 7.1.17 Let R be a rectangle in $\mathbb{R}^n$. Let P and P' be partitions of R with subrectangles $\{R_i\}_{i \in I}$ and $\{R'_j\}_{j \in J}$ respectively. If P' is a refinement of P, then all of the following hold:

(a) For every $j \in J$, there exists a unique $i \in I$ such that $R'_j \subseteq R_i$.

(b) For every $i \in I$, the set $J_i = \{j \in J : R'_j \subseteq R_i\}$ satisfies $R_i = \bigcup_{j \in J_i} R'_j$ and

$$\operatorname{vol}(R_i) = \sum_{j \in J_i} \operatorname{vol}(R'_j).$$

(c) The index set J is the disjoint union[1] of sets J_i for $i \in I$.

Proof. (a) This is left as an exercise. The existence of $i \in I$ is the key item. Once you prove its existence, the index $i \in I$ is necessarily unique by Theorem 7.1.10. Indeed, if $R_{i'}$ is another subrectangle of P containing R'_j, then $R_i \cap R_{i'}$ contains R'_j which has non-empty interior, so it must be that $i = i'$. To prove existence, start by proving the $n = 1$ case and then apply this result to the general case. Notation might get a bit messy, but it all works out.

(b) Fix $i = (i_1, \dots, i_n) \in I$. It suffices to prove that there exists a partition Q of the rectangle

[1] The phrase "disjoint union" means here that J is the union of J_i for $i \in I$, and also $J_i \cap J_{i'} = \varnothing$ whenever $i \neq i'$.

R_i such that the subrectangles of Q is equal to $\{R'_j : j \in J_i\}$. If so, then the desired result follows from Theorem 7.1.10.

Now, write $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$. Write its partition $P = (P_1, \ldots, P_n)$ with $P_t = \{x_0^{(t)}, x_1^{(t)}, \ldots, x_{k_t}^{(t)}\}$ for $t \in \{1, \ldots, n\}$, where

$$a_t = x_0^{(t)} < x_1^{(t)} < \cdots < x_n^{(t)} = b_t. \tag{7.1.7}$$

Then, per Definition 7.1.8, write

$$R_i = R_{(i_1,\ldots,i_n)} := \left[x_{i_1-1}^{(1)}, x_{i_1}^{(1)}\right] \times \left[x_{i_2-1}^{(2)}, x_{i_2}^{(2)}\right] \times \cdots \times \left[x_{i_n-1}^{(n)}, x_{i_n}^{(n)}\right].$$

Similarly, write the refinement $P' = (P'_1, \ldots, P'_n)$ with $P'_t = \{y_0^{(t)}, y_1^{(t)}, \ldots, y_{\ell_t}^{(t)}\}$ for $t \in \{1, \ldots, n\}$. Since $P_t \subseteq P'_t$, it follows by (7.1.7) that there exists $p_t, q_t \in \{1, \ldots, \ell_t\}$ with $p_t < q_t$ such that

$$x_{i_t-1}^{(t)} = y_{p_t}^{(t)} < y_{p_t+1}^{(t)} < \cdots < y_{q_t}^{(t)} = x_{i_t}^{(t)}. \tag{7.1.8}$$

Now, define $Q = (Q_1, \ldots, Q_n)$ by $Q_t = \{y_{p_t}^{(t)}, y_{p_t+1}^{(t)}, \ldots, y_{q_t}^{(t)}\}$ for $t \in \{1, \ldots, n\}$, so by (7.1.8)

$$\left\{x_{i_t-1}^{(t)}, x_{i_t}^{(t)}\right\} \subseteq Q_t \subseteq \left[x_{i_t-1}^{(t)}, x_{i_t}^{(t)}\right].$$

Hence, Q is a partition of R_i. It remains to show the subrectangles of Q are equal to $\{R'_j : j \in J_i\}$.

The subrectangles of P' are of the form

$$R'_j = R'_{(j_1,\ldots,j_n)} = \left[y_{j_1-1}^{(1)}, y_{j_1}^{(1)}\right] \times \left[y_{j_2-1}^{(2)}, y_{j_2}^{(2)}\right] \times \cdots \times \left[y_{j_n-1}^{(n)}, y_{j_n}^{(n)}\right].$$

where $1 \le j_t \le \ell_t$ for $t \in \{1, \ldots, n\}$. The subrectangles of Q are of the same form except with $p_t < j_t \le q_t$ for $t \in \{1, \ldots, n\}$. Notice that

$$j = (j_1, \ldots, j_n) \in J_i \iff R'_j \subseteq R_i \iff \forall t \in \{1, \ldots, n\},, [y_{j_t-1}^{(t)}, y_{j_t}^{(t)}] \subseteq [x_{i_t-1}, x_{i_t}]$$

From (7.1.8), this occurs if and only if $p_t < j_t \le q_t$ for all $t \in \{1, \ldots, n\}$. These correspond precisely to the subrectangles of Q, as required.

(c) This is left as an exercise. Combine (a) and (b). ■

As you might expect, the refinement of a refinement is a refinement.

Lemma 7.1.18 Let P, P', P'' be partitions of a rectangle R in $\mathbb{R}^n$. If P'' is a refinement of P' and P' is a refinement of P then P'' is a refinement of P.

Proof. This is left as a straightforward exercise. ■

Given two different partitions of a rectangle, you will want to construct a partition that simultaneously refines both of them. There is a natural way to do so.

Definition 7.1.19 Let $P' = (P'_1, \ldots, P'_n)$ and $P'' = (P''_1, \ldots, P''_n)$ be partitions of a rectangle R. The **common refinement of** P' **and** P'' is the partition $P = (P_1, \ldots, P_n)$ given by $P_j = P'_j \cup P''_j$ for every $j \in \{1, \ldots, n\}$.

This definition is really just describing a natural way to combine two partitions.

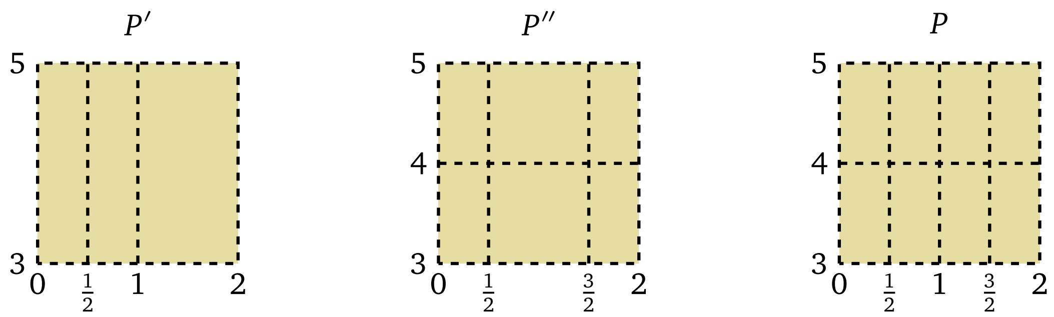
Example 7.1.20 Consider the rectangle $R = [0,2] \times [3,5]$ and two partitions

$$P' = (\{0, \tfrac{1}{2}, 1, 2\}, \{3,5\}), \qquad P'' = (\{0, \tfrac{1}{2}, \tfrac{3}{2}, 2\}, \{3,4,5\})$$

Let P be the common refinement between P' and P'', so

$$P = (\{0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2\}, \{3,4,5\}).$$

You can visualize this process below.



From your experience with $\mathbb{R}$, you may be tempted to say that P is the union of the two other partitions, but this is not correct. It is more correct to say that P is the *component-by-component* union of P' and P''.

The above example illustrates how the common refinement refines both partitions. Moreover, the common refinement is the "smallest possible" refinement of two partitions.

Lemma 7.1.21 Let P' and P'' be partitions of the rectangle R in $\mathbb{R}^n$. Let P be the common refinement of P' and P''.
(a) P is a refinement of both P' and P''.
(b) If Q is a refinement of both P' and P'', then Q is a refinement of P.

Proof. (a) As $P = (P_1, \dots, P_n)$ is the common refinement of $P' = (P'_1, \dots, P'_n)$ and $P'' = (P''_1, \dots, P''_n)$, it follows for every $j \in \{1, \dots, n\}$ that $P'_j \subseteq P'_j \cup P''_j = P_j$ and similarly $P''_j \subseteq P_j$. Thus, by Definition 7.1.15, P is a refinement of P' and a refinement of P''.

(b) Since $Q = (Q_1, \dots, Q_n)$ is a refinement of P' and P'', Definition 7.1.15 implies for every $j \in \{1, \dots, n\}$ that $P'_j \subseteq Q_j$ and $P''_j \subseteq Q_j$, in which case $P_j = P'_j \cup P''_j \subseteq Q_j$ by definition of the common refinement. Thus, by Definition 7.1.15, Q is a refinement of P. ■

Finally, how can you measure the level of refinement of a partition? There are many ways to do so. A common choice is its norm.

Definition 7.1.22 Let P be a partition of a rectangle R in $\mathbb{R}^n$. The **norm of** P, denoted $||P||$, is the maximum diameter of all of its subrectangles.

Remark 7.1.23 The **diameter** of a rectangle is the maximum distance between any two points in the rectangle.

Example 7.1.24 Consider the rectangle $R = [2,4] \times [1,3]$ and partition P in Example 7.1.16 with subrectangles $\{R_{11}, R_{12}, R_{21}, R_{22}\}$. Notice that the diameters of R_{11} and R_{21} are both equal to

$$\sqrt{(1.5)^2 + 1^2} = \sqrt{3.25}$$

and the diameters of R_{12} and R_{22} are both equal to

$$\sqrt{(0.5)^2 + 1^2} = \sqrt{1.25}.$$

The maximum diameter of all the subrectangles is therefore $\sqrt{3.25}$, so $||P|| = \sqrt{3.25}$.

Crucially, you can produce partitions with arbitrarily small norm, and the norm reduces or remains the same after refinement.

Lemma 7.1.25 Let R be a rectangle in $\mathbb{R}^n$.

(a) For every $\delta > 0$, there exists a partition P of R with norm $||P|| < \delta$.

(b) Let P and P' be partitions of R. If P' is a refinement of P then $||P'|| \leq ||P||$.

Proof. This is left as an exercise. For (a), use a regular partition with N^n subrectangles and choose N large enough. For (b), use Theorem 7.1.17(a). ■

This concludes step one of integration over a rectangle.

Chop the rectangle into little pieces.

There were many new features in higher dimensions and things got a bit complicated, but you can always rely on your intuition and pictures in $\mathbb{R}^2$. Next, you must create estimates for a function on each of your chopped pieces and total your estimates.

Exercises for Section 7.1

Concepts and definitions

7.1.1 Which of the following are partitions of the rectangle $R = [0,1] \times [2,5]$?

(a) $\big(\{0, 0.4, 1\}, \{2, 3, 5\}\big)$

(b) $\big(\{2, 3, 5\}, \{0, 0.4, 1\}\big)$

(c) $\big\{[0,0.4] \times [2,3], [0.4,1] \times [3,5], [0,0.4] \times [3,5], [0.4,1] \times [2,3]\big\}$

(d) $\big(\{0, \sqrt{2}, 1\}, \{2, \pi, 5\}\big)$

(e) $\big(\{0, 1\}, \{2, 5\}\big)$

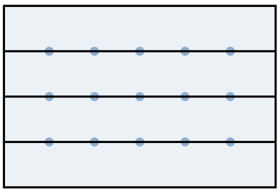
7.1.2 Consider the partition $P = \big(\{0, 0.4, 1\}, \{2, 3, 5\}\big)$ of the rectangle $R = [0,1] \times [2,5]$.

(a) What is the index set of P?

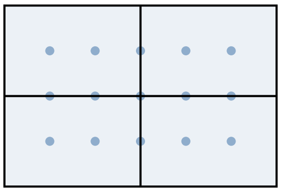
(b) What are the subrectangles of P?

7.1.3 Let $R = [1,7] \times [-2,2]$. Which of these are partitions of R? And which are regular partitions?

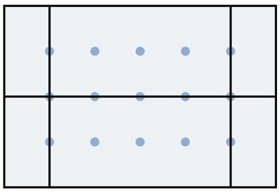
(a)



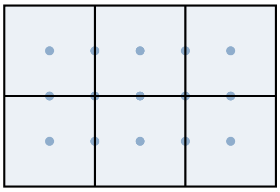
(b)



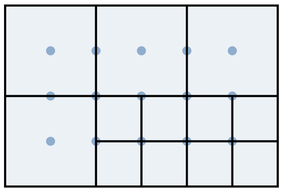
(c)



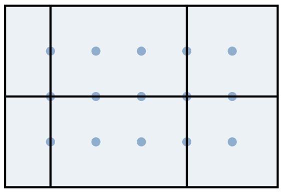
(d)



(e)



(f)



7.1.4 Let R be a rectangle in $\mathbb{R}^n$. Let P be a partition of the rectangle R. Let $\{R_i\}_{i \in I}$ be the subrectangles of P. Which statements are true or false? If true, briefly justify. If false, state a counterexample.

(a) The rectangle R is the union of R_i over $i \in I$.

(b) The number of subrectangles of P is equal to $|I|$, the size of the index set.

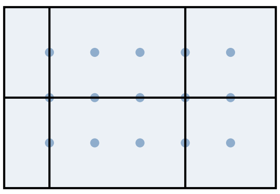
(c) The volume of each subrectangle of P is equal to $\dfrac{\operatorname{vol}(R)}{|I|}$.

(d) If R_i and R_j are distinct subrectangles of P, then $R_i \cap R_j = \varnothing$.

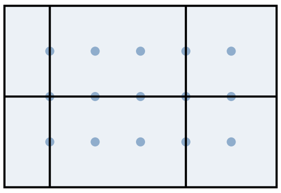
(e) If R_i and R_j are distinct subrectangles of P, then $\operatorname{int}(R_i) \cap \operatorname{int}(R_j) = \varnothing$.

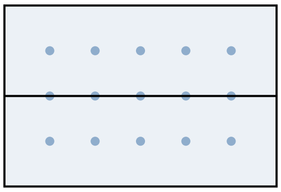
(f) If $x \in R$ then there exists a unique $i \in I$ such that $x \in R_i$.

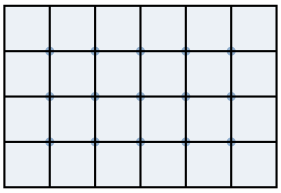
7.1.5 Let P be this partition of the rectangle $R = [1,7] \times [0,4]$.

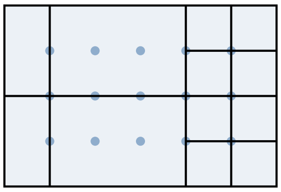


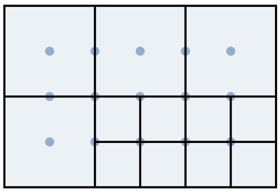
Which of these are a refinement of P?

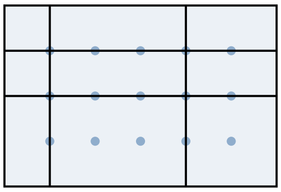
(a) 

(b) 

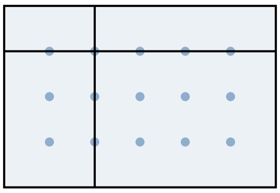
(c) 

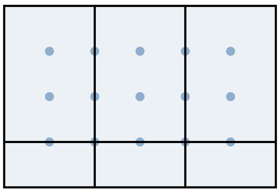
(d) 

(e) 

(f) 

7.1.6 Let P' and P'' be these two partitions. Sketch their **common refinement** P.





7.1.7 Let P be the partition of the rectangle $R = [1,7] \times [0,4]$ given by

$$P = \big(\{1,2,5,7\}, \{0,2,4\}\big)$$

Which of these partitions are a refinement of P? Which of these are regular partitions of R?

(a) $P_1 = \big(\{1,2,5,7\}, \{0,2,4\}\big)$

(b) $P_2 = \big(\{1,7\}, \{0,2,4\}\big)$

(c) $P_3 = \big(\{1,2,3,4,5,6,7\}, \{0,1,2,3,4\}\big)$

(d) $P_4 = \big(\{1,2,3,4,5,6,7\}, \{0,4\}\big)$

(e) $P_5 = \big(\{1,2,3,4,5,6,7\}, \{0,2,4\}\big)$

(f) $P_6 = \big(\{1,2,5,7\}, \{0,2,3,4\}\big)$

7.1.8 Let P' and P'' be two partitions of $R = [1,7] \times [0,4]$ given by

$$P' = \big(\{1,3,7\}, \{0,3,4\}\big) \qquad P'' = \big(\{1,3,5,7\}, \{0,1,4\}\big)$$

Define their common refinement P. Is $P = P' \cup P''$?

7.1.9 Let R be a rectangle in $\mathbb{R}^n$. Assume:

- $P = (P_1, \dots, P_n)$ is a partition of R with index set I and subrectangles $\{R_i\}_{i \in I}$
- $P' = (P'_1, \dots, P'_n)$ is a partition of R with index set J and subrectangles $\{R'_j\}_{j \in J}$
- P' is a refinement of P

Which of the following are true or false? If true, briefly justify. If false, state a counterexample.

(a) $P_j \subseteq P'_j$ for every $j \in \{1, \dots, n\}$.

(b) $P \subseteq P'$

(c) $\sum_{i\in I} \text{vol}(R_i) = \sum_{j\in J} \text{vol}(R'_j)$

(d) $|I| \leq |J|$

(e) For each $j \in J$, there exists $i \in I$ such that $R'_j \subseteq R_i$.

(f) For each $i \in I$, there exists $j \in J$ such that $R_i = R'_j$.

(g) If Q is the common refinement of P and P', then $Q = P'$.

7.1.10 Fill in the blanks below.

(a) A regular partition $\{x_0, \dots, x_M\}$ of $[a, b]$ into M subintervals is given by $x_i =$ ________________ for $0 \leq i \leq M$. The width of each subinterval is $\Delta x =$ ________________ .

(b) A regular partition $\{y_0, \dots, y_N\}$ of $[c, d]$ into N subintervals is given by $y_j =$ ________________ for $0 \leq j \leq N$. The width of each subinterval is $\Delta y =$ ________________ .

(c) A regular partition of $R = [a, b] \times [c, d]$ created from these regular partitions of $[a, b]$ and $[c, d]$ has exactly __________ subrectangles $\{R_{ij}\}_{i,j}$ each with area equal to ________________ .

7.1.11 Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. Fix $N \in \mathbb{N}^+$. For each $j \in \{1, \dots, n\}$, let P_j be the regular partition of the interval $[a_j, b_j]$ into N subintervals. Define $P = (P_1, P_2, \dots, P_n)$.

(a) For $j \in \{1, \dots, n\}$, explicitly write the elements of P_j. What is the width of a subinterval of P_j?

(b) Explicitly write the index set I of P using set builder notation. How many indices does it have?

(c) Explicitly write the subrectangles $\{R_i\}_{i\in I}$ of P in terms of your notation. How many subrectangles are there and what is the area of each subrectangle?

Computations

7.1.12 Prove (7.1.3) using sigma notation.

7.1.13 Let $R = [a, b] \times [c, d]$. Let $P = (P_1, P_2)$ be a partition of R so $P_1 = \{x_0, \dots, x_M\}$ and $P_2 = \{y_0, \dots, y_N\}$ are partitions of $[a, b]$ and $[c, d]$ respectively. Denote $\Delta x_i = x_i - x_{i-1}$ for $1 \leq i \leq M$ and $\Delta y_j = y_j - y_{j-1}$ for $1 \leq j \leq N$. Explain the identity

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \Delta x_i \Delta y_j = \text{area}(R)$$

using geometry and then prove it by direct computation with sigma notation.

7.1.14 Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. Let P be a partition of R with index set I and subrectangles $\{R_i\}_{i\in I}$. Show that

$$\sum_{i\in I} \text{vol}(R_i) = \text{vol}(R)$$

by direct computation with sigma notation.

7.1.15 Partition the unit cube $[0,1]^n$ using a regular partition with N^n subrectangles.

7.1.16 Using the norm of partitions, you will construct partitions in $\mathbb{R}^n$ with subrectangles that get arbitrarily small. However, "small" means something more than just "small volume".

(a) What is the diameter of the rectangle $R = [a_1, b_1] \times [a_2, b_2] \times \cdots [a_n, b_n]$?

(b) Prove or disprove: There exists a rectangle in $\mathbb{R}^2$ with volume ≥ 100 and diameter $\leq \frac{1}{100}$.

(c) Prove or disprove: There exists a rectangle in $\mathbb{R}^2$ with diameter ≥ 100 and volume $\leq \frac{1}{100}$.

(d) Fix $N \in \mathbb{N}^+$. Let P_N be a regular partition of R constructed from N subintervals for each interval $[a_1, b_1], \ldots, [a_n, b_n]$. What is the diameter of each subrectangle belonging to P_N?

(e) Prove that for every $\delta > 0$ there exists a partition P of the rectangle R with norm $||P|| < \delta$.

Proofs

7.1.17 You will verify the fundamental properties of partitions in $\mathbb{R}$ and $\mathbb{R}^2$.

(a) Prove (7.1.1) using the definition of a partition.

(b) Prove (7.1.2) using standard topological facts and the definition of a partition.

(c) Prove (7.1.4) using (a) and standard topological facts.

(d) Prove (7.1.5) using (b) and standard topological facts.

7.1.18 You will analyze a minor variant of the textbook's proof of the following theorem.

> **Theorem.** *Let R be a rectangle in $\mathbb{R}^n$. Let P be a partition of R. Let I be the index set of P and let $\{R_i : i \in I\}$ be the subrectangles of P. For $i, j \in I$, if $i \neq j$, then* $\operatorname{int}(R_i \cap R_j) = \varnothing$.

1. *Fix $i, j \in I$ and assume $i = (i_1, \ldots, i_n) \neq (j_1, \ldots, j_n) = j$.*
2. *Without loss of generality, assume $i_1 \neq j_1$.*
3. *By definition of a partition over $\mathbb{R}$, the subset of real numbers $A := [x^{(1)}_{i_1-1}, x^{(1)}_{i_1}] \cap [x^{(1)}_{j_1-1}, x^{(1)}_{j_1}]$ is either $\varnothing$ or equal to $\{x^{(1)}_{i_1}\}$ or equal to $\{x^{(1)}_{j_1}\}$.*
4. *Since $R_i \cap R_j \subseteq A \times \mathbb{R}^{n-1}$, it suffices to prove that $\{a\} \times \mathbb{R}^{n-1}$ has empty interior for any fixed $a \in \mathbb{R}$.*
5. *You can verify that $\{a\} \times \mathbb{R}^{n-1}$ has empty interior by standard arguments in topology.*

(a) Line 2 uses "without loss of generality" to avoid writing an equivalent step. State that step.

(b) Line 3 concludes A is equal to 3 possible cases. Each case corresponds to equalities (or inequalities) involving the indices i_1 and j_1. State those corresponding equalities (or inequalities).

(c) Line 4 could use a bit more detail. Justify this line with facts from topology and line 3.

(d) Justify Line 5 by proving $\{a\} \times \mathbb{R}^{n-1}$ has empty interior. *Hint:* Find a topological lemma.

7.1.19 Let P, P', P'' be partitions of a rectangle R in $\mathbb{R}^n$. Prove that if P'' is a refinement of P' and P' is a refinement of P, then P'' is a refinement of P.

7.1.20 Prove or disprove: The common refinement of two regular partitions is a regular partition.

7.1.21 Let P and P' be partitions of the rectangle R in $\mathbb{R}^n$. Show that if P' is a refinement of P, then $||P'|| \leq ||P||$.

7.2. Upper sums and lower sums

You will continue your preparation to integrate over a rectangle in $\mathbb{R}^n$. In this section, you will unfold the next step in the process.

> *Chop the rectangle into little pieces. Estimate the value of the function on each piece. Total your estimates to approximate the total value over the rectangle.*

You have learned to chop and now you will learn to estimate. You will want to create an over-estimate (upper sums) and an under-estimate (lower sums) almost exactly as they appear for the single variable integral[2]. You shall use these similarities with the one-dimensional case to your advantage when generalizing to higher dimensions. The new challenges are almost entirely brought on by the increased complexity of partitions in $\mathbb{R}^n$, but it is mostly notational.

7.2.1 Definition of upper and lower sums

For the sake of notational brevity, you shall begin with definitions specialized to $\mathbb{R}^2$.

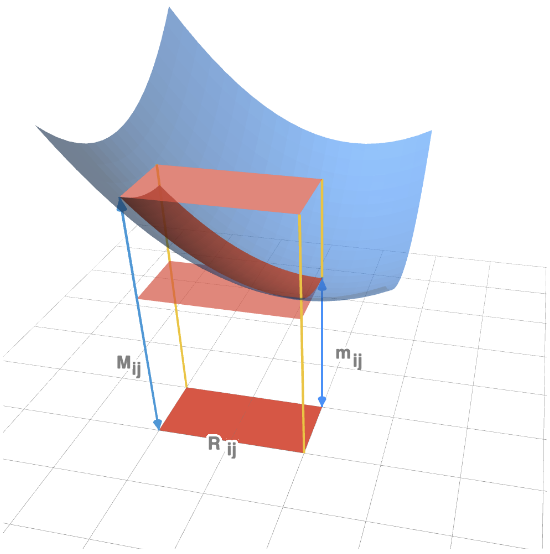
Definition 7.2.1 Let R be a rectangle in $\mathbb{R}^2$. Let $f : R \to \mathbb{R}$ be a bounded function. Let P be a partition of R and let $\{R_{ij} : 1 \le i \le k, 1 \le j \le \ell\}$ be the subrectangles of P. The P-**lower sum** of f and P-**upper sum** of f are respectively defined by

$$L_P(f) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} m_{ij}\,\text{area}(R_{ij}) \qquad \text{and} \qquad U_P(f) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} M_{ij}\,\text{area}(R_{ij}),$$

where

$$\forall 1 \le i \le k, 1 \le j \le \ell, \qquad m_{ij} = \inf_{x \in R_{ij}} f(x) \qquad \text{and} \qquad M_{ij} = \sup_{x \in R_{ij}} f(x).$$

Notice the upper and lower sums are always defined for a bounded function. This graph below which you can also view on Math3D demonstrates the visual intuition behind these definitions.



The upper sum is an over-estimate for the total value, and the lower sum is an under-estimate for the total value. You can explicitly calculate these in some basic examples.

[2]If you want to review some single variable calculus, watch this MAT137 video on the definition of the single-variable integral [9]. It gives the basic intuition and the technical details.

Example 7.2.2 Define the function $f:\mathbb{R}^2\to\mathbb{R}$ as $f(x,y)=2x+y$ for all $(x,y)\in\mathbb{R}^2$ and the rectangle $R=[0,1]\times[2,4]$. To estimate the total value of f on R, you can use the partition P of the rectangle R given by $P=\left(\{0,\frac{1}{2},1\},\{2,3,4\}\right)$. This creates four subrectangles, namely

$$R_{11}=[0,\tfrac{1}{2}]\times[2,3],\quad R_{12}=[0,\tfrac{1}{2}]\times[3,4],\quad R_{21}=[\tfrac{1}{2},1]\times[2,3],\quad R_{22}=[\tfrac{1}{2},1]\times[3,4].$$

Notice $\operatorname{area}(R_{ij})=1/2$ for all $i,j\in\{1,2\}$. To calculate $U_p(f)$ and $L_p(f)$, notice that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are always positive, so f is increasing as x increases with y fixed, or as y increases with x fixed. Hence, for any $i,j\in\{1,2\}$, the quantity m_{ij} is f evaluated at the bottom-left corner of R_{ij}, and the quantity M_{ij} is f evaluated at the top-right corner of R_{ij}. Therefore,

$$U_P(f)=\sum_{i=1}^{2}\sum_{j=1}^{2}M_{ij}\operatorname{area}(R_{ij})=\tfrac{1}{2}\left(f\left(\tfrac{1}{2},3\right)+f\left(\tfrac{1}{2},4\right)+f(1,3)+f(1,4)\right)=10,$$

$$L_p(f)=\sum_{i=1}^{2}\sum_{j=1}^{2}m_{ij}\operatorname{area}(R_{ij})=\tfrac{1}{2}\left(f(0,2)+f(0,3)+f\left(\tfrac{1}{2},2\right)+f\left(\tfrac{1}{2},3\right)\right)=6.$$

It is usually not easy to calculate the infimum and supremum of a function on a set, but the function in the previous example is easy to understand. This allows you to explicitly calculate its upper and lower sums with a more complicated partition.

Example 7.2.3 Let $f(x,y)=2x+y$. Fix $N\in\mathbb{N}^+$. Let $\{x_0,x_1,\dots,x_N\}$ and $\{y_0,y_1,\dots,y_N\}$ be regular partitions of $[0,1]$ and $[2,4]$, respectively. That is, $x_i=\frac{i}{N}$ and $y_j=2+\frac{2j}{N}$ for all $1\le i,j\le N$. Define the regular partition P_N of the rectangle $R=[0,1]\times[2,4]$ by

$$P_N=(\{x_0,x_1,\dots,x_N\},\{y_0,y_1,\dots,y_N\}),$$

so its subrectangles are given by $R_{ij}=[x_{i-1},x_i]\times[y_{j-1},y_j]$ for $1\le i,j\le N$.

For $1\le i,j\le N$, notice that $\operatorname{area}(R_{ij})=(x_i-x_{i-1})(y_j-y_{j-1})=\frac{1}{N}\cdot\frac{2}{N}=\frac{2}{N^2}$ and, since $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are always positive,

$$m_{ij}=f(x_{i-1},y_{j-1})\quad\text{and}\qquad M_{ij}=f(x_i,y_j).$$

It follows by definition that

$$L_{P_N}(f)=\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}f(x_{i-1},y_{j-1})\qquad\text{and}\qquad U_{P_N}(f)=\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}f(x_i,y_j).$$

Using standard formulas for sums, notice

$$\begin{aligned}U_{P_N}(f)=\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(2x_i+y_j)&=\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{2i}{N}+2+\frac{2j}{N}\right)\\&=\frac{4}{N^3}\cdot N\cdot\sum_{i=1}^{N}i+\frac{4}{N^2}\cdot N^2+\frac{4}{N^3}\cdot N\sum_{j=1}^{N}j\\&=\frac{2(N+1)}{N}+4+\frac{2(N+1)}{N}=8+\frac{4}{N}\end{aligned}$$

and similarly $L_{P_N}(f)=8-\frac{4}{N}$.

The definition for upper and lower sums in $\mathbb{R}^n$ is a straightforward generalization.

Definition 7.2.4 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Let P be a partition of R with index set I and subrectangles $\{R_i : i \in I\}$. The **P-lower sum** of f and **P-upper sum** of f are respectively

$$L_P(f) = \sum_{i \in I} m_i \operatorname{vol}(R_i) \qquad \text{and} \qquad U_P(f) = \sum_{i \in I} M_i \operatorname{vol}(R_i),$$

where

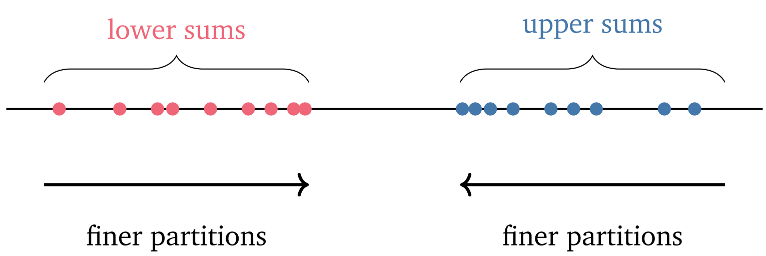
$$\forall i \in I, \qquad m_i = \inf_{x \in R_i} f(x) \qquad \text{and} \qquad M_i = \sup_{x \in R_i} f(x).$$

Remark 7.2.5 The upper sum and lower sum have only one sum $\sum$ since the sum runs over i from the index set I. An element of I is a multi-index $i = (i_1, \dots, i_n)$ so this hides the n iterated sums $\sum \cdots \sum$ which would appear if you wrote out each index $i_1, \dots, i_n$ separately.

Explicit calculations with upper sums and lower sums in $\mathbb{R}^n$ for $n \geq 3$ is really no different than the previous examples. It is mostly a matter of more notation and more computations.

7.2.2 Properties of upper and lower sums

Thus far, you have defined partitions, upper sums, and lower sums. How do upper sums and lower sums behave with respect to partitions and refinements? This will be critical to analyzing your definition of the integral. The following figure[3] neatly summarizes their relationship.



You will start by formally stating and verifying these properties in the same vein as single variable calculus. For a given partition, its lower sum is always at most its upper sum.

Lemma 7.2.6 Let P be a partition of a rectangle $R \subseteq \mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Then $L_P(f) \leq U_P(f)$.

Proof. This is left as a short exercise. Use the definitions of supremum and infimum. ∎

Next, you will see why a refinement of a partition always improves your estimate.

Lemma 7.2.7 Let P and P' be partitions of a given rectangle $R \subseteq \mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. If P' is a refinement of P, then $L_P(f) \leq L_{P'}(f)$ and $U_{P'}(f) \leq U_P(f)$.

Proof. It suffices to check the lower sum inequality since the upper sum inequality is similar. Let $\{R_i\}_{i \in I}$ and $\{R'_j\}_{j \in J}$ be the subrectangles of P and P' respectively. For $i \in I$, let J_i be the subset of indices $j \in J$ such that $R'_j \subseteq R_i$. By Theorem 7.1.17, since P' is a refinement of P, each $j \in J$ belongs to a subset J_i for a *unique* index $i \in I$ and, moreover, J is the *disjoint* union of J_i over $i \in I$.

[3] Modified from and originally created by Alfonso Gracia-Saz [9].

Fix a given subrectangle R_i of P for some $i \in I$. By Theorem 7.1.17(b),

$$\operatorname{vol}(R_i) = \sum_{j \in J_i} \operatorname{vol}(R_j).$$

Also, for $j \in J_i$, we have that $R'_j \subseteq R_i$ so $m_i = \inf_{x \in R_i} f(x) \leq \inf_{x \in R'_j} f(x) = m'_j$. These observations imply that

$$m_i \operatorname{vol}(R_i) = \sum_{j \in J_i} m_i \operatorname{vol}(R'_j) \leq \sum_{j \in J_i} m'_j \operatorname{vol}(R'_j).$$

Summing this over $i \in I$, it follows that

$$L_P(f) = \sum_{i \in I} m_i \operatorname{vol}(R_i) \leq \sum_{i \in I} \sum_{j \in J_i} m'_j \operatorname{vol}(R'_j) = \sum_{j \in J} m'_j \operatorname{vol}(R'_j) = L_{P'}(f).$$

The last equality holds since J is the disjoint union of J_i over $i \in I$. ■

Finally, you can explain why *all* lower sums are smaller than *all* upper sums.

Lemma 7.2.8 If P' and P'' are two partitions of a rectangle $R \subseteq \mathbb{R}^n$ then $L_{P'}(f) \leq U_{P''}(f)$.

Proof. Let P be the common refinement of P' and P'', so P is a refinement of P' and P'' by Lemma 7.1.21. By Lemmas 7.2.6 and 7.2.7, it follows that

$$L_{P'}(f) \leq L_P(f) \leq U_P(f) \leq U_{P''}(f),$$

as required. ■

The above three lemmas will form the foundation for the existence of integrals in $\mathbb{R}^n$, since they inform you how these estimates change with refinements of partitions. Now, there are other properties about upper sums and lower sums that will be vital for establishing properties of the integral itself. The upper sum properties are summarized in the lemma below.

Lemma 7.2.9 Let P be a partition of a rectangle R in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ and $g : R \to \mathbb{R}$ be bounded functions. All of the following hold:

(a) $U_P(f+g) \leq U_P(f) + U_P(g)$.
(b) $U_P(\lambda f) = \lambda U_P(f)$ for any $\lambda > 0$.
(c) $U_P(-f) = -L_P(f)$.
(d) If $f \leq g$ on R, then $U_P(f) \leq U_P(g)$.

Proof. This is left as an exercise. Each of them follow from properties of the supremum. You are encouraged to also state the corresponding lemma for lower sums to Lemma 7.2.9. ■

You have collected all of the key tools about estimates for total value to move forward with the theory of the integral. Upper sums and lower sums will be the linchpin of its definition and the theory because of the technical lemmas above. Now, explicitly calculating upper sums and lower sums requires you to compute the suprema and infima on each subrectangle. That is equivalent to trying to solve an optimization problem *on each* subrectangle! This is computationally infeasible. Thus, while upper and lower sums provide many theoretical advantages via the above lemmas, other contexts may demand methods for estimating total value that possess computational advantages.

7.2.3 Definition of Riemann sums

Another method for estimating total value is by Riemann sums. With all of your new language for partitions, these can be effortlessly generalized to higher dimensions.

Definition 7.2.10 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Let P be a partition of R with index set I and subrectangles $\{R_i\}_{i\in I}$. For each $i \in I$, choose a **sample point** $x_i^* \in R_i$. The quantity

$$S_P^*(f) = \sum_{i\in I} f(x_i^*)\,\mathrm{vol}(R_i)$$

is a **Riemann sum** for f with P and these sample points.

Remark 7.2.11 Partitions and the associated sample points are also called **tagged partitions**. Each sample point is "tagged" to the corresponding subrectangle.

Riemann sums are also an estimate for the total value, but the crucial new feature is the choice of sample points. They are selected in advance, and do not require the computation of any suprema or infima. This makes Riemann sums comparatively much easier to calculate.

Example 7.2.12 Let $f(x, y) = 2x + y$ and define the rectangle $R = [0, 3] \times [0, 6]$. Fix the partition

$$P = (\{0, 2, 3\}, \{0, 4, 6\})$$

which has four subrectangles

$$R_{11} = [0, 2] \times [0, 4], \quad R_{12} = [0, 2] \times [4, 6], \quad R_{21} = [2, 3] \times [0, 4], \quad R_{22} = [2, 3] \times [4, 6].$$

To compute a Riemann sum for f with P, you must choose some sample points. A sample point can be any point in each subrectangle of P, so you can choose, for instance,

$$x_{11}^* = (1, 2), \quad x_{12}^* = (1, 5), \quad x_{21}^* = (2, 2), \quad x_{22}^* = (3, 5).$$

By definition, the Riemann sum for f with P and these sample points is given by

$$S_P^*(f) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(x_{ij}^*)\,\mathrm{area}(R_{ij}) = 8f(1, 2) + 4f(1, 5) + 4f(2, 2) + 2f(3, 5) = 106.$$

The sample points in the above example were haphazardly chosen. You can choose them more systematically.

Example 7.2.13 Let $f(x, y) = 2x + y$. Fix $N \in \mathbb{N}^+$. Let $\{x_0, x_1, \ldots, x_N\}$ and $\{y_0, y_1, \ldots, y_N\}$ be regular partitions of $[0, 1]$ and $[2, 4]$, respectively. That is, $x_i = \frac{i}{N}$ and $y_j = 2 + \frac{2j}{N}$ for all $1 \le i, j \le N$. Define the regular partition P_N of the rectangle $R = [0, 1] \times [2, 4]$ by

$$P_N = (\{x_0, x_1, \ldots, x_N\}, \{y_0, y_1, \ldots, y_N\}).$$

Its subrectangles are $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ for $1 \le i, j \le N$ in which case $\mathrm{area}(R_{ij}) = \frac{2}{N^2}$.

Again, you must choose a sample point for each subrectangle R_{ij}. You could choose them randomly for each subrectangle, but it may be preferable to choose the sample points in an easy-to-describe way. For instance, you could choose the “center” of each rectangle. Namely, for $1 \le i, j \le N$, take the sample point (x_i^*, y_j^*) in R_{ij} to be given by

$$x_i^* = \frac{x_{i-1} + x_i}{2} = \frac{2i-1}{2N}, \qquad y_j^* = \frac{y_{j-1} + y_j}{2} = 2 + \frac{2j-1}{N}.$$

The Riemann sum for f with partition P_N and these sample points is therefore equal to

$$S_{P_N}^*(f) = \sum_{i=1}^{N} \sum_{j=1}^{N} f(x_i^*, y_j^*) \operatorname{area}(R_{ij}) = \frac{2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{4i-2}{2N} + 2 + \frac{2j-1}{N} \right).$$

Similar to Example 7.2.3, you can verify that $S_{P_N}^*(f) = 8$ for all $N \in \mathbb{N}^+$. It is a coincidence that this does not depend on N.

These examples demonstrate that Riemann sums are straightforward and relatively quick to calculate, especially for a computer. They also more easily produce a version of Lemma 7.2.9.

Lemma 7.2.14 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ and $g : R \to \mathbb{R}$ be bounded functions. Let P be a partition of R with index set I and subrectangles $\{R_i\}_{i \in I}$. For each $i \in I$, choose a sample point $x_i^* \in R_i$. Both of the following hold:

(a) $S_P^*(f + \lambda g) = S_P^*(f) + \lambda S_P^*(g)$ for any $\lambda \in \mathbb{R}$.
(b) If $f \le g$ on R, then $S_P^*(f) \le S_P^*(g)$.

Proof. This is left as a very short exercise. ■

These achievements mark a computational advantage over upper and lower sums! However, these gains comes at a serious price. While Riemann sums estimate the total value, you usually have no idea whether it is an over-estimate or an under-estimate. This depends on the choice of sample points and there is no easy way to check. This issue occurs for both Examples 7.2.12 and 7.2.13. Unfortunately, this problem makes it much more annoying to use Riemann sums for developing theory. It is hard, for instance, to establish a reasonable analogue to Lemmas 7.2.6 to 7.2.8. Does the Riemann sum estimate improve as you refine the partition? You cannot guarantee it will. Thus, it would be unnecessarily complicated to use Riemann sums to prove the existence of integrals, so you will temporarily[4] put them aside.

To develop the theory of the integral, you will rely solely on upper and lower sums for estimation. This concludes step two of integration over a rectangle.

> *Chop the rectangle into little pieces. Estimate the value of the function on each piece. Total your estimates to approximate the total value over the rectangle.*

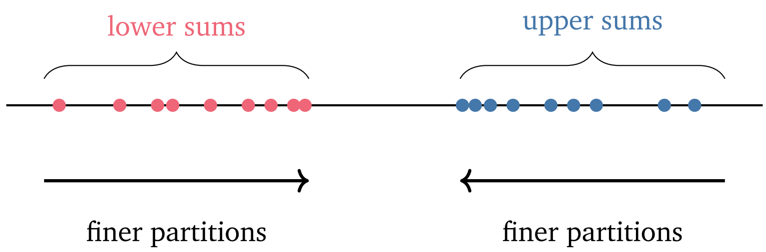
You are ready for the third and final step: arbitrarily refining your estimates to the exact value.

[4]After you have established the foundational theory of the integral, Riemann sums will return to help you naturally derive integral formulas in physics, geometry, statistics, and other areas.

Exercises for Section 7.2

Concepts and definitions

7.2.1 Consider the diagram below.



Three key lemmas about upper sums and lower sums explain the truths behind it, namely Lemmas 7.2.6 to 7.2.8.

(a) Which lemma(s) explain why all the lower sums are smaller than all of the upper sums?

(b) Which lemma(s) explain why finer partitions improve your approximation?

7.2.2 Let $R = [a,b] \times [c,d]$. Let $P = (P_1, P_2)$ be a partition of R, so $P_1 = \{x_0, \dots, x_M\}$ and $P_2 = \{y_0, \dots, y_N\}$ are partitions of $[a,b]$ and $[c,d]$ respectively. The subrectangles $\{R_{ij}\}_{i,j}$ of P are given by

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j] \qquad \forall 1 \le i \le M, 1 \le j \le N.$$

Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and satisfies $\dfrac{\partial f}{\partial x} \le 0$ and $\dfrac{\partial f}{\partial y} \ge 0$ everywhere on R.

Consider the six double sums below.

$$\text{(A)} \sum_{i=1}^{M}\sum_{j=1}^{N} f(x_i, y_j)\Delta x_i \Delta y_j \quad \text{(B)} \sum_{i=1}^{M}\sum_{j=1}^{N} f(x_{i-1}, y_j)\Delta x_i \Delta y_j \quad \text{(C)} \sum_{i=1}^{M}\sum_{j=1}^{N} f(x_i, y_{j-1})\Delta x_i \Delta y_j$$

$$\text{(D)} \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(x_i, y_{j+1})\Delta x_i \Delta y_j \quad \text{(E)} \sum_{i=0}^{M}\sum_{j=1}^{N} f(x_i, y_j)\Delta x_{i+1} \Delta y_j \quad \text{(F)} \sum_{i=0}^{M-1}\sum_{j=1}^{N} f(x_i, y_j)\Delta x_{i+1} \Delta y_j$$

(a) Which of these six sums correspond to the upper sum $U_P(f)$? Identify all that apply.

(b) Choose sample points to be the *upper right* corners of each subrectangle. Which of these six sums correspond to the Riemann sum $S_P^*(f)$? Identify all that apply.

7.2.3 State the lemma for lower sums corresponding to Lemma 7.2.9.

Computations

7.2.4 Consider $R = [0,1] \times [2,5]$. Let $f(x,y) = -x + 3y$.

(a) Sketch the subrectangles of the partition $P = (\{0, 0.4, 1\}, \{2, 3, 5\})$ of R.

(b) Compute $U_P(f)$ and $L_P(f)$.

7.2.5 Let $R=[0,1]\times[2,5]$ and let $f(x,y)=-x+3y$. Fix $N\in\mathbb{N}^+$. Let P_N be a regular partition of R with N subintervals of $[0,1]$ and N subintervals of $[2,5]$.

(a) Illustrate this partition P_N with a sketch. Include a typical subrectangle R_{ij} and label your diagram with standard notation.

(b) Express the upper sum $U_{P_N}(f)$ and lower sum $L_{P_N}(f)$ each as a double sum.

(c) Give an explicit formula for $U_{P_N}(f)$ and $L_{P_N}(f)$ without any summations.

7.2.6 Let $R=[0,2]\times[3,7]$. Consider the discontinuous function $f(x,y)=\begin{cases}-1 & \text{if } x=0,\\ 2 & \text{otherwise.}\end{cases}$

(a) Explicitly compute all upper sums of f.

(b) For $\varepsilon>0$, construct a partition P_ε of R such that $L_{P_\varepsilon}(f)\geq 16-4\varepsilon$. Sketch the subrectangles of your partition and label it according to the notation in your argument.

7.2.7 Let P be the partition of the region $R=[1,3]\times[2,5]$ into six 1×1 squares. Define $f(x,y)=xy^2$.

(a) Calculate the Riemann sum $S_P^*(f)$ using the lower right corners as sample points.

(b) Without calculating, will this Riemann sum be greater or smaller than the Riemann sum with upper right corners? Briefly explain.

7.2.8 Approximate the volume of the sand in 6×4 m rectangular sandbox whose four corners have depths of sand 30 cm, 60 cm, 70 cm and 100 cm. Explain the choices you made for your approximation. Is your approximation an over-estimate or an under-estimate?

Proofs

7.2.9 Abbi and Bevers are arguing about properties of upper sums. They each make a claim.

(a) Abbi makes a claim and scribbles down a rough proof sketch for her claim.

Claim A. *Let P be a partition of a rectangle R in $\mathbb{R}^n$. with subrectangles $\{R_i\}_{i\in I}$. Let $f:R\to\mathbb{R}$ be a bounded function. Fix $\lambda\in\mathbb{R}$. Then $U_P(\lambda f)=\lambda U_P(f)$.*

1. By definition,

$$U_P(\lambda f)=\sum_{i\in I}\Big[\sup_{x\in R_i}\big(\lambda f(x)\big)\Big]\operatorname{vol}(R_i) \tag{7.2.1}$$

$$=\sum_{i\in I}\Big[\lambda\sup_{x\in R_i}\big(f(x)\big)\Big]\operatorname{vol}(R_i) \tag{7.2.2}$$

$$=\lambda\sum_{i\in I}\Big[\sup_{x\in R_i}\big(f(x)\big)\Big]\operatorname{vol}(R_i)=\lambda U_P(f). \tag{7.2.3}$$

Bevers is skeptical of Abbi and finds an error. Identify it and fix the claim.

(b) Bevers makes a different claim and jots down an argument.

Claim B. *Let P be a partition of a rectangle R in $\mathbb{R}^n$ with subrectangles $\{R_i\}_{i\in I}$. Let $f:R\to\mathbb{R}$ and $g:R\to\mathbb{R}$ be bounded functions. Then $U_P(f+g)=U_P(f)+U_P(g)$.*

1. *By definition,*

$$\begin{aligned}
U_P(f+g) &= \sum_{i\in I}\Big[\sup_{x\in R_i}\big(f(x)+g(x)\big)\Big]\mathrm{vol}(R_i) && (7.2.4)\\
&= \sum_{i\in I}\Big[\sup_{x\in R_i}\big(f(x)\big)+\sup_{x\in R_i}\big(g(x)\big)\Big]\mathrm{vol}(R_i) && (7.2.5)\\
&= \sum_{i\in I}\Big[\sup_{x\in R_i}\big(f(x)\big)\Big]\mathrm{vol}(R_i)+\sum_{i\in I}\Big[\sup_{x\in R_i}\big(g(x)\big)\Big]\mathrm{vol}(R_i) && (7.2.6)\\
&= U_P(f)+U_P(g). && (7.2.7)
\end{aligned}$$

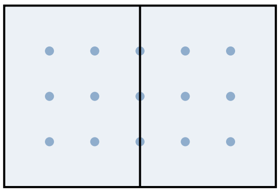
Abbi doesn't believe anything Bevers says, and spots a flaw. Identify it and fix the claim.

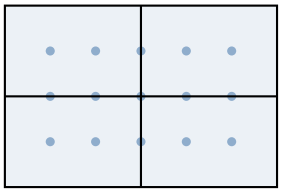
7.2.10 Recall the key lemma on upper sums and lower sums.

Lemma. *Let P and P' be partitions of a rectangle $R\subseteq\mathbb{R}^n$. Let $f:R\to\mathbb{R}$ be a bounded function. If P' is a refinement of P, then $L_P(f)\le L_{P'}(f)$ and $U_{P'}(f)\le U_P(f)$.*

The textbook proof is tough to read because of all the indices and notation. Whenever this happens, your best strategy is try to verify the simplest meaningful example. This illuminates the main ideas without all the clutter.

(a) Here is a drawing of a partition P and a refinement P' of the rectangle $R=[a,b]\times[c,d]$ in $\mathbb{R}^2$.





Label the 2 subrectangles $\{R_i\}_{i\in I}$ of P and 4 subrectangles $\{R'_j\}_{j\in J}$ of P'.

(b) Using your labels, show that for every $j\in J$, there exists $i\in I$ such that $R'_j\subseteq R_i$. What assumption implies this property?

(c) Show that $U_{P'}(f)\le U_P(f)$ for any bounded function $f:R\to\mathbb{R}$.

7.2.11 Let P be a partition for a rectangle R in $\mathbb{R}^n$. Let $f:R\to\mathbb{R}$ be bounded. Prove $L_P(f)\le U_P(f)$.

7.2.12 Let P be a partition for a rectangle R in $\mathbb{R}^n$. Let $f:R\to\mathbb{R}$ be bounded. Prove $U_P(-f)=-L_P(f)$.

7.2.13 Let P be a partition for a rectangle R in $\mathbb{R}^n$. Let $f:R\to\mathbb{R}$ and $g:R\to\mathbb{R}$ be bounded.

(a) Prove or disprove: if $f\le g$ on R, then $U_P(f)\le L_P(g)$.

(b) Prove or disprove: if $f\le g$ on R, then $U_P(f)\le U_P(g)$.

7.2.14 Let P be a partition for a rectangle R in $\mathbb{R}^n$. Let $\{R_i\}_{i\in I}$ be the subrectangles of P. For every $i\in I$, choose a sample point $x_i^*\in R_i$. Let $f:R\to\mathbb{R}$ and $g:R\to\mathbb{R}$ be bounded. Fix $\lambda\in\mathbb{R}$. Prove that

$$S_P^*(f+\lambda g)=S_P^*(f)+\lambda S_P^*(g).$$

7.3. Integration over rectangles

With all of your preparations, you can now define integration over a rectangle in $\mathbb{R}^n$. In this section, you will unfold the final step in the process.

> *Chop the rectangle into little pieces. Estimate the value of the function on each piece. Total your estimates to approximate the total value over the rectangle. Repeat this process for finer and finer pieces to approach the exact answer.*

You have learned to chop and estimate, and in this section you will learn to how to arbitrarily refine your estimates. You will want to take a "limit" to refine your over-estimates (upper integral) and to refine your under-estimates (lower integral). This limiting process will be performed with the supremum and infimum in a nearly identical fashion as with the single variable integral[5]. Since you have carefully generalized partitions and upper/lower sums as well as their properties, the proofs and arguments here will have minor distinctions between one variable and many variables. Keep this mind and use it to guide your intuition.

7.3.1 Definition of the integral

Before defining the integral, it will be helpful to define the "best possible" over-estimate as the upper integral and the "best possible" under-estimate as the lower integral.

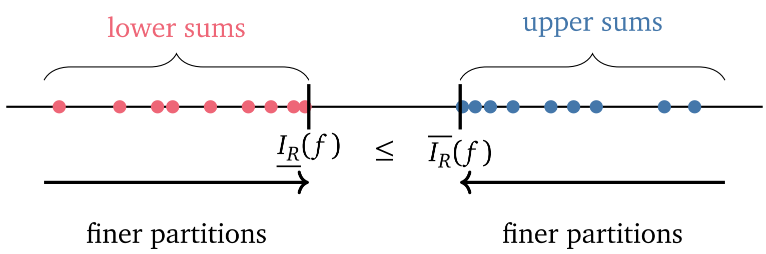
Definition 7.3.1 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. The **lower integral** of f on R and **upper integral** of f on R are defined by

$$\underline{I_R}(f) = \sup_P L_P(f) \qquad \text{and} \qquad \overline{I_R}(f) = \inf_P U_P(f),$$

where the supremum and infimum are over all partitions P of the rectangle R.

Remark 7.3.2 You can verify that the upper and lower integrals are always defined (and not infinite) for a bounded function on R.

The diagram below captures the spirit of this definition.



Remember there are inherent assumptions in this diagram which you have verified. All the lower sums are smaller than all the upper sums (Lemma 7.2.8). Finer partitions create better estimates; upper sums decrease and the lower sums increase (Lemma 7.2.7). Hence, the best possible over-estimate should be the *infimum* of all the upper sums (i.e. the greatest lower bound of all upper sums) and the best possible under-estimate should be the *supremum* of all the lower sums (i.e. the smallest upper bound of all lower sums). This collection of ideas lead to Definition 7.3.1 and prove that the lower integral is less than or equal to the upper integral.

[5]Again, if you want to review some single variable calculus, watch this MAT137 video on the definition of the single-variable integral [9]. It gives the basic intuition and the technical details.

Lemma 7.3.3 Let R be a rectangle in $\mathbb{R}^n$. If $f : R \to \mathbb{R}$ is a bounded function, then both $\underline{I_R}(f)$ and $\overline{I_R}(f)$ exist, and $\underline{I_R}(f) \le \overline{I_R}(f)$.

Proof. Let P' be a partition of the rectangle R. By Lemma 7.2.8, it follows that $L_P(f) \le U_{P'}(f)$ for any partition P of R. This implies that $U_{P'}(f)$ is an upper bound for the set of lower sums of f on R. Hence, the supremum of this set exists and satisfies

$$\underline{I_R}(f) = \sup_P L_P(f) \le U_{P'}(f).$$

This proves $\underline{I_R}(f)$ exists. Note the partition P' is arbitrary in the above inequality. Hence, the lower integral $\underline{I_R}(f)$ is a lower bound for the set of upper sums of f on R. Hence, the infimum of this set exists and satisfies $\underline{I_R}(f) \le \inf_P U_P(f) = \overline{I_R}(f)$. ■

You can directly calculate the upper and lower integral in some basic examples.

Example 7.3.4 Recall Example 7.2.3. Let $f(x,y) = 2x + y$. Fix $N \in \mathbb{N}^+$. Let $\{x_0, x_1, \ldots, x_N\}$ and $\{y_0, y_1, \ldots, y_N\}$ be regular partitions of $[0,1]$ and $[2,4]$, respectively. Define the regular partition P_N of the rectangle $R = [0,1] \times [2,4]$ by $P_N = (\{x_0, x_1, \ldots, x_N\}, \{y_0, y_1, \ldots, y_N\})$.

You verified

$$L_{P_N}(f) = 8 - \frac{4}{N}, \quad U_{P_N}(f) = 8 + \frac{4}{N}.$$

Since P_N is a partition of R, it follows that $L_{P_N}(f) \le \sup_P L_P(f) = \underline{I_R}(f)$ and similarly, $U_{P_N}(f) \ge \inf_P U_P(f) = \overline{I_R}(f)$. Combining all of these observations, you can conclude that

$$8 - \frac{4}{N} = L_{P_N}(f) \le \underline{I_R}(f) \le \overline{I_R}(f) \le U_{P_N}(f) = 8 + \frac{4}{N}$$

for $N \in \mathbb{N}^+$. Taking $N \to \infty$, the squeeze theorem implies that $\underline{I_R}(f) = \overline{I_R}(f) = 8$.

The lower integral and upper integral each represent a different way to chop, estimate, and refine. Are these processes the same? That is, do they yield the exact same answer? In any reasonable scenario, you want the lower and upper integrals be equal. This ultimately motivates the definition that you have been working very hard to build.

Definition 7.3.5 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. If $\underline{I_R}(f) = \overline{I_R}(f)$, then f is **integrable** on R and the **integral** of f on R is defined by

$$\int_R f\, dV := \underline{I_R}(f) = \overline{I_R}(f).$$

If $\underline{I_R}(f) < \overline{I_R}(f)$, then f is **non-integrable**.

Remark 7.3.6 The dV stands for "volume element" but it is only notation and has no meaning[6] on its own. Notice the integrand is written as f and not $f(x)$; that choice is intentional to avoid confusion with future notation.

Remark 7.3.7 There are many other definitions of the integral; this is the **Darboux integral**.

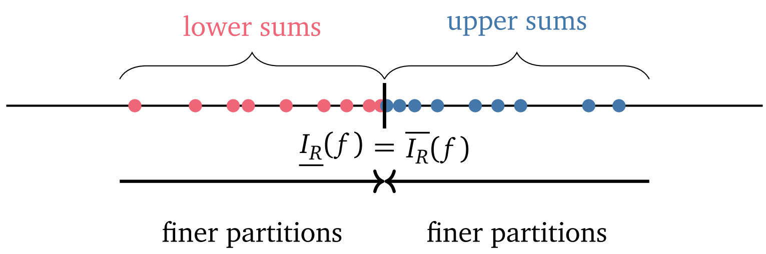
[6]There is a way to formalize this relationship with differential forms but that is beyond the scope of this course.

Example 7.3.8 From your previous example, you showed that $f(x,y) = 2x + y$ is integrable on $R = [0,1] \times [2,4]$ with

$$\int_R f\, dV = 8,$$

because $\underline{I_R}(f) = \overline{I_R}(f) = 8$.

If f is integrable on the rectangle R then the previous diagram is what you might expect.



As you shall see, many reasonable functions are integrable, but there are exceptions.

Example 7.3.9 Define the unit rectangle $R = [0,1]^2$ and the function $f : \mathbb{R}^2 \to \mathbb{R}$ as

$$f(x,y) = \begin{cases} 1 & x \in \mathbb{Q} \text{ and } y \in \mathbb{Q} \\ -1 & x \notin \mathbb{Q} \text{ or } y \notin \mathbb{Q} \end{cases}$$

Is f integrable on R? No, it is not for a critical reason: any open interval of $\mathbb{R}$ contains both rational and irrational numbers. Here is an outline[7] of the proof. Let P be a partition of R with index set I and subrectangles $\{R_i\}_{i \in I}$. Any subrectangle of R contains coordinate pairs that either both belong to $\mathbb{Q}$ or neither do. This implies that the infimum of f on any subrectangle R_i is -1 and the supremum is 1. Thus,

$$U_P(f) = \sum_{i \in I} M_i \operatorname{area}(R_i) = 1 \qquad L_P(f) = \sum_{i \in I} m_i \operatorname{area}(R_i) = -1.$$

Since our partition is arbitrary then this holds for any partition of R. Therefore,

$$-1 = \underline{I_R}(f) < \overline{I_R}(f) = 1,$$

so f is not integrable on R. See this MAT137 video for a similar example.

You have also the necessary properties for upper and lower sums to establish an equivalent definition for integrability, which is quite useful for proofs.

Lemma 7.3.10 (ε-characterization for integrability) Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. The function f is integrable on R if and only if

$$\forall \varepsilon > 0, \exists \text{ a partition } P \text{ of } R \text{ such that } U_P(f) - L_P(f) < \varepsilon.$$

Proof. This is left as an exercise. It is depicted nicely through the familiar diagram above for integrable functions, so use this to your advantage. ∎

You have finally completed the delicate definition of the integral over a rectangle!

[7] A full proof of this example requires a rigorous construction of the real numbers. See an introductory course on real analysis for details; only a sketch is provided here.

Remark 7.3.11 A few remarks on notation are in order. Other notation includes

$$\int \cdots \int_R f\,dV \qquad \text{or} \qquad \int_R f$$

For rectangles in $\mathbb{R}^3$, you may also write

$$\iiint_R f\,dV$$

For rectangles in $\mathbb{R}^2$, you may also write

$$\int_R f\,dA \qquad \text{or} \qquad \iint_R f\,dA$$

because dA stands for "area element". Do **not** write $dx\,dy$ instead of dA or $dx_1 \dots dx_n$ instead of dV. This will have a different meaning.

7.3.2 Properties of integrals over rectangles

With the integral over a rectangle defined, you can reflect on some of the original questions about volumes, net totals, and mass at the start of Chapter 7. If this definition of the integral is sensible, then it should possess some natural properties about volumes and mass. For example, if you cut a block of cheese into two blocks, then the total mass of the block should be the sum of the mass of the two blocks. There should be a corresponding property of the integral for it to be a valid interpretation of mass.

One key goal is to formally establish these properties. You must usually unpack everything including partitions, refinements, upper and lower sums, and upper and lower integral. Proving these facts are excellent exercises, so you will notice that this subsection has few proofs but many sketches. The first property is the perfect warm up.

Lemma 7.3.12 Let R be a rectangle in $\mathbb{R}^n$. The identity function 1 is integrable on R and

$$\int_R 1\,dV = \mathrm{vol}(R).$$

Proof. This is left as an exercise. This is straightforward but it is not trivial; remember $\mathrm{vol}(R)$ is defined independently of the integral itself. ■

Second, the integral acts like a linear transformation on functions.

Theorem 7.3.13 (Linearity) Let R be a rectangle in $\mathbb{R}^n$. Let f and g be bounded functions on R. Let $\lambda \in \mathbb{R}$. If f and g are integrable on R and $\lambda \in \mathbb{R}$, then $f + \lambda g$ is integrable on R and

$$\int_R (f + \lambda g)\,dV = \int_R f\,dV + \lambda \int_R g\,dV.$$

Proof. This is left as an exercise. Divide into cases depending on whether λ is positive or negative, and use Lemma 7.2.9. ■

Third, the integral is monotone.

Theorem 7.3.14 (Monotonicity) Let R be a rectangle in $\mathbb{R}^n$. Let f and g be bounded functions on R. If f and g are integrable on R and $f \leq g$ on R, then

$$\int_R f\,dV \leq \int_R g\,dV.$$

Proof. This is left as an exercise. Compare one integral with its lower sums and the other integral with its upper sums. Then use the definition of integrability. ■

Fourth, integrability on a rectangle implies absolute integrability[8] on a rectangle.

Theorem 7.3.15 (Triangle inequality) Let R be a rectangle in $\mathbb{R}^n$. Let f be a bounded function on R. If f is integrable on R, then $|f|$ is integrable on R and

$$\left|\int_R f\,dV\right| \leq \int_R |f|\,dV.$$

Proof. This is left as an exercise. Start by proving that $U_P(|f|) - L_P(|f|) \leq U_P(f) - L_P(f)$ for any partition P of R. For each subrectangle R_i of P, try to consider cases depending on the positivity or negativity of $\inf\limits_{x\in R_i} f(x)$ or $\sup\limits_{x\in R_i} f(x)$. Draw a picture with $n = 1$. ■

Fifth, the product of integrable functions is integrable by Cauchy-Schwarz.

Theorem 7.3.16 (Cauchy-Schwarz) Let R be a rectangle in $\mathbb{R}^n$. Let f and g be bounded functions on R. If f and g are integrable on R, then their product fg is integrable on R and

$$\int_R fg\,dV \leq \left(\int_R f^2\,dV\right)^{1/2}\left(\int_R g^2\,dV\right)^{1/2}.$$

Proof. This is left as a tricky exercise; it requires a neat trick and does not follow the same trend as the other proofs. Define the function $H(\lambda) = \int_R (f + \lambda g)^2\,dV$. Write $H(\lambda)$ as a quadratic in λ. What can you conclude about the discriminant of H as a quadratic in λ? ■

Sixth, the integral is additive; you can break up the domain into two other rectangles.

Theorem 7.3.17 (Additivity) Let R be a rectangle in $\mathbb{R}^n$. Let f be a bounded function on R. Suppose $R = R' \cup R''$ is a union of two subrectangles R' and R'' with disjoint interiors. The function f is integrable on R if and only if f is integrable on both R' and R'', in which case

$$\int_R f\,dV = \int_{R'} f\,dV + \int_{R''} f\,dV.$$

Proof. This is left as a challenging exercise. The basic idea is that any partition of R can be refined so that it is the union of a partition of R' and R''. ■

Finally, although the integral is defined with lower and upper sums, you can relate its definition to Riemann sums via the norm of partitions. Although you will not use this for establishing its properties, it can be quite helpful for deriving integral formulas in applications.

[8] This implication may appear alarming when you consider single-variable *improper* integrals. However, in this case, the functions are bounded and the domain is a rectangle (and hence bounded), so there is no such concern.

Theorem 7.3.18 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Let $P_1, P_2, \ldots$ be a sequence of partitions of R such that $||P_N|| \to 0$ as $N \to \infty$. For each partition P in the sequence, pick a sample point $x_i^* \in R_i$ for every subrectangle R_i of P. If f is integrable on R then

$$\int_R f\,dV = \lim_{N\to\infty} S^*_{P_N}(f)$$

Proof. Omitted. You will need prove that this textbook's definition of the integral (the Darboux integral) is equivalent to the Riemann integral. This technical issue is beyond the scope of this text and, in all honesty, it is not especially illuminating. ■

This completes the intricate definition of the integral over a rectangle and a long list of its properties! You have achieved quite a lot after some significant labour. Unfortunately, you do not yet have a simple criteria for proving that a function is integrable. Your current methods are limited and complex. Thus far, you can try a direct proof from definition as in Example 7.3.8, but that is quite involved. You have an equivalent definition (Lemma 7.3.10) but this does not appear any simpler. For applications, it will be vital to possess a broad class of functions which you can quickly check are integrable[9]. In the next section, you will prove a fantastic criteria for integrability by advancing your knowledge of topology.

[9]This parallels the problem you faced with differentiability. In that case, you later learned that C^1 functions are differentiable and it is easy to verify whether a function is C^1.

Exercises for Section 7.3

Concepts and definitions

7.3.1 Let $R = [0,2] \times [3,7]$. Consider the discontinuous function

$$f(x,y) = \begin{cases} -1 & \text{if } x = 0 \\ 2 & \text{otherwise} \end{cases}.$$

(a) You have proven that $U_P(f) = 16$ for all partitions P of R.
Using only this information about the upper sums, what can you conclude about $\overline{I_R} f$?

(b) For all $0 < \varepsilon < 1$, you have constructed a partition P_ε of R such that $L_{P_\varepsilon}(f) = 16 - 4\varepsilon$.
Using only this information about the lower sums, what can you conclude about $\underline{I_R} f$?

(c) Is f integrable on R?

7.3.2 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be bounded. Which statements are true or false?

(a) For every partition P of R, $L_P(f) \leq \underline{I_R} f$.

(b) There exists a partition P of R such that $\underline{I_R} f = L_P(f)$.

(c) For $\varepsilon > 0$, there exists a partition P of R such that $\underline{I_R} f - \varepsilon \leq L_P(f)$.

(d) If f is integrable on R, then there exists a partition P of R such that $\underline{I_R} f = L_P(f)$.

7.3.3 Let $R = [0,2] \times [3,7]$. Define the discontinuous function

$$f(x,y) = \begin{cases} 5 & \text{if } (x,y) = (0,3), \\ 2 & \text{otherwise.} \end{cases}$$

(a) For $\varepsilon > 0$, construct a partition P_ε of R such that $U_{P_\varepsilon}(f) \leq 16 + \varepsilon^2$. Sketch your partition and label it according to the notation in your argument.

(b) Compute $L_{P_\varepsilon}(f)$. From this information only, what can you conclude about $\underline{I_R} f$?

(c) Is f integrable on R?

7.3.4 Let $R = [1,3] \times [1,3]$. Define the discontinuous function

$$f(x,y) = \begin{cases} 1 & \text{if } x \geq y, \\ 0 & \text{otherwise.} \end{cases}$$

For $N \in \mathbb{N}^+$, let P_N be a regular partition of R constructed from N subintervals for each interval $[1,3]$. Hence, P_N has N^2 subrectangles.

(a) Illustrate the subrectangles $\{R_{ij}\}_{i,j}$ of P_N with a 2D sketch of R. Label it with standard notation. Also separately draw and label a typical subrectangle R_{ij}.

(b) Compute m_{ij} and M_{ij} for all subrectangles R_{ij}.

(c) Express $U_{P_N}(f)$ and $L_{P_N}(f)$ as double sums. Then compute them.

(d) Is f integrable on R?

Proofs

7.3.5 Let $R = [0,1] \times [2,5]$ and let $f(x,y) = -x + 3y$. Doja is explaining to her friend, Ariana, how to prove that f is integrable on R. She sketches down the argument below.

1. *Fix $N \in \mathbb{N}^+$. Let P_N be a regular partition of R with N subintervals of $[0,1]$ and N subintervals of $[2,5]$, so there are N^2 subrectangles of P_N.*
2. *You can check that the upper sum and lower sum are given by*
$$U_{P_N}(f) = 30 + \frac{15}{N}, \quad L_{P_N}(f) = 30 - \frac{15}{N}.$$
3. *It follows that*
$$\overline{I_R} f = \lim_{N\to\infty} U_{P_N}(f) = \lim_{N\to\infty} \left(30 + \frac{15}{N}\right) = 30,$$
and
$$\underline{I_R} f = \lim_{N\to\infty} L_{P_N}(f) = \lim_{N\to\infty} \left(30 - \frac{15}{N}\right) = 30.$$
4. *Since $\overline{I_R} f = 30 = \underline{I_R} f$, you can conclude that f is integrable on R.*

Ariana agrees that the upper and lower sums are correctly calculated, but she notices that Doja's argument is flawed. One line is missing some important justifications. Identify that line and fix it.

7.3.6 Remember two key lemmas on partitions.

Lemma A. *Let P be a partition of a rectangle $R \subseteq \mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Then $L_P(f) \le U_P(f)$.*

Lemma B. *Let P' and P'' be partitions of a rectangle $R \subseteq \mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Then $L_{P'}(f) \le U_{P''}(f)$.*

(a) Here is a WRONG proof of the fact $\underline{I_R} f \le \overline{I_R} f$.

1. *By Lemma A, $L_P(f) \le U_P(f)$ for every partition P of R.*
2. *Taking the supremum of one side and the infimum of the other,*
$$\underline{I_R} f = \sup_P L_P(f) \le \inf_P U_P(f) = \overline{I_R} f.$$

Identify the line with the flaw and explain why.

(b) Here is a BAD proof of the fact $\underline{I_R} f \le \overline{I_R} f$.

1. *By Lemma B, $L_{P'}(f) \le U_{P''}(f)$ for any partitions P' and P'' of R.*
2. *Taking the supremum over P' and the infimum over P'',*
$$\underline{I_R} f = \sup_{P'} L_{P'}(f) \le \inf_{P''} U_{P''}(f) = \overline{I_R} f.$$

There are not enough details in one line. Identify the line and add enough details.

7.3.7 The ε-characterization for integrability is a standard tool and uses several of the key partition lemmas. You will establish it for rectangles in $\mathbb{R}^n$. The proofs are nearly identical to the single variable case. Let f be a bounded real-valued function on a rectangle R in $\mathbb{R}^n$.

(a) Prove if $\forall \varepsilon > 0, \exists$ a partition P of R s.t. $U_P(f) - L_P(f) < \varepsilon$, then f is integrable on R.
Hint: For any partition P, order the quantities $U_P(f), L_P(f), \overline{I_R}f$, and $\underline{I_R}f$.

(b) Prove that if f is integrable on R then $\forall \varepsilon > 0, \exists$ partition P of R such that $U_P(f) - L_P(f) < \varepsilon$. (To get you started, the first couple of lines are given. Finish the proof.)

> 1. *Fix $\varepsilon > 0$. By definition,*
> $$\overline{I_R}f = \inf_P U_P(f) \quad \textit{and} \quad \underline{I_R}f = \sup_P L_P(f).$$
> *where the infimum and supremum are over all partitions P of R.*
> 2. *By the definition of infimum and supremum, there exists partitions P' and P'' such that*
> $$U_{P'}(f) < \overline{I_R}f + \frac{\varepsilon}{2} \quad \textit{and} \quad L_{P''}(f) > \underline{I_R}f - \frac{\varepsilon}{2}.$$

7.3.8 Proofs of integral properties have a natural setup, but there are often subtleties.

(a) Consider the following WRONG proof of additivity over a rectangular set R.

> 1. *Fix a partition P of the rectangle R.*
> 2. *Notice that $U_P(f+g) = U_P(f) + U_P(g)$ and $L_P(f+g) = L_P(f) + L_P(g)$.*
> 3. *Since f, g are integrable, taking supremums and infimums gives*
> $$\underline{I_R}(f+g) = \underline{I_R}f + \underline{I_R}g \qquad \textit{and} \qquad \overline{I_R}(f+g) = \overline{I_R}f + \overline{I_R}g$$
> 4. *Since f and g are integrable on R, this implies $\underline{I_R}(f+g) = \underline{I_R}f + \underline{I_R}g = \overline{I_R}f + \overline{I_R}g = \overline{I_R}(f+g)$.*
> 5. *So $f+g$ is integrable and $\displaystyle\int_R (f+g)\,dV = \int_R f\,dV + \int_R g\,dV$.*

This argument has multiple errors. Identify the line with the first error and explain the error.

(b) Write a corrected proof. (This is not easy. You will need at least one page of space.)

7.3.9 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded function. Ilana makes a claim.

Ilana's claim. *If $|f|$ is integrable on R, then f is integrable on R.*

(a) Ilana gives an attempted WRONG proof.

> 1. *Suppose $|f|$ is integrable on the rectangle R.*
> 2. *Since $-|f| \le f \le |f|$ on R, by monotonicity, $\displaystyle\int_R -|f|\,dV \le \int_R f\,dV \le \int_R |f|\,dV$.*
> 3. *Therefore, f is integrable on R.*

Identify the line with the error and explain why it is incorrect.

(b) Is the statement true or false? If true, prove it. If false, give a counterexample.

7.3.10 Fix $\lambda \in \mathbb{R}$. Show that the constant function $f(x) = \lambda$ is integrable on any rectangle R in $\mathbb{R}^n$ and

$$\int_R \lambda dV = \lambda \operatorname{vol}(R).$$

7.3.11 Let R be a rectangle in $\mathbb{R}^n$. Prove that if f, g are integrable on R and $f \leq g$ then

$$\int_R f dV \leq \int_R g dV.$$

Hint: Use a lower sum for one integral and an upper sum for the other.

7.3.12 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be bounded. Prove or disprove.

(a) If f is integrable on R, then there exists a partition P such that $U_P(f) = L_P(f)$.

(b) If there exists a partition P such that $U_P(f) = L_P(f)$, then f is integrable on R.

7.4. Uniform continuity and integration

You have developed a solid definition of the integral over a rectangle by chopping, estimating, and refining. The integral has many natural properties which are relatively reasonable to prove with the theory that you have developed. There is, however, a serious obstacle.

> *How can you verify whether a function is integrable on a rectangle?*

It is impractical to check directly by the definition for every function that you encounter. There would be a lot to calculate and prove each time. Your goal in this section is to establish a simple criterion for verifying integrability. A new concept in topology – uniform continuity – will be pivotal to your success, so you will temporarily put integration aside and independently unravel this concept. Once your investigation is complete, you will return to this section's motivating question and prove a great theorem.

> *Continuous functions on rectangles are integrable.*

Your new tool will power the proof of this landmark result on integration.

7.4.1 Uniform continuity

Continuity is as a *local* property of a function $f: A \to \mathbb{R}^m$ with domain $A \subseteq \mathbb{R}^n$; it holds at a given point in A. This can be seen from its formal definition. Recall f is continuous exactly when

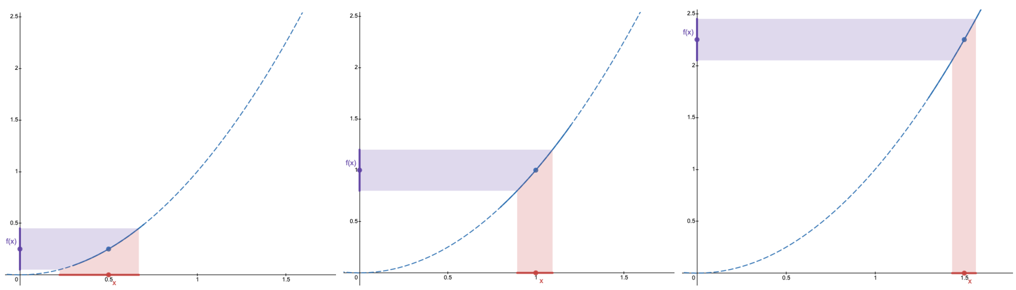
$$\forall x \in A, \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall y \in A, ||x - y|| < \delta \implies ||f(x) - f(y)|| < \varepsilon. \qquad (7.4.1)$$

Informally speaking, once y is δ-close to x, the value $f(y)$ is ε-close to $f(x)$. The order of quantifiers reveals a critical observation. The quantity δ appears after ε and after x.

> *To prove f is continuous at x, the choice of δ may also depend on the point x.*

If the point x changes, then you may need a smaller δ to ensure $f(y)$ is ε-close to $f(x)$. This can occur if f changes rapidly. You can view this phenomenon in an example.

Example 7.4.1 Consider the continuous function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Illustrated below is a verification of continuity with $\varepsilon = 0.2$ and three choices of x, namely $x = 0.5, 1,$ and 1.5. For each point x, a corresponding δ is chosen to satisfy (7.4.1).



Play with the Desmos graph for a better visual. Notice that for larger x, the same choice of $\varepsilon = 0.2$ requires smaller and smaller choices of δ. Heuristically speaking, this occurs since the parabola changes more rapidly for larger values of x.

Since integration requires you to approximate by f on its entire domain, you need a version of continuity that does not depend on a given point. In other words, you want δ to depend only on ε and not on a particular point. This can be achieved by simply swapping the order of quantifiers in (7.4.1), which creates a new fundamental notion.

Definition 7.4.2 Let $A \subseteq \mathbb{R}^n$ be a set. A function $f : A \to \mathbb{R}^m$ is **uniformly continuous** if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x, y \in A, ||x - y|| < \delta \implies ||f(x) - f(y)|| < \varepsilon.$$

Remark 7.4.3 It will be convenient to discuss uniform continuity on subsets. For a subset $S \subseteq A$, the function f is **uniformly continuous on** S if its restriction $f|_S : S \to \mathbb{R}^m$ is uniformly continuous.

Notice the definition is nearly identical to (7.4.1) aside from the order of quantifiers. Consequently, uniform continuity is a *global* property of a function; it does not depend on any particular point in the domain. You can explore some examples.

Example 7.4.4 Let $f(x) = x^2$. Recall Example 7.4.1 suggests f is not uniformly continuous on its domain $\mathbb{R}$, because the parabola grows more and more rapidly for large values of x. By restricting its domain, can you restrict this growth and force it to be uniformly continuous? Indeed you can in this example. For instance, you can prove by definition that f is uniformly continuous on the open interval $I = (1, 2)$.

Proof. Let $\varepsilon > 0$ be arbitrary and fix $\delta = \frac{\varepsilon}{4}$. Let $x, y \in (1, 2)$ and assume $|x - y| < \delta$. It suffices to show $|f(x) - f(y)| < \varepsilon$. Notice that since $x, y \in (1, 2)$ then $|x + y| < 4$. As $|x + y| < 4$, this implies that

$$|f(x) - f(y)| = |x^2 - y^2| = |x + y||x - y| < 4|x - y| < 4\delta = \varepsilon,$$

as required. ■

Example 7.4.5 Let $f(x) = x^2$. You can prove by definition that f is not uniformly continuous on the open interval $J = (1, \infty)$. By negating the definition, you must therefore show that

$$\exists \varepsilon > 0 \text{ s.t. } \forall \delta > 0, \exists x, y \in (1, \infty) \text{ s.t. } |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \varepsilon.$$

Proof. Fix $\varepsilon = 1$ and let $\delta > 0$. Fix $x = \frac{5\delta^2+4}{4\delta} + 1$ and $y = \frac{3\delta^2+4}{4\delta} + 1$. Note that the choice of x and y are valid since $\delta > 0$ thus $x, y \in (1, \infty)$. It suffices to show $|x - y| < \delta$ and $|f(x) - f(y)| \geq \varepsilon$. First, you can see that $|x - y| < \delta$, because

$$|x - y| = \left| \frac{5\delta^2 + 4}{4\delta} + 1 - \frac{3\delta^2 + 4}{4\delta} - 1 \right| = \frac{2\delta^2}{4\delta} = \frac{\delta}{2} < \delta.$$

Next, you can see that $|f(x) - f(y)| \geq \varepsilon$, because

$$|f(x) - f(y)| = |x^2 - y^2| = |x + y||x - y| = \left| 2\delta + \frac{2}{\delta} + 2 \right| \left| \frac{\delta}{2} \right| = \delta^2 + \delta + 1 \geq 1 = \varepsilon.$$

Hence, both $|x - y| < \delta$ and $|f(x) - f(y)| \geq \varepsilon$ as required. ■

The previous example exhibits a continuous function that is not uniformly continuous.

Lemma 7.4.6 If $f : A \to \mathbb{R}^m$ is uniformly continuous, then f is continuous.

Proof. This is left as an exercise. Proceed directly from definitions. ■

Uniform continuity is therefore strictly stronger than continuity. This lemma does not help you verify continuity, but its contrapositive can sometimes be useful. As you can see from the last example, proving a function is not uniformly continuous using the full definition can be a dreadful exercise. The lemma provides a quick shortcut for functions that are discontinuous.

Example 7.4.7 The function $H : \mathbb{R}^2 \to \mathbb{R}$ given by

$$H(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{otherwise} \end{cases}$$

is not continuous at $(0,0)$ as explained in Example 2.6.12. By Lemma 7.4.6, H is not uniformly continuous on $\mathbb{R}^2$. In fact, this shows that H is not uniformly continuous on any subset containing $(0,0)$.

Although uniform continuity is stronger than continuity, the following brilliant theorem gives a partial converse.

Theorem 7.4.8 Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}^m$ be a function. If A is compact and f is continuous, then f is uniformly continuous.

Its proof will transport you back to the chapter on topology.

Proof. Suppose, for a contradiction, that f is not uniformly continuous on A. By definition, there exists $\varepsilon > 0$ such that

$$\forall \delta > 0, \exists x, y \in A \text{ s.t. } ||x-y|| < \delta \text{ and } ||f(x) - f(y)|| \geq \varepsilon.$$

For each $n \in \mathbb{N}^+$, take $\delta = \frac{1}{n}$ above to produce $x_n, y_n \in A$ such that

$$||x_n - y_n|| < \frac{1}{n} \text{ and } ||f(x_n) - f(y_n)|| \geq \varepsilon. \tag{7.4.2}$$

Since A is compact, the set $A \times A$ is compact by Lemma 2.4.11. Thus, the sequence $\{(x_n, y_n)\}_n$ in $A \times A$ has a convergent subsequence $\{(x_{k(n)}, y_{k(n)})\}_n$ for some monotone increasing function $k : \mathbb{N}^+ \to \mathbb{N}^+$. This subsequence converges to a point $(x,y) \in A \times A$, so $x_{k(n)} \to x$ and $y_{k(n)} \to y$ as $n \to \infty$. We claim that $x = y$.

By the triangle inequality and (7.4.2),

$$\begin{aligned} ||x-y|| &\leq ||x - x_{k(n)}|| + ||x_{k(n)} - y_{k(n)}|| + ||y_{k(n)} - y|| \\ &< ||x - x_{k(n)}|| + \frac{1}{k(n)} + ||y_{k(n)} - y||. \end{aligned}$$

Taking $n \to \infty$, notice the lefthand side is constant whereas each term on the righthand side tends to zero. Thus, by the monotonicity limit law, $||x-y|| \leq 0$ in which case $x = y$.

Finally, notice (7.4.2) implies that

$$\varepsilon \leq \lim_{n\to\infty} ||f(x_{k(n)}) - f(y_{k(n)})||.$$

Since f is continuous, the righthand side is equal to $||f(x) - f(y)|| = 0$ as $x = y$. This shows $\varepsilon \leq 0$, a contradiction. ∎

This theorem gives you a method for verifying uniform continuity on compact sets.

Example 7.4.9 The function $f(x) = x^2$ is continuous on $\mathbb{R}$. Thus, f is uniformly continuous on any compact subset of $\mathbb{R}$ by Theorem 7.4.8.

With a minor addition, you can also verify uniform continuity on non-compact sets.

Example 7.4.10 Let $f(x) = x^2$. You verified in Example 7.4.4 that f is uniformly continuous on the open interval $I = (1,2)$ using the definition of uniform continuity. Here is a short alternate proof with your new tool.

The function $f(x) = x^2$ is continuous on the compact set $A = [1,2]$. By Theorem 7.4.8, f is uniformly continuous on $[1,2]$. Since the open interval $I = (1,2)$ is a subset of $A = [1,2]$, it follows that f is uniformly continuous on I.

This concludes your study of uniform continuity.

7.4.2 Continuous functions are integrable

Now, you may return to the world of integration and claim a fantastic reward.

Theorem 7.4.11 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a real-valued function. If f is continuous on R, then f is integrable on R.

This immediately yields a classic result on single-variable integration.

Corollary 7.4.12 Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a,b] \to \mathbb{R}$ be a real-valued function. If f is continuous on $[a,b]$, then f is integrable on $[a,b]$.

Uniform continuity will play a crucial role in the proof.

Proof. By the ε-criterion for integrability (Lemma 7.3.10), f is integrable if and only if for every $\varepsilon > 0$ there exists a partition P of R such that $U_P(f) - L_P(f) < \varepsilon$. It is sufficient to verify this property.

Let $\varepsilon > 0$. Write $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$. The set R is compact and f is continuous on R so f must be uniformly continuous on R by Theorem 7.4.8. Thus, there exists $\delta > 0$ such that

$$\forall x, y \in R, \|x - y\| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{\mathrm{vol}(R)}. \tag{7.4.3}$$

Pick $N \in \mathbb{N}^+$ as $N = \lceil \delta^{-1} \sum_{i=1}^{n} (b_i - a_i) \rceil + 1$ so that

$$\sum_{i=1}^{n} \frac{b_i - a_i}{N} < \delta.$$

Let P_N be a regular partition of $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ into N^n subrectangles constructed from a regular partition of each interval $[a_1, b_1], \ldots, [a_n, b_n]$ into N subintervals of equal width. Let $\{R_i\}_{i \in I}$ be the subrectangles of P_N.

Fix $i \in I$ temporarily. Since f is continuous on R and hence on the compact subset R_i, by the extreme value theorem, there exists $c_i, C_i \in R_i$ such that

$$m_i = \inf_{x \in R_i} f(x) = f(c_i), \qquad M_i = \sup_{x \in R_i} f(x) = f(C_i).$$

For any $x, y \in R_i$, it follows by the triangle inequality and our choice of N that

$$\|x - y\| \leq \sum_{j=1}^{n} |x_j - y_j| \leq \sum_{j=1}^{n} \frac{b_j - a_j}{N} < \delta.$$

Therefore, by (7.4.3), $c_i, C_i \in R_i$ implies that

$$M_i - m_i = f(C_i) - f(c_i) = |f(C_i) - f(c_i)| < \frac{\varepsilon}{\mathrm{vol}(R)}.$$

Since this estimate holds for any $i \in I$, it follows that

$$\begin{aligned} U_{P_N}(f) - L_{P_N}(f) = \sum_{i \in I}(M_i - m_i)\mathrm{vol}(R_i) &< \sum_{i \in I} \frac{\varepsilon}{\mathrm{vol}(R)}\mathrm{vol}(R_i) \\ &= \frac{\varepsilon}{\mathrm{vol}(R)} \sum_{i \in I} \mathrm{vol}(R_i) \\ &= \frac{\varepsilon}{\mathrm{vol}(R)}\mathrm{vol}(R) \qquad \text{by Theorem 7.1.10} \\ &= \varepsilon. \end{aligned}$$

Therefore $U_{P_N}(f) - L_{P_N}(f) < \varepsilon$, as desired. ■

You can now easily verify many functions are integrable without breaking a sweat.

Example 7.4.13 Let $f(x,y) = 2x + y$ and $R = [0,1] \times [2,4]$. Remember that, in a series of examples concluding with Example 7.3.8, you proved directly from the definition that f is integrable on R, and it was quite a lot of work. Theorem 7.4.11 now makes this effortless: since f is continuous, it follows that f is integrable on R.

Theorem 7.4.11 is an outstanding achievement but it is not the end of the story. It is unfortunately too restrictive in higher dimensions.

Example 7.4.14 Consider the unit disk

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

and the continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = 2 - \sin(xy).$$

What is the total value of f on S? You ideally you want to integrate f over S, but S is not a rectangle. There is a very clever idea to get around this non-rectangular issue.

Define $R = [-3,3] \times [-3,3]$ and define the **indicator function** on S by

$$\chi_S(x,y) = \begin{cases} 1 & \text{for } (x,y) \in S, \\ 0 & \text{for } (x,y) \notin S. \end{cases}$$

The disk S is contained inside the rectangle R and the product of functions $\chi_S f : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\chi_S f(x,y) = \chi_S(x,y)f(x,y) = \begin{cases} f(x,y) & \text{for } (x,y) \in S, \\ 0 & \text{for } (x,y) \notin S. \end{cases}$$

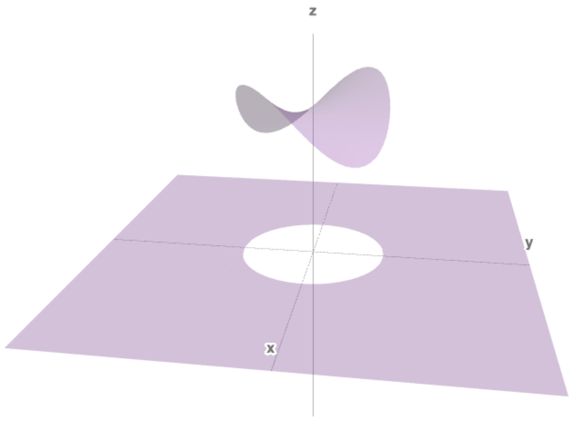
The total value of f on S should be given by

$$\iint_R \chi_S f \, dA.$$

Because of the way $\chi_S f$ is defined, you can expect anything inside R but outside S contributes nothing to the integral whereas anything inside S contributes according to the size of f. That's a pretty cute trick! Sadly, there's a serious issue. The function $\chi_S f$ is *discontinuous* at every point of the unit circle

$$\partial S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

A Math3D plot of $z = \chi_S f(x, y)$ is included below to illustrate this phenomenon.



Hence, $\chi_S f$ is discontinuous at infinitely many points inside the rectangle $R = [-3, 3] \times [-3, 3]$. Even with this exceptionally clever trick, you cannot apply Theorem 7.4.11.

On the other hand, there is reason to be optimistic. The set of discontinuities of $\chi_S f$ is the unit circle which is a 1-dimensional smooth manifold inside $\mathbb{R}^2$. You might expect that any contribution from f on this thin curve is negligible because the "area" of the curve should be zero. How can you show that this contribution is negligible? This will require a deeper investigation into volumes of non-rectangular sets.

You have successfully applied the concept of uniform continuity to produce a classic and powerful theorem on integrating over rectangles. This completes the first stage of defining the integral in higher dimensions! However, as you have just seen, the theory so far has serious limitations if you try to integrate over non-rectangular sets. You will need to overcome this hurdle because finding the volume of non-rectangular sets (or averaging over them) is an essential skill in higher dimensions.

This motivates a transition to the second stage: defining integrals over *non-rectangular sets*. Your new goal will be to prove a more powerful version of Theorem 7.4.11 that permits a "negligible" set of discontinuities as in Example 7.4.14. This will be achieved in the next few sections, and you will finally arrive at the full fledged definition of the integral equipped with some incredible theorems.

Exercises for Section 7.4

Concepts and definitions

7.4.1 Give a brief answer and explanation for each question below.

(a) Is $f(x) = x^2$ uniformly continuous on $S = [0,1]$?

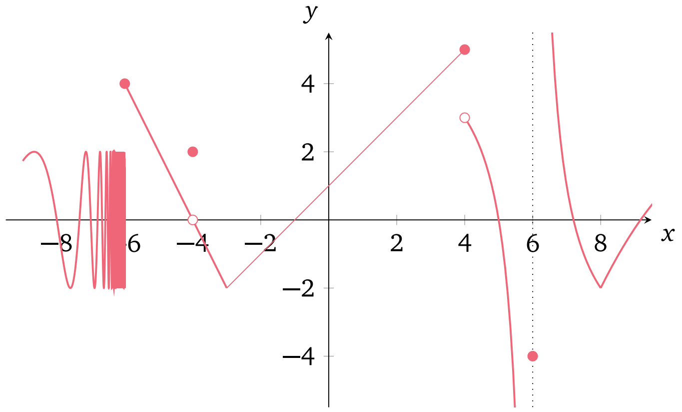
(b) Is $f(x) = x^2$ uniformly continuous on $S = \mathbb{R}$?

(c) Is $g(x) = \frac{1}{||x||}$ uniformly continuous on $S = \{x \in \mathbb{R}^n : 0 < ||x|| \leq 3\}$?

(d) Is $g(x) = \frac{1}{||x||}$ uniformly continuous on $S = \{x \in \mathbb{R}^n : 1 \leq ||x|| \leq 3\}$?

(e) Is $g(x) = \frac{1}{||x||}$ uniformly continuous on $S = \{x \in \mathbb{R}^n : 1 < ||x|| \leq 3\}$?

7.4.2 Consider the graph of $y = f(x)$ for $-9 \leq x \leq 9$. On which intervals $I \subseteq [-9,9]$ does f appear to be uniformly continuous?



(a) $I = [-6,-3]$

(b) $I = [-3,4]$

(c) $I = [-2,5]$

(d) $I = (4,6)$

(e) $I = [-8,-6)$

(f) $I = [-8,-6.1]$

(g) $I = [-8,-6.1)$

(h) $I = (6.1,9)$

7.4.3 Let $A \subseteq \mathbb{R}^n$ and $a \in A$. Let f be an $\mathbb{R}^m$-valued function on A. Which statements are true or false or nonsense? If true, state whether by definition or a theorem. If false, list a counterexample.

(a) If f is uniformly continuous at a, then f is continuous at a.

(b) If f is continuous on A, then f is uniformly continuous on A.

(c) If f is uniformly continuous on A, then f is continuous on A.

(d) If f is uniformly continuous on A, then f is uniformly continuous on any subset of A.

(e) If f is differentiable on A, then f is continuous on A.

(f) If f is differentiable on A, then f is uniformly continuous on A.

7.4.4 Let $f : D \to \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$. Which are equivalent to "f is uniformly continuous on D"?

(a) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall x, y \in D, ||x-y|| < \delta \implies ||f(x) - f(y)|| < \varepsilon$

(b) $\forall a \in D, \forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall x \in D, ||x-a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon$

(c) $\forall \varepsilon > 0, \forall a \in D, \exists \delta > 0$ s.t. $\forall x \in D, ||x-a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon$

(d) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall a \in D, \forall x \in D, ||x-a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon$

(e) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall x \in D, \forall a \in D, ||x-a|| < \delta \implies ||f(x) - f(a)|| < \varepsilon$

(f) $\forall a \in D, \forall \varepsilon > 0, \exists \delta > 0$ s.t. $f(B_\delta(a) \cap D) \subseteq B_\varepsilon(f(a))$

(g) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall a \in D, f(B_\delta(a) \cap D) \subseteq B_\varepsilon(f(a))$

Proofs

7.4.5 Prove that $f(x) = -x + 3y$ is integrable on $R = [0,1] \times [2,5]$. *Hint:* It is now very very short.

7.4.6 Prove if $f : \mathbb{R} \to \mathbb{R}$ is differentiable and f' is bounded, then f is uniformly continuous.
Hint: Use the single-variable mean value theorem.

7.4.7 Prove directly by definition that $f(x) = x^3 + x$ is uniformly continuous on $I = [-10, 10]$.

7.4.8 Prove directly by definition that $g(x) = \sin(1/x)$ is not uniformly continuous on $I = (0, 1]$.

7.4.9 Let $f : A \to \mathbb{R}^m$ be a function. Prove or disprove: if f is uniformly continuous on every compact subset of A, then f is uniformly continuous.

7.4.10 Let $A \subseteq \mathbb{R}^n$. Let $f : A \to \mathbb{R}^m$ and $g : A \to \mathbb{R}^m$. Fix $\lambda \in \mathbb{R}$. Show that if f and g are uniformly continuous, then $f + \lambda g$ is uniformly continuous.

Applications and beyond

7.4.11 It can be helpful to re-prove theorems in special cases to better understand them.

(a) Prove Corollary 7.4.12 by specializing the proof of Theorem 7.4.11. Your argument and the notation should simplify a bit, but the overall structure should be similar and your proof will crucially apply uniform continuity.

(b) Using continuity alone is not enough. Here is a WRONG proof of Corollary 7.4.12.

1. *Fix $\varepsilon > 0$. Let $N \in \mathbb{N}^+$ be arbitrary.*
2. *Let $P = \{x_0, x_1, \dots, x_N\}$ be the regular partition of $[a, b]$.*
   The width of each subinterval $[x_{i-1}, x_i]$ is equal to $\frac{b-a}{N}$.
3. *Since f is continuous on the compact subinterval $[x_{i-1}, x_i]$, by the extreme value theorem, f attains a maximum and a minimum on the subinterval. That is, there exists $c_i, C_i \in [x_{i-1}, x_i]$ such that*

$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x) = f(c_i) \qquad \textit{and} \qquad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) = f(C_i).$$

4. *Since the width of $[x_{i-1}, x_i]$ is $\frac{1}{N}$, it follows that $|C_i - c_i| \le |x_{i-1} - x_i| = \frac{1}{N}$.*
5. *Hence, as $N \to \infty$, $|C_i - c_i| \to 0$.*
6. *By continuity of f, $M_i - m_i = |f(C_i) - f(c_i)| \to 0$ as $N \to \infty$.*
7. *It follows by the definition of upper and lower sums that*

$$U_P(f) - L_P(f) = \sum_{i=1}^{N} (M_i - m_i)(x_i - x_{i-1}) = \frac{1}{N} \sum_{i=1}^{N} (M_i - m_i).$$

8. *Since $M_i - m_i \to 0$ as $N \to \infty$, it follows that*

$$U_P(f) - L_P(f) \to 0 \qquad \textit{as } N \to \infty.$$

9. *Therefore, for N large enough, $U_P(f) - L_P(f) < \varepsilon$ as required.*

This proof has false claims on two lines. Identify these lines and explain the flaws.

7.4.12 Uniform continuity is critical to Theorem 7.4.11. Here is a condensed CORRECT proof.

1. *Fix $\varepsilon > 0$. Since f is continuous on the compact set R, f is uniformly continuous on R.*
2. *Thus, there exists $\delta > 0$ such that*

$$\forall x, y \in R, ||x - y|| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{\text{vol}(R)}. \qquad (\star)$$

3. *Set $N = \lceil \delta^{-1} \sum_{i=1}^{n} (b_i - a_i) \rceil + 1$ so that $\displaystyle\sum_{i=1}^{n} \frac{b_i - a_i}{N} < \delta$.*
4. *Let $P_N = \{R_i\}_{i\in I}$ be a regular partition of $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ into N^n subrectangles made from a regular partition of each interval $[a_1, b_1], \dots, [a_n, b_n]$ into N subintervals of equal width.*
5. *Fix a subrectangle R_i. Since f is continuous on R and hence on the compact subset R_i, by the extreme value theorem, there exists $c_i, C_i \in R_i$ such that*

$$m_i = \inf_{x\in R_i} f(x) = f(c_i), \qquad M_i = \sup_{x\in R_i} f(x) = f(C_i).$$

6. *It follows by the triangle inequality and our choice of N that*

$$\forall x, y \in R_i, \qquad ||x - y|| \le \sum_{j=1}^{n} |x_j - y_j| \le \sum_{j=1}^{n} \frac{b_j - a_j}{N} < \delta.$$

7. *Therefore, by $(\star)$, $c_i, C_i \in R_i$ implies that*

$$M_i - m_i = f(C_i) - f(c_i) = |f(C_i) - f(c_i)| < \frac{\varepsilon}{\text{vol}(R)}.$$

8. *Since this estimate holds for any subrectangle R_i, it follows that*

$$U_{P_N}(f) - L_{P_N}(f) = \sum_{i\in I} (M_i - m_i)\text{vol}(R_i) < \sum_{i\in I} \frac{\varepsilon}{\text{vol}(R)} \text{vol}(R_i) = \frac{\varepsilon}{\text{vol}(R)} \sum_{i\in I} \text{vol}(R_i) = \varepsilon.$$

9. *This establishes the ε-criterion for integrability and hence completes the proof.*

(a) Line 3 defines N. Find a line later in the proof which uses this definition. What does this choice ensure about the norm of the partition P_N?

(b) You can modify the proof to avoid defining a regular partition. Fill in the blank to replace lines 3 and 4.

Choose a partition P of R such that ________________________.

(c) For $i \in I$, give an exact expression for the difference **in the heights of** f between your overestimate and underestimate on the subrectangle R_i. *Hint:* The expression appears in one of the later lines.

(d) For $i \in I$, the proof estimates this difference by what quantity? This quantity does not depend on $i \in I$. Explain how the author ensures this lack of dependence.

7.5. Jordan measurable sets and volume

To continue your pursuit of integrating over a non-rectangular set, Example 7.4.14 suggests that you must formalize the idea that a set of discontinuities is negligible or, equivalently, has "zero volume". This issue forces you to first resolve one of this chapter's motivating questions.

What is the volume of a set S in $\mathbb{R}^n$? Is it always defined?

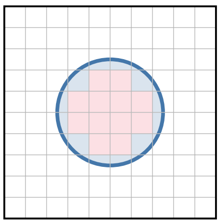
Since rectangles are the only object for which you have defined volume, you must use rectangles to define the area of non-rectangles. Indeed, integration on rectangles and the indicator function will generate a solution.

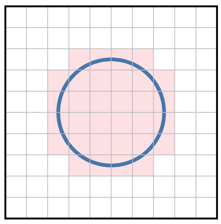
Definition 7.5.1 (Indicator function) The **indicator function** of a set $S \subseteq \mathbb{R}^n$ is denoted[10] $\chi_S : \mathbb{R}^n \to \mathbb{R}$ and defined by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

To define volume, you will proceed with your usual process of chopping, estimating and refining. An illustrative example with the unit disk reveals the core intuition.

Example 7.5.2 Let $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ be the unit disk in $\mathbb{R}^2$. Your goal is to rigorously define and compute the area of S. Start by placing S inside a larger rectangle, say $R = [-2,2]^2$ so $S \subseteq R$. Let P be a partition of R with subrectangles $\{R_i\}_{i\in I}$, so this chops up S into many pieces. You can two estimates for the area of S. The total area of subrectangles lying inside S gives an *underestimate*. The total area of subrectangles intersecting S somewhere gives an *overestimate*. These regions are illustrated below.





More formally, the expressions

$$\sum_{\substack{i\in I \\ R_i \subseteq S}} \text{area}(R_i) \quad \text{and} \quad \sum_{\substack{i\in I \\ R_i \cap S \neq \varnothing}} \text{area}(R_i) \tag{7.5.1}$$

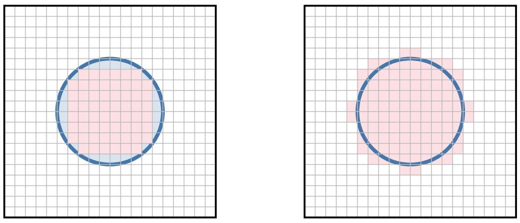
are respectively the underestimate and overestimate for the area of S. Now, here is the critical observation: if $\chi_S : \mathbb{R}^2 \to \mathbb{R}$ is the indicator function of S, then

$$m_i := \inf_{x\in R_i} \chi_S(x) = \begin{cases} 1 & \text{if } R_i \subseteq S, \\ 0 & \text{otherwise,} \end{cases} \qquad M_i := \sup_{x\in R_i} \chi_S(x) = \begin{cases} 1 & \text{if } R_i \cap S \neq \varnothing, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the expressions in (7.5.1) are precisely the lower sum $L_P(\chi_S)$ and upper sum $U_P(\chi_S)$ of the indicator function!

Refining the partition P should improve these estimates.

[10] The Greek letter χ, pronounced *chi*, is used because χ_S is also called the characteristic function of S.



This suggests that the area of S will be defined if and only if $\sup_P L_P(\chi_S) = \inf_P U_P(\chi_S)$. Equivalently, the area of S should be defined when the integral

$$\int_R \chi_S dV$$

exists and its area should be given by the value of this integral.

You can observe this same idea from another less formal perspective. Heuristically speaking, every point $(x, y) \in R$ contributes 0 to the integral if $(x, y) \notin S$ and contributes 1 if $(x, y) \in S$. Integrating over R is equivalent to totalling all of the contribution, so the integral should presumably represent the area of S in $\mathbb{R}^2$.

7.5.1 Volume

This prototypical example inspires a definition of volume.

Definition 7.5.3 (Volume) A set $S \subseteq \mathbb{R}^n$ is **Jordan measurable** if there exists a rectangle R in $\mathbb{R}^n$ such that $S \subseteq R$ and the indicator function χ_S is integrable on R. If so, the **volume of** S (or **Jordan measure of** S) is defined as

$$\mathrm{vol}(S) = \int_R \chi_S dV.$$

Remark 7.5.4 This extends the basic idea that the interval $[a, b] \subseteq \mathbb{R}$ satisfies

$$\mathrm{length}([a, b]) = b - a = \int_a^b 1 dx = \int_{[a,b]} 1 dV.$$

Also, for $\mathbb{R}^2$, volume can be referred to as the **area** of S and denoted area(S).

The phrase "measurable" is used because you can measure the volume of these sets. The adjective "Jordan" comes from the French mathematician Camille Jordan who defined volume in an equivalent manner[11].

Now, there is a serious issue in the definition of volume. Does the value of the integral depend on the rectangle R? The expression vol(S) seems to suggest it should not depend on the rectangle. Mathematicians equivalently ask:

> *Is the volume* ***well-defined****?*

Indeed it is, but you must formally prove that the volume does not depend on the choice of rectangle in the integral. As a first fundamental special case, you can simultaneously verify that rectangles are Jordan measurable and the volume of a rectangle is well-defined.

[11] Our definition does not match the usual notion of volume introduced by Jordan in the 1890s. His notion was referred to as the *Jordan measure* and it is equivalent to the Darboux integral definition here. The definition for volume has a long and beautiful history, and Jordan made major contributions. The story continues with the Lebesgue measure; see a course on real analysis or measure theory for details.

Lemma 7.5.5 (Rectangles are Jordan measurable) If $S = [c_1, d_1] \times \cdots \times [c_n, d_n]$ is a rectangle in $\mathbb{R}^n$, then for any rectangle R containing S,

$$\int_R \chi_S \, dV = (d_1 - c_1) \cdots (d_n - c_n).$$

In particular, S is Jordan measurable and volume is well-defined for rectangles.

Remark 7.5.6 The volume of a rectangle has been defined in two different ways, namely with the righthand formula (Definition 7.1.8) and with the lefthand integral (Definition 7.5.3). This lemma confirms that there is no abuse of notation.

Proof. This is left as an exercise. It is excellent practice with the definition of the integral. ■

This lemma on rectangles and their indicator functions will be a surprisingly useful engine for many proofs. Indeed, it is essential in the proof of the invariance of volume.

Theorem 7.5.7 (Invariance of volume) Let $S \subseteq \mathbb{R}^n$ be a set. Let R and R' be rectangles each containing S. The indicator function χ_S is integrable on R if and only if χ_S is integrable on R'. If so,

$$\int_R \chi_S \, dV = \int_{R'} \chi_S \, dV.$$

Remark 7.5.8 Thus, a set S is Jordan measurable if and only if χ_S is integrable on *every* rectangle containing S. If so, you can compute the volume of S using *any* such rectangle.

Proof. By symmetry, it suffices to prove one direction. Assume χ_S is integrable on R. By definition, there exists a partition P of R with subrectangles $\{R_i\}_{i \in I}$ such that

$$\sum_{i \in I} M_i \operatorname{vol}(R_i) < \int_R \chi_S \, dV + \varepsilon, \qquad \text{where } M_i = \sup_{x \in R_i} \chi_S(x).$$

For each $i \in I$, define $R_i' := R' \cap R_i$ so R_i' is a rectangle or the empty set. Using the convention that $\operatorname{vol}(\varnothing) = 0$, it follows that $\operatorname{vol}(R_i) \geq \operatorname{vol}(R_i')$ so, since $M_i \in \{0, 1\}$ for all $i \in I$,

$$\sum_{i \in I} M_i \operatorname{vol}(R_i') \leq \sum_{i \in I} M_i \operatorname{vol}(R_i).$$

Notice by the invariance of rectangle volume (Lemma 7.5.5) and linearity (Theorem 7.3.13),

$$\sum_{i \in I} M_i \operatorname{vol}(R_i') = \sum_{i \in I} M_i \int_{R'} \chi_{R_i'} \, dV = \int_{R'} \Big(\sum_{i \in I} M_i \chi_{R_i'} \Big) dV.$$

Since $S \subseteq R'$, you can verify that $\chi_S \leq \sum_{i \in I} M_i \chi_{R_i'}$ so, by monotonicity of upper integrals,

$$\overline{I_{R'}}(\chi_S) \leq \overline{I_{R'}}\Big(\sum_{i \in I} M_i \chi_{R_i'} \Big) = \int_{R'} \Big(\sum_{i \in I} M_i \chi_{R_i'} \Big) dV.$$

Combining all of our observations, we have shown that $\overline{I_{R'}}(\chi_S) < \int_R \chi_S \, dV + \varepsilon$. A similar estimate holds for the lower integral so we may conclude that

$$\int_R \chi_S \, dV - \varepsilon < \underline{I_{R'}}(\chi_S) \leq \overline{I_{R'}}(\chi_S) < \int_R \chi_S \, dV + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this proves that $\underline{I_{R'}}(\chi_S) = \overline{I_{R'}}(\chi_S) = \int_R \chi_S \, dV$ as required. ■

Theorem 7.5.7 confirms volume is well-defined. Since the integral does not depend on the rectangle, you may equivalently write

$$\operatorname{vol}(S) = \int \chi_S dV \qquad \text{or} \qquad \operatorname{vol}(S) = \int_S 1 dV$$

and drop the rectangle notation entirely. It is always understood that you integrate over a rectangle containing S.

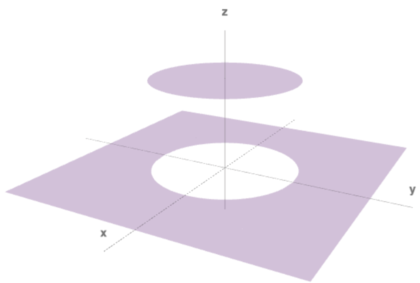
7.5.2 Jordan measurable sets

While the definition of volume is natural, the above example reveals two immediate concerns.

> *How can efficiently decide whether a set $S \subseteq \mathbb{R}^n$ is Jordan measurable? If so, how can you quickly compute its volume?*

The second question will be postponed for several chapters. The first question is more urgent. Unfortunately, you cannot use that continuous functions on rectangles are integrable (Theorem 7.4.8) since χ_S is discontinuous on the boundary ∂S.

Example 7.5.9 Let S be the unit disk in $\mathbb{R}^2$. The graph of its indicator function χ_S is illustrated below as $z = \chi_S(x, y)$ using Math3D.



Notice the discontinuities of χ_S are precisely equal to the boundary of S.

Currently, you can therefore only confirm Jordan measurability by definition. This can be surprisingly hard for non-rectangular sets. A triangle is one of the simplest non-rectangular sets, yet it is also quite cumbersome.

Example 7.5.10 You can show by definition that the triangle

$$S = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq x\}$$

is Jordan measurable and its area is $\frac{1}{2}$. A sketch is provided here.

Proof. (Sketch) The rectangle $R = [0,1]^2$ contains S so, by Theorem 7.5.7, it suffices to prove that χ_S is integrable on R. Let $N \in \mathbb{N}^+$ be arbitrary. Define

$$P_N = \big(\{x_0, \dots, x_N\}, \{y_0, \dots, y_N\}\big)$$

where $\{x_0, \dots, x_N\}$ and $\{y_0, \dots, y_N\}$ are both regular partitions of $[0,1]$. Hence, P_N is a regular partition of R with subrectangles $\{R_{ij} : 1 \leq i, j \leq N\}$ given by

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j] = \Big[\frac{i-1}{N}, \frac{i}{N}\Big] \times \Big[\frac{j-1}{N}, \frac{j}{N}\Big].$$

with $\operatorname{area}(R_{ij}) = N^{-2}$. You can verify that

$$m_{ij} := \inf_{(x,y)\in R_{ij}} \chi_S(x,y) = 1 \iff y_j \le x_{i-1} \iff j \le i-1,$$
$$M_{ij} := \sup_{(x,y)\in R_{ij}} \chi_S(x,y) = 1 \iff y_{j-1} \le x_i \iff j-1 \le i,$$

in which case

$$L_{P_N}(\chi_S) = \sum_{i=1}^{N}\sum_{j=1}^{N} m_{ij}\operatorname{area}(R_{ij}) = \sum_{i=1}^{N}\sum_{j=1}^{i-1}\frac{1}{N^2} = \frac{1}{N^2}\sum_{i=1}^{N}(i-1) = \frac{1}{2} - \frac{1}{N},$$
$$U_{P_N}(\chi_S) = \sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\operatorname{area}(R_{ij}) = \sum_{i=1}^{N}\sum_{j=1}^{i+1}\frac{1}{N^2} = \frac{1}{N^2}\sum_{i=1}^{N}(i+1) = \frac{1}{2} + \frac{1}{N}.$$

Therefore, by taking $N \to \infty$, the inequality

$$\frac{1}{2} - \frac{1}{N} = L_{P_N}(\chi_S) \le \underline{I_R}\chi_S \le \overline{I_R}\chi_S \le U_{P_N}(\chi_S) = \frac{1}{2} + \frac{1}{N},$$

implies that $\underline{I_R}\chi_S = \overline{I_R}\chi_S = \frac{1}{2}$. Hence, χ_S is integrable on R and $\operatorname{area}(S) = \frac{1}{2}$. ■

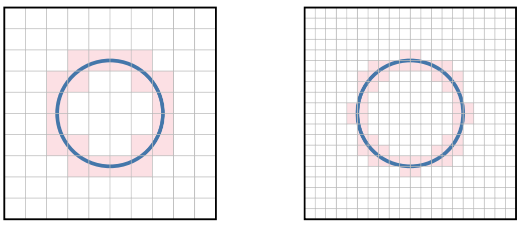
There are also examples of sets which are not Jordan measurable.

Example 7.5.11 The plane $S = \{(x,y,z) \in \mathbb{R}^3 : z = 0\}$ in $\mathbb{R}^3$ is not Jordan measurable since S is not bounded and hence cannot be contained in any rectangle.

Example 7.5.12 The set $S = [0,1] \cap \mathbb{Q}$ is not Jordan measurable in $\mathbb{R}$. For any partition P of $R = [0,1]$, you can confirm that $L_P(\chi_S) = 0$ and $U_P(\chi_S) = 1$ by the density of the rationals and irrationals in $\mathbb{R}$. Thus, $\underline{I_R}(\chi_S) = 0 \neq 1 = \overline{I_R}(\chi_S)$ so χ_S is not integrable on R.

The last two examples demonstrate some limitations of the Jordan measure but these issues are mild for your current goals[12]. The unit disk example will again uncover a pivotal insight.

Example 7.5.13 Let $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ be the unit disk. Let P be a partition of the rectangle $R = [-2,2]^2$ which contains S. You have already seen in Example 7.5.2 how the lower sum $L_P(\chi_S)$ and upper sum $U_P(\chi_S)$ respectively underestimate and overestimate its area. If S is Jordan measurable, then the difference $U_P(\chi_S) - L_P(\chi_S)$ can be made arbitrarily small (Lemma 7.3.10) for a suitably chosen partition P. The diagram below illustrates the difference $U_P(\chi_S) - L_P(\chi_S)$ for two different partitions.



Here is the pivotal insight: *the remaining rectangles cover the boundary!* This suggests that S is Jordan measurable provided its boundary ∂S is Jordan measurable and $\operatorname{vol}(\partial S) = 0$.

You have stumbled upon a deep equivalent definition.

[12] See a course on real analysis with measure theory for a resolution to these drawbacks.

Theorem 7.5.14 A set $S \subseteq \mathbb{R}^n$ is Jordan measurable if and only if S is bounded, its boundary ∂S is Jordan measurable, and $\mathrm{vol}(\partial S) = 0$.

Remark 7.5.15 You can refer to Definition 7.5.3 as the *integral definition* of Jordan measurability, and to Theorem 7.5.14 as the *topological definition* of Jordan measurability. These phrases are not standard, but will be convenient for referencing.

Proof. Its proof is a sophisticated combination of topology and integration, so it is postponed to the last subsection. ∎

This theorem opens new avenues, which you can see by revisiting an example.

Example 7.5.16 Consider again the notorious set $S = [0,1] \cap \mathbb{Q}$ which has boundary $\partial S = [0,1]$. By Lemma 7.5.5, the rectangle $[0,1]$ is Jordan measurable and $\mathrm{vol}(\partial S) = 1$. Thus, by Theorem 7.5.14, the set S is not Jordan measurable.

By the topological definition of Jordan measurability, if a set S is Jordan measurable, then the set of discontinuities for the indicator function χ_S (i.e. the boundary ∂S) has zero volume. This marks significant progress towards your ultimate goal of integrating over non-rectangular sets.

7.5.3 Properties of volume and Jordan measurable sets

There are many natural properties of volume. Identities between indicator functions and properties of integration over rectangles act as stepping stones to prove such facts.

Lemma 7.5.17 (Monotonicity) If S and T are Jordan measurable sets in $\mathbb{R}^n$ and $S \subseteq T$ then $\mathrm{vol}(S) \le \mathrm{vol}(T)$.

Proof. For $x \in \mathbb{R}^n$, if $x \notin S$ then $\chi_S(x) = 0 \le \chi_T(x)$ by definition and if $x \in S$ then $S \subseteq T$ implies $x \in T$, so $\chi_S(x) = 1 \le 1 = \chi_T(x)$. This proves that $\chi_S \le \chi_T$ everywhere. Now, let R be a rectangle containing T and hence containing S. By assumption, S and T are Jordan measurable so χ_S and χ_T are both integrable on R. By monotonicity of the integral (Theorem 7.3.14) and the identity $\chi_S \le \chi_T$, we have that $\mathrm{vol}(S) = \int_R \chi_S dV \le \int_R \chi_T dV = \mathrm{vol}(T)$ as desired. ∎

The topological definition of Jordan measurability (Theorem 7.5.14) and properties of the boundary are also excellent tools.

Lemma 7.5.18 (Union and intersection) If S and T are Jordan measurable sets in $\mathbb{R}^n$, then $S \cup T$ and $S \cap T$ are Jordan measurable sets in $\mathbb{R}^n$ and

$$\mathrm{vol}(S \cup T) = \mathrm{vol}(S) + \mathrm{vol}(T) - \mathrm{vol}(S \cap T),$$

Proof. This is left as an exercise. The Jordan measurability follows from properties of the boundary (Lemma 2.1.21). The volume identity follows from linearity of the integral (Theorem 7.3.13) and an identity involving $\chi_{S\cup T}, \chi_S, \chi_T$, and $\chi_{S\cap T}$. Draw pictures. ∎

Lemma 7.5.19 (Subadditivity) If $S_1, \ldots, S_k$ are Jordan measurable sets in $\mathbb{R}^n$, then their union $S_1 \cup \cdots \cup S_k$ is Jordan measurable and

$$\mathrm{vol}(S_1 \cup \cdots \cup S_k) \le \mathrm{vol}(S_1) + \cdots + \mathrm{vol}(S_k).$$

Proof. This is left as an exercise. Proceed by induction and apply Lemma 7.5.18. ∎

Lemma 7.5.20 If S is a Jordan measurable set in $\mathbb{R}^n$, then its closure $\overline{S}$, its interior S^o, and its boundary ∂S are Jordan measurable with $\operatorname{vol}(S) = \operatorname{vol}(\overline{S}) = \operatorname{vol}(S^o)$ and $\operatorname{vol}(\partial S) = 0$.

Proof. This is left as an exercise. See Section 2.1 for useful lemmas. ■

Overall, you have rigorously defined volume in higher dimensions. You have produced many desirable properties for volume, but are still struggling to verify whether a set is Jordan measurable. For example, you still have not yet proved whether the unit disk in $\mathbb{R}^2$ is Jordan measurable nor computed its area by definition. Your topological definition of Jordan measurability (Theorem 7.5.14) has reduced the problem to a new question.

> *How can you quickly decide whether $\partial S \subseteq \mathbb{R}^n$ is Jordan measurable **and** $\operatorname{vol}(\partial S) = 0$?*
> *In other words, when are the discontinuities of χ_S a set with zero volume?*

This may not seem simpler than before, but you have made significant progress by defining volume. In the next section, you will craft elegant criteria for showing a set has zero volume.

7.5.4 Proof of topological definition of Jordan measurable sets

The proof of Theorem 7.5.14 is a powerful blend of integration and topology, and relies crucially on Lemma 7.5.5. It may take many readings to fully digest all the ideas.

Proof of Theorem 7.5.14. Before proceeding, we make two initial observations. First, S is bounded if and only there exists a rectangle R containing S. Second, for any partition P of R with subrectangles $\{R_i\}_{i \in I}$, notice that for $i \in I$

$$M_i := \sup_{x \in R_i} \chi_S(x) \qquad \text{and} \qquad m_i := \inf_{x \in R_i} \chi_S(x)$$

are either 0 or 1 since the indicator function only takes values of 0 or 1. Moreover, you can verify that exactly one of the following holds:

- $M_i = m_i = 1$ in which case $R_i \subseteq S$
- $M_i = m_i = 0$ in which case $R_i \subseteq \mathbb{R}^n \setminus S = S^c$
- $M_i = 1$ and $m_i = 0$ in which case $R_i \cap S \neq \varnothing$ and $R \cap S^c \neq \varnothing$.

This motivates the definition of the subset of indices

$$J := \{i \in I : R_i \cap S \neq \varnothing \text{ and } R_i \cap S^c \neq \varnothing\} \tag{7.5.2}$$

which satisfies

$$\sum_{j \in J} \operatorname{vol}(R_j) = \sum_{i \in I} (M_i - m_i) \operatorname{vol}(R_i).$$

These observations and notations will be used throughout both directions of proof.

($\Longrightarrow$) Assume $S \subseteq \mathbb{R}^n$ is bounded, ∂S is Jordan measurable, and $\operatorname{vol}(\partial S) = 0$. Since S is bounded, there exists a rectangle such that $S \subseteq R$. By the invariance of volume (Theorem 7.5.7) and the Jordan measurability of ∂S, the indicator $\chi_{\partial S}$ is integrable on R. Fix $\varepsilon > 0$ arbitrary. By Lemma 7.3.10, there exists a partition P of R such that $U_P(\chi_{\partial S}) < \varepsilon$.

Let $\{R_i\}_{i \in I}$ be the subrectangles of P. We claim for any rectangle $T \subseteq \mathbb{R}^n$, if $T \cap S \neq \varnothing$ and $T \cap S^c \neq \varnothing$ then $T \cap \partial S \neq \varnothing$. Assuming the claim, our observations imply that

$$U_P(\chi_S) - L_P(\chi_S) = \sum_{i \in I} (M_i - m_i) \operatorname{vol}(R_i) = \sum_{j \in J} \operatorname{vol}(R_j) \leq \sum_{\substack{i \in I \\ R_i \cap \partial S \neq \varnothing}} \operatorname{vol}(R_i) = U_P(\chi_{\partial S}) < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows by the ε-characterization of integrability (Lemma 7.3.10) that χ_S is integrable on R and hence the set S is Jordan measurable.

It remains to prove the claim. Let $T \subseteq \mathbb{R}^n$ be a rectangle such that $T \cap S \neq \varnothing$ and $T \cap S^c \neq \varnothing$. You can verify that the indicator function $\chi_S : \mathbb{R}^n \to \mathbb{R}$ is discontinuous at x if and only if $x \in \partial S$, so it suffices to show that χ_S is not continuous on T. Since T is a rectangle, it is path-connected. However, by our assumption on T, the image $\chi_S(T) = \{0,1\}$ is not path-connected so by Theorem 2.7.8, χ_S is not continuous on T as required.

($\Longleftarrow$) Assume $S \subseteq \mathbb{R}^n$ is a Jordan measurable set. By definition of Jordan measurability of S, it follows that χ_S is integrable on a rectangle R containing S. Since the rectangle R is bounded and contains S, the set S is bounded. From the invariance of volume (Theorem 7.5.7), we may enlarge the rectangle and assume $\overline{S} \subseteq R^o$ without loss of generality. We wish to show that the indicator of the boundary $\chi_{\partial S}$ is integrable on R and $\mathrm{vol}(\partial S) = 0$.

Fix $\varepsilon > 0$ arbitrary. By Lemma 7.3.10, there exists a partition P of R such that $U_P(\chi_S) - L_P(\chi_S) < \frac{\varepsilon}{2}$. Let $\{R_i\}_{i \in I}$ be the subrectangles of P. We claim that

$$\partial S \subseteq \bigcup_{j \in J} R_j \quad \text{implying} \quad \chi_{\partial S} \leq \sum_{j \in J} \chi_{R_j}.$$

Assuming the claim, each χ_{R_j} is integrable on R and $\int_R \chi_{R_j} dV = \mathrm{vol}(R_j)$ by Lemma 7.5.5. Thus, for each $j \in J$, there exists a partition Q_j of R such that $U_{Q_j}(\chi_{R_j}) < \mathrm{vol}(R_j) + \frac{\varepsilon}{2|J|}$. Let Q be the common refinement of P and Q_j for every $j \in J$. It follows by a property of suprema, the claim, and Lemma 7.1.21 that

$$U_Q(\chi_{\partial S}) \leq \sum_{j \in J} U_Q(\chi_{R_j}) \leq \sum_{j \in J} U_{Q_j}(\chi_{R_j})$$

From our choices of Q_j and P, the above is

$$< \sum_{j \in J} \Big(\mathrm{vol}(R_j) + \frac{\varepsilon}{2|J|} \Big) = \Big(\sum_{i \in I} (M_i - m_i) \mathrm{vol}(R_i) \Big) + \frac{\varepsilon}{2} = U_P(\chi_S) - L_P(\chi_S) + \frac{\varepsilon}{2} < \varepsilon.$$

Since $\chi_S \geq 0$, a property of infima implies that $L_Q(\chi_{\partial S}) \geq 0$ in which case

$$0 \leq L_Q(\chi_{\partial S}) \leq \underline{I_R}(\chi_{\partial S}) \leq \overline{I_R}(\chi_{\partial S}) \leq U_Q(\chi_{\partial S}) < \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this proves that $\underline{I_R}(\chi_{\partial S}) = \overline{I_R}(\chi_{\partial S}) = 0$ so $\chi_{\partial S}$ is integrable on R and $\mathrm{vol}(\partial S) = \int_R \chi_{\partial S} dV = 0$ as desired.

It remains to prove the claim. Recall $\overline{S} \subseteq R^o$ by construction and hence $\partial S \subseteq R^o$. Fix $p \in \partial S$ and proceed by cases depending on whether $p \in S$ or not. Assume $p \in S \cap \partial S$. By the sequential definition of boundary point, there exists a sequence lying inside S^c converging to p. As p is an interior point of R, the tail of this sequence must lie inside R. As $R = \cup_{i \in I} R_i$ and the union is finite, infinitely many terms of this tail sequence must lie inside R_j for a fixed $j \in I$, proving $R_j \cap S^c \neq \varnothing$. This subsequence denoted $\{x(k)\}_k \subseteq R_j \cap S^c$ must converge to p, and $x(k) \neq p$ for any $k \in \mathbb{N}^+$ as $p \in S$ by assumption. By the sequential definition of limit point, the limit point p must belong to R_j as the rectangle R_j is closed. As $p \in S$ by assumption, $R_j \cap S \neq \varnothing$ which proves that $j \in J$. The case $p \in S^c \cap \partial S$ is similar. ■

Exercises for Section 7.5

Concepts and definitions

7.5.1 Let $S \subseteq \mathbb{R}^n$ be a set. Which statements are equivalent to "S is Jordan measurable"?

(a) The indicator function χ_S is integrable on R for some rectangle R containing S.

(b) The indicator function χ_S is integrable on R for every rectangle R containing S.

(c) The volume of S exists.

(d) S is bounded, its boundary ∂S is Jordan measurable, and $\operatorname{vol}(\partial S) = 0$.

(e) S is bounded and the set of discontinuities of χ_S is Jordan measurable with zero volume.

7.5.2 Exhibit an example for each part below. No proof required.

(a) A set $A \subseteq \mathbb{R}^n$ such that A is not Jordan measurable, ∂A is Jordan measurable, and $\operatorname{vol}(\partial A) = 0$.

(b) A set $B \subseteq \mathbb{R}^n$ such that B is not Jordan measurable, and B is bounded.

(c) A set $C \subseteq \mathbb{R}^n$ such that C is not Jordan measurable, and both $\overline{C}$ and C° are Jordan measurable.

(d) A set $D \subseteq \mathbb{R}^n$ such that D is not Jordan measurable, and D is compact.[13]

7.5.3 Let $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ be the unit disk in $\mathbb{R}^2$. Assume D is Jordan measurable. Which of the following quantities are equal to $\operatorname{vol}(D)$?

(a) $\displaystyle\iint_{[-1,1]^2} \chi_D \, dA$

(b) $\displaystyle\iint_{[-1,2]\times[-7,3]} \chi_D \, dA$

(c) $\displaystyle\iint_{[0,1]\times[-1,0]} \chi_D \, dA$

(d) $\displaystyle\int \chi_D \, dV$

(e) $\displaystyle\int_D 1 \, dV$

7.5.4 Let P be a partition of a rectangle R in $\mathbb{R}^n$ with subrectangles $\{R_i\}_{i\in I}$. Let $S \subseteq \mathbb{R}^n$ be a set. There are four quantities listed below.

- One of them is always equal to $\operatorname{vol}(R)$.
- One of them is always equal to the upper sum $U_P(\chi_S)$.
- One of them is always equal to the lower sum $L_P(\chi_S)$.
- One of them is *nearly* equal to the upper sum $U_P(\chi_{\partial S})$.

Identify which quantity corresponds to which property.

(a) $\displaystyle\sum_{i\in I} A_i \operatorname{vol}(R_i)$ where, for $i \in I$, $A_i = 1$ if $R_i \subseteq S$, and $A_i = 0$ otherwise.

(b) $\displaystyle\sum_{i\in I} B_i \operatorname{vol}(R_i)$ where, for $i \in I$, $B_i = 1$ if $R_i \cap S \neq \varnothing$ and $B_i = 0$ otherwise.

(c) $\displaystyle\sum_{i\in I} C_i \operatorname{vol}(R_i)$ where, for $i \in I$, $C_i = 1$ if $R_i \cap S \neq \varnothing$ or $R_i \cap S^c \neq \varnothing$, and $C_i = 0$ otherwise.

(d) $\displaystyle\sum_{i\in I} D_i \operatorname{vol}(R_i)$ where, for $i \in I$, $D_i = 1$ if $R_i \cap S \neq \varnothing$ and $R_i \cap S^c \neq \varnothing$, and $D_i = 0$ otherwise.

[13] This question is intentionally too hard, but think about it for a minute to see why your attempted counterexamples might fail. Afterwards, look at the hint.

7.5.5 The language about volume of rectangles is confusing Bert and Ernie.

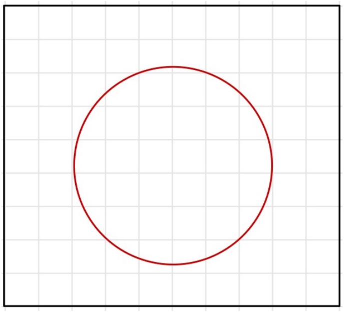
> Bert asserts: *"If $S = [a,b] \times [c,d]$ is a rectangle then $\int_R \chi_S dV = (d-c)(b-a)$ by definition. It's trivial!"*
>
> Ernie disagrees: *"No, there's something to prove and you shouldn't use the word 'trivial'."*

Who is correct? If you agree with Bert, then cite the corresponding definition. If you agree with Ernie, then write the formal statement of what you need to prove.

7.5.6 You can study the following example to gain some better intuition for the integral definition of Jordan measurable sets. Define

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9\}$$

so the disk S lies inside the rectangle R below. Use the following diagram for all parts below.



Let P be a partition of R with subrectangles $\{R_i\}_i$ so $m_i = \inf_{x \in R_i} \chi_S(x)$ and $M_i = \sup_{x \in R_i} \chi_S(x)$.

(a) If $\text{area}(R_i) = 1$ for each subrectangle, then compute $L_P(\chi_S)$.

(b) If $\text{area}(R_i) = 1$ for each subrectangle, then compute $U_P(\chi_S)$.

(c) If $\text{area}(R_i) = 1$ for each subrectangle, then compute $U_P(\chi_S) - L_P(\chi_S)$.

(d) As you refine your partition, what do you expect to occur to $U_P(\chi_S) - L_P(\chi_S)$?

Computations

7.5.7 Indicator functions of sets have helpful identities depending on set properties and operations. You will create some of those identities. Let A and B be sets in $\mathbb{R}^n$.

(a) What is $\chi_\emptyset$ and $\chi_{\mathbb{R}^n}$?

(b) Show that χ_A is bounded.

(c) Relate χ_A and χ_{A^c}.

(d) Relate χ_A and χ_B if $A \subseteq B$.

(e) Relate $\chi_{A \cup B}$, $\chi_{A \cap B}$, χ_A, and χ_B.

7.5.8 Define

$$S = \{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq x^2\}.$$

Show by definition of the integral that χ_S is integrable on $R = [0,1]^2$ and $\text{vol}(S) = 1/3$.

7.5.9 You can prove a set is Jordan measurable by the integral definition, but it is quite challenging. The topological definition of Jordan measurability is much better, but you need one extra ingredient. Here you will get two previews of this upcoming technique.

(a) Let $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ be the unit disk.

> **Sard's lemma.** *If $f : \mathbb{R} \to \mathbb{R}^2$ is C^1 then the image $f([a, b])$ is Jordan measurable in $\mathbb{R}^2$ with zero volume.*

Assuming Sard's lemma, show S is Jordan measurable in $\mathbb{R}^2$.

(b) Let $S = \{(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 + z^2 \leq 1\}$ be the unit ball. Conjecture a generalization of Sard's lemma for maps $f : \mathbb{R}^2 \to \mathbb{R}^3$. Assuming your conjecture, show S is Jordan meaurable in $\mathbb{R}^3$.

Proofs

7.5.10 Properties about volume can be proved by several methods. Here you analyze one such property.

> **Lemma A.** *If $S \subseteq \mathbb{R}^n$ is Jordan measurable, then its interior S° is Jordan measurable and* $\text{vol}(S) = \text{vol}(S^\circ)$.

One method of proof follows the integral definition of Jordan measurable.

1. Let R be a rectangle containing S and hence containing S°.   *2. Since S is Jordan measurable, $\chi_S - \chi_{S^\circ}$ is integrable on R and $\int_R (\chi_S - \chi_{S^\circ}) dV = 0$.*   *3. Since $\chi_{S^\circ} = \chi_S - (\chi_S - \chi_{S^\circ})$, it follows that χ_{S° is integrable on R.*   *4. Moreover, $\int_R \chi_{S^\circ} dV = \int_R \chi_S dV - \int_R (\chi_S - \chi_{S^\circ}) dV = \int_R \chi_S dV$.*

This attempted proof is correct but missing many details.

(a) Lines 2 has almost no justification. Add sufficient details. *Hint:* Squeeze $\chi_S - \chi_{S^\circ}$.

(b) Line 3 follows from a property of integrals. Identify that property and give the details.

7.5.11 Properties about volume can also be established using the topological definition of Jordan measurability. You will apply this method to the following property.

> **Lemma B.** *If $S \subseteq \mathbb{R}^n$ is Jordan measurable, then its interior S° is Jordan measurable.*

(a) State three topological properties you will need to prove the above lemma.

- **Property I (Bounded Sets)**
- **Property II (Boundary)**
- **Property III (Zero Volume)**

(b) Assuming these three properties, prove the lemma.

7.5.12 Prove or disprove: any subset of a Jordan measurable set is Jordan measurable.

7.5.13 Let S be a Jordan measurable set in $\mathbb{R}^n$.

(a) Prove that the closure $\overline{S}$ is Jordan measurable.

(b) Prove that $\text{vol}(S) = \text{vol}(\overline{S})$.

7.5.14 Let S and T be Jordan measurable sets in $\mathbb{R}^n$.

(a) Show that $S \cup T$ and $S \cap T$ are Jordan measurable.

(b) Prove that $\text{vol}(S \cup T) = \text{vol}(S) + \text{vol}(T) - \text{vol}(S \cap T)$.

7.5.15 Let $S_1, \ldots, S_k$ be Jordan measurable sets in $\mathbb{R}^n$. Define $S = S_1 \cup \cdots \cup S_k$.

(a) Prove that S is Jordan measurable.

(b) Prove that $\chi_S \leq \chi_{S_1} + \cdots + \chi_{S_k}$.

(c) Prove that $\operatorname{vol}(S) \leq \operatorname{vol}(S_1) + \cdots + \operatorname{vol}(S_k)$.

Applications and beyond

7.5.16 The invariance of volume (Theorem 7.5.7) has a subtle proof with many great ideas, so you will analyze it closely.

1. *By symmetry, it suffices to prove one direction. Assume χ_S is integrable on R.*
2. *By definition, there exists a partition P of R with subrectangles $\{R_i\}_{i\in I}$ such that*

$$\sum_{i\in I} M_i \operatorname{vol}(R_i) < \int_R \chi_S \, dV + \varepsilon, \qquad \text{where } M_i = \sup_{x\in R_i} \chi_S(x).$$

3. *For each $i \in I$, define $R'_i := R' \cap R_i$ so R'_i is a rectangle or the empty set. Using the convention that $\operatorname{vol}(\varnothing) = 0$, it follows that $\operatorname{vol}(R_i) \geq \operatorname{vol}(R'_i)$*
4. *Since $M_i \in \{0,1\}$ for all $i \in I$, $\sum_{i\in I} M_i \operatorname{vol}(R'_i) \leq \sum_{i\in I} M_i \operatorname{vol}(R_i)$.*
5. *Notice by the invariance of rectangle volume (Lemma 7.5.5) and linearity (Theorem 7.3.13),*

$$\sum_{i\in I} M_i \operatorname{vol}(R'_i) = \sum_{i\in I} M_i \int_{R'} \chi_{R'_i} \, dV = \int_{R'} \Big(\sum_{i\in I} M_i \chi_{R'_i}\Big) dV.$$

6. *Since $S \subseteq R'$, you can verify that $\chi_S \leq \sum_{i\in I} M_i \chi_{R'_i}$.*
7. *By monotonicity of upper integrals,*

$$\overline{I_{R'}}(\chi_S) \leq \overline{I_{R'}}\Big(\sum_{i\in I} M_i \chi_{R'_i}\Big) = \int_{R'} \Big(\sum_{i\in I} M_i \chi_{R'_i}\Big) dV.$$

8. *We conclude that $\overline{I_{R'}}(\chi_S) < \int_R \chi_S \, dV + \varepsilon$.*
9. *By a similar argument for lower integrals,*

$$\int_R \chi_S \, dV - \varepsilon < \underline{I_{R'}}(\chi_S) \leq \overline{I_{R'}}(\chi_S) < \int_R \chi_S \, dV + \varepsilon.$$

10. *As $\varepsilon > 0$ was arbitrary, this proves that $\underline{I_{R'}}(\chi_S) = \overline{I_{R'}}(\chi_S) = \int_R \chi_S \, dV$ as required.*

(a) Line 2 says "by definition" but it refers to two definitions. Identify those definitions.

(b) Line 3 introduces rectangles R'_i instead of continuing with R_i. This was necessary for a later line. Identify that line and explain why it was necessary.

(c) Line 7 equates the upper integral to its integral. How is this justified by previous lines?

(d) Prove Line 6 using casework and Lines 1 to 5.

7.5.17 The topological definition of Jordan measurable sets serves as a special case of the ultimate theorem on integration which you shall soon learn. You will extract the big ideas of its proof in one direction.

Theorem C. *If $S \subseteq \mathbb{R}^n$ is bounded, ∂S is Jordan measurable, and* $\operatorname{vol}(\partial S) = 0$ *then its indicator function χ_S is integrable on any rectangle R containing S.*

You will compare an informal outline with a formal proof.

(a) Below is an informal outline.

> A. *The upper and lower sums of χ_S only differ near* ______________________________.
>
> B. *By assumption, this set* ____ .
>
> C. *Hence, the difference between the upper and lower sums can be made arbitrarily small by* ____.

Fill in the blanks for this summary.

(b) Below is a condensed formal proof. There are no errors. The argument is mostly complete aside from one claim and some other details.

> 1. *Assume $S \subseteq \mathbb{R}^n$ is bounded, ∂S is Jordan measurable, and* $\operatorname{vol}(\partial S) = 0$.
> 2. *Since S is bounded, there exists a rectangle such that $S \subseteq R$.*
> 3. *By the invariance of volume and by assumption, the indicator $\chi_{\partial S}$ is integrable on R.*
> 4. *Fix $\varepsilon > 0$ arbitrary. There exists a partition P of R such that $U_P(\chi_{\partial S}) < \varepsilon$.*
> 5. *Let $\{R_i\}_{i\in I}$ be the subrectangles of P. Define*
>
> $$J := \{i \in I : R_i \cap S \neq \varnothing \text{ and } R_i \cap S^c \neq \varnothing\}.$$
>
> 6. *We claim for any rectangle $T \subseteq \mathbb{R}^n$, if $T \cap S \neq \varnothing$ and $T \cap S^c \neq \varnothing$ then $T \cap \partial S \neq \varnothing$.*
> 7. *Assuming the claim, our observations imply that*
>
> $$U_P(\chi_S) - L_P(\chi_S) = \sum_{i\in I}(M_i - m_i)\operatorname{vol}(R_i) = \sum_{j\in J}\operatorname{vol}(R_j) \leq \sum_{\substack{i\in I \\ R_i\cap\partial S\neq\varnothing}} \operatorname{vol}(R_i) = U_P(\chi_{\partial S}) < \varepsilon.$$
>
> 8. *Since $\varepsilon > 0$ was arbitrary, χ_S is integrable on R and hence the set S is Jordan measurable.*

Match each step in the informal outline from (a) to parts of the above formal proof. A step may correspond to one or more line(s), an equality of expressions, or some combination of these. *Hint:* Start with steps B and C first.

7.6. Sets with zero volume

You have constructed a rigorous definition of volume (Definition 7.5.3) and established an equivalent topological formulation for when the volume exists (Theorem 7.5.14). Unfortunately, you do not have any simple method for verifying a set is Jordan measurable beyond its definition. The topological definition of Jordan measurable sets reduces this question to a simpler one.

How can you verify whether a set $S \subseteq \mathbb{R}^n$ is Jordan measurable ***and*** $\operatorname{vol}(S) = 0$*?*

Remember from Example 7.4.14 that this consitutes a core obstacle to integration over non-rectangular sets. You will want to prove that the boundary of many non-rectangular has zero volume and hence the discontinuities are negligible.

In this section, you will unlock two potent methods for proving sets have zero volume. First, you will use rectangle coverings. Second, you will discover Sard's theorem. To discover these methods, you will begin by exploring some basic properties of zero volume sets.

7.6.1 Properties of sets with zero volume

Basic properties about Jordan measurable sets with volume zero are intuitive if you view them from a physical standpoint or if you draw some pictures. Their proofs are often straightforward consequences of properties of Jordan measurable sets and some topology.

Lemma 7.6.1 (Subset) Any subset of a Jordan measurable set with zero volume is also Jordan measurable with zero volume.

Proof. This is left as an exercise. Adapt the proof of monotonicity of Jordan measurable sets (Lemma 7.5.17) but here you cannot assume the subset is Jordan measurable. Use an upper integral to avoid this issue. ■

Lemma 7.6.2 (Union) Any finite union of Jordan measurable sets with zero volume is also Jordan measurable with zero volume.

Proof. This follows by subadditivity of Jordan measurable sets (Lemma 7.5.19). ■

Lemma 7.6.3 (Closure) The closure of a Jordan measurable set with zero volume is also Jordan measurable with zero volume.

Proof. This is left as an exercise. Use the topological definition of Jordan measurable sets (Theorem 7.5.14) and Lemma 7.6.1 for a quick proof. ■

Lemma 7.6.4 (Empty interior) A Jordan measurable set with zero volume has empty interior.

Proof. This is left as an exercise. The empty interior can be established with Theorem 7.6.6. Remember that you can always fit a closed rectangle inside an open ball. ■

Somewhat counterintuitively, an empty interior is not sufficient.

Example 7.6.5 The set $S = [0,1] \cap \mathbb{Q}$ is bounded and has empty interior (see Example 2.1.10). However, its boundary $\partial S = [0,1]$ is a rectangle and $\operatorname{vol}(\partial S) = 1$ by Lemma 7.5.5, so S is not Jordan measurable by Theorem 7.5.14.

While the converse of Lemma 7.6.4 fails, you can still establish a powerful tool for verifying whether a set has zero volume.

7.6.2 Covering with rectangles

The subset and union properties (Lemmas 7.6.1 and 7.6.2) inspire a nice informal idea.

> *A set should be Jordan measurable with zero volume if it can be covered by Jordan measurable sets with zero volume.*

This idea may seem a bit useless because rectangles are the only family of sets which you have proven to be Jordan measurable (Lemma 7.5.5) and their volume is not zero. You can instead adapt this idea to rectangles with a small modification.

> *A set should be Jordan measurable with zero volume if it can be covered by* ***finitely many rectangles*** *with* ***arbitrarily small total*** *volume.*

This heuristic idea can be formalized into a useful equivalent definition.

Theorem 7.6.6 (Rectangular covering) A set $S \subseteq \mathbb{R}^n$ is Jordan measurable with $\operatorname{vol}(S) = 0$ if and only if, for every $\varepsilon > 0$, there exist finitely many rectangles $R_1, \ldots, R_N \subseteq \mathbb{R}^n$ such that

$$S \subseteq \bigcup_{i=1}^{N} R_i \qquad \text{and} \qquad \sum_{i=1}^{N} \operatorname{vol}(R_i) < \varepsilon. \tag{7.6.1}$$

Remark 7.6.7 Although no standard name exists, you can refer to Theorem 7.6.6 as the *rectangular covering* definition of Jordan measurable sets with zero volume. Unlike subrectangles of a partition, these rectangles are permitted to have overlapping interiors. This aspect is a positive feature in many contexts.

The proof depends on the subadditivity of Jordan measurable sets and the proof of Lemma 7.6.1.

Proof. ($\Longrightarrow$) From (7.6.1), S is bounded so there exists a rectangle R containing S. Fix $\varepsilon > 0$. By assumption, there exist rectangles $R_1, \ldots, R_N$ satisfying (7.6.1). For $i \in \{1, \ldots, N\}$, define $R_i' = R_i \cap R$ so R_i' is either a rectangle or empty. Moreover, as $S \subseteq R$,

$$S = S \cap R \subseteq \left(\bigcup_{i=1}^{N} R_i \right) \cap R = \bigcup_{i=1}^{N} (R_i \cap R) = \bigcup_{i=1}^{N} R_i'$$

We claim this implies that $\chi_S \leq \chi_{R_1'} + \cdots + \chi_{R_N'}$.

Indeed, if $x \in S$, then $\chi_S(x) = 1$ and for some $i \in \{1, \ldots, N\}$, $\chi_{R_i'}(x) = 1$ as $S \subseteq \cup_{j=1}^{N} R_j'$. This implies $\chi_S(x) = 1 = \chi_{R_i'}(x) \leq \chi_{R_1'}(x) + \cdots \chi_{R_N'}(x)$ since $\chi_{R_j'}(x) \geq 0$ for all $j \in \{1, \ldots, N\}$. Similarly, if $x \notin S$, then $\chi_S(x) = 0 \leq \chi_{R_1'}(x) + \cdots \chi_{R_N'}(x)$. This proves the claim.

Thus, by monotonicity of upper integrals,

$$\overline{I_R}(\chi_S) \leq \overline{I_R}(\chi_{R_1'} + \cdots + \chi_{R_N'})$$

By Lemma 7.5.5, the indicator functions $\chi_{R_1}, \ldots, \chi_{R_N}$ are all integrable on R and their integrals equal to the corresponding rectangle's volume. By linearity of integration,

$$\overline{I_R}(\chi_{R_1} + \cdots + \chi_{R_N}) = \int_R \left(\chi_{R_1'} + \cdots + \chi_{R_N'} \right) dV = \int_R \chi_{R_1'}\, dV + \cdots + \int_R \chi_{R_N'}\, dV = \sum_{i=1}^{N} \operatorname{vol}(R_i').$$

Since $\chi_S \geq 0$ and $\operatorname{vol}(R_i') \leq \operatorname{vol}(R_i)$ for each $i \in \{1, \ldots, N\}$, it follows by monotonicity of lower integrals and our assumption (7.6.1) that

$$0 \leq \underline{I_R}(\chi_S) \leq \overline{I_R}(\chi_S) \leq \sum_{i=1}^{N} \operatorname{vol}(R_i) < \varepsilon$$

Since $\varepsilon > 0$ was arbitrary, this proves that $\underline{I_R}(\chi_S) = \overline{I_R}(\chi_S) = 0$. Therefore, S is a Jordan measurable with $\mathrm{vol}(S) = \int_R \chi_S\, dV = 0$.

($\Longleftarrow$) This is left as an exercise. Choose a partition P of a rectangle R with a small upper sum $U_P(\chi_S)$. Take a subset of its subrectangles as your choice for $R_1, \ldots, R_N$. ■

Now, the rectangular covering definition has a special advantage over the original definition: the choice of rectangles! They do not need to arise from a partition, and they are permitted to overlap. This grants you substantial flexibility and finally allows you to verify examples of Jordan measurable sets with zero volume.

Example 7.6.8 Let S be a singleton set in $\mathbb{R}$ so $S = \{x\}$ for some $x \in \mathbb{R}$. Fix $\varepsilon > 0$. Set $I = [x - \frac{\varepsilon}{4}, x + \frac{\varepsilon}{4}]$. Then I contains S and $\mathrm{length}(I) = \frac{\varepsilon}{2} < \varepsilon$. Thus, by the rectangular covering definition, S is Jordan measurable with $\mathrm{vol}(S) = 0$.

Example 7.6.9 Any finite subset of $\mathbb{R}^n$ is Jordan measurable with zero volume. The proof is left as an exercise; adapt the approach of the previous argument.

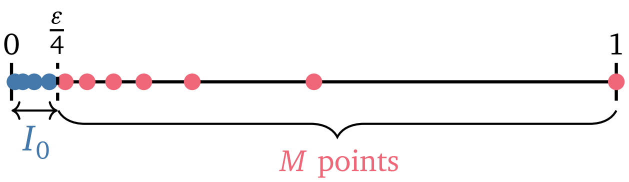
This example hopefully matches your intuition. A single point in $\mathbb{R}$ has zero length. Infinite sets in $\mathbb{R}$ can also have zero volume.

Example 7.6.10 The set

$$S = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\} = \left\{\frac{1}{k} : k \in \mathbb{N}^+\right\}$$

contains an infinite number of points, but you can prove that S is Jordan measurable with zero volume. To prove this claim with a rectangular covering (Theorem 7.6.6), you cannot create an interval around every point in S, because the equivalent definition allows only *finitely* many intervals. Instead, you must fit infinitely many points into a single small interval. Notice that a majority of the points in S clump up near the origin $0 \in \mathbb{R}$ while a finite number of points remain scattered in between 0 and 1, away from the origin. This suggests you can place a small rectangle near $0 \in \mathbb{R}$ to capture the clump, and cover the remaining points each with a small rectangle. Here is a proof.

Proof. Fix $\varepsilon > 0$. Without loss of generality, you may assume $\varepsilon < 4$. Define the interval $I_0 = [0, \frac{\varepsilon}{4}]$. A point in S belongs to I_0 or the interval $[\frac{\varepsilon}{4}, 1]$. There are finitely many points of S in the interval $[\frac{\varepsilon}{4}, 1]$, say $M \in \mathbb{N}^+$ points so $M = |S \cap [\frac{\varepsilon}{4}, 1]|$ by definition[14]. (In fact, you know that $M = \lfloor 4/\varepsilon \rfloor$ but this is not necessary.)



Notice that

$$S \cap \left[\frac{\varepsilon}{4}, 1\right] = \left\{\frac{1}{k} : k \leq \frac{4}{\varepsilon}\right\} = \left\{\frac{1}{k} : 1 \leq k \leq M\right\}.$$

For each of these M remaining points, create an interval around each point. Namely, define the kth interval as

$$I_k = \left[\frac{1}{k} - \frac{\varepsilon}{8M}, \frac{1}{k} + \frac{\varepsilon}{8M}\right] \quad \text{for } k \in \{1, \ldots, M\}.$$

Notice that for every $k \in \{1, \ldots, M\}$, I_k contains the element $\frac{1}{k}$ from S. Thus, these M intervals cover all M points outside of $[0, \frac{\varepsilon}{4}]$. Therefore,

$$S \subseteq \bigcup_{k=0}^{M} I_k$$

and, moreover,

$$\begin{aligned}
\sum_{k=0}^{M} \text{length}(I_k) &= \text{length}\left(\left[0, \frac{\varepsilon}{4}\right]\right) + \sum_{k=1}^{M} \text{length}\left(\left[\frac{1}{k} - \frac{\varepsilon}{8M}, \frac{1}{k} + \frac{\varepsilon}{8M}\right]\right) \\
&= \frac{\varepsilon}{4} + \sum_{k=1}^{M} \frac{\varepsilon}{4M} \\
&= \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon
\end{aligned}$$

Thus, S is Jordan measurable with $\text{vol}(S) = 0$ by the rectangular covering definition. ∎

A few brief remarks on this proof may be helpful. First, notice that it does not matter whether the intervals overlap. Second, the choice of intervals I_k is not unique. It was selected so that you can easily verify the identity $\sum_{k=0}^{M} \text{length}(I_k) = \frac{\varepsilon}{2}$. You have a lot of flexibility when choosing these intervals to satisfy the two conditions in the rectangular covering definition.

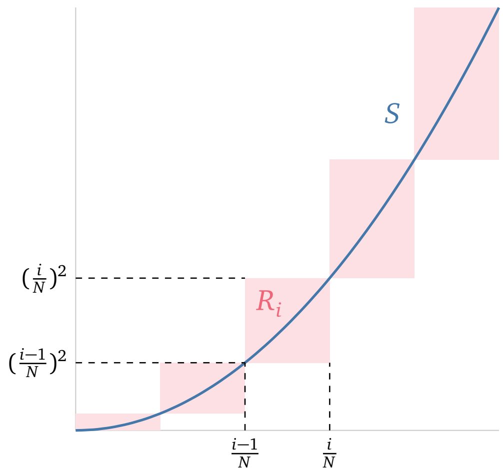
You can apply the same ideas to this higher dimensional setting, but the geometry and notation can become quite a bit more complicated even with the simplest examples.

Example 7.6.11 The set $S = \{(x, x^2) : x \in [0, 1]\}$ in $\mathbb{R}^2$ is Jordan measurable with $\text{vol}(S) = 0$.

Proof. Fix $\varepsilon > 0$. Set $N = \lceil \frac{4}{\varepsilon} \rceil$ so $\frac{2}{N} < \varepsilon$. For $i \in \{1, \ldots, N\}$, define the rectangles

$$R_i = \left[\frac{i-1}{N}, \frac{i}{N}\right] \times \left[\left(\frac{i-1}{N}\right)^2, \left(\frac{i}{N}\right)^2\right] \quad \text{so} \quad \text{area}(R_i) = \frac{1}{N} \cdot \frac{2i-1}{N^2} = \frac{2i-1}{N^3}.$$

For $x \in [0, 1]$, there exists $i \in \{1, \ldots, N\}$ such that $\frac{i-1}{N} \leq x \leq \frac{i}{N}$ in which case $(x, x^2) \in R_i$.



[14]Not to be confused with the notation for absolute value, vertical bars around a set denotes cardinality, i.e. the number of elements in a set.

Therefore, S is contained in the union of $R_1, \ldots, R_N$ and

$$\sum_{i=1}^{N} \text{area}(R_i) = \sum_{i=1}^{N} \frac{2i-1}{N^3} \leq \sum_{i=1}^{N} \frac{2N}{N^3} = \frac{2}{N} < \varepsilon$$

by our choice of N. This proves S is Jordan measurable with zero volume. ∎

Now, you have succeeded in formulating another natural definition for zero volume but it still seems unwieldy to use in practice. Despite these intricacies, the rectangular covering definition for sets with zero volume plays a pivotal role in building your newest hammer.

7.6.3 Sard's theorem

One of this section's motivations was to formally establish a heuristic.

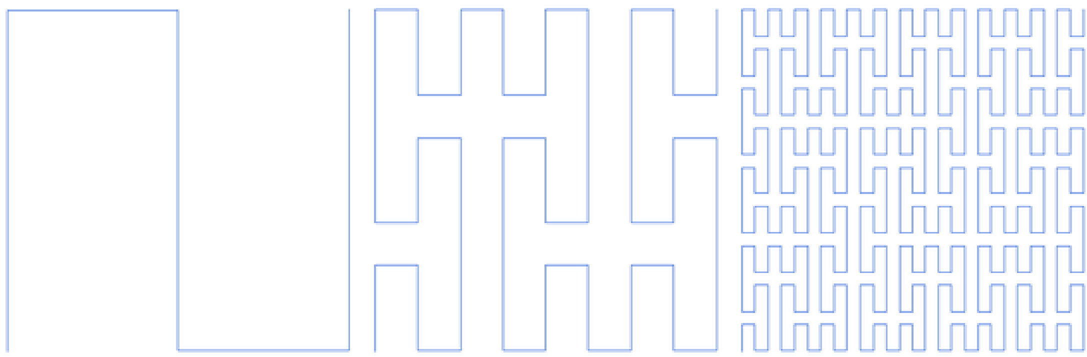
> *Lower dimensional manifolds have zero volume in $\mathbb{R}^n$.*

They should intuitively have zero volume since they are "thin". For instance, Example 7.6.11 shows that a given curve (i.e. 1-dimensional smooth manifold) in $\mathbb{R}^2$ is Jordan measurable with zero volume. This leads to a more formal question.

> *If R is a rectangle in a lower dimensional space $\mathbb{R}^k$ and $f : R \to \mathbb{R}^n$ is a continuous map to higher dimensional space $\mathbb{R}^n$, then does the image $f(R)$ have zero volume?*

Shockingly, the answer is not necessarily!

Example 7.6.12 There exists a *continuous* bijective map $f : [0,1] \to [0,1]^2$ from the unit interval in $\mathbb{R}$ to the unit square in $\mathbb{R}^2$. This was first discovered in 1890 by Giuseppe Peano and such curves are called space filling curves. The construction of this curve f is through a limiting process of other curves; the first few iterations are below.[15]



The proof that the limiting curve fills the unit square is beyond the scope of this text, but you should at least be familiar with the existence of this counterexample. The argument is an intricate and beautiful display of analysis. Amazingly, this process has applications to computer graphics (see this 3Blue1Brown video for details).

This surprising example demonstrates the extreme subtlety of such questions. Luckily, if you assume a little bit more than continuity, then Sard's theorem[16] shows many parametrized sets have zero volume.

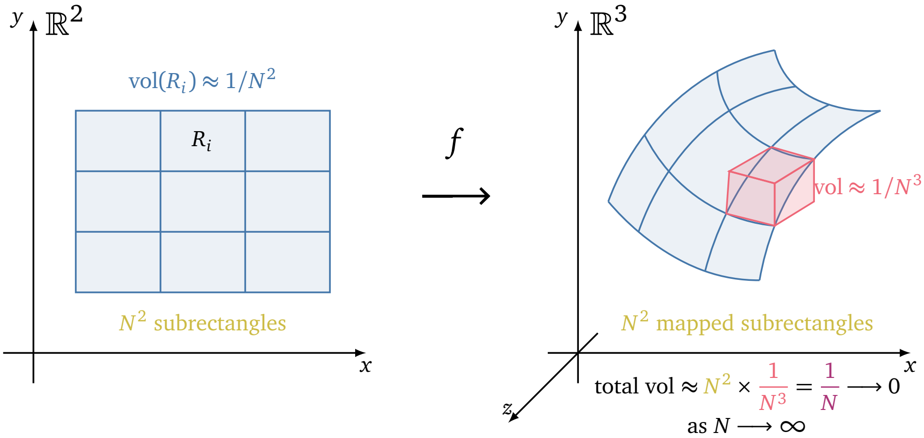
[15] Image retrieved from Wikimedia Commons on 2024-07-22 licensed under PD.
[16] The traditional statement of Sard's theorem (or Sard's lemma) is much deeper, which you can learn in an advanced course on differential geometry or multivariable analysis. Theorem 7.6.13 is a simplified version.

Theorem 7.6.13 (Sard) Fix $k, n \in \mathbb{N}^+$ with $k < n$. Let R be a rectangle in $\mathbb{R}^k$. If f is a $\mathbb{R}^n$-valued function that is C^1 on an open set containing R, then its image

$$f(R) = \{f(x) : x \in R\}$$

is Jordan measurable with zero volume in $\mathbb{R}^n$.

A picture proof of Sard's theorem with the case $k = 2$ and $n = 3$ is illustrated below.



The basic idea is to apply the rectangular covering definition (Theorem 7.6.6). First, you can partition the k-dimensional rectangle R into N^k subrectangles $\{R_i\}_i$ in $\mathbb{R}^k$ and map these subrectangles under f. The mean value theorem controls the volume of these mapped subrectangles in $\mathbb{R}^n$ so that the parametrized set can be covered with N^k rectangles in $\mathbb{R}^n$ of volume $\approx N^{-n}$. The total volume is $\approx N^{k-n}$ which is small as $k < n$.

The formal proof masterfully combines these ideas with the proof that continuous functions on rectangles are integrable (Theorem 7.4.11).

Proof. Write $f = (f^1, \dots, f^n)$ in components, so $f^1, \dots, f^n$ are real-valued functions that are C^1 on an open set containing R. You can prove by the multivariable mean value theorem that there exists $B > 0$ such that for all $j \in \{1, \dots, n\}$,

$$\forall x, y \in R, |f^j(x) - f^j(y)| < B||x - y||. \tag{7.6.2}$$

Fix $0 < \varepsilon < 1$. Choose $N \in \mathbb{N}^+$ as

$$N = \left\lceil \frac{1}{\varepsilon}\Big(B \sum_{\ell=1}^{k} (b_\ell - a_\ell)\Big)^{\frac{n}{n-k}} \right\rceil + 1.$$

Let P_N be the regular partition of the rectangle $R = [a_1, b_1] \times \cdots \times [a_k, b_k]$ constructed from regular partitions of each interval $[a_i, b_i]$ into N subintervals. Let $\{R_i\}_{i \in I}$ be the subrectangles of P_N, so their number is equal to $|I| = N^k$ by construction.

Temporarily fix $i \in I$ and $j \in \{1, \dots, n\}$. Since f is continuous on R, its jth component f^j is continuous on the compact subset R_i. By the extreme value theorem, there exists $c_i^j, C_i^j \in R_i$ such that

$$m_i^j := \inf_{x \in R_i} f^j(x) = f^j(c_i^j), \qquad M_i := \sup_{x \in R_i} f^j(x) = f^j(C_i^j)$$

For any $x, y \in R_i$, it follows by the triangle inequality that

$$||x - y|| \le \sum_{\ell=1}^{n} |x_\ell - y_\ell| \le \sum_{\ell=1}^{n} \frac{b_\ell - a_\ell}{N}.$$

Therefore, by (7.6.2),

$$M_i^j - m_i^j = f^j(C_i^j) - f^j(c_i^j) = |f^j(C_i^j) - f^j(c_i^j)| < \frac{B}{N}\sum_{\ell=1}^{k}(b_\ell - a_\ell).$$

Now, for each index $i \in I$, define the n-dimensional rectangle $S_i \subseteq \mathbb{R}^n$ by

$$S_i = [m_i^1, M_i^1] \times \cdots \times [m_i^n, M_i^n].$$

Fix $x \in R$, so $x \in R_i$ for some $i \in I$. Note $f(x) = (f^1(x), \cdots, f^n(x)) \in f(R_i)$ and, by definition, $m_i^j \le f^j(x) \le M_i^j$ for each $j \in \{1, \ldots, n\}$. This implies that $f(x) \in S_i$ and thus,

$$f(R) \subseteq \bigcup_{i \in I} S_i.$$

For each $i \in I$,

$$\mathrm{vol}(S_i) = (M_i^1 - m_i^1)\cdots(M_i^n - m_i^n) \le \Big(\frac{B}{N}\sum_{\ell=1}^{k}(b_\ell - a_\ell)\Big)^n.$$

As $|I| = N^k$, our choice of N therefore implies that

$$\sum_{i \in I}\mathrm{vol}(S_i) \le \sum_{i \in I}\Big(\frac{B}{N}\sum_{\ell=1}^{k}(b_\ell - a_\ell)\Big)^n = \frac{1}{N^{n-k}}\Big(B\sum_{\ell=1}^{n}(b_\ell - a_\ell)\Big)^n < \varepsilon^{n-k} \le \varepsilon.$$

By the rectangular covering definition (Theorem 7.6.6), the image $f(R)$ is Jordan measurable with zero volume, as desired. ■

Sard's theorem (Theorem 7.6.13) allows you to effortlessly identify many zero volume sets.

Example 7.6.14 The set $S = \{(x, x^2) : x \in [0,1]\}$ in $\mathbb{R}^2$ from Example 7.6.11 can be written as $S = f([0,1])$ for the C^1 function $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(x) = (x, x^2)$. By Theorem 7.6.13, S is Jordan measurable with zero volume.

Combined with the topological definition of Jordan measurability (Theorem 7.5.14), this means you can swiftly prove many sets are Jordan measurable.

Example 7.6.15 The unit disk $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ in $\mathbb{R}^2$ is bounded and ∂D is the unit circle. You can verify by Sard's theorem that ∂D is Jordan measurable and $\mathrm{vol}(\partial D) = 0$. Thus, by Theorem 7.5.14, D is Jordan measurable.

Overall, you have created a natural description for zero volume sets (Theorem 7.6.6) using rectangles, can quickly verify many sets have zero volume (Theorem 7.6.13), and can therefore prove many sets are Jordan measurable (Theorem 7.5.14). These are outstanding accomplishments, but you might not be convinced that much has been achieved. You might wonder:

How can you actually compute the volume of a Jordan measurable set?

You could unpack the enormous definition of the integral but that is neither practical nor entertaining. There are slick computational methods but those are postponed to a future chapter. Right now, you must enter the endgame of defining the integral in higher dimensions. You have tackled the obstacle of discontinuities for indicator functions, so you can finally apply these ideas to achieve this chapter's ultimate goal: define the integral of real-valued functions over *non-rectangular* sets.

Exercises for Section 7.6

For convenience and the sake a brevity, we shall say a set has **zero Jordan measure** if the set is Jordan measurable with zero volume.

Concepts and definitions

7.6.1 Which of the following subsets of $\mathbb{R}$ have zero Jordan measure?

(a) $\{2, \pi, 7\}$

(b) $\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \}$

(c) $\mathbb{Z}$

(d) $\mathbb{R}$

(e) $(0,1)$

(f) $[0,1]$

(g) $[0,1] \cap \mathbb{Q}$

(h) $\partial[0,1]$

7.6.2 Which of the following subsets S of $\mathbb{R}^n$ have zero Jordan measure?

(a) A finite set S of $\mathbb{R}^n$

(b) $S = B_r(a)$ for $r > 0$ and $a \in \mathbb{R}^n$

(c) $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ in $\mathbb{R}^2$

(d) $S = \{(t, t^2) : t \in [0,1]\}$ in $\mathbb{R}^2$

(e) $S = \{(x,y,z) : z = 2x + 3y\}$ in $\mathbb{R}^3$

(f) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 - y^2 - z^2 = 1\}$ in $\mathbb{R}^3$

(g) $S = [0,1]^n \cap \mathbb{Q}^n$ in $\mathbb{R}^n$

(h) $S = [2,3] \times \{7\}$ in $\mathbb{R}^2$

(i) $S = ([2,3] \cap \mathbb{Q}) \times \{7\}$ in $\mathbb{R}^2$

(j) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = 1, 0 \le z \le 3\}$

7.6.3 Which of the following sets are Jordan measurable?

(a) $S = [-3,4] \times (-1,2)$

(b) $S = B_r(a)$ for $r > 0$ and $a \in \mathbb{R}^n$

(c) $S = \overline{B_r(a)}$ for $r > 0$ and $a \in \mathbb{R}^n$

(d) $S = \mathbb{R}^n$

(e) A finite set S in $\mathbb{R}^n$.

(f) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \le 1, 0 \le z \le 3\}$

(g) $S = \{(x,y) \in \mathbb{R}^2 : x^2 - 1 \le y \le 1 - x^2\}$

(h) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 - y^2 - z^2 \le 1\}$

(i) $S = ([0,1] \cap \mathbb{Q})^n$

(j) $S = ([2,3] \cap \mathbb{Q}) \times \{7\}$ in $\mathbb{R}^2$

7.6.4 Let $S \subseteq \mathbb{R}^n$ be a set in $\mathbb{R}^n$. Which of the following are true or false? If true, provide a brief justification. If false, provide a counterexample.

(a) If S has zero Jordan measure, then $\overline{S}$ has zero Jordan measure.

(b) If $\overline{S}$ has zero Jordan measure, then S has zero Jordan measure.

(c) If S has zero Jordan measure, then ∂S has zero Jordan measure.

(d) If ∂S has zero Jordan measure, then S has zero Jordan measure.

(e) If S has zero Jordan measure, then S° is empty and S is bounded.

(f) If S° is empty and S is bounded, then S has zero Jordan measure.

Proofs

7.6.5 Sard's theorem is a powerful tool for verifying parametrized sets have zero Jordan measure. It therefore allows you to prove that sets are Jordan measurable.

(a) Show that the unit circle in $\mathbb{R}^2$ has zero Jordan measure.

(b) Conclude that the unit disk in $\mathbb{R}^2$ is Jordan measurable.

(c) Show that the unit sphere in $\mathbb{R}^3$ has zero Jordan measure.

(d) Conclude that the unit ball in $\mathbb{R}^3$ is Jordan measurable.

7.6.6 Show that the cylinder of radius R and height H has zero Jordan measure.

7.6.7 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a C^1 function. Show that the set $S = \{(x, f(x)) : x \in (0,1)^n\}$ has zero Jordan measure in $\mathbb{R}^{n+m}$.

7.6.8 Define $S = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 1, x^2 \le y < x\}$. Prove that S is Jordan measurable.

7.6.9 Define $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \le 1, 0 \le z \le 3\}$. Prove that S is Jordan measurable.

7.6.10 A picture proof can help discover how to show a set has zero Jordan measure using the rectangular covering theorem.

(a) Draw a picture proof that illustrates $S = \{1, \frac{1}{2}, \frac{1}{3}\}$ has zero Jordan measure in $\mathbb{R}$.

(b) Write a formal proof that $S = \{1, \frac{1}{2}, \frac{1}{3}\}$ has zero Jordan measure in $\mathbb{R}$.

7.6.11 Define the set $S = \{(x,y,z) \in \mathbb{R}^3 : 0 \le x \le 1, 0 \le y \le 1, z = 0\}$.

(a) Prove by the rectangular covering theorem that S has zero Jordan measure.

(b) Prove by Sard's theorem that S has zero Jordan measure.

7.6.12 Define the set $S = \{(x,y,z) \in \mathbb{R}^3 : 0 \le x \le 1, 0 \le y \le 2, z = 2x + 3y\}$ in $\mathbb{R}^3$.

(a) Prove by the rectangular covering theorem that S has zero Jordan measure.
Hint: Cover S with N^2 rectangles in $\mathbb{R}^3$ each of volume $\approx \frac{1}{N^3}$. Sketch a picture to help.

(b) Prove by Sard's theorem that S has zero Jordan measure.

7.6.13 Let $S, T \subseteq \mathbb{R}^n$. Use the rectangular covering theorem to prove that if $S \subseteq T$ and T has zero Jordan measure, then S has zero Jordan measure.

7.6.14 Let $S, T \subseteq \mathbb{R}^n$. Use the rectangular covering theorem to prove that if S and T has zero Jordan measure, then $S \cup T$ has zero Jordan measure.

7.6.15 Disprove that any infinite union of zero volume sets has zero volume.

7.6.16 Let $S \subseteq \mathbb{R}^n$. Prove that if S has zero Jordan measure, then $\overline{S}$ has zero Jordan measure.

7.6.17 Let $S \subseteq \mathbb{R}^n$. Prove that if S has zero Jordan measure, then S° is empty.

Applications and beyond

7.6.18 The proof of Sard's theorem blends together some nice ideas. You will discover those ideas by analyzing an incomplete proof that the quartercircle

$$S = \left\{(\cos x, \sin x) : 0 \le x \le \frac{\pi}{2}\right\}$$

in $\mathbb{R}^2$ is Jordan measurable with zero volume. The argument has no errors but is missing details.

1. *Fix* $\varepsilon > 0$. *Set* $N = \left\lceil \dfrac{8}{\pi^2 \varepsilon} \right\rceil$ *so* $\dfrac{\pi^2}{4N} < \varepsilon$.
2. *For* $i \in \{0, 1, \ldots, N\}$, *define* $\theta_i = \dfrac{\pi i}{2N}$ *so* $\{\theta_0, \theta_1, \ldots, \theta_N\}$ *is a regular partition of* $\left[0, \dfrac{\pi}{2}\right]$
3. *For* $i \in \{1, \ldots, N\}$, *define the rectangles*
$$R_i = \Big[\cos(\theta_i), \cos(\theta_{i-1})\Big] \times \Big[\sin(\theta_{i-1}), \sin(\theta_i)\Big].$$
4. *By the mean value theorem,*
$$\text{area}(R_i) \leq \frac{\pi^2}{4N^2}.$$
5. *For any* $\theta \in [0, \frac{\pi}{2}]$, *there exists* $i \in \{1, \ldots, N\}$ *such that* $\theta_{i-1} \leq \theta \leq \theta_i$ *so* $(\cos\theta, \sin\theta) \in R_i$.
6. *Therefore,* S *is contained in the union of* $R_1, \ldots, R_N$.
7. *Moreoer,*
$$\sum_{i=1}^{N} \text{area}(R_i) \leq \sum_{i=1}^{N} \frac{\pi^2}{4N^2} = \frac{\pi^2}{4N} < \varepsilon$$
by our choice of N.
8. *By the rectangular covering theorem,* S *is Jordan measurable with zero volume.*

(a) Draw a picture proof illustrating this formal proof. Label your diagram using the same notation.

(b) Line 3 did not justify the definition of R_i. Add that justification.

(c) Line 4 did not explain how it used the mean value theorem. Add those details.

(d) Line 5 does not justify why $(\cos\theta, \sin\theta) \in R_i$. Add those justifications.

7.7. Integration over non-rectangles

The primary objective of this chapter has been to rigorously define integration of real-valued functions over bounded sets in $\mathbb{R}^n$. You have chopped, estimated, and refined your way here, so you can finally accomplish this monumental achievement and resolve a fundamental question.

What is the total value of a real-valued function f over a region S in $\mathbb{R}^n$?

Jordan measurable sets were the last piece required to handle the boundaries of non-rectangular sets. In this section, you will define the integral in its final form, establish an elegant criterion for integrability, and formulate standard properties. This sounds like a lot but you are highly prepared from your earlier efforts.

7.7.1 Definition of the integral

Recall χ_S is the indicator function of S. The function $\chi_S f : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\chi_S f(x) = \begin{cases} f(x) & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

You can say that $\chi_S f$ "extends f by zero" because even if f is not defined outside of S, this function is defined everywhere in $\mathbb{R}^n$. Inspired by Example 7.4.14 and its clever trick, you can define the integral in its final form.

Definition 7.7.1 Let $S \subseteq \mathbb{R}^n$ be a bounded set. Let f be a bounded real-valued function on S. The function f is **integrable on** S if the function $\chi_S f : \mathbb{R}^n \to \mathbb{R}$ is integrable for some rectangle R containing S. If so, the **integral of f over** S is defined as

$$\int_S f\, dV := \int_R \chi_S f\, dV.$$

Remark 7.7.2 This definition may be confusing since it seems to define an integral with an integral. However, the integral over the set S is simply a new piece of notation whereas the integral over the rectangle R is independently defined with partitions, upper and lower sums, and upper and lower integrals. There is nothing circular about it. The same notation is used for convenience and it does not cause any conflict.

Again, similar to volume, you may ask:

Is the integral well defined? Does the value of the integral depend on the rectangle R?

It is again not obvious, but the integral does not depend on the choice of rectangle. In other words, it is well-defined.

Theorem 7.7.3 (Invariance of integrals) Let $S \subseteq \mathbb{R}^n$ be a bounded set. Let f be a bounded real-valued function on S. Let R and R' be rectangles each containing S. The function $\chi_S f$ is integrable on R if and only if $\chi_S f$ is integrable on R'. If so,

$$\int_R \chi_S f\, dV = \int_{R'} \chi_S f\, dV.$$

Proof. This is left as an exercise. Follow the same strategy as in the proof of Theorem 7.5.7. ■

7.7.2 Criteria for integrability

You have built a sensible definition of the integral, which opens up a new question.

How can you prove a function is integrable on a non-rectangular set?

Your efforts with the Jordan measure will now pay off in the ultimate theorem which supersedes all of your previous theorems on integration.

Theorem 7.7.4 Let $S \subseteq \mathbb{R}^n$ be a bounded set. Let f be a bounded real-valued function on S. If S is Jordan measurable and the set

$$\{x \in S : f \text{ is not continuous at } x\}$$

is Jordan measurable with volume zero, then f is integrable on S.

Remark 7.7.5 Note Theorem 7.7.4 implies Theorem 7.4.11 and one direction of Theorem 7.5.14. The converse of Theorem 7.7.4 is false; the proof is left as an exercise.

Its proof is postponed to the final subsection. At the moment, you can see how to apply this powerful theorem in some examples.

Example 7.7.6 The function $f(x,y,z) = e^{-xyz}$ is continuous inside the unit ball

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}.$$

The set of discontinuities of f on S is empty and hence is Jordan measurable with zero volume. Notice its boundary

$$\partial S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

is the unit sphere. It can be written as $\partial S = g([0,2\pi] \times [0,\pi])$ where $g : \mathbb{R}^2 \to \mathbb{R}^3$ is the C^1 function defined by

$$g(\theta,\phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi).$$

Thus, by Theorem 7.6.13, ∂S is Jordan measurable with zero volume and, as S is bounded, the set S is Jordan measurable. Therefore, f is integrable on S by Theorem 7.7.4.

Example 7.7.7 The function $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f(x,y) = \begin{cases} 237 & \text{if } x = y, \\ 6 & \text{otherwise.} \end{cases}$$

Notice f is continuous on the rectangle $R = [0,1] \times [0,1]$ except on the set

$$D = \{(x,y) \in R : x = y\}.$$

The set D can be written as $D = g([0,1])$ where $g : \mathbb{R} \to \mathbb{R}^2$ is the C^1 function defined by $g(t) = (t,t)$. By Theorem 7.6.13, D is Jordan measurable with zero volume. As rectangles are Jordan measurable, Theorem 7.7.4 therefore implies f is integrable on R.

Notice in both examples how Sard's theorem (Theorem 7.6.13) is used to quickly verify that a set has zero volume. Theorems 7.6.13 and 7.7.4 form a potent combination for demonstrating integrability. There are also two special situations where you can verify integrability: when the region has zero volume, or when the integrand is almost always zero.

Theorem 7.7.8 Let S be a bounded set in $\mathbb{R}^n$. Let f be a bounded function on S.

(a) If S is a set with zero volume, then f is integrable on S and $\displaystyle\int_S f\,dV = 0$.

(b) If $f = 0$ on S except on a set with zero volume, then f is integrable on S and $\displaystyle\int_S f\,dV = 0$.

Remark 7.7.9 For (a), this holds for *any* bounded function f, no matter how strange. For (b), this holds for *any* bounded set S, even if it is not Jordan measurable.

Proof. Let R be a rectangle containing S. For (a), since S has zero volume, both $\chi_S f$ and χ_S are zero on R except on a set of zero Jordan measure. By Theorem 7.7.4, both $\chi_S f$ and χ_S are integrable on R. Since f is bounded on S, there exists $M > 0$ such that $|f| \le M$ on S. Hence, on the rectangle R,

$$-M\chi_S \le \chi_S f \le M\chi_S.$$

By linearity of the integral (Theorem 7.3.13), both $-M\chi_S$ and $M\chi_S$ are integrable on R. By monotonicity and linearity (Theorems 7.3.13 and 7.3.14), the definition of volume implies that

$$-M\,\mathrm{vol}(S) = -M\int_R \chi_S dV \le \int_R \chi_S f\,dV \le M\int_R \chi_S dV = M\,\mathrm{vol}(S).$$

Since S has zero Jordan measure, $\mathrm{vol}(S) = 0$ and therefore

$$\int_S f\,dV = \int_R \chi_S f\,dV = 0.$$

For (b), this is left as an exercise. It follows quickly from (a). ∎

These criteria for integrability, especially Theorem 7.7.4, will be your standard tools.

7.7.3 Properties of integrals over any set

The properties of the integral over a rectangle transfer in a straightforward fashion to the integral over non-rectangular sets. The proofs all rely on these earlier properties over a rectangle. First, the constant function is integrable on Jordan measurable sets.

Lemma 7.7.10 Let S be a Jordan measurable set in $\mathbb{R}^n$. Fix $\lambda \in \mathbb{R}$. The constant function λ is integrable on S, and

$$\int_S \lambda dV = \lambda\,\mathrm{vol}(S).$$

Proof. This is left as an exercise. Use linearity on rectangles (Theorem 7.3.13). ∎

Second, the integral acts like a linear transformation on functions.

Theorem 7.7.11 (Linearity) Let S be a bounded set in $\mathbb{R}^n$. Let f and g be bounded functions on S. Let $\lambda \in \mathbb{R}$. If f and g are integrable on S and $\lambda \in \mathbb{R}$, then $f + \lambda g$ is integrable on S and

$$\int_S (f + \lambda g)dV = \int_S f\,dV + \lambda\int_S g\,dV.$$

Proof. This is left as a short exercise. Use linearity on rectangles (Theorem 7.3.13). ∎

Third, the integral is monotone.

Theorem 7.7.12 (Monotonicity) Let S be a bounded set in $\mathbb{R}^n$. Let f and g be bounded functions on S. If f and g are integrable on S and $f \leq g$ on S, then

$$\int_S f\,dV \leq \int_S g\,dV.$$

Proof. This is left as a short exercise. Use monotonicity on rectangles (Theorem 7.3.14). ■

Fourth, integrability implies absolute integrability[17] on bounded sets.

Theorem 7.7.13 (Triangle inequality) Let S be a bounded set in $\mathbb{R}^n$. Let f be a bounded function on S. If f is integrable on S, then $|f|$ is integrable on S and

$$\left|\int_S f\,dV\right| \leq \int_S |f|\,dV.$$

Proof. This is left as a short exercise. Use the triangle inequality (Theorem 7.3.15). ■

Fifth, the product of integrable functions is integrable by Cauchy-Schwarz.

Theorem 7.7.14 (Cauchy-Schwarz) Let S be a bounded set in $\mathbb{R}^n$. Let f and g be bounded functions on S. If f and g are integrable on S, then their product fg is integrable on S and

$$\int_S fg\,dV \leq \left(\int_S f^2\,dV\right)^{1/2}\left(\int_S g^2\,dV\right)^{1/2}.$$

Proof. This is left as a short exercise. Use Cauchy-Schwarz (Theorem 7.3.16) on rectangles. ■

Sixth, the integral is additive; the domain can be divided into two sets which are nearly disjoint.

Theorem 7.7.15 (Additivity) Let S be a bounded set in $\mathbb{R}^n$. Let f be a bounded function on S. Suppose $S = S' \cup S''$ is a union of two sets S' and S'' such that $S' \cap S''$ has zero Jordan measure. If f is integrable on both S' and S'', then f is integrable on S and

$$\int_S f\,dV = \int_{S'} f\,dV + \int_{S''} f\,dV.$$

Proof. This is left as an exercise. Use an identity between four indicator functions, linearity on rectangles, and Theorem 7.7.8. ■

Finally, Theorem 7.7.8 introduces a brand new property of integrals.

Theorem 7.7.16 Let S be a Jordan measurable set of $\mathbb{R}^n$. Let f and g be bounded functions on S. Assume $f = g$ on S except on a set of zero volume. then f is integrable on S if and only if g is integrable on S. If so,

$$\int_S f\,dV = \int_S g\,dV.$$

Proof. This is left as an exercise. Use Theorems 7.7.8 and 7.7.11 ■

[17]This implication may appear alarming when you consider single-variable *improper* integrals. However, in this case, the functions are bounded and the domain is bounded, so there is no such concern.

The informal statement of this theorem is captured by the motto below.

An integral does not change value if it is modified on a set of zero volume.

This brief summary captures all the wonderful properties of integrals.

7.7.4 Integration on a rectangle with few discontinuities

The proof of Theorem 7.7.4 can be reduced to the special case of a rectangle.

Theorem 7.7.17 Let R be a rectangle in $\mathbb{R}^n$. Let f be a bounded real-valued function on R. If the set

$$\{x \in R : f \text{ is not continuous at } x\}$$

is Jordan measurable with zero volume, then f is integrable on R.

Proof of Theorem 7.7.4 assuming Theorem 7.7.17. Assume S is a Jordan measurable set of $\mathbb{R}^n$. Let R be a rectangle in $\mathbb{R}^n$ containing S. Let D be the set of discontinuities of f inside S, so

$$D = \{x \in S : f \text{ is discontinuous at } x\}.$$

By assumption, D has zero Jordan measure. The boundary ∂S is the set of discontinuities of χ_S and has zero Jordan measure since, by assumption, S is Jordan measurable. Since $\chi_S f$ is a bounded real-valued function on R, it suffices to prove that the discontinuities of $\chi_S f$ on R are contained in $D \cup \partial S$. Assuming this claim, note $D \cup \partial S$ has zero Jordan measure by Lemma 7.6.2 since both ∂S and D have zero Jordan measure. Theorem 7.7.17 therefore implies $\chi_S f$ is integrable which proves f is integrable on S. It remains to prove the claim.

Let p be a discontinuity of $\chi_S f$. Notice $\chi_S f$ is identically zero outside of $\overline{S} = S^o \cup \partial S$ and hence continuous in that region so it must be that $p \in S^o \cup \partial S$. If $p \in \partial S$ then we are done so suppose $p \in S^o$. The indicator function χ_S is continuous on S^o so, as p is a discontinuity of $\chi_S f$, it must be that f is discontinuous at p. In this case, $p \in D$. Overall, in all cases, $p \in D \cup \partial S$ which establishes the claim. ■

All that remains is to prove Theorem 7.7.17. The ideas are a beautiful blend from the proofs of Theorems 7.4.11 and 7.5.14 but rather technical to execute. A detailed sketch is provided. Constructing a partition with all the desired properties is the key new hurdle.

Proof of Theorem 7.7.17. (Sketch) Fix $\varepsilon > 0$. Let $D \subseteq R$ be the set of discontinuities of f lying inside the rectangle R. Assume D is Jordan measurable with zero volume. Since f is bounded, there exists $B \geq 1$ such that $|f| \leq B$ on the rectangle R. Since $\mathrm{vol}(D) = \int_R \chi_D dV = 0$, by definition of the integral, there exists a partition P of R such that

$$U_P(\chi_D) < \int_R \chi_D dV + \frac{\varepsilon}{4B} = \frac{\varepsilon}{4B}. \tag{7.7.1}$$

Let $\{R_i : i \in I\}$ be the subrectangles of P. The set $X \subseteq R$ defined by

$$X := \bigcup_{\substack{i \in I \\ R_i \subseteq R \setminus D}} R_i.$$

is compact because it is a finite union of rectangles. Moreover, f is continuous on X because $X \subseteq R \setminus D$ and D is the set of discontinuities of f on R. Thus, by Theorem 7.4.8, f is uniformly continuous on X so there exists $\delta > 0$ such that

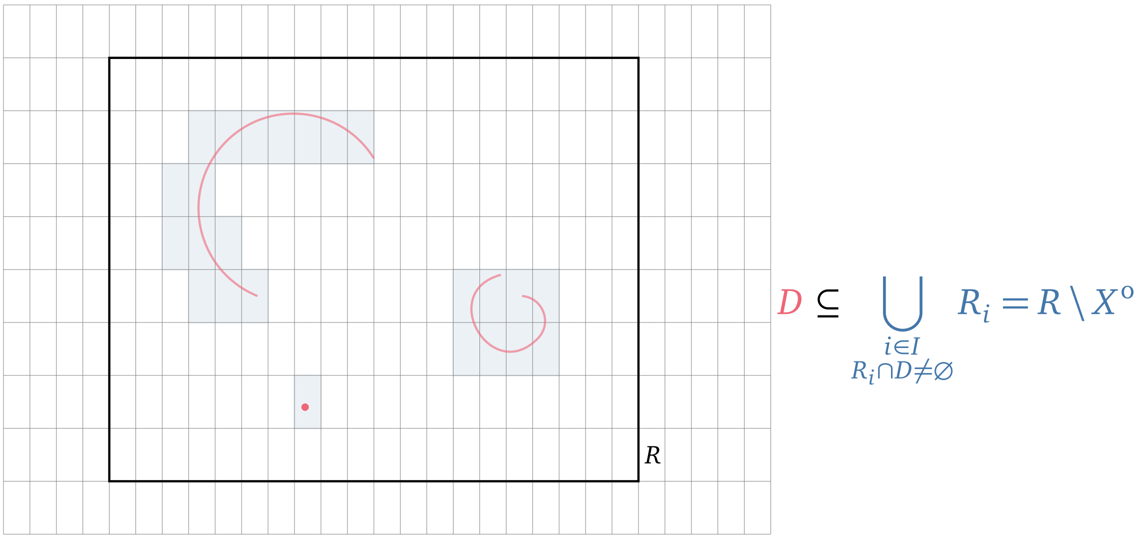
$$\forall x, y \in X, \quad ||x - y|| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{2\,\mathrm{vol}(R)}. \tag{7.7.2}$$

By Lemma 7.1.25, there exists a partition P' of R with norm satisfying $||P'|| < \delta$.

Finally, let P'' be the common refinement of P and P' so $||P''|| \leq ||P'|| < \delta$ by Lemma 7.1.25. Let $\{R''_j : j \in J\}$ be the subrectangles of P''. For each $i \in I$, define the subset

$$J_i = \{j \in J : R''_j \subseteq R_i\}$$

which satisfies the properties outlined in Theorem 7.1.17 as P'' is a refinement of P. An illustration of this partition P'' and the surrounding setup is included below.



For $j \in J$, define $M''_j = \sup_{x \in R''_j} f(x)$ and $m''_j = \inf_{x \in R''_j} f(x)$. Since $M_j - m_j \leq 2B$, you can verify that

$$\sum_{\substack{i \in I \\ R_i \cap D \neq \varnothing}} \sum_{j \in J_i} (M''_j - m''_j) \operatorname{vol}(R''_j) \leq 2B \sum_{\substack{i \in I \\ R_i \cap D \neq \varnothing}} \sum_{j \in J_i} \operatorname{vol}(R''_j) = 2B \sum_{\substack{i \in I \\ R_i \cap D \neq \varnothing}} \operatorname{vol}(R_i) = 2B \cdot U_P(\chi_D) < \frac{\varepsilon}{2}$$

by (7.7.1). Following the strategy in Theorem 7.4.11, you can show the choice of δ in (7.7.2) and Theorems 7.1.10 and 7.1.17 imply that

$$\sum_{\substack{i \in I \\ R_i \subseteq R \setminus D}} \sum_{j \in J_i} (M''_j - m''_j) \operatorname{vol}(R''_j) < \frac{\varepsilon}{2 \operatorname{vol}(R)} \sum_{\substack{i \in I \\ R_i \subseteq R \setminus D}} \sum_{j \in J_i} \operatorname{vol}(R''_j) = \frac{\varepsilon}{2 \operatorname{vol}(R)} \sum_{\substack{i \in I \\ R_i \subseteq R \setminus D}} \operatorname{vol}(R_i) \leq \frac{\varepsilon}{2}.$$

By Lemma 7.3.10, these inequalities imply that f is integrable on R. ■

You have reached the top of the mountain for integration! All the definitions have been laid out in full gory detail. The core theorems allow you to happily verify integrability with relative ease. These achievements are monumental. Indeed, your painstaking efforts have truly produced a rich and deep theory of integration. This journey with integrals, however, might have felt a bit too abstract at times.

What does the integral really represent?

Lucky for you, excellent theory is rewarded with marvelous applications. In the next chapter, you will discover how this robust theory of the integral can be applied to averages, totals, volumes, mass, and probability. These viewpoints and interpretations are vital for many scientific pursuits, and for a deep understanding of what the integral really means.

Exercises for Section 7.7

Concepts and definitions

7.7.1 For each example, determine whether f is integrable on S. Briefly explain why or why not.

(a) $S = B_1(0)$ in $\mathbb{R}^3$ and $f(x,y,z) = \dfrac{1}{(x-2)(y-3)(z-4)}$.

(b) $S = B_1(0)$ in $\mathbb{R}^3$ and $f(x,y,z) = \dfrac{1}{1-x^2-y^2-z^2}$.

(c) $S = [2,3]\times\{7\}$ in $\mathbb{R}^2$ and $f(x,y) = x+2$.

(d) $S = [2,3]\times\{7\}$ in $\mathbb{R}^2$ and $f(x,y) = \begin{cases} -1 & x\in\mathbb{Q} \\ 1 & x\notin\mathbb{Q} \end{cases}$.

(e) $S = ([0,1]\cap\mathbb{Q})^n$ in $\mathbb{R}^n$ and $f(x) = 237$ for all $x\in\mathbb{R}^n$.

7.7.2 Let D be the unit disk in $\mathbb{R}^2$ and let g be integrable on D. Which of these are equal to $\displaystyle\int_D g\,dV$?

$$\int_{[-1,1]^2} g\,dV \qquad \int_{[-1,1]^2} g\chi_D\,dV \qquad \int_{[-1,2]\times[-7,3]} g\chi_D\,dV \qquad \int_{[0,1]\times[-1,0]} g\chi_D\,dA \qquad \int g\chi_D\,dA$$

7.7.3 The big theorem on integration implies your previous results, but its converse is false.

(a) Show that Theorem 7.7.4 implies Theorem 7.4.11.

(b) Show that Theorem 7.7.4 implies one direction of Theorem 7.5.14.

(c) Show that the converse of Theorem 7.7.4 is false.

Proofs

7.7.4 Recall Sard's theorem is one of your main tools to show that a function is integrable on a set. You will often use Theorem 7.7.4 along with Sard's theorem to prove a function is integrable. Define

$$g(x,y) = \begin{cases} \sin(x+y) & y\neq x \\ 12 & y = x \end{cases}$$

and let $S = \{(x,y)\in\mathbb{R}^2 : x^2+y^2\leq 8\}$.

(a) Prove that the set S is Jordan measurable.

(b) Let $D\subseteq\mathbb{R}^2$ be the set of discontinuities of g on S. Show that D has zero Jordan measure.

(c) Conclude that g is integrable on S.

7.7.5 Let $D = \{(x,y)\in\mathbb{R}^2 : x^2+y^2<1\}$. Prove $f = \chi_D$ is integrable on

$$S = \{(x,y)\in\mathbb{R}^2 : (x-1)^2+y^2<1\}.$$

7.7.6 Define the function

$$g(x,y)=\begin{cases}237 & xy\neq 0\\ 12 & x=0 \text{ or } y=0\end{cases}$$

and the set $S=[0,2]\times[-1,3]$.

(a) Prove g is integrable on S using Theorem 7.7.4.

(b) Prove g is integrable on S using Theorem 7.7.17.

(c) Which proof would you prefer and why?

7.7.7 Define $f(x,y)=e^{x^2-y^2}$ and $S=\{(x,y)\in\mathbb{R}^2 : x^2+4y^2\le 1\}$.

(a) Prove f is integrable on S using Theorem 7.7.4.

(b) Prove f is integrable on S using Theorem 7.7.17.

(c) Which proof would you prefer and why?

7.7.8 Theorem 7.7.4 only applies when the set S is Jordan measurable. If S is not Jordan measurable then you do not have many options to verify integrability on S, but there is a special and somewhat silly possibility.

(a) Let $S=[-1,1]^2\cup([2,3]\cap\mathbb{Q})^2$. Briefly explain why S is not Jordan measurable.

(b) Define

$$f(x,y)=\begin{cases}xy & x^2+y^2\le 1\\ 0 & x^2+y^2>1\end{cases}.$$

Prove that f is integrable on S. *Hint:* What are the discontinuities of $\chi_S f$?

7.7.9 Proving properties of the integral over non-rectangular sets relies almost entirely on properties of the integral over rectangular sets. Since most of the hard work is done for rectangles, the proofs over non-rectangles are often quite short. You will prove linearity of integration on non-rectangular sets (Theorem 7.7.11).

(a) Precisely state the theorem for integration on rectangles that you will need in your proof.

(b) Prove the theorem for integration over non-rectangles.

7.7.10 The introduction of zero volume sets has created a new property for integrals (Theorem 7.7.8). This implies that integrals do not change if you modify integrands on a zero volume set (Theorem 7.7.16). Prove Theorem 7.7.16 assuming Theorem 7.7.8. *Hint:* Use linearity.

7.7.11 Let S be a bounded set in $\mathbb{R}^n$. Let f and g be integrable functions on S.

(a) Prove that if $f\le g$ on S, then $\displaystyle\int_S f\,dV\le\int_S g\,dV$.

(b) Prove that if $f\le g$ on S except on a zero volume set, then $\displaystyle\int_S f\,dV\le\int_S g\,dV$.

7.7.12 Let S be a bounded set in $\mathbb{R}^n$. Let f be a bounded function on S. Suppose $S=S'\cup S''$ is a union of two sets S' and S'' such that $S'\cap S''$ has zero Jordan measure. Prove that if f is integrable on both S' and S'', then f is integrable on S and

$$\int_S f\,dV=\int_{S'} f\,dV+\int_{S''} f\,dV.$$

Applications and beyond

7.7.13 The proof of Theorem 7.7.4 was reduced to Theorem 7.7.17, i.e. you proved it by assuming Theorem 7.7.17. Here is a simplified version of this proof needing some details in addition to excluding a claim.

1. *Let R be a rectangle in $\mathbb{R}^n$ containing S.*
2. *Let D be the set of discontinuities of f inside S which has zero Jordan measure by assumption.*
3. *The set of discontinuities of χ_S has zero Jordan measure.*
4. *Since $\chi_S f$ is a bounded real-valued function on R, it suffices to prove that the discontinuities of $\chi_S f$ on R are contained in $D \cup \partial S$. Assume this claim for now.*
5. *Then $D \cup \partial S$ has zero Jordan measure since both ∂S and D have zero Jordan measure.*
6. *Theorem B therefore implies $\chi_S f$ is integrable on R which proves f is integrable on S.*
7. *It remains to prove the claim. [...]*

(a) Line 3 is missing a justification. Identify it.

(b) Line 6 follows from lines 4 and 5 but it uses a property about zero Jordan measure sets. State this property and explain how it is used to conclude that $\chi_S f$ Is integrable on R.

(c) The claim after in line 4 is proved in your readings. Here is an example illustrating the claim.

Let $R = [-4,4] \times [-4,4]$ and let $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9\}$. Let $f(x,y)$ be given by this graph on Math3D and let D be the set of discontinuities of f in the rectangle R. Draw the set of discontinuities of $\chi_S f$ in R and draw $D \cup \partial S$ separately. Compare your two drawings and verify the claim.

7.7.14 A full proof of Theorem 7.7.17 is technical and finicky, so only a sketch was provided. The main ideas behind the proof are inspired by breakthroughs for two prior theorems, namely continuous functions on compact sets are uniformly continuous (Theorem 7.4.8) and the topological definition of Jordan measurability (Theorem 7.5.14). Notice Theorem 7.7.17 actually implies both of these results, so its not so surprising that its proof relies on the same breakthroughs. A condensed proof sketch of Theorem 7.7.17 is below.

1. *Fix $\varepsilon > 0$. Since f is bounded, there exists $B \geq 1$ such that $|f| \leq B$ on R.*
2. *Since the set D of discontinuities has zero volume, there exists a partition P of R with subrectangles $\{R_i\}_{i \in I}$ such that $U_P(\chi_D) < \dfrac{\varepsilon}{4B}$.*
3. *The set $X := \displaystyle\bigcup_{\substack{i \in I \\ R_i \subseteq R \setminus D}} R_i$ is compact because it is a finite union of rectangles.*
4. *Since X is compact and f is continuous on $X \subseteq R \setminus D$, note f is uniformly continuous on X.*
5. *By uniform continuity, there exists $\delta > 0$ such that*
$$\forall x, y \in X, \quad ||x - y|| < \delta \implies |f(x) - f(y)| < \frac{\varepsilon}{2\,\mathrm{vol}(R)}.$$
6. *There exists a partition P' of R with norm satisfying $||P'|| < \delta$, that is, the maximum length inside any of the subrectangles of P' is at most δ.*

7. *Now, let P'' be the common refinement of P and P' so $||P''|| \leq ||P'|| < \delta$.*
8. *Let $\{R''_j : j \in J\}$ be the subrectangles of P'' and, for $i \in I$, define $J_i = \{j \in J : R''_j \subseteq R_i\}$.*
9. *For $j \in J$, define $M''_j = \sup_{x \in R''_j} f(x)$ and $m''_j = \inf_{x \in R''_j} f(x)$.*
10. *Since $M_j - m_j \leq 2B$, you can verify that*

$$\sum_{\substack{i \in I \\ R_i \cap D \neq \varnothing}} \sum_{j \in J_i} (M''_j - m''_j) \operatorname{vol}(R''_j) \leq 2B \sum_{\substack{i \in I \\ R_i \cap D \neq \varnothing}} \sum_{j \in J_i} \operatorname{vol}(R''_j) = 2B \cdot U_P(\chi_D) < \frac{\varepsilon}{2}$$

11. *You can show the choice of δ implies that*

$$\sum_{\substack{i \in I \\ R_i \subseteq R \setminus D}} \sum_{j \in J_i} (M''_j - m''_j) \operatorname{vol}(R''_j) < \frac{\varepsilon}{2 \operatorname{vol}(R)} \sum_{\substack{i \in I \\ R_i \subseteq R \setminus D}} \sum_{j \in J_i} \operatorname{vol}(R''_j) \leq \frac{\varepsilon}{2}.$$

12. *This combines to imply the ε-characterization for integrability, so f is integrable on R.*

Here you will analyze the proof sketch, and try to identify the origins of key steps.

(a) Which lines are inspired by the proof on uniform continuity and compactness?
Hint: There are 4 lines that stand out.

(b) Which lines are inspired by the proof on the topological definition of Jordan measurability?
Hint: There are 3 lines that stand out.

(c) Two lines are critical to successfully merge these proof ideas. Identify those two lines.

(d) Explain how Lines 10 and 11 imply that f is integrable.

8. Integral applications

You have thus far carefully constructed a definition for the multivariable integral and established some powerful theoretical tools. As remarked at the end of the last chapter, your (justified) obsession with theory has left a vital question on integrals unexplored.

> *What does the integral really represent?*

To reveal the nature of the integral, you still need to address some fundamental questions related to the previous chapter.

> *What is the volume of the region lying under a graph $f : S \to [0, \infty)$?*
> *What is the average value of a real-valued function f over a region $S \subseteq \mathbb{R}^n$?*
> *What is the total mass of an object S in $\mathbb{R}^n$ with variable density $f : S \to [0, \infty)$?*

By exploring these applications, you will develop multiple sophisticated viewpoints of the integral and how to interpret it in practical terms. The parallels with single variable integrals will continue, so remember to reflect those similarities. At the end of this chapter, you will also uncover a historically pivotal application of the multivariable integral: probability.

> *How can you interpret probabilities with an integral?*

This problem is quite deep and the solution will wield all of your recently developed theory. Remarkably, this application will also reveal the limitations of your current theory.

Now, these four interpretations of the integral are essential, but they are not exhaustive. You may rediscover the integral in other applied contexts beyond this text. Therefore, a major goal of this chapter is to practice deriving integral formulas as a limit of discrete approximations.

> *How do I build a discrete approximation from natural principles? And how can I derive an integral formula as a limit of discrete approximations?*

Note "derivations" are not necessarily proofs; they mix rigorous theory, natural principles, and heuristic reasoning. You will learn to apply these mixed techniques and, by the end of the chapter, you will have successfully chopped, estimated, and refined to produce wonderful applications of the integral.

8.1. Volume under a graph

The first application of the integral generalizes a classic interpretation from single variable calculus, namely the area under a curve.

> *Let $S \subseteq \mathbb{R}^n$ be a bounded set. What is the volume of the region $T \subseteq \mathbb{R}^{n+1}$ under the graph of $f : S \to [0, \infty)$?*

From Section 7.5, you already have one reasonable guess.

> *Is the region $T \subseteq \mathbb{R}^{n+1}$ Jordan measurable? If so, its volume is* $\operatorname{vol}(T)$.

Remember $\operatorname{vol}(T)$ is defined via an $(n+1)$-dimensional integral. However, based on your intuition from one dimension, you may have another guess.

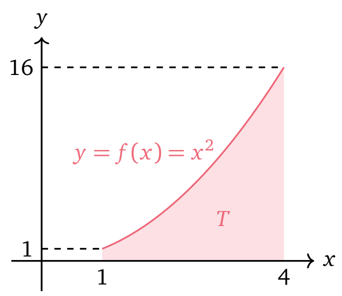
> *Is the n-dimensional integral $\displaystyle\int_S f\,dV$ also the volume of this region $T \subseteq \mathbb{R}^{n+1}$?*

You can better explore this question through a couple of illustrative examples.

Example 8.1.1 If $S = [1,4] \subseteq \mathbb{R}$ and $f(x) = x^2$ then $f \geq 0$ on S and the region under the graph is the set

$$T = \{(x,y) \in \mathbb{R}^2 : 1 \leq x \leq 4, 0 \leq y \leq x^2\} \subseteq \mathbb{R}^2.$$

This set is the usual region in $\mathbb{R}^2$ under the curve $y = x^2$ for $1 \leq x \leq 4$.



You can decompose ∂T into several pieces and use Sard's theorem to prove that T is Jordan measurable. This means that $\operatorname{area}(T)$ is defined. On the other hand, from single-variable calculus, it seems plausible that

$$\iint_T 1 dA = \operatorname{area}(T) = \int_1^4 x^2 dx.$$

Notice this claim is an unproven identity between the two-dimensional lefthand integral and the one-dimensional righthand integral. These are defined independently yet you expect that they are equal.

Of course, you can also use the same ideas to express areas between two graphs. This example with a region in $\mathbb{R}^2$ can be extended to regions in $\mathbb{R}^3$. The geometry may be harder to visualize, but it is the same principle.

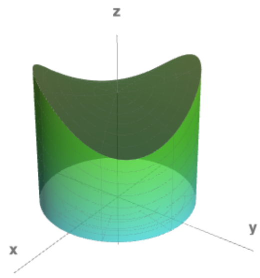
Example 8.1.2 If

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}, \qquad f(x,y) = 2 - \sin(xy),$$

then $f \geq 0$ on S. The solid under the graph of f is the set

$$T = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, 0 \leq z \leq 2 - \sin(xy)\} \subseteq \mathbb{R}^3$$

which is illustrated below via Math3D.



Again, you can verify that $T \subseteq \mathbb{R}^3$ is Jordan measurable by using Sard's theorem and decomposing the boundary into pieces. Hence, $\mathrm{vol}(T)$ is defined. Taking this property for granted, you may heuristically guess that

$$\iiint_T 1\,dV = \mathrm{vol}(T) = \iint_S f\,dA$$

but this requires a formal proof. Notice the lefthand integral is 3-dimensional whereas the righthand integral is 2-dimensional. This equality may be geometrically natural but it is not obviously true.

You can similarly express 3-dimensional volumes between two graphs. In general, for a bounded set $S \subseteq \mathbb{R}^n$, the solid under the graph of $f : S \to [0, \infty)$ is the set $T \subseteq \mathbb{R}^{n+1}$ given by

$$T = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in S, 0 \leq y \leq f(x)\}.$$

This creates your first interpretation of the integral.

Theorem 8.1.3 Let $S \subseteq \mathbb{R}^n$ be a compact Jordan measurable set. Let $f : S \to [0, \infty)$ be a continuous non-negative function on S. The $(n+1)$-dimensional set

$$T = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in S, 0 \leq y \leq f(x)\} \tag{8.1.1}$$

is a compact Jordan measurable set and satisfies

$$\mathrm{vol}(T) = \int_S f\,dV. \tag{8.1.2}$$

Proof. This is left as a very challenging exercise. Try a one-dimensional example first. First, show T is Jordan measurable by writing its boundary ∂T into three pieces; you will need uniform continuity. Second, define a union of rectangles contained in T and a union of rectangles containing T. Refine these sets and squeeze the volumes. ∎

A formal proof of Theorem 8.1.3 would be mathematically satisfying, but you may not always be seeking such a high level of precision and justification. Right now, you will be better served by deriving this integral formula (8.1.2) with an incomplete but convincing argument. This skill will be essential as you encounter more applications where you must construct other such formulas. The basic strategy is the usual one: chop, estimate, and refine.

Example 8.1.4 (Derivation of volume under a graph) Let R be a rectangle in $\mathbb{R}^n$ containing the compact Jordan measurable set S. Assume $f : S \to [0,\infty)$ is continuous on S, so $\chi_S f : \mathbb{R}^n \to [0,\infty)$ is integrable on R. Let P be a partition of R with subrectangles $\{R_i\}_{i\in I}$. For each $i \in I$, choose a sample point $x_i^* \in R_i$. The rectangle

$$R_i \times [0, \chi_S f(x_i^*)] = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in R_i, 0 \le y \le \chi_S f(x_i^*)\}$$

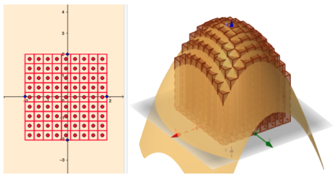
in $\mathbb{R}^{n+1}$ approximates the solid under the graph of f above the subrectangle R_i, that is,

$$T_i = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in R_i, 0 \le y \le \chi_S f(x)\}.$$

By totaling these estimates, it follows that

$$\mathrm{vol}(T) = \sum_{i\in I} \mathrm{vol}(T_i) \approx \sum_{i\in I} \chi_S f(x_i^*)\,\mathrm{vol}(R_i) = S_P^*(\chi_S f).$$

Hence, the Riemann sum is a *discrete* approximation for the volume of T defined by (8.1.1). You can visualize this process with the Geogebra demo below.



You can refine this discrete approximation by taking a sequence of partitions P_N with $||P_N|| \to 0$. Since f is continuous on the compact Jordan measurable set S, it follows that $\chi_S f$ is integrable on R by Theorem 7.7.4. By Theorem 7.3.18, the volume of T should therefore be given by

$$\mathrm{vol}(T) = \lim_{N\to\infty} S_{P_N}^*(\chi_S f) = \int_R \chi_S f\, dV = \int_S f\, dV.$$

This derivation heuristically justifies (8.1.2).

Remark 8.1.5 The above argument is not a formally justified proof. It omits any discussion of Jordan measurability, but it describes the essential ideas for the volume identity (8.1.2). First, you chop up the solid. Second, you discretely approximate the solid with a bunch of rectangles. Third, you refine your discrete approximations to get the formula. Notice that each step could be done differently. You could have chopped with only regular partitions. You could have estimated using upper and lower sums instead of Riemann sums. You could have refined using infima and suprema. Regardless, you should end up with the same formula; the above choices were only made for simplicity. A formal proof of Theorem 8.1.3 would be inspired by the same ideas but it would need to rigorously use upper and lower sums.

This concludes your first application of the integral. The basic idea was hopefully familiar, but the derivation of the integral formula was not so straightforward. In many applications, you will want to construct integral formulas that are valid for many choices of functions and regions. You also will want any sensible choice of chopping, estimating, and refining to lead to the same answer. Luckily, your definitions and theorems are so robust that you can satisfy these requirements and arrive at the same integral formula from many different paths. This trend will continue as you unravel more perspectives of the integral.

Exercises for Section 8.1

Concepts and definitions

8.1.1 Define $f : [-2,2] \to \mathbb{R}$ by $f(x) = x^2$. Assume the set

$$S = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 2, 0 \le y \le f(x)\}.$$

is Jordan measurable without proof. Which of the following integrals are equal to area(S)?

(a) $\displaystyle\iint_S 1 dA$

(b) $\displaystyle\iint_{[0,2]\times[0,4]} \chi_S \, dA$

(c) $\displaystyle\int_S f \, dV$

(d) $\displaystyle\int_0^2 x^2 dx$

(e) $\displaystyle\int_0^4 (2-\sqrt{y}) \, dy$

(f) $\displaystyle\int_{[0,2]} f \, dV$

8.1.2 Let $T \subseteq \mathbb{R}^2$ be the region between $f(x) = x^2$ and $g(x) = x$ along $1 \le x \le 4$.

(a) Express T in set builder notation.

(b) Give expressions for area(T) as a 2-dimensional integral and as a 1-dimensional integral.

8.1.3 Fardin and Alisa are calculating the volume of the solid T in $\mathbb{R}^3$ below the graph of $z = y^2 + 1$ and above the disk $x^2 + y^2 \le 4$ in the $z = 0$ plane. See this Math3D graph for a visual.

(a) Fardin says "*The volume of T is a **three** dimensional integral by definition.*" Write a candidate integral that supports Fardin's claim. Explicitly define any functions or sets.

(b) Alisa says "*That's ridiculous! The volume of T is a **two** dimensional integral by definition!*" Write a candidate integral that supports Alisa's claim. Explicitly define any functions or sets.

(c) Who is correct? Briefly explain.

8.1.4 Let $T \subseteq \mathbb{R}^3$ be the closed ball of radius 4 centered at $(0,0,0)$.

(a) Express T in set builder notation as the region between two graphs.

(b) Give expressions for vol(T) as a 3-dimensional integral and as a 2-dimensional integral.

Proofs

8.1.5 You will analyze a derivation for the integral formula of the volume under a graph. Let R be a rectangle in $\mathbb{R}^n$. Assume $f : R \to [0,\infty)$ is continuous on R. Define

$$T = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in R, 0 \le y \le f(x)\} \subseteq \mathbb{R}^{n+1}.$$

1. *Let P be a partition of R with subrectangles $\{R_i\}_{i\in I}$.*
2. *For each $i \in I$, choose a sample point $x_i^* \in R_i$.*
3. *The solid under the graph of f above the subrectangle R_i is given by*

$$T_i = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : x \in R_i, 0 \le y \le f(x)\}.$$

4. *The rectangle*

$$R_i \times [0, f(x_i^*)] = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : x \in R_i, 0 \le y \le f(x_i^*)\}$$

   in $\mathbb{R}^{n+1}$ approximates T_i.
5. *By totaling these estimates,*

$$\mathrm{vol}(T) = \sum_{i\in I} \mathrm{vol}(T_i) \approx \sum_{i\in I} f(x_i^*)\,\mathrm{vol}(R_i) = S_P^*(f),$$

   so the Riemann sum $S_P^*(f)$ is a discrete approximation for the volume of T.
6. *Refine this discrete approximation by taking a sequence of partitions P_N with $||P_N|| \to 0$.*
7. *Thus, the volume of T is given by*

$$\mathrm{vol}(T) = \lim_{N\to\infty} S_{P_N}^*(f) = \int_R f\,dV.$$

Remember the above is a heuristic justification. It is not a proof.

(a) Line 4 claims the rectangle approximates the solid T_i. Why should you believe this claim? Identify any relevant assumptions and properties.

(b) Line 5 approximates the volume of T. Why should you believe this approximation? Identify any relevant assumptions and properties.

(c) Line 7 has two equalities; one can be quickly justified and the other cannot. Identify and briefly explain.

8.1.6 Let S be a Jordan measurable subset in $\mathbb{R}^n$. Assume $f : S \to [0, \infty)$ is continuous on S. Define

$$T = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : x \in S, 0 \le y \le f(x)\} \subseteq \mathbb{R}^{n+1}.$$

You may assume T is Jordan measurable without proof. Use upper sums and lower sums to derive the integral formula for the volume of T as the volume under the graph, namely

$$\mathrm{vol}(T) = \int_S f\,dV.$$

Hint: Follow the textbook's proof ideas, but you will need two discrete estimates.

Applications and beyond

8.1.7 Let $T = \{(x, y) \in \mathbb{R}^2 : 0 \le y \le -x^2 + 6x\}$.

(a) Prove that T is Jordan measurable.

(b) Express T as a 2-dimensional integral.

(c) Express T as a 1-dimensional integral in two different ways.

8.1.8 Define the solid ellipsoid

$$T = \left\{ (x, y, z) \in \mathbb{R}^3 : x^2 + \frac{y^2}{4} + \frac{z^2}{9} \le 1 \right\}.$$

(a) Prove that T is Jordan measurable.

(b) Express T as a 3-dimensional integral.

(c) Express T as a 2-dimensional integral in three different ways.

8.1.9

(a) Conjecture a version of Theorem 8.1.3 for regions between two graphs.

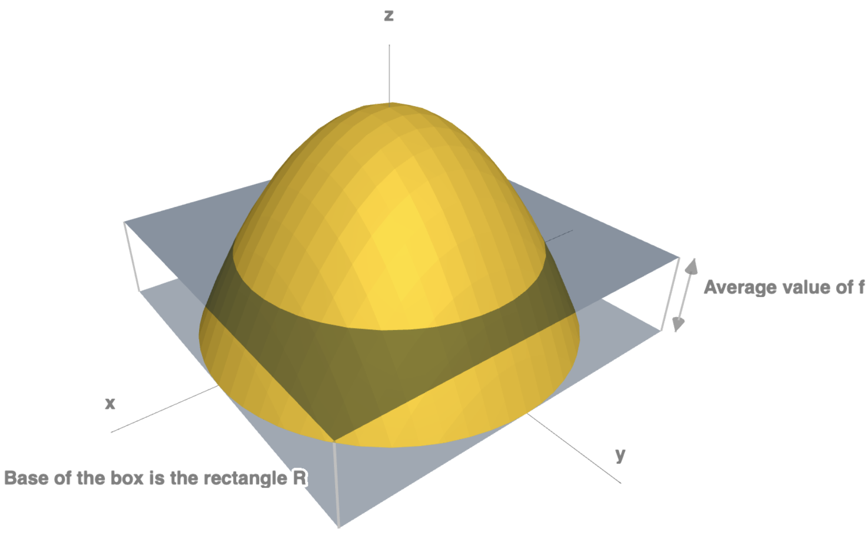
(b) Derive your conjectured integral formula using Riemann sums by following Example 8.1.4.

8.2. Average value

A second interpretation of the integral is motivated by an elegant statistical perspective.

What is the average value of a real-valued function f over a region $S \subseteq \mathbb{R}^n$?

The average value has a nice physical interpretation for rectangles S in $\mathbb{R}^2$. Assume the region below the graph of $f : S \to [0, \infty)$ is made out of uniformly dense wax. If this object melts within the perimeter of S, then the melted wax will become a box with height equal to the average value of f. See Math3D graph below, where S is the rectangle R.



Your goal in this section is to construct a formal definition of "average value" and analyze some of its properties. To begin, you can derive its definition using a limit of discrete approximations. For simplicity, start with the special case where S is a rectangle.

Example 8.2.1 (Derivation of average value) Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$ and let f be a real-valued function that is integrable on R. For $N \in \mathbb{N}^+$, let P_N be a regular partition of R into N^n subrectangles constructed from a regular partition of each subinterval $[a_j, b_j]$ into N subintervals for $1 \leq j \leq n$. Thus, the subrectangles $\{R_i\}_{i \in I_N}$ of P_N each have the same volume $\mathrm{vol}(R_i) = \mathrm{vol}(R)/N^n$ for $i \in I$.

Now, you want to construct a discrete approximation for the average value of f on R. For each subrectangle R_i, choose a sample point $x_i^* \in R_i$. The single value $f(x_i^*)$ approximates the average value of f on R_i. Since each subrectangle has the same volume, this implies that the *discrete* average of the values $f(x_i)$ for $i \in I_N$ should approximate the *continuous* average of f on R. In other words, the quantity

$$\frac{1}{|I_N|} \sum_{i \in I_N} f(x_i^*)$$

is a *discrete* approximation for the average value of f on R. By taking $N \to \infty$, you expect that this approaches the exact answer. Notice that $|I_N| = N^n$, so by multiplying top and bottom by $\mathrm{vol}(R)$, you can see that the above is equal to

$$\frac{1}{\mathrm{vol}(R)} \sum_{i \in I_N} f(x_i^*) \frac{\mathrm{vol}(R)}{N^n} = \frac{1}{\mathrm{vol}(R)} \sum_{i \in I_N} f(x_i^*) \, \mathrm{vol}(R_i) = \frac{1}{\mathrm{vol}(R)} S_{P_N}^*(f),$$

which is precisely the Riemann sum of f with partition P_N and sample points $\{x_i^*\}_{i\in I}$. By taking the limit as $N \to \infty$, the continuous average of f on R should presumably be equal to

$$\frac{1}{\operatorname{vol}(R)} \int_R f \, dV \tag{8.2.1}$$

by Theorem 7.3.18 , because f is integrable on R and $||P_N|| \to 0$ as $N \to \infty$.

Remark 8.2.2 Again, this derivation is an informal yet convincing argument to define average value using (8.2.1). There are many equivalent methods for deriving the same integral.

Inspired by (8.2.1), you can formally define average value on other sets $S \subseteq \mathbb{R}^n$.

Definition 8.2.3 Let $S \subseteq \mathbb{R}^n$ be a Jordan measurable set with non-zero volume. Let f be integrable on S. The **average value of f on S** is defined to be

$$\frac{1}{\operatorname{vol}(S)} \int_S f \, dV.$$

Now, there is a complementary question to averages.

What is the total value of $f : S \to \mathbb{R}$ over a region S in $\mathbb{R}^n$?

By following Example 8.2.1, you can conjecture that the integral

$$\int_S f \, dV$$

represents total value. You have been implicitly holding this perspective for the entirety of this chapter, so a heuristic derivation can provide supporting evidence.

All this theory culminates to confirm the desired intuitive principle:

$$(\textit{total value of } f \textit{ over a region } S) = (\textit{average value of } f \textit{ on } S) \times (\textit{volume of } S)$$

Your definition of the integral convincingly satisfies this principle from your heuristic derivations. This relationship is a triumph of your painstaking efforts to rigorously define integrals. This principle is also confirmed formally: continuous functions attain their average value on compact path-connected sets.

Theorem 8.2.4 (Integral mean value theorem) Let S be a Jordan measurable set in $\mathbb{R}^n$. Let f be a continuous function on S. If S is compact and path-connected, then there exists a point $p \in S$ such that

$$\int_S f \, dV = f(p) \operatorname{vol}(S).$$

Proof. This is left as an exercise. Use the extreme value theorem and monotonicity of the integral to show that the integral is a real number between two values. Then use a version of the intermediate value theorem. ■

The integral mean value theorem provides a beautiful interpretation of the integrand. Namely, the value of a continuous function is the limit of its average value on shrinking balls.

Corollary 8.2.5 Fix $p \in \mathbb{R}^n$. Let f be a real-valued function. If f is continuous on an open set containing p, then

$$f(p) = \lim_{\varepsilon \to 0^+} \left[\frac{1}{\operatorname{vol}(B_\varepsilon(p))} \int_{B_\varepsilon(p)} f \, dV \right].$$

Proof. This is left as an exercise. Use the integral mean value theorem. ∎

Notice how Corollary 8.2.5 parallels the fundamental theorem of calculus in one-dimension.

A limit of shrinking integrals acts like a derivative!

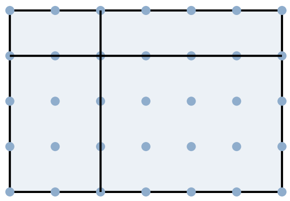
This insight is far more powerful than you might expect and it will recur many times when you explore vector calculus and generalizations of the fundamental theorem of calculus. Now, if you pause for a moment, then it is worthwhile to appreciate what has just happened. You were studying another application of the integral and, in doing so, you discovered several great theorems. This phenomenon is a perfect example of how applied and pure mathematics mutually enrich each other. Indeed, the insight from Corollary 8.2.5 will already impact your next perspective of the integral as the mass of an object.

Exercises for Section 8.2

Computations

8.2.1 Any partition can be used to produce a discrete approximation. It does not need to be regular.

(a) A rectangle $R=[0,6]\times[0,4]$ is partitioned into four pieces $R_{11}, R_{21}, R_{12}, R_{22}$ below. Let x_i be a point in R_i. Let $f : R \to \mathbb{R}$ be integrable. Consider the table of function values.



x	x_{11}	x_{21}	x_{12}	x_{22}
$f(x)$	0.5	-0.6	0.75	1.0

Use this data to approximate the average value of f of R using this non-regular partition.

(b) Let R be any rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be integrable. Construct *three* different discrete approximations of the average value of f on R using *any* partition P of R.

(c) For each approximation you created, what "limit" should equal the integral formula for the average value of f on R?

Proofs

8.2.2 Here you derive the average value of a function f using regular partitions, upper sums, and lower sums.

- Let $R=[a_1,b_1]\times\cdots\times[a_n,b_n]$ be a rectangle in $\mathbb{R}^n$.
- Let $f : R \to \mathbb{R}$ be an integrable real-valued function.
- For $N\in\mathbb{N}^+$, let P_N be a regular partition of R into N^n subrectangles constructed from a regular partition of each subinterval $[a_j,b_j]$ into N subintervals for $1\le j\le n$.
- Let $\{R_i\}_{i\in I_N}$ be the subrectangles of P_N, so each subrectangle has volume $\mathrm{vol}(R)/N^n$.

(a) Construct a discrete over-estimate for the average value of f on R.

(b) Construct a discrete under-estimate for the average value of f on R.

(c) Explain why your estimates should agree as $N\to\infty$ and derive the integral formula for average value.

8.2.3 Prove the mean value theorem for integrals, namely Theorem 8.2.4. *Hint:* Show that the average value of f lies between the global extreme values of f.

8.2.4 Prove that the value of a continuous function is the limit of its average value on shrinking balls, namely Corollary 8.2.5.

Applications and beyond

8.2.5 Let $S\subseteq\mathbb{R}^n$ be a Jordan measurable set with positive volume. Let $f : S\to\mathbb{R}$ be integrable on S. Derive the integral formula for the average value of f on S using Riemann sums. Use *any* partition of a rectangle R containing S, not necessarily regular partitions.

8.2.6 Derive (8.2.1) using any sequence of partitions P_N with $||P_N||\to 0$.

8.2.7 Derive the integral formula for average value of a function $f : S \to \mathbb{R}$ (Definition 8.2.3) for any compact Jordan measurable set S. Assume f is integrable on S.

8.2.8 Derive the formula for total value of a function $f : S \to \mathbb{R}$ over a compact Jordan measurable set $S \subseteq \mathbb{R}^n$. Assume f is integrable on S.

8.3. Mass

A third interpretation of the integral is physical. There are many such perspectives, and the concept of mass is a classic situation.

> *What is the mass* mass(S) *of an object* S *in* $\mathbb{R}^n$ *with* ***variable*** *density* $\delta : S \to [0, \infty)$*?*

The "mass" function is a physical notion that you are taking for granted; it is not formally defined. The goal of this section is to construct definitions of mass and density that agree with fundamental physical principles and that can be heuristically derived with your theory of the integral. This investigation will build one of the most versatile viewpoints of the integral.

8.3.1 Density

This question about mass has a subtle yet immediately troubling issue.

> *For a given point* x*, what does the density* $\delta(x)$ *represent at that single point?*

On one hand, a point has zero volume in $\mathbb{R}^n$ so it should have zero mass. This seems to suggest that density at a point cannot be rigorously defined, but all is not lost. The key insight can be derived from a fundamental physical principle:

$$\textit{density} = \frac{\textit{mass}}{\textit{volume}} \quad \text{for objects with } \textit{uniform} \text{ (i.e. constant) density.}$$

This principle also generalizes to objects with non-uniform density:

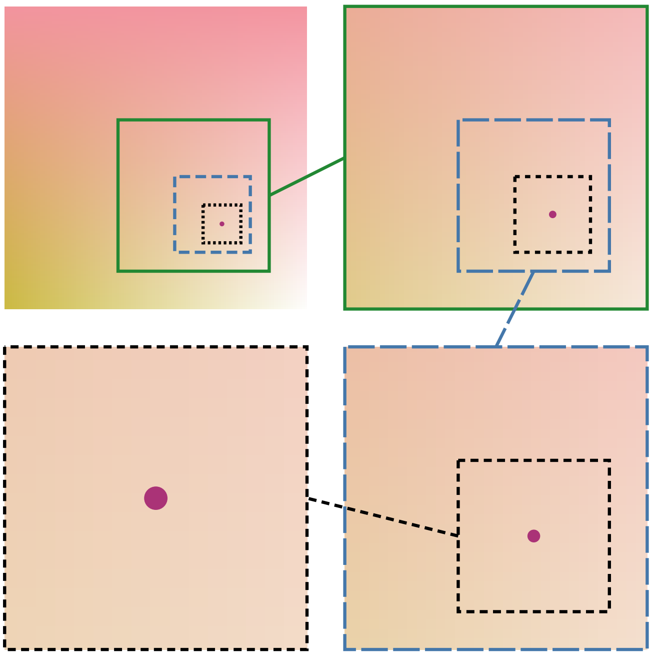
$$\textit{average density} = \frac{\textit{mass}}{\textit{volume}} \quad \text{for any object.}$$

Now, the extra inspiration comes from Corollary 8.2.5, which informally states that the average value of a function on shrinking balls centered at $x \in \mathbb{R}^n$ is actually the function value at x.

> *For a given point* x*, can the density* $\delta(x)$ *be thought of as the average density on shrinking balls centered at* x*?*

Absolutely! A visual example can help illuminate this great idea.

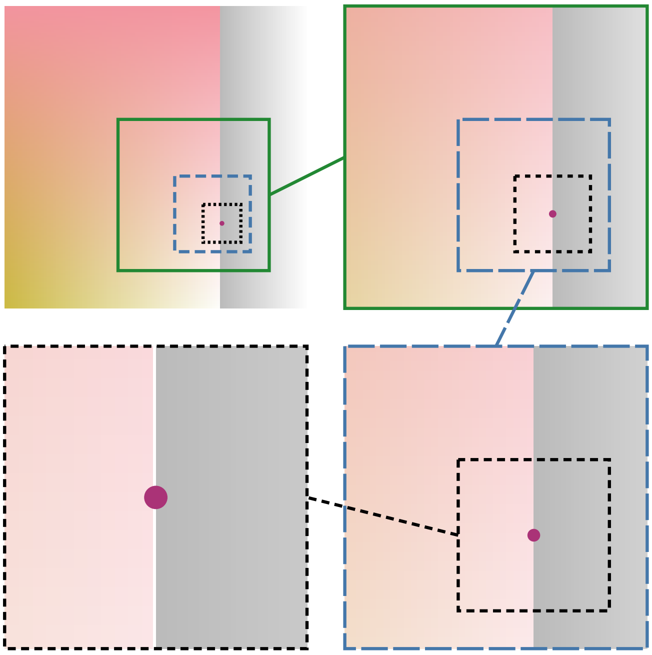
Example 8.3.1 Consider zooming onto a slice of soft cheese in $\mathbb{R}^2$ with variable density.



Suppose you are zooming into the point $p \in \mathbb{R}^2$ on the slice of cheese. The average density of the entire square is not a good measure of how dense the cheese will be near p, because the density varies so much. If you cut out a smaller subrectangle near p, then the density will be less variable across that subrectangle. As you cut smaller and smaller subrectangles centred at p, the density will presumably be *approximately uniform* on the smaller subrectangles. This means that the average density on a very small subrectangle centred at p could be considered a good approximate value for the "density at p".

Now, the above example has one caveat: the figure assumes the density is *continuous*. Surprisingly, there are many examples of non-continuous density functions in everyday life.

Example 8.3.2 A knife blade is touching the edge of your cheese slice.



Suppose you are zooming into the point $q \in \mathbb{R}^2$ on this edge. The lefthand side is the cheese with low average density that is roughly uniform, say 1 gram per cm^3. The righthand side is the steel blade with high average density that is roughly uniform, say 8 grams per cm^3. A small subrectangle centred at q has roughly half cheese and half steel. Therefore, the "density at q" should be the average of these two values, namely $\frac{1+8}{2} = 4.5$ grams per cm^3. Thus, the density function of this space is presumably discontinuous at q.

Notice this heuristic holds true for *any point along the edge*, so you can reasonably model this space with a density function that is discontinuous along the entire edge and continuous everywhere else. There is a "jump" in the density from the cheese to the knife. On the other hand, the set of these discontinuities form a set of zero volume in $\mathbb{R}^2$ since it is only a line segment. According to your theorems on integrability, this implies your density function should be integrable on the space.

These awesome observations harken back to the many examples of densities from Section 1.2.1. It is worthwhile to try and describe this idea somewhat more formally with the language of limits, cubes, and volume.

Example 8.3.3 Imagine $S \subseteq \mathbb{R}^3$ is a fancy block of cheese with varying density. You can measure its mass and volume to get its average density

$$\frac{\text{mass}(S)}{\text{vol}(S)}$$

This approximates the density at a specific point $(x, y, z) \in S$. To improve this approximation, chop a small ε-cube

$$C_\varepsilon = [x-\varepsilon, x+\varepsilon] \times [y-\varepsilon, y+\varepsilon] \times [z-\varepsilon, z+\varepsilon]$$

which contains $(x, y, z) \in S$. The average density of this cube

$$\frac{\text{mass}(S \cap C_\varepsilon)}{\text{vol}(S \cap C_\varepsilon)}$$

is presumably a better approximation of the density. By taking smaller and smaller cubes, i.e. $\varepsilon \to 0^+$, you expect that the limit of average densities

$$\lim_{\varepsilon \to 0^+} \frac{\text{mass}(S \cap C_\varepsilon)}{\text{vol}(S \cap C_\varepsilon)}$$

exists and should represent the mass density at the point $(x, y, z) \in S$.

Heuristically speaking, the mass density function $\delta : S \to [0, \infty)$ should satisfy

$$\delta(x) = \lim_{\varepsilon \to 0^+} \frac{\text{mass}(S \cap B_\varepsilon(x))}{\text{vol}(S \cap B_\varepsilon(x))}.$$

This conjectured identity uses solid balls, but the shape does not really matter; shrinking solid cubes would be fine too. While this limit has no formal meaning, it can serve as the basis for your intuitive understanding of density. The bottomline can be succinctly summarized.

Mass density is the limit of shrinking average densities.

It is quite remarkable how Corollary 8.2.5 on shrinking averages inspired this idea.

8.3.2 Mass and average density

With the interpretation of density settled, you can return to the original question on mass.

How can you define the mass of an object $S \subseteq \mathbb{R}^n$ with variable density $\delta : S \to [0, \infty)$?

There are many possible derivations for the formula and, as usual, they all follow the same basic strategy. The special case of a rectangle reveals the key items.

Example 8.3.4 (Derivation of mass of a rectangular solid) Let $R \subseteq \mathbb{R}^n$ be a rectangular solid with variable density $\delta : R \to [0, \infty)$. Assume δ is integrable on R. Let P be a partition of R with subrectangles $\{R_i\}_{i \in I}$. For $i \in I$, let m_i be the mass of the object inside R_i. Since average density on R_i is the mass divided by volume, it follows for $i \in I$ that

$$\inf_{x \in R_i} \delta(x) \leq \frac{m_i}{\text{vol}(R_i)} \leq \sup_{x \in R_i} \delta(x).$$

Multiplying both sides by $\mathrm{vol}(R_i)$ and summing over $i \in I$, you get that

$$\sum_{i\in I}\Big[\inf_{x\in R_i}\delta(x)\Big]\mathrm{vol}(R_i) \le \sum_{i\in I} m_i \le \sum_{i\in I}\Big[\sup_{x\in R_i}\delta(x)\Big]\mathrm{vol}(R_i).$$

The lefthand expression is the lower sum $L_P(\delta)$ and the righthand expression is the upper sum $U_P(\delta)$. The middle expression is the mass of the solid $m = \mathrm{mass}(S)$. Thus,

$$L_P(\delta) \le m \le U_P(\delta)$$

for every partition P of R. Hence, m is an upper bound for all lower sums, implying by definition of supremum that $\underline{I_R}(\delta) \le m$ and similarly $m \le \overline{I_R}(\delta)$. Since δ is integrable on R by assumption, the upper and lower integral are equal, so

$$m = \int_R \delta\, dV.$$

These heuristic arguments inspire integral formulas for classical notions in physics and give a new perspective on what the integral really represents.

Definition 8.3.5 Let $\delta : S \to [0,\infty)$ be the density function for an object $S \subseteq \mathbb{R}^n$. Assume S is bounded and δ is integrable on S. Its **mass** $m = \mathrm{mass}(S)$ and **average density** ρ are respectively given by

$$m = \int_S \delta\, dV, \qquad \rho = \frac{1}{\mathrm{vol}(S)}\int_S \delta\, dV.$$

This concludes your investigation of mass, but the fun with integrals and mass does not stop there. You can derive other fundamental physical notions with the same heuristics.

8.3.3 Centre of mass

Integral formulas are abundant in physics and they are not always so simple. A derivation should be clearly based on physical principles and discrete approximations. A more complicated example involves the centre of mass which is a *weighted* average.

Example 8.3.6 (Centre of mass for a discrete system) Consider a system of point masses at positions $x_1,\dots,x_N \in \mathbb{R}^n$ with masses $m_1,\dots,m_N$. The centre of mass of this discrete system is given by the weighted average

$$\frac{\sum_{i=1}^N m_i x_i}{\sum_{i=1}^N m_i}.$$

The weights are the masses at each point; objects with heavier weight will impact the average greater than those with small weight. There is a physical explanation for this weighted average, but you may take it for granted.

This discrete approximation leads to a continuous version for centre of mass.

Definition 8.3.7 Let $\delta : S \to [0,\infty)$ be the density of an object $S \subseteq \mathbb{R}^n$. Assume S is bounded and δ is integrable on S. Its **centre of mass** $\overline{x} = (\overline{x}_1,\dots,\overline{x}_n) \in \mathbb{R}^n$ is given by

$$\overline{x} = \frac{1}{m}\int_S x\delta(x)\, dV$$

where $m > 0$ is the mass of S. Its **centroid** is the centre of mass assuming density is uniform.

Remark 8.3.8 If $f : S \to \mathbb{R}^k$ with components $f_1, \ldots, f_k$, then the integral of f over $S \subseteq \mathbb{R}^n$ is defined to be the vector

$$\int_S f\, dV := \left(\int_S f_1 dV \,,\ \ldots\ ,\ \int_S f_k dV \right) \in \mathbb{R}^k,$$

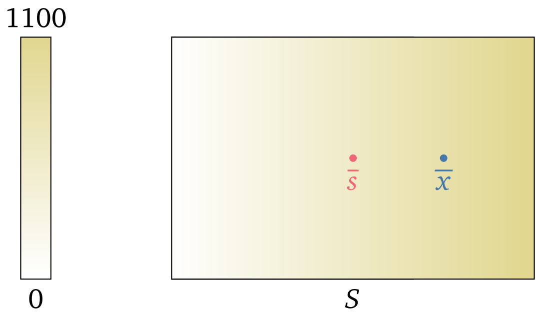
if every component exists. This is mostly for notational convenience.

Remark 8.3.9 You can show that the centroid of S is given by

$$\frac{1}{\mathrm{vol}(S)} \int_S x dV \in \mathbb{R}^n.$$

Informally speaking, the centre of mass is the "weighted centre of your object" and the centroid is the "geometric centre of your object". This physical intuition can actually assist you in better understanding the integral and how it depends on the mass density.

Example 8.3.10 Let $S \subseteq \mathbb{R}^2$ be a slice of cheese with variable density $\delta : S \to [0, \infty)$ measured in mg/mL^3. This slice is illustrated below.



Its centre of mass $\bar{x} \in \mathbb{R}^2$ and centroid $\bar{s} \in \mathbb{R}^2$ are labelled. Notice how the centre of mass shifts according to the density δ and its symmetries. The centroid only depends on S.

This finishes the standard collection of viewpoints on the integral as volume under a graph, average value, and mass. Your theory of the integral has finally confirmed the three core questions of integration. These applications are important in their own right, and they expand on your classical understanding of the single variable integral. Each perspective will be useful for your geometric and physical intuition. However, you currently have almost no tools to actually calculate higher dimensional integrals yet. This will be remedied in the next chapter. Before you conclude this chapter, you will explore one of the most powerful applications of the multivariable integral: probability.

Exercises for Section 8.3

Concepts and definitions

8.3.1 Language describing mass density is quite subtle and can quickly go wrong. Eeyore, Owl, Piglet, Roo, and Tigger are discussing the mass density function $\delta : S \to [0, \infty)$ at a point $x \in S$.

> Eeyore says *"The value $\delta(x)$ is equal to the density of S at the point x."*
>
> Owl says *"The value $\delta(x)$ is approximately the density of S at the point x."*
>
> Piglet says *"The value $\delta(x)$ is equal to the average density of S at the point x."*
>
> Roo says *"The value $\delta(x)$ is approximately the average density of S at the point x."*
>
> Tigger says *"None of these claims are accurate!"*

Who is the closest to the truth? Explain why.

8.3.2 You list physical examples of objects $S \subseteq \mathbb{R}^3$ with a mass density function $\delta : S \to [0, \infty)$ that satisfies the given property. These examples will help you interpret these mathematical properties from a physical viewpoint. Try to come up with as many different examples as you can. Assume $\operatorname{vol}(S) > 0$.

(a) List examples of objects S where the mass density δ is uniform.

(b) List examples of objects S where the mass density δ is continuous.

(c) List examples of objects S where the mass density δ is continuous except on a zero volume set.

(d) List examples of objects S where the mass density δ is not continuous anywhere.

8.3.3 You will often need to translate between the language of integrals and language of mass. This fluency will be quite rewarding, so it is worthwhile to practice. Let $A \subseteq \mathbb{R}^3$ and $B \subseteq \mathbb{R}^3$ be solids with mass densities $\delta_A : A \to [0, \infty)$ and $\delta_B : B \to [0, \infty)$ respectively. For each part below, translate the informal sentence into a formal mathematical statement, or vice versa.

(a) The solid A weighs more than solid B.

(b) The volume of A is equal to the volume of B.

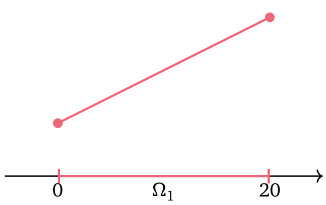
(c) The average density of A exceeds the average density of B.

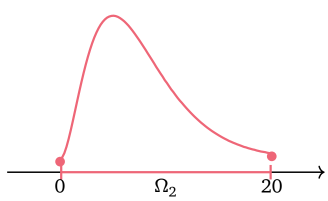
(d) $\exists k \in (0, \infty)$ s.t. $\forall x \in B, \delta_B(x) = k$

(e) $\forall x \in A, \quad \delta(x) = \lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{vol}(A \cap B_\varepsilon(x))} \int_{A \cap B_\varepsilon(x)} \delta \, dV$

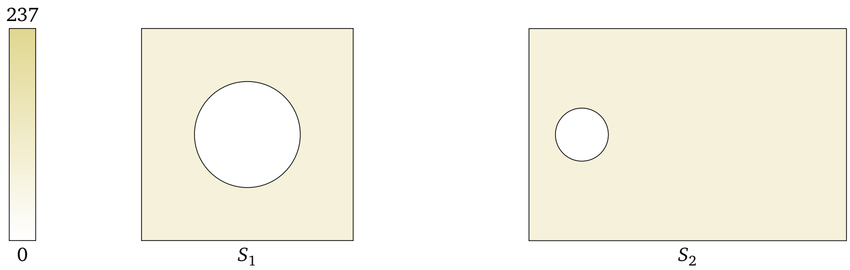
8.3.4 For each object, approximate the position of its centroid and centre of mass.

(a) Two wires Ω_1 and Ω_2 with mass density functions δ_1 and δ_2 respectively.

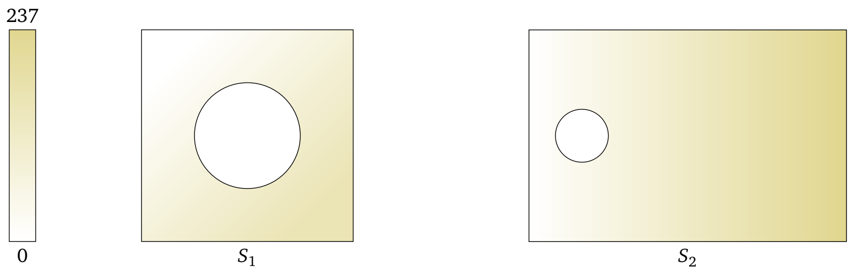




(b) Two slices of cheese S_1 and S_2 in $\mathbb{R}^2$ with a mass density grayscale picture.



(c) Two slices of cheese S_3 and S_4 in $\mathbb{R}^2$ with a mass density grayscale picture.



8.3.5 You now have four interpretations of the integral. It is good to summarize those viewpoints again. Let f be integrable on a Jordan measurable set $S \subseteq \mathbb{R}^n$. Fill in the blanks.

(a) The integral $\displaystyle\int_S f\,dV$ represents the volume of the region ________________ provided ________________.

(b) The volume of the set S is defined by the integral ________________.

(c) The average value of f over S is given by ________________ provided ________________.

(d) The mass of an object S with ________________ is given by $\displaystyle\int_S f\,dV$ provided ________________ .

8.3.6 Let $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9\}$ be a slice of cheese with a C^1 variable density $\delta : C \to [0, \infty)$. Let $(\overline{x}, \overline{y})$ be the centre of mass of C. Define the subsets

$$N = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9, y \geq 0\} \qquad E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9, x \geq 2\}$$
$$S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9, y \leq 0\} \qquad W = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9, x \leq 2\}$$

Notice E and W are not symmetric about the vertical axis. Which statements are true or false? *Hint:* Do not try to argue everything mathematically. Use your physical intuition.

(a) If $\frac{\partial \delta}{\partial y} > 0$ on C, then the mass of N exceeds the mass of S.

(b) If $\frac{\partial \delta}{\partial x} > 0$ on C, then the mass of E exceeds the mass of W.

(c) If the mass of N is greater than the mass of S, then $\frac{\partial \delta}{\partial y} > 0$ on C.

(d) If the mass of N is greater than the mass of S, then the volume of N exceeds the volume of S.

(e) If the mass of N is greater than the mass of S, then the centroid of C lies above the y-axis.

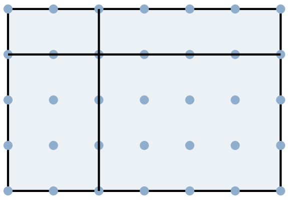
(f) If the mass of N is greater than the mass of S, then $\overline{y} > 0$.

(g) If the mass of N equals the mass of S, then $\overline{y} = 0$.

(h) If the mass of E is greater than the mass of W, then $\overline{x} > 2$.

Computations

8.3.7 A rectangular slice of cheese $R \subseteq \mathbb{R}^2$ with dimensions 6 cm by 4 cm is partitioned into four pieces $R_{11}, R_{21}, R_{12}, R_{22}$. Let x_i be a point R_i for each index i. Let $\delta : R \to [0, \infty)$ be the continuous mass density of R measured in grams per cm^2. Consider the table of function values.



x	x_{11}	x_{21}	x_{12}	x_{22}
$f(x)$	0.5	0.2	0.75	1.0

(a) Use this data and diagram below to approximate the mass of R.

(b) Is your approximation an underestimate or overestimate?

Proofs

8.3.8 You will analyze a derivation for the integral formula of the mass of a rectangular solid. Remember this is a heuristic justification. It is not a proof.

1. *Let $R \subseteq \mathbb{R}^n$ be a rectangular solid with continuous density $\delta : R \to [0, \infty)$.*
2. *Let P be a partition of R with subrectangles $\{R_i\}_{i \in I}$.*
3. *For $i \in I$, let m_i be the mass of the object inside R_i.*
4. *Since average density on R_i is the mass divided by volume, it follows for $i \in I$ that*

$$\inf_{x \in R_i} \delta(x) \leq \frac{m_i}{\operatorname{vol}(R_i)} \leq \sup_{x \in R_i} \delta(x).$$

5. *Multiplying both sides by $\operatorname{vol}(R_i)$ and summing over $i \in I$, you get that*

$$\sum_{i \in I} \Big[\inf_{x \in R_i} \delta(x) \Big] \operatorname{vol}(R_i) \leq \sum_{i \in I} m_i \leq \sum_{i \in I} \Big[\sup_{x \in R_i} \delta(x) \Big] \operatorname{vol}(R_i).$$

6. *This implies that*

$$L_P(\delta) \leq m \leq U_P(\delta)$$

7. *As the partition P was arbitrary,*

$$\underline{I_R}(\delta) \leq m \leq \overline{I_R}(\delta)$$

8. *Since δ is continuous and hence integrable on R by assumption, it follows that*

$$m = \int_R \delta \, dV.$$

(a) Line 4 requires a leap of faith, unless you assume there exists $x_i^* \in R_i$ such that $\delta(x_i^*) = \dfrac{m_i}{\operatorname{vol}(R_i)}$. Based on your assumptions on δ and physical intuition, why does this assumption seem sensible?

(b) Why does Line 7 follow from Line 6? Identify one detail that is missing.

(c) This derivation only applies for a rectangular solid R. How can you modify it to apply to a non-rectangular solid S? Identify any assumptions on S or δ that will be necessary.

Applications and beyond

8.3.9 You can construct different discrete approximations and still derive the same integral formula for mass. Let R be any rectangle in $\mathbb{R}^n$ with mass m. Let $\delta : R \to [0, \infty)$ be an integrable mass density function. Let P be an arbitrary partition of R with subrectangles $\{R_i\}_{i \in I}$.

(a) Construct a discrete approximation of the mass m using a sample point $x_i^* \in R_i$ for each $i \in I$.

(b) Construct a discrete over-estimate of the mass m.

(c) Construct a discrete under-estimate of the mass m.

(d) For each estimate above, what "limit" should derive the integral formula for the mass m?

8.3.10 What is the mass density function and why does it make any sense? A basic principle is the key.

If an object has ***uniform*** *density then its mass is equal to its density times its volume.*

With this physical law, you can define the mass density function for objects with **non-uniform** density. Let $S \subseteq \mathbb{R}^n$ be a block of compact Jordan measurable n-dimensional cheese. For $x \in S$, the mass density $\delta(x)$ at x can be defined as

$$\delta(x) = \lim_{\varepsilon \to 0^+} \frac{\text{mass}(C_\varepsilon(x) \cap S)}{\text{vol}(C_\varepsilon(x) \cap S)}, \tag{$\star$}$$

where $C_\varepsilon(x) = \{x + y : y \in [-\varepsilon, \varepsilon]^n\}$ is a cube centered at x with sidelengths 2ε.

(a) What does the quantity inside the limit represent?

(b) Based on the physical principle, under what assumptions on δ do you expect this limit to exist at $x \in S$? Explain why those assumptions are often reasonable.

(c) Is ($\star$) a mathematically rigorous definition? Explain why or why not.

8.3.11 A sheet $R \subseteq \mathbb{R}^2$ has continuous mass density $\delta : R \to [0, \infty)$. Let $(\overline{x}, \overline{y})$ be its centre of mass.

(a) Conjecture an integral formula for $\overline{x}$. You will derive it here.

(b) Use a partition of R to construct a discrete approximation for $\overline{x}$ using Riemann sums.

(c) Show an appropriate limit of your discrete approximations yields an integral formula for $\overline{x}$.

8.3.12 Derive the integral formula for mass of a non-rectangular solid $S \subseteq \mathbb{R}^n$ with mass density $\delta : S \to [0, \infty)$ using Riemann sums. Identify any assumptions you may need for S or δ.

8.3.13 Derive the formula for average density (Definition 8.3.5) of a rectangular solid using upper and lower sums and *any* partition.

8.3.14 Show that the centroid of S is given by $\dfrac{1}{\text{vol}(S)} \displaystyle\int_S x \, dV \in \mathbb{R}^n$.

8.4. Probability

Physical applications of the integral are usually in dimensions one, two, or three, because these are natural choices for describing your physical reality. Yet you have spent enormous efforts to develop the integral in $\mathbb{R}^n$ for arbitrary $n \geq 1$. Why do you need all of this careful construction of integrals and the Jordan measure for higher dimensions? It may be satisfying to generalize ideas, but you should certainly demand better motivation for such abstract machinery.

Historically, the development of integration in higher dimensions was intensely motivated by probability. Probabilities involving finite discrete outcomes (e.g. flipping a coin three times) are fairly intuitive to describe and do not require any integrals. You can do some counting and express everything using finite sums. Here you will be focused on a complementary setting.

How do you rigorously describe probabilities of ***continuous*** *outcomes?*

For instance, what is the probability that you will wait more than 15 minutes in a line-up for your student ID card? Time is a continuous quantity, so counting is not a viable method. Similarly, quantum mechanics suggests that the position of an electron is a probabilistic notion; what is the probability that an electron for a hydrogen atom belongs to a specific orbital? Space is treated as a continuous quantity, so the same obstacle occurs.

[1]


[2]

Mathematicians in the late 1800s were struggling to define probability with continuous random quantities. This struggle pushed them to axiomatically describe what properties were required and subsequently develop a deep theory of integration that satisfies these axioms. Your efforts thus far have been secretly making progress towards this greater goal. In this section, you will use the rigorous theory of the integral and the Jordan measure to propose a definition for continuous probabilities. This proposal will have significant theoretical limitations, but you can still use it to analyze interesting problems in probability. As a bonus, you will also generate a new and powerful viewpoint for the integral.

You will begin by studying some standard terminology and notation. These terms are usually axiomatically defined in a rigorous probability course[3], but an informal introduction will suffice for your purposes in this textbook. Throughout this section, you will use a running example of coin flips to solidify the intuition.

[1]Image taken by Geoffrey Vendeville, used with permission and retrieved from University of Toronto on 2024-07-22

[2]Image retrieved from Wikimedia Commons on 2024-07-23 licensed under CC BY-SA.

[3]You might have already been introduced to probability elsewhere. Other texts usually introduce probability as a bunch of axioms about "events" and "probability functions". It may later be explained that integrals can be used as an example which satisfies these axioms. In this text, you will take the *opposite* approach. First, you will define "events" and "probability functions" using integrals. Afterwards, you will prove that it satisfies the desired properties. This approach is somewhat narrow, but that is okay since the intention is to give a brief glimpse of probability.

8.4.1 Sample space and events

First, a random phenomenon has many possible outcomes. A **sample space** Ω is an arbitrary non-empty set that contains all possible outcomes.

Example 8.4.1 (Coin flips - sample space) Flipping a fair coin twice is a random phenomenon. If H denotes heads and T denotes tails, then the sample space Ω contains all 4 possible outcomes, namely

$$\Omega = \{HH, HT, TH, TT\}.$$

Second, you may want to track a certain subset of possible outcomes. The **event space** Σ is a collection of subsets of the sample space Ω. Every element $A \in \Sigma$ is an **event**, so $A \subseteq \Omega$ by definition. Note the event space does not need to be all subsets of Ω.

Example 8.4.2 (Coin flips - events) There are many choices for the event space of flipping a fair coin twice . The most natural choice Σ is all possible subsets of Ω. Since $|\Omega| = 4$, there are $2^4 = 16$ such subsets so explicitly

$$\Sigma = \big\{\varnothing, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}, \{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}, \Omega\big\}.$$

The event A that you flip exactly 1 head and 1 tails is the set

$$A = \{HT, TH\}$$

and notice $A \subseteq \Omega$. The event B that you flip at least 1 head or at least 1 tails always occurs, so $B = \Omega$. The event C that you flip neither head nor tails is impossible, so $C = \varnothing$.

Now, Σ is not the only choice of event space. You could also choose

$$\Sigma' = \{\varnothing, A, \Omega \setminus A, \Omega\}$$

if you are only interested in whether or not A occurs. Notice the event A does not occur if and only if the event $\Omega \setminus A$ occurs.

The event space Σ should satisfy some basic properties, such as:

- $\Omega \in \Sigma$ *"Some outcome must happen."*
- If $A \in \Sigma$ then $\Omega \setminus A \in \Sigma$. *"An event either occurs or does not occur."*
- If $A_1, \ldots, A_N \in \Sigma$ then $A_1 \cup \cdots \cup A_N \in \Sigma$. *"Any finite union of events is an event."*

You can verify that the previous example satisfied all three of these properties.

With your theory of Jordan measurable sets, you can define a *continuous* sample space and an event space with these properties.

Theorem 8.4.3 Let $\Omega \subseteq \mathbb{R}^n$ be a Jordan measurable set. Define the event space Σ to be

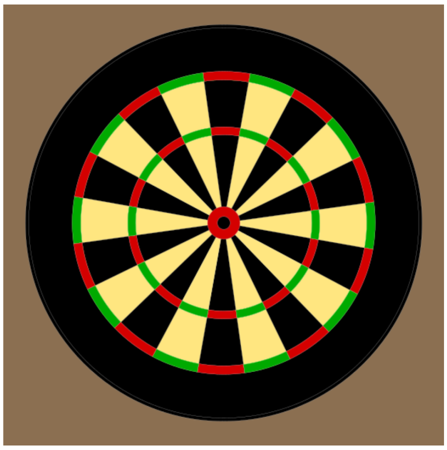
$$\Sigma = \{A \subseteq \Omega : A \text{ is Jordan measurable}\}.$$

All of the following hold:
(a) $\Omega \in \Sigma$
(b) If $A \in \Sigma$, then $\Omega \setminus A \in \Sigma$.
(c) If $A_1, \ldots, A_N \in \Sigma$, then $A_1 \cup \cdots \cup A_N \in \Sigma$.

Proof. (a) follows by assumption. (b) is left as an exercise. (c) follows from Lemma 7.5.18 and a straightforward induction. ■

This creates a useful definition for continuous random phenomena.

Example 8.4.4 (Dart board – events) Suppose you are throwing a single dart onto a 4 by 4 square board. Assume you always hit the board. The location where the dart lands is a random phenomenon that you can model in $\mathbb{R}^2$. The sample space $\Omega = [-2, 2]^2$ can represent the board. Since Ω is a rectangle, it is Jordan measurable. Theorem 8.4.3 suggests you should define Σ to be all Jordan measurable subsets of Ω.[4]



You want to get better at throwing darts so you keep track of when you land within unit distance of the centre. In other words, the set

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$$

should be an event belonging to the event space Σ. Indeed, A is the unit disk so $A \subseteq \Omega$ and you can verify that A is Jordan measurable, so $A \in \Sigma$ by definition.

Now that you can describe random phenomena and track the occurrence of events, your third and final task is to define the likelihood that an event occurs. Your goal is therefore to assign a probability to each event in your event space.

8.4.2 Probability spaces and densities

Given a sample space Ω and an event space Σ, a **probability function** $\mathbb{P} : \Sigma \to [0, 1]$ is a function assigning each event in the event space a probability between 0 and 1. Thus, if A is an event in Σ, then $\mathbb{P}(A)$ is the probability that A occurs. A **probability space** is the triple $(\Omega, \Sigma, \mathbb{P})$.

Example 8.4.5 (Coin flips – probability space) Assume you are flipping a *fair* coin twice. Let Ω be the sample space from Example 8.4.1 and let Σ be all possible subsets of Ω. Since you are flipping a fair coin, the probability that any of the 4 possible outcomes HH, HT, TH, and TT occur are equally likely, so they each occur with probability $\frac{1}{4}$. More generally, $\mathbb{P} : \Sigma \to [0, 1]$ is defined to be

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{4}$$

for any event $E \in \Sigma$. For instance, if A is the event that you flip exactly 1 head and 1 tail, then $\mathbb{P}(A) = \frac{2}{4} = \frac{1}{2}$.

Theorem 8.4.3 provides a choice for a continuous sample space Ω and event space Σ, so you only need to choose a probability function. Since the event space consists of Jordan measurable

[4] Image modified from Wikimedia Commons on 2024-07-22 licensed under CC0.

subsets of Ω, you can use the integral to define *continuous* probabilities!

Definition 8.4.6 A triple $(\Omega, \Sigma, \mathbb{P})$ is a **continuous probability space**[5] **in** $\mathbb{R}^n$ provided:

- The **sample space** Ω is a Jordan measurable subset of $\mathbb{R}^n$.
- The **event space** Σ is the set of all Jordan measurable subsets of Ω.
- The **probability function** $\mathbb{P} : \Sigma \to [0,1]$ is given by

$$\mathbb{P}(A) = \int_A \phi \, dV \quad \text{for } A \in \Sigma,$$

  for some bounded non-negative function $\phi : \Omega \to [0, \infty)$ such that ϕ is continuous on Ω except for a set of zero Jordan measure and $\int_\Omega \phi = 1$.

The function ϕ is the **probability density function** of $\mathbb{P}$.

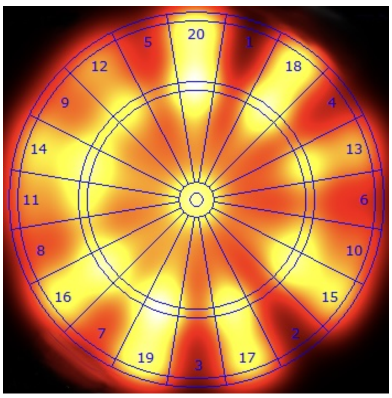
A continuous probability space satisfies natural properties in addition to Theorem 8.4.3.

Theorem 8.4.7 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. Then
(a) $\mathbb{P}(\Omega) = 1$.
(b) $\mathbb{P}(A)$ exists and $0 \le \mathbb{P}(A) \le 1$ for every $A \in \Sigma$.
(c) If $A_1, \dots, A_N \in \Sigma$ are pairwise disjoint, then $\mathbb{P}(A_1 \cup \cdots \cup A_N) = \displaystyle\sum_{i=1}^N \mathbb{P}(A_i)$.

Proof. Let ϕ be the probability density function of $\mathbb{P}$. Note (a) follows by assumption. For (b), fix $A \in \Sigma$ so A is a Jordan measurable subset of Ω. By assumption, ϕ is continuous on Ω (and hence A) except a set of zero Jordan measure. By Theorem 7.7.4, it follows that $\mathbb{P}(A) = \int_A \phi \, dV$ exists. It is left as an exercise to verify that $0 \le \mathbb{P}(A) \le 1$. (c) is also left as an exercise. ■

This formal notion can be applied to model many situations.

Example 8.4.8 (Dart board – density) Continue with the 4 by 4 dart board from Example 8.4.4. Throwing a single dart can be modeled with a continuous probability space $(\Omega, \Sigma, \mathbb{P})$ with the same space $\Omega = [-2,2]^2$ and event space Σ as before. What does the probability density function $\phi : \Omega \to [0, \infty)$ of $\mathbb{P}$ represent? Assume ϕ has the following heat map[6].



[5]The terminology "continuous probability space" and this definition are not standard, but this convention will be suitable for your purposes. See a course on measure theory for the modern mathematical standard.

The "hot areas" indicate regions where the dart is likely to land. The integral of ϕ over a region $A \subseteq [-2,2]$ gives the probability $\mathbb{P}(A)$ that the dart lands inside A. Although $\phi \geq 0$, notice that ϕ can have values exceeding 1.

The probability density function ϕ gives you yet another way to interpret the integral using probability. Unlike discrete contexts, the value of the probability density function $\phi(x)$ is *not* the probability that $x \in \Omega$ occurs. You can verify that

$$\mathbb{P}(\{x\}) = 0, \qquad \text{for all } x \in \Omega.$$

Indeed, it is possible that ϕ is much larger than 1. It is better to think of ϕ like you would mass density; for instance, a 1 kilogram object can have a mass density of 3 kg/m^3 for a small proportion of the object. There are many reasonable choices of probability density functions, but one choice is especially common.

Definition 8.4.9 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. The probability function $\mathbb{P}$ is **uniform** if its probability density function $\phi : \Omega \to [0, \infty)$ is constant. That is,

$$\phi(x) = \frac{1}{\mathrm{vol}(\Omega)} \qquad \text{for all } x \in \Omega.$$

Remark 8.4.10 The phrase "uniform" or "uniformly distributed" can be equivalently applied to the probability density function ϕ itself.

Example 8.4.11 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. You can show that if $\mathbb{P}$ is uniform, then for every event $A \in \Sigma$,

$$\mathbb{P}(A) = \frac{\mathrm{vol}(A)}{\mathrm{vol}(\Omega)}.$$

Example 8.4.12 (Dart board – uniform density) Continue with the dart board from Example 8.4.8 and a continuous probability space $(\Omega, \Sigma, \mathbb{P})$. What does it mean if your probability density function ϕ is uniform? Formally,

$$\phi(x) = \frac{1}{\mathrm{vol}([-2,2]^2)} = \frac{1}{16}$$

for every $x \in \Omega$. Informally, since ϕ is constant, the heat map of ϕ would be a single colour. Example 8.4.11 implies that the probability you throw a dart in a region $A \subseteq \Omega$ is *directly proportional* to the area of A. You are therefore throwing uniformly at random. It is like closing your eyes when throwing the dart!

You are now equipped to study more complicated situations.

Example 8.4.13 Choose a vector (x,y,z) uniformly at random from inside the ball $B_3(0)$ in $\mathbb{R}^3$. How can you express the probability that $x+y+z > 0$? Use Definition 8.4.6 as your guide. Define the sample space $\Omega = B_3(0)$. Define the event space Σ to be all Jordan measurable subsets of Ω. The desired event is therefore

$$A = \{(x,y,z) \in B_3(0) : x+y+z > 0\}$$

but notice you must show that $A \in \Sigma$. It is immediate that $A \subseteq \Omega$ but you need to show A is

[6] Image taken from DataGenetics.com.

Jordan measurable. This is left as an exercise. Now, by definition, $\mathbb{P}$ is uniform so its density function is given by

$$\phi(x) = \frac{1}{\text{vol}(B_3(0))} = \frac{1}{36\pi}$$

so $(\Omega, \Sigma, \mathbb{P})$ is a continuous probability space in $\mathbb{R}^3$ by Theorem 8.4.7. Assuming A is an event in Σ, it follows that the desired probability is given by

$$\mathbb{P}(A) = \int_A \frac{1}{36\pi} dV = \frac{\text{vol}(A)}{36\pi}.$$

To actually calculate this probability, you will need some computational techniques.

You have proposed a rigorous definition for continuous probabilities using integrals in higher dimensional space. This suggestion connects elegantly to your theory of the integral and the Jordan measure, and satisfies quite a few desirable properties. There are, however, significant limitations.

8.4.3 Limitations of the Darboux integral and the Jordan measure

Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$ as in Theorem 8.4.7. Here are the key limitations of this definition.

1. *The sample space Ω is Jordan measurable and hence bounded.*

This prevents you from utilizing distributions like the Gaussian, exponential, logistic, or Gumbel which are defined on unbounded sets. Each of these are standard and naturally occur in all kinds of statistical and scientific phenomena, such as logistic regression in statistics, neural networks in data science, ideal gases in statistical mechanics, or harmonic oscillators in quantum mechanics.

2. *The sample space Ω is a subset of $\mathbb{R}^n$ for a finite $n \in \mathbb{N}^+$.*

This inhibits your ability to talk about limiting behaviour, infinitely many random variables, or anything more general than a vector of real numbers. For instance, you cannot establish standard results in probability and statistics like the strong law of large numbers, and the law of the iterated logarithm. You also cannot sensibly discuss more complicated random processes, like fractals or choosing "a random function", because these are cannot be naturally viewed as subsets of $\mathbb{R}^n$.

3. *The event space Σ of Jordan measurable subsets excludes many reasonable sets.*

This issue does not permit more subtle and intricate events. For instance, the set $[0,1] \cap \mathbb{Q}$ is not Jordan measurable yet you might more intuitively expect that its "volume" should be zero[7] and hence it should be an event. Aside from this strange example, this challenge is also a theoretical barrier to verifying whether a set A is an event or not. If you could permit a wider class of sets, then it would be easier to verify.

4. *The probability function $\mathbb{P}$ satisfies finite additivity, but not countable additivity.*

This concern arises from both the Darboux integral and Jordan measurable sets. If $\{A_k\}_{k=1}^{\infty}$ is a countable collection of pairwise disjoint events in Σ, then it is not necessarily true that

$$\mathbb{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

[7] Informally speaking, there are far fewer rationals compared to irrationals. Formally speaking, $\mathbb{Q}$ is countable whereas irrationals are not.

This identity fails for several reasons. First, the Darboux integral is not sufficiently robust to handle convergence issues arising from the infinite sum of integrals on the righthand side. Second, the countable union $\bigcup_{k=1}^{\infty} A_k$ is not necessarily an event in Σ. The property of countable additivity is essential for analyzing tail events like in Kolmogorov's zero-one law.

Collectively, this list of grievances may feel like a defeat as you conclude this chapter. Has all your hard work been for nothing? Not at all! The Darboux integral and Jordan measure were the historical predecessor to the resolution of these issues, namely the Lebesgue measure. This list therefore serves as inspiration for deeper investigations into the world of probability and measure theory, both of which are adventures beyond this text. There are still plenty of interesting problems and applications within reach of your existing results. You have amassed powerful technology on integrals that can be broadly applied in many scientific situations. However, one question has persisted throughout its development.

How do you efficiently compute integrals?

This inquiry will be your focus in the next chapter.

Exercises for Section 8.4

Concepts and definitions

8.4.1 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$ with probability density function ϕ. Which of the following are true or false? If true, briefly justify. If false, write a corrected statement.

(a) Σ is the set of all subsets of Ω.

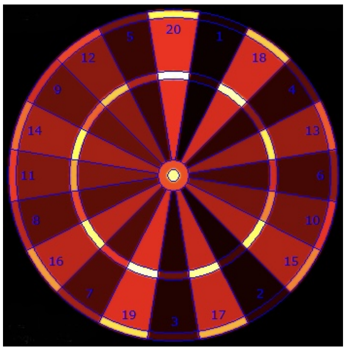
(b) $\mathbb{P}(A) = \displaystyle\int_A \phi \, dV$ for every subset $A \subseteq \Omega$.

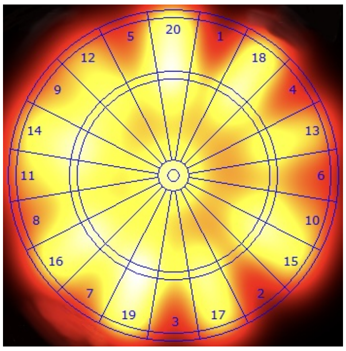
(c) $0 \leq \mathbb{P}(A) \leq 1$ for all events $A \in \Sigma$.

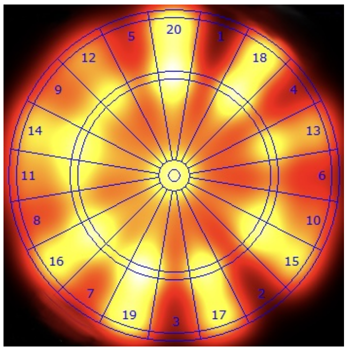
(d) $0 \leq \phi(x) \leq 1$ for all $x \in \Omega$.

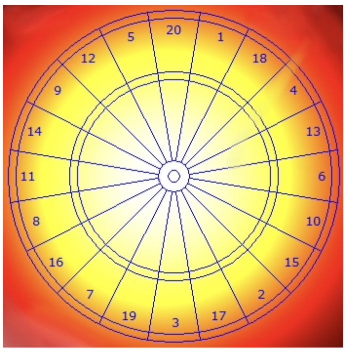
(e) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for every $A, B \in \Sigma$.

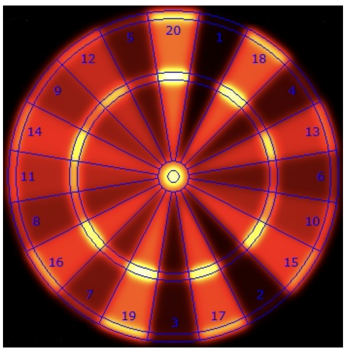
8.4.2 Six nerdy friends play games of darts. They have individually tracked their throws over thousands of games and generated a probability density function for each of them. Below are the heat maps [8]

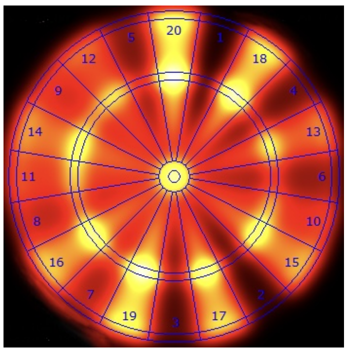

Abed


Brita


Jeff


Pierce


Shirley


Troy

(a) Who throws with the most precision?

(b) Whose probability density function is closest to uniform?

(c) Is Brita more likely to hit on the left or on the right?

(d) If the point $(0,0) \in \mathbb{R}^2$ represents the perfect bullseye on the dartboard, then who has the highest probability of hitting a perfect bullseye?

[8] Images modified from DataGenetics.com.

8.4.3 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}$ with sample space $\Omega = [-1, 1]$. Which of the following choices could be the probability density function for $\mathbb{P}$?

(a) $\phi_1(x) = \frac{1}{2}$ for $-1 \le x \le 1$

(b) $\phi_2(x) = x + \frac{1}{2}$ for $-1 \le x \le 1$.

(c) $\phi_3(x) = 100$ for $-\frac{1}{200} < x < \frac{1}{200}$ and $\phi_3(x) = 0$ for $|x| \ge \frac{1}{200}$.

(d) $\phi_4(x) = 0$ for $x \in [-1, 1] \cap \mathbb{Q}$ and $\phi_4(x) = \frac{1}{2}$ for $x \in [-1, 1] \cap \mathbb{Q}^c$.

8.4.4 A grocery store sells two different brands of pasta. The same grocery store also sells one brand of tomato sauce. You usually eat pasta with sauce, otherwise it's very dry and not tasty! Let $\phi(x, y)$ be the probability density function of the quantity x of pasta and quantity y of tomato sauce that they will sell in a week, measured in kg [9].

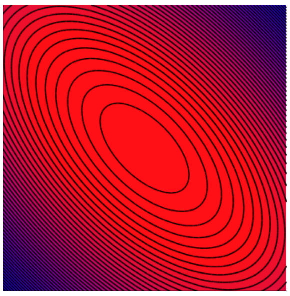


Diagram A

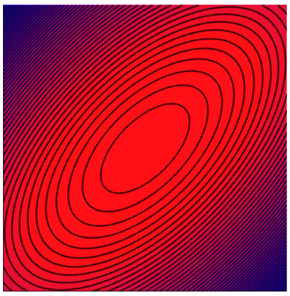


Diagram B

The two contour diagrams above are colored with red representing higher curves and blue representing lower curves. Which contour diagram corresponds to $\phi(x, y)$? Exactly one of them is correct.

8.4.5 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^2$ for selecting a vector inside the square $\Omega = [-7, 7]^2$. Assume the probability density function ϕ is arbitrary. Which events occur with probability zero?

(a) The vector lies inside the square.

(b) The vector lies outside the square.

(c) The vector lies on the edge of the square.

(d) The vector is the zero vector.

(e) The vector lies along the horizontal axis.

(f) The vector lies along the unit circle.

(g) The vector points upward but not directly up.

(h) The vector has magnitude between 1 and 2.

(i) The vector has two rational components.

(j) The vector has one rational component.

[9] Images created by Cindy Blois and used with permission.

8.4.6 Let $(\Omega, \Sigma, \mathbb{P})$ be the continuous probability space in $\mathbb{R}^3$ for choosing a vector inside the unit ball $B_1(0)$ with a uniform distribution.

(a) Define a set $A \subseteq \Omega$ that detects whether the vector lies in the upper hemisphere.

(b) Verify that $A \in \Sigma$.

(c) Conclude that $\mathbb{P}(A) = \frac{1}{2}$. No calculation required.

Proofs

8.4.7 Unlike discrete probabilities, the likelihood of a single outcome is always zero.

Lemma B. *Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. For $x \in \Omega$, $\mathbb{P}(\{x\}) = 0$.*

Fill in the blanks of the following incomplete proof by citing specific lemmas and including an additional detail.

1. *Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. Fix $x \in \Omega$.*
2. *The set $\{x\}$ has zero Jordan measure by* ______________________________.
3. ____.
4. *Therefore, by* ______________________________,

$$\mathbb{P}(\{x\}) = \int_{\{x\}} \phi \, dV = 0.$$

8.4.8 Celine is belting out a proof of the following lemma.

Lemma A. *Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. For $A \in \Sigma$, $0 \leq \mathbb{P}(A) \leq 1$.*

1. *By definition, $\Omega \subseteq \mathbb{R}^n$ is Jordan measurable.*
2. *Also, Σ is the collection of all Jordan measurable subsets of Ω.*
3. *Let $\phi : \Omega \to [0, \infty)$ be the probability density function of $\mathbb{P}$, so ϕ is bounded, continuous except on a zero volume set, and satisfies*

$$\int_{\Omega} \phi \, dV = 1.$$

4. *Fix $A \in \Sigma$. It is obvious that $\mathbb{P}(A) \geq 0$.*
5. *Moreover,*

$$\mathbb{P}(A) = \int_A \phi \, dV \leq \int_{\Omega} \phi \, dV = 1.$$

While there are no errors, Celine's "proof" is incomplete in addition to being a bit rude. You will fill in the details with additional details on integrals.

(a) Line 4 needs more details. Justify it with a property of integrals and one of your assumptions.

(b) Line 5 has an unexplained inequality. Explain it in more detail.

8.4.9 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. Using properties of Jordan measurable sets and zero Jordan measure, prove that if $A \in \Sigma$, then $\Omega \setminus A \in \Sigma$.

8.4.10 Let $(\Omega, \Sigma, \mathbb{P})$ be a continuous probability space in $\mathbb{R}^n$. Prove that if $A, B \in \Sigma$, then $A \cup B \in \Sigma$ and $A \cap B \in \Sigma$ and, moreover,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Applications and beyond

8.4.11 Random matrices possess a beautiful interplay between linear algebra, calculus, and probability. They also have remarkable applications across many scientific areas. Here you will study one such question.

> *The set of* 2×2 *matrices with all entries in* $[-100, 100]$ *can be viewed as a subset of* $\mathbb{R}^4$. *Choose a matrix* M *from this subset of* $\mathbb{R}^4$ *with a uniform probability function. What is the probability that the matrix* M *is* ***not*** *invertible?*

(a) Precisely define the probability space $(\Omega, \Sigma, \mathbb{P})$ needed to answer this question.

(b) Express the invertibility question in terms of $\mathbb{P}(A)$ for some well chosen event A.

(c) What do you conjecture is the value of $\mathbb{P}(A)$? Give a heuristic explanation.

8.4.12 The Darboux integral and Jordan measure have limitations for defining probability spaces. Construct a continuous probability space $(\Omega, \Sigma, \mathbb{P})$ in $\mathbb{R}^n$ such that $\{A_n\}_{n=1}^{\infty}$ is a collection of pairwise disjoint events in Σ but the countable union $\bigcup_{k=1}^{\infty} A_k$ is not an event in Σ.

9. Integration methods

In the previous chapters, you developed a robust theory for the integral in higher dimensions and interpreted its meaning in core applications. For a Jordan measurable set $S \subseteq \mathbb{R}^n$ and a bounded function $f : S \to \mathbb{R}$ continuous on S aside from a set of zero volume, the integral

$$\int_S f \, dV$$

naturally represents volumes, averages, mass, and probabilities. This accomplishment is monumental for multivariable calculus, and is poised to collect a bounty of applications. There is, however, a troubling obstacle.

How do you efficiently compute integrals?

The definition with partitions, upper sums, lower sums, upper integrals, and lower integrals seems far too complicated for everyday calculations. You will need substantially better computational methods to effectively apply your newfound theory.

In differential calculus, you successfully reduced higher dimensional derivatives to single variable derivatives. This triumphant strategy will continue here by reducing all of your calculations to single variable integrals and exploiting the *fundamental theorem of calculus*. Afterwards, inspired by linear algebra and determinants, you will add a new ingredient to calculate integrals by switching coordinate systems. By the end of this chapter, you will have developed powerful techniques to calculate integrals and applied them in an incredible variety of situations.

9.1. Fubini's theorem in 2D

To keep things simple, your computational endeavours with integrals will begin in two dimensions. In this section, you will uncover a foundational result that connects integrals in $\mathbb{R}^2$ to single-variable integrals, namely *Fubini's theorem*. Before you do so, you can take a moment to compare your "new" definition of integration over $\mathbb{R}$ with your "old" definition.

Lemma 9.1.1 Let $f : [a,b] \to \mathbb{R}$ be a bounded real-valued function on the interval $[a,b]$. If f is integrable on $[a,b]$ then $\displaystyle\int_{[a,b]} f\,dV = \int_a^b f(x)dx.$

Proof. There is nothing to prove here. Both integrals are defined in the exact same way with upper sums and lower sums. This "lemma" is written to bring your attention to the notational distinction because, while it does not matter over $\mathbb{R}$, it will matter for all higher dimensions. The integral on the left is using the notation from the previous chapter, whereas the one of the right is using classical notation for integrating on an interval from single variable calculus. ∎

The source of your computational power comes from single variable integrals and the fundamental theorem of calculus: if $f : [a,b] \to \mathbb{R}$ and $F : [a,b] \to \mathbb{R}$ are continuous and $F' = f$ on (a,b) then

$$F(b) - F(a) = \int_a^b f(x)dx.$$

Remember that the quantity on the left is defined using derivatives, whereas the quantity on the right represents net area. You have a trove of techniques to calculate antiderivatives (parts, substitution, partial fractions, trigonometric substitution, etc.), so the fundamental theorem of calculus creates a bridge for calculating single-variable integrals!

How can you use single variable integrals to integrate over two-dimensional sets?

You will propose a solution, namely iterated double integrals, and verify when this proposal is equivalent to the two-dimensional integral. Fubini's theorem will be the gateway.

9.1.1 Integrals of slices

To reduce integration over rectangles in $\mathbb{R}^2$ to single-variable calculus, you will use the standard strategy of slicing or, equivalently, fixing a variable.

Definition 9.1.2 Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a function.

- A **x-slice of f (at $x = x_0$)** is a function $g : [c,d] \to \mathbb{R}$ of the form
$$g(y) = f(x_0, y) \qquad \text{for some fixed } x_0 \in [a,b].$$
- A **y-slice of f (at $y = y_0$)** is a function $h : [a,b] \to \mathbb{R}$ of the form
$$h(x) = f(x, y_0) \qquad \text{for some fixed } y_0 \in [c,d].$$

Remark 9.1.3 Sometimes an x-slice of f will be denoted[1] $f^x : [c,d] \to \mathbb{R}$ where $f^x(y) = f(x,y)$ for $y \in [c,d]$. This abuses notation since the phrase "x-slice" treats x as a variable whereas the expression $f^x(y)$ treats x as a fixed value $x \in [a,b]$. Nonetheless, it will be convenient in some contexts, so be careful to pay attention to this subtle distinction. Similarly y-slices of f can be denoted $f^y : [a,b] \to \mathbb{R}$ where $f^y(x) = f(x,y)$ for $x \in [a,b]$.

You explored this technique in Section 1.2 using graphs of functions (see Definition 1.2.19 and Example 1.2.20). This new definition is for the *function* itself. The slice of a *two* variable function is a *one* variable function.

Example 9.1.4 Define $f : [-1,1] \times [0,2] \to \mathbb{R}$ by $f(x,y) = x^2 + y$. The x-slice of f at $x = 0$ is the function $g : [0,2] \to \mathbb{R}$ defined by $g(y) = f(0,y) = y$ for $y \in [0,2]$. Similarly, the y-slice of f at $y = 1$ is the function $h : [-1,1] \to \mathbb{R}$ defined by $h(x) = f(x,1) = x^2 + 1$ for $x \in [-1,1]$.

Some properties of functions naturally transfer to their slices.

Lemma 9.1.5 Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a function.
(a) If f is bounded, then every slice of f is bounded.
(b) If f is continuous, then every slice of f is continuous.

Proof. These are short exercises. Note f and a slice of f have different domains. ■

Now, if every x-slice of $f : [a,b] \times [c,d] \to \mathbb{R}$ is integrable on $[c,d]$, then for every $x \in [a,b]$, you can define $F : [a,b] \to \mathbb{R}$ by

$$F(x) = \int_c^d f(x,y)\,dy. \tag{9.1.1}$$

Similarly, if every y-slice of $f : [a,b] \times [c,d] \to \mathbb{R}$ is integrable on $[a,b]$, then for every $y \in [c,d]$, you can define $G : [c,d] \to \mathbb{R}$ by

$$G(y) = \int_a^b f(x,y)\,dx. \tag{9.1.2}$$

The fundamental theorem of calculus allows you to compute these integrals with ease.

Example 9.1.6 As before, define $f : [-1,1] \times [0,2] \to \mathbb{R}$ by $f(x,y) = x^2 + y$. Since f is continuous on $[-1,1] \times [0,2]$, every slice of f is continuous and hence integrable. In particular, the functions $F : [-1,1] \to \mathbb{R}$ and $G : [0,2] \to \mathbb{R}$ given by

$$F(x) = \int_0^2 f(x,y)\,dy, \qquad G(y) = \int_{-1}^1 f(x,y)\,dx$$

are well-defined, i.e. $F(x)$ is defined for $x \in [-1,1]$ and $G(y)$ is defined for $y \in [0,2]$. Thus,

$$F(0) = \int_0^2 f(0,y)\,dy = \int_0^2 y\,dy = \frac{y^2}{2}\Big|_{y=0}^{y=2} = 2$$

by the fundamental theorem of calculus. More generally, for any fixed $x \in [-1,1]$,

$$F(x) = \int_0^2 f(x,y)\,dy = \int_0^2 x^2 + y\,dy = \Big[x^2 y + \frac{y^2}{2}\Big]\Big|_{y=0}^{y=2} = 2x^2 + 2.$$

Notice x is treated as a constant when integrating since it is fixed. You can similarly verify that $G(y) = \frac{2}{3} + 2y$ for $0 \le y \le 2$. Notice both F and G are also continuous.

The functions F and G defined by integrals inherit properties from the integrand f.

Theorem 9.1.7 If $f : [a,b]\times[c,d] \to \mathbb{R}$ is continuous, then both F and G respectively defined by (9.1.1) and (9.1.2) are continuous.

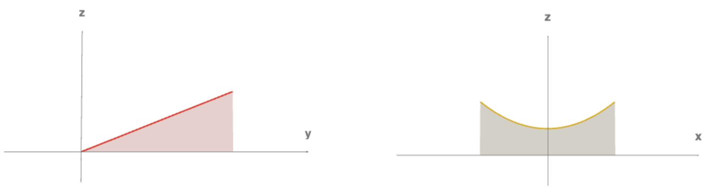
Proof. This follows from the fundamental theorem of calculus and Lemma 9.1.5. ■

Now, what do these functions defined as integrals geometrically represent in $\mathbb{R}^3$?

Example 9.1.8 Continue with the same $f : [-1,1]\times[0,2] \to \mathbb{R}$ given by $f(x,y) = x^2 + y$ and $F : [-1,1] \to \mathbb{R}$ and $G : [0,2] \to \mathbb{R}$ given by

$$F(x) = \int_0^2 f(x,y)dy, \qquad G(y) = \int_{-1}^1 f(x,y)dx.$$

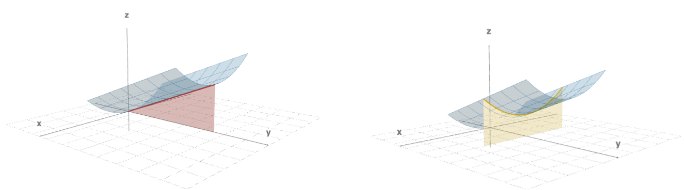
In $\mathbb{R}^2$, the integral $F(0)$ represents the area under the curve $z = f(0,y) = y$, and the integral $G(1)$ represents the area under the curve $z = f(x,1) = x^2 + 1$.



In $\mathbb{R}^3$, these regions respectively correspond to the sets S and T in $\mathbb{R}^3$ given by

$$S = \{(0,y,z) \in \mathbb{R}^3 : 0 \le y \le 2, 0 \le z \le f(0,y)\},$$
$$T = \{(x,1,z) \in \mathbb{R}^3 : -1 \le x \le 1, 0 \le z \le f(x,1)\}.$$

View this Math3D demo for a visual.



Varying x from -1 to 1, the lefthand slices "sweep out" the solid under the graph $z = f(x,y)$. Varying y from 0 to 2, the righthand slices "sweep out" the same solid in a perpendicular fashion. You have just discovered a critical observation!

$$\textit{Do } \int_{-1}^1 F(x)dx \textit{ and } \int_0^2 G(y)dy \textit{ both exist and equal } \iint_{[-1,1]\times[0,2]} f\,dA?$$

Equivalently you may ask:

$$\textit{Do } \int_{-1}^1 \Big(\int_0^2 f(x,y)dy\Big)dx \textit{ and } \int_0^2 \Big(\int_{-1}^1 f(x,y)dx\Big)dy \textit{ exist and equal } \iint_{[-1,1]\times[0,2]} f\,dA?$$

After a tedious calculation with upper sums and lower sums, you can manually verify that

$$\iint_{[-1,1]\times[0,2]} f\,dA = \frac{16}{3}.$$

Alternatively, by the fundamental theorem of calculus and Example 9.1.6, it follows that

$$\int_{-1}^{1} F(x)dx = \int_{-1}^{1} 2x^2 + 2dx = \frac{16}{3}, \quad \text{and} \quad \int_{0}^{2} G(y)dy = \int_{0}^{2} \frac{2}{3} + 2ydy = \frac{16}{3}$$

as well. This gives evidence that the answer should be "yes", but there is a conceptual problem: each slice has *zero volume* in $\mathbb{R}^3$! They are two-dimensional surfaces living in $\mathbb{R}^3$. How can the volume of a solid in $\mathbb{R}^3$ be calculated as the total of zero volume slices?

9.1.2 Iterated double integrals

By calculating single variable integrals of single variable integrals, you have discovered a potential avenue to efficiently calculate two-dimensional integrals.

Definition 9.1.9 Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be a bounded function. The quantities

$$\int_a^b \Big(\int_c^d f(x,y)dy \Big) dx, \qquad \int_c^d \Big(\int_a^b f(x,y)dx \Big) dy$$

are **iterated double integrals**.

Remark 9.1.10 The lefthand iterated integral exists if and only if every x-slice of f is integrable on $[c,d]$ and the function $F(x) = \int_c^d f(x,y)dy$ is integrable on $[a,b]$.

Example 9.1.11 Let $f : [-1,1] \times [0,2] \to \mathbb{R}$ be given by $f(x,y) = x^2 + y$. Applying the fundamental theorem of calculus several times,

$$\begin{aligned} \int_{-1}^{1} \Big(\int_{0}^{2} f(x,y)dy \Big) dx &= \int_{-1}^{1} \Big(\int_{0}^{2} x^2 + ydy \Big) dx \\ &= \int_{-1}^{1} \Big[x^2 y + \frac{y^2}{2} \Big]_{y=0}^{y=2} dx = \int_{-1}^{1} 2x^2 + 2dx = \frac{16}{3}. \end{aligned}$$

You can formalize the central question. Let $R = [a,b] \times [c,d]$ and $f : R \to \mathbb{R}$ be bounded. Assume every x-slice of f is integrable on $[c,d]$ and every y-slice of f is integrable on $[a,b]$.

$$\textit{When do } \int_a^b \Big(\int_c^d f(x,y)dy \Big) dx \textit{ and } \int_c^d \Big(\int_a^b f(x,y)dx \Big) dy \textit{ exist and equal } \iint_R f\, dA?$$

Sadly, there is no obvious relationship between these three integrals [18, Chapter 7].

Example 9.1.12 Let $R = [0,1]^2$ and define

$$f(x,y) = \begin{cases} 1 & \text{if } x = 0, y \in \mathbb{Q}, \\ 0 & \text{otherwise} \end{cases}$$

so $\int_0^1 f(0,y)dy$ does not exist implying the iterated integral $\int_0^1 \int_0^1 f(x,y)dydx$ does not exist. On the other hand, f is the indicator function for $\{0\} \times ([0,1] \cap \mathbb{Q})$ which has zero volume. Hence, f is integrable on R and $\iint_R f\, dA = 0$.

Example 9.1.13 There also exists an integrable function where neither iterated integral exists. It only requires a small variation of the previous example, so this is left as an exercise.

Example 9.1.14 Let $R=[0,1]^2$ and define

$$f(x,y)=\begin{cases}1 & \text{if } y\in\mathbb{Q},\\ 2x & \text{if } y\notin\mathbb{Q}.\end{cases}$$

For any $y\in\mathbb{R}$, notice that $\int_0^1 f(x,y)dx=1$, so $\int_0^1\int_0^1 f(x,y)dxdy=1$. It is more challenging to verify that f is not integrable on R. View this Math3D demo illustrating how the lower integral and upper integral would be calculated.

Example 9.1.15 The previous example also shows that one iterated integral may exist whereas the other may not. Indeed, notice $\int_0^1\int_0^1 f(x,y)dxdy$ equals one and hence exists whereas you can verify that $\int_0^1\int_0^1 f(x,y)dydx$ does not exist. This is left as a short exercise.

Example 9.1.16 Let $R=[0,1]^2$ and define

$$f(x,y)=\begin{cases}1 & \text{if } x=\frac{m}{q}, y=\frac{n}{q} \text{ for some } m,n,q\in\mathbb{N}^+ \text{ with } q \text{ prime}\\ 0 & \text{otherwise.}\end{cases}$$

It is not easy but you can prove that $L_P(f)=0$ and $U_P(f)=1$ for every partition P of R. It follows that f is not integrable on R. On the other hand, if $x\notin\mathbb{Q}$, then $f(x,y)=0$ for all $y\in\mathbb{R}$ so $\int_0^1 f(x,y)dy=0$. If $x\in\mathbb{Q}$ (e.g. $x=\frac{1}{7}$) then there are only *finitely* many other $y\in[0,1]$ such that $f(x,y)=1$ (e.g. $y=\frac{1}{7},\frac{2}{7},\cdots,\frac{7}{7}$) and otherwise $f(x,y)=0$, so $\int_0^1 f(x,y)dy=0$ again. Thus, $\int_0^1\int_0^1 f(x,y)dydx=\int_0^1 0dx=0$. The same holds for the other iterated integral. Thus, both iterated integrals exist yet f may not be integrable!

These perverse examples are rather disappointing. The existence of one integral does not necessarily imply the existence of others. Do not fear, for Fubini is here.

9.1.3 Fubini's theorem in two dimensions

Fubini's theorem gives an elegant criterion when these three integrals all exist and are all equal.

Theorem 9.1.17 (Fubini) Let $R=[a,b]\times[c,d]$ and let $f:R\to\mathbb{R}$ be bounded. For $x\in[a,b]$, define $f^x:[c,d]\to\mathbb{R}$ by $f^x(y)=f(x,y)$. Assume

- f^x is integrable on $[c,d]$ for every $x\in[a,b]$.
- f is integrable on $[a,b]\times[c,d]$.

Then

- $\displaystyle\int_c^d f(x,y)dy$ exists for every $x\in[a,b]$.
- $\displaystyle\int_a^b\left(\int_c^d f(x,y)dy\right)dx$ exists and equals $\displaystyle\iint_R f\,dA$.

Remark 9.1.18 You can also rewrite Fubini's theorem with y-slices instead of x-slices.

The proof is temporarily postponed to Section 9.1.4. Notice the assumptions are based on integrability. You have powerful theorems to verify integrability (Theorems 7.6.13 and 7.7.4) so Fubini's theorem is fantastically useful! Also, you obtain an elegant corollary.

Corollary 9.1.19 (Fubini) If $f : [a,b]\times[c,d]\to\mathbb{R}$ is continuous, then

$$\int_a^b \Big(\int_c^d f(x,y)dy\Big)dx \quad \text{and} \quad \int_c^d \Big(\int_a^b f(x,y)dx\Big)dy$$

both exist and are equal to $\displaystyle\iint_{[a,b]\times[c,d]} f\,dA$.

This corollary can be applied like butter.

Example 9.1.20 Define $R = [-1,1]\times[0,2]$ and $f(x,y) = x^2+y$. Since f is continuous on the rectangle R, it follows by Fubini's theorem (Corollary 9.1.19) that

$$\iint_R f\,dA = \int_0^2\int_{-1}^1 x^2+y\,dx\,dy = \int_{-1}^1\int_0^2 x^2+y\,dy\,dx.$$

These latter two integrals can be computed easily as in Example 9.1.11.

Fubini's theorem (Theorem 9.1.17) allows you to write iterated integrals over non-rectangular regions, but this takes more effort to implement.

Example 9.1.21 Define $S = \{(x,y)\in\mathbb{R}^2 : 0\le x\le 1, 0\le y\le x\}$. Here it is proved that

$$\iint_S e^{xy}\,dA = \int_0^1\int_0^x e^{xy}\,dy\,dx.$$

You can check that S is Jordan measurable since it is bounded inside $R = [0,1]\times[0,1]$ and ∂S is the triangle with vertices $(0,0),(1,0)$ and $(1,1)$ which has zero volume by Sard's theorem (Theorem 7.6.13) and properties of zero volume sets. Since e^{xy} is continuous on S, it follows by Theorem 7.7.4 that e^{xy} is integrable on S and hence $\iint_S e^{xy}\,dA$ exists. Define $f:\mathbb{R}^2\to\mathbb{R}$ by

$$f(x,y) = \chi_S(x,y)e^{xy} \quad \text{for } (x,y)\in\mathbb{R}^2.$$

Temporarily fix $x\in[0,1]$ and define $f^x : [0,1]\to\mathbb{R}$ given by $f^x(y) = f(x,y)$. Notice f^x satisfies by definition of S that

$$f^x(y) = \chi_S(x,y)e^{xy} = \chi_{[0,x]}(y)e^{xy}.$$

Therefore $f^x(y)$ is continuous on $[0,1]$ except at $y=x$ (a set of zero Jordan measure). By Theorem 7.7.17, f^x is integrable on $[0,1]$ for every $x\in[0,1]$.

This verifies all the assumptions of Fubini's theorem (Theorem 9.1.17) so the iterated integral $dy\,dx$ exists and

$$\begin{aligned}\iint_S e^{xy}\,dA = \iint_R f\,dA &= \int_0^1\int_0^1 f(x,y)\,dy\,dx \\ &= \int_0^1\int_0^1 \chi_{[0,x]}(y)e^{xy}\,dy\,dx = \int_0^1\int_0^x e^{xy}\,dy\,dx.\end{aligned}$$

Remark 9.1.22 In Example 9.1.21, notice how you verified all the integrability conditions using some combination of Theorems 7.6.13, 7.7.4 and 7.7.17. This illustrates again how powerful those theorems can be.

That seems like quite a lot of work to check what seems obvious. Do not worry. You can deduce some easy-to-use theorems from Fubini's theorem that apply in many cases. Moreover, for basic computations, you will usually be permitted (unless specified otherwise) to use Fubini's theorem without justification.

9.1.4 Proof of Fubini's theorem in 2D

The proof of Fubini's theorem is quite technically sophisticated and uses many properties of suprema and infima.

Proof of Theorem 9.1.17. Fix $x \in [a,b]$ and define $f^x : [c,d] \to \mathbb{R}$ by $f^x(y) = f(x,y)$. Since f^x is integrable on $[c,d]$, it follows by definition that the integral

$$\int_c^d f^x(y)dy = \int_c^d f(x,y)dy$$

exists for $x \in [a,b]$. This proves the first item. For the second item, define $F : [a,b] \to \mathbb{R}$ by

$$F(x) = \int_c^d f(x,y)dy.$$

We wish to show that F is integrable on $[a,b]$ and moreover

$$\int_a^b F(x)dx = \int_a^b \int_c^d f(x,y)dy\,dx = \iint_R f\,dA.$$

Fix $\varepsilon > 0$. Since f is integrable on R, by the ε-characterization of integrability, there exists a partition $P = \{R_{ij}\}_{ij}$ of R such that

$$0 \le U_P(f) - L_P(f) < \varepsilon. \tag{9.1.3}$$

Since $\iint_R f\,dA$ is between $U_P(f)$ and $L_P(f)$, this implies

$$U_P(f) - \varepsilon \le \iint_R f\,dA \le L_P(f) + \varepsilon. \tag{9.1.4}$$

Now, P is constructed from partitions $P_1 = \{x_0,\dots,x_k\}$ of $[a,b]$ and $P_2 = \{y_0,\dots,y_\ell\}$ of $[c,d]$. We estimate the value $F(x)$ with the P_2-upper sum for the integral over $[c,d]$. For $x \in [a,b]$,

$$F(x) \le \sum_{j=1}^{\ell} \Big(\sup_{y\in[y_{j-1},y_j]} f(x,y) \Big) \Delta y_j.$$

For $1 \le i \le k$, we estimate F on the subinterval $[x_{i-1},x_i]$. The above inequality implies that

$$\begin{aligned}
\sup_{x\in[x_{i-1},x_i]} F(x) &\le \sup_{x\in[x_{i-1},x_i]} \left(\sum_{j=1}^{\ell} \Big(\sup_{y\in[y_{j-1},y_j]} f(x,y) \Big) \Delta y_j \right) \\
&\le \sum_{j=1}^{\ell} \sup_{x\in[x_{i-1},x_i]} \left(\Big(\sup_{y\in[y_{j-1},y_j]} f(x,y) \Big) \Delta y_j \right) \\
&\le \sum_{j=1}^{\ell} \left(\sup_{x\in[x_{i-1},x_i]} \Big(\sup_{y\in[y_{j-1},y_j]} f(x,y) \Big) \right) \Delta y_j && \text{as } \Delta y_j > 0 \\
&\le \sum_{j=1}^{\ell} \Big(\sup_{(x,y)\in R_{ij}} f(x,y) \Big) \Delta y_j = \sum_{j=1}^{\ell} M_{ij} \Delta y_j.
\end{aligned}$$

The second inequality uses this property of suprema for bounded functions g and h on $[x_{i-1}, x_i]$:

$$\sup_{x\in[x_{i-1},x_i]} (g+h)(x) \leq \sup_{x\in[x_{i-1},x_i]} g(x) + \sup_{x\in[x_{i-1},x_i]} h(x).$$

Multiplying the prior inequality by Δx_i and summing over $1 \leq i \leq k$, it follows that

$$U_{P_1}(F) = \sum_{i=1}^{k} \Big(\sup_{x\in[x_{i-1},x_i]} F(x) \Big) \Delta x_i \leq \sum_{i=1}^{k}\sum_{j=1}^{\ell} M_{ij} \Delta x_i \Delta y_j = U_P(f).$$

Similarly,

$$L_{P_1}(F) = \sum_{i=1}^{k} \Big(\inf_{x\in[x_{i-1},x_i]} F(x) \Big) \Delta x_i \geq \sum_{i=1}^{k}\sum_{j=1}^{\ell} m_{ij} \Delta x_i \Delta y_j = L_P(f)$$

From (9.1.3), we conclude that

$$U_{P_1}(F) - L_{P_1}(F) \leq U_P(f) - L_P(f) < \varepsilon$$

so F is integrable on $[a, b]$ and $\int_a^b F(x)dx$ exists by the ε-characterization for integrability. From the two prior inequalities and (9.1.4), it follows that

$$\iint_R f\,dA - \varepsilon \leq L_P(f) \leq L_{P_1}(F) \leq \int_a^b F(x)dx \leq U_{P_1}(F) \leq U_P(f) \leq \iint_R f\,dA + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this proves the integrals are equal. ■

Fubini's theorem in two dimensions is a major breakthrough for integration. You found three different integrals that you could compute on a rectangle $R = [a, b] \times [c, d]$. The first integral is the original one, namely

$$\iint_R f\,dA.$$

Its definition and properties are geometrically natural and it represents exactly what you would expect, e.g. volume, total value, average. You also have a flexible criterion to check when this integral exists. There are also two iterated integrals you can compute:

$$\int_c^d \Big(\int_a^b f(x, y)dx \Big) dy \qquad\qquad \int_a^b \Big(\int_c^d f(x, y)dy \Big) dx.$$

These can often be easily computed by applying single variable calculus techniques and the fundamental theorem of calculus. Fubini's theorem verifies all three of these integrals represent the exact same quantity so you can get the best of both worlds: natural properties, flexible criterion for existence, and computational ease. This connection prepares you to finally build computational integration techniques in two dimensions.

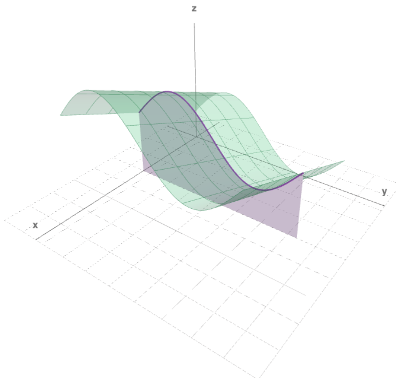
Exercises for Section 9.1

Concepts and definitions

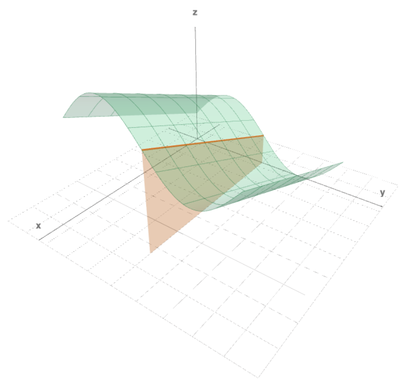
9.1.1 Define $\varphi(x,y) = \sin(\pi y) + x + 1$. Match the pair of integrals

$$I = \int_0^2 \varphi(1,y)\,dy, \qquad J = \int_0^2 \varphi(x,1)\,dx$$

to the corresponding sliced region in $\mathbb{R}^3$ under the graph $z = \varphi(x,y)$.

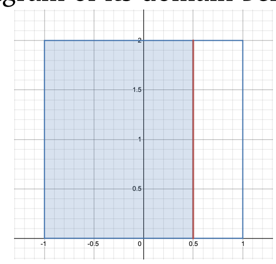


Region A



Region B

9.1.2 Let $\varphi : [-1,1] \times [0,2] \to \mathbb{R}$ be continuous. Consider the diagram of its domain below.



(a) Write the integral of φ over the red vertical line.

(b) Write the integral of φ over the blue shaded region.

9.1.3 Let $R = [0,2] \times [3,7]$ and $f : R \to \mathbb{R}$ be bounded. Which statements are true or false? If true, justify with a theorem or definition. If false, cite a counterexample.

(a) If f is continuous on R, then $\int_0^2 \int_3^7 f(x,y)\,dy\,dx$, $\int_3^7 \int_0^2 f(x,y)\,dx\,dy$, $\int_R f\,dA$ exist and are equal.

(b) If f is integrable on R, then both $\int_0^2 \int_3^7 f(x,y)\,dy\,dx$ and $\int_3^7 \int_0^2 f(x,y)\,dx\,dy$ exist.

(c) If f is integrable on R, then $\int_R f\,dA$ exists.

(d) If $\int_0^2 \int_3^7 f(x,y)\,dy\,dx$ exists, then f is integrable on R.

(e) If $\int_0^2 \int_3^7 f(x,y)\,dy\,dx$ exists, then $\int_3^7 \int_0^2 f(x,y)\,dx\,dy$ exists.

(f) If $\int_0^2 \int_3^7 f(x,y)\,dy\,dx$ and $\int_3^7 \int_0^2 f(x,y)\,dx\,dy$ both exist, then f is integrable on R.

9.1.4 Let $R=[a,b]\times[c,d]$ be a rectangle in $\mathbb{R}^2$. Let $\varphi : R\to\mathbb{R}$ be a bounded function. Consider 3 different assumptions you might use to apply Fubini's theorem in $\mathbb{R}^2$.

(A) φ is integrable on $[a,b]\times[c,d]$.

(B) For every $x\in[a,b]$, the x-slice φ^x is integrable on $[c,d]$

(C) For every $y\in[c,d]$, the y-slice φ^y is integrable on $[a,b]$

(a) $\displaystyle\int_a^b \varphi(x,y)dx$ exists provided which assumption(s) hold?

(b) $\displaystyle\int_c^d\int_a^b \varphi(x,y)dxdy$ exists provided which assumption(s) hold?

(c) $\displaystyle\int_R \varphi dV=\int_c^d\int_a^b \varphi(x,y)dxdy$ provided which assumption(s) hold?

(d) $\displaystyle\int_R \varphi dV=\int_c^d\int_a^b \varphi(x,y)dxdy=\int_a^b\int_c^d \varphi(x,y)dydx$ provided which assumption(s) hold?

9.1.5 Let $f:[a,b]\times[c,d]\to\mathbb{R}$ be a function. True or false? If true, prove it. If false, give a counterexample. (For the true statements, write out the formal definitions with quantifiers; the proof will be very short.)

(a) If f is bounded, then every x-slice of f is bounded.

(b) If f is continuous, then every x-slice of f is continuous.

(c) If f is integrable, then every x-slice of f is integrable.

Computations

9.1.6 Evaluate each of the following integrals. *Hint:* You can use Fubini twice.

(a) $\displaystyle\int_0^1\int_0^2 xy^2\,dxdy$

(b) $\displaystyle\int_0^1\int_0^2 xy^2\,dydx$

(c) $\displaystyle\int_0^2\int_0^1 xy^2\,dydx$

(d) $\displaystyle\iint_{[0,1]\times[0,2]} xy^2\,dA$

Proofs

9.1.7 If the integrand is not continuous, you must carefully verify several assumptions to apply Fubini's theorem. Define $R=[0,2]\times[4,6]$ and $\varphi:R\to\mathbb{R}$ by

$$\varphi(x,y)=\begin{cases}-1 & 0\le x\le 2, 4\le y<5,\\ 3 & 0\le x\le 2, 5\le y\le 6.\end{cases}$$

(a) What are the y-slices of φ? Prove that they are integrable on their domains.

(b) What are the x-slices of φ? Prove that they are integrable on their domains.

(c) Prove that φ is integrable on R.

(d) Conclude that $\displaystyle\int_0^2\int_4^6 \varphi(x,y)dydx$ exists and $\displaystyle\iint_R \varphi dA=\int_0^2\int_4^6 \varphi(x,y)dydx$.

(e) Conclude that $\int_4^6 \int_0^2 \varphi(x,y)dxdy$ exists and $\iint_R \varphi dA = \int_4^6 \int_0^2 \varphi(x,y)dxdy$.

(f) Evaluate $\iint_R \varphi dA$ by either choice of iterated double integral.

9.1.8 To apply Fubini's theorem in two dimensions, you have to verify several assumptions. Suppose you want to calculate the integral of $\varphi(x,y) = x\cos y$ over the region

$$S = \{(x,y) \in \mathbb{R}^2 : 1 \le y \le 3, \cos y \le x \le \sin y\}.$$

To compute the integral of φ on S, you want to express it as an iterated integral by Fubini's theorem. View this Desmos graph to help you visualize.

(a) Fill in the blank: "*I want to show by Fubini's theorem that* $\int_S \varphi dA =$ ________________."

(b) Fubini's theorem is only stated for rectangles so you must rephrase the previous statement.

Fill in the blank: "*Equivalently, I want to show by Fubini's theorem that* $\int_R \chi_S \varphi dA =$ ___________, *where* $R =$ ___________ *is a rectangle containing* S."

(c) The **red line** in Desmos corresponds to an integral of a slice of $\chi_S\varphi$. Write this integral for an *arbitrary* slice and define the slice as a 1-variable function.

(d) Move the slider in Desmos to generate a **blue shaded region**. This animation corresponds to the integral of a 1-variable function. Define the 1-variable function and the integral.

(e) Now, create the checklist of assumptions that you must verify to apply Fubini's theorem.

- ☐ The 2-variable function _____ is integrable on R.
  That is, the integral _____ exists.
- ☐ The 1-variable functions _____ are integrable on _____.
  That is, the integrals _____ all exist.

(f) After verifying the above assumptions of Fubini's theorem, fill in your conclusion.

- The 1-variable function _____ is integrable on _____.
  That is, the double integral _____ exists.
- Moreover, this double integral is equal to _____.

(g) Cameron claims *"If I prove φ is continuous on S and S is Jordan measurable, I can apply Fubini's theorem."* What is wrong with Cameron's reasoning?

(h) Verify the assumptions in Exercise 9.1.8.5 and apply Fubini's theorem.

Applications and beyond

9.1.9 The proof of Fubini's theorem in 2D is a technical yet natural sequence of steps. The most challenging part of the proof is a sequence of inequalities involving suprema. Here you will identify exactly what properties of suprema are used without mention. The excerpt below skips forward to the relevant part. Do not worry about the definition of $F(x)$ or what comes before or after this excerpt.

1. *Now, the partition $P = \{R_{ij}\}_{i,j}$ of the rectangle $R = [a,b] \times [c,d]$ is constructed from partitions $P_1 = \{x_0, \ldots, x_k\}$ of $[a,b]$ and $P_2 = \{y_0, \ldots, y_\ell\}$ of $[c,d]$.*
2. *The value of $F(x)$ is estimated by the P_2-upper sum for the integral over $[c,d]$. For any $x \in [a,b]$,*
$$F(x) \le \sum_{j=1}^{\ell} \Big(\sup_{y \in [y_{j-1}, y_j]} f(x,y) \Big) \Delta y_j$$
3. *Next, for $1 \le i \le k$, we estimate F on the subinterval $[x_{i-1}, x_i]$. The above inequality implies that*

$$\sup_{x \in [x_{i-1}, x_i]} F(x) \le \sup_{x \in [x_{i-1}, x_i]} \left(\sum_{j=1}^{\ell} \Big(\sup_{y \in [y_{j-1}, y_j]} f(x,y) \Big) \Delta y_j \right) \tag{9.1.5}$$
$$\le \sum_{j=1}^{\ell} \sup_{x \in [x_{i-1}, x_i]} \left(\Big(\sup_{y \in [y_{j-1}, y_j]} f(x,y) \Big) \Delta y_j \right) \tag{9.1.6}$$
$$\le \sum_{j=1}^{\ell} \left(\sup_{x \in [x_{i-1}, x_i]} \Big(\sup_{y \in [y_{j-1}, y_j]} f(x,y) \Big) \right) \Delta y_j \qquad \text{as } \Delta y_j > 0 \tag{9.1.7}$$
$$\le \sum_{j=1}^{\ell} \Big(\sup_{(x,y) \in R_{ij}} f(x,y) \Big) \Delta y_j = \sum_{j=1}^{\ell} M_{ij} \Delta y_j. \tag{9.1.8}$$

Let g and h be bounded functions on $[x_{i-1}, x_i]$. For each part, state the desired property using g and h. Be as precise as possible.

(a) Equation (9.1.5) follows from Line 2 by which property of suprema?

(b) Equation (9.1.6) follows from equation (9.1.5) by which property of suprema?

(c) Equation (9.1.7) follows from equation (9.1.6) by which property of suprema?

(d) Equation (9.1.8) follows from equation (9.1.7) by which property of suprema? You will need to refer to f directly here.

9.2. Double integrals

It is finally time to develop strategies for calculating iterated double integrals and identifying their symmetries. Your former practice with single-variable integration techniques will be handy both technically and experientially; you may remember that integration techniques are not executed as a step-by-step algorithm. Multivariable integrals are no different. There will also be new challenges, especially with the geometry of two-dimensional regions and describing them using iterated integrals.

Much like optimization from differential calculus, there is no fixed algorithm that will always succeed in evaluating or manipulating iterated double integrals. This section will describe some possible approaches in selected examples. You are encouraged to focus on how these core tools operate, rather than trying to classify the problems. Remember that you will apply Fubini's theorem without justification.

First, you can directly calculate with the fundamental theorem of calculus.

Example 9.2.1 Suppose you want to find the mass of a rectangular sheet $R = [0,2]\times[-2,-1]$ with mass density $\varphi(x,y) = e^{x-y}$. By definition of R and Fubini's theorem, it follows that the mass m is equal to that

$$m = \iint_R e^{x-y}\,dA = \int_0^2 \int_{-2}^{-1} e^{x-y}\,dy\,dx = \int_{-2}^{-1}\int_0^2 e^{x-y}\,dx\,dy.$$

This gives two possible ways to evaluate the integral. You can calculate directly $dy\,dx$, so

$$\begin{aligned}\iint_R e^{x-y}\,dA = \int_0^2 \int_{-2}^{-1} e^{x-y}\,dy\,dx &= \int_0^2 -e^{x-y}\Big|_{y=-2}^{y=-1} dx\\ &= \int_0^2 e^x(e^2-e)\,dx\\ &= (e^2-1)(e^2-e).\end{aligned}$$

Equivalently, you can calculate directly $dx\,dy$ and obtain that

$$\begin{aligned}\iint_R e^{x-y}\,dA = \int_{-2}^{-1}\int_0^2 e^{x-y}\,dx\,dy &= \int_{-2}^{-1}(e^2-1)e^{-y}\,dy\\ &= (e^2-1)(e^2-e).\end{aligned}$$

The answers are unsurprisingly the same.

This prior example was straightforward since the region was a rectangle. Often the region will be *non*-rectangular. This can complicate things even in simple scenarios. Luckily, you have a brand new technique: *swapping the order of integration!*

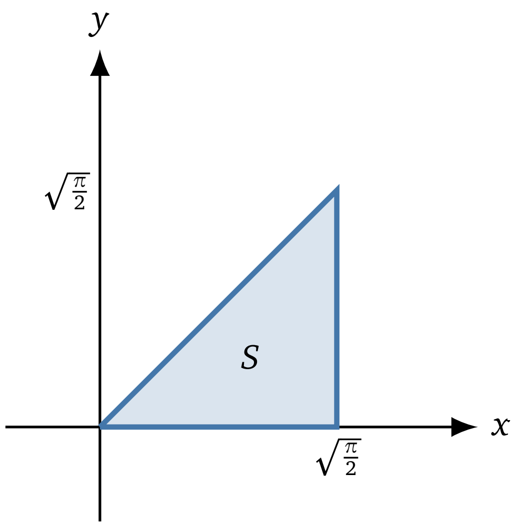
Example 9.2.2 Suppose you want to calculate the volume below the graph $z = \sin(x^2)$ and above $z = 0$ in $\mathbb{R}^3$ for $0 \le y \le \sqrt{\frac{\pi}{2}}$ and $y \le x \le \sqrt{\frac{\pi}{2}}$. You therefore wish to compute

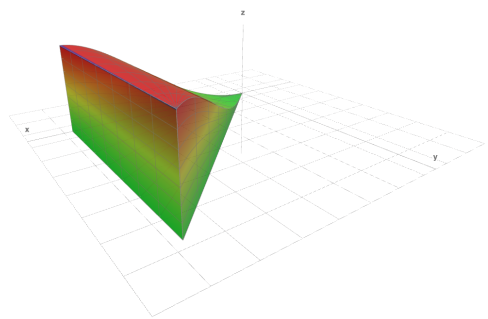
$$I = \iint_S \sin(x^2)\,dA,$$

where

$$S = \left\{(x,y)\in\mathbb{R}^2 : 0 \le y \le \sqrt{\tfrac{\pi}{2}},\, y \le x \le \sqrt{\tfrac{\pi}{2}}\right\}.$$

It is always helpful to sketch your region of integration and, if possible, use software to sketch the solid. See the region S and this Math3D graph of the solid.





Notice S can be naturally described with y-slices, so this integral can be rewritten as an iterated double integral

$$I = \int_0^{\sqrt{\frac{\pi}{2}}} \int_y^{\sqrt{\frac{\pi}{2}}} \sin(x^2) dx dy,$$

but this integral is impossible to evaluate in its current form as $\sin(x^2)$ does not have an elementary antiderivative. This seems daunting, but notice that S has a special property: you can also describe it with x-slices. This means that you can apply Fubini's theorem to swap the order of integration! Namely, since you can rewrite the triangular region S as

$$S = \left\{(x, y) \in \mathbb{R}^2 : 0 \le y \le x \text{ and } 0 \le x \le \sqrt{\frac{\pi}{2}}\right\},$$

it follows by Fubini's theorem that

$$I = \int_0^{\sqrt{\frac{\pi}{2}}} \int_0^x \sin(x^2) dy dx.$$

Miraculously, this is now tractable. You can conclude that

$$I = \int_0^{\sqrt{\frac{\pi}{2}}} \int_0^x \sin(x^2) dy dx = \int_0^{\sqrt{\frac{\pi}{2}}} x \sin(x^2) dx = -\frac{1}{2} \cos(x^2)\Big|_{x=0}^{x=\sqrt{\frac{\pi}{2}}} = \frac{1}{2}.$$

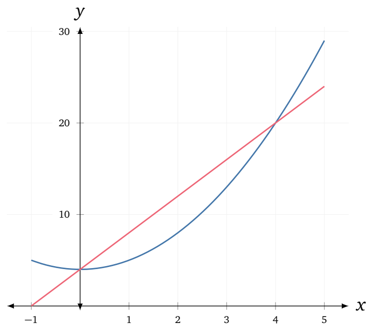
Remark 9.2.3 Remember that Fubini's theorem as stated in Section 9.1 applies only to rectangular regions. To justify the swap in Example 9.2.2, you would need to veriy Fubini's assumptions for the function $\chi_S(x, y) \sin(x^2)$ on a rectangle containing the region S. As usual, you will not check this unless specifically asked.

Determining the region of integration can itself be a challenge.

Example 9.2.4 Suppose you want to compute

$$\iint_S \frac{x-y}{x+y}\,dA$$

where S is the region bounded by the curves $y = x^2+4$ and $y = 4x+4$. You will want to express the region $S \subseteq \mathbb{R}^2$ so that you can write the integral as an iterated double integral. As a matter of principle, you should first sketch the region whenever possible.



You can see that the line is always above the parabola for the region S, that is,

$$x^2+4 \le y \le 4x+4$$

whenever $(x,y) \in S$. Next, you will need to know the intersection points of the curves to describe the region S. Notice that $x^2+4 = 4x+4$ if and only if $x=0$ or $x=4$, so $(0,4)$ and $(4,20)$ are the intersection points. From the sketch, it therefore follows that

$$S = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 4, x^2+4 \le y \le 4x+4\}.$$

Hence, by Fubini's theorem,

$$\iint_S \frac{x-y}{x+y}\,dA = \int_0^4 \int_{x^2+4}^{4x+4} \frac{x-y}{x+y}\,dy\,dx.$$

By a similar process, you can express the same integral as an iterated double integral $dx\,dy$; this is left as an exercise.

Reflecting on Example 9.2.4, you may notice that the set description of the region was not rigorously explained. The sketch acts as informal justification, which is acceptable when the curves are simple enough. By carefully manipulating inequalities, it is feasible to formally prove the region of integration is equal to the claimed set. This level of formality is too tedious so a well-labelled and accurate sketch will usually suffice.

Thus far, you have integrated over sets that are ideal for iterated double integrals.

Definition 9.2.5 A set $S \subseteq \mathbb{R}^2$ is ***x*-simple** if there exist continuous $f : [a,b] \to \mathbb{R}$ and $g : [a,b] \to \mathbb{R}$ such that

$$S = \{(x,y) \in \mathbb{R}^2 : a \le x \le b, f(x) \le y \le g(x)\}.$$

A set $T \subseteq \mathbb{R}^2$ is ***y*-simple** if there exist continuous $p : [c,d] \to \mathbb{R}$ and $q : [c,d] \to \mathbb{R}$ such that

$$T = \{(x,y) \in \mathbb{R}^2 : c \le y \le d, p(y) \le x \le q(y)\}.$$

Remark 9.2.6 These were perfectly made to describe slices. Assuming Fubini's theorem holds, you can express integrals over x-simple sets S as an iterated double integral $dydx$.

$$\iint_S \varphi dA = \int_a^b \int_{f(x)}^{g(x)} \varphi(x,y)dydx$$

Similarly, you can express integrals over y-simple sets T as an iterated double integral $dxdy$.

$$\iint_T \varphi dA = \int_c^d \int_{p(y)}^{q(y)} \varphi(x,y)dxdy.$$

Thus far, you have only encountered these types of sets.

Example 9.2.7 The triangular region S in Example 9.2.2 is both x-simple and y-simple. The same is true for the region in Example 9.2.4. However, the region

$$T = \{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 2\pi, 0 \leq y \leq \sin(x)+1\}$$

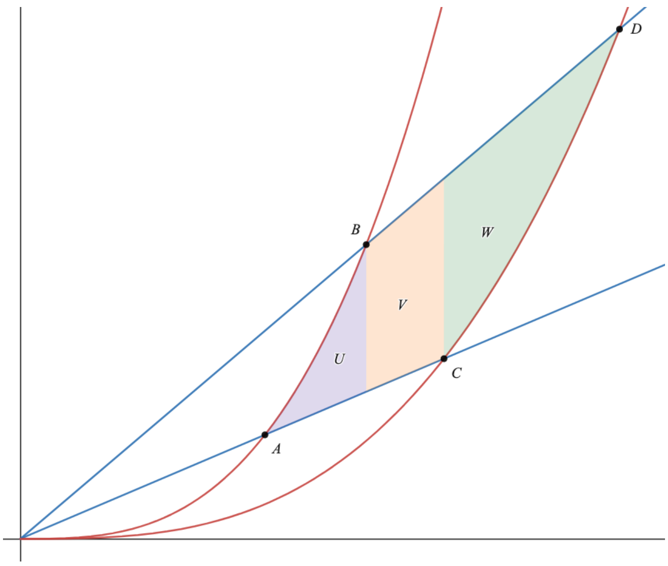
is x-simple, but it is not y-simple. Finally, the annulus

$$A = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 4\}$$

is neither x-simple nor y-simple.

Even if a region is not x-simple or y-simple, most regions can be *broken up into a union of simple pieces*. This introduces another strategy.

Example 9.2.8 Suppose you want to find the mass of a metal alloy plate P bounded between the curves $y = x$, $y = 2x$, $y = x^3$, and $y = 3x^3$ in the first quadrant with mass density given by $\delta(x,y) = \frac{y^2}{x^6}$. As usual, you can sketch this region; see this Desmos graph for details.



You can verify that the region P is both x-simple and y-simple, but it is easier to express it as a sum of double integrals by writing P as a union of three x-simple regions. These three regions $U, V, W \subseteq \mathbb{R}^2$ can be described using the four intersection points $A, B, C, D \in \mathbb{R}^2$ of the given curves, as labelled in the diagram above. These can be found by individually solving the four equations $x = 3x^3$, $2x = 3x^3$, $x = x^3$, and $2x = x^3$. Since they lie in the first quadrant, you can solve these equations to find that the four intersection points are therefore $A = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, $B = (\frac{\sqrt{2}}{\sqrt{3}}, \frac{2\sqrt{2}}{\sqrt{3}})$, $C = (1,1)$, and $D = (\sqrt{2}, 2\sqrt{2})$.

Based on the diagram above and these calculations, it follows that

$$\begin{aligned}
U &= \{(x,y) \in \mathbb{R}^2 : \tfrac{1}{\sqrt{3}} \le x \le \tfrac{\sqrt{2}}{\sqrt{3}}, x \le y \le 3x^3\} \\
V &= \{(x,y) \in \mathbb{R}^2 : \tfrac{\sqrt{2}}{\sqrt{3}} \le x \le 1, x \le y \le 2x\} \\
W &= \{(x,y) \in \mathbb{R}^2 : 1 \le x \le \sqrt{2}, x^3 \le y \le 2x\}.
\end{aligned}$$

and $P = U \cup V \cup W$. Since U, V, W intersect on zero Jordan measure sets, it follows that

$$\iint_P \delta \, dA = \iint_U \delta \, dA + \iint_V \delta \, dA + \iint_W \delta \, dA.$$

Since each of U, V, W are x-simple, you can conclude by Fubini's theorem that

$$\iint_P \delta \, dA = \int_{\frac{1}{\sqrt{3}}}^{\frac{\sqrt{2}}{\sqrt{3}}} \int_x^{3x^3} \frac{y^2}{x^6} dy dx + \int_{\frac{\sqrt{2}}{\sqrt{3}}}^{1} \int_x^{2x} \frac{y^2}{x^6} dy dx + \int_1^{\sqrt{2}} \int_{x^3}^{2x} \frac{y^2}{x^6} dy dx.$$

You can finally evaluate these three integrals; this calculation is left as a tedious exercise.

Another strategy is to apply *symmetry* of the integrand and the region of integration. In two dimensions, there are more symmetries that you can exploit.

Example 9.2.9 Suppose you want to evaluate

$$\iint_S \sin(x+y) dA.$$

where $S = \overline{B_1(0)} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ is the closed unit disk. You can rewrite this as

$$S = \{(x,y) \in \mathbb{R}^2 : -\sqrt{1-y^2} \le x \le \sqrt{1-y^2} \text{ and } -1 \le y \le 1\},$$

so the unit disk is y-simple. By Fubini's theorem, it follows that

$$\iint_S \sin(x+y) dA = \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sin(x+y) dx dy$$

Instead of directly evaluating this integral, you can use symmetry. Notice that the integrand satisfies a nice symmetry, namely

$$\forall (x,y) \in \mathbb{R}^2, \qquad \sin(x+y) = -\sin(-x-y).$$

The region of integration S is symmetric about the origin, that is, you can verify that

$$\forall (x,y) \in \mathbb{R}^2, (x,y) \in S \iff (-x,-y) \in S.$$

These symmetries are compatible so you can exploit it to evaluate the integral. You can split the unit disk and corresponding iterated integrals into its four quadrants, so that

$$\iint_{\overline{B_1(0)}} \sin(x+y)dA = \int_0^1 \int_0^{\sqrt{1-y^2}} \sin(x+y)dxdy + \int_{-1}^0 \int_0^{\sqrt{1-y^2}} \sin(x+y)dxdy + \int_0^1 \int_{-\sqrt{1-y^2}}^0 \sin(x+y)dxdy + \int_{-1}^0 \int_{-\sqrt{1-y^2}}^0 \sin(x+y)dxdy$$

Applying the symmetry $\sin(x+y) = -\sin(-x-y)$ is equivalent to doing the substitutions $x \mapsto -x$ for the inner dx interal, and then $y \mapsto -y$ for the outer dy integral. Following these steps, you can verify that

$$\int_{-1}^0 \int_{-\sqrt{1-y^2}}^0 \sin(x+y)dxdy = \int_0^1 \int_0^{\sqrt{1-y^2}} -\sin(x+y)dxdy$$

and

$$\int_0^1 \int_{-\sqrt{1-y^2}}^0 \sin(x+y)dxdy = \int_{-1}^0 \int_0^{\sqrt{1-y^2}} -\sin(x+y)dxdy$$

Combined with the earlier identity, you conclude that

$$\iint_S \sin(x+y)dA = 0.$$

Amazingly, you did this without computing any integrals directly.

Sometimes you may want to only identify whether an integral is non-negative. Instead of calculating the integral, you can often analyze the integrand and the region of integration to make such conclusions.

Example 9.2.10 The iterated integral

$$I = \int_0^2 \int_{-2}^{-3} xy^2 dydx$$

is non-negative since the integrand xy^2 is non-negative on the domain $R = [0,2] \times [-2,-3]$. Indeed, $x \geq 0$ for $x \in [0,2]$ and $y^2 \geq 0$ always, so $xy^2 \geq 0$ on R. By monotonicity, I is non-negative.

Finally, certain iterated double integrals can be interpreted as the volume of a solid or area of a region. If the solid or region is classical, then you can use well-known formulas to determine the integral without any calculation.

Example 9.2.11 Consider the iterated double integrals

$$I = \int_0^1 \int_0^{\sqrt{1-x^2}} 1dydx, \qquad J = \int_0^1 \int_0^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}dydx.$$

The region of integration of I is

$$S = \{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq \sqrt{1-x^2}\}.$$

After some sketching, you may notice that this is the quarter unit disk in the first quadrant. Thus, using the area of a disk and its symmetry, you can conclude that

$$I = \text{area}(S) = \frac{\pi}{4}.$$

For J, notice that $\sqrt{1-x^2-y^2} \geq 0$ so J is the volume of the set $T \subseteq \mathbb{R}^3$ given by

$$T = \{(x,y,z) \in \mathbb{R}^3 : 0 \leq x \leq 1, 0 \leq y \leq \sqrt{1-x^2}, 0 \leq z \leq \sqrt{1-x^2-y^2}\}.$$

After some sketching, you may notice that the set T is the unit ball in the first octant of $\mathbb{R}^3$. Thus, using the volume of the unit ball and its symmetry,

$$J = \text{vol}(T) = \frac{1}{8} \cdot \frac{4\pi}{3} = \frac{\pi}{6}.$$

These examples showcase some flexible and interesting ideas for analyzing and evaluating iterated double integrals. The main strategies can be succinctly described.

> *Sketch the region and describe it in several ways; directly calculate with FTC; swap the order of integration with Fubini's theorem; apply symmetries of the integrand or region; break up the region into simpler pieces; interpret geometrically as a volume or area of a classic object.*

You may be able to succeed in more than one way. You may sometimes need to combine several of these ideas. You may need to create new ones. This list is an excellent starting point, but it is not exhaustive by any means. In fact, it is missing a revolutionary strategy.

> *Convert the integral from rectangular coordinates to polar coordinates.*

You will investigate this fundamental tool and its advantages in the next section.

Exercises for Section 9.2

Concepts and definitions

9.2.1 Decide whether the following integrals are positive, zero, or negative.

(a) $\displaystyle\int_{-1}^{1}\int_{0}^{1} xy\,dx\,dy$

(b) $\displaystyle\int_{-1}^{1}\int_{-1}^{1} xy\,dx\,dy$

(c) $\displaystyle\int_{0}^{1}\int_{0}^{1} xy\,dx\,dy$

(d) $\displaystyle\int_{-\pi/2}^{\pi/2}\int_{-1}^{1} x^2\sin(y)\,dx\,dy$

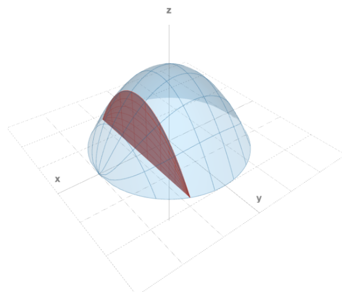
(e) $\displaystyle\int_{-\pi/2}^{\pi/2}\int_{-2}^{2} x^2\cos(y)\,dx\,dy$

(f) $\displaystyle\int_{2}^{4}\int_{0}^{1} e^x(1-y^2)\,dx\,dy$

9.2.2 Iterated double integrals can be used to find the volume of solids. Remember there is a natural two-step process: first, determine the cross-sectional area of a slice of the solid, and then integrate along those slices to get the total volume. Consider the solid

$$T = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 \le 4, 0 \le z \le f(x,y)\}$$

The picture below illustrates a slice of T.

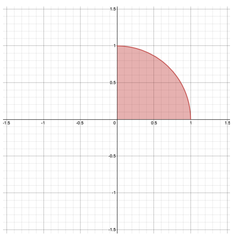


(a) Label the area of the cross section $A(x)$ or $A(y)$ depending on which variable is appropriate. Give an explicit integral formula for the cross-sectional area.

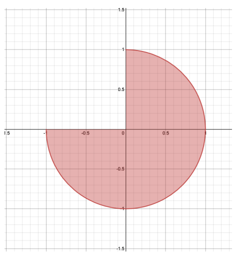
(b) Express $\text{vol}(T)$ as an iterated double integral using your area function.

9.2.3 Consider the following six regions.

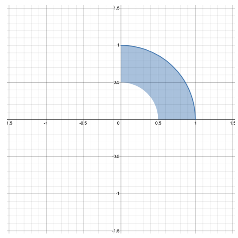
A.



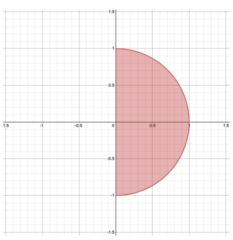
C.



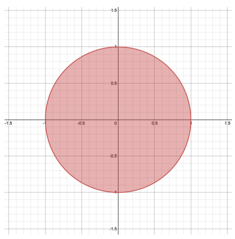
E.



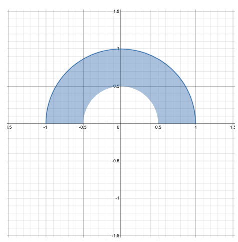
B.



D.



F.



(a) Which regions are x-simple, y-simple, neither, or both?

(b) Which regions can be written as a finite union of x-simple regions?

(c) Which regions can be written as a finite union of y-simple regions?

9.2.4 Let $f:\mathbb{R}^2\to\mathbb{R}$ be a continuous function. Consider four possible symmetries of f.

A) $\forall(x,y)\in\mathbb{R}^2, f(-x,y)=f(x,y)$
B) $\forall(x,y)\in\mathbb{R}^2, f(-x,y)=-f(x,y)$
C) $\forall(x,y)\in\mathbb{R}^2, f(x,-y)=-f(x,y)$
D) $\forall(x,y)\in\mathbb{R}^2, f(-x,-y)=-f(x,y)$

(a) Which symmetries would guarantee that $\displaystyle\int_{-1}^{1}\int_{-2}^{2} f(x,y)dydx=0$?

(b) Which symmetries would guarantee that $\displaystyle\int_{-1}^{1}\int_{0}^{2} f(x,y)dydx=0$?

(c) Which symmetries would guarantee that $\displaystyle\int_{-1}^{1}\int_{-2}^{2} f(x,y)dydx=2\int_{0}^{1}\int_{-2}^{2} f(x,y)dydx$?

(d) Which symmetries would guarantee that $\displaystyle\int_{-1}^{1}\int_{-2}^{2} f(x,y)dydx=2\int_{-1}^{1}\int_{0}^{2} f(x,y)dydx$?

(e) Which symmetries would guarantee that $\displaystyle\int_{-1}^{1}\int_{0}^{\sqrt{4-x^2}} f(x,y)dydx=0$?

(f) Which symmetries would guarantee that $\displaystyle\int_{0}^{1}\int_{0}^{x} f(x,y)dydx+\int_{-1}^{0}\int_{x}^{0} f(x,y)dydx=0$?

Computations

9.2.5 Switching the order of integration is a new and amazing integration technique. Sketch the region of integration and compute

$$\int_0^1\int_y^1 e^{-x^2}dxdy$$

by swapping to the $dydx$ order.

9.2.6 Express the integral in Example 9.2.4 as an iterated double integral $dxdy$.

9.2.7 For each iterated integral, i) draw the region of integration; ii) write the integral in both $dxdy$ and $dydx$ orders; and iii) compute the integral. Assume Fubini's theorem applies.

(a) $\displaystyle\int_0^1\int_0^x xy^2dydx$

(a) $\displaystyle\int_{-1}^1\int_0^{\sqrt{1-y^2}} ydxdy$

9.2.8 Iterated double integrals can sometimes be interpreted as the volumes of familiar solids. For each iterated integral below, describe the solid and use that knowledge (rather than integration) to determine the value of the integral.

(a) $\displaystyle\int_0^L\int_0^W Hdxdy$ for fixed $L,W,H>0$.

(b) $\displaystyle\int_0^1\int_0^{1-x} 4dydx$

(c) $\displaystyle\iint_{x^2+y^2\le R^2}\sqrt{R^2-x^2-y^2}\,dA$ for fixed $R>0$.

9.2.9 Sketch the region and evaluate the integral $\int_0^1 \int_x^{x^2+2} (1-x)dydx$.

9.2.10 Sketch the region and evaluate the integral $\int_0^\pi \int_0^{x^3} e^{y/x} dydx$.

9.2.11 Calculate

$$\int_{-3}^{3} \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} 3-\sqrt{x^2+y^2}dxdy$$

by interpreting it as the volume of a familiar solid.

9.2.12 Evaluate

$$\int_0^\pi \int_x^\pi \frac{\sin(y)}{y} dydx$$

by sketching the region, swapping to $dxdy$, and then calculating.

Proofs

9.2.13 Thad wants to express the area of the sheet $S = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2+y^2 \le 4\}$ using iterated double integrals. He loves algebra, so he mindlessly manipulates inequalities to find an expression.

> 1. *Note* $1 \le x^2+y^2 \le 4 \implies 1-x^2 \le y^2 \le 4-x^2$ *so*
>
> $$\sqrt{1-x^2} \le y \le \sqrt{4-x^2} \quad \text{or} \quad -\sqrt{4-x^2} \le y \le -\sqrt{1-x^2}$$
>
> 2. *Also since* $y^2 \ge 0$, *it follows that* $1 \le x^2 \le 4$ *so* $-2 \le x \le -1$ *or* $1 \le x \le 2$.
> 3. *Combining all cases,*
>
> $$\text{area}(S) = \int_1^2 \int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}} 1dydx + \int_1^2 \int_{-\sqrt{4-x^2}}^{-\sqrt{1-x^2}} 1dydx$$
> $$+ \int_{-2}^{-1} \int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}} 1dydx + \int_{-2}^{-1} \int_{-\sqrt{4-x^2}}^{-\sqrt{1-x^2}} 1dydx$$

This terrible write up has many flaws, but do not worry about fixing the reasoning.

(a) Thad's final answer is wrong. Explain why. *Hint:* What do his iterated integrals represent?

(b) What advice would you give Thad so he does not make the same mistake next time?

9.2.14 Define $S = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2+y^2 \le 4\}$

(a) Express S as a union of x-simple sets.

(b) Express the area of S as a sum of four double integrals $dydx$.

9.2.15 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuous. Assume $f(-x,-y) = -f(x,y)$ for $(x,y) \in \mathbb{R}^2$. For $a, b \in \mathbb{R}$, prove that

$$\int_{-a}^{a} \int_{-b}^{b} f(x,y)dxdy = 0.$$

Applications and beyond

9.2.16 You are determining the mass of a metal alloy plate P bounded between the curves $y = x$, $y = 2x$, $y = x^3$, and $y = 3x^3$ in the first quadrant with mass density given by $\delta(x, y)$.

(a) First, sketch the curves. Label the curves in your sketch.

(b) Next, after some calculations, you find four intersection points $A = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, $B = (\frac{\sqrt{2}}{\sqrt{3}}, \frac{2\sqrt{2}}{\sqrt{3}})$, $C = (1, 1)$, and $D = (\sqrt{2}, 2\sqrt{2})$. Express the mass as a sum of iterated integrals $dy\,dx$.

9.2.17 Let M be the solid which is fully encapsulated inside the graph of $f(x, y) = x^2y^3$, the xy-plane, and the planes $x = -2$, $x = 1$, $y = -1$ and $y = 1$.

(a) The volume of M is not equal to $\displaystyle\int_{-2}^{1}\int_{-1}^{1} x^2y^3\,dy\,dx$. Briefly explain why.

(b) Find the volume of M.

9.2.18 Find the mass of the sheet $S = \{(x, y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 4, x \ge 0, y \ge 0\}$ with density $\delta(x, y) = y$.

9.2.19 Find the average value of $\varphi(x, y) = 5xy^2 - y^4$ on the region bounded by $x = 2y$ and $y = \sqrt{x}$.

9.2.20 Find the centroid of the triangle $T = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1,\ 0 \le y \le x\}$.

9.2.21 Find the centre of mass of $T = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1,\ 0 \le y \le x\}$ with mass density $\delta(x, y) = x$.

9.3. Double integrals in polar coordinates

Iterated double integrals ($dxdy$ and $dydx$) are best suited for calculating integrals of regions in $\mathbb{R}^2$ which are rectangular (e.g. rectangles, areas under curves $y = f(x)$), but not all shapes are naturally described using rectangular coordinates. Another coordinate system in $\mathbb{R}^2$ is *polar coordinates*; see Section 1.4.1 for an introduction. For the entirety of this section, the map $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$g(r, \theta) = (r\cos\theta, r\sin\theta) \tag{9.3.1}$$

is the polar coordinate transformation. Recall that its domain is referred to as the (r, θ)-plane and its codomain is referred to as the (x, y)-plane. Radially symmetric objects are often best described with polar coordinates. For instance, the unit disk in the (x, y)-plane corresponds to a rectangle in the (r, θ)-plane; how does integrating over this rectangle correspond to integrating over the unit disk? This can be phrased more generally.

Can you integrate over regions in the (x, y)-plane by integrating in the (r, θ)-plane?

You will explore this question here and heuristically derive a positive answer. This creates a brand new integration technique: integration in polar coordinates! Although the explanations and calculations will be not be fully rigorous yet, this example will reveal some deep insights and set the stage for an incredible theorem on integration with different coordinate systems.

9.3.1 Regions in polar coordinates

Before attempting to address this section's core question, it will be helpful to recognize how polar coordinates describe regions in the (x, y)-plane. Recall the standard relationship

$$(x, y) = g(r, \theta) = (r\cos\theta, r\sin\theta)$$

for any $r, \theta \in \mathbb{R}$. If r is positive, then $r = |r|$ corresponds to the distance of (x, y) from the origin and θ is the angle, measured counterclockwise from the positive horizontal axis, of the point (x, y). If r is negative, then $-r = |r|$ is the distance of (x, y) from the origin and $\theta + \pi$ is the angle, measured counterclockwise from the positive horizontal axis, of the point (x, y). See Section 1.4.1 for details on this distinction between positive and negative values of r.

You will often construct integrals where r is non-negative, but this sign issue will have an effect on how you describe polar regions. For now, you can study a few examples.

Example 9.3.1 Informally speaking, the polar region $0 \le r \le 1$ corresponds to the unit disk $x^2 + y^2 \le 1$. More formally, the rectangle $[0,1] \times \mathbb{R}$ in the (r, θ)-plane is mapped under g to the unit disk

$$g([0,1] \times \mathbb{R}) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$$

in the (x, y)-plane. Notice this mapping is not at all one-to-one; the image of $[0,1]\times(-\infty, 0]$ and $[0,1] \times [0, \infty)$ under g will both be the unit disk too.

To describe regions in polar coordinates, it will be more sensible to use polar regions that almost correspond one-to-one to a rectangular region. For instance, the rectangle $[0,1] \times [0, 2\pi]$ in the (r, θ)-plane is mapped by g to the unit disk $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ in the (x, y)-plane and this mapping is one-to-one *aside from a set of zero Jordan measure*. In particular, the line segments $[0,1] \times \{0\}$ and $[0,1] \times \{2\pi\}$ both map to the positive horizontal axis and the line segment $\{0\} \times [0, 2\pi]$ maps to the origin. All of these sets have zero Jordan measure in $\mathbb{R}^2$. As you have seen, sets of zero Jordan measure will not affect the value of an integral so this minor defect will be permissible when integrating.

When studying the polar coordinate transformation (9.3.1) in Section 1.4.1, you were extremely careful to distinguish between regions described in its domain, the (r,θ)-plane, and its codomain, the (x,y)-plane. For instance, the set

$$A = \{(r,\theta) \in \mathbb{R}^2 : 0 \le r \le 1, 0 \le \theta \le 2\pi\}$$

is *not* the same as the set

$$B = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}.$$

Formally, B is the image of A under g, that is $B = g(A)$. This level of formality is valuable in some contexts, but it can be unnecessarily pedantic especially for less rigorous calculations. To relieve some of this pedantry, you may more informally say, for instance,

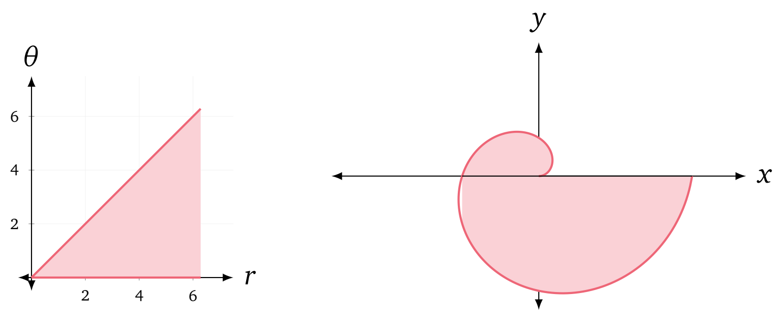
$$0 \le r \le 1, 0 \le \theta \le 2\pi \textit{ is the unit disk } x^2 + y^2 \le 1 \textit{ in polar form.}$$

This can always be formally interpreted with the polar coordinate transformation between a set in its domain and a set in its codomain, but it will be acceptable to write in this style for less formal arguments. Remember some texts always assume that $r \ge 0$ but you shall not do so here; if you want to require $r \ge 0$, then you should specify it.

Example 9.3.2 The polar curve $r = \theta$ for $0 \le \theta \le 2\pi$ corresponds to a straight line in the (r,θ)-plane and an outward counterclockwise spiral in the (x,y)-plane. The polar region

$$0 \le r \le \theta, 0 \le \theta \le 2\pi$$

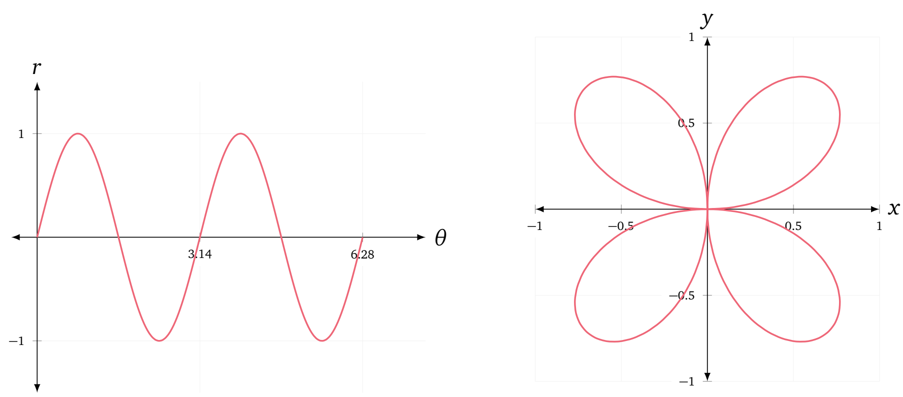
therefore corresponds to a triangle in the (r,θ)-plane and the spiral region in the (x,y)-plane.



View this Desmos animation to watch how the θ-slices trace out the region.

Polar regions are more subtle when they involve negative values of the radius r.

Example 9.3.3 The polar curve $r = \sin(2\theta)$ for $0 \le \theta \le 2\pi$ corresponds to the usual sine plot in the (r,θ)-plane, and it traces out the shape of four leaves in the (x,y)-plane.



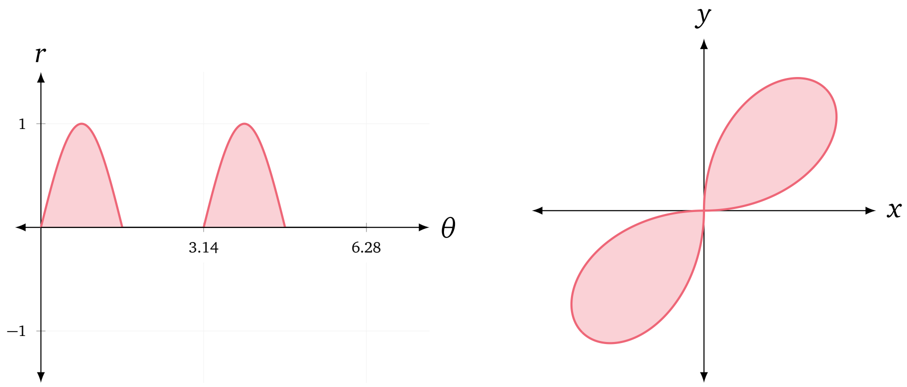
See Example 1.4.8 for a similar curve. Notice, in particular, that $r = \sin 2\theta \le 0$ when $\frac{\pi}{2} \le \theta \le \pi$ and $\frac{3\pi}{2} \le \theta \le 2\pi$. Thus the polar region

$$0 \le r \le \sin(2\theta), 0 \le \theta \le 2\pi$$

is equivalent to

$$0 \le r \le \sin(2\theta), 0 \le \theta \le \frac{\pi}{2} \quad \text{and} \quad 0 \le r \le \sin(2\theta), \pi \le \theta \le \frac{3\pi}{2}.$$

This corresponds to two hills in the (r,θ)-plane and the two leaves in the (x,y)-plane.



If you want to include the other two leaves, you need to describe those regions separately with the other intervals of θ, namely $\sin(2\theta) \le r \le 0$ and either $\frac{\pi}{2} \le \theta \le \pi$ or $\frac{3\pi}{2} \le \theta \le 2\pi$. View this Desmos animation to watch how the θ-slices trace out the two leaves.

9.3.2 Derivation of integrals in polar coordinates

Now, you can approach the core problem. Given a real-valued function $f : D \to \mathbb{R}$ that is integrable on a Jordan measurable set $D \subseteq \mathbb{R}^2$, you want to calculate its integral

$$\iint_D f\, dA.$$

Thus far, your only strategy is to try and express this as a sum of integrals $dx\,dy$ or $dy\,dx$. Such an expression is referred to as *rectangular* coordinates since x-slices and y-slices can describe rectangular regions. If you want to switch to polar coordinates, you must be able to describe D in terms of polar region.

In other words, assume there exists a Jordan measurable set $\Omega \subseteq \mathbb{R}^2$ such that

$$D = g(\Omega) = \{(r\cos\theta, r\sin\theta) : (r,\theta) \in \Omega\}$$

and further assume the restricted polar coordinates transformation $g|_\Omega : \Omega \to g(\Omega)$ is *bijective*. Notice an integral over the rectangular region $D = g(\Omega)$ lives in the (x,y)-plane, whereas the polar region Ω lives in the (r,θ)-plane. Equivalently, Ω is the polar form of $g(\Omega)$.

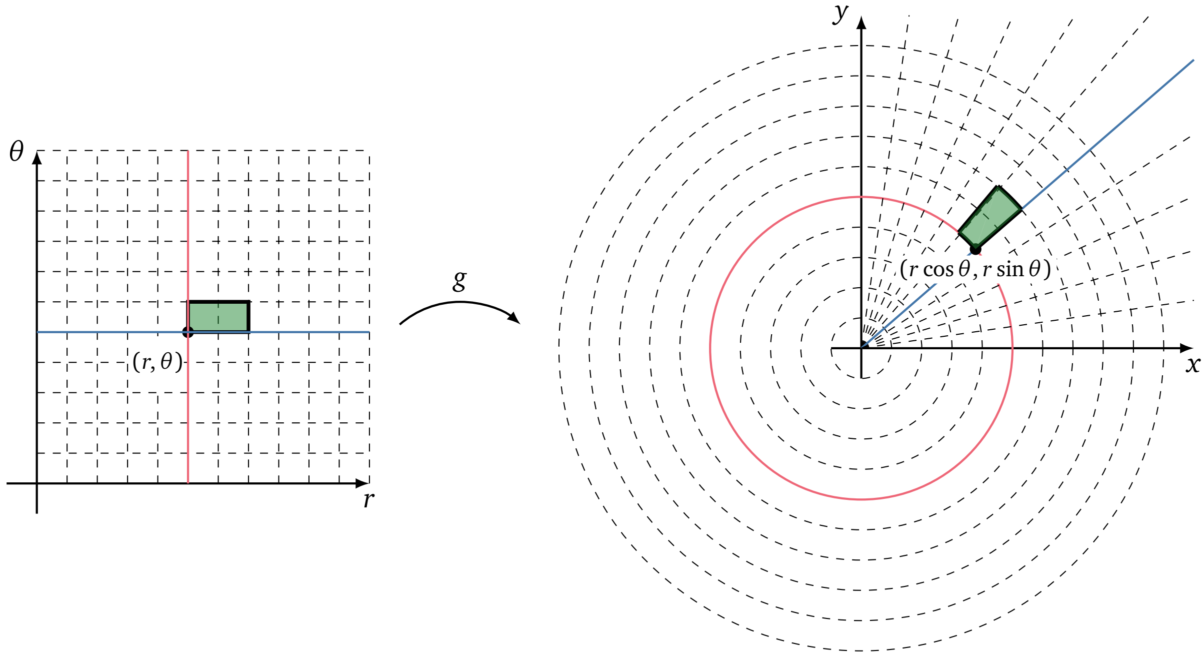
> *How does integration over $g(\Omega)$ in the (x,y)-plane correspond to integration over Ω in the (r,θ)-plane?*

This question is a more formal version of this section's motivating concern. Using your mantra of chopping, estimating, and refining, you can heuristically derive a method for integrating with polar coordinates.

Before chopping up Ω, consider how a small rectangle in the (r,θ)-plane transforms into the (x,y)-plane. For a fixed $r>0$ and $\theta\in\mathbb{R}$, the small rectangle

$$R=[r,r+\Delta r]\times[\theta,\theta+\Delta\theta]$$

in the (r,θ)-plane transforms under g to a piece of washer $g(R)$ in the (x,y)-plane. This transformation is illustrated below.



Since the rectangle is small, the thickness of the washer is approximately

$$||g(r+\Delta r,\theta)-g(r,\theta)||\approx\left|\left|\frac{\partial g}{\partial r}(r,\theta)\right|\right|\Delta r=\left|\left|\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}\right|\right|\Delta r=\Delta r$$

and its width is approximately

$$||g(r,\theta+\Delta\theta)-g(r,\theta)||\approx\left|\left|\frac{\partial g}{\partial\theta}(r,\theta)\right|\right|\Delta\theta=\left|\left|\begin{bmatrix}-r\sin\theta\\ r\cos\theta\end{bmatrix}\right|\right|\Delta\theta=r\Delta\theta$$

Overall, this implies that

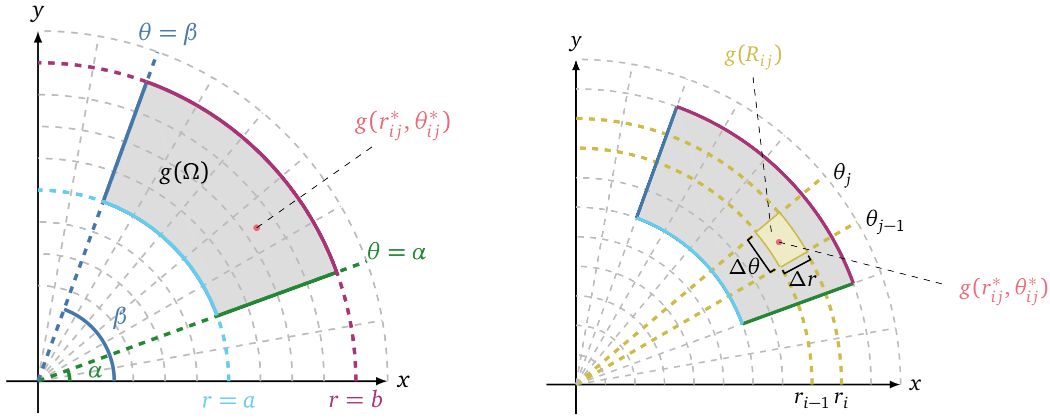
$$\text{area}(g(R))\approx r\Delta r\Delta\theta=r\,\text{area}(R). \tag{9.3.2}$$

Notice the area is scaled by the radius $r=|r|$ as $r>0$. If $r<0$, then the area would be scaled by $-r=|r|$. Now, you can proceed to chop your region.

Without loss of generality, assume Ω is a rectangle, that is,

$$\Omega=\{(r,\theta):a\le r\le b,\alpha\le\theta\le\beta\}.$$

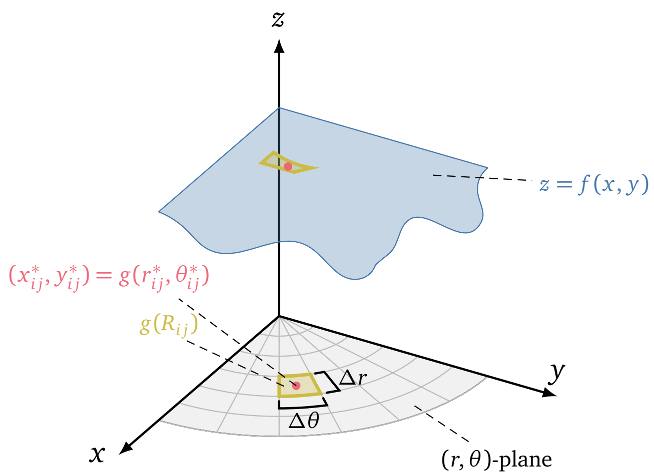
The transformed region $D=g(\Omega)$ in the (x,y)-plane is illustrated below in diagram (a).



(a) (b)

Let $P = \{R_{ij} : 1 \le i \le N, 1 \le j \le N\}$ be a regular partition of Ω into N^2 rectangles $R_{ij} = [r_{i-1}, r_i] \times [\theta_{j-1}, \theta_j]$ of width Δr and height $\Delta\theta$. A transformed subrectangle $g(R_{ij})$ is illustrated in diagram (b) above. Assume the norm $||P||$ is small so that each transformed subrectangle $g(R^*_{ij})$ is small.

Choose a sample point $(r^*_{ij}, \theta^*_{ij}) \in R_{ij}$, say the centre of R_{ij} as in diagram (b). This sample point in the (r, θ)-plane corresponds to $(x^*_{ij}, y^*_{ij}) = g(r^*_{ij}, \theta^*_{ij})$ in the (x, y)-plane. Since $||P||$ is small, the function f should be roughly constant on the transformed subrectangle $g(R_{ij})$ and hence approximated by its value $f(x^*_{ij}, y^*_{ij})$ at this sample point. This idea is illustrated below with the graph of $z = f(x, y)$.



Thus, the contribution of f on the transformed subrectangle $g(R_{ij})$ is approximately

$$f(x^*_{ij}, y^*_{ij})\,\text{area}(g(R_{ij})) \approx f \circ g(r^*_{ij}, \theta^*_{ij})|r^*_{ij}|\Delta r \Delta\theta$$

by (9.3.2). Summing this over all subrectangles, the total value of f on $D = g(\Omega)$ satisfies

$$\iint_{g(\Omega)} f\,dA \approx \sum_{i=1}^{N}\sum_{j=1}^{N} f \circ g(r^*_{ij}, \theta^*_{ij})|r^*_{ij}|\Delta r \Delta\theta.$$

Taking $N \to \infty$ implies $||P_N|| \to 0$ (i.e. $\Delta r \to 0$ and $\Delta\theta \to 0$) and thus, it seems plausible that

$$\iint_{g(\Omega)} f\,dA = \lim_{N\to\infty} \sum_{i=1}^{N}\sum_{j=1}^{N} f \circ g(r^*_{ij}, \theta^*_{ij})|r^*_{ij}|\Delta r \Delta\theta = \int_{\alpha}^{\beta}\int_{a}^{b} (f \circ g)(r, \theta)|r|\,dr\,d\theta.$$

Voilà! You have heuristically derived the desired relation and can create a conjecture.

Theorem 9.3.4 Let $\Omega \subseteq \mathbb{R}^2$ be a Jordan measurable set such that the restricted polar coordinate transformation $g|_\Omega : \Omega \to g(\Omega)$ given by (9.3.1) is a bijection. If $f : g(\Omega) \to \mathbb{R}$ is integrable on $g(\Omega)$, then $F : \Omega \to \mathbb{R}$ given by $F(r,\theta) = (f \circ g)(r,\theta) \cdot |r|$ is integrable on Ω and

$$\iint_{g(\Omega)} f\,dA = \iint_\Omega F\,dA.$$

Proof. Postponed. This will follow from a more general theorem. ■

Theorem 9.3.4 can be remembered using the "area" element dA, namely you have that

$$dxdy = dA = |r|drd\theta.$$

The "identity" between the symbols above has no formal meaning, but you can interpret this as:

> *Infinitesimal area dA can be calculated using the area of an infinitesimal rectangle $dxdy$ or using the area of an infinitesimal polar rectangle $|r|drd\theta$.*

This mnemonic device is a useful trick that will re-appear many times.

9.3.3 Examples of integrals in polar coordinates

You can finally calculate many double integrals in polar coordinates. This new technique is especially powerful for regions with radial symmetry.

Example 9.3.5 The washer $1 \le x^2 + y^2 \le 9$ is given by $1 \le r \le 3, 0 \le \theta \le 2\pi$ in polar form. These regions do not map bijectively under the polar coordinates transformation, but the only issue occurs on their boundaries. This can be resolved by removing this boundary and adding its contribution later because this set has zero Jordan measure.

Set $\Omega = (1,3) \times (0,2\pi)$, so $g(\Omega) = \{(x,y) \in \mathbb{R}^2 : 1 < x^2 + y^2 < 9\}$ where g is given by (9.3.1). You can verify that this mapping $g|_\Omega : \Omega \to g(\Omega)$ is a bijection. Also, you can verify that $g(\Omega)$ is a Jordan measurable set, so by Theorem 7.7.4, the constant 1 function is integrable on $g(\Omega)$. By switching to polar coordinates (Theorem 9.3.4) and Fubini's theorem, the area of the washer is given by

$$\text{area}(g(\Omega)) = \iint_{g(\Omega)} 1dA = \iint_\Omega F\,dA$$

where $F : \mathbb{R}^2 \to \mathbb{R}$ is given by $F(r,\theta) = |r|$.

Now, since Ω is Jordan measurable, Theorem 7.5.14 implies that $\text{vol}(\partial\Omega) = 0$ and so by Theorem 7.7.8, $\iint_{\partial\Omega} F\,dA = 0$. Since $\Omega \cap \partial\Omega = \varnothing$, it follows by additivity (Theorem 7.3.17) that

$$\text{area}(g(\Omega)) = \iint_\Omega F\,dA = \iint_\Omega F\,dA + \iint_{\partial\Omega} F\,dA = \iint_{\overline{\Omega}} F\,dA.$$

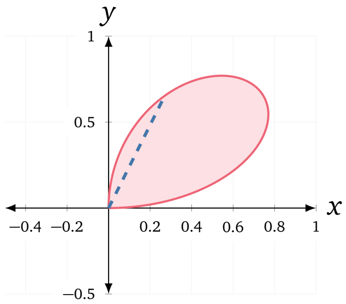
As $\overline{\Omega} = [1,3] \times [0,2\pi]$, we may finally deduce that

$$\text{area}(g(\Omega)) = \int_0^{2\pi} \int_1^3 r\,drd\theta = \int_0^{2\pi} \frac{1}{2}r^2\Big|_{r=1}^{r=3} d\theta = \int_0^{2\pi} 4d\theta = 8\pi.$$

This last calculation is dramatically easier compared to rectangular coordinates!

Rigorously explaining how to apply Theorem 9.3.4 requires many details to be checked, so most calculations will be less formal and skip such details. Nonetheless, you should at least make clear choices which will satisfy its assumptions. Bijectivity issues along the boundary can be safely ignored since the sets are Jordan measurable.

Example 9.3.6 Let D be the region in the *first quadrant* bounded by the polar curve $r = \sin(2\theta)$ for $0 \le \theta \le 2\pi$. You want to find the mass of D which has density $\delta(x, y) = y$. By inspecting Example 9.3.2, notice that D is described as $0 \le r \le \sin(2\theta)$, $0 \le \theta \le \frac{\pi}{2}$ in polar form.



View this Desmos animation. Note that each θ-slice originates at the origin and terminates at $r = \sin(2\theta)$, which is always non-negative for any fixed $0 \le \theta \le \frac{\pi}{2}$. Thus, the θ-slice is the line segment at angle θ with $0 \le r \le \sin(2\theta)$. Overall, by changing variables to polar coordinates (Theorem 9.3.4), you obtain that

$$\begin{aligned} \iint_D \delta(x,y)dA &= \int_0^{\frac{\pi}{2}} \int_0^{\sin(2\theta)} \delta(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\ &= \int_0^{\frac{\pi}{2}} \int_0^{\sin(2\theta)} r^2 \sin\theta\,dr\,d\theta \\ &= \int_0^{\frac{\pi}{2}} \frac{1}{3}(\sin(2\theta))^3 \sin(\theta)\,d\theta. \end{aligned}$$

Using the trigonometric identities $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ and $\sin^2(\theta) + \cos^2(\theta) = 1$, it follows that the above is equal to

$$\int_0^{\frac{\pi}{2}} \frac{8}{3}\sin^4(\theta)\cos^3(\theta)\,d\theta = \int_0^{\frac{\pi}{2}} \frac{8}{3}(1-\sin^2(\theta))\sin^4(\theta)\cos(\theta)\,d\theta.$$

Using the substitution $u = \sin(\theta)$ and $du = \cos(\theta)d\theta$, you may conclude that

$$\iint_D \delta(x,y)dA = \int_0^1 \frac{8}{3}(1-u^2)u^4\,du = \frac{8}{15} - \frac{8}{21}.$$

Remark 9.3.7 Unlike Example 9.3.5, notice the above example does not explain how it applies Theorem 9.3.4. This level of explanation is routine, but you should be able to fill in the details in a similar fashion. For instance, the region $0 \le r \le \sin(2\theta), 0 \le \theta \le \frac{\pi}{2}$ does not map bijectively under the polar coordinates maps, but it only fails along the boundary. This effect is harmless because you can carefully apply to Theorem 9.3.4 as in Example 9.3.5 to complete the details.

Now, you may witness true mathematical beauty by evaluating the Gaussian integral.

Example 9.3.8 Using polar coordinates, you can sketch a proof of a magical identity:

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Define $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$. You can verify that this integral converges, e.g. by comparing with e^{-x} and using even symmetry. Now, assuming some liberties with improper integrals, you may square I and reasonably believe that

$$\begin{aligned} I^2 = I \cdot I &= \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right) e^{-y^2/2} dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy. \end{aligned}$$

These manipulations are not justified but they are plausible. Indeed, the final quantity is a *two-dimensional* improper integral which you have not yet studied.

Nonetheless, you may take this opportune moment to wield integration in polar coordinates. Note that any point in $\mathbb{R}^2$ can be represented by some angle $0 \le \theta \le 2\pi$ and radius $0 \le r < \infty$. Then, since $r^2 = x^2 + y^2$, it follows by switching to polar coordinates that

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta = \int_0^{2\pi} \lim_{r \to \infty} (1 - e^{-r^2/2}) d\theta = \int_0^{2\pi} d\theta = 2\pi$$

This informally shows that $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ as desired.

This concludes the creation of a powerful new technique for integration in $\mathbb{R}^2$. Equipped with Fubini's theorem, you can calculate double integrals in $\mathbb{R}^2$ using rectangular coordinates or polar coordinates. This toolbox contains the essentials and it is already fairly versatile. Your journey continues by generalizing these ideas to integration in $\mathbb{R}^3$, starting with Fubini's theorem.

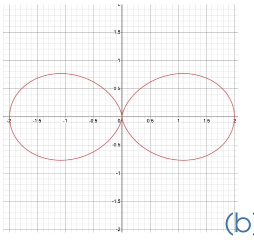
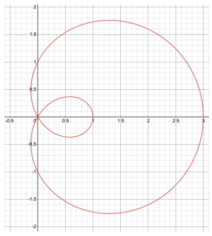
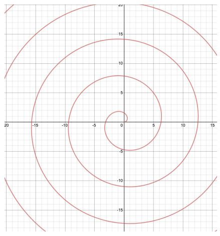
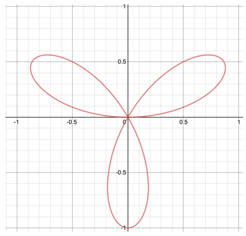
Exercises for Section 9.3

Concepts and definitions

9.3.1 Here is a list of six equations in polar coordinates.

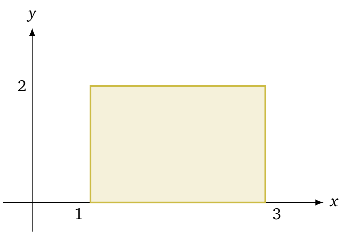
(A) $r = \theta, \theta > 0$ (B) $r = \theta, \theta < 0$ (C) $r = 1 + \cos 2\theta$
(D) $r = \cos 3\theta$ (E) $r = \sin 3\theta$ (F) $r = 1 + 2\cos\theta$

For each plot below, match them to exactly 1 of the above equations.

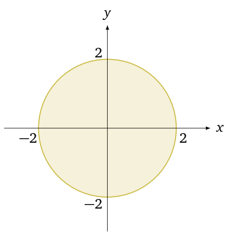
(a)

(b)

(c)

(d)


9.3.2 Sketch the region described by the cardiod $0 \le r \le 1 + \cos\theta$ for $0 \le \theta \le 2\pi$.

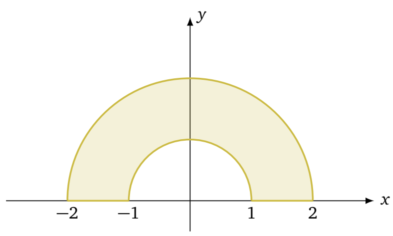
9.3.3 Write an integrated double integral for the area of the following regions after deciding which coordinate system to integrate in.



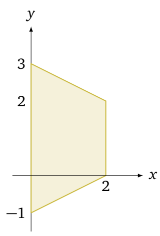
(a)



(b)



(c)



(d)

9.3.4 To set up an integral in polar coordinates, you need to know how the geometry relates to the integral. Here you will calculate the mass of a region using polar coordinates via **two** equivalent approaches. Each one corresponds to a different geometric perspective.

Let R be the region in $\mathbb{R}^2$ bounded by $r = \sin(3\theta)$ in the 1st quadrant. This leaf R has mass density $\delta(x, y) = y$ and is illustrated with a Desmos polar graphing calculator.

(a) Sketch a typical θ-slice of the leaf R. For which values of $\theta \in [0, 2\pi]$ will this θ-slice be non-empty aside from the origin?

(b) Express the "1-dimensional mass density" of each θ-slice of the leaf R as a single integral dr.

(c) Express the 2-dimensional mass of the leaf R as a double integral $dr\,d\theta$.

(d) Sketch a typical r-slice of the leaf R. For which values of $r \ge 0$ will this r-slice be non-empty?

(e) Express the "1-dimensional mass density" of each r-slice of the leaf R as a single integral $d\theta$.

(f) Express the 2-dimensional mass of the leaf R as a double integral $d\theta\,dr$.

(g) Use WolframAlpha to calculate the mass of R using one of your two double integrals above.

9.3.5 Let $P = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \le 4\}$ be a plate with mass density $f(x,y)$. Which of the following expressions are equal to the mass of P?

(a) $\displaystyle\iint_P f\,dA$

(b) $\displaystyle\int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} f(x,y)\,dy\,dx$

(c) $\displaystyle\int_{-2}^{2}\int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} f(x,y)\,dx\,dy$

(d) $\displaystyle\int_{0}^{2\pi}\int_{0}^{2} f(r,\theta)\,r\,dr\,d\theta$

(e) $\displaystyle\int_{0}^{2\pi}\int_{0}^{2} f(r\cos\theta, r\sin\theta)\,dr\,d\theta$

(f) $\displaystyle\int_{0}^{2\pi}\int_{0}^{2} f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta$

9.3.6 Three friends are calculating the mass of two leaves bounded by the curve $r = \sin(2\theta)$ for $0 \le \theta \le \pi$. The mass density is given by a function δ. See this Math3D graph for a visual.
Archie says: *"The region lies between $r = 0$ and $r = \sin(2\theta)$, so the mass of the leaf is given by*

$$\int_0^{\pi}\int_0^{\sin(2\theta)} \delta(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta."$$

Betty replies: *"Not quite. The radius can be negative so the mass of the leaves is given by*

$$\int_0^{\pi}\int_0^{\sin(2\theta)} \delta(r\cos\theta, r\sin\theta)\,|r|\,dr\,d\theta."$$

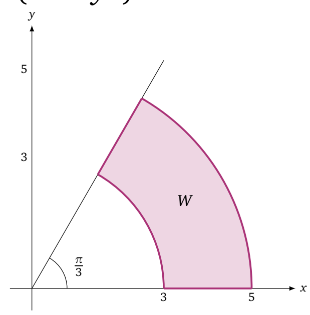
Chuck clarifies *"Still we have to be more careful. The mass of the leaves is given by*

$$\int_0^{\pi/2}\int_0^{\sin(2\theta)} \delta(r\cos\theta, r\sin\theta)\,|r|\,dr\,d\theta + \int_{\pi/2}^{\pi}\int_{\sin(2\theta)}^{0} \delta(r\cos\theta, r\sin\theta)\,|r|\,dr\,d\theta."$$

One of them is truly correct and one of them is accidentally correct. Identify who and explain.

Computations

9.3.7 Let R be the region in $\mathbb{R}^2$ illustrated below. Define $f(x,y) = (x^2+y^2)^{-1/2}$.



(a) Describe W in polar form.

(b) Evaluate $\displaystyle\iint_W f\,dA$ using polar coordinates.

9.3.8 Let R be the region in $\mathbb{R}^2$ bounded by $r = 1+\cos(2\theta)$ with $x > 0$. Evaluate $\displaystyle\iint_R x\,dA$.

9.3.9 Let R be the disk of radius 3 centered at the origin in $\mathbb{R}^2$. Evaluate $\displaystyle\iint_R \cos(x^2+y^2)\,dA$.

9.3.10 Evaluate the integral $\displaystyle\int_0^1\int_0^{\sqrt{1-x^2}} \frac{1}{\sqrt{4-x^2-y^2}}\,dy\,dx$ by converting to polar coordinates

9.3.11 Evaluate the integral $\displaystyle\int_0^2 \int_0^{\sqrt{4-x^2}} (x^2+y^2)^{3/2}\,dy\,dx$ by converting to polar coordinates

9.3.12 Calculate the area of the region of the disk $x^2+y^2 \le 4$ that is to the left of the line $x = -\sqrt{3}$

9.3.13 Let S be the region bounded by $r = 5\sin(2\theta)$ when restricted to the first quadrant. Find the average value of $g(x,y) = \frac{x^2+y^2}{25}$ over S.

9.3.14 Two regions are bounded by the cardiod $r = 1+2\cos\theta$. Find the area of each region.

9.3.15 Find the mass and centre of mass of the region bounded by the cardioid with polar equation $r = 1+\cos\theta$ and the density function $\delta(r,\theta) = r$.

9.3.16 Calculate the volume of the solid below $y = 9-x^2-z^2$ and above the xz-plane.

Proofs

9.3.17 The formal justification for switching to polar coordinates is encapsulated in Theorem 9.3.4. This theorem can express the area of $D = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2+y^2 \le 4\}$ in polar coordinates.

(a) Here is a first attempt.

> 1. *Note that* $D = g([1,2]\times[0,2\pi])$.
> 2. *Since the constant function* $f(x,y) = 1$ *is integrable on* D*, it follows by Theorem 9.3.4 that* $F(r,\theta) = |r| = r$ *is integrable on* $[1,2]\times[0,2\pi]$.
> 3. *Moreover, by Theorem 9.3.4,* $\operatorname{area}(D) = \displaystyle\iint_D 1\,dA = \iint_{[1,2]\times[0,2\pi]} F\,dA$.
> 4. *As* F *is continuous on* $[1,2]\times[0,2\pi]$*, Fubini's theorem implies* $\operatorname{area}(D) = \displaystyle\int_0^{2\pi}\int_1^2 r\,dr\,d\theta$.

In addition to missing details, this argument incorrectly applies a theorem. Identify this error.

(b) Here is a second attempt.

> 1. *Note that* $D = g([1,2]\times[0,2\pi))$.
> 2. *Since the constant function* $f(x,y) = 1$ *is integrable on* D*, it follows by Theorem 9.3.4 that* $F(r,\theta) = |r| = r$ *is integrable on* $[1,2]\times[0,2\pi)$.
> 3. *Moreover,* $\operatorname{area}(D) = \displaystyle\iint_D 1\,dA = \iint_{[1,2]\times[0,2\pi)} F\,dA$.
> 4. *As* F *is continuous on* $[1,2]\times[0,2\pi)$*, Fubini's theorem implies* $\operatorname{area}(D) = \displaystyle\int_0^{2\pi}\int_1^2 r\,dr\,d\theta$.

In addition to missing details, this argument incorrectly applies a theorem. Identify this error.

(c) Fix the flaws of both previous arguments and add any missing missing details. That is, prove

$$\operatorname{area}(D) = \int_0^{2\pi}\int_1^2 r\,dr\,d\theta.$$

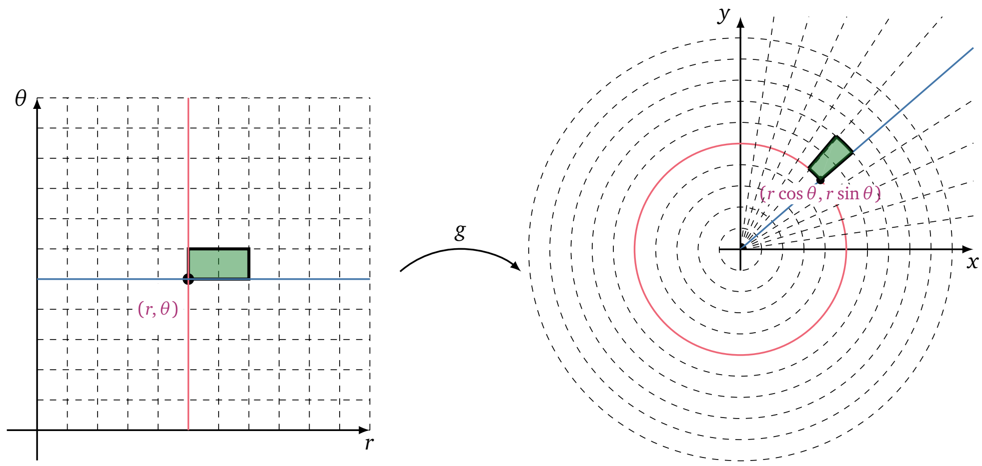
Hint: After applying Theorem 9.3.4, add a second integral.

Applications and beyond

9.3.18 Here you will investigate a key step to deriving the method for integrating using the polar coordinate transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ given by $g(r, \theta) = (r\cos\theta, r\sin\theta)$.

1. *For a fixed* $r > 0$ *and* $\theta \in \mathbb{R}$, *the small rectangle*
$$R = [r, r + \Delta r] \times [\theta, \theta + \Delta\theta]$$
in the (r, θ)*-plane transforms under* g *to a piece of washer* $g(R)$ *in the* (x, y)*-plane.*
2. *Since the rectangle is small, the thickness of the washer is approximately*
$$||g(r + \Delta r, \theta) - g(r, \theta)|| \approx \left|\left|\frac{\partial g}{\partial r}(r, \theta)\right|\right| \Delta r = \left|\left|\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}\right|\right| \Delta r = \Delta r.$$
3. *Similarly, its width is approximately*
$$||g(r, \theta + \Delta\theta) - g(r, \theta)|| \approx \left|\left|\frac{\partial g}{\partial \theta}(r, \theta)\right|\right| \Delta\theta = \left|\left|\begin{bmatrix}-r\sin\theta \\ r\cos\theta\end{bmatrix}\right|\right| \Delta\theta = r\Delta\theta$$
4. *Overall, this implies that* $\operatorname{area}(g(R)) \approx r\Delta r\Delta\theta = r\operatorname{area}(R)$.

(a) Label the four corners of R and four corners of $g(R)$ according to the description in Line 1.



(b) Explain why the author can plausibly use the ≈ symbol in Lines 2 and 3.

(c) Line 4 follows from Lines 2 and 3 with an assumption about $\dfrac{\partial g}{\partial r}$ and $\dfrac{\partial g}{\partial \theta}$. Identify that assumption.

9.4. Fubini's theorem in 3D and higher

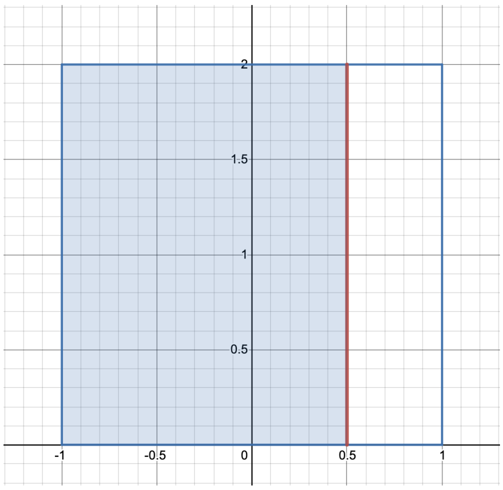
You have related integrals in $\mathbb{R}^2$ to iterated double integrals by Fubini's theorem, which opens up an avenue for efficient calculations. To extend these new ideas to integrals in $\mathbb{R}^3$, you will need to define iterated triple integrals and formulate Fubini's theorem for three dimensional space. Afterwards, you can generalize the statement of Fubini's theorem to any dimension, so that integrals in $\mathbb{R}^n$ can be efficiently calculated by iterated n-fold integrals.

Since you are studying these problems in higher dimensions, you will have fewer or no visuals. For instance, you can interpret iterated double integrals $\int_a^b \int_c^d \varphi dy dx$ in $\mathbb{R}^3$ using the *graph* of φ, as in Example 9.1.6, but this will no longer be possible for triple integrals. You can instead view iterated double integrals in $\mathbb{R}^2$ using the *domain* of φ.

Example 9.4.1 Let $R = [-1,1] \times [0,2]$ and define $\varphi : R \to \mathbb{R}$ by $\varphi(x,y) = x^2 + y$. The iterated double integrals

$$\int_{-1}^{1} \int_{0}^{2} \varphi(x,y) dy dx, \qquad \int_{0}^{2} \int_{-1}^{1} \varphi(x,y) dx dy$$

can be respectively visualized with this $dydx$ Desmos demo and this $dxdy$ Desmos demo. Below are individual frames from this animation.



These pair of animations capture the process of iterated double integrals. Informally speaking, the $dydx$ integral first integrates dy along a fixed x-slice and then totals all of these slices by integrating dx. Similarly, the $dxdy$ integral first integrates dx along a fixed y-slice and then totals all of these slices by integrating dy. Since φ is continuous on the rectangle, Fubini's theorem implies that both iterated double integrals will yield the same result.

This basic intuition with iterated integrals and Fubini's theorem $\mathbb{R}^2$ will carry forward to $\mathbb{R}^3$ and $\mathbb{R}^n$ for $n \geq 4$. There will be more algebraic, combinatorial, and geometric complexities, but remember to rely on this basic intuition.

9.4.1 Integrals of slices in 3D

Before generalizing slices to any dimension, you can see how the process inductively builds from $\mathbb{R}^2$ to $\mathbb{R}^3$. Reducing integration in $\mathbb{R}^3$ to single-variable calculus will take several steps. First, you will need to fix 2 variables and integrate those slices.

Definition 9.4.2 Let $R = [a,b] \times [c,d] \times [e,f]$ be a rectangle in $\mathbb{R}^3$. Let $\varphi : R \to \mathbb{R}$.
A (x,y)**-slice of** φ is a function $\varphi^{x,y} : [e,f] \to \mathbb{R}$ of the form

$$\varphi^{x,y}(z) = \varphi(x,y,z) \qquad \text{for some fixed } x \in [a,b], y \in [c,d].$$

A (x,z)**-slice of** φ is a function $\varphi^{x,z} : [c,d] \to \mathbb{R}$ of the form

$$\varphi^{x,z}(y) = \varphi(x,y,z) \qquad \text{for some fixed } x \in [a,b], z \in [e,f].$$

A (y,z)**-slice of** φ is a function $\varphi^{y,z} : [a,b] \to \mathbb{R}$ of the form

$$\varphi^{y,z}(x) = \varphi(x,y,z) \qquad \text{for some fixed } y \in [c,d], z \in [e,f].$$

You can visualize the corresponding single-variable integrals using the domain of φ in $\mathbb{R}^3$.

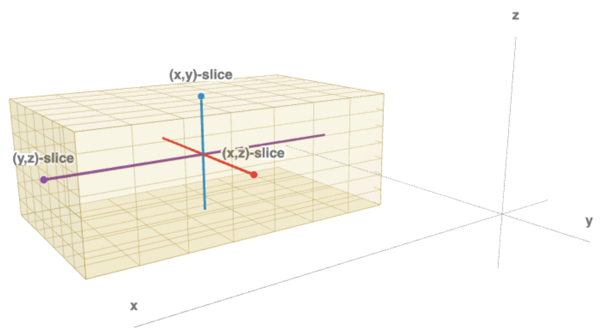
Example 9.4.3 (Cheese – fix two) Let $R = [1,5] \times [-3,-1] \times [0,2]$ be a block of cheese with mass density $\varphi(x,y,z) = ze^{xy}$. Some slices of φ on R with two fixed variables include:

- $f : [0,2] \to \mathbb{R}$ given by $f(z) = \varphi(3,-2,z) = e^{-6}z$ is a (x,y)-slice of φ.
- $g : [-3,-1] \to \mathbb{R}$ given by $g(y) = \varphi(3,y,1) = e^{3y}$ is a (x,z)-slice of φ.
- $h : [1,5] \to \mathbb{R}$ given by $h(x) = \varphi(x,-2,1) = e^{-2x}$ is a (y,z)-slice of φ.

You can interpret the three single-variable integrals

$$\int_0^2 \varphi(3,-2,z)dz, \qquad \int_{-3}^{-1} \varphi(3,y,1)dy, \qquad \int_1^5 \varphi(x,-2,1)dx$$

as the "1D mass density" of three different 1-dimensional cheese strings; see Math3D.



This concludes the first step. Second, you will need to fix 1 variable.

Definition 9.4.4 Let $R = [a,b] \times [c,d] \times [e,f]$ be a rectangle in $\mathbb{R}^3$. Let $\varphi : R \to \mathbb{R}$.
A x**-slice of** φ is a function $\varphi^{x} : [c,d] \times [e,f] \to \mathbb{R}$ of the form

$$\varphi^{x}(y,z) = \varphi(x,y,z) \qquad \text{for some fixed } x \in [a,b].$$

A y**-slice of** φ is a function $\varphi^{y} : [a,b] \times [e,f] \to \mathbb{R}$ of the form

$$\varphi^{y}(x,z) = \varphi(x,y,z) \qquad \text{for some fixed } y \in [c,d].$$

A z**-slice of** φ is a function $\varphi^{y,z} : [a,b] \times [c,d] \to \mathbb{R}$ of the form

$$\varphi^{z}(x,y) = \varphi(x,y,z) \qquad \text{for some fixed } z \in [e,f].$$

You can visualize the two-variable integrals using the domain of φ in $\mathbb{R}^3$. By Fubini's theorem in $\mathbb{R}^2$, these integrals can be viewed as more than one iterated double integral.

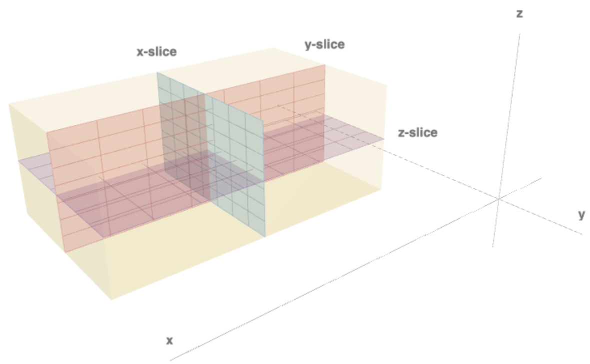
Example 9.4.5 (Cheese – fix one) Continue with Example 9.4.3 so $R = [1,5]\times[-3,-1]\times[0,2]$ is a block of cheese with mass density $\varphi(x,y,z) = ze^{xy}$. Some examples of slices of φ on R with one fixed variable include:

- $F : [-3,-1]\times[0,2] \to \mathbb{R}$ given by $F(y,z) = \varphi(3,y,z) = ze^{3y}$ is a x-slice of φ.
- $G : [1,5]\times[0,2] \to \mathbb{R}$ given by $G(x,z) = \varphi(x,-2,z) = ze^{-2x}$ is a y-slice of φ.
- $H : [1,5]\times[-3,-1] \to \mathbb{R}$ given by $H(x,y) = \varphi(x,y,1) = e^{xy}$ is a z-slice of φ.

The three double integrals

$$\iint_{[-3,-1]\times[0,2]} F\,dA, \qquad \iint_{[1,5]\times[0,2]} G\,dA, \qquad \iint_{[1,5]\times[-3,1]} H\,dA \tag{9.4.1}$$

can be interpreted as the "2D mass density" of three different 2-dimensional cheese sheets. View this Math3D demo for a better visual.



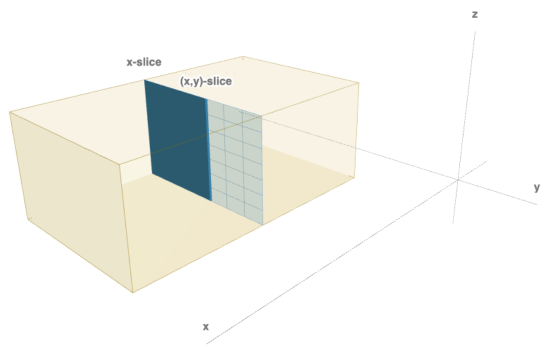
You can express these integrals as iterated double integrals with Fubini's theorem in two dimensions. For instance, if $F(y,z) = \varphi(3,y,z)$ is integrable on $[-3,-1]\times[0,2]$, and for each $y \in [-3,-1]$, the y-slice $F^y(z) = \varphi(3,y,z)$ is integrable on $[0,2]$, then

$$\iint_{[-3,-1]\times[0,2]} F\,dA = \int_{-3}^{-1}\int_{0}^{2} \varphi(3,y,z)\,dz\,dy$$

by Fubini's theorem in 2D. Notice that these assumptions are equivalent to:

- For $x = 3$, the x-slice of φ is integrable on $[-3,-1]\times[0,2]$.
- For $x = 3$ and $y \in [-3,1]$, the (x,y)-slice of φ is integrable on $[0,2]$.

The diagram below illustrates how the iterated double integral $dz\,dy$ is calculated.



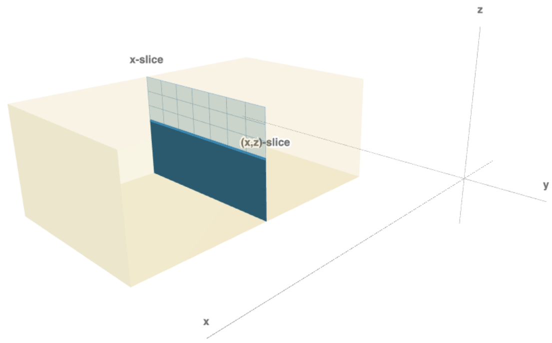
You can repeat the same argument to express this integral $dydz$ by modifying the assumptions. If instead you assume:

- For $x = 3$, the x-slice of φ is integrable on $[-3,-1] \times [0,2]$,
- For $x = 3$ and $z \in [0,2]$, the (x,z)-slice of φ is integrable on $[-3,-1]$,

then by Fubini's theorem in 2D,

$$\iint_{[-3,-1]\times[0,2]} F\,dA = \int_0^2 \int_{-3}^{-1} \varphi(3,y,z)dydz.$$

The diagram below illustrates how the iterated double integral $dydz$ is calculated.



Play with this Math3D demo to visualize the iterated double integrals $dzdy$ and $dydz$. With the correct assumptions, you can use 2D Fubini to express the integrals of G and H from (9.4.1) as iterated double integrals, namely either $dxdz$ or $dzdx$ for the integral of G, and either $dxdy$ or $dydx$ for the integral of H.

This showcases all possible integrals of slices in $\mathbb{R}^3$. Notice how this process inductively applied Fubini's theorem in 2D. Each integral of a slice can be expressed as many different iterated integrals provided you can apply Fubini's theorem.

9.4.2 Iterated triple integrals and Fubini's theorem in 3D

By iteratively applying these ideas, you can propose a method to efficiently calculate three-dimensional integrals.

Definition 9.4.6 Let $\varphi : [a,b] \times [c,d] \times [e,f] \to \mathbb{R}$ be a bounded function. The quantity

$$\int_a^b \int_c^d \int_e^f \varphi(x,y,z)dzdydx$$

is an **iterated triple integral**.

Remark 9.4.7 It is left as an exercise to write down the 5 other iterated triple integrals, namely $dxdydz$, $dxdzdy$, $dydxdz$, $dydzdx$, and $dzdxdy$.

As expected, you can swiftly calculate these integrals.

Example 9.4.8 (Cheese – iterated triple integral) Continue with Example 9.4.5 so $R = [1,5] \times [-3,-1] \times [0,2]$ is a block of cheese with mass density $\varphi(x,y,z) = ze^{xy}$. By applying the fundamental theorem of calculus three times, you can compute the iterated triple integral

$dzdydx$ as

$$\begin{aligned}
\int_1^5 \int_{-3}^{-1} \int_0^2 ze^{xy}\,dzdydx &= \int_1^5 \int_{-3}^{-1} \frac{z^2 e^{xy}}{2}\bigg|_{z=0}^{z=2} dydx \\
&= \int_1^5 \int_{-3}^{-1} 2e^{xy}\,dydx \\
&= \int_1^5 \frac{2e^{xy}}{x}\bigg|_{y=-3}^{y=-1} dx = \int_1^5 \frac{2e^{-x} - 2e^{-3x}}{x} dx \approx 0.410375.
\end{aligned}$$

The last integral was numerically approximated with WolframAlpha. Does this value represent the mass of the cheese $\iiint_R \varphi dV$? You need to generalize Fubini's theorem.

Notice Definition 9.4.6 requires that the iterated integral $\int_c^d \int_e^f \varphi(x,y,z)dzdy$ exists for every $x \in [a,b]$ and the integral $\int_e^f \varphi(x,y,z)dz$ exists for every $x \in [a,b], y \in [c,d]$. It is not obvious when these conditions will be satisfied, and even if they are satisfied, it is not clear whether the iterated triple integral will exist either. Finally, even if all of these iterated integrals exist, the integral $\iiint_R \varphi dV$ may not exist and, even if it does, it may not be equal to the iterated integrals. The counterexamples from Section 9.1 can be generalized to show that all of these concerns are genuine. Fubini's theorem in $\mathbb{R}^3$ allows you to avoid all of these pathological counterexamples and confirm your natural intuition.

Theorem 9.4.9 (Fubini) Let $R = [a,b] \times [c,d] \times [e,f]$ and let $\varphi : R \to \mathbb{R}$ be bounded. If

- For every $x \in [a,b], y \in [c,d]$, the (x,y)-slice $\varphi^{(x,y)}$ is integrable on $[e,f]$.
- For every $x \in [a,b]$, the x-slice φ^x is integrable on $[c,d] \times [e,f]$.
- φ is integrable on $R = [a,b] \times [c,d] \times [e,f]$.

Then the iterated triple integral

$$\int_a^b \int_c^d \int_e^f \varphi(x,y,z)dzdydx$$

exists and is equal to the integral $\displaystyle\iiint_R \varphi dV$.

Proof. Omitted. The ideas are the same as the two-dimensional case; you must proceed by a carefully phrased induction. ■

By rewriting this theorem for all 6 iterated triple integrals, you obtain a nice corollary.

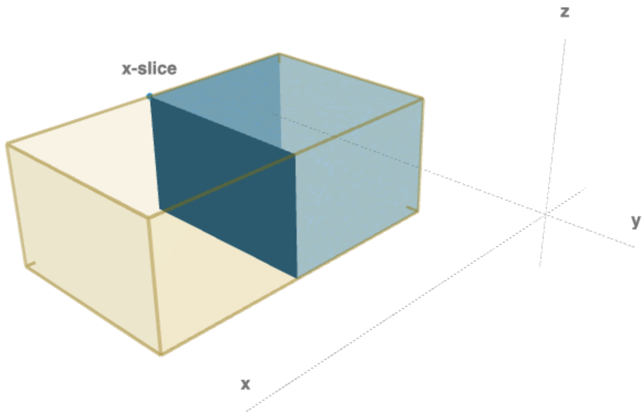
Corollary 9.4.10 Let $R = [a,b]\times[c,d]\times[e,f]$ be a rectangle in $\mathbb{R}^3$. If $\varphi : R \to \mathbb{R}$ is continuous, then every iterated triple integral of φ on R exists and they are all equal to $\iiint_R \varphi dV$.

You can apply Fubini's theorem to your running example of cheese.

Example 9.4.11 (Cheese – Fubini) Again, $R = [1,5] \times [-3,-1] \times [0,2]$ is a block of cheese with mass density $\varphi(x,y,z) = ze^{xy}$. Since φ is continuous, Fubini's theorem (Corollary 9.4.10) implies that the mass of the cheese can be expressed as an iterated triple integral; for instance, its mass is given by

$$\iiint_R \varphi\, dV = \int_1^5 \int_{-3}^{-1} \int_0^2 \varphi(x,y,z)\,dz\,dy\,dx \approx 0.410375$$

from Example 9.4.8. The outermost integral dx can be viewed as totaling the mass of all x-slices, one of which appeared in Example 9.4.5; see this Math3D demo.



9.4.3 Iterated integrals and Fubini's theorem in any dimension

Now, you can generalize these ideas any dimensions. Unsurprisingly, the notation needed to precisely describe the assumptions in Fubini's theorem is rather hideous. You would need a way of describing "all possible slices" of a function. Instead of writing this down with a web of indices, here is a less precise definition of a slice.

Definition 9.4.12 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a real-valued function. A function g is a **slice** of f on R if g is defined by fixing one or more coordinates of f in R.

Remark 9.4.13 Notice f is not a slice of itself. Also, all coordinates cannot be fixed; otherwise, g would have empty domain.

Example 9.4.14 Let $\varphi : [a,b] \times [c,d] \times [e,f] \to \mathbb{R}$. There are 6 different types of slices of φ, namely x-slices, y-slices, z-slices, (x,y)-slices, (x,z)-slices, and (y,z)-slices.

These definitions and examples illustrate how the geometry and combinatorics gets more complicated in higher dimensions. Nonetheless, you can verify standard properties of slices.

Lemma 9.4.15 Let R be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a real-valued function.
(a) If f is bounded, then every slice of f is bounded.
(b) If f is continuous, then every slice of f is continuous.

Proof. This is left as a short exercise. It follows directly from definitions. ■

Integrals of slices can be expressed as iterated integrals. The number of iterated integrals depends on how many variables you fix.

Example 9.4.16 Let $\varphi : [a,b] \times [c,d] \times [e,f] \to \mathbb{R}$. The integral of an (x,y)-slice can be expressed as 1 single-variable integral dz. The integral of a x-slice can possibly be expressed as 2 iterated double integrals $dy\,dz$ or $dz\,dy$, depending on whether Fubini's theorem applies.

By slicing and integrating iteratively as you did for $\mathbb{R}^3$, you can construct the proposed quantity that will efficiently compute any n-dimensional integral.

Definition 9.4.17 Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. Let $\varphi : R \to \mathbb{R}$ be a bounded function. The quantity

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} \varphi(x_1, \ldots, x_n) dx_n \cdots dx_2 \, dx_1$$

is an **iterated n-fold integral**.

There are a total of $n!$ iterated integrals depending on how you order the symbols $dx_1, \ldots, dx_n$. Computationally, these behave the same as iterated triple integrals where you repeatedly apply the fundamental theorem of calculus. If the dimension $n \in \mathbb{N}^+$ is arbitrary, then induction is a common technique.

Example 9.4.18 It is left as an exercise to prove by induction on $n \in \mathbb{N}^+$ that

$$\int_0^1 \int_0^1 \cdots \int_0^1 x_1 x_2^2 \cdots x_n^n dx_n \cdots dx_2 dx_1 = \frac{1}{(n+1)!}.$$

Finally, you can conjecture Fubini's theorem by analyzing the statements for $\mathbb{R}^2$ (Theorem 9.1.17) and $\mathbb{R}^3$ (Theorem 9.4.9). The key observation can be informally summarized.

If the function and its slices are integrable, then the iterated integral equals the integral.

This insight can be made rigorous into Fubini's theorem for any dimension.

Theorem 9.4.19 (Fubini) Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. Let $f : R \to \mathbb{R}$ be a bounded real-valued function. If f is integrable on R and every slice of f on R is integrable on its domain, then every iterated integral of f on R exists and they are all equal to the integral of f on R. That is, the iterated n-fold integral

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_n \cdots dx_1$$

exists and all $n!$ orderings of this iterated integral exist and are all equal to $\displaystyle\int_R f \, dV$.

Proof. Omitted. The proof uses induction and a key theorem is listed below. ∎

Combined with Lemma 9.4.15, this theorem immediately yields an elegant corollary.

Corollary 9.4.20 Let R be a rectangle in $\mathbb{R}^n$. If $f : R \to \mathbb{R}$ is continuous, then every iterated integral of f on R exists and they are all equal to the integral of f on R.

The proof of Theorem 9.4.19 rests on the following theorem.

Theorem 9.4.21 (Fubini) Let R be a rectangle in $\mathbb{R}^n$. Let $\varphi : R \times [a, b] \to \mathbb{R}$ be bounded. For every $t \in [a, b]$, define the slice $\varphi^t : R \to \mathbb{R}$ by $\varphi^t(x) = \varphi(x, t)$. If the function φ is integrable on $R \times [a, b]$, and for every $t \in [a, b]$ the slice φ^t is integrable on R, then the function $t \mapsto \int_R \varphi^t dV$ is integrable on $[a, b]$ and

$$\int_{R\times[a,b]} \varphi dV = \int_a^b \left(\int_R \varphi^t dV \right) dt.$$

Proof. Omitted. Follow the proof of Fubini in $\mathbb{R}^2$ (Theorem 9.1.17) in Section 9.1.4. ■

Rigorously verifying the assumptions in Fubini's theorem can simple if your integrand is continuous but it can often be quite hard. At least the hypotheses are only about integrability of functions on rectangles; this reduction is manageable with your theorems on integration from the previous chapter (e.g. Theorem 7.7.4). You should be able to verify the hypotheses, but it is not practical to do so in every situation.

> *For the remainder of this text, you will assume that the hypotheses of Fubini's theorem hold for any integral that you encounter.*

Overall, you have made monumental progress towards a viable strategy for computing integrals in any dimension. The many versions of Fubini's theorem provide a gateway for expressing the integral as an iterated integral, where you can unleash the fundamental theorem of calculus and your plethora of integration techniques. Paralleling your progress in two-dimensions, you can now focus on building strategies for calculating integrals in three dimensions, especially by using linear algebra, symmetries, and geometry.

Exercises for Section 9.4

Concepts and definitions

9.4.1 Let R be a rectangle in $\mathbb{R}^3$ and $f : R \to \mathbb{R}$ be bounded. True or false? Do not attempt to justify.

(a) If f is integrable on R, then all 6 iterated triple integrals of f over R exist.

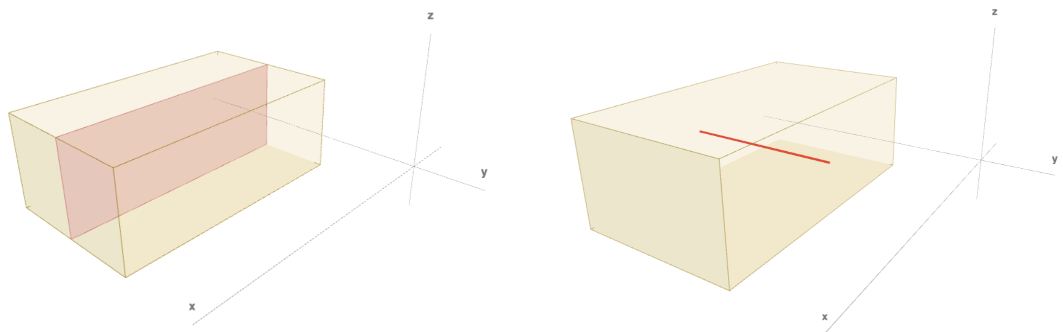
(b) If all 6 iterated triple integrals of f over R exist, then f is integrable on R.

9.4.2 Determine the assumptions on φ needed to express the integral over G in equation (9.4.1) as an iterated double integral $dxdz$ or as an iterated double integral $dzdx$.

9.4.3 In addition to Definition 9.4.6, write down the 5 other iterated triple integrals, namely $dxdydz$, $dxdzdy$, $dydxdz$, $dydzdx$, and $dzdxdy$.

9.4.4 Rewrite Fubini's theorem for the $dxdzdy$ iterated triple integral.

9.4.5 The rectangle $R = [1,5] \times [-3,-1] \times [0,2]$ is a block of cheese with *continuous* mass density $\varphi(x,y,z)$. Inside the block of cheese, you have sliced a cheese string (one-dimensional) and a cheese sheet (two-dimensional) as illustrated below.



(a) Which integral represents the 1D mass density of the cheese string? Select ONE.

$$A = \int_0^2 \varphi(3,-2,z)dz \qquad B = \int_{-3}^{-1} \varphi(3,y,1)dy \qquad C = \int_1^5 \varphi(x,-2,1)dx$$

(b) Which integral represents the 2D mass density of the cheese sheet? Select ONE.

$$A = \int_1^5 \int_{-3}^{-1} \varphi(x,y,1)dxdy \qquad B = \int_1^5 \int_0^2 \varphi(x,-2,z)dxdz \qquad C = \int_1^5 \int_{-3}^{-1} \varphi(3,y,z)dydz$$

$$D = \int_1^5 \int_{-3}^{-1} \varphi(x,y,1)dydx \qquad E = \int_1^5 \int_0^2 \varphi(x,-2,z)dzdx \qquad F = \int_1^5 \int_{-3}^{-1} \varphi(3,y,z)dzdy$$

(c) Which integral(s) represent the 3D mass of the cheese block? Select ALL.

$$A = \int_1^5 \int_{-3}^{-1} \int_0^2 \varphi(x,y,z)dxdydz \qquad B = \iiint_{[1,5]\times[-3,-1]\times[0,2]} \varphi dV$$

$$C = \int_0^2 \int_{-3}^{-1} \int_1^5 \varphi(x,y,z)dxdydz \qquad D = \iiint_{[0,2]\times[-3,-1]\times[1,5]} \varphi dV$$

9.4.6 Let $R = [a,b] \times [c,d] \times [e,f]$ be a rectangle in $\mathbb{R}^3$ and let $\varphi : R \to \mathbb{R}$ be bounded. Consider 7 different assumptions you might use to apply Fubini's theorem in $\mathbb{R}^3$.

(A) φ is integrable on $[a,b] \times [c,d] \times [e,f]$.
(B) For every $x \in [a,b]$, the x-slice φ^x is integrable on $[c,d] \times [e,f]$
(C) For every $y \in [c,d]$, the y-slice φ^y is integrable on $[a,b] \times [e,f]$
(D) For every $z \in [e,f]$, the z-slice φ^z is integrable on $[a,b] \times [c,d]$
(E) For every $x \in [a,b], y \in [c,d]$, the (x,y)-slice $\varphi^{x,y}$ is integrable on $[e,f]$
(F) For every $x \in [a,b], z \in [e,f]$, the (x,z)-slice $\varphi^{x,z}$ is integrable on $[c,d]$
(G) For every $y \in [c,d], z \in [e,f]$, the (y,z)-slice $\varphi^{y,z}$ is integrable on $[a,b]$.

(a) $\displaystyle\int_a^b \varphi(x,y,z)dx$ exists provided which assumption(s) hold?

(b) $\displaystyle\int_c^d \int_a^b \varphi(x,y,z)dxdy$ exists provided which assumption(s) hold?

(c) $\displaystyle\int_e^f \int_c^d \int_a^b \varphi(x,y,z)dxdydz$ exists provided which assumption(s) hold?

(d) $\displaystyle\int_R \varphi dV = \int_e^f \int_c^d \int_a^b \varphi(x,y,z)dxdydz$ provided which assumption(s) hold?

Computations

9.4.7 Calculate $\displaystyle\iiint_{[0,1]\times[0,2]\times[0,3]} 6xz^2 dV.$

9.4.8 Express the volume of the unit ball in $\mathbb{R}^3$ as an iterated triple integral and evaluate it. You may apply Fubini's theorem without justification.

9.4.9 Calculate by induction on $n \in \mathbb{N}^+$ that

$$\int_0^1 \int_0^1 \cdots \int_0^1 x_1 x_2^2 \cdots x_n^n dx_n \cdots dx_2 dx_1 = \frac{1}{(n+1)!}.$$

Proofs

9.4.10 If the integrand is not continuous, you must carefully verify several assumptions to apply Fubini's theorem. Define $R = [0,2] \times [4,6] \times [1,3]$ and $\varphi : R \to \mathbb{R}$ by

$$\varphi(x,y,z) = \begin{cases} -1 & 0 \le x \le 2, 4 \le y \le 6, 1 \le z < 2 \\ 7 & 0 \le x \le 2, 4 \le y \le 6, 2 \le z \le 3. \end{cases}$$

(a) What are the (x,y)-slices of φ? Prove that they are integrable on their domains.

(b) What are the x-slices of φ? Prove that they are integrable on their domains.

(c) Prove that φ is integrable on R.

(d) Conclude that $\displaystyle\int_0^2 \int_4^6 \int_1^3 \varphi(x,y,z)dzdydx$ exists and $\displaystyle\iiint_R \varphi dV = \int_0^2 \int_4^6 \int_1^3 \varphi(x,y,z)dzdydx.$

(e) Evaluate $\iiint_R \varphi \, dV$.

9.4.11 Let S be the solid in $\mathbb{R}^3$ bounded by the cylinder $y = x^2$, the plane $z = 0$, and the plane $y + z = 4$. Suppose you want to calculate the integral of $\varphi(x,y,z) = ye^{xz}$ over S. You will want to express it as an iterated integral by Fubini's theorem. You will first set up the desired iterated integral and then list what you need to verify to apply Fubini's theorem. View this Math3D demo for each part to help you visualize.

(a) The rectangle $R =$ ____________________ in $\mathbb{R}^3$ contains S so $\int_S \varphi \, dV = \int_R \chi_S \varphi \, dV$.

(b) Toggle the "*single*" button ON. Move the "*single integral*" slider. This animation represents the integral of a slice of $\chi_S \varphi$ with 2 coordinates fixed. Write this integral for an *arbitrary* slice and define the slice as a 1-variable function.

(c) Toggle the "*single*" button OFF and the "*double*" button ON. Move the "*double integral*" slider. This animation represents the integral of an arbitrary slice of $\chi_S \varphi$ with 1 coordinate fixed. Define the slice as a 2-variable function. Write its integral in **two different ways**: as an integral over a set in $\mathbb{R}^2$ and as a (double) iterated integral.

(d) Toggle the "*double*" button OFF and the "*triple*" button ON. Move the "*triple integral*" slider. This animation represents the integral of $\chi_S \varphi$. Write its integral in **two different ways**: as an integral over a set in $\mathbb{R}^3$ and as a (triple) iterated integral.

(e) Fill in the blank: "*I want to show by Fubini's theorem that*

$$\int_R \chi_S \varphi \, dV = \underline{\hspace{8cm}}."$$

(f) Create the checklist of assumptions that you must verify to apply Fubini's theorem.

- ☐ The 3-variable function ________________ is integrable on R.
  That is, the integral ________________ exists.
- ☐ The 2-variable functions ________________ are integrable on __________.
  That is, the integrals ________________ all exist.
- ☐ The 1-variable functions ________________ are integrable on __________.
  That is, the integrals ________________ all exist.

9.4.12 Rigorously verify the assumptions in Exercise 9.4.11 to apply Fubini's theorem as described. Beware that this exercise is rather long.

9.4.13 Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a rectangle in $\mathbb{R}^n$. For $i \in \{1, \ldots, n\}$, let $\varphi_i : [a_i, b_i] \to \mathbb{R}$ be continuous. Define $\varphi : R \to \mathbb{R}$ by $\varphi(x_1, \ldots, x_n) = \varphi_1(x_1) \cdots \varphi_n(x_n)$. Prove by induction on $n \in \mathbb{N}^+$ that

$$\int_R \varphi \, dV = \left(\int_{a_1}^{b_1} \varphi_1(x_1) dx_1 \right) \left(\int_{a_2}^{b_2} \varphi_2(x_2) dx_2 \right) \cdots \left(\int_{a_n}^{b_n} \varphi_n(x_n) dx_n \right).$$

Applications and beyond

9.4.14 The classic form of Fubini's theorem (Theorem 9.4.21) is used to deduce the other versions. Assuming Theorem 9.4.21, Korra hastily attempts to prove Fubini's theorem in 3D (Theorem 9.4.9).

1. *Let* $f : [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \to \mathbb{R}$ *be bounded.*
2. *Then*

$$\iiint_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f\,dV = \int_{a_3}^{b_3} \Big(\iint_{[a_1,b_1]\times[a_2,b_2]} f^z dV \Big) dz$$
$$= \int_{a_3}^{b_3} \Big(\int_{a_2}^{b_2} \Big(\int_{[a_1,b_1]} f^{(y,z)} dV \Big) dy \Big) dz$$
$$= \int_{a_3}^{b_3} \Big(\int_{a_2}^{b_2} \Big(\int_{a_1}^{b_1} f(x,y,z) dx \Big) dy \Big) dz.$$

Her argument has no serious errors, but she really needs to add more explanation.

(a) Korra forgets to define two pieces of notation. Define them precisely with quantifiers.

(b) Korra first applies Theorem 9.4.21 in the first equality. State the choices of R, φ, and $[a, b]$ that she uses. What assumption must she make?

(c) Korra again applies Theorem 9.4.21 in the second equality. State the choices of R, φ, and $[a, b]$ that she uses. What assumption must she make?

9.4.15 Fix $n \in \mathbb{N}^+$. Let R be a rectangle in $\mathbb{R}^n$ and $\varphi : R \to \mathbb{R}$.

(a) Count the number of types of slices of φ.

(b) Fix $k \in \mathbb{N}^+$. Count the number of types of slices of φ fixing k coordinates.

(c) Count the number of iterated integrals of φ over R.

9.4.16 Conjecture a version of Fubini's theorem (Theorem 9.4.19) which only assumes n slices are integrable. You do not need to prove it.

9.4.17 For $n \in \mathbb{N}^+$ and $R > 0$, let $V_n(R)$ be the volume of the n-dimensional ball $B_R(0)$ in $\mathbb{R}^n$. You may apply Fubini's theorem without justification.

(a) Express $V_n(R)$ recursively in terms of the $(n-2)$-dimensional volume function $V_{n-2}(\cdot)$.

(b) Express $V_n(R)$ as an n-dimensional iterated integral with respect to $dx_n dx_{n-1} \cdots dx_2 dx_1$.

(c) For $k \in \mathbb{N}^+$, prove by induction that $V_{2k}(R) = \dfrac{\pi^k}{k!} R^{2k}$.

(d) Conclude the rather bizarre result: the volume of the unit $(2k)$-ball tends to zero as $k \to \infty$. (The same is true for odd dimensions but the volume formula is a bit different.)

9.5. Triple integrals

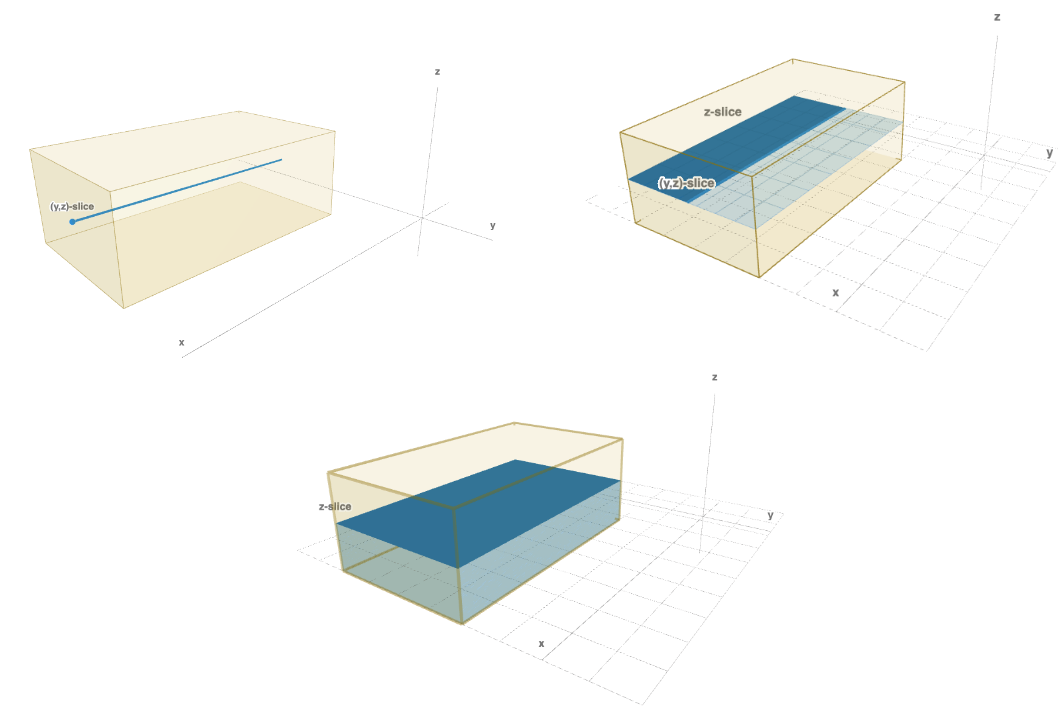
With your experience in two dimensions, you can practice strategies for calculating iterated triple integrals and identifying their symmetries. Many of the same ideas will carry over. You can still evaluate by the fundamental theorem of calculus, swap the order of integration, break up the region into a few pieces, interpret as volumes of classic objects, or apply symmetries. The core new challenge will be the increased geometric and algebraic complexity of $\mathbb{R}^3$.

Again, there is no fixed algorithm that will always succeed in evaluating or manipulating iterated triple integrals. Instead of practicing all possible strategies, you will explore selected examples in this section that focus on this increased geometric complexity and how to address it. As usual, you will apply Fubini's theorem without justification.

Example 9.5.1 Let $R = [1,5] \times [-3,-1] \times [0,2]$ be your favourite block of cheese with mass density $\varphi(x,y,z) = ze^{xy}$. You calculated its mass in Examples 9.4.8 and 9.4.11 by expressing its mass as an integral $dz\,dy\,dx$. By Fubini's theorem, you can calculate the mass using a different order of integration, such as $dx\,dy\,dz$. Since R is a rectangle, swapping the order of integration takes no effort. The mass can be expressed as

$$\iiint_R \varphi\,dV = \int_0^2 \int_{-3}^{-1} \int_1^5 ze^{xy}\,dx\,dy\,dz = \int_0^2 \int_{-3}^{-1} z\frac{e^{5y}-e^{y}}{y}\,dy\,dz \approx 0.410375.$$

The iterated double integral was evaluated numerically since there is no elementary antiderivative for the integrand. There was no advantage to switching the order of integration, but the geometry of how the ordering $dx\,dy\,dz$ differs compared to $dz\,dy\,dx$. In particular, for $dx\,dy\,dz$, the outermost integral dz can be viewed as totaling the 2D mass density of the z-slices to compute the 3D mass; dy totals the 1D mass density of the (y,z)-slices to compute 2D mass density; and dx computes 1D mass density of a single (y,z)-slice.



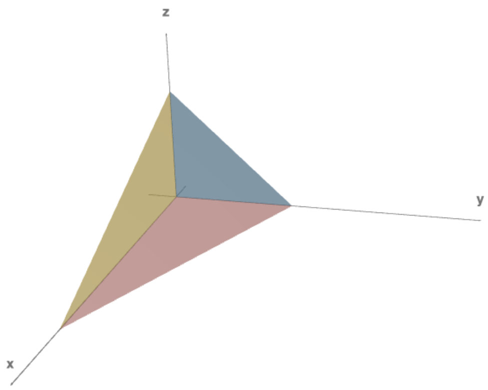
Watch this Math3D dx demo, Math3D dy demo, and Math3D dz demo. Keep these visuals in mind as you continue to more intricate examples.

If your region is not a rectangle then, broadly speaking, there are two key approaches which can help you reduce dimension. The next example outlines one of them: the *projection method*[2]. The basic idea is to project the solid into a coordinate plane, and integrate over that lower-dimensional projection.

Example 9.5.2 Let S be a solid bounded by $x+2y+6z=4$ in the first octant with mass density $\delta(x,y,z)=e^{x+y-z}$. Formally speaking,

$$S=\{(x,y,z)\in\mathbb{R}^3 : x+2y+6z\le 4, x\ge 0, y\ge 0, z\ge 0\}.$$

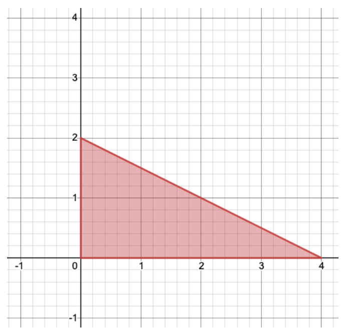
View this Math3D demo for a better visual of the solid.



You can find its mass m in six different ways by changing the order of integration. Here you will see three different ways.

As a first approach, you can project S into the (x,y)-plane, which gives the triangle

$$T_1=\{(x,y)\in\mathbb{R}^2 : 0\le y\le 2, 0\le x\le 4-2y\}\subseteq\mathbb{R}^2.$$



For a fixed $(x,y)\in T_1$, the point (x,y,z) lies in S if and only if (x,y,z) lies above $z=0$ and below $x+2y+6z=4$. This gives the condition $0\le z\le \frac{2}{3}-\frac{1}{3}y-\frac{1}{6}x$ and so

$$m=\int_0^2\int_0^{4-2y}\int_0^{\frac{2}{3}-\frac{1}{3}y-\frac{1}{6}x} e^{x+y-z}\,dz\,dx\,dy.$$

This gives one expression. Alternatively, you can describe the same triangle T_1 as

$$T_1=\{(x,y)\in\mathbb{R}^2 : 0\le x\le 4, 0\le y\le 2-\frac{x}{2}\}.$$

[2]This method name is not standard in other sources, but this naming hopefully helps you remember the approach.

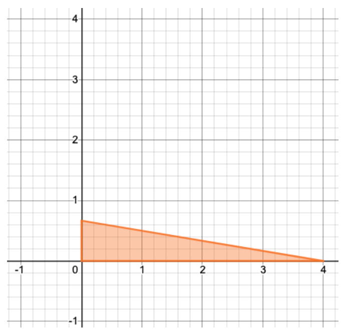
By the same reasoning for the dz-integral, it follows that

$$m = \int_0^4 \int_0^{2-\frac{1}{2}x} \int_0^{\frac{2}{3}-\frac{1}{3}y-\frac{1}{6}x} e^{x+y-z}\, dz\, dy\, dx.$$

This gives a second expression.

For a third expression, you can also project S into the (x,z)-plane, giving the triangle

$$T_2 = \{(x,z) \in \mathbb{R}^2 : 0 \le x \le 4, 0 \le z \le \frac{2}{3} - \frac{1}{6}x\}.$$



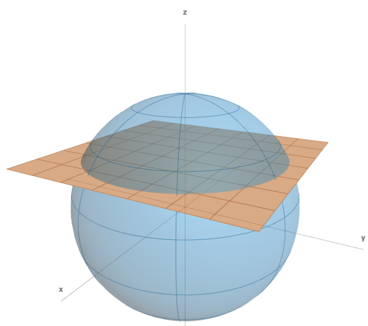
For a fixed $(x,z) \in T_2$, the point (x,y,z) lies in S if and only if (x,y,z) lies between $y = 0$ and $x + 2y + 6z = 4$. This gives the condition $0 \le y \le 2 - 3z - \frac{1}{2}x$ and so

$$m = \int_0^4 \int_0^{\frac{2}{3}-\frac{1}{6}x} \int_0^{2-3z-\frac{1}{2}x} e^{x+y-z}\, dy\, dz\, dx.$$

This gives a third expression. It is left as an exercise to find the three other expressions.

The projections are strong supporting evidence for your claims but you may feel they are redundant and you could find it by simply manipulating inequalities. For a simple solid like in Example 9.5.2, you probably can find these integral expressions without drawing the projections. However, for more complicated solids, this *projection method* can be essential. The second key approach to reducing dimensions for calculating integrals is the *slicing method*[3].

Example 9.5.3 Here you will setup a triple integral which gives the volume of the solid S bounded by the sphere $x^2 + y^2 + z^2 = 4$ and above the $z = 1$ plane.



The projection of S into the (x,y)-plane corresponds to bottom of the solid cut by $z = 1$, so this projection is the disk $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 3\}$. The disk can be rewritten with

[3]Again, this method name is not standard but is intended to be good mnemonic.

routine inequalities, namely

$$D = \{(x,y) \in \mathbb{R}^2 : -\sqrt{3} \le x \le \sqrt{3}, -\sqrt{3-x^2} \le y \le \sqrt{3-x^2}\}.$$

Now, for a fixed $(x,y) \in D$, the point $(x,y,z) \in \mathbb{R}^3$ lies inside S if and only if (x,y,z) lies above the plane $z = 1$ and below the top half of the sphere $x^2+y^2+z^2 = 4$, that is, $1 \le z \le \sqrt{4-x^2-y^2}$. Thus,

$$\mathrm{vol}(S) = \int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{3-x^2}}^{\sqrt{3-x^2}} \int_1^{\sqrt{4-x^2-y^2}} 1 dz dy dx,$$

as required.

Another method is by slicing. The solid S lies between $1 \le z \le 2$. For each fixed $z \in [1,2]$, the z-slice of S is given by the disk

$$S_z = \{(x,y) \in \mathbb{R}^2 : x^2+y^2+z^2 \le 4\} = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \le 4-z^2\} \subseteq \mathbb{R}^2.$$

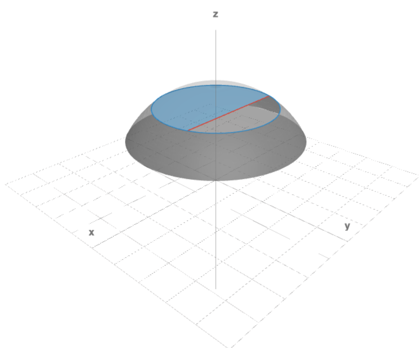
View this Desmos animation of these slices. The radius of this disk is $r(z) = \sqrt{4-z^2}$ so this z-slice can be described as

$$\begin{aligned} S_z &= \{(x,y) \in \mathbb{R}^2 : -r(z) \le y \le r(z),\ -\sqrt{r(z)^2-y^2} \le x \le \sqrt{r(z)^2-y^2}\} \\ &= \{(x,y) \in \mathbb{R}^2 : -\sqrt{4-z^2} \le y \le \sqrt{4-z^2},\ -\sqrt{4-z^2-y^2} \le x \le \sqrt{4-z^2-y^2}\} \end{aligned}$$

The volume of S can therefore be given by

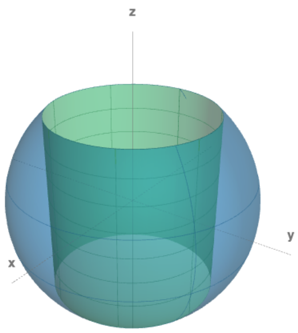
$$\mathrm{vol}(S) = \int_1^2 \mathrm{area}(S_z) dz = \int_1^2 \int_{-\sqrt{4-z^2}}^{\sqrt{4-z^2}} \int_{-\sqrt{4-z^2-y^2}}^{\sqrt{4-z^2-y^2}} 1 dx dy dz.$$

Geometrically speaking, this method is equivalent to finding the area of a cross section of the region (the inner two integrals), and then adding up the contributions of each cross section to get the total volume. View this Math3D demo illustrating this slicing method in $\mathbb{R}^3$.



The underlying principle of the *projection method* is to determine for each point in the projection, which surfaces do the projected points lie between; this may vary depending on the point's location in the projection. This requires careful sketching and geometric considerations. The underlying principle of the *slicing method* is to determine the two-dimension slice after fixing a variable; this may vary depending on the location of the slice. This reduces your problem to $\mathbb{R}^2$ but still requires careful sketching and geometric considerations.

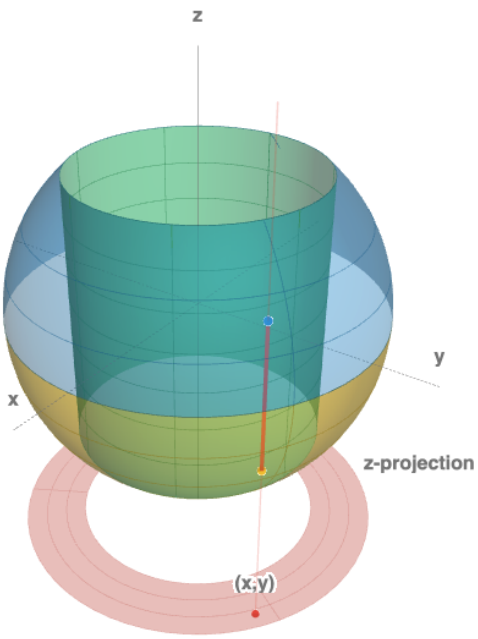
Example 9.5.4 Here you will set up a triple integral which gives the volume of the solid S bounded between the sphere $x^2+y^2+z^2=9$ and the cylinder $x^2+y^2=4$. You will do this in two different ways. View this Math3D demo for a better visual.



First, proceed by the *projection method*. Projecting S into the (x,y)-plane, you get the region bounded between two circles, namely

$$T=\{(x,y)\in\mathbb{R}^2 : 4\leq x^2+y^2\leq 9\}.$$

For any fixed $(x,y)\in T$, the point $(x,y,z)\in\mathbb{R}^3$ lies inside S if and only if (x,y,z) lies between the upper hemisphere $z=\sqrt{9-x^2-y^2}$ and the lower hemisphere $z=-\sqrt{9-x^2-y^2}$. This Math3D demo illustrates this phenomenon.



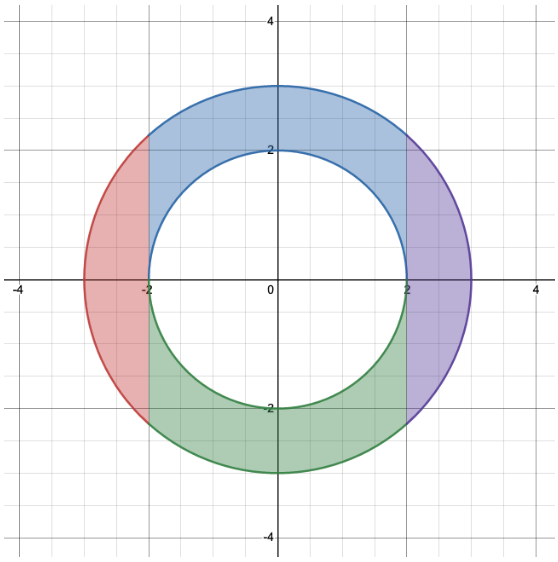
In terms of integrals, this implies that

$$\operatorname{vol}(S)=\iint_{4\leq x^2+y^2\leq 9}\left(\int_{-\sqrt{9-x^2-y^2}}^{\sqrt{9-x^2-y^2}} dz\right)dA \qquad (9.5.1)$$

It only remains to integrate over $T\subseteq\mathbb{R}^2$. This falls under the usual strategies for double integrals. In this case, you can break up T into four pieces, namely

- $-3 \le x \le -2$ and $-\sqrt{9-x^2} \le y \le \sqrt{9-x^2}$
- $-2 \le x \le 2$ and $-\sqrt{9-x^2} \le y \le -\sqrt{4-x^2}$
- $-2 \le x \le 2$ and $\sqrt{4-x^2} \le y \le \sqrt{9-x^2}$
- $2 \le x \le 3$ and $-\sqrt{9-x^2} \le y \le \sqrt{9-x^2}$

This can be observed in the Desmos graph below.



Combining all of these cases with (9.5.1), it follows that

$$\begin{aligned}\text{vol}(S) = &\int_2^3 \int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}} \int_{-\sqrt{9-y^2-x^2}}^{\sqrt{9-y^2-x^2}} 1 dz dy dx + \int_{-3}^{-2} \int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}} \int_{-\sqrt{9-y^2-x^2}}^{\sqrt{9-y^2-x^2}} 1 dz dy dx \\ &+ \int_{-2}^{2} \int_{\sqrt{4-x^2}}^{\sqrt{9-x^2}} \int_{-\sqrt{9-y^2-x^2}}^{\sqrt{9-y^2-x^2}} 1 dz dy dx + \int_{-2}^{2} \int_{-\sqrt{9-x^2}}^{-\sqrt{4-x^2}} \int_{-\sqrt{9-y^2-x^2}}^{\sqrt{9-y^2-x^2}} 1 dz dy dx.\end{aligned}$$

Second, you can arrive at an equivalent expression by the *slicing method*. It seems that z-slices of S will be the most natural to describe. The cylinder $x^2+y^2=4$ and sphere $x^2+y^2+z^2=9$ intersect precisely when $z^2=5 \iff z=\pm\sqrt{5}$. From the above diagrams, it follows that the solid S lies between $-\sqrt{5} \le z \le \sqrt{5}$.

For $-\sqrt{5} \le z \le 5$, the z-slice of S is given by

$$\begin{aligned}S_z &= \{(x,y) \in \mathbb{R}^2 : x^2+y^2+z^2 \le 9, x^2+y^2 \ge 4\} \\ &= \{(x,y) \in \mathbb{R}^2 : 4 \le x^2+y^2 \le 9-z^2\}.\end{aligned}$$

The volume of S is therefore given by

$$\text{vol}(S) = \int_{-\sqrt{5}}^{\sqrt{5}} \text{area}(S_z) dz.$$

To calculate area(S_z) for each fixed $-\sqrt{5} \le z \le \sqrt{5}$, you can follow the same strategy as above by breaking up this washer region into four pieces, such as

- $-\sqrt{9-z^2} \le x \le -2$ and $-\sqrt{9-x^2-z^2} \le y \le \sqrt{9-x^2-z^2}$
- $-2 \le x \le 2$ and $-\sqrt{9-x^2-z^2} \le y \le -\sqrt{4-x^2}$

- $-2 \leq x \leq 2$ and $\sqrt{4-x^2} \leq y \leq \sqrt{9-x^2-z^2}$
- $2 \leq x \leq \sqrt{9-z^2}$ and $-\sqrt{9-x^2-z^2} \leq y \leq \sqrt{9-x^2-z^2}$

After doing so, it follows that

$$\begin{aligned}\mathrm{vol}(S) = &\int_{-\sqrt{5}}^{\sqrt{5}} \int_{-\sqrt{9-z^2}}^{-2} \int_{-\sqrt{9-z^2-x^2}}^{\sqrt{9-z^2-x^2}} 1\,dy\,dx\,dz + \int_{-\sqrt{5}}^{\sqrt{5}} \int_{-2}^{2} \int_{-\sqrt{9-x^2-z^2}}^{-\sqrt{4-x^2}} 1\,dy\,dx\,dz \\ &+ \int_{-\sqrt{5}}^{\sqrt{5}} \int_{-2}^{2} \int_{\sqrt{4-x^2}}^{\sqrt{9-x^2-z^2}} 1\,dy\,dx\,dz + \int_{-\sqrt{5}}^{\sqrt{5}} \int_{2}^{\sqrt{9-z^2}} \int_{-\sqrt{9-x^2-z^2}}^{\sqrt{9-x^2-z^2}} 1\,dy\,dx\,dz.\end{aligned}$$

There are more methods by mixing these ideas with geometry. For instance, to calculate the volume of S, you can take the volume of the ball B and then *subtract* the volume of the region R bounded within the cylinder and sphere. That is quite clever! The details of this approach will be left as an exercise, but the outline will be sketched here. For the ball B, you can use classic geometry formulas but it is good to practice the projection method and find that

$$\mathrm{vol}(B) = \int_{-3}^{3} \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} \int_{-\sqrt{9-x^2-y^2}}^{\sqrt{9-x^2-y^2}} 1\,dz\,dx\,dy.$$

For the region R between the cylinder and the sphere caps, you can again use projections and that find that

$$\mathrm{vol}(R) = \int_{-2}^{2} \int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} \int_{-\sqrt{9-x^2-y^2}}^{\sqrt{9-x^2-y^2}} 1\,dz\,dx\,dy.$$

Thus, you may conclude that

$$\mathrm{vol}(S) = \int_{-3}^{3} \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} \int_{-\sqrt{9-x^2-y^2}}^{\sqrt{9-x^2-y^2}} 1\,dz\,dx\,dy - \int_{-2}^{2} \int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} \int_{-\sqrt{9-x^2-y^2}}^{\sqrt{9-x^2-y^2}} 1\,dz\,dx\,dy.$$

This gives three equivalent but completely different-looking expressions for the volume of S. There are certainly many more!

As you can see from these examples, the geometry of $\mathbb{R}^3$ can be much more complicated and make it a more difficult compute integrals over solids. There are many equivalent approaches and general techniques, but there is no systematic method. It is also not apparent what approach may be tractable or reasonable. You will need to sketch many pictures, be flexible, and get a lot of practice. These are classic traits of a mathematician and you are encouraged to continue this training mindset. In the next section, you will learn how to integrate in other three-dimensional coordinate systems, specifically cylindrical and spherical coordinates. This powerful technique will alleviate some of the geometric challenges for integration in $\mathbb{R}^3$, especially for solids with radial or spherical symmetries.

Exercises for Section 9.5

Concepts and definitions

9.5.1 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be continuous. Which iterated triple integrals are valid expressions?

(a) $\displaystyle\int_0^1 \int_0^2 \int_0^3 f(x,y,z)dxdydz$

(b) $\displaystyle\int_0^x \int_0^{2y} \int_0^{3z} f(x,y,z)dxdydz$

(c) $\displaystyle\int_0^{3z} \int_0^{2y} \int_0^{x} f(x,y,z)dxdydz$

(d) $\displaystyle\int_0^1 \int_0^x \int_0^{3z} f(x,y,z)dxdydz$

(e) $\displaystyle\int_0^1 \int_0^{2y} \int_0^{3z} f(x,y,z)dxdydz$

(f) $\displaystyle\int_0^1 \int_0^2 \int_{y+z}^{3z} f(x,y,z)dxdydz$

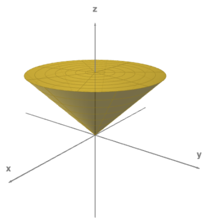
(g) $\displaystyle\int_0^1 \int_0^2 \int_{y+z}^{3z} f(x,y,z)dxdydz$

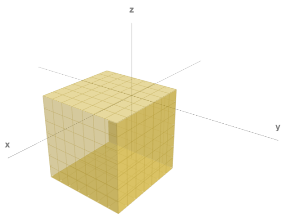
(h) $\displaystyle\int_0^1 \int_0^x \int_{y+z}^{3z} f(x,y,z)dxdydz$

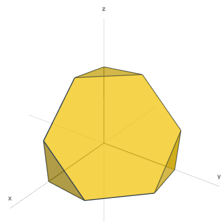
(i) $\displaystyle\int_0^1 \int_0^{2y} \int_{y+z}^{3z} f(x,y,z)dxdydz$

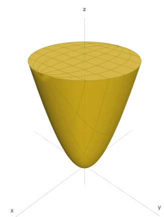
(j) $\displaystyle\int_0^1 \int_0^{3z} \int_{y+z}^{3z} f(x,y,z)dxdydz$

9.5.2 Choose the corresponding picture (A)–(D) of the solid described by each of the given inequalities. Then set up a triple integral to compute the volume. **Do not compute the integrals.**


(A)


(B)


(C)


(D)

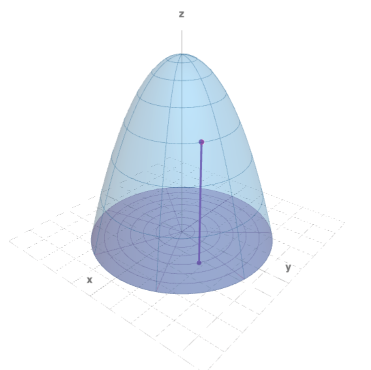
(a) $x+y+z \le 3,\ 0 \le x \le 2,\ 0 \le y \le 2,\ 0 \le z \le 2$

(b) $x^2+y^2 \le z \le 4$

(c) $4(x^2+y^2) \le z^2 \le 4,\ z \ge 0$

(d) $1 \le x \le 3,\ 0 \le y \le 2,\ -1 \le z \le 1$

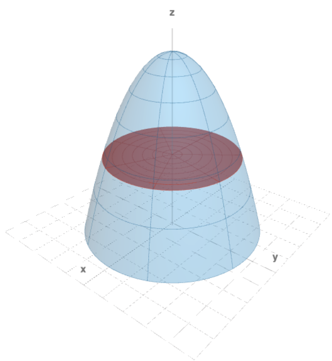
9.5.3 Let S be the solid below the paraboloid $z = 9 - x^2 - y^2$ and above the plane $z = 0$ with density $f(x,y,z)$. You will express its mass in three different ways. View this Math3D demo.



(a) Define the projection of S in the (x,y)-plane as a subset P of $\mathbb{R}^2$.

(b) Fix $(x,y) \in \mathbb{R}^2$ in the projection. Express the "1D mass density" of the vertical wire inside the solid lying above (x,y) as a single integral dz.

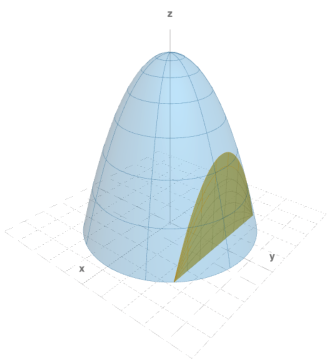
(c) Express the mass of S as a triple integral by totalling these wires.



(d) Sketch a typical z-slice of S in $\mathbb{R}^2$, and define it with set builder notation. For which values of z will this slice be non-empty?

(e) Express the "2D mass density" of a z-slice of S as a double integral $dydx$.

(f) Express the mass of S as a triple integral by totalling these z-slices.



(g) Sketch a typical y-slice of S in $\mathbb{R}^2$, and define it with set builder notation. For which values of y will this slice be non-empty?

(h) Express the "2D mass density" of a y-slice of S as a double integral $dzdx$.

(i) Express the mass of S as a triple integral by totalling these y-slices.

Computations

9.5.4

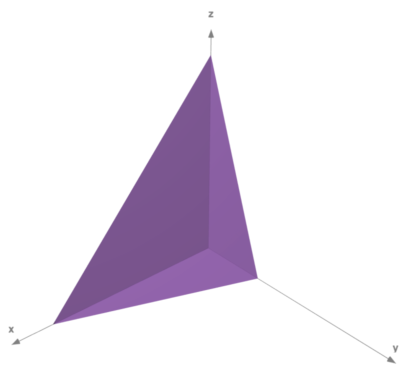
Let T be the tetrahedron with vertices at $(0,0,0), (0,0,10), (0,2,0)$ and $(5,0,0)$. Let $g(x,y,z)$ be a continuous function on T.
Express

$$I = \int\int\int_T g(x,y,z)dV$$

in two different ways.

(a) $dzdydx$

(b) $dxdzdy$



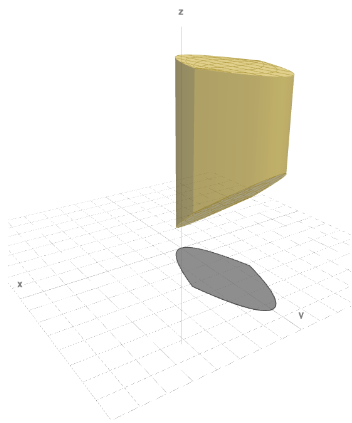
9.5.5 Let S be the solid between the parabolic cylinders $y = 2x^2$ and $y = 4 - 2x^2$ that is bounded below by $z = y + 1$ and above by $z = 8$.
For a function $f(x, y, z)$, you will express

$$I = \iiint_S f(x, y, z)dV$$

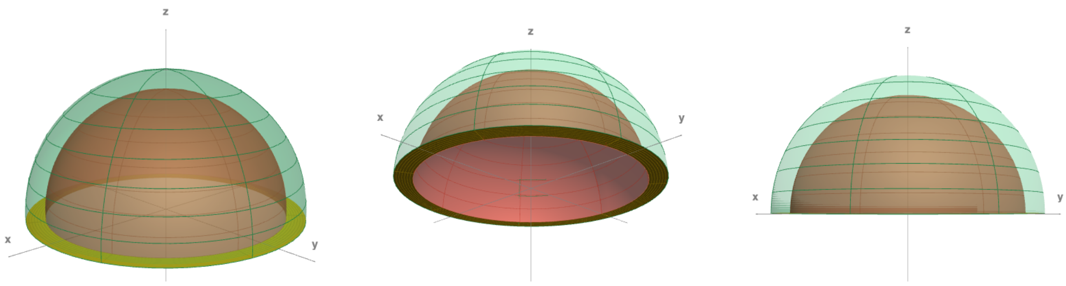
using iterated triple integral(s) in two ways.

(a) Express I by projecting the solid S. Remember to sketch and label your projection.

(b) Express I by using z-slices of S. Remember to sketch typical slices.



9.5.6 A carved melon M is described by the region $3 \leq x^2 + y^2 + z^2 \leq 4$ lying above the $z = 0$ plane. This carved melon has variable density $\delta(x, y, z) = x^2 + y^2 + z^2$.



(a) Find the mass of M using iterated triple integral(s) by totaling z-slices.
Hint: Sketch typical slices. More than one will be necessary.

(b) Find the mass of M using iterated triple integral(s) by projecting into the (x, y)-plane. *Hint:* Break up the projected region in $\mathbb{R}^2$ into 5 pieces.

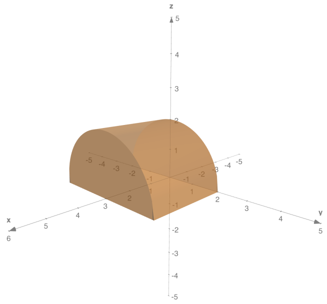
9.5.7 Let S be the solid bounded by the cylinder $x^2 + y^2 = 4$, the plane $y + z = 2$ and the horizontal plane $z = 0$. View this Math3D demo of S. Set up a triple integral or sum of triple integrals representing the volume of S in the following three ways. Include sketches of slices or projections to justify your reasoning. Evaluate the integrals using computer algebra software.

(a) with respect to $dz\, dx\, dy$.

(b) with respect to $dx\, dy\, dz$.

(c) with respect to $dy\, dx\, dz$.

9.5.8 Let S be bounded by $y^2 + z^2 = 4$, $z = 0$, $x = 0$, and $x = 3$, as shown below. Let $f(x, y, z) = 5z^3$ and evaluate the integral of f over S using an iterated integral with order $dzdydx$.



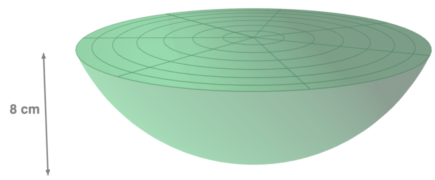
Applications and beyond

9.5.9 Express the volume of the solid cone S bounded by $z = 5$ as an integral with respect to $dzdydx$ and also as an integral with respect to $dxdydz$

9.5.10 Express the volume of the solid S bounded between $2x + y + z = 5$ and $3x + 5y + z = 5$ and above the region $x \leq 2 - y, x \geq 0, y \geq 0$ in the xy-plane as a triple integral.

9.5.11 Express the volume of a pyramid with a square base $[-3, 3] \times [-3, 3]$ in the xy-plane and its peak at $(0, 0, 3)$ as a triple integral and evaluate it.

9.5.12 Let S be a portion of a spherical ball with radius 13 cm after being cut to a height of 8 cm. Express the volume of S using a triple integral.



9.5.13 Let $S = \{(x, y, z) \colon x^2 + y^2 \leq 16, 0 \leq z \leq 1, x \geq 0, y \geq 0\}$ be a solid with mass density equal to the distance from the xy-plane squared. Using computer algebra software as necessary, find the centre of mass of S.

9.6. Triple integrals in cylindrical coordinates

Iterated triple integrals (e.g. $dxdydz$) are best suited for calculating integrals of regions in $\mathbb{R}^3$ which are rectangular (e.g. rectangles, regions under surfaces $z = f(x,y)$), but not all shapes are naturally described using rectangular coordinates. As you did in $\mathbb{R}^2$ with polar coordinates, you can consider other coordinate systems. There are two other classical coordinate systems in $\mathbb{R}^3$, namely cylindrical coordinates and spherical coordinates; see Section 1.4 for an introduction. Over the next two sections, you will investigate both coordinate systems for integration. These discussions will strongly parallel the relationship between rectangular and polar coordinates in $\mathbb{R}^2$, but the geometry and visuals are naturally more complicated in $\mathbb{R}^3$.

Here you will focus on *cylindrical coordinates*; see Section 1.4.2 for an introduction. For the entirety of this section, the map $g : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$g(r,\theta,z) = (r\cos\theta, r\sin\theta, z) \tag{9.6.1}$$

is the cylindrical coordinate transformation. Recall that its domain is referred to as (r,θ,z)-space and its codomain is referred to as (x,y,z)-space. Objects that are radially symmetric about the z-axis are often best described in cylindrical coordinate systems.

Can you integrate over regions in (x,y,z)-space by integrating in (r,θ,z)-space?

Your explorations will create another new integration technique: integration in cylindrical coordinates. Again, the explanations and calculations will be not be fully rigorous but you will continue to build evidence for a future spectacular theorem on integration with different coordinate systems.

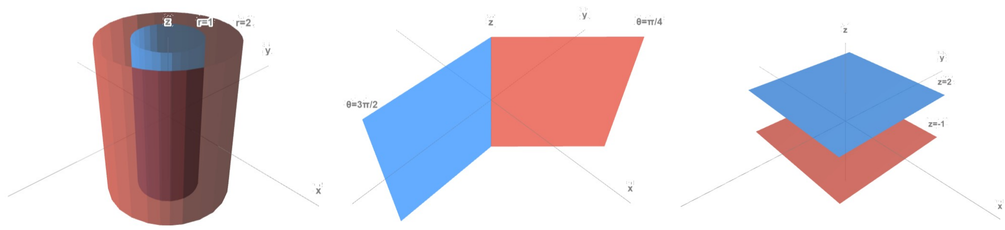
9.6.1 Regions in cylindrical coordinates

Recall the standard relationship

$$(x,y,z) = g(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$$

for any $r,\theta,z \in \mathbb{R}$. See Section 1.4.1 for details on the distinction between positive and negative values of r. You will almost always construct integrals where r is non-negative, in which case you do not need to worry about this distinction.

Example 9.6.1 Slices in cylindrical coordinates will dictate how the geometry of integration will operate. The diagram below shows the *image* of r-slices, θ-slices, and z-slices from (r,θ,z)-space to (x,y,z)-space under the cylindrical coordinate transformation. For the sake of brevity, you will simply refer to them as r-slices, θ-slices, and z-slices.



Note the displayed θ-slices implicitly assume that $r \geq 0$.

Using these slices, you can describe cylindrical regions. Again, when studying the cylindrical coordinate transformation (9.6.1) in Section 1.4.2, you carefully distinguished between regions described in its domain, (r,θ,z)-space, and its codomain, (x,y,z)-space. For instance, the set

$$A = \{(r,\theta,z) \in \mathbb{R}^3 : 0 \leq r \leq 1, 0 \leq \theta \leq 2\pi, 0 \leq z \leq 1\}$$

is *not* the same as the set

$$B = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, 0 \leq z \leq 1\}.$$

Formally, B is the image of A under g, that is $B = g(A)$. You may more informally say

$0 \leq r \leq 1, 0 \leq \theta \leq 2\pi, 0 \leq z \leq 1$ *is the solid cylinder* $x^2 + y^2 \leq 1, 0 \leq z \leq 1$ *in cylindrical form*.

Remember some texts always assume that $r \geq 0$ but you shall not do so here; if you want to require $r \geq 0$, then you should specify it.

Example 9.6.2 The solid ball $x^2 + y^2 + z^2 \leq 9$ of radius 3 can be expressed in cylindrical coordinates since $r^2 = x^2 + y^2$. Thus, $r^2 + z^2 \leq 9, 0 \leq \theta \leq 2\pi$ is the solid ball of radius 3 in cylindrical form. Notice you must specify $0 \leq \theta \leq 2\pi$.

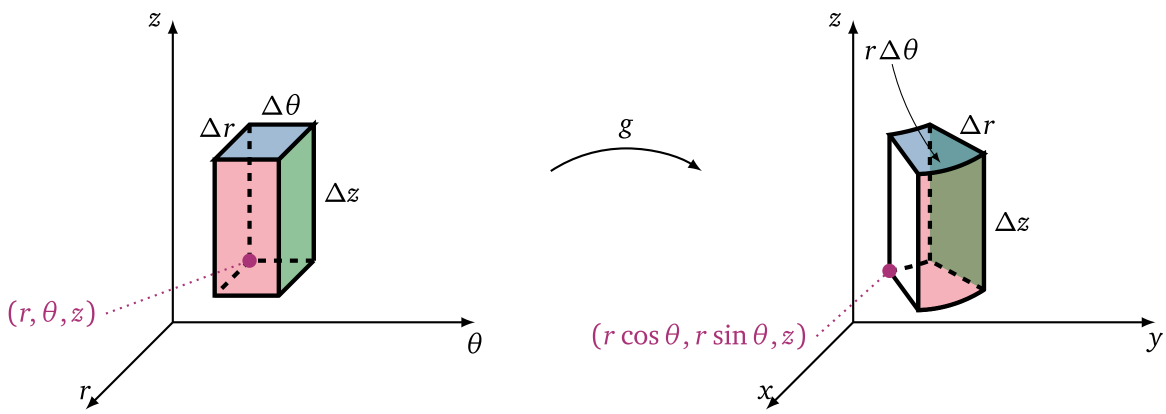
9.6.2 Derivation of integrals in cylindrical coordinates

The derivation of integration with cylindrical coordinates in $\mathbb{R}^3$ closely mimics the ideas for polar coordinates in $\mathbb{R}^2$ found in Section 9.3.2, so only the highlights will be discussed. The details will be left as an exercise.

For a fixed $r > 0, \theta \in \mathbb{R}$, and $z \in \mathbb{R}$, the small rectangle

$$R = [r, r + \Delta r] \times [\theta, \theta + \Delta\theta] \times [z, z + \Delta z]$$

in (r, θ, z)-space transforms under g to a cylindrical rectangle $g(R)$ in (x, y, z)-space. This transformation is illustrated below.



You can argue that

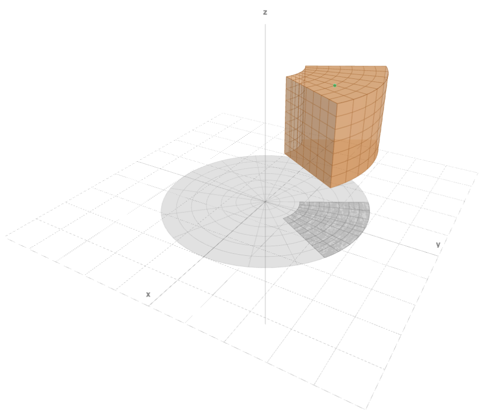
$$\operatorname{vol}(g(R)) \approx r \Delta r \Delta\theta \Delta z = r \operatorname{vol}(R). \tag{9.6.2}$$

Again, the volume is scaled by the radius $r = |r|$ as $r > 0$. If $r < 0$, then the volume would be scaled by $-r = |r|$. Equipped with (9.6.2), you can address the main problem.

Without loss of generality, assume Ω is a rectangle, that is,

$$\Omega = \{(r, \theta, z) : a \leq r \leq b, \alpha \leq \theta \leq \beta, c \leq z \leq d\}.$$

The transformed region $D = g(\Omega)$ in (x, y, z)-space is illustrated below.



Using this visual as a guide and Section 9.3.2 as a template argument, you can derive for the region $\Omega = [a, b] \times [\alpha, \beta] \times [c, d]$ the following relationship

$$\iiint_{g(\Omega)} f\,dV = \int_c^d \int_\alpha^\beta \int_a^b (f \circ g)(r, \theta, z)|r|\,dr\,d\theta\,dz \tag{9.6.3}$$

provided $f : g(\Omega) \to \mathbb{R}$ is integrable.

This allows you to formulate a conjecture.

Theorem 9.6.3 Let $\Omega \subseteq \mathbb{R}^3$ be a Jordan measurable set such that the restricted cylindrical coordinate transformation $g|_\Omega : \Omega \to g(\Omega)$ given by (9.6.1) is a bijection. If the real-valued function $f : g(\Omega) \to \mathbb{R}$ is integrable on $g(\Omega)$, then the real-valued function $F : \Omega \to \mathbb{R}$ given by $F(r, \theta, z) = (f \circ g)(r, \theta, z) \cdot |r|$ is integrable on Ω and

$$\iiint_{g(\Omega)} f\,dV = \iiint_\Omega F\,dV.$$

Proof. Postponed. You will later construct a more general theorem. ■

Remark 9.6.4 In some cases, you may want to apply this theorem to a region Ω where $g|_\Omega : \Omega \to g(\Omega)$ is nearly a bijection aside from a set of zero Jordan measure. The issue usually occurs on the boundary $\partial\Omega$ in which case you can instead apply Theorem 9.6.3 to the interior of Ω and use the ideas in Example 9.3.5 to "add back" the boundary.

Theorem 9.6.3 can be remembered using the volume element dV, namely you have that

$$dx\,dy\,dz = dV = |r|\,dr\,d\theta\,dz.$$

The "identity" between the symbols above has no formal meaning, but you can interpret this as:

> *Infinitesimal volume dV is calculated as the volume of an infinitesimal rectangle $dx\,dy\,dz$ or as the volume of an infinitesimal cylindrical rectangle $|r|\,dr\,d\theta\,dz$.*

This notion will return again later.

9.6.3 Examples of integrals in cylindrical coordinates

Cylindrical coordinates are perfect for integrating over regions with radial symmetry about the z-axis. The method of slices, especially with z-slices, is a common approach.

Example 9.6.5 You can calculate the volume of the ball S of radius $R > 0$ in $\mathbb{R}^3$ using cylindrical coordinates. Recall that $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 \le R^2\}$. As discussed in Example 9.6.2, the region S can be described as

$$r^2+z^2 \le R^2, \quad 0 \le \theta \le 2\pi$$

in cylindrical form. Thus, a z-slice of S is described in polar form as

$$0 \le r \le \sqrt{R^2-z^2}, \quad 0 \le \theta \le 2\pi.$$

As the ball is radius R, the z-slices of S occur for $-R \le z \le R$. View this Math3D demo for a visual. Thus, you may express the volume of S in cylindrical coordinates as

$$\begin{aligned}\mathrm{vol}(S) = \iiint_S 1 dV = \int_{-R}^{R}\int_0^{2\pi}\int_0^{\sqrt{R^2-z^2}} r dr d\theta dz &= \int_{-R}^{R}\int_0^{2\pi} \frac{1}{2}r^2\Big|_{r=0}^{r=\sqrt{R^2-z^2}} d\theta dz \\ &= \frac{1}{2}\int_{-R}^{R}\int_0^{2\pi} R^2-z^2 d\theta dz \\ &= \pi\int_{-R}^{R} R^2-z^2 dz \\ &= \pi R^2 z - \frac{\pi}{3}z^3\Big|_{z=-R}^{z=R} = \frac{4}{3}\pi R^3.\end{aligned}$$

You can also calculate the volume of the ball of radius R in cylindrical coordinates by expressing it as a triple integral $drdzd\theta$ or $d\theta dzdr$. These are left as an exercise.

If the solid has nice symmetries, it can be much easier to express integrals in cylindrical coordinates compared to rectangular coordinates.

Example 9.6.6 You can revisit Example 9.5.4 in cylindrical coordinates. Recall S is the solid between the sphere $x^2+y^2+z^2 = 9$ and the cylinder $x^2+y^2 = 4$. As you verified earlier, the z-slices of S occur for $-\sqrt{5} \le z \le \sqrt{5}$ and a z-slice $S_z \subseteq \mathbb{R}^2$ is given by the washer

$$S_z = \{(x,y) \in \mathbb{R}^2 : 4 \le x^2+y^2 \le 9-z^2\}.$$

This z-slice can be expressed as $4 \le r^2 \le 9-z^2$ and $0 \le \theta \le 2\pi$. Thus, S is written as

$$-\sqrt{5} \le z \le \sqrt{5}, 2 \le r \le \sqrt{9-z^2}, 0 \le \theta \le 2\pi.$$

Thus, the volume of S is given by

$$\mathrm{vol}(S) = \int_{-\sqrt{5}}^{\sqrt{5}}\int_2^{\sqrt{9-z^2}}\int_0^{2\pi} r d\theta dr dz.$$

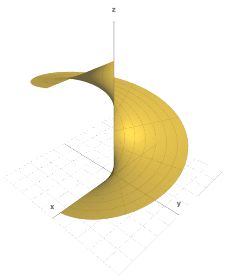
Compared to Example 9.5.4, that was dramatically simpler!

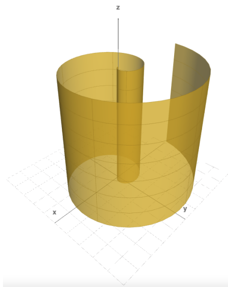
For objects with rotational symmetry, cylindrical coordinates provides another avenue for integration in $\mathbb{R}^3$. You shall next explore how to integrate with spherical coordinates in $\mathbb{R}^3$.

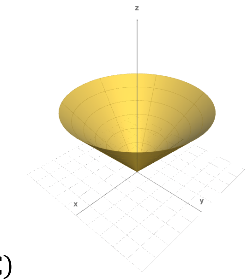
Exercises for Section 9.6

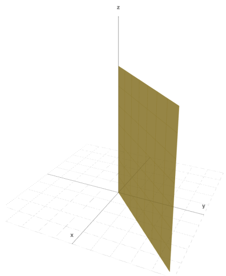
Concepts and definitions

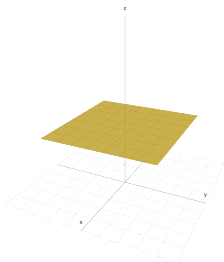
9.6.1 Match the surfaces described in cylindrical coordinates with the figures.

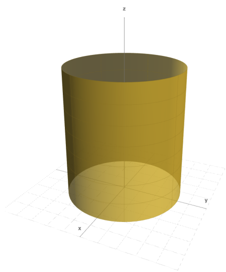

(A)


(B)


(C)


(D)


(E)


(F)

(a) $r = 3$

(b) $\theta = \frac{\pi}{4}$, $r \geq 0$

(c) $z = 3$

(d) $r = \theta$, $\theta \geq 0$

(e) $r = z$, $z \geq 0$

(f) $\theta = z$, $z \geq 0$, $r \geq 0$

9.6.2 Each set of equations in cylindrical coordinates below corresponds to one of the shapes:

(A) Cylinder
(B) Half cylinder
(C) Solid between two cylinders
(D) Cylinder hollowed out by cone
(E) Cone
(F) Cone with flat bottom and flat top
(G) Cone hollowed by cylinder
(H) Half cone

Fix $0 < a < A$ and $0 < h < H$. Identify which shape corresponds to which set of equations.

(a) $0 \leq r \leq A$
$0 \leq \theta \leq 2\pi$
$0 \leq z \leq H$

(b) $0 \leq r \leq z$
$0 \leq \theta \leq \pi$
$0 \leq z \leq H$

(c) $a \leq r \leq A$
$0 \leq \theta \leq 2\pi$
$0 \leq z \leq H$

(d) $a \leq r \leq z$
$0 \leq \theta \leq 2\pi$
$0 \leq z \leq H$

(e) $0 \leq r \leq A$
$0 \leq \theta \leq \pi$
$0 \leq z \leq H$

(f) $0 \leq r \leq z$
$0 \leq \theta \leq 2\pi$
$h \leq z \leq H$

9.6.3 Let $f : \mathbb{R}^3 \to \mathbb{R}$ be continuous. Which iterated triple integrals are valid expressions?

(a) $\displaystyle\int_0^1\int_0^2\int_0^3 f(r,\theta,z)drd\theta dz$

(b) $\displaystyle\int_0^1\int_0^2\int_0^3 f(r\cos\theta,r\sin\theta,z)drd\theta dz$

(c) $\displaystyle\int_0^1\int_0^2\int_0^3 f(x,y,z)drd\theta dz$

(d) $\displaystyle\int_0^1\int_0^2\int_0^3 f(r\cos\theta,r\sin\theta,z)dxdydz$

(e) $\displaystyle\int_0^1\int_0^{4\pi}\int_{-3}^3 f(r,\theta,z)drd\theta dz$

(f) $\displaystyle\int_0^{z\sin\theta}\int_0^{2z}\int_0^3 f(r,\theta,z)drd\theta dz$

(g) $\displaystyle\int_0^3\int_0^{2z}\int_0^{z\sin\theta} f(r,\theta,z)drd\theta dz$

(h) $\displaystyle\int_0^3\int_0^{2r}\int_0^{r\sin\theta} f(r,\theta,z)drd\theta dz$

9.6.4

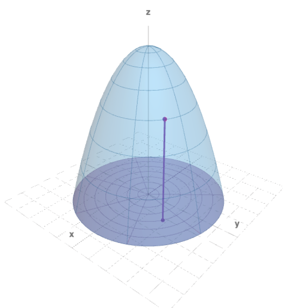
The *projection method* and *slicing method* in cylindrical coordinates follows the same strategy as before, except now the geometry looks a bit different. Let S be the solid below the paraboloid $z=9-x^2-y^2$ and above the plane $z=0$ with density $f(x,y,z)$. You will express its mass in four different ways in cylindrical coordinate; see this Math3D demo.

First you will explore the projection method. Second, you will use the slicing method in three different ways by fixing 1 variable in cylindrical coordinates.

(a) Let $P\subseteq\mathbb{R}^2$ be the projection of S in the (x,y)-plane. Express P in (r,θ)-coordinates.

(b) Fix $(r,\theta)\in\mathbb{R}^2$ in the projection. Express the "1D mass density" of the vertical wire inside the solid lying above $(r\cos\theta, r\sin\theta)$ as a single integral dz.

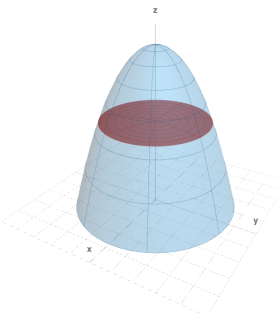
(c) Express the mass of S as a triple integral by totaling these wires.



(d) Express a typical z-slice of S using (r,θ)-coordinates. For which values of z will this slice be non-empty?

(e) Express the "2D mass density" of a z-slice of S as a double integral $drd\theta$.

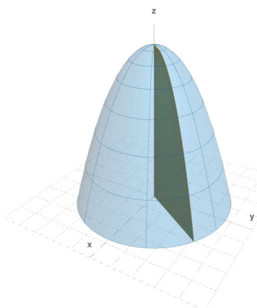
(f) Express the mass of S as a triple integral in cylindrical coordinates by totaling these z-slices.



(g) Express a typical θ-slice of S with $r\geq 0$ using (r,z)-coordinates. For which values of θ will this slice be non-empty?

(h) Express the "2D mass density" of a θ-slice with $r\geq 0$ as a double integral $drdz$.

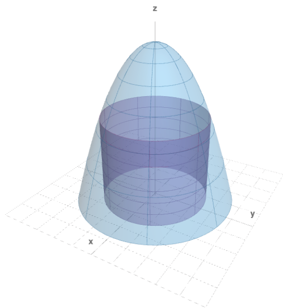
(i) Express the mass of S as a triple integral by totaling these θ-slices.



(j) Express a typical r-slice of S with $r \geq 0$ using (θ, z)-coordinates. For which values of r will this slice be non-empty?

(k) Express the "2D mass density" of a r-slice of S as a double integral $d\theta dz$.

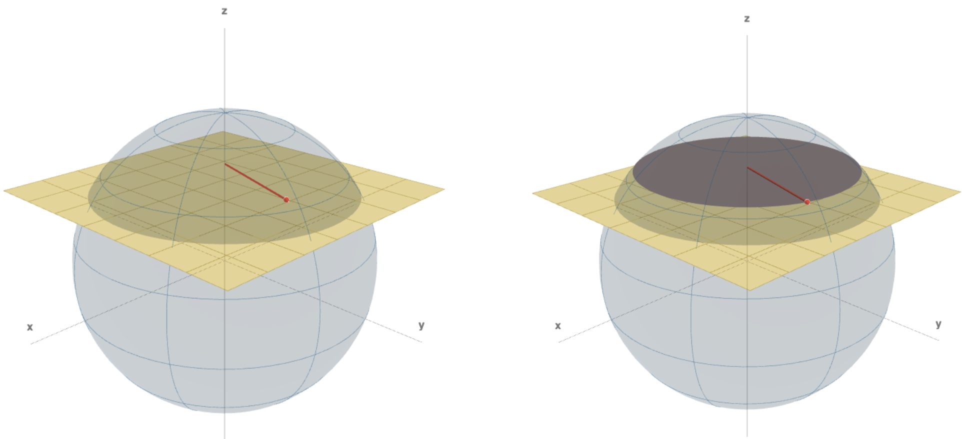
(l) Express the mass of S as a triple integral by totaling these r-slices.



9.6.5 The geometry of the slicing method in cylindrical coordinates can be better understood by breaking down the steps even further. Here you will look at slices with 2 fixed variables, then 1 fixed variable, and then 0 fixed variables. Let S be the solid inside the sphere $x^2 + y^2 + z^2 = 4$ and above the horizontal plane $z = 1$ with mass density $f(x, y, z)$. You will express the mass of S in three different ways.

(a) Express S in rectangular form and in cylindrical form.

For (b),(c),(d), play with this Math3D demo to see how the slices sweep out the solid.

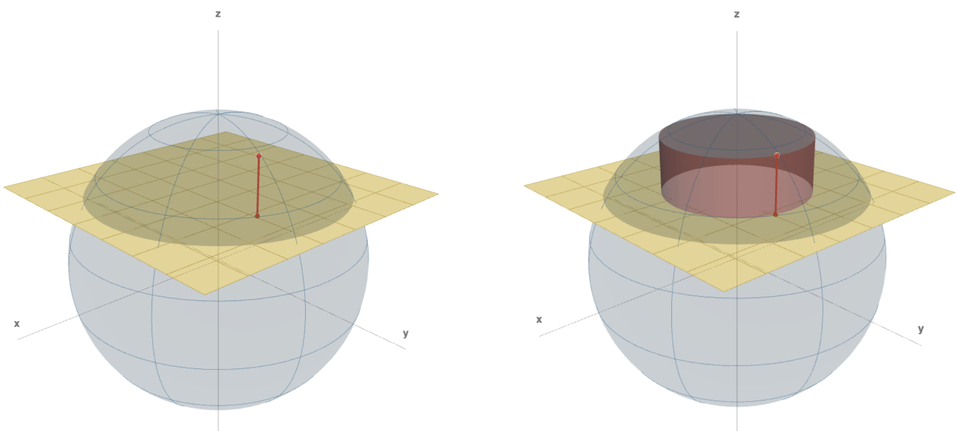


(b) Express the "1D mass density" of a (θ, z)-slice as a single integral dr.

(c) Express the "2D mass density" of a z-slice as a double integral $drd\theta$.

(d) Set up a triple integral in cylindrical coordinates $drd\theta dz$ representing the mass of S.

For (e),(f),(g), play with this Math3D demo to see how the slices sweep out the solid.

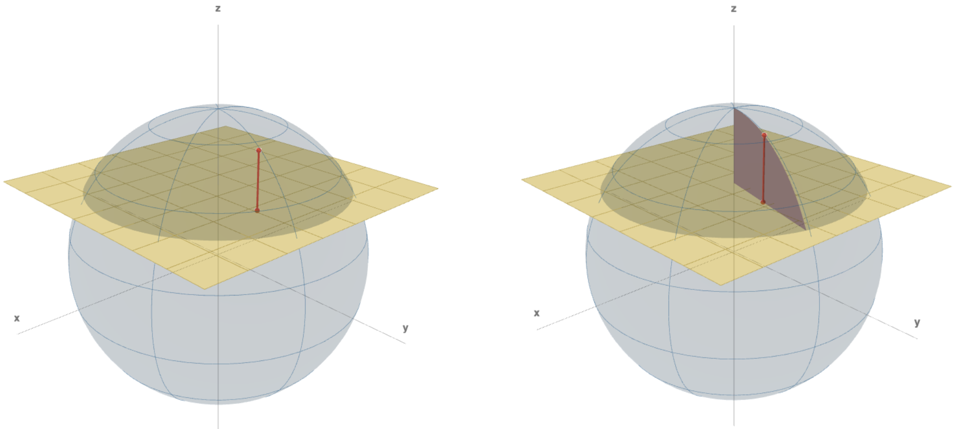


(e) Express the "1D mass density" of a (r, θ)-slice as a single integral dz.

(f) Express the "2D mass density" of a r-slice as a double integral $dzd\theta$.

(g) Set up a triple integral in cylindrical coordinates $dzd\theta dr$ representing the mass of S.

For (h),(i),(j), play with this Math3D demo to see how the slices sweep out the solid.



(h) Express the "1D mass density" of a (r,θ)-slice as a single integral dz.

(i) Express the "2D mass density" of a θ-slice as a double integral $dzdr$.

(j) Set up a triple integral in cylindrical coordinates $dzdrd\theta$ representing the mass of S.

(k) There 3 remaining orders of integration in cylindrical coordinates, and the geometry will look a bit different for each one. Repeat the approach above for each of the remaining 3 orders. Include corresponding sketches of your intermediate steps. This might take quite a while.

Computations

For the following questions, you are encouraged to plot regions on Desmos or Math3D to help you set up your integrals.

9.6.6 Let $B = B_R((0,0,0))$ be the ball in $\mathbb{R}^3$ with radius R centered at the origin.

(a) Express the volume of B in cylindrical coordinates $drd\theta dz$.

(b) Express the volume of B in cylindrical coordinates $drdzd\theta$.

(c) Express the volume of B in cylindrical coordinates $dzd\theta dr$.

(d) Express the volume of B in cylindrical coordinates $dzd\theta dr$.

9.6.7 Using cylindrical coordinates, describe a slice cut at an angle of $\pi/3$ from a cake 8 cm tall and with a radius of 10 cm.

9.6.8 Use cylindrical coordinates to express the following integral.

$$\int_0^2 \int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} \int_{x^2+y^2}^{2\sqrt{x^2+y^2}} yz\,dz\,dx\,dy$$

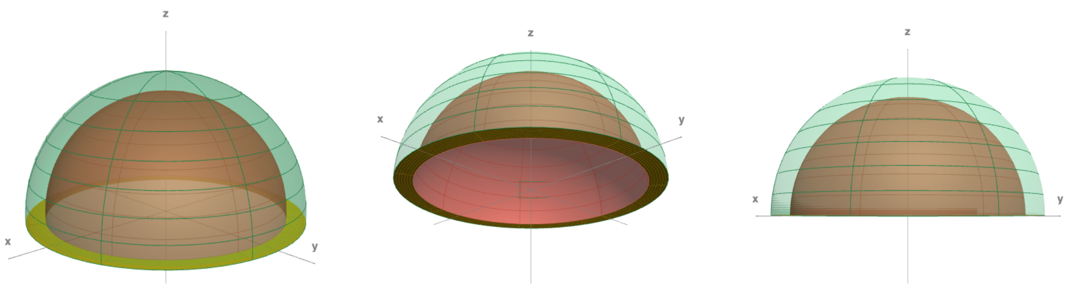
9.6.9 Let S be the solid above the cone $z = \sqrt{x^2+y^2}$ and below the sphere of radius 4 centered at the origin. Write a triple integral for the volume of S using cylindrical coordinates.

9.6.10 Let S be the solid formed between two spheres centered at the origin with radii 3 and 4. Write a triple integral for the volume of S using cylindrical coordinates.

9.6.11 Let B be the solid region inside the sphere of radius 3 centered at the origin and outside the cylinder given by $x^2 + y^2 = 5$. Calculate the volume of B.

Applications and beyond

9.6.12 A carved melon M is described by the region $3 \leq x^2 + y^2 + z^2 \leq 4$ lying above the $z = 0$ plane. This carved melon has variable density $\delta(x, y, z) = x^2 + y^2 + z^2$. You have previously expressed its mass in rectangular coordinates, and now you will do so in cylindrical coordinates.



Once you finish, notice how much simpler this becomes in cylindrical coordinates.

(a) Express the mass of M with iterated triple integral(s) in cylindrical coordinates by totaling z-slices. *Hint:* Sketch typical slices. You will have two different types.

(b) Express the mass of M with iterated triple integral(s) in cylindrical coordinates by projecting M. *Hint:* Break up the projection in $\mathbb{R}^2$ into 2 pieces and describe them in (r, θ)-coordinates.

(c) Calculate the mass of M using either expression.

9.6.13 Here you will heuristically justify a key step to deriving the method for integrating using the cylindrical coordinate transformation $g : \mathbb{R}^3 \to \mathbb{R}^3$ given by $g(r, \theta, z) = (r \cos\theta, r \sin\theta, z)$.

1. *For a fixed $r > 0, \theta \in [0, 2\pi], z \in \mathbb{R}$, the small rectangle*

$$R = [r, r + \Delta r] \times [\theta, \theta + \Delta\theta] \times [z, z + \Delta z]$$

in (r, θ, z)-space transforms under g to a cylindrical rectangle $g(R)$ in (x, y, z)-space.

2. *Since the rectangle is small, the radial thickness of the transformed rectangle is approximately*

$$||g(r + \Delta r, \theta, z) - g(r, \theta, z)|| \approx \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_ = \Delta r.$$

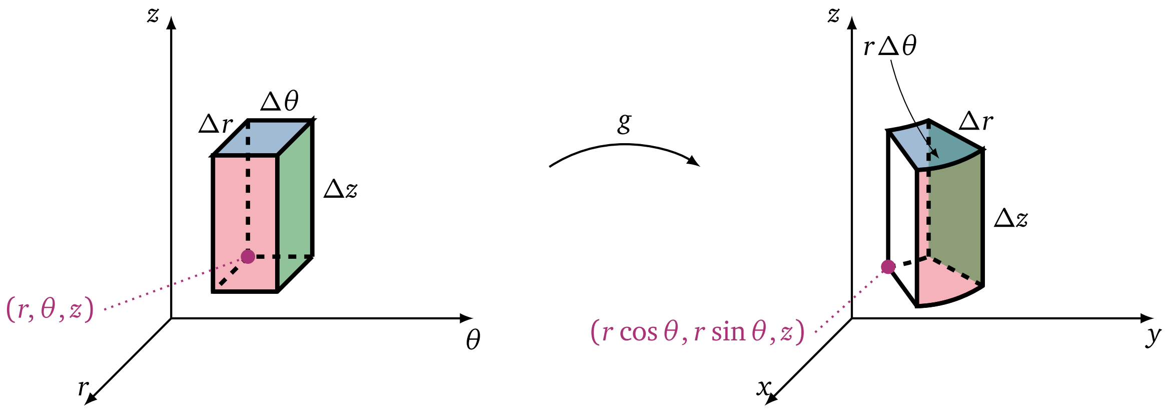
3. *Similarly,*

$$||g(r, \theta + \Delta\theta, z) - g(r, \theta, z)|| \approx \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_ = r\Delta\theta,$$

$$||g(r, \theta, z + \Delta z) - g(r, \theta, z)|| \approx \_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_ = \Delta z.$$

4. *Overall, this implies that* $\mathrm{vol}(g(R)) \approx r\Delta r\Delta\theta\Delta z = r\,\mathrm{vol}(R)$.

(a) Label three points of R and the corresponding three points of $g(R)$ appearing in Lines 2 and 3..



(b) Fill in the 6 blanks in the above argument.

(c) Line 4 follows from Lines 2 and 3 with an assumption about $\frac{\partial g}{\partial r}$, $\frac{\partial g}{\partial \theta}$, and $\frac{\partial g}{\partial z}$. Identify that assumption. *Hint:* What linear algebra property do those 3 vectors satisfy?

9.6.14 Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the cylindrical coordinate transformation. Let $\Omega = [a,b] \times [\alpha,\beta] \times [c,d]$. Assume f is integrable on $g(\Omega)$ and $f \circ g$ is integrable on Ω. Heuristically derive the formula

$$\iiint_{g(\Omega)} f\,dV = \int_c^d \int_\alpha^\beta \int_a^b (f \circ g)(r,\theta,z)|r|\,dr\,d\theta\,dz.$$

Remember you must first explain how the volume of a small rectangle transforms.

9.7. Triple integrals in spherical coordinates

Thus far, you have learned how to integrate in rectangular coordinates (e.g. $dxdydz$) and cylindrical coordinates (e.g. $rdrd\theta dz$). The best choice of coordinate system will be influenced by the shape and symmetries of your region. In this section, you will add another classical coordinate system in $\mathbb{R}^3$ to your factory of techniques, namely *spherical coordinates*. The themes and questions from cylindrical coordinates will repeat, but the geometry will be quite different. This exploration will deepen your spatial and geometric understanding of three dimensions.

For the entirety of this section, the map $g : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$g(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi) \tag{9.7.1}$$

is the spherical coordinate transformation. Recall that its domain is referred to as (ρ,θ,ϕ)-space and its codomain is referred to as (x,y,z)-space; see Section 1.4.3 for an introduction. Objects that possess many rotational symmetries about the origin are often best described in spherical coordinates. The key question arises again.

Can you integrate over regions in (x,y,z)-space by integrating in (ρ,θ,ϕ)-space?

As usual, explanations and calculations will be not be fully rigorous but, by the end, you will be on the cusp of a grand theorem on integration with different coordinate systems.

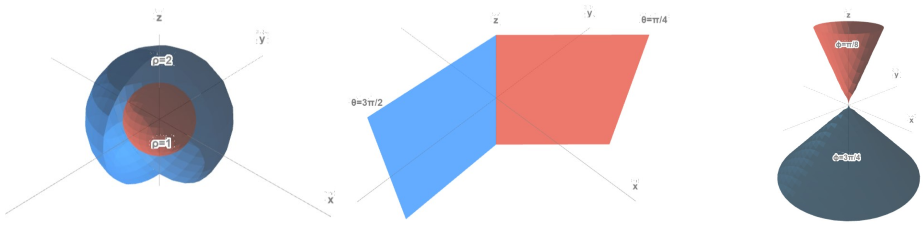
9.7.1 Regions in spherical coordinates

Recall the standard relationship

$$(x,y,z) = g(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$$

for any $\rho,\theta,\phi \in \mathbb{R}$. You will almost always construct integrals where $0 \le \rho < \infty$, $0 \le \theta \le 2\pi$, and $0 \le \phi \le \pi$. Nonetheless, you should always be clear with your choices.

Example 9.7.1 Slices in spherical coordinates decide how the geometry of integration will operate. The diagram below shows the *image* of ρ-slices, θ-slices, and ϕ-slices from (ρ,θ,ϕ)-space to (x,y,z)-space under the spherical coordinate transformation. For the sake of brevity, you will simply refer to them as ρ-slices, θ-slices, and ϕ-slices.



Note the displayed θ-slices and ϕ-slices implicitly assume that $\rho \ge 0$.

Using these slices, you can describe spherical regions. Again, when studying the spherical coordinate transformation (9.7.1) in Section 1.4.3, you carefully distinguished between regions described in its domain, (ρ,θ,ϕ)-space, and its codomain, (x,y,z)-space. For instance, the set

$$A = \{(\rho,\theta,\phi) \in \mathbb{R}^3 : 0 \le \rho \le 1, 0 \le \theta \le 2\pi, 0 \le \phi \le \pi\}$$

is *not* the same as the set

$$B = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \le 1\}.$$

Formally, B is the image of A under g, that is $B = g(A)$. You may more informally say

$$0 \le \rho \le 1, 0 \le \theta \le 2\pi, 0 \le \phi \le \pi \textit{ is the solid ball } x^2 + y^2 + z^2 \le 1 \textit{ in spherical form.}$$

Remember some texts always assume that $\rho \ge 0$ but you shall not do so here; if you want to require $\rho \ge 0$, then you should specify it.

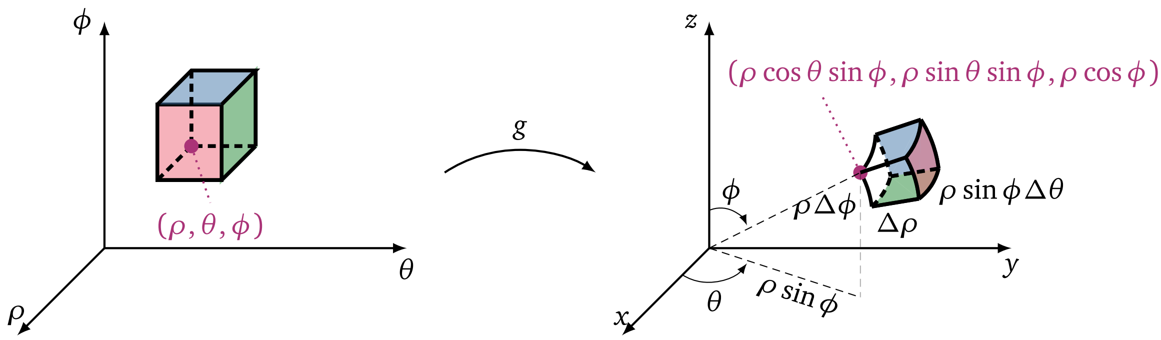
9.7.2 Derivation of integrals in spherical coordinates

The derivation of integration with spherical coordinates in $\mathbb{R}^3$ is nearly identical to the ideas for polar coordinates in $\mathbb{R}^2$ (Section 9.3.2) and cylindrical coordinates in $\mathbb{R}^3$ (Section 9.6.2), so only the highlights will be discussed. The details will be left as an exercise.

For a fixed $\rho > 0, \theta \in \mathbb{R}$, and $\phi \in \mathbb{R}$, the small rectangle

$$R = [\rho, \rho + \Delta\rho] \times [\theta, \theta + \Delta\theta] \times [\phi, \phi + \Delta\phi]$$

in (ρ, θ, ϕ)-space transforms under g to a spherical rectangle $g(R)$ in (x, y, z)-space. This transformation is illustrated below.



As the diagram suggests, you can argue that

$$\operatorname{vol}(g(R)) \approx \rho^2 \sin\phi \Delta\rho\Delta\theta\Delta\phi = \rho^2 \sin\phi \operatorname{vol}(R). \tag{9.7.2}$$

Note the volume is scaled by $\rho^2 \sin\phi = |\rho^2 \sin\phi|$ if $\sin\phi > 0$. If $\sin\phi < 0$, then the volume would be scaled by $-\rho^2 \sin\phi = |\rho^2 \sin\phi|$.

Equipped with (9.7.2) and using Sections 9.3.2 and 9.6.2 as template arguments, you can derive for the region $\Omega = [a, b] \times [\alpha, \beta] \times [\lambda, \mu]$ the following relationship

$$\iiint_{g(\Omega)} f\,dV = \int_\lambda^\mu \int_\alpha^\beta \int_a^b (f \circ g)(\rho, \theta, \phi) |\rho^2 \sin\phi|\, d\rho\, d\theta\, d\phi \tag{9.7.3}$$

provided $f : g(\Omega) \to \mathbb{R}$ is integrable. This allows you to formulate a conjecture.

Theorem 9.7.2 Let $\Omega \subseteq \mathbb{R}^3$ be a Jordan measurable set such that the restricted spherical coordinate transformation $g|_\Omega : \Omega \to g(\Omega)$ given by (9.7.1) is a bijection. If the real-valued function $f : g(\Omega) \to \mathbb{R}$ is integrable on $g(\Omega)$, then the real-valued function $F : \Omega \to \mathbb{R}$ given by $F(\rho, \theta, \phi) = (f \circ g)(\rho, \theta, \phi) \cdot |\rho^2 \sin\phi|$ is integrable on Ω and

$$\iiint_{g(\Omega)} f\,dV = \iiint_\Omega F\,dV.$$

Proof. Postponed. You will soon discover a much more general theorem. ■

Remark 9.7.3 As usual, you may sometimes want to apply this theorem to a region Ω where $g|_\Omega : \Omega \to g(\Omega)$ is nearly a bijection aside from a set of zero Jordan measure. The issue usually occurs on the boundary $\partial\Omega$ in which case you can instead apply Theorem 9.7.2 to the interior of Ω and use the ideas in Example 9.3.5 to "add back" the boundary.

Theorem 9.7.2 is informally realized using the volume element dV, namely you have that

$$dx\,dy\,dz = dV = |\rho^2 \sin\phi|\,d\rho\,d\theta\,d\phi.$$

The "identity" between the symbols above has no formal meaning, but you can interpret this as:

> *Infinitesimal volume dV can be calculated as the volume of an infinitesimal spherical rectangle $|\rho^2 \sin\phi|\,d\rho\,d\theta\,d\phi$.*

Now, it is time to review some basic examples.

9.7.3 Examples of integrals in spherical coordinates

Spherical coordinates are perfect for integrating over regions with rotational symmetry about the origin.

Example 9.7.4 You can swiftly calculate the volume of the ball S of radius $R > 0$ in $\mathbb{R}^3$ using spherical coordinates. Recall that $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq R^2\}$. As $\rho^2 = x^2 + y^2 + z^2$, the region S can be described as

$$0 \leq \rho \leq R, \quad 0 \leq \theta \leq 2\pi, \qquad 0 \leq \phi \leq \pi$$

in spherical form. It is simply a rectangle! Thus, you may express the volume of S in spherical coordinates as

$$\begin{aligned}
\mathrm{vol}(S) = \iiint_S 1\,dV &= \int_0^{\pi} \int_0^{2\pi} \int_0^{R} \rho^2 \sin\phi\,d\rho\,d\theta\,d\phi \\
&= \int_0^{\pi} \int_0^{2\pi} \frac{R^3}{3} \sin\phi\,d\theta\,d\phi \\
&= \int_0^{\pi} \frac{2\pi R^3}{3} \sin\phi\,d\phi \\
&= \frac{4\pi R^3}{3}.
\end{aligned}$$

If the solid has nice symmetries, it can be easier to express integrals in spherical coordinates compared to other coordinate systems.

Example 9.7.5 You can revisit the spherical cap from Example 9.5.3 in spherical coordinates. Recall that S is the solid bounded by the sphere $x^2 + y^2 + z^2 = 4$ and above the $z = 1$ plane. These surfaces can be converted into spherical form as

$$\rho = 2 \qquad \text{and} \qquad \rho\cos\phi = 1.$$

For a fixed polar angle θ and inclination angle ϕ, the radius ρ must be above the $\rho\cos\phi = 1$ plane and below the sphere $\rho = 2$. In other words, it must be that

$$\frac{1}{\cos\phi} \leq \rho \leq 2$$

for any fixed θ, ϕ. Next, notice the spherical cap is rotationally symmetric about the z-axis so the polar angle θ is not restricted and must satisfy

$$0 \leq \theta \leq 2\pi$$

for any fixed ϕ. Finally, the inclination angle $\phi \in [0, \pi]$ is largest where the plane and sphere intersect. In other words, it is required to satisfy

$$\frac{1}{\cos\phi} \leq 2 \implies 0 \leq \phi \leq \frac{\pi}{3}$$

since $\frac{1}{\cos\phi} = 2$ implies that $\phi = \frac{\pi}{3}$. Overall, the volume of the region is given by

$$\mathrm{vol}(S) = \int_0^{2\pi} \int_0^{\frac{\pi}{3}} \int_{1/\cos\phi}^{2} \rho^2 \sin\phi \, d\rho \, d\phi \, d\theta.$$

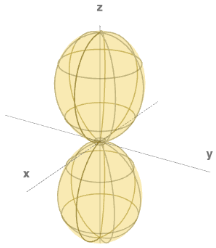
Although the geometry is quite tricky, the final expression is rather simple.

You have studied integration with three different coordinate systems in $\mathbb{R}^3$. Each one has its own advantages and their diversity allows you to handle many different geometric situations. Truthfully, you can only understand and apply them through ample practice. Beyond the calculations, you also studied their origins. After deriving the integral formulas for switching coordinate systems, you may have noticed many patterns and repetitions in the argument. This repetition suggests that you may be able to integrate by switching to *any* coordinate system! In the next section, this optimistic outlook will uncover the most powerful integration technique. It is so powerful that it will shift your geometric and analytic perspective on integration.

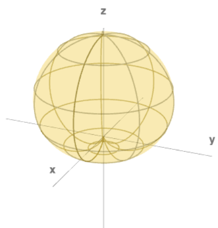
Exercises for Section 9.7

Concepts and definitions

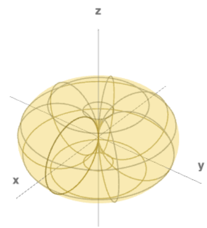
9.7.1 Match the surfaces described in spherical coordinates with the figures.



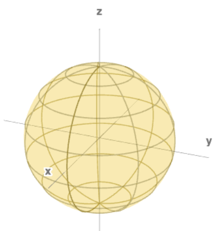
(A)



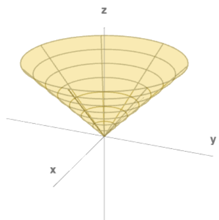
(B)



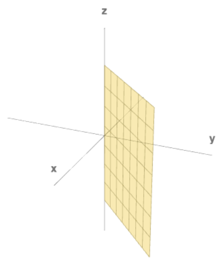
(C)



(D)



(E)



(F)

(a) $\rho = 2$

(b) $\theta = \frac{\pi}{4}$, $\rho \geq 0$

(c) $\phi = \frac{\pi}{4}$, $\rho \geq 0$

(d) $\rho = \sin(\phi)$

(e) $\rho = 1 + \cos(\phi)$

(f) $\rho = 1 + \cos(2\phi)$

9.7.2 Each set of equations in spherical coordinates below corresponds to one of the following shapes:

(A) Solid between two spheres
(B) Upper hemisphere
(C) Lower hemisphere
(D) Right half-sphere
(E) Quarter-sphere
(F) $\frac{1}{8}$-sphere
(G) Cone with flat cap
(H) Cone with spherical cap

Fix $0 < r < R$. Identify which shape corresponds to which set of equations.

(a) $0 \leq \rho \leq R$
$0 \leq \theta \leq 2\pi$
$0 \leq \phi \leq \pi/2$

(b) $0 \leq \rho \leq R$
$0 \leq \theta \leq \pi/2$
$0 \leq \phi \leq \pi/2$

(c) $r \leq \rho \leq R$
$0 \leq \theta \leq 2\pi$
$0 \leq \phi \leq \pi$

(d) $0 \leq \rho \leq R$
$0 \leq \theta \leq \pi$
$0 \leq \phi \leq \pi$

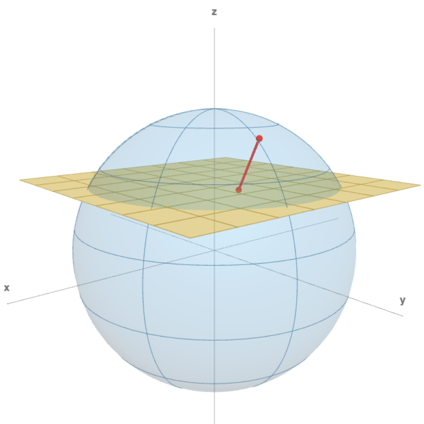
(e) $0 \leq \rho \leq R$
$0 \leq \theta \leq \pi$
$0 \leq \phi \leq \pi/2$

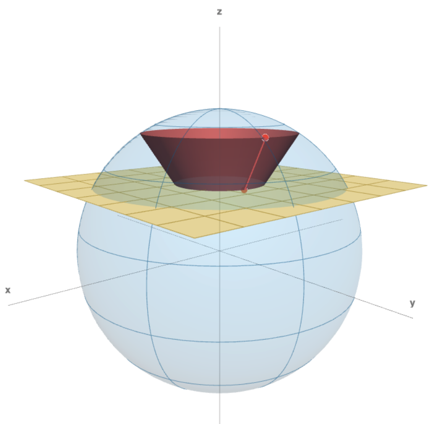
(f) $0 \leq \rho \leq R$
$0 \leq \theta \leq 2\pi$
$0 \leq \phi \leq \pi/6$

9.7.3 The geometry of the *slicing method* in spherical coordinates follows the same strategy as previous coordinate systems, but it can be difficult to visualize. It helps to break down the steps slice by slice. Let S be the solid inside the sphere $x^2 + y^2 + z^2 = 4$ and above the horizontal plane $z = 1$ with mass density $f(x, y, z)$. You will express the mass of S in three different ways using spherical coordinates.

(a) Express S in rectangular form and in spherical form.

For (b), (c), (d), play with this Math3D demo to see how the slices sweep out the solid.



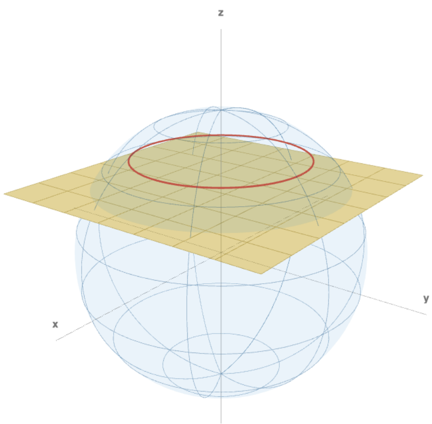


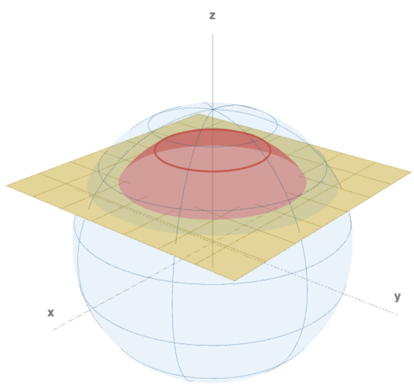
(b) Express the "1D mass density" of a (θ, ϕ)-slice as a single integral $d\rho$.

(c) Express the "2D mass density" of a ϕ-slice as a double integral $d\rho d\theta$.

(d) Set up a triple integral in spherical coordinates $d\rho d\theta d\phi$ representing the mass of S.

For (e), (f), (g), play with this Math3D demo to see how the slices sweep out the solid.



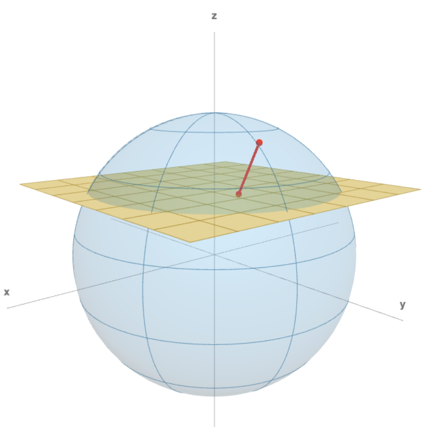


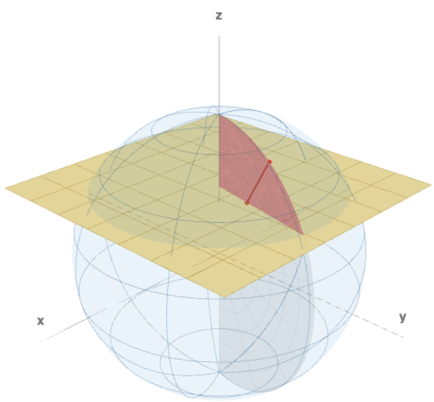
(e) Express the "1D mass density" of a (ρ, ϕ)-slice as a single integral $d\theta$.

(f) Express the "2D mass density" of a ρ-slice as a double integral $d\theta d\phi$.

(g) Set up a triple integral in spherical coordinates $d\theta d\phi d\rho$ representing the mass of S.

For (h), (i), (j), play with this Math3D demo to see how the slices sweep out the solid.





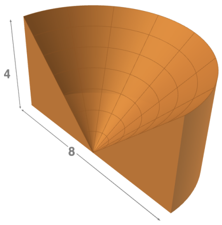
(h) Express the "1D mass density" of a (θ, ϕ)-slice as a single integral $d\rho$.

(i) Express the "2D mass density" of a θ-slice as a double integral $d\rho\, d\phi$.

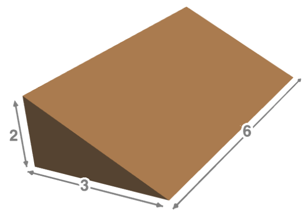
(j) Set up a triple integral in spherical coordinates $d\rho\, d\phi\, d\theta$ representing the mass of S.

(k) There 3 remaining orders of integration in spherical coordinates, and the geometry will look a bit different for each one. Repeat the approach above for each of the remaining 3 orders. Include corresponding sketches of your intermediate steps. This might take quite a while.

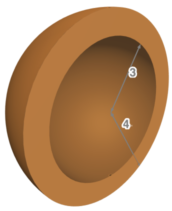
9.7.4 Describe each region with one of rectangular, cylindrical, or spherical coordinates. Based on your choice, write an iterated integral for the volume of each solid.



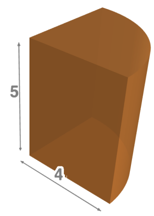
(a)



(c)



(b)



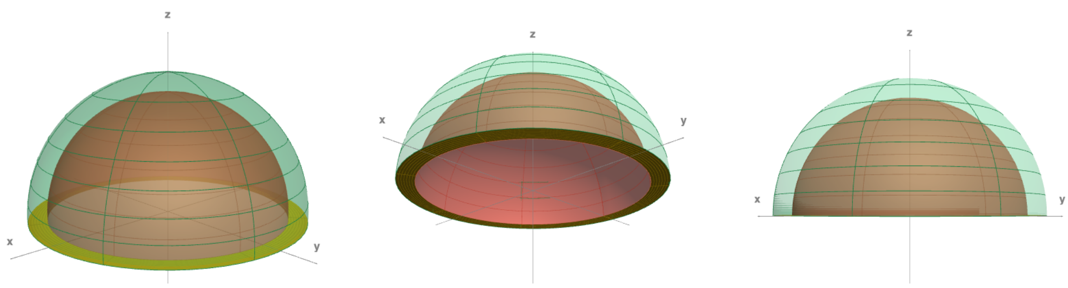
(d)

Computations

9.7.5 Use spherical coordinates to show the volume of a ball of radius R in $\mathbb{R}^3$ is equal to $\dfrac{4\pi R^3}{3}$.

9.7.6 Let S be the solid formed between two spheres with radii 3 and 4 centered at the origin. Write a triple integral for the volume of S using spherical coordinates.

9.7.7 A carved melon M is described by the region $3 \le x^2 + y^2 + z^2 \le 4$ lying above the $z = 0$ plane. This carved melon has variable density $\delta(x, y, z) = x^2 + y^2 + z^2$. You have previously expressed its mass in rectangular coordinates and spherical coordinates, and now you will do so in spherical coordinates.



Express the mass of M with iterated triple integral(s) in spherical coordinates. Evaluate your expression. (Once you finish, you can appreciate how much simpler this calculation was compared to rectangular or cylindrical coordinates!)

9.7.8

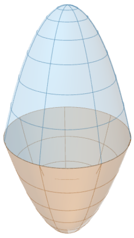
Let S be the solid above the cone $z = \sqrt{x^2 + y^2}$ and below the sphere of radius 4 centered at the origin. Write a triple integral for the volume of S using spherical coordinates.

Applications and beyond

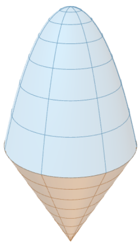
9.7.9

Based on the symmetries of the object, which coordinate system(s) in $\mathbb{R}^3$ would be a good choice? Choose between rectangular, cylindrical, and spherical. More than one may be reasonable.

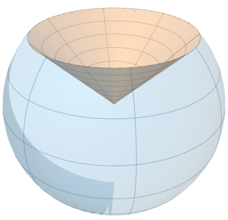
(a) solid between two paraboloids



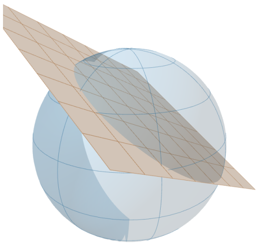
(b) solid between a paraboloid and a cone



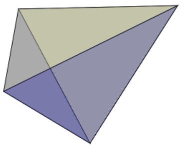
(c) solid inside sphere and below a cone



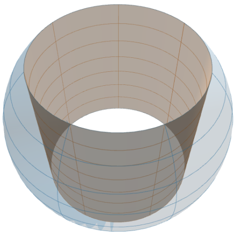
(d) solid sphere cut by plane



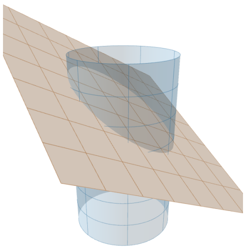
(e) polyhedron



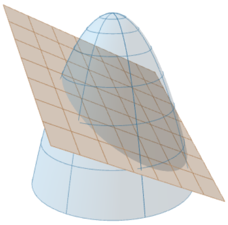
(f) solid sphere cut out by cylinder



(g) solid cylinder cut by plane



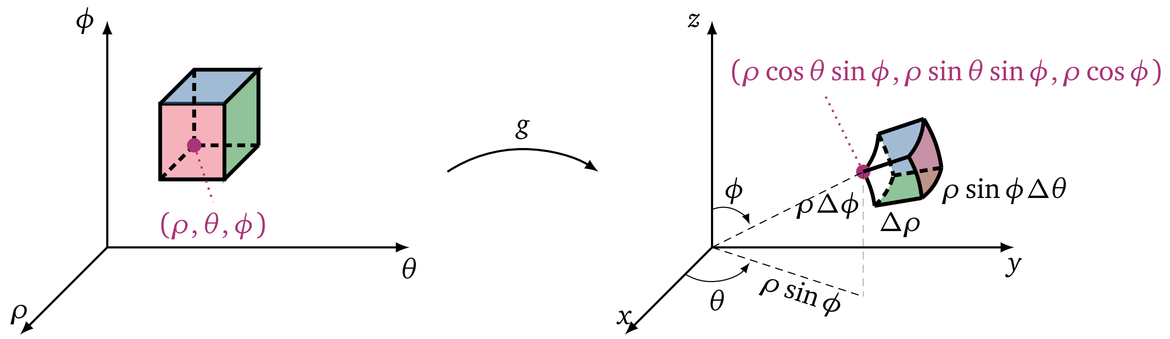
(h) solid paraboloid cut by plane



9.7.10 Here you will justify a key step to deriving the method for integrating using the spherical coordinate transformation $g : \mathbb{R}^3 \to \mathbb{R}^3$ given by $g(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$.

> 1. *For a fixed* $\rho > 0, \theta \in [0, 2\pi, \phi \in [0, \pi]$*, the small rectangle*
>
> $$R = [\rho, \rho + \Delta\rho] \times [\theta, \theta + \Delta\theta] \times [\phi, \phi + \Delta\phi]$$
>
> *in* (ρ, θ, ϕ)*-space transforms under* g *to a spherical rectangle* $g(R)$ *in* (x, y, z)*-space.*
> 2. *Since the rectangle is small, the radial thickness of the transformed rectangle is approximately*
>
> $$||g(\rho + \Delta\rho, \theta, \phi) - g(\rho, \theta, \phi)|| \approx \underline{\hspace{6em}} = \underline{\hspace{6em}} = \Delta\rho.$$
>
> 3. *Similarly,*
>
> $$||g(\rho, \theta + \Delta\theta, \phi) - g(\rho, \theta, \phi)|| \approx \underline{\hspace{6em}} = \underline{\hspace{6em}} = \rho \sin\phi \Delta\theta,$$
>
> $$||g(\rho, \theta, \phi + \Delta\phi) - g(\rho, \theta, \phi)|| \approx \underline{\hspace{6em}} = \underline{\hspace{6em}} = \rho \Delta\phi.$$
>
> 4. *Overall, this implies that* $\mathrm{vol}(g(R)) \approx \rho^2 \sin\phi \Delta\rho \Delta\theta \Delta\phi = \rho^2 \sin\phi \, \mathrm{vol}(R)$.

(a) Label the four corners of R and four corners of $g(R)$ according to the description in Line 1.



(b) Fill in the 6 blanks in the above argument.

(c) Line 4 follows from Lines 2 and 3 with an assumption about $\frac{\partial g}{\partial \rho}, \frac{\partial g}{\partial \theta}$, and $\frac{\partial g}{\partial \phi}$. Identify that assumption.

9.7.11 Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the spherical coordinate transformation. Let $\Omega = [a, b] \times [\alpha, \beta] \times [\lambda, \mu]$. Assume f is integrable on $g(\Omega)$ and $f \circ g$ is integrable on Ω. Heuristically derive the formula

$$\iiint_{g(\Omega)} f \, dV = \int_\lambda^\mu \int_\alpha^\beta \int_a^b (f \circ g)(\rho, \theta, \phi) |\rho^2 \sin\phi| d\rho \, d\theta \, d\phi.$$

Remember you must first explain how the volume of a small rectangle transforms.

9.8. Change of variables

Using Fubini's theorem, iterated integrals, and some heuristic derivations, you have learned to integrate in many different ways so far. One of the most novel techniques was changing coordinate systems. In two dimensions, you can integrate in rectangular or polar coordinates. In three dimensions, you can integrate in rectangular, cylindrical, or spherical coordinates. The derivations of integration in other coordinate systems were uncannily similar. This repetition suggests an informal yet deep question.

Can you integrate with respect to any coordinate system?

As discussed in Section 1.4, the phrase "coordinate system" is not rigorous. You can formalize this notion using a diffeomorphism (see Section 5.1). Throughout this section, let $U, V \subseteq \mathbb{R}^n$ be open sets and let

$$g : U \to V$$

be a diffeomorphism from U to V. The map g gives a "smooth" bijective correspondence between subsets of $U \subseteq \mathbb{R}^n$ and subsets of $V \subseteq \mathbb{R}^n$. Indeed, by restricting its domains, the polar coordinate transformation can also be viewed as a diffeomorphism; see, for instance, Example 5.1.9. The same holds for cylindrical and spherical coordinate transformations. The generalized framework of diffeomorphisms allows you to construct a more formal question.

For a diffeomorphism $g : U \to V$, does integration over the set $\Omega \subseteq U$ correspond to integration over its image $g(\Omega) \subseteq V$?

With a spectacular blend of linear algebra and multivariable differential calculus, you will establish a marvelous theorem commonly known as *change of variables*. It will provide a new remarkable perspective on integration and act as one of your most powerful integration tools.

9.8.1 Derivation of change of variables

Recall that each coordinate system (polar, cylindrical, spherical) came with its own "stretch factor" since the volume of a transformed rectangle will be "stretched" according to the coordinate transformation. To switch coordinate systems via a diffeomorphism $g : U \to V$, you will need to investigate the same core question for rectangles lying inside U. Diffeomorphisms are typically *nonlinear* maps so, as usual with multivariable calculus, it is better to first ask the same question for *linear* maps.

How does a linear map stretch the volume of a rectangle? Determinants from linear algebra were built exactly for this purpose.

Theorem 9.8.1 Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. For every rectangle R in $\mathbb{R}^n$,

$$\operatorname{vol}(T(R)) = |\det(T)| \operatorname{vol}(R).$$

Proof. While there is technically something to prove here using your definition of volume and determinants, you may freely assume this consequence from linear algebra. ■

This has an elegant informal interpretation.

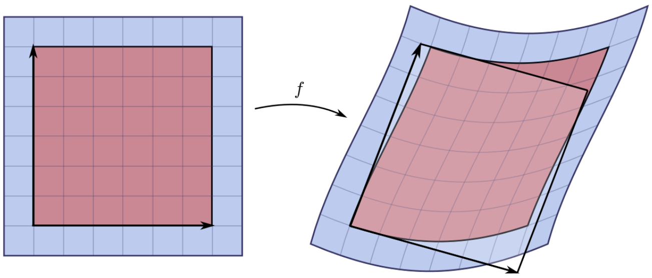
The "stretch factor" of a linear map is the absolute value of its determinant.

Now, your efforts with multivariable differential calculus will pay off in fantastic fashion. You can transfer this fact about linear maps to nonlinear maps via the *differential*.

Let R be a small rectangle lying inside the open subset $U \subseteq \mathbb{R}^n$. Applying the diffeomorphism, this becomes the transformed rectangle $g(R)$ lying inside the open subset $V \subseteq \mathbb{R}^n$. The transformed rectangle $g(R)$ is not necessarily a rectangle in any sense, but you can *linearly approximate* it! Fix a point $a \in R$. The linear approximation of g at a is given by

$$\forall x \in \mathbb{R}^n, \ell_a(x) = g(a) + dg_a(x-a)$$

where dg_a is the differential of g at a. The set $g(R)$ should presumably be linearly approximated by the set $\ell_a(R)$, but what is the shape of this new set? Since the differential $dg_a : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map and R is a rectangle, this set will be a (translated) parallelopiped!



This miraculous insight suggests by Theorem 9.8.1 that

$$\operatorname{vol}(g(R)) \approx \operatorname{vol}(dg_a(R)) = |\det(dg_a)| \operatorname{vol}(R) = |\det Dg(a)| \operatorname{vol}(R), \tag{9.8.1}$$

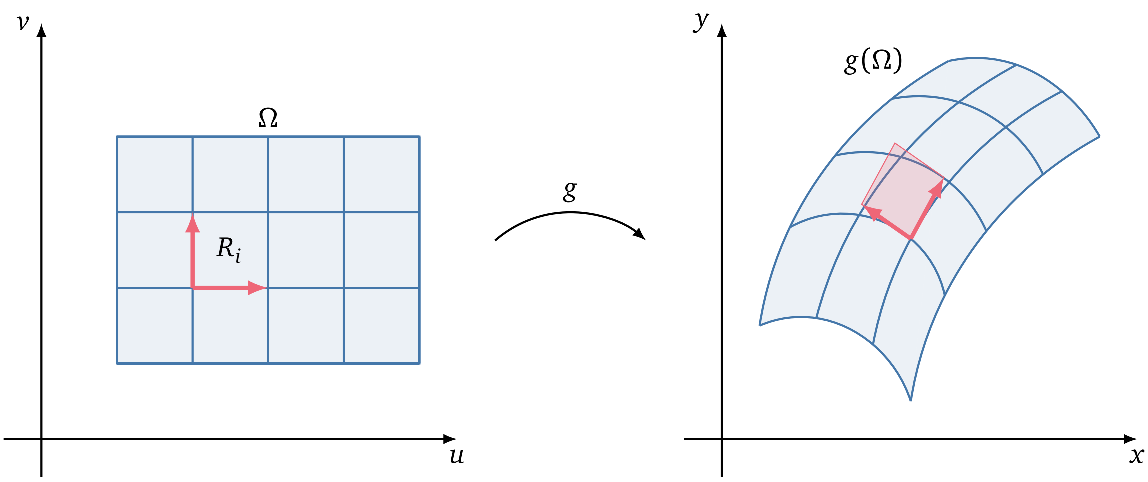
since the Jacobian $Dg(a)$ is the $n \times n$ matrix of the differential dg_a. This heuristic leads to a beautiful conclusion.

> *The "stretch factor" of a nonlinear map is the absolute value of its linear approximation's determinant.*

Equipped with (9.8.1) and the template argument in Section 9.3.2, you can attempt to integrate over the coordinate system defined by g.

For simplicity, assume $\Omega = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is a rectangle lying inside the domain $U \subseteq \mathbb{R}^n$ of $g : U \to V$. Let $P_N = \{R_i\}_{i \in I}$ be a regular partition of Ω into N^n rectangles such that each interval $[a_i, b_i] \subseteq \mathbb{R}$ is divided into N subintervals of width $\Delta x_i = \frac{b_i - a_i}{N}$. Since $g : U \to V$ is a diffeomorphism, the set of transformed rectangles $\{g(R_i)\}_{i \in I}$ should presumably be a collection of compact Jordan measurable sets (not necessarily rectangles) satisfying

$$\bigcup_{i \in I} g(R_i) = g(\Omega) \qquad \text{and} \qquad \forall i, j \in I, i \neq j \implies g(R_i)^\circ \cap g(R_j)^\circ = \varnothing.$$



For each subrectangle R_i of P, choose a sample point $x_i^* \in R_i$, so it presumably follows that

$$\begin{aligned}\int_{g(\Omega)} f\,dV = \sum_{i\in I}\int_{g(R_i)} f\,dV &\approx \sum_{i\in I} f(g(x_i^*))\,\mathrm{vol}\big(g(R_i)\big)\\ &\approx \sum_{i\in I} f\circ g(x_i^*)|\det Dg(x_i^*)|\,\mathrm{vol}(R_i) \qquad \text{by (9.8.1)}\end{aligned}$$

This last sum is a Riemann sum for the function $F = (f\circ g)|\det Dg| : \Omega \to \mathbb{R}$. Taking $N \to \infty$ (and hence $||P_N|| \to 0$) and assuming $(f\circ g)|\det Dg|$ is integrable on Ω, you can heuristically derive by Theorem 7.3.18 that

$$\int_{g(\Omega)} f\,dV = \lim_{N\to\infty} S^*_{P_N}(F) = \int_{\Omega} (f\circ g)|\det Dg|dV$$

for the rectangle $\Omega = [a_1, b_1]\times\cdots\times[a_n, b_n]$. That was a slick derivation.

9.8.2 Statement and consequences

You can finally formulate a grand theorem.

Theorem 9.8.2 (Change of variables) Let $U, V \subseteq \mathbb{R}^n$ be open sets. Let $g : U \to V$ be a diffeomorphism. Let $\Omega \subseteq U$ be a Jordan measurable set. Assume $|\det Dg|$ is bounded on Ω. The function f is integrable on $g(\Omega)$ if and only if $(f\circ g)|\det Dg|$ is integrable on Ω. If so,

$$\int_{g(\Omega)} f\,dV = \int_{\Omega} (f\circ g)|\det Dg|dV.$$

If additionally Fubini's theorem is satisfied for both integrals, then

$$\int\cdots\int_{g(\Omega)} f(x)dx_1\cdots dx_n = \int\cdots\int_{\Omega} (f\circ g)(u)|\det Dg(u)|du_1\cdots du_n.$$

Proof. The formal proof is beyond the scope of this text; see [18, Section 7.5] for details. ■

Remark 9.8.3 Notice this theorem requires that $g|_{\Omega} : \Omega \to g(\Omega)$ is a bijection, but that is not sufficient for many applications, e.g. polar coordinates. As usual, the issue usually occurs on the boundary $\partial\Omega$ which is a set of zero Jordan measure; thus, you can instead apply Theorem 9.8.2 to Ω^{o} and use the ideas in Example 9.3.5 to "add back" the boundary.

Remark 9.8.4 Many other texts refer to the function $|\det Dg| : U \to (0,\infty)$ as the Jacobian and Dg is the Jacobian matrix, but you will refer to $|\det Dg|$ as the absolute determinant of the Jacobian, and Dg as the Jacobian. More informally, $|\det Dg|$ is the "stretch factor" of g.

Informally speaking, the change of variables theorem has a profound meaning.

> *Integration in $\mathbb{R}^n$ does not depend on your choice of coordinate system!*

This philosophy will be a prominent feature for physical or geometric quantities defined by integrals, especially in the next chapters on vector calculus. You can also remember the change of variables theorem by the symbolic identity

$$x = g(u) \implies dx_1\cdots dx_n = dV = |\det Dg(u)|du_1\cdots du_n.$$

This incredible theorem also has a number of intriguing consequences. First, change of variables over $\mathbb{R}^n$ is the familiar integration technique of substitution over $\mathbb{R}$.

Corollary 9.8.5 Let $g : [a,b] \to \mathbb{R}$ be continuous, C^1 on (a,b) with g' bounded and $g' > 0$. Then f is integrable on $[g(a), g(b)]$ if and only if $f \circ g$ is integrable on $[a,b]$. If so,

$$\int_{g(a)}^{g(b)} f(x)dx = \int_a^b f(g(u))g'(u)du.$$

Proof. This is left as an exercise. ■

Change of variables describes how volume transforms under a diffeomorphism.

Corollary 9.8.6 Let $U, V \subseteq \mathbb{R}^n$ be open sets. Let $g : U \to V$ be a diffeomorphism. If $\Omega \subseteq U$ is Jordan measurable and $|\det Dg|$ is bounded on Ω, then

$$\mathrm{vol}(g(\Omega)) = \int_\Omega |\det Dg| dV.$$

Proof. This is left as an exercise. ■

You can also strengthen Theorem 9.8.1 for linear maps.

Corollary 9.8.7 Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear map. If $\Omega \subseteq \mathbb{R}^n$ is Jordan measurable, then

$$\mathrm{vol}(T(\Omega)) = |\det(T)|\,\mathrm{vol}(\Omega).$$

Proof. This is left as an exercise. ■

As you may recall from Section 5.1, there is a challenge in applying the change of variables theorem: how do you verify whether a map is a diffeomorphism? It is not easy, but the inverse function theorem slightly simplifies the process.

Lemma 9.8.8 Let $U, V \subseteq \mathbb{R}^n$ be open sets. If $g : U \to V$ is C^1, bijective, and $\det Dg(x) \neq 0$ for all $x \in U$, then g is a diffeomorphism.

Proof. By Definition 5.1.4, it suffices to prove that the inverse of g is C^1. Let $f = g^{-1} : V \to U$ be the inverse of g. Fix $y \in V$ and set $x = f(y)$ so $y = g(x)$. We wish to prove f is C^1 at y.

Since $\det Dg(x) \neq 0$, the Jacobian $Dg(x)$ is invertible so, by the inverse function theorem (Theorem 5.2.8), g is a local diffeomorphism at x. By Definition 5.1.16, there exists an open set $W \subseteq U$ such that $x \in W$ and $g|_W : W \to g(W)$ is a diffeomorphism. Since inverses are unique, it must be that $f|_{g(W)}$ is the inverse of $g|_W$. As $g|_W$ is a diffeomorphism, it follows that $f|_{g(W)}$ is C^1 on its domain $g(W)$, which contains $y = g(x)$. As the restriction $f|_{g(W)}$ is C^1 at y, you may conclude that f is C^1 at y. ■

Lemma 9.8.8 is a standard tool for applying a change of variables.

9.8.3 Examples with change of variables

First, you can revisit some familiar coordinate systems.

Example 9.8.9 The polar coordinate transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $g(r,\theta) = (r\cos(\theta), r\sin(\theta))$. You can apply the change of variables theorem to various restrictions of g provided they are diffeomorphisms, such as $g|_{(0,\infty)\times(0,2\pi)}$ or $g|_{(0,\infty)\times(-\pi,\pi)}$. For any

$(r, \theta) \in \mathbb{R}^2$, the stretch factor in the change of variables theorem will be

$$|\det Dg(r,\theta)| = \left|\det \begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix}\right| = |r\cos^2\theta + r\sin^2\theta| = |r|.$$

This confirms the stretch factor for the polar coordinate transformation.

Example 9.8.10 Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the cylindrical coordinate transformation given by $g(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$. You can also apply the change of variables theorem to many restrictions of g provided it is a diffeomorphism, such as $g|_{(0,\infty)\times(0,2\pi)\times(-\infty,\infty)}$. It is left as an exercise to verify that $|\det Dg(r,\theta,z)| = |r|$.

Example 9.8.11 Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the spherical coordinate transformation given by $g(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$. You can also apply the change of variables theorem to many restrictions of g provided it is a diffeomorphism, such as $g|_{(0,\infty)\times(0,2\pi)\times(0,\pi)}$. It is left as an exercise to verify that $|\det Dg(\rho,\theta,\phi)| = |\rho^2\sin\phi|$.

Change of variables allows you to shift your approach when integrating over complicated regions. By choosing a suitable coordinate system, you can simplify your region of integration. It is tedious to verify all the details for a change of variables, so for now you will focus on the calculation without any justifications.

Example 9.8.12 Let $P \subseteq \mathbb{R}^2$ be the region bounded between the curves $y = x$, $y = 2x$, $y = x^3$, and $y = 3x^3$ in the first quadrant with density $f(x,y) = \frac{y^2}{x^6}$. As shown in Example 9.2.8, you can express the mass of P a sum of iterated integrals $dy\,dx$ in rectangular coordinates. The complexity of the region forced you to break it up into several pieces. You will find the mass of P by applying a change of variables. You will choose the relevant quantities in Theorem 9.8.2 but, for the sake of brevity, you will not verify all of the conditions.

Notice by definition

$$P = \{(x,y) \in (0,\infty)^2 : x \le y \le 2x, x^3 \le y \le 3x^3\}$$

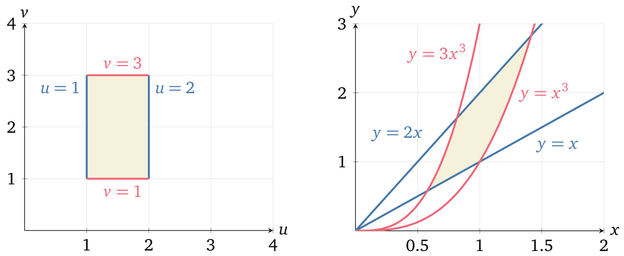
Define $h : (0,\infty)^2 \to (0,\infty)^2$ by

$$h(x,y) = \left(\frac{y}{x}, \frac{y}{x^3}\right).$$

You can verify that h is a diffeomorphism. For clarity, use the variables $(u,v) = h(x,y)$ to refer to the codomain of h. In this case, it follows that the image

$$h(P) = \{(u,v) \in (0,\infty)^2 : 1 \le u \le 2, 1 \le v \le 3\} = [1,2]\times[1,3] \tag{9.8.2}$$

is a rectangle! This choice is illustrated below.



Notice h maps from the (x,y)-plane to the (u,v)-plane. Define the diffeomorphism

$$g := h^{-1} : (0,\infty)^2 \to (0,\infty)^2$$

and take $\Omega = [1,2]\times[1,3]$. By the change of variables theorem (Theorem 9.8.2) and (9.8.2), the mass of the plate $P = g(\Omega)$ is given by

$$m = \iint_{g(\Omega)} f\,dA = \iint_{\Omega} (f\circ g)|\det Dg|dA.$$

By Fubini's theorem, it follows that

$$m = \int_1^3 \int_1^2 (f\circ g)(u,v)|\det Dg(u,v)|du dv. \tag{9.8.3}$$

Since the region is a rectangle, this will be a piece of cake to calculate once you can express the integrand in terms of the variables u and v. There are two reasonable methods at this stage. A first option is to explicitly calculate $g(u,v)$ in terms of u and v and plug everything into the integrand. This is left as an exercise.

A second option is to calculate $(f\circ g)(u,v)$ and the stretch factor $|\det Dg(u,v)|$ in terms of u and v without finding any inverses. This requires some luck and cleverness with algebraic manipulations, but it can often make for more concise calculations. For $(x,y)\in(0,\infty)^2$ and $(u,v) = h(x,y) = (y/x, y/x^3)$, it follows that

$$f(x,y) = \frac{y^2}{x^6} = \left(\frac{y}{x^3}\right)^2 = v^2$$

so $f\circ g(u,v) = v^2$. Since

$$\det Dh(x,y) = \det\begin{bmatrix} -y/x^2 & 1/x \\ -3y/x^4 & 1/x^3 \end{bmatrix} = \frac{2y}{x^5}$$

and the Jacobian $Dg(u,v)$ is the inverse of the Jacobian $Dh(x,y)$ by Theorem 5.2.1,

$$|\det Dg(u,v)| = \frac{1}{|\det Dh(x,y)|} = \left|\frac{2y}{x^5}\right|^{-1} = \frac{x^5}{2y} = \frac{1}{2}\left(\frac{x^3}{y}\right)^2\left(\frac{y}{x}\right) = \frac{u}{2v^2}.$$

The last step took a bit of brute force cleverness. Combining these calculations in (9.8.3), you can conclude that

$$m = \int_1^3\int_1^2 v^2\frac{u}{2v^2}du dv = \int_1^3\int_1^2 \frac{1}{2}u du dv = \int_1^3 \frac{3}{4}dv = \frac{3}{2}.$$

There are a lot of details to verify to rigorously apply a change of variables.

Example 9.8.13 Here you will justify the change of variables applied in Example 9.8.12. The goal is to apply the change of variables theorem with the following choices:

- $U = V = (0,\infty)^2$ and $g : U\to V$ given by $g(u,v) = \left(u^{1/2}v^{-1/2}, u^{3/2}v^{-1/2}\right)$
- $f(x,y) = y^2/x^6$
- $\Omega = [1,2]\times[1,3]$.

The formula for g takes a bit of effort to find, but that is left as an exercise. To formally apply this change of variables (Theorem 9.8.2), it suffices to prove all of the following.

(a) U and V are open
(b) g is C^1 and $\det Dg$ is never zero.
(c) Ω is a Jordan measurable subset of U and $|\det Dg|$ is bounded on Ω
(d) $(f \circ g)|\det Dg|$ is integrable on Ω
(e) g is invertible on U

Notice (a), (b), and (e) imply that g is a diffeomorphism by Lemma 9.8.8. Also, (d) can be equivalently replaced with showing f is integrable on $g(\Omega)$. A mostly complete proof is provided for each for these items.

(a) $U = V = (0, \infty)^2$ are Cartesian products of open intervals and hence they are open.
(b) The components of $g(u, v)$ are both C^1 on their domains since $(u, v) \in (0, \infty)^2$ implies $v \neq 0$. Hence, g is C^1. Moreover,

$$\det Dg(u, v) = \det\left(\frac{1}{2}\begin{bmatrix} u^{-1/2}v^{-1/2} & -u^{1/2}v^{-3/2} \\ 3u^{1/2}v^{-1/2} & -u^{3/2}v^{-3/2} \end{bmatrix}\right) = \frac{u}{2v^2}$$

so $\det Dg(u, v) \neq 0$ as $u \neq 0$ for $(u, v) \in (0, \infty)^2$. Then $\det Dg$ is non-zero on U.
(c) $\Omega = [1, 2] \times [1, 3]$ is a rectangle, which is Jordan measurable and $\Omega \subseteq (0, \infty)^2 = U$. Moreover, $|\det Dg(u, v)| = \frac{u}{2v^2} \leq \frac{2}{2\cdot 1^2} \leq 1$ for $(u, v) \in \Omega$, so $|\det Dg|$ is bounded on Ω.
(d) Notice that $f \circ g(u, v) = v^2$ and $|\det Dg(u, v)| = \frac{u}{2v^2}$. Thus, $(f \circ g)|\det Dg|$ is a ratio of polynomials with non-zero denominator on $U = (0, \infty)^2$. Hence, $(f \circ g)|\det Dg|$ continuous on U and thus integrable on the rectangle $\Omega = [1, 2] \times [1, 3] \subseteq U$.
(e) Verifying that $g : U \to V$ is invertible is the most tedious. Often it will be happily assumed without proof, but a full proof is included here for the sake of completeness. If you are bored, you may skip the details. You can first prove g is injective.

Proof. Let $(u_0, v_0), (u_1, v_1) \in U$ so $u_0, u_1, v_0, v_1 > 0$. By definition of g, we have

$$u_0^{1/2}v_0^{-1/2} = u_1^{1/2}v_1^{-1/2} \text{ and } u_0^{3/2}v_0^{-1/2} = u_1^{3/2}v_1^{-1/2}$$
$$\Longrightarrow v_0 = \frac{u_0}{u_1}v_1 \text{ and } u_0 = \left(\frac{v_0}{v_1}\right)^{\frac{1}{3}}u_1$$

after squaring both sides and rearranging. Using the second equation in the first gives $v_0^{\frac{2}{3}} = v_1^{\frac{2}{3}}$. Since v_0, v_1 are positive, we have $v_0 = v_1$. Plugging this into the second equation gives $u_0 = u_1$. Then $(u_0, v_0) = (u_1, v_1)$. ■

Second you can prove that $g : U \to V$ is surjective.

Proof. Let $(x, y) \in V$ so $x, y > 0$. Fix $(u, v) = (\frac{y}{x}, \frac{y}{x^3})$. It suffices to show that $(u, v) \in U$ and $g(u, v) = (x, y)$. Since $x, y > 0$, it is immediate that $u, v > 0$ so $(u, v) \in (0, \infty)^2 = U$. Now, by definition of V we have

$$\begin{aligned} g(u, v) &= \left(\sqrt{\tfrac{u}{v}}, \sqrt{\tfrac{u^3}{v}}\right) = \left(\sqrt{\tfrac{y}{x}\tfrac{x^3}{y}}, \sqrt{\left(\tfrac{y}{x}\right)^3\tfrac{x^3}{y}}\right) && \text{by choice of } u \text{ and } v \\ &= (\sqrt{x^2}, \sqrt{y^2}) = (x, y) && \text{as } x, y > 0. \end{aligned}$$

This completes the proof that g is surjective. ■

Thus, g is invertible.

This completes the verification of the assumptions in the change of variables theorem. Phew!

A change of variables can be used to quickly evaluate integrals via symmetry.

Example 9.8.14 You informally used symmetry and Fubini's theorem in Example 9.2.9 to argue for the unit disk $S = \overline{B_1(0)}$ and $f(x,y) = \sin(x+y)$ that

$$\iint_S f(x,y)dA = 0. \tag{9.8.4}$$

You will follow a more elegant argument with a change of variables. Define $g : \mathbb{R}^2 \to \mathbb{R}^2$ by $g(x,y) = (-x,-y)$. Notice g is an invertible linear map, so it is a diffeomorphism.

Next, you check that $g(S) = S$. If $(x,y) \in S$ then $(-x)^2 + (-y)^2 = x^2 + y^2 \le 1$, so $g(x,y) = (-x,-y) \in S$. Hence, $S \subseteq g(S)$. Similarly, if $(x,y) \in g(S)$ then there exists $(x',y') \in S$ such that $(x,y) = g(x',y')$ so $x = -x'$ and $y = -y'$. Thus, $x^2 + y^2 = (-x')^2 + (-y')^2 = (x')^2 + (y')^2 \le 1$ implying $(x,y) \in S$. Hence, $g(S) \subseteq S$ so $S = g(S)$.

By definition of the Jacobian,

$$Dg(x,y) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \implies |\det Dg(x,y)| = 1.$$

Therefore, it follows that

$$\begin{aligned} \iint_S f\,dA = \iint_{g(S)} f\,dA &= \iint_S (f \circ g)|\det Dg|dA && \text{by change of variables} \\ &= \iint_S \sin(-x-y)dA \\ &= -\iint_S \sin(x+y)dA = -\iint_S f\,dA. \end{aligned}$$

This proves (9.8.4).

The change of variables theorem is a magnificent triumph of integral calculus. Its proof is a beautiful blend of linear algebra and differential calculus. Its statement shows that *integration does not depend on the choice of coordinates*. You can chop your region up using rectangles, cylinders, spheres, or whatever shape you want! With an appropriate stretch factor, the integral will be the same. This consequence displays the robustness of your definition of the integral and the Jordan measure from Chapter 7. You will apply change of variables in many scenarios. Your first opportunity will be improper integrals in higher dimensions.

Exercises for Section 9.8

Concepts and definitions

9.8.1 Let $\Omega \subseteq \mathbb{R}^2$ be a Jordan measurable set with $\operatorname{vol}(\Omega) = 0.1$. Assume $(1,2) \in \Omega$.

(a) Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map given by $T(x,y) = \begin{bmatrix} 2 & 7 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$. Evaluate $\operatorname{vol}(T(\Omega))$.

(b) Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be a diffeomorphism. Assume $Dg(1,2) = \begin{bmatrix} 2 & 7 \\ 0 & 3 \end{bmatrix}$. Approximate $\operatorname{vol}(g(\Omega))$.

9.8.2 A good change of variables allows you to transform a complicated region S into a simpler one. Here you will learn how to decide whether a given change of variables is good. You must choose a map g and a region Ω such that:

I) $g : U \to V$ is diffeomorphism
II) $g(\Omega) \subseteq V$ is your current region of integration, so $g(\Omega) = S$.
III) $\Omega \subseteq U$ is your new region and easy to describe, e.g. a rectangle.

Then $(x,y) = g(u,v)$ is how you change variables. You want to apply a change of variables to

$$S = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 4, x \ge 0, y \ge 0\}.$$

Which choices satisfy I, II, and III?

(a) $\Omega_A = [1,2] \times [1,2]$ and $g_A : \mathbb{R}^2 \to \mathbb{R}^2$ given $g_A(u,v) = (u,v)$.

(b) $\Omega_B = [1,2] \times [0,2\pi]$ and $g_B : \mathbb{R}^2 \to \mathbb{R}^2$ given by $g_B(u,v) = (u\cos v, u\sin v)$.

(c) $\Omega_C = [1,2] \times [0,\frac{\pi}{2}]$ and $g_C : \mathbb{R}^2 \to \mathbb{R}^2$ given by $g_C(u,v) = (u\cos v, u\sin v)$.

9.8.3 In practice, you will often choose a change of variables by first defining the *inverse* change of variables. You must choose a map h and a region Ω such that:

I) $h : V \to U$ is diffeomorphism
II) $h^{-1}(\Omega) \subseteq V$ is your current region of integration, so $h(S) = \Omega$.
III) $\Omega \subseteq U$ is your new region and easy to describe, e.g. a rectangle.

Then $(u,v) = h(x,y)$ is how you change variables. You want to apply a change of variables to

$$S = \{(x,y) \in \mathbb{R}^2 : 2 \le x - y \le 8, 0 \le 2x + y \le 3\}.$$

Which choices satisfy I, II, and III? If so, determine the change of variables map $g = h^{-1}$.

(a) $\Omega_A = [2,8] \times [0,3]$ and $h_A : \mathbb{R}^2 \to \mathbb{R}^2$ given $h_A(x,y) = (x,y)$.

(b) $\Omega_B = [2,8] \times [0,3]$ and $h_B : \mathbb{R}^2 \to \mathbb{R}^2$ given $h_B(x,y) = (x-y, 2x+y)$.

(c) $\Omega_C = [0,3] \times [2,8]$ and $h_C : \mathbb{R}^2 \to \mathbb{R}^2$ given by $h_C(x,y) = (2x+y, x-y)$.

(d) $\Omega_D = [1,4] \times [0,1]$ and $h_D : \mathbb{R}^2 \to \mathbb{R}^2$ given by $h_D(x,y) = \left(\dfrac{x-y}{2}, \dfrac{2x+y}{3}\right)$.

9.8.4 You want to apply a change of variables to

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 \le y \le 2x^2, 1 \le xy \le 3\}.$$

You must choose a map h and a region Ω such that:

I) $h : V \to U$ is diffeomorphism
II) $h^{-1}(\Omega) \subseteq V$ is your current region of integration, so $h(S) = \Omega$.
III) $\Omega \subseteq U$ is your new region and easy to describe, e.g. a rectangle.

Which choices satisfy I, II, and III? If so, determine the change of variables map $g = h^{-1}$.

(a) $\Omega_A = [1,2]\times[1,3]$ and $h_A : \mathbb{R}^2 \to \mathbb{R}^2$ given $h_A(x,y) = (x,y)$.

(b) $\Omega_B = [1,2]\times[1,3]$ and $h_B : \mathbb{R}^2 \to \mathbb{R}^2$ given $h_B(x,y) = (x^2, xy)$.

(c) $\Omega_C = [1,2]\times[1,3]$ and $h_C : \mathbb{R}^2 \to \mathbb{R}^2$ given by $h_C(x,y) = \left(\frac{y}{x^2}, xy\right)$.

(d) $\Omega_D = [\frac{1}{2},1]\times[1,3]$ and $h_D : (0,\infty)^2 \to (0,\infty)^2$ given by $h_D(x,y) = \left(\frac{x^2}{y}, xy\right)$.

Computations

9.8.5 Here you will confirm the stretch factors for cylindrical and spherical coordinates.

(a) Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $g(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$. By direct calculations, verify that $|\det Dg(r,\theta,z)| = |r|$.

(b) Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $g(\rho,\phi,\theta) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)$. By direct calculation, verify that $|\det Dg(\rho,\theta,\phi)| = |\rho^2\sin\phi|$.

9.8.6 Redo Example 9.8.12 starting from (9.8.3) and follow the first option of explicitly finding a formula for $g(u,v)$.

9.8.7 Here you will step-by-step apply a change of variables to calculate

$$I := \iint_S (x^3 + y^3)\,dA$$

where

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 \le y \le 2x^2, 1 \le xy \le 3\}.$$

Do not attempt to justify the details of the change of variables.

(a) Define the inverse change of variables map $h : (0,\infty)^2 \to (0,\infty)^2$ by

$$h(x,y) = \left(\frac{y}{x^2}, xy\right).$$

Assume h is a diffeomorphism without proof. Identify a set Ω such that $h(S) = \Omega$ and confirm it has the desired properties.

(b) The set S can be described with coordinates $(u,v) = h(x,y)$. Compute the inverse $g = h^{-1}$ and sketch the transformation g from the (u,v)-plane to the (x,y)-plane.

(c) Find the inverse Jacobian determinant $\det\left[Dh(x,y)\right]$ for $(x,y) \in (0,\infty)^2$.

(d) Find the Jacobian determinant $\det\left[Dg(u,v)\right]$ for $(u,v) \in (0,\infty)^2$.

(e) Calculate I using this change of variables.

9.8.8 Let P be the 3-dimensional parallelopiped whose sides lie in the planes

$$z - y = 0, \qquad z - y = 3, \qquad x - y = 0, \qquad x - y = 2,$$

$$x + y + z = 1 \qquad \text{and} \qquad x + y + z = 2.$$

Evaluate $\iiint f\,dV$ where $f(x,y,z) = (x+y+z)^2$.

9.8.9 Let P be the parallelogram in the (x, y) plane with vertices $(0,0)$, $(3,1)$, $(-1,1)$, and $(2,2)$.

(a) Find coordinates u and v so that P can be described as the unit square $0 \le u \le 1$, $0 \le v \le 1$. Give formulas for x and y in terms of u and v, and formulas for u and v in terms of x and y.

(b) Calculate the Jacobian for the transformation from (u, v)-coordinates to (x, y)-coordinates.

(c) Calculate the Jacobian for the transformation from (x, y) coordinates to (u, v) coordinates:

(d) Calculate

$$\int\int_P 3y - x dx dy$$

using the (u, v) coordinates.

9.8.10 Let Q be the solid 3-dimensional parallelopiped with one vertex at the origin spanned by the vectors $(3,0,0)$, $(1,0,1)$, and $(0,2,1)$. Calculate

$$\int\int\int_Q xy dx\, dy\, dz$$

.

9.8.11 Find the area of the first-quadrant region bounded by the curves $y = x^2$, $y = 4x^2$ and $x = 2y^2$, $x = 6y^2$.

9.8.12 Find the area of the first quadrant region bounded by the lines $y = x$, $y = 2x$, and the hyperbolas $xy = 1$ and $xy = 2$.

9.8.13 Find the volume of the region in $\mathbb{R}^3$ bounded by the xy-plane, the paraboloid $z = x^2 + y^2$, and the elliptic cylinder

$$\frac{x^2}{9} + \frac{y^2}{25} = 1$$

9.8.14 Use 4-dimensional spherical coordinates to compute the volume of the unit 4-ball.

Proofs

9.8.15 To verify whether a map is a diffeomorphism, you might use Lemma 9.8.8 for some coordinate systems. When applying this lemma here, assume your choices are bijective without proof.

(a) Choose open sets $U, V \subseteq \mathbb{R}^2$ as large as possible such that the polar coordinate transformation $g : U \to V$ given by $g(r, \theta) = (r\cos\theta, r\sin\theta)$ is C^1, invertible, and $\det Dg \neq 0$ on U. Is your answer unique?

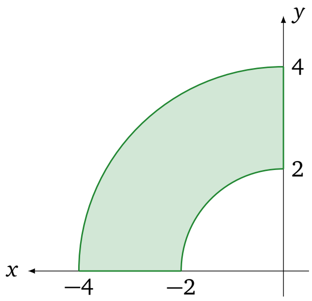
(b) Choose open sets $U, V \subseteq \mathbb{R}^3$ as large as possible such that the cylindrical coordinate transform $g : U \to V$ given by $g(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$ is C^1, invertible, and $\det Dg \neq 0$ on U. Is your answer unique?

(c) Choose open sets $U, V \subseteq \mathbb{R}^3$ as large as possible such that the spherical coordinate transform $g : U \to V$ given by $g(\rho, \phi, \theta) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)$ is C^1, invertible, and $\det Dg \neq 0$ on U. Is your answer unique?

9.8.16 Use a change of variables theorem to rewrite the integral

$$\iint_S y\,dA$$

in polar coordinates where S is the illustrated region. Justify your change of variables.



9.8.17 Justify your change of variables in Exercise 9.8.7.

9.8.18 Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear map. Justify a change of variables to prove that for any Jordan measurable set $\Omega \subseteq \mathbb{R}^n$,

$$\mathrm{vol}(T(\Omega)) = |\det(T)|\,\mathrm{vol}(\Omega).$$

9.8.19 Fix $a, b, c > 0$. Use a change of variables to show that the volume of an ellipsoid in $\mathbb{R}^3$ given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1$$

is equal to $\dfrac{4\pi abc}{3}$. Justify your change of variables.

9.8.20 Let $g : [a, b] \to \mathbb{R}$ be continuous, C^1 on (a, b) with g' bounded and $g' > 0$. Prove that f is integrable on $[g(a), g(b)]$ if and only if $f \circ g$ is integrable on $[a, b]$. If so,

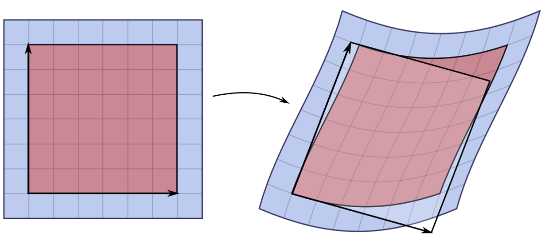
$$\int_{g(a)}^{g(b)} f(x)dx = \int_a^b f(g(u))g'(u)du.$$

Applications and beyond

9.8.21 Let $U, V \subseteq \mathbb{R}^2$ be open sets. Let $g : U \to V$ be a diffeomorphism. Fix $(a, b) \in U$. Let $R \subseteq U$ be a small rectangle such that

$$R = [a, a + \Delta a] \times [b, b + \Delta b]$$

An illustration of g mapping R to $g(R)$ is included below.



By labelling this picture in stages, you will illuminate the big idea for change of variables.

(a) Label the sets $U, V, R, g(R)$, and the mapping g.

(b) Label the point (a, b) and its image $g(a, b)$.

(c) Define $\ell : \mathbb{R}^2 \to \mathbb{R}^2$ to be the linear approximation of g at (a,b). Express ℓ using g.

$$\forall (x,y) \in \mathbb{R}^2, \quad \ell(x,y) = \underline{\hspace{8cm}}$$

(d) Label the set $\ell(R)$.

(e) Use the standard basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ to label the vectors adjacent to (a,b) in U.

(f) Label the images of those vectors.

(g) Use your labelled diagram and the vectors to heuristically explain why

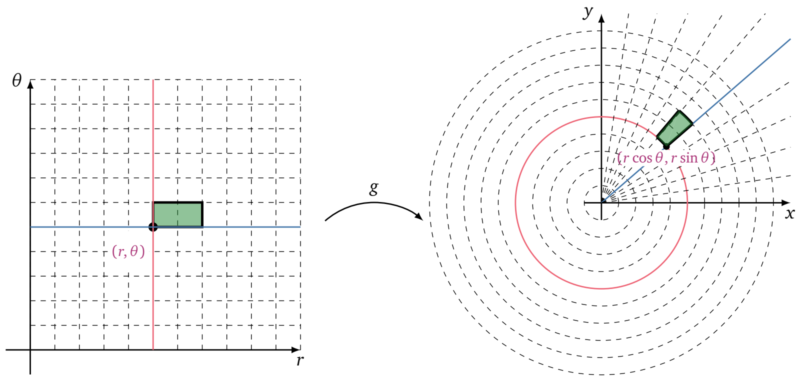
$$\operatorname{vol}(g(R)) \approx |\det Dg(a,b)| \operatorname{vol}(R).$$

9.8.22 To solidify ideas, you can repeat Problem 9.8.21 for polar coordinates and spherical coordinates.

(a) Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinate transformation, so

$$g(r,\theta) = (r\cos\theta, r\sin\theta).$$

Fix $(r,\theta) \in \mathbb{R}^2$ and define $R = [r, r+\Delta r] \times [\theta, \theta + \Delta\theta]$. Let $\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear approximation of g at (r,θ). Use this setup to label the diagram below as in Problem 9.8.21.

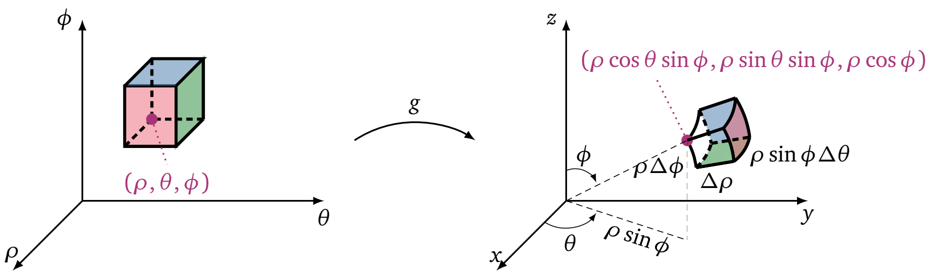


You will need to draw some vectors and a parallelogram.

(b) Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the spherical coordinate transformation so

$$g(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi).$$

Fix $(\rho,\theta,\phi) \in \mathbb{R}^3$ and define $R = [\rho, \rho+\Delta\rho] \times [\theta, \theta+\Delta\theta] \times [\phi, \phi+\Delta\phi]$. Let $\ell : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear approximation of g at (ρ,θ,ϕ). Again, label the diagram as in Problem 9.8.21.



You will need to draw some vectors and a parallelopiped.

9.8.23 The proof of change of variables is highly technical but the basic idea relies on Problem 9.8.21. To simplify the proof a bit, assume stronger hypotheses:

- f is continuous on $g(\Omega)$ and $g(\Omega)$ is Jordan measurable.
- $(f \circ g)|\det Dg|$ is continuous on Ω and Ω is Jordan measurable.

Even so, the proof is hard. Below is a WRONG proof of change of variables with these extra assumptions.

1. *Fix $\varepsilon > 0$. Let R be a rectangle containing Ω.*
2. *Define $F = (f \circ g)|\det Dg|$ so F is continuous on Ω and hence integrable on Ω.*
3. *Since F is continuous on Ω and hence integrable on Ω, there exists a partition $P = \{R_i\}_{i\in I}$ of R such that*
$$\int_\Omega F\,dV - \varepsilon \le L_P(F) \le U_P(F) \le \int_\Omega F\,dV + \varepsilon$$
4. *For each rectangle R_i, by the extreme value theorem, choose a point $t_i \in R_i$ such that $\sup\limits_{x\in R_i} F(x) = F(t_i)$ and a point $s_i \in R_i$ such that $\inf\limits_{x\in R_i} F(x) = F(s_i)$.*
5. *Thus, by definition of F,*
$$U_P(F) = \sum_{i\in I} F(t_i)\operatorname{vol}(R_i) = \sum_{i\in I}(f \circ g)(t_i)|\det Dg(t_i)|\operatorname{vol}(R_i).$$
6. *From the definition of the differential dg_{t_i} and Jacobian $Dg(t_i)$,*
$$U_P(F) = \sum_{i\in I}(f \circ g)(t_i)\operatorname{vol}(dg_{t_i}(R_i)) = \sum_{i\in I} f(g(t_i))\operatorname{vol}(g(R_i)).$$
7. *Define $s_i' = g(s_i)$ and $t_i' = g(t_i)$ and $R_i' = g(R_i)$. Then*
$$U_P(F) = \sum_{i\in I} f(t_i')\operatorname{vol}(R_i') \quad \text{and similarly} \quad L_P(F) = \sum_{i\in I} f(s_i')\operatorname{vol}(R_i').$$
8. *These two sums are equal to $U_{P'}(f)$ and $L_{P'}(f)$ where $P' = \{R_i'\}_{i\in I}$ is a partition of $R' = g(R)$.*
9. *Since f is continuous on $g(\Omega)$, it follows f is integrable on $g(\Omega)$.*
10. *Thus, by line 3, $\displaystyle\int_\Omega F\,dV - \varepsilon \le L_P(F) = L_{P'}(f) \le \int_{g(\Omega)} f\,dV \le U_{P'}(f) = U_P(F) \le \int_\Omega F\,dV + \varepsilon$*
11. *As $\varepsilon > 0$ was arbitrary, this implies the integrals are equal, as desired.*

Wrong proofs are not always bad. This provides a sketch of the basic ideas so you can learn a lot from critiquing it. Identify the critical errors and gaps which need to be filled. Do not try to fix the proof. *Hint:* Lines 4, 6, and 8 all have independent serious errors.

10. Improper integrals

All of your integration methods thus far enable you to calculate integrals over *bounded* sets and with *bounded* functions. This suffices for many situations but it is not always enough. For applications in probability and physics, it is natural to integrate a function where the region may be unbounded or the function may blow up near a point.

How do you integrate over an unbounded region or with an unbounded integrand?

Such integrals are referred to as *improper*, because they do not satisfy the original definition of integrals. Over $\mathbb{R}$, some classic examples include

$$\int_1^\infty \frac{1}{x}dx, \qquad \int_1^\infty e^{-x}dx, \qquad \int_0^1 \frac{1}{x}dx, \qquad \int_{-1}^1 \frac{1}{x^{1/3}}dx.$$

These are defined using single-variable limits. You will generalize this definition, but there will be several new obstacles in higher dimensions. Once you have formulated a rigorous definition, you will compile a list of convergence theorems that naturally generalize from one dimension. This outcome will constitute a great success for applications in probability and physics.

10.1. Local integrability and exhaustions

A couple of examples illustrate the immediate issues with attempting to integrate over unbounded sets or unbounded functions in higher dimensions.

Example 10.1.1 The integrals

$$\int_{\mathbb{R}^n} \frac{1}{1+||x||^2} dV \qquad \text{and} \qquad \int_{0<||x||<1} \frac{1}{||x||} dV$$

do not make any sense with our current definition. The first integral is not yet defined because the region of integration is unbounded. The second integral is currently undefined because the integrand $||x||^{-1}$ is unbounded on $B_1(0) \setminus \{0\}$. On other hand, $\frac{1}{1+||x||^2}$ is continuous on $\mathbb{R}^n$ so by Theorem 7.7.4, $\frac{1}{1+||x||^2}$ is integrable on the open ball $B_R(0)$ for any $R > 0$. Can you define

$$\int_{\mathbb{R}^n} \frac{1}{1+||x||^2} dV = \lim_{R\to\infty} \int_{||x||\le R} \frac{1}{1+||x||^2} dV? \tag{10.1.1}$$

Similarly, $||x||^{-1}$ is continuous on $B_1(0) \setminus \{0\}$ so by Theorem 7.7.4, $||x||^{-1}$ is integrable on $B_1(0) \setminus B_\varepsilon(0)$ for any $0 < \varepsilon < 1$. Can you therefore define

$$\int_{0<||x||<1} \frac{1}{||x||} dV = \lim_{\varepsilon\to 0^+} \int_{\varepsilon\le||x||<1} \frac{1}{||x||} dV? \tag{10.1.2}$$

The answer is not yet clear.

These are examples of improper integrals in $\mathbb{R}^n$. To define improper integrals more generally, you will need two basic properties illustrated by these cases.

10.1.1 Local integrability

First, to define an improper integral, you need the function to be integrable when restricted to any compact Jordan measurable subset of the region. Otherwise, as in (10.1.1) and (10.1.2) from Example 10.1.1, you cannot necessarily calculate a limit of "proper" integrals. This suggests a new definition.

Definition 10.1.2 Let $\Omega \subseteq \mathbb{R}^n$ be a set. A real-valued function f is **locally integrable** on Ω if f is integrable on every compact Jordan measurable subset of Ω.

Remark 10.1.3 Notice Ω does not need to be bounded and the function f does not need to be bounded. This choice is always intentional for improper integrals.

There are some standard ways to verify whether a function is locally integrable. The proofs tend to borrow from the many results about integration that you have already developed.

Lemma 10.1.4 Let $\Omega \subseteq \mathbb{R}^n$ be a set. If a real-valued function f is continuous on Ω, then f is locally integrable on Ω.

Proof. This is left as a short exercise. Use Theorem 7.7.4. ■

Lemma 10.1.5 Let $\Omega \subseteq \mathbb{R}^n$ be a Jordan measurable set. Let f be a bounded real-valued function on Ω. If f is integrable on Ω, then f is locally integrable on Ω.

Proof. This is left as an exercise. Use Cauchy-Schwarz and Definition 7.5.3. ■

You can swiftly apply both of these results in explicit examples.

Example 10.1.6 Consider three different scenarios.

- By Lemma 10.1.4, the function $\frac{1}{1+||x||^2}$ is locally integrable on $\mathbb{R}^n$ since $\frac{1}{1+||x||^2}$ is continuous on $\mathbb{R}^n$.
- By Lemma 10.1.4, the function $||x||^{-1}$ is locally integrable on $B_1(0)\setminus\{0\}$ since $||x||^{-1}$ is continuous on $B_1(0)\setminus\{0\}$. However, notice $||x||^{-1}$ is not integrable on $B_1(0)\setminus\{0\}$ because $||x||^{-1}$ is not bounded on $B_1(0)\setminus\{0\}$.
- By Lemma 10.1.5, the indicator function $\chi_{B_1(0)}$ is locally integrable on $[-1,1]^n$ since $\chi_{B_1(0)}$ is integrable on $[-1,1]^n$ from Definition 7.5.3.

Locally integrable functions satisfy some standard properties which you can verify.

Exercise 10.1.7 Let $\Omega \subseteq \mathbb{R}^n$ be a set. Fix $\lambda \in \mathbb{R}$. Show that if f and g are real-valued functions that are locally integrable on Ω, then $f+\lambda g$ and $f \cdot g$ are locally integrable on Ω.

Local integrability is the first step towards defining improper integrals.

10.1.2 Exhaustions

Second, you want to “fill up" or “exhaust" the region of integration Ω with compact Jordan measurable subsets. This allows you to leverage local integrability and create a sequence of approximations by integrating on subsets of Ω.

Definition 10.1.8 Let $\Omega \subseteq \mathbb{R}^n$ be a set. A sequence of compact Jordan measurable sets $\{\Omega_k\}_{k=1}^{\infty}$ is an **exhaustion** of Ω if $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$ and $\forall k \geq 1, \Omega_k \subseteq \Omega_{k+1}^{\circ}$.

Remark 10.1.9 Notice there are some topological subtleties of open sets and interiors. These details will be technically important, but do not worry about them for now.

Constructing exhaustions in explicit examples is usually straightforward.

Example 10.1.10 Here are a few examples of exhaustions.

- $\left\{[2+\frac{1}{k}, 7-\frac{1}{k}]\right\}_{k=1}^{\infty}$ is an exhaustion of the open interval $(2,7)$.
- $\left\{\overline{B_k(0)}\right\}_{k=1}^{\infty}$ is an exhaustion of $\mathbb{R}^n$.
- $\{[-k,k]^n\}_{k=1}^{\infty}$ is an exhaustion of $\mathbb{R}^n$.
- $\{\Omega_k\}_{k=1}^{\infty}$ where $\Omega_k = \{x \in \mathbb{R}^n : \frac{1}{k} \leq ||x|| \leq 1-\frac{1}{k}\}$ is an exhaustion of $\Omega = B_1(0)\setminus\{0\}$.

The definition of exhaustions poses a new problem..

Which sets in $\mathbb{R}^n$ can be exhausted by compact Jordan measurable sets?

This is surprisingly subtle and better reserved for a course in topology, so you will ignore it. You will only handle explicit examples like in Example 10.1.10. You can at least make one conclusion about sets in $\mathbb{R}^n$ that can be exhausted by compact Jordan measurable sets.

Lemma 10.1.11 Let $\Omega \subseteq \mathbb{R}^n$ be a set. If there exists an exhaustion of Ω by compact Jordan measurable sets $\{\Omega_k\}_{k=1}^{\infty}$, then Ω is open.

Proof. Since $\Omega_k^{\circ} \subseteq \Omega_k \subseteq \Omega_{k+1}^{\circ}$ for all $k \geq 1$, it follows that

$$\bigcup_{k=1}^{\infty} \Omega_k^{\circ} \subseteq \Omega = \bigcup_{k=1}^{\infty} \Omega_k \subseteq \bigcup_{k=1}^{\infty} \Omega_{k+1}^{\circ} = \bigcup_{k=1}^{\infty} \Omega_k^{\circ}.$$

The latter equality holds since $\Omega_1^{\circ} \subseteq \Omega_1 \subseteq \Omega_2^{\circ}$. From the above, it follows that Ω is the countable union of open sets $\{\Omega_k^{\circ}\}_{k=1}^{\infty}$ and hence Ω is open. ■

The concept of exhaustion is a new feature needed for higher dimensions. For $\mathbb{R}$, there is a topologically natural way to exhaust an interval like $(2,7)$ since you can only approach 2 from the right and 7 from the left. In $\mathbb{R}^n$, you can exhaust an open set in uncountably many ways and this poses serious issues.

Example 10.1.12 Define

$$\Omega = \{(x,y) \in \mathbb{R}^2 : x > 1, y > 1\}$$

and

$$f(x,y) = \frac{y-x}{x^2+y^2}.$$

Notice f is continuous on Ω and hence locally integrable on Ω by Lemma 10.1.4. Let

$$\Omega_k = [1+\frac{1}{k}, k] \times [1+\frac{1}{k}, k] \quad \text{and} \quad \Omega_k' = [1+\frac{1}{k}, k] \times [1+\frac{1}{k}, 4k]$$

for $k \geq 1$, so both $\{\Omega_k\}_{k=1}^\infty$ and $\{\Omega_k'\}_{k=1}^\infty$ are exhaustions of Ω. Here is the critical question:

Do the limits $\lim_{k\to\infty} \int_{\Omega_k} f\,dV$ *and* $\lim_{k\to\infty} \int_{\Omega_k'} f\,dV$ *both produce the same result?*

If an integral f over Ω can be defined via a limit, you should get the same value regardless of which regions you use to exhaust Ω. You can compute both limits directly in this example. Notice Ω_k is symmetric across the line $y = x$, that is, $(x,y) \in \Omega_k$ if and only if $(y,x) \in \Omega_k$. Since f is antisymmetric across the line $y = x$, i.e. $f(x,y) = -f(y,x)$, you can verify by Fubini's theorem and a change of variables that

$$\forall k \geq 1, \int_{\Omega_k} f\,dV = 0 \quad \text{so} \quad \lim_{k\to\infty} \int_{\Omega_k} f\,dV = 0.$$

However, Ω_k' is not symmetric across the line $y = x$. Since Ω_k' is the disjoint union of Ω_k and $[1+\frac{1}{k}, k] \times (k, 4k]$, and $f(x,y) \geq 0$ for $y \geq 0$, it follows by Fubini's theorem and monotonicity that

$$\begin{aligned}
\int_{\Omega_k'} f\,dV &= \int_{\Omega_k} f\,dV + \int_{1+1/k}^{k} \int_{k}^{4k} \frac{y-x}{x^2+y^2}\,dy\,dx \\
&\geq 0 + \int_{1+1/k}^{k} \int_{2k}^{4k} \frac{y-x}{x^2+y^2}\,dy\,dx \\
&\geq \int_{1+1/k}^{k} \int_{2k}^{4k} \frac{k}{k^2+16k^2}\,dy\,dx \\
&= \frac{2}{17}\big(k-1-\frac{1}{k}\big).
\end{aligned}$$

By the squeeze theorem for sequences, this implies that $\int_{\Omega_k'} f\,dV \to \infty$ as $k \to \infty$. Both limits have wildly different answers! Sadly, you cannot expect to sensibly define the integral of f over Ω.

The above example is really quite troubling.

10.1.3 Definition of improper integrals

To avoid the issue in Example 10.1.12, you can perform a classic mathematician's trick: create a definition that assumes this never happens. That is how you will define improper integrals.

Definition 10.1.13 Let $\Omega \subseteq \mathbb{R}^n$ be a set. Let $\{\Omega_k\}_{k=1}^{\infty}$ be an exhaustion of Ω by compact Jordan measurable sets. Let $f : \Omega \to \mathbb{R}$ be locally integrable. The **improper integral** of f on Ω is given by

$$\int_{\Omega} f\,dV = \lim_{k\to\infty} \int_{\Omega_k} f\,dV$$

provided the limit does not depend on the choice of exhaustion and, if so,

- the improper integral **converges** when the limit exists.
- the improper integral **diverges** when the limit does not exist.
- the improper integral **diverges to** ∞ when the limit is ∞.
- the improper integral **diverges to** $-\infty$ when the limit is $-\infty$.

If the limit depends on the choice of exhaustion, then the improper integral **diverges**.

Remark 10.1.14 As with the single variable calculus terminology, the case where the improper integral "diverges" includes the cases where it "diverges to ∞" or "diverges to $-\infty$".

Remark 10.1.15 The limit does not depend on the choice of exhaustion provided for any two exhaustions $\{\Omega_k\}_{k=1}^{\infty}$ and $\{\Omega'_k\}_{k=1}^{\infty}$ of Ω by compact Jordan measurable sets,

$$\lim_{k\to\infty} \int_{\Omega_k} f\,dV = \lim_{k\to\infty} \int_{\Omega'_k} f\,dV.$$

These limits may both equal a real number, both equal ∞, or both equal $-\infty$.

Now, before diving into any examples, notice there is an abuse of notation by writing

$$\int_{\Omega} f\,dV$$

to denote the improper integral of f on Ω (using limits and exhaustions) and also the "proper" integral of f on Ω (as in Chapter 7). You may assume the notation is consistent, but notice there is something to prove. If f is integrable on Ω, then you must check that the improper integral definition above should also be satisfied and yield the same value; this ensures the notation is consistent across both definitions. Below is a formal statement of this property.

Lemma 10.1.16 Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. If f is integrable on Ω, then the improper integral of f on Ω converges and its value is equal to the integral of f on Ω.

Proof. Omitted. You may assume this fact without proof. ■

If f is not integrable on Ω then its integral does not exist but the improper integral may exist. This is fine but you cannot apply properties of the integral directly to the improper integral; you must write everything as a limit. Keeping this distinction in mind, you can prove linearity for convergent improper integrals.

Lemma 10.1.17 (Linearity) Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be locally integrable. Fix $\lambda \in \mathbb{R}$. If the improper integrals $\int_\Omega f\,dV$ and $\int_\Omega g\,dV$ both converge, then the improper integral $\int_\Omega (f + \lambda g)\,dV$ converges and

$$\int_\Omega (f + \lambda g)\,dV = \int_\Omega f\,dV + \lambda \int_\Omega g\,dV.$$

Proof. This is left as an exercise. Use limit laws for sequences and linearity for integrals. ∎

Exercise 10.1.18 Formulate and prove versions of linearity for improper integrals diverging to ∞. Be careful with the possible choices for the scalars.

Overall, this settles the definition for improper integrals. This definition seems to handle the key obstacles, but it creates a new and very serious obstacle.

How can you compute limits of integrals of all possible exhaustions?

That seems foolish, and makes it impossible to verify Definition 10.1.13 in even the simplest examples. Luckily, there is a wonderful theorem for non-negative functions that you will explore in the next section.

Exercises for Section 10.1

Concepts and definitions

10.1.1 Determine which of the following functions below are locally integrable on the specified set Ω.

(a) $\Omega = \mathbb{R}^n \setminus \{0\}$ and $f(x) = \dfrac{1}{\|x\|^{1/2}}$.

(b) $\Omega = \mathbb{R}^n$ and $f(x) = \dfrac{1}{1+\|x\|^{1/3}}$.

(c) $\Omega = \mathbb{R}^2$ and $f(x,y) = \begin{cases} x & \text{if } x \geq y \\ -y & \text{if } x < y. \end{cases}$

(d) $\Omega = (0,1)$ and $f = \chi_{\mathbb{Q}}$.

(e) $\Omega = \mathbb{R}^n$ and $f(x) = e^{\|x\|}$

(f) $\Omega = B_1(0)$ and $f(x) = \dfrac{1}{1-\|x\|}$.

(g) $\Omega = \mathbb{R}^n$ and $f(x) = \begin{cases} \dfrac{1}{\|x\|} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$

(h) Ω Jordan measurable, f integrable on Ω.

10.1.2 Let f be locally integrable on the open set Ω. Fix two exhaustions $\{\Omega_k\}_{k=1}^\infty$ and $\{\Omega_k'\}_{k=1}^\infty$ of Ω. Determine which of the following statements are true and which are false.

(a) If $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV$ exists, then the improper integral $\displaystyle\int_\Omega f\,dV$ converges.

(b) If $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV$ does not exist, then the improper integral $\displaystyle\int_\Omega f\,dV$ diverges.

(c) If $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV = \lim_{k\to\infty}\int_{\Omega_k'} f\,dV$ exists, then the improper integral $\int_\Omega f\,dV$ converges.

(d) If the improper integral converges, then $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV = \lim_{k\to\infty}\int_{\Omega_k'} f\,dV$.

(e) If the limit value $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV$ is independent of the exhaustion $\{\Omega_k\}_{k=1}^\infty$, then the improper integral must converge.

Computations

10.1.3

(a) Create an exhaustion for the interval (a,b)

(b) Create two different exhaustions for $\mathbb{R}^n$.

(c) Create two different exhaustions for $\mathbb{R}^n \setminus \{0\}$.

(d) Create two different exhaustions for the open upper half plane, $\Omega = \{(x,y) \in \mathbb{R}^2 : y > 0\}$.

Proofs

10.1.4 Standard integral properties (such as linearity and monotonicity) can get quite complicated in the setting of improper integrals. Let f and g be locally integrable on the open set $\Omega \subseteq \mathbb{R}^n$. Prove or disprove each of the following statements.

(a) If $\displaystyle\int_\Omega f\,dV$ and $\displaystyle\int_\Omega g\,dV$ converge, then $\displaystyle\int_\Omega (f+g)dV$ converges and

$$\int_\Omega (f+g)dV = \int_\Omega f\,dV + \int_\Omega g\,dV.$$

(b) If $\displaystyle\int_\Omega (f+g)\,dV$ converges, then $\displaystyle\int_\Omega f\,dV$ and $\displaystyle\int_\Omega g\,dV$ converge and

$$\int_\Omega (f+g)\,dV = \int_\Omega f\,dV + \int_\Omega g\,dV.$$

10.1.5 Let $\Omega \subseteq \mathbb{R}^n$ be a set and let $f : \Omega \to \mathbb{R}$. Prove that if f is continuous on Ω, then f is locally integrable on Ω.

10.1.6 Let $\Omega \subseteq \mathbb{R}^n$ be a Jordan measurable set. Let $f : \Omega \to \mathbb{R}$ be bounded. Prove that if f is integrable on Ω, then f is locally integrable on Ω.

10.1.7 Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be locally integrable. Prove or disprove each of the following:

(a) If $\displaystyle\int_\Omega f\,dV = \infty$ and $\displaystyle\int_\Omega g\,dV = \infty$, then $\displaystyle\int_\Omega (f+g)\,dV = \infty$.

(b) If $\displaystyle\int_\Omega f\,dV$ diverges and $\displaystyle\int_\Omega g\,dV$ diverges, then $\displaystyle\int_\Omega (f+g)\,dV$ diverges.

(c) If $\displaystyle\int_\Omega f\,dV = \infty$ and $\displaystyle\int_\Omega g\,dV$ converges, then $\displaystyle\int_\Omega (f+g)\,dV = \infty$.

Applications and beyond

10.1.8 Let $\Omega \subseteq \mathbb{R}^n$ be open and Jordan measurable. Assume f is integrable on Ω. Sokka and Toph are confused about the relationship between the improper integral and the usual integral. Sokka says:

> *Since both the improper integral and the usual integral are defined in different ways, they may not equal so we shouldn't write them both as $\int_\Omega f\,dV$. It's notational abuse!*

Toph suggests:

> *But f is integrable on Ω. That means it doesn't matter which definition you use. It'll give you the same answer.*

Who is correct? Explain why.

10.2. Monotone convergence theorem

With a solid definition of improper integrals in higher dimensions, you want to proceed to prove all the theorems. However, the definition of improper integrals (Definition 10.1.13) seems impossible to verify because it requires you to check limits over all possible exhaustions. More precisely, let Ω be a set with an exhaustion $\{\Omega_k\}_{k=1}^{\infty}$ by compact Jordan measurable sets, and let $f : \Omega \to \mathbb{R}$ be locally integrable. For any exhaustion $\{\Omega_k\}_{k=1}^{\infty}$ of Ω by compact Jordan measurable sets, the improper integral of f on Ω is given by

$$\int_{\Omega} f\, dV := \lim_{k\to\infty} \int_{\Omega_k} f\, dV$$

provided the limit does not depend on the choice of exhaustion. In particular, it converges if the limits always exists and always equals the same value. A priori, this appears to be a ludicrous assumption but the monotone converge theorem shall save you from despair.

Theorem 10.2.1 (Monotone convergence theorem) Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f be a real-valued locally integrable function on Ω. If $f \geq 0$ on Ω then the improper integral $\displaystyle\int_{\Omega} f\, dV$ either converges or diverges to ∞.

Proof. Postponed to Section 10.2.3. ■

Informally speaking, the monotone convergence theorem says:

> *Improper integrals of non-negative functions never depend on the choice of exhaustion. That is, you can compute them any way you like.*

Non-negative integrands are quite common (e.g. mass density or probability density functions) so this is a wonderful outcome. It also leads to some natural questions.

> *How do you compute improper integrals with the monotone convergence theorem? What are classic examples of convergent and divergent improper integrals in $\mathbb{R}^n$? How do you prove the monotone convergence theorem?*

You will address these over the next few subsections.

10.2.1 Improper integrals of non-negative functions

The monotone convergence theorem allows you to directly compute an improper integral by definition because you only need to consider a *single* exhaustion. You can see how by exploring a couple of illustrative examples.

Example 10.2.2 Suppose you want to determine whether this improper integral converges:

$$\int_{||(x,y)||\geq 1} \frac{1}{x^2+y^2}\, dA.$$

Notice $\frac{1}{x^2+y^2}$ is continuous on $\Omega = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 > 1\}$ and hence locally integrable on Ω. Since $\frac{1}{x^2+y^2} \geq 0$, by the monotone convergence theorem, the improper integral $\int_{\Omega} \frac{1}{x^2+y^2}\, dA$ either exists or equals ∞. Hence,

$$\int_{\Omega} \frac{1}{x^2+y^2}\, dA = \lim_{k\to\infty} \int_{\Omega_k} \frac{1}{x^2+y^2}\, dA$$

for *any fixed* choice of exhaustion $\{\Omega_k\}_{k=1}^{\infty}$ of Ω. Take $\Omega_k = \{(x,y) \in \mathbb{R}^2 : (1+\frac{1}{k})^2 \leq x^2+y^2 \leq k^2\}$ for $k \geq 1$. You can verify that $\{\Omega_k\}_{k=1}^{\infty}$ is an exhaustion of Ω. Now, by Fubini's theorem and a change of variables to polar coordinates, you can check that

$$\int_{\Omega_k} \frac{1}{x^2+y^2} dA = \int_0^{2\pi} \int_{1+1/k}^{k} \frac{1}{r^2} \cdot r\,dr\,d\theta = 2\pi\big(\ln k - \ln(1+\tfrac{1}{k})\big).$$

Therefore,

$$\int_{\Omega} \frac{1}{x^2+y^2} dA = \lim_{k\to\infty} \int_{\Omega_k} \frac{1}{x^2+y^2} dA = \lim_{k\to\infty} 2\pi\big(\ln k - \ln(1+\tfrac{1}{k})\big) = \infty,$$

so the improper integral diverges.

The strategy above for unbounded sets applies equally to unbounded functions.

Example 10.2.3 Suppose you want to determine whether this improper integral converges:

$$\int_{0<||(x,y,z)||<1} \frac{1}{||(x,y,z)||} dV.$$

Set

$$\Omega = \{(x,y,z) \in \mathbb{R}^3 : 0 < x^2+y^2+z^2 < 1\}.$$

Notice that $\frac{1}{\sqrt{x^2+y^2+z^2}}$ is continuous on Ω and hence locally integrable on Ω. Since the integrand $(x^2+y^2+z^2)^{-1/2}$ is non-negative, it follows by the monotone convergence theorem (Theorem 10.2.1) that the improper integral $\int_\Omega \frac{1}{x^2+y^2+z^2} dV$ either exists or equals ∞. In other words,

$$\int_{\Omega} \frac{1}{||(x,y,z)||} dV = \lim_{k\to\infty} \int_{\Omega_k} \frac{1}{\sqrt{x^2+y^2+z^2}} dV,$$

for *any fixed* choice of exhaustion $\{\Omega_k\}_{k=1}^{\infty}$ of Ω. Take

$$\Omega_k = \{(x,y,z) \in \mathbb{R}^3 : \tfrac{1}{k^2} \leq x^2+y^2+z^2 \leq (1-\tfrac{1}{k})^2\}$$

for $k \geq 1$. You can verify that $\{\Omega_k\}_{k=1}^{\infty}$ is an exhaustion of Ω. Now, by Fubini's theorem and a change of variables to spherical coordinates, you can check that

$$\begin{aligned}\int_{\Omega_k} \frac{1}{\sqrt{x^2+y^2+z^2}} dV &= \int_0^{2\pi} \int_0^{\pi} \int_{1/k}^{1-1/k} \frac{1}{\rho} \cdot \rho^2 \sin\phi\, d\rho\, d\phi\, d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{\pi} \sin\phi\, d\phi \int_{1/k}^{1-1/k} \rho\, d\rho \\ &= 2\pi\Big((1-\frac{2}{k})\Big)\end{aligned}$$

Therefore,

$$\int_{\Omega} \frac{1}{\sqrt{x^2+y^2+z^2}} dV = \lim_{k\to\infty} 2\pi\Big(1-\frac{2}{k}\Big) = 2\pi,$$

so the improper integral converges and its value is 2π.

10.2.2 A family of improper integrals

With a basic toolkit for calculating improper integrals, you can study a classic family of improper integrals known as *p-integrals*.

Example 10.2.4 Recall from single variable calculus that

$$\int_1^\infty \frac{1}{x^p}dx \begin{cases} \text{converges if } p > 1 \\ \text{diverges to } \infty \text{ if } p \le 1 \end{cases} \qquad \text{and} \qquad \int_0^1 \frac{1}{x^p}dx \begin{cases} \text{diverges to } \infty \text{ if } p \ge 1 \\ \text{converges if } p < 1 \end{cases}$$

You can verify both of these statements by direct calculation.

This family can be generalized to two dimensions, but something changes.

Example 10.2.5 For which $p \in \mathbb{R}$, does the integral

$$\iint_{1<x^2+y^2<\infty} \frac{1}{||(x,y)||^p}dA$$

converge? The case $p = 2$ was formally considered in Example 10.2.2 and it barely diverged. If you repeat the argument there, you will find that the integral converges if $p > 2$ and diverges if $p \le 2$.

Notice that the "threshold" value $p = 2$ for p-integrals in $\mathbb{R}^2$ is different compared to $p = 1$ for p-integrals in $\mathbb{R}$. Why exactly does this occur? Of course, one justification is a formal calculation and proof. Another justification is heuristic. Over $\mathbb{R}$, the interval $B_r(0) = (-r, r)$ has length $2r$, so $f(x) = |x|^{-p}$ must decay faster than $\frac{1}{r} = \frac{1}{|x|}$ as $r = |x| \to \infty$ in order to converge. Over $\mathbb{R}^2$, the open disk $B_r((0,0))$ of radius r has area πr^2, so

$$f(x,y) = ||(x,y)||^{-p}$$

must decay faster than $\dfrac{1}{r^2} = \dfrac{1}{x^2+y^2}$ as $r = ||(x,y)|| \to \infty$ in order to converge.

Exercise 10.2.6 Heuristically explain why

$$\iint_{x^2+y^2<1} ||(x,y)||^{-p}dA$$

converges if $p < 2$ and diverges if $p \ge 2$. Then formally prove it.

You can generalize this family of improper integrals to any dimension.

Theorem 10.2.7 Fix $p \in \mathbb{R}$. For the given improper integrals in $\mathbb{R}^n$, one has

$$\int_{||x||>1} \frac{1}{||x||^p}dV \begin{cases} \text{converges if } p > n, \\ \text{diverges to } \infty \text{ if } p \le n. \end{cases}$$

$$\int_{0<||x||<1} \frac{1}{||x||^p}dV \begin{cases} \text{diverges to } \infty \text{ if } p \ge n, \\ \text{converges if } p < n. \end{cases}$$

Exercise 10.2.8 Prove Theorem 10.2.7 for $n=2$ using polar coordinates.

Exercise 10.2.9 Prove Theorem 10.2.7 for $n=3$ using spherical coordinates.

Proof. This is left as an exercise. Spherical coordinates can be generalized to n-dimensions. Namely, the map $g:\mathbb{R}^n\to\mathbb{R}^n$ defined by $(x_1,\dots,x_n)=g(r,\phi_1,\dots,\phi_{n-1})$, where

$$\begin{aligned}
x_1 &= r\cos\phi_1\\
x_2 &= r\sin\phi_1\cos\phi_2\\
x_3 &= r\sin\phi_1\sin\phi_2\cos\phi_3\\
&\ \ \vdots\\
x_{n-1} &= r\sin\phi_1\cdots\sin\phi_{n-2}\cos\phi_{n-1}\\
x_n &= r\sin\phi_1\cdots\sin\phi_{n-2}\sin\phi_{n-1}
\end{aligned}$$

is the n-dimensional spherical coordinate transformation. You may assume that the ball $B_R(0)$ of radius $R>0$ in $\mathbb{R}^n$ is parametrized by g with domain $[0,R]\times[0,\pi]^{n-2}\times[0,2\pi]$. You may also assume that

$$\det Dg(r,\phi_1,\cdots,\phi_{n-1}) = r^{n-1}\sin^{n-2}\phi_1\sin^{n-3}\phi_2\cdots\sin\phi_{n-2}.$$

Using these assumptions, you can finish the proof. ∎

This classic family will later guide your intuition for other improper integrals.

10.2.3 Proof of the monotone convergence theorem

The proof of the monotone convergence theorem (Theorem 10.2.1) needs a revolutionary theorem in topology and compact sets.

Theorem 10.2.10 (Heine-Borel) Let $A\subseteq\mathbb{R}^n$ be compact. Let $\{V_j\}_{j=1}^{\infty}$ be a sequence of open sets such that $V_j\subseteq V_{j+1}$ for $j\geq 1$. If $A\subseteq\bigcup_{j=1}^{\infty}V_j$, then there exists $k\in\mathbb{N}^+$ such that $A\subseteq V_k$.

Proof. Omitted. This goes beyond the scope of this text. See, for example, [6, Theorem I.8.10] or a course in topology. ∎

Equipped with Heine-Borel, you can embark on the main proof of Theorem 10.2.1.

Proof of monotone convergence theorem. Let $\{\Omega_k\}_{k=1}^{\infty}$ and $\{\Omega_k'\}_{k=1}^{\infty}$ be two exhaustions of Ω. Define the limits

$$I=\lim_{k\to\infty}\int_{\Omega_k} f\,dV,\qquad I'=\lim_{k\to\infty}\int_{\Omega_k'} f\,dV.$$

Since $f\geq 0$ on Ω and $\Omega_k\subseteq\Omega_{k+1}$, it follows that $0\leq \chi_{\Omega_k}f\leq\chi_{\Omega_{k+1}}f$ so by monotonicity of the integral,

$$0\leq\int_{\Omega_k} f\,dV\leq\int_{\Omega_{k+1}} f\,dV.$$

By the monotone convergence theorem for sequences, this increasing sequence converges if it is bounded above. Hence, either I exists or equals ∞. The same is true for I'. It suffices to prove that $I=I'$.

We claim for any $i \geq 1$, there exists $j = j(i)$ such that $\Omega'_i \subseteq \Omega_j$. Assuming the claim, it follows again by monotonicity that

$$\forall i \geq 1, \qquad 0 \leq \int_{\Omega'_i} f\,dV \leq \int_{\Omega_{j(i)}} f\,dV.$$

Recall $I' = \lim_{i\to\infty} \int_{\Omega'_i} f\,dV$ exists or equals infinity. Since I either exists or equals ∞, the limit along the subsequence $\{\int_{\Omega_{j(i)}} f\,dV\}_{i=1}^{\infty}$ must be the same so $\lim_{i\to\infty} \int_{\Omega_{j(i)}} f\,dV = I$. Hence, by monotonicity of the limit, the previous inequality implies

$$I' = \lim_{i\to\infty} \int_{\Omega'_i} f\,dV \leq \lim_{i\to\infty} \int_{\Omega_{j(i)}} f\,dV = I.$$

This holds by the squeeze theorem even if one of the limits is ∞. By symmetry, $I \leq I'$ so we may conclude $I = I'$ as desired. It remains to prove the claim.

Since $\{\Omega_k\}_{k=1}^{\infty}$ is an exhaustion of Ω, it follows by definition that $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$. As Ω_k is contained in the interior of Ω_{k+1}, you can verify that $\Omega = \bigcup_{k=1}^{\infty} \Omega_k^o$. Now, fix $i \geq 1$. Since $\{\Omega'_k\}_{k=1}^{\infty}$ is also an exhaustion of Ω, we have $\Omega'_i \subseteq \Omega = \bigcup_{k=1}^{\infty} \Omega_k^o$. Therefore, the compact set Ω'_i is covered by the open sets $\{\Omega_k^o\}_{k=1}^{\infty}$ which are increasing (i.e. $\Omega_k^o \subseteq \Omega_{k+1}^o$). By the Heine-Borel theorem, there exists $j \geq 1$ such that Ω'_i is contained in Ω_j^o and hence inside Ω_j. This proves the claim. ■

Equipped with the monotone convergence theorem, you can compute improper integrals by definition and you will next see how to establish great results with this fantastic tool.

Exercises for Section 10.2

Concepts and definitions

10.2.1 The monotone convergence theorem implies the choice of exhaustion does not matter for improper integrals with non-negative integrands. This saves you from the absurd task of checking every exhaustion!

Let f be a real-valued function on the open set $\Omega \subseteq \mathbb{R}^n$ and let $\{\Omega_k\}_{k=1}^{\infty}$ be a fixed exhaustion of Ω. Determine which of the following statements are true and which are false.

(a) If f is non-negative on Ω then $\displaystyle\int_{\Omega} f\,dV$ converges or diverges to ∞.

(b) If f is locally integrable and $f \geq 0$ on Ω, then $\displaystyle\int_{\Omega} f\,dV$ converges or diverges to ∞.

(c) If f is locally integrable and $f \geq 0$ on Ω and $\displaystyle\lim_{k\to\infty}\int_{\Omega_k} f\,dV$ exists, then $\displaystyle\int_{\Omega} f\,dV$ converges.

(d) If f is locally integrable and $f \geq 0$ on Ω, then $\displaystyle\int_{\Omega} f\,dV = \lim_{k\to\infty}\int_{\Omega_k} f\,dV$.

(e) If f is locally integrable and the improper integral $\displaystyle\int_{\Omega} f\,dV$ converges, then $f \geq 0$ on Ω.

Proofs

10.2.2 Determine whether the following improper integrals converge or diverge. If it converges, find its value. If it diverges, specify how it diverges.

(a) $\displaystyle\int_{1<\|(x,y)\|<\infty} \frac{1}{(x^2+y^2)^{\frac{3}{2}}}\,dA$

(b) $\displaystyle\int_{0<\|(x,y,z)\|<1} \frac{1}{(x^2+y^2+z^2)^2}\,dV$

10.2.3 Here is more practice in computing improper integrals. Determine whether the following improper integrals converge or diverge. If it converges, compute its value. If it diverges, specify how.

(a) $\displaystyle\int_{\mathbb{R}^2} \frac{1}{1+x^2+y^2}\,dA.$

(b) $\displaystyle\int_{\Omega} e^{-(x^2+y^2+z^2)^{3/2}}\,dV$ where $\Omega = \{(x,y,z) \in \mathbb{R}^3 : z > 0\}$.

10.2.4 Here you will study a classic family of integrals in $\mathbb{R}^2$.

(a) Determine for which $p \in \mathbb{R}$ the improper integral $\displaystyle\int_{x^2+y^2>1} \frac{1}{\|(x,y)\|^p}\,dA$ converges.

(b) Determine for which $p \in \mathbb{R}$ the improper integral $\displaystyle\int_{0<x^2+y^2<1} \frac{1}{\|(x,y)\|^p}\,dA$ converges.

10.2.5 Spherical coordinates can be generalized to n-dimensions via the transformation:

$$\begin{aligned}
x_1 &= r\cos\phi_1 \\
x_2 &= r\sin\phi_1\cos\phi_2 \\
x_3 &= r\sin\phi_1\sin\phi_2\cos\phi_3 \\
&\vdots \\
x_{n-1} &= r\sin\phi_1\cdots\sin\phi_{n-2}\cos\phi_{n-1} \\
x_n &= r\sin\phi_1\cdots\sin\phi_{n-2}\sin\phi_{n-1}
\end{aligned}$$

where $r \geq 0, 0 \leq \phi_1, \ldots, \phi_{n-2} \leq \pi, 0 \leq \phi_{n-1} \leq 2\pi$. Denote this by $(x_1, \ldots, x_n) = g(r, \phi_1, \ldots, \phi_{n-1})$. You may assume that

$$|\det Dg| = r^{n-1}\sin^{n-2}\phi_1\sin^{n-3}\phi_2\cdots\sin\phi_{n-2}.$$

Use generalized spherical coordinates in $\mathbb{R}^n$ to show that

$$\int_{0<\|x\|<1}\frac{1}{\|x\|}dV = \infty.$$

10.2.6 Determine for which $\beta \in \mathbb{R}$ the improper integral $\displaystyle\int_{\mathbb{R}^2} e^{-\beta(x^2+y^2)}dA$ converges.

10.3. Convergence tests

Thus far, you have determined the convergence of improper integrals by direct computation. You also required the integrands to be non-negative, due to the monotone convergence theorem.

Can you verify if an improper integral converges without direct computation? How do you analyze improper integrals of functions which are both positive and negative?

Good answers are much easier to establish since the ideas are almost identical to those for series or improper integrals in single variable calculus. Your main tool will still be the monotone convergence theorem, which you will apply in several neat situations.

10.3.1 Basic comparison test

Exactly as in single variable calculus, you have a basic comparison test.

Theorem 10.3.1 (Basic comparison test) Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f and g be real-valued locally integrable functions on Ω.

- If $0 \le f \le g$ on Ω and $\displaystyle\int_\Omega g\,dV$ converges, then $\displaystyle\int_\Omega f\,dV$ converges.
- If $0 \le f \le g$ on Ω and $\displaystyle\int_\Omega f\,dV$ diverges, then $\displaystyle\int_\Omega g\,dV$ diverges.

Proof. This is left as an exercise. It follows from the monotone convergence theorem. ∎

The basic comparison test is useful because you can determine convergence of improper integrals without any computation. Using your standard set of examples (Theorem 10.2.7), you can apply the basic comparison test.

Example 10.3.2 The n-dimensional Gaussian $e^{-||x||^2}$ is fundamental to probability. Does

$$I = \int_{||x||>1} e^{-||x||^2}\,dV$$

converge? From single variable calculus, you can verify that $\displaystyle\lim_{t\to\infty}\frac{t^{n+1}}{e^{t^2}} = 0$ so there exists a constant $C > 0$ such that $e^{-t^2} \le Ct^{-n-1}$ for $t \ge 1$. Thus, for $x \in \mathbb{R}^n$ with $||x|| \ge 1$,

$$0 \le e^{-||x||^2} \le C||x||^{-n-1}.$$

Both $e^{-||x||^2}$ and $||x||^{-n-1}$ are continuous on the open set $\Omega = \{x \in \mathbb{R}^n : ||x|| > 1\}$, and hence locally integrable on Ω. Thus, by the basic comparison test and Theorem 10.2.7, the improper integral I converges.

Comparing improper integrals may require breaking up the region of integration. This is geometrically natural but a proper justification requires careful use of the definitions.

Example 10.3.3 Does the integral

$$J = \int_{\mathbb{R}^n} e^{-||x||^2}\,dV$$

converge? Geometrically, you might want to say that

$$\int_{\mathbb{R}^n} e^{-||x||^2} dV = \int_{||x||\leq 1} e^{-||x||^2} dV + \int_{||x||>1} e^{-||x||^2} dV \tag{10.3.1}$$

and conclude, by the previous example, that the integral converges. But how does this follow from the definition of an improper integral? There's a lot of missing details.

Fix the exhaustion $\{\Omega_k\}_{k=2}^{\infty}$ of $\mathbb{R}^n$ given by $\Omega_k = \{x \in \mathbb{R}^n : ||x|| \leq k\}$. Notice Ω_k can be written as a disjoint union

$$\Omega_k = \{x \in \mathbb{R}^n : ||x|| \leq 1\} \cup \{x \in \mathbb{R}^n : 1 < ||x|| < 1 + \tfrac{1}{k}\} \cup \Omega_k'$$

where $\Omega_k' = \{x \in \mathbb{R}^n : 1 + \frac{1}{k} \leq ||x|| \leq k\}$. As $e^{-||x||^2}$ is non-negative and continuous on $\mathbb{R}^n$, the monotone convergence theorem (Theorem 10.2.1) and linearity of the integral (Theorem 7.7.11) imply that

$$\int_{\mathbb{R}^n} e^{-||x||^2} dV = \lim_{k\to\infty} \int_{\Omega_k} e^{-||x||^2} dV$$

$$= \lim_{k\to\infty} \Big(\int_{||x||\leq 1} e^{-||x||^2} dV + \int_{1<||x||<1+\frac{1}{k}} e^{-||x||^2} dV + \int_{\Omega_k'} e^{-||x||^2} dV \Big).$$

To apply the limit sum law and conclude (10.3.1), you must verify all three limits exist. The leftmost integral is a constant with respect to k, so its limit exists. Since $\{\Omega_k'\}_{k=2}^{\infty}$ is an exhaustion of the region $||x|| > 1$, the rightmost integral is the improper integral $\int_{||x||>1} e^{-||x||^2} dV$ which is convergent by the previous example. Thus, to justify (10.3.1), it remains to show that

$$\lim_{k\to\infty} \int_{1<||x||<1+\frac{1}{k}} e^{-||x||^2} dV = 0.$$

The integral is at most a constant times the volume of the region $1 < ||x|| < 1 + \frac{1}{k}$. You can verify that this volume tends to zero as $k \to \infty$ (e.g. by a direct computation with n-dimensional spherical coordinates) but, for simplicity, these last few details are omitted.

This completes the core tools for improper integrals of non-negative functions, namely the monotone convergence theorem, the family of p-integrals, and the basic comparison test.

10.3.2 Absolute convergence

How do you analyze improper integrals for functions which may be positive or negative? A similar issue occurs with infinite series of real numbers. In that case, there is a rather nice criterion of *absolute convergence*[1].

Let $\{a_n\}_n \subseteq \mathbb{R}$. *If* $\sum_n |a_n|$ *converges, then* $\sum_n a_n$ *converges.*

Building on this intuition and the analogy between series and integrals, you may ask:

Let f *be locally integrable on* Ω. *If* $\int_\Omega |f| dV$ *converges, then does* $\int_\Omega f\, dV$ *converge?*

[1] See this MAT137 video for a more detailed explanation.

This question is excellent, but there is a hidden assumption. Namely, it also assumes that $|f|$ is locally integrable in order for the improper integral $\int_\Omega |f|\,dV$ to be defined. You can prove, however, that this extra assumption is not actually needed.

Lemma 10.3.4 Let $\Omega \subseteq \mathbb{R}^n$. If f is locally integrable on Ω, then $|f|$ is locally integrable on Ω.

Proof. This is left as an exercise. Use the integral triangle inequality (Theorem 7.7.13). ∎

Your question, along with this lemma, suggests a natural definition.

Definition 10.3.5 Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f be a real-valued locally integrable on Ω. The improper integral $\int_\Omega f\,dV$ **absolutely converges**[2] if the improper integral $\int_\Omega |f|\,dV$ converges.

There are some benchmark examples for this definition.

Example 10.3.6 Let $f \geq 0$ be locally integrable on a set $\Omega \subseteq \mathbb{R}^n$ with an exhaustion by compact Jordan measurable sets. Since $f = |f|$, the improper integral $\int_\Omega f\,dV$ converges if and only if it absolutely converges. This scenario is silly but it is a good reminder of earlier definitions.

Example 10.3.7 The improper integral

$$\int_1^\infty \frac{\sin x}{x^2}\,dx$$

absolutely converges. You can prove this using the basic comparison test (Theorem 10.3.1) with $f(x) = \frac{|\sin x|}{x^2}$ with $g(x) = \frac{1}{x^2}$. The details are left as an exercise. Now, does the integral also converge? As you shall see from the upcoming Theorem 10.3.9, indeed it does.

Example 10.3.8 The improper integral

$$\int_1^\infty \frac{\sin x}{x}\,dx$$

converges but it is does not absolutely converge. The proof is fairly technical and better reserved for a course in analysis.

In analogy with infinite series, you can show absolute convergence implies convergence.

Theorem 10.3.9 (Absolute convergence test) Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f be a real-valued locally integrable function on Ω. If the improper integral $\int_\Omega |f|\,dV$ converges, then the improper integral $\int_\Omega f\,dV$ converges.

The monotone convergence theorem will, as usual, play a crucial role in its proof.

Proof. The clever idea is to split f into positive and negative parts. Define the non-negative functions

$$f^+(x) = \max\{f(x), 0\} \quad \text{and} \quad f^-(x) = \max\{-f(x), 0\}.$$

You can verify that f^+ and f^- are both locally integrable since f is locally integrable. As $|f| = f^+ + f^-$ and both functions are non-negative, it follows that $0 \leq f^+ \leq |f|$ and $0 \leq f^- \leq |f|$.

[2]You can also define conditional convergence but it is beyond the scope of this course. Those integrals have all the same horrible issues in higher dimensions as they do in one dimension. If you see one, run the other way!

Since $\int_\Omega |f| dV$ converges by assumption, the basic comparison test implies that both improper integrals

$$\int_\Omega f^+ dV \quad \text{and} \quad \int_\Omega f^- dV$$

converge. As $f = f^+ - f^-$, the linearity of convergent improper integrals (Lemma 10.1.17) implies $\int f\, dV$ converges. ■

Theorem 10.3.9 is your chief mechanism for verifying convergence of integrals with integrands that are both positive and negative.

Example 10.3.10 Consider the integral

$$I = \int_{||(x,y,z)||>1} \frac{xyz\sin(xy+z)}{(x^2+y^2+z^2)^5} dV.$$

Since sine is bounded in magnitude by 1 and $|x|,|y|,|z| \leq ||(x,y,z)||$, it follows that

$$0 \leq \left|\frac{xyz\sin(xy+z)}{(x^2+y^2+z^2)^5}\right| \leq \frac{|x||y||z|}{||(x,y,z)||^{10}} \leq \frac{1}{||(x,y,z)||^7}$$

These functions are continuous and hence locally integrable on the region $||(x,y,z)|| > 1$ by Lemma 10.1.4. By Theorem 10.2.7,

$$\int_{||(x,y,z)||>1} \frac{1}{||(x,y,z)||^7} dV$$

converges. The basic comparison test (Theorem 10.3.1) therefore implies that

$$\int_{||(x,y,z)||>1} \left|\frac{xyz\sin(xy+z)}{(x^2+y^2+z^2)^5}\right| dV$$

converges. By Theorem 10.3.9, I absolutely converges implies that I converges.

Absolutely convergent integrals behave nicely and satisfy the basic properties that you hope they would. Sadly, there are convergent integrals that do not behave so nicely.

Example 10.3.11 The converse to Theorem 10.3.9 is false. For example,

$$\int_{x^2+y^2>1} \frac{\sin(x^2+y^2)}{x^2+y^2} dA$$

converges but it does not absolutely converge. You can verify this by converting to polar coordinates. You will need that the single variable integral

$$\int_1^\infty \frac{\sin r}{r} dr$$

converges but not absolutely. Verifying the details of this example are optional but it is important that you are aware such examples exist in higher dimensions.

This concludes your brief foray into improper integrals. The definition was intricate, but carefully supported by the monotone convergence theorem. Most importantly, you can finally integrate over a large class of unbounded sets and unbounded functions.

Now, your odyssey through multivariable integral calculus has reached its end. You have constructed robust definitions of integrals, established powerful theorems on integrability and the Jordan measure, accumulated a wealth of applications for integration to many areas of science, distilled elegant computational techniques via Fubini's theorem, and discovered deep perspectives with change of variables, and extended improper integrals to higher dimensions. Give yourself a second to bask in this glory, because the final part of this book will transport all of your achievements to the magical land of vector calculus.

Exercises for Section 10.3

Concepts and definitions

10.3.1 Determine whether the following improper integrals converge or diverge.

(a) $\displaystyle\int_{|x|>1} \frac{1}{|x|^2}\,dx$

(b) $\displaystyle\iint_{\|x\|>1} \frac{1}{\|x\|^2}\,dA$

(c) $\displaystyle\iiint_{\|x\|>1} \frac{1}{\|x\|^2}\,dV$

(d) $\displaystyle\int_{0<|x|<1} \frac{1}{|x|^2}\,dx$

(e) $\displaystyle\iint_{0<\|x\|<1} \frac{1}{\|x\|^2}\,dA$

(f) $\displaystyle\iiint_{0<\|x\|<1} \frac{1}{\|x\|^2}\,dV$

(g) $\displaystyle\int_{|x|>1} \frac{1}{|x|^2+|x|^3}\,dx$

(h) $\displaystyle\iint_{\|x\|>1} \frac{1}{\|x\|^2+\|x\|^3}\,dA$

(i) $\displaystyle\iiint_{\|x\|>1} \frac{1}{\|x\|^2+\|x\|\|^3}\,dV$

10.3.2 Let f and g be real-valued locally integrable functions on the open set $\Omega \subseteq \mathbb{R}^n$. Consider the claims.

(A) If $\displaystyle\int_\Omega f\,dV$ converges then $\displaystyle\int_\Omega g\,dV$ converges.

(B) If $\displaystyle\int_\Omega f\,dV = \infty$ then $\displaystyle\int_\Omega g\,dV = \infty$.

(C) If $\displaystyle\int_\Omega g\,dV$ converges then $\displaystyle\int_\Omega f\,dV$ converges.

(D) If $\displaystyle\int_\Omega g\,dV = \infty$ then $\displaystyle\int_\Omega f\,dV = \infty$.

(a) Assume $f \leq g$ on Ω. Which of these claims can you conclude?

(b) Assume $0 \leq f \leq g$ on Ω. Which of these claims can you conclude?

10.3.3 Let f be a real-valued locally integrable function on the open set $\Omega \subseteq \mathbb{R}^n$.

(a) Assume $\displaystyle\int_\Omega |f|\,dV$ converges. Which of these must converge? Explain.

(A) $\displaystyle\int_\Omega f^+\,dV$

(B) $\displaystyle\int_\Omega f^-\,dV$

(C) $\displaystyle\int_\Omega f\,dV$

(b) Assume $\displaystyle\int_\Omega f^+\,dV$ and $\displaystyle\int f^-\,dV$ converge. Which of these must converge? Explain.

(A) $\displaystyle\int_\Omega |f|\,dV$

(B) $\displaystyle\int_\Omega f\,dV$

(c) Assume $\displaystyle\int_\Omega f\,dV$ converges. Which of these must converge? Explain.

(A) $\displaystyle\int_\Omega f^+\,dV$

(B) $\displaystyle\int_{\Omega} f^{-}\,dV$

(C) $\displaystyle\int_{\Omega} |f|\,dV$

Proofs

10.3.4 Fix $p \in \mathbb{R}$. By direct calculation, prove that

$$\iiint_{0<\|(x,y,z)\|<1} \frac{1}{\|(x,y,z)\|^p}\,dV \quad \begin{cases} \text{diverges to } \infty \text{ if } p \geq 3 \\ \text{converges if } p < 3 \end{cases}$$

10.3.5 Use the basic comparison test to determine whether the improper integral converges or diverges.

(a) $\displaystyle\iint_{1<x^2+y^2<\infty} \frac{2+e^{-x^2-y^2}}{x^2+y^2}\,dA.$

(b) $\displaystyle\iiint_{0<\|(x,y,z)\|<1} \frac{x^2y^2}{\|(x,y,z)\|^5}\,dV$

10.3.6 The basic comparison test is an essential tool for determining convergence.

Basic comparison test. *Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f and g be real-valued locally integrable functions on Ω.*

- *If $0 \leq f \leq g$ on Ω and $\displaystyle\int_{\Omega} g\,dV$ converges then $\displaystyle\int_{\Omega} f\,dV$ converges.*
- *If $0 \leq f \leq g$ on Ω and $\displaystyle\int_{\Omega} f\,dV$ diverges then $\displaystyle\int_{\Omega} g\,dV$ diverges.*

Here is an essentially correct proof of the first implication. Details are missing.

1. *Let $\{\Omega_k\}_{k=1}^{\infty}$ be an exhaustion of Ω.*
2. *Since $0 \leq f \leq g$ on Ω, it follows that $\displaystyle\int_{\Omega_k} f\,dV \leq \int_{\Omega_k} g\,dV.$*
3. *Thus, $\displaystyle\int_{\Omega} f\,dV = \lim_{k\to\infty} \int_{\Omega_k} f\,dV \leq \lim_{k\to\infty} \int_{\Omega_k} g\,dV = \int_{\Omega} g\,dV.$*
4. *Since $\displaystyle\int_{\Omega} g\,dV < \infty$ we get $\displaystyle\int_{\Omega} f\,dV < \infty.$*
5. *Hence $\displaystyle\int_{\Omega} f\,dV$ converges as needed.*

Two lines are missing important justifications. Identify those lines and add the details.

10.3.7 Show that the improper integral

$$\iiint_{0<\|(x,y,z)\|<1} \frac{\sin(x)\cos(y)}{\|(x,y,z)\|^{5/2}}\,dV$$

absolutely converges.

10.3.8 Some simple properties of the integral do not obviously hold for improper integrals without more work. Let $f(x) = \dfrac{1}{1+\|x\|^3}$. Here is a WRONG proof that $\displaystyle\iint_{\mathbb{R}^2} f\,dA$ converges.

> 1. $\displaystyle\iint_{\mathbb{R}^2} f\,dA = \iint_{\|x\|\le 1} f\,dA + \iint_{\|x\|>1} f\,dA.$
> 2. *The first integral exists since f is continuous on the Jordan measurable set $\overline{B_1(0)}$.*
> 3. *The second integral converges by the basic comparison test.*
> 4. *Hence* $\displaystyle\iint_{\mathbb{R}^2} f\,dA$ *must converge as well.*

(a) One line is missing a critical justification. Identify the line and briefly explain what needs to be done.

(b) Line 3 claims to use the basic comparison test but does not specify how. Suggest how to do so.

10.3.9 Most of our theorems deal with non-negative functions only. To deal with functions that assume positive and negative values, it suffices to restrict our attention to their absolute value.

> **Theorem.** *Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f be a real-valued locally integrable function on Ω. If the improper integral $\int_\Omega |f|\,dV$ converges, then the improper integral $\int_\Omega f\,dV$ converges.*

Here is a WRONG proof of this theorem:

> 1. *Fix an exhaustion $\{\Omega_k\}_{k=1}^{\infty}$ of Ω.*
> 2. *Since f is locally integrable on Ω, $|f|$ is locally integrable on Ω so* $\displaystyle\int_{\Omega_k} |f|\,dV$ *exists.*
> 3. *By linearity, $-|f|$ is integrable on Ω_k and* $\displaystyle\int_{\Omega_k} -|f|\,dV = -\int_{\Omega_k} |f|\,dV$.
> 4. *Hence, both* $\displaystyle\int_{\Omega_k} |f|\,dV$ *and* $\displaystyle\int_{\Omega_k} -|f|\,dV$ *converge as $k \to \infty$.*
> 5. *Since $-|f| \le f \le |f|$ on Ω, the basic comparison test implies that* $\displaystyle\int_{\Omega_k} f\,dV$ *converges at $k \to \infty$.*

One line in the above proof contains a fatal mistake. Identify that line and explain the fatal mistake.

10.3.10 Determine for which values of $\alpha \in \mathbb{R}$ the improper integral

$$\iint_{\mathbb{R}^2} e^{-\alpha(x^2+y^2)}\,dA$$

converges. If it converges, calculate its value.

10.3.11 Formally prove that

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}.$$

10.3.12 Determine for which values of $\beta \in \mathbb{R}$ the integral

$$\int_{||x||>1} e^{-\beta ||x||} dV$$

converges.

10.3.13 Determine for which values of $\beta \in \mathbb{R}$ the integral

$$\int_{\mathbb{R}^n} e^{-\beta ||x||} dV$$

converges.

10.3.14 Prove the basic comparison test for divergent improper integrals.

10.3.15 Show that the improper integral

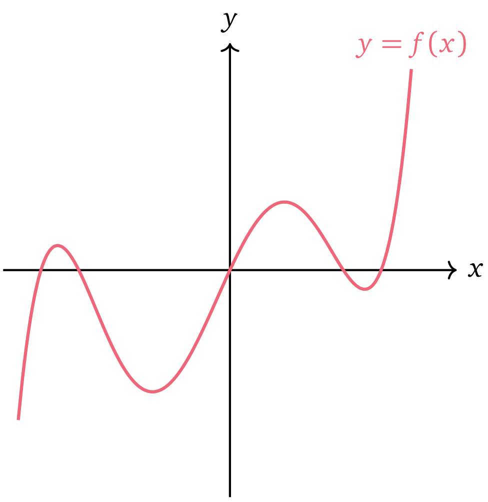
$$\int_{1<||(x,y,z,w)||<\infty} \frac{xyze^{-w^2}}{||(x,y,z,w)||^8} dV$$

absolutely converges.

10.3.16 Let $\Omega \subseteq \mathbb{R}^n$ be a set. Prove that if f is locally integrable on Ω, then $|f|$ is locally integrable on Ω.

Applications and beyond

10.3.17 The graph of $f : \mathbb{R} \to \mathbb{R}$ is given below.



(a) Sketch $|f|$ and $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$.

(b) How can you relate f, f^+, f^-, and $|f|$ with inequalities? You will need more than one inequality.

IV

Vector calculus

11. Integration on curves

The physics of a particle moving through a vector field is the origin of vector calculus. For instance, an electrical charge could be moving through a static electromagnetic field. A water molecule could be flowing through a steady river. This leads to some natural physical questions.

> *What is the velocity of the particle? What is its direction of motion? What is the length of its path? How much work is done on the particle by the force field? Does the work done depend on the path taken by the particle?*

The particle's path is described using a *curve*. To answer these questions, you will therefore need to develop a theory for calculus with curves. Luckily, you are quite prepared.

Differentiation in $\mathbb{R}^n$ was built up from derivatives in one variable (Section 3.1), which heavily relies on your knowledge from single variable calculus. This special case is designed for differentiating on curves and can be swiftly applied to describe a particle's motion (Section 1.1), including its velocity, acceleration, unit tangent, and Frenet frame. Thus, you have already acquired and employed the key tools for *differential* calculus with curves.

Integration in $\mathbb{R}^n$, however, was designed for n-dimensional solids. The basic building blocks were rectangles in $\mathbb{R}^n$ which have an intrinsic notion of volume and can approximate these solids. This theory led to the intuitive result that lower dimensional surfaces have zero volume since they can be covered by few thin rectangles. This suggests every curve in $\mathbb{R}^n$ should have zero volume. That is sensible, but you are still left with some unresolved issues.

> *What is the length of a curve? How can you integrate along a curve?*

These core questions will power this chapter's development for *integral* calculus with curves and ultimately lead to the definition of line integrals.

Afterwards, you will return to the original physics motivations and study the interplay between curves, vector fields, and gradients. This investigation will culminate with a beautifully potent generalization of the fundamental theorem of calculus: the fundamental theorem of line integrals in $\mathbb{R}^n$.

11.1. Curves

To develop calculus with curves, you must begin with the most essential question.

What is a curve?

This may seem innocuous, but you may recall it is quite troublesome. Indeed, your study of 1-dimensional smooth manifolds in Section 4.6 required a massive effort in Chapter 5 for *implicitly* described sets. Those deep tools are not immediately sufficient for your purposes here, because you will want a definition that can model the motion of a particle, namely *parametrically* described sets with *one* variable. You must therefore slightly modify the question.

What is a parametrized curve?

Some serious obstacles will again arise in trying to formulate a good definition. Fortunately, the implicit function theorem will again be your saviour and elegantly resolve this question with some straightforward criteria.

However, once you have done so, you will uncover an unsettling issue.

Will "equivalent" parametrizations of the same curve yield different answers?

As you develop calculus with curves in this chapter, this question will repeatedly arise, so you will need to be ready to face it. This introduces a preliminary question.

How are two parametrizations of the same curve considered equivalent?

By resolving these questions, you will lay the foundation for calculus with curves.

11.1.1 Simple regular parametrizations

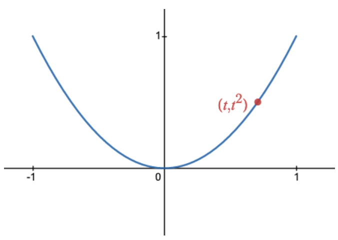
You must begin by deciding what is a valid parametrization of a curve. From your informal study of curves (Section 1.1), you may initially define the following.

Definition 11.1.1 A map $\gamma : [a, b] \to \mathbb{R}^n$ is a **(1-variable) parametrization** of a set $C \subseteq \mathbb{R}^n$ if $C = \gamma([a, b])$ and γ is continuous on $[a, b]$.

This first property appears sensible.

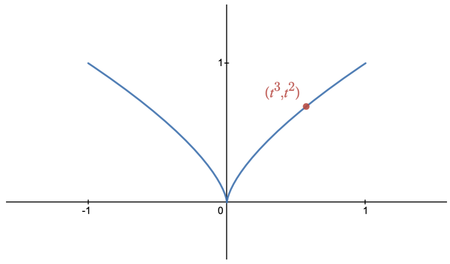
Example 11.1.2 The unit circle $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ has a parametrization $\gamma : [0, 2\pi] \to \mathbb{R}^2$ given by $\gamma(t) = (\cos t, \sin t)$. Watch this Desmos animation to view the particle trace C counterclockwise at constant speed, i.e. $||\gamma'(t)|| = 1$ for $t \in (0, 2\pi)$.

Example 11.1.3 The full parabola $P = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$ does not have a parametrization, since P is not compact and the image of any parametrization must be compact. However, the piece of the parabola $C = \{(x, y) \in \mathbb{R}^2 : y = x^2, -1 \leq x \leq 1\}$ has a parametrization $\gamma : [-1, 1] \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$. Watch this Desmos animation to view the particle trace out C at varying speed.



Unfortunately, Definition 11.1.1 alone is riddled with undesirable complications. Most egregiously, it allows sets to be parametrized which are *not* 1-dimensional smooth manifolds.

Example 11.1.4 The cusp $C = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3, -1 \le x \le 1\}$ is *not* a 1-dimensional smooth manifold, as seen in Examples 4.6.1 and 4.6.11 . A singularity occurs at the origin.



Nonetheless, the map $\gamma : [-1,1] \to \mathbb{R}^2$ given by $\gamma(t) = (t^3, t^2)$ is a parametrization of C. Indeed, the function γ is continuous and it is left as an exercise to verify that $\gamma([-1,1]) = C$ by directly showing the sets contain each other.

Although this example exposes a drawback of Definition 11.1.1, you can view this Desmos animation to watch the parametrization and make a critical observation: the particle at position momentarily *stops* at the origin! In other words, $\gamma(t) = (t^3, t^2)$ has *zero velocity* when $t = 0$. You can in fact verify that

$$\forall t \in (-1,1), \gamma'(t) = (0,0) \iff t = 0.$$

The cusp at the origin can occur precisely because $\gamma'(t)$ vanishes at $t = 0$.

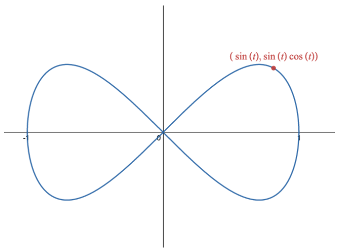
Example 11.1.4 has a nice physical interpretation: if you momentarily stop, then you can rigidly change direction in a discontinuous fashion. This suggests two extra properties.

Definition 11.1.5 A map $\gamma : [a,b] \to \mathbb{R}^n$ is **smooth**[1] if γ is continuous on $[a,b]$, γ is C^1 on (a,b), and its derivative γ' is bounded on (a,b).

Definition 11.1.6 A map $\gamma : [a,b] \to \mathbb{R}^n$ is **regular** if γ is differentiable on (a,b) and $\gamma'(t) \neq 0$ for $t \in (a,b)$.

Sadly, all of these properties combined is still not enough.

Example 11.1.7 The figure eight $C = \{(x,y) \in \mathbb{R}^2 : x^4 = x^2 - y^2\}$ is *not* a 1-dimensional smooth manifold, as seen in Examples 4.6.1 and 4.6.10. A singularity occurs at the origin.



Nonetheless, the map $\gamma : [0, 2\pi] \to \mathbb{R}^2$ given by $\gamma(t) = (\sin t, \sin t \cos t)$ is a smooth regular parametrization of C. The function γ is continuous and it is left as an exercise to verify that $\gamma([0,2\pi]) = C$. Moreover, you can verify that $\gamma'(t) = (\cos t, \cos^2 t - \sin^2 t) = (\cos t, \cos 2t)$ never vanishes on $(0, 2\pi)$.

[1]The notion of a "smooth" map depends a lot on the context, so there is no standard definition. Always check your source's definition. Our choice will be convenient for applying change of variables (Theorem 9.8.2) later.

Again, this example exposes a drawback to Definition 11.1.6, but you can view Desmos animation to watch the parametrization and make another critical observation: the particle visits the origin 3 times! Namely, $\gamma(t) = (0,0)$ at $t = 0, \pi$, and 2π. In other words, γ is not injective on its domain $[0, 2\pi]$. The matching endpoints $t = 0$ and $t = 2\pi$ is not so concerning, since the same occurs with the unit circle (Example 11.1.2). The vital issue is that the interior point $t = \pi$ matches other points in the domain $[0, 2\pi]$.

This last example suggests a fourth property.

Definition 11.1.8 A map $\gamma : [a,b] \to \mathbb{R}^n$ is **simple** if γ is injective except possibly $\gamma(a) = \gamma(b)$. In other words,

$$\forall s, t \in [a,b], \gamma(s) = \gamma(t) \implies s = t \text{ or } \{s,t\} = \{a,b\}.$$

Notice the endpoints $\gamma(a)$ and $\gamma(b)$ are permitted to be equal in order to allow parametrizations like the unit circle in Example 11.1.2. This warrants an extra property name.

Definition 11.1.9 A map $\gamma : [a,b] \to \mathbb{R}^n$ is **closed** if $\gamma(a) = \gamma(b)$.

These definitions finally eliminate the bad parametrizations and keep the good ones.

Example 11.1.10 The parametrization $(\cos t, \sin t)$ for $0 \leq t \leq 2\pi$ of the unit circle in Example 11.1.2 is smooth, simple, regular, and closed. The parametrization (t^3, t^2) for $-1 \leq t \leq 1$ of the cusp in Example 11.1.4 is smooth and simple, but not regular or closed. The parametrization $(\sin t, \sin t \cos t)$ for $0 \leq t \leq 2\pi$ of the figure eight in Example 11.1.7 is smooth, regular, and closed, but not simple.

These examples reveal the principal theorem on parametrized curves.

Theorem 11.1.11 If $\gamma : [a,b] \to \mathbb{R}^n$ is a smooth regular simple parametrization of a set $C \subseteq \mathbb{R}^n$, then the set C is a 1-dimensional smooth manifold at $\gamma(c)$ for every $c \in (a,b)$.

Remark 11.1.12 This can be viewed as the parametric analogue of the relationship between the gradient and sets in implicit form Theorem 5.5.3.

Proof. Omitted. This follows from the implicit function theorem and compactness. See a key step, for example, in Theorem 2 of [13, Section 3.2] or Theorem 3.13 of [11]. ■

Be careful that this criterion is *sufficient* but not necessary.

Example 11.1.13 The unit circle $C = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ can be parametrized in many ways. For instance, $\gamma(t) = (\cos t, \sin t)$ for $0 \leq t \leq 2\pi$ is a smooth simple regular parametrization so, by Theorem 11.1.11, C is a 1-dimensional smooth manifold at $\gamma(t)$ for $0 < t < 2\pi$ and hence at every $p \in C$ except possibly $p = (1,0)$. By choosing another smooth simple regular parametrization, you can conclude that C is a 1-dimensional smooth manifold, including at $(1,0)$. This is left as an exercise.

Example 11.1.14 The parametrizations of cusp in Example 11.1.4 and figure eight in Example 11.1.7 are not smooth simple regular parametrizations, so you cannot apply Theorem 11.1.11. In fact, neither curve has a simple smooth regular parametrization but this claim is not easy to prove with your available tools. While you should be able to identify when the existence of such parametrizations appear impossible, you will not need to prove such claims. This issue of non-existence is beyond the scope of this text.

11.1.2 Curves and piecewise curves

You have achieved your first goal of identifying valid parametrizations and Theorem 11.1.11 yields a straightforward criteria for verifying whether a parametrization defines a 1-dimensional smooth manifold. This paves the way for defining curves.

Definition 11.1.15 A set $C \subseteq \mathbb{R}^n$ is a **curve** if there exists a smooth simple regular parametrization of C. The curve is also **closed** if the parametrization is closed.

Remark 11.1.16 The word "curve" does not have a standard definition, so beware when you venture into other sources. This choice[2] is made for convenience because, for the remainder of vector calculus, you will be satisfied with studying these types of curves.

Example 11.1.17 By Example 11.1.13, the unit circle is a closed curve. By Example 11.1.14 the cusp and figure eight are not curves.

Example 11.1.18 Any line segment in $\mathbb{R}^n$ is a curve. This is left as an exercise.

Theorem 11.1.11 can be rephrased into the following corollary.

Corollary 11.1.19 Every curve is a 1-dimensional smooth manifold everywhere except possibly two points.

Sometimes you may want to consider sets which can be broken up into pieces.

Definition 11.1.20 A set $C \subseteq \mathbb{R}^n$ is a **piecewise curve** if there exists a map $\gamma : [a,b] \to \mathbb{R}^n$ and a partition $\{t_0, t_1, \dots, t_k\}$ of $[a,b]$ such that γ is a parametrization of C, and its restriction $\gamma|_{[t_{i-1},t_i]}$ is a smooth simple regular map for $i \in \{1,\dots,k\}$.

This expands the class of curves under consideration to many natural possibilities.

Example 11.1.21 The cusp $C = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3, -1 \le x \le 1\}$ in Example 11.1.4 is not a curve, but it is a piecewise curve. Define $\gamma : [-1,1] \to \mathbb{R}^2$ by $\gamma(t) = (t^3, t^2)$, so you can verify that γ is a 1-variable parametriation of C. Moreover, the partition $\{-1,0,1\}$ of $[-1,1]$ defines the restrictions $\gamma_1 := \gamma|_{[-1,0]}$ and $\gamma_2 := \gamma|_{[0,1]}$. You can prove γ_1 and γ_2 are smooth simple regular maps. Thus, C is a piecewise curve.

Notice that γ_1 and γ_2 are smooth simple regular parametrizations of the curves

$$C_1 = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3, -1 \le x \le 0\}, \qquad C_2 = \{(x,y) \in \mathbb{R}^2 : x^2 = y^3, 0 \le x \le 1\}.$$

respectively, and $C = C_1 \cup C_2$.

Example 11.1.22 The figure eight in Example 11.1.7 and the unit square $\partial[0,1]^2$ are both piecewise curves. These are left as exercises to verify. It is a bit tedious but worthwhile to check at least one of them.

Statements for curves almost always have very similar statements for piecewise curves, but the precise distinctions are often tedious to write. This observation is especially true for proofs involving curves compared to piecewise curves. Thus, for the sake of simplicity, many future theorems and definitions will be stated only for *curves* but you may freely apply all of them for *piecewise curves* with suitable modifications. In any case, these definitions are an excellent starting point for performing calculus with curves.

[2] Definition 11.1.15 predates the modern definition of a curve, which you can see in a course on differential geometry, or a course on curves and surfaces.

11.1.3 Reparametrizations and orientation

Definition 11.1.15 has one lingering predicament. A curve is defined by the *existence* of a parametrization. You will not want different choices of parametrizations to lead to different conclusions[3]. This creates one last question.

How are two parametrizations considered equivalent?

The choice of parametrization should not affect your future definitions with curves, so you must rigorously define this equivalence.

Definition 11.1.23 Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be smooth simple regular parametrizations of a set $C \subseteq \mathbb{R}^n$. The map γ_1 is a **reparametrization** of γ_2 if there exists a continuous invertible $\varphi : [a,b] \to [c,d]$ such that φ is C^1 on (a,b), its derivative φ' is bounded and never zero, and $\gamma_1 = \gamma_2 \circ \varphi$.

- If $\varphi' > 0$ on (a,b), then γ_1 has the **same orientation** as γ_2.
- If $\varphi' < 0$ on (a,b), then γ_1 has the **opposite orientation** as γ_2.

Remark 11.1.24 Lemma 9.8.8 implies that restriction $\varphi|_{(a,b)} : (a,b) \to (c,d)$ is a diffeomorphism.

As usual, the unit circle possesses all the crucial behaviour.

Example 11.1.25 The parametrization $\gamma(t) = (\cos t, \sin t)$ for $0 \le t \le 2\pi$ is smooth, simple, and regular. It traces out the circle counterclockwise exactly once. Consider four other parametrizations tracing out the circle in $\mathbb{R}^2$.

- $\gamma_1(t) = (\cos 2t, \sin 2t)$ for $0 \le t \le \pi$ is a reparametrization of γ with the same orientation. You can verify Definition 11.1.23 with $\gamma_1 = \gamma \circ \varphi$ and $\varphi(t) = 2t$.
- $\gamma_2(t) = (\cos 2t, \sin 2t)$ for $0 \le t \le 2\pi$ is not a reparametrization of γ, because it is not simple. It traces the circle twice.
- $\gamma_3(t) = (\cos t, -\sin t)$ for $0 \le t \le 2\pi$ is a reparametrization of γ with the opposite orientation. You can verify Definition 11.1.23 with $\gamma_3 = \gamma \circ \varphi$ and $\varphi(t) = -t$.
- $\gamma_4(t) = (\cos t, \sin t)$ for $-\pi \le t \le \pi$ is not a reparametrization of γ. It is simple and regular, but the endpoints differ. For instance, Definition 11.1.23 requires that $\gamma_4(\pi) = (-1,0)$ is equal to either $\gamma(0)$ or $\gamma(2\pi)$, but that is not the case.

The property of reparametrization is symmetric, transitive, and reflexive.

Lemma 11.1.26 Let $\gamma_1 : [a,b] \to \mathbb{R}^n, \gamma_2 : [c,d] \to \mathbb{R}^n, \gamma_3 : [e,f] \to \mathbb{R}^n$ be smooth simple regular parametrizations of a set $C \subseteq \mathbb{R}^n$. All of the following hold:

(a) (**Reflexive**) γ_1 is a reparametrization of itself.
(b) (**Symmetry**) If γ_1 is a reparametrization of γ_2, then γ_2 is a reparametrization of γ_1.
(c) (**Transitive**) If γ_1 is a reparametrization of γ_2 and γ_2 is a reparametrization of γ_3, then γ_1 is a reparametrization of γ_3.

Proof. These are left as exercises. Use properties of inverse functions and the chain rule. ■

These three properties allow you to think of a curve[4] using any "equivalent" smooth simple regular parametrization. Orientation will be studied more closely in a later section.

[3] This parallels the relationship between integral calculus and coordinate systems. Different ways of chopping a solid should lead to the same value for its volume. Indeed, the change of variables theorem miraculously proved that your answer did not depend on the choice of coordinates.

[4] The formal terminology is *equivalence class* of simple regular parametrizations. For more on equivalence classes, see a course on mathematical proofs. These same ideas commonly appear in algebra or number theory.

Through a careful scrutiny of many examples, you have brilliantly identified smooth simple regular parametrizations as the ideal way to describe curves. You will omit the adjectives "smooth simple regular" because you will exclusively study these parametrizations for the entirety of vector calculus.

*A **parametrization** will henceforth be a smooth simple regular parametrization.*

This selection will act as the perfect model for motion of particles in $\mathbb{R}^n$, because Theorem 11.1.11 guarantees that the tangent space of these curves are 1-dimensional subspaces of $\mathbb{R}^n$. This creates a gateway for approximating *curves* by *line segments*, and that is precisely how all of your existing calculus tools will thrive.

Exercises for Section 11.1

Concepts and definitions

11.1.1 Let $S = \{(x, y) \in \mathbb{R}^2 : y = x^2, 0 \le x \le 1\}$. Define the following six maps.

- $\gamma_A(t) = (t, t^2)$ for $0 \le t \le 1$.
- $\gamma_B(t) = (t, t^2)$ for $-1 \le t \le 1$.
- $\gamma_C(t) = (t^2, t^4)$ for $0 \le t \le 1$.
- $\gamma_D(t) = (t^2, t^4)$ for $-1 \le t \le 0$.
- $\gamma_E(t) = (t^2, t^4)$ for $-1 \le t \le 1$.
- $\gamma_F(t) = (t^{1/2}, t)$ for $0 \le t \le 1$.

(a) Which of these maps are parametrizations of S?

(b) Which of these maps are smooth?

(c) Which of these maps are regular?

(d) Which of these maps are simple?

(e) Which of these maps are closed?

11.1.2 Let $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Define the following five maps.

- $\gamma_A(t) = (\cos t, \sin t)$ for $0 \le t \le 2\pi$.
- $\gamma_B(t) = (\cos(t^2), \sin(t^2))$ for $0 \le t \le 2\pi$.
- $\gamma_C(t) = (\cos(2t), \sin(2t))$ for $0 \le t \le \pi$.
- $\gamma_D(t) = (\sin t, \cos t)$ for $0 \le t \le 2\pi$.
- $\gamma_E(t) = (\sin t, \cos t)$ for $\frac{\pi}{2} \le t \le \frac{5\pi}{2}$.

Determine each of the following.

(a) Which of these maps are parametrizations of S?

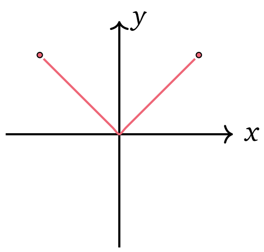
(b) Which of these maps are smooth?

(c) Which of these maps are regular?

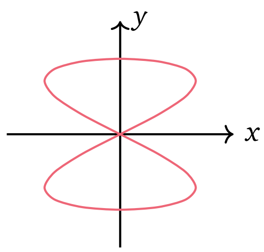
(d) Which of these maps are simple?

(e) Which of these maps are closed?

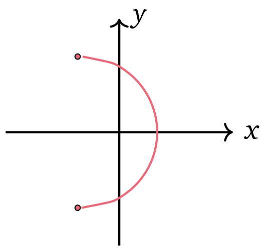
11.1.3 For each set $C \subseteq \mathbb{R}^2$, determine if C is a curve, a piecewise curve, both, or neither.



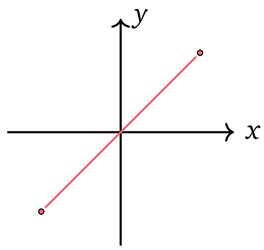
(a)



(c)



(b)



(d)

11.1.4 Define $\gamma : [-1,1] \to \mathbb{R}^2$ by $\gamma(t) = (t^3, t^3)$. Define the set $C = \gamma([-1,1]) \subseteq \mathbb{R}^2$.

(a) Is γ a smooth simple regular parametrization of C?

(b) Is C a curve?

11.1.5 Let $C \subseteq \mathbb{R}^n$. Determine which of the following are true or false.
If true, briefly justify. If false, state a counterexample.

(a) If there exists a smooth simple regular parametrization of C, then C is a curve.

(b) If C is a curve, then C is a piecewise curve.

(c) If C is a piecewise curve, then C is a curve.

(d) If C is a curve, then C is compact.

(e) If C is a curve, then C is a 1-dimensional smooth manifold.

(f) If C is a 1-dimensional smooth manifold, then C is a curve.

11.1.6 The map $\gamma(t) = (t, t^2)$ for $0 \le t \le 2$ parametrizes the path along the parabola from $(0,0)$ to $(2,4)$. For each parametrization below, determine if it is a reparametrization of γ. If so, determine whether it has the same orientation or opposite orientation.

(a) $\gamma_1(t) = (2t, 4t^2)$ for $0 \le t \le 1$.

(b) $\gamma_2(t) = (t, t^2)$ for $-2 \le t \le 0$.

(c) $\gamma_3(t) = (2-t, (2-t)^2)$ for $0 \le t \le 2$.

(d) $\gamma_4(t) = (2\sin t, 4\sin^2 t)$ for $0 \le t \le \pi$.

(e) $\gamma_5(t) = (2\sin t, 4 - 4\cos^2 t)$ for $0 \le t \le \pi/2$.

Proofs

11.1.7 Smooth simple regular parametrizations are fantastic because of Theorem 11.1.11. In other words, these can be used to prove that a set is locally the graph of a 1-variable function. By analyzing some incorrect arguments, you will discover some intricacies of this theorem.

(a) Jinan attempts to prove that the cusp in $\mathbb{R}^2$ is not a 1-dimensional smooth manifold.

> 1. *Let $C = \{(x,y) \in \mathbb{R}^2 : x^3 = y^2, -1 \le x \le 1\}$.*
> 2. *Define $\gamma : [-1,1] \to \mathbb{R}^2$ by $\gamma(t) = (t^2, t^3)$.*
> 3. *Notice γ parametrizes C since $\gamma([-1,1]) = C$ and γ is continuous.*
> 4. *However, γ is not regular since $\gamma'(t) = (2t, 3t^2)$ implying $\gamma'(0) = (0,0)$.*
> 5. *Thus, by Theorem 11.1.11, C is not a 1-dimensional smooth manifold.*

Aside from missing details, Jinan makes a mistake in one line. Identify the flaw and briefly explain how to fix the argument. Do not fix it.

(b) Cameron attempts to prove that the unit circle in $\mathbb{R}^2$ is a 1-dimensional smooth manifold.

1. *Let* $C = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.
2. *Define* $\gamma : [0,2\pi] \to \mathbb{R}^2$ *by* $\gamma(t) = (\cos t, \sin t)$*, so* γ *is* C^1 *on* $(0,2\pi)$.
3. *Notice* $\gamma([0,2\pi]) = C$ *and* γ *is continuous.*
4. *As* $\gamma'(t) = (-\sin t, \cos t)$*, we have that* $||\gamma'(t)||^2 = \sin^2 t + \cos^2 t = 1$ *so* $\gamma'(t)$ *is bounded and never vanishes.*
5. *Finally,* γ *is injective except that* $\gamma(0) = (1,0) = \gamma(2\pi)$.
6. *Thus,* γ *is a smooth simple regular parametrization of* C.
7. *By Theorem 11.1.11,* C *is a 1-dimensional smooth manifold.*

Aside from missing details, Cameron makes a subtle mistake in one line. Identify the flaw and briefly explain how to fix the argument. Do not fix it.

11.1.8 Define $\gamma : [0,4\pi] \to \mathbb{R}^3$ by $\gamma(t) = (\cos t, \sin t, t)$. Show that the helix $C = \gamma([0,4\pi])$ is a curve.

11.1.9 Fix distinct $p, q \in \mathbb{R}^n$. Prove that the compact line segment C from p to q is a curve.

11.1.10 Define $C = \partial\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, y \geq 0\}$. Prove that C is a piecewise curve.

11.1.11 Show that the lemniscate $C = \{(x,y) \in \mathbb{R}^2 : x^4 = x^2 - y^2\}$ is a piecewise curve.

11.1.12 The definitions of reparametrization and orientation are a mouthful (Definition 11.1.23). To digest it, you can start with some simple examples.

(a) Let $\gamma : [a,b] \to \mathbb{R}^n$ be a smooth simple regular parametrization of a curve $C \subseteq \mathbb{R}^n$. Prove that γ is a reparametrization of itself with the same orientation.

(b) Let $C = \{(x,y) \in \mathbb{R}^2 : y = x^2, 0 \leq x \leq 2\}$. You may assume without proof that

$$\gamma_1(t) = (2-t, (2-t)^2) \quad \text{for } 0 \leq t \leq 2 \qquad \text{and} \qquad \gamma_2(t) = (2t, 4t^2) \quad \text{for } 0 \leq t \leq 1$$

are smooth simple regular parametrizations of C. Prove γ_1 is a reparametrization of γ_2 with opposite orientation.

11.1.13 Let $\gamma_1 : [a,b] \to \mathbb{R}^n, \gamma_2 : [c,d] \to \mathbb{R}^n, \gamma_3 : [e,f] \to \mathbb{R}^n$ be smooth simple regular parametrizations of a set $C \subseteq \mathbb{R}^n$. Prove that if γ_1 is a reparametrization of γ_2 and γ_2 is a reparametrization of γ_3, then γ_1 is a reparametrization of γ_3.

11.1.14 Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be simple regular parametrizations of a set $C \subseteq \mathbb{R}^n$. Prove that if γ_1 is a reparametrization of γ_2, then γ_2 is a reparametrization of γ_1.

Applications and beyond

11.1.15 The properties of symmetry and transivity for reparametrizations can be refined to include orientations. You will formulate these conjectures here.

(a) Fill in the blanks.

Conjecture A. *Let* $\gamma_1 : [a,b] \to \mathbb{R}^n, \gamma_2 : [c,d] \to \mathbb{R}^n$ *be smooth simple regular parametrizations of a set* $C \subseteq \mathbb{R}^n$.
(i) If γ_1 *is a reparametrization of* γ_2 *with same orientation, then* γ_2 *is a reparametrization of* γ_1 *with* ______.

(ii) If γ_1 *is a reparametrization of* γ_2 *with opposite orientation, then* γ_2 *is a reparametrization of* γ_1 *with* ______.

(b) Fill in the blanks.

> **Conjecture B.** *Let* $\gamma_1 : [a,b] \to \mathbb{R}^n, \gamma_2 : [c,d] \to \mathbb{R}^n, \gamma_3 : [e,f] \to \mathbb{R}^n$ *be smooth simple regular parametrizations of* $C \subseteq \mathbb{R}^n$.
>
> *(i) If* γ_1 *is a reparametrization of* γ_2 *with same orientation and* γ_2 *is a reparametrization of* γ_3 *with same orientation, then* γ_1 *is a reparametrization of* γ_3 *with* ______.
>
> *(ii) If* γ_1 *is a reparametrization of* γ_2 *with opposite orientation and* γ_2 *is a reparametrization of* γ_3 *with same orientation, then* γ_1 *is a reparametrization of* γ_3 *with* ______.
>
> *(iii) If* γ_1 *is a reparametrization of* γ_2 *with same orientation and* γ_2 *is a reparametrization of* γ_3 *with opposite orientation, then* γ_1 *is a reparametrization of* γ_3 *with* ______.
>
> *(iv) If* γ_1 *is a reparametrization of* γ_2 *with opposite orientation and* γ_2 *is a reparametrization of* γ_3 *with opposite orientation, then* γ_1 *is a reparametrization of* γ_3 *with* ______.

(c) Prove Conjecture B(iv). *Hint:* You already did most of it in a previous problem.

11.2. Arc length

You can begin to develop integral calculus for curves. The first question is fundamental.

What is the length of a curve in $\mathbb{R}^n$?

You will quickly create its definition by relying purely on physical intuition[5], and you will later heuristically derive its integral definition again by chopping, estimating, and refining. This fundamental question can be generalized to permit a curve with varying density.

What is the mass of a wire in $\mathbb{R}^n$?

This physical problem is begging you to define integrals of scalar functions over curves. The basic ideas and derivations will repeat here and many times in other places. These repetitive arguments can be tedious to do rigorously, so you will learn to leverage infinitesimals and build these formulas by informal geometric explanations.

11.2.1 Definition and invariance

The length of a curve relies a common physical principle from single variable calculus.

The distance a particle travels is the integral of its speed.

This motivates the definition of arc length in higher dimensions.

Definition 11.2.1 The **arc length** (or **length**) of a curve $C \subseteq \mathbb{R}^n$ is defined as

$$\ell(C) = \int_a^b ||\gamma'(t)||dt,$$

where $\gamma : [a, b] \to \mathbb{R}^n$ is a parametrization of C.

A rigorous derivation of this definition involves polygonal approximations to a curve; this is postponed to a later subsection. Right now, you have a more pressing concern. The definition of arc length seems to depend on the choice of parametrization for the curve.

Does the length depend on the choice of parametrization?

In mathematical terminology, is length well-defined? The length of a curve should not depend on how it is parametrized and indeed it does not.

Theorem 11.2.2 (Invariance of arc length) Let $C \subseteq \mathbb{R}^n$ be a curve. Let $\gamma_1 : [a, b] \to \mathbb{R}^n$ and $\gamma_2 : [c, d] \to \mathbb{R}^n$ be parametrizations of C. If γ_1 is a reparametrization of γ_2, then

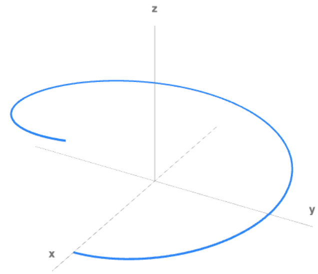
$$\int_a^b ||\gamma_1'(t)||dt = \int_c^d ||\gamma_2'(t)||dt.$$

Proof. This is left as an exercise. Use chain rule and integration by substitution. ∎

Arc length is therefore well-defined, i.e. independent of the choice of parametrization. Now, computing arc length is often a very messy or impossible calculation due to the square root function. The next example is one of the few that work out so simply.

[5] Depending on your background with physics, this feature may be a delight or may be uncomfortable. Vector calculus will repeatedly utilize physical intuition to derive formulas, so do not shy away from this practice. Fortunately, you only need your everyday experiences – not a university course in physics – to build this skill.

Example 11.2.3 The helix C is parametrized $\gamma(t) = (5\cos t, 5\sin t, 2t)$ for $0 \le t \le 2\pi$. You can verify by definition that γ is a smooth simple regular parameterization.



By definition, its length is equal to

$$\ell(C) = \int_0^{2\pi} ||\gamma'(t)|| dt = \int_0^{2\pi} ||(-5\sin t, 5\cos t, 2)|| dt = \int_0^{2\pi} \sqrt{5^2 + 2^2} dt = 2\pi\sqrt{29}.$$

This establishes the basics for arc length, so you can dig a little deeper.

11.2.2 Arc length parametrization

Physically speaking, arc length is equivalent to asking for the total distance travelled. You may be interested in slightly more refined time-dependent information.

How much distance did you travel as a function of time?

The answer is a minor but important variant on Definition 11.2.1.

Definition 11.2.4 Let $\gamma : [a, b] \to \mathbb{R}^n$ be a parametrization of a curve $C \subseteq \mathbb{R}^n$. The **arc length parameter of** γ is the function $s : [a, b] \to [0, \infty)$ given by

$$s(t) = \int_a^t ||\gamma'(u)|| du, \qquad a \le t \le b.$$

Remark 11.2.5 Notice that the arc length parameter depends on the parametrization γ. It would be appropriate to write s_γ instead of s

Informally speaking, the arc length parameter $s(t)$ is the length of γ on the interval $[a, t]$ for $a \le t \le b$. Physically speaking, $s(t)$ is the distance travelled by the particle from time a to time t. This parameter is geometrically convenient.

Example 11.2.6 The helix $\gamma_1(t) = (5\cos t, 5\sin t, 2t)$ for $0 \le t \le 2\pi$ has arclength parameter given by

$$s(t) = \int_0^t ||\gamma_1'(u)|| du = \int_0^t \sqrt{29} dt = t\sqrt{29}.$$

Given this notion, you may wonder about the existence of a special parametrization.

Can you parametrize a curve so you can travel 1 unit distance per 1 unit time?

Such a parametrization has its own name.

Definition 11.2.7 Let $\gamma : [a,b] \to \mathbb{R}^n$ be a parametrization of a curve in $\mathbb{R}^n$. The map γ is **parametrized by arc length** if $||\gamma'(t)|| = 1$ for $a < t < b$.

Remark 11.2.8 You can show that γ is parametrized by arc length if and only if its arc length parameter satisfies $s(t) = t - a$ for $a \leq t \leq b$.

Remark 11.2.9 If γ is parametrized by arc length, then the variable s is usually used as the parameter for γ. That is, you usually write $\gamma(s)$ instead of $\gamma(t)$. This common convention is an abuse of notation since $s(t)$ also represents the arc length parameter. This will rarely be an issue but you should pay attention to context.

Example 11.2.10 The helix from Example 11.2.6 is parametrized by arc length via

$$\gamma_2(s) = \Big(5\cos(\frac{s}{\sqrt{29}}), 5\sin(\frac{s}{\sqrt{29}}), \frac{2s}{\sqrt{29}}\Big), \qquad 0 \leq s \leq 2\pi\sqrt{29}.$$

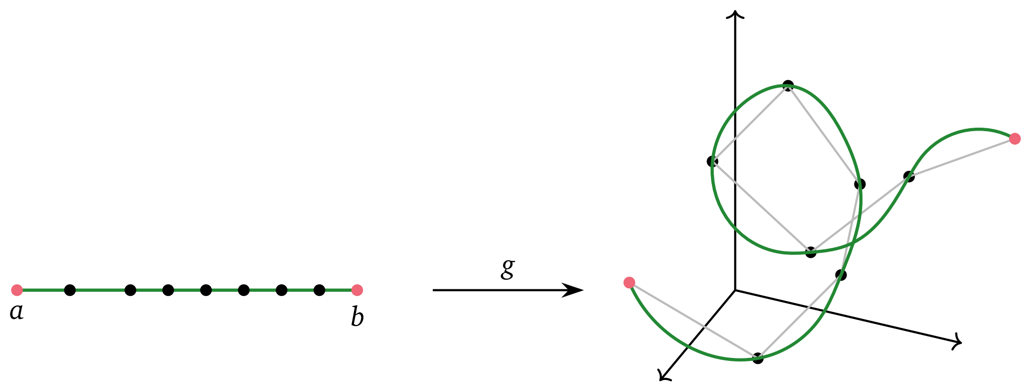
You can verify this fact since $||\gamma_2'(s)|| = 1$ for all $0 \leq s \leq 2\pi\sqrt{29}$. Notice the curves γ_1 is a reparametrization of γ_2. You can verify this by Definition 11.1.23 with the diffeomorphism $\varphi(t) = t\sqrt{29}$; as seen in Example 11.2.6, this is exactly the arc length parameter of γ_1! By direct computation, you can also verify that $\ell(\gamma_2) = \ell(\gamma_1)$.

This completes the geometric applications and computational aspects of arc length. It remains to explore the promised rigorous derivation of arc length.

11.2.3 Derivation of arc length

Finding the length of a "curvy" curve will follow the usual procedure of chopping, estimating, and refining, but you will also rely on two vital assumptions. First, the length of a straight line segment between a point $p \in \mathbb{R}^n$ and a point $q \in \mathbb{R}^n$ is exactly $||p - q||$. Second, this is the shortest path between these two points.

Let $\gamma : [a,b] \to \mathbb{R}^n$ parametrize a curve $C \subseteq \mathbb{R}^n$. A discrete approximation of the curve γ is a polygonal curve, so you can begin by chopping up the curve. Let $P = \{t_0, t_1, \ldots, t_N\}$ be a partition of the interval $[a,b]$. The figure[6] below illustrates this setup.



The length of the polygonal approximation to γ using the partition P is

$$\sum_{i=1}^{N} ||\gamma(t_i) - \gamma(t_{i-1})||.$$

[6] Inspired by Shifrin [18].

Since the shortest distance between two points is a straight line, this must be an *underestimate* for the length of γ. By taking a supremum, this discrete approximation should approach the length of γ. Notice, however, it is not at all clear whether the supremum is bounded! Nonetheless, you can derive the definition for arc length via a formal theorem.

Theorem 11.2.11 If $\gamma : [a,b] \to \mathbb{R}^n$ parametrizes a curve $C \subseteq \mathbb{R}^n$, then

$$\int_a^b ||\gamma'(t)||dt = \sup_P \left\{ \sum_{i=1}^N ||\gamma(t_i) - \gamma(t_{i-1})|| \right\},$$

where the supremum is over all partitions $P = \{t_0, t_1, \ldots, t_N\}$ of $[a,b]$.

Proof. Without loss of generality, assume C is a curve and hence γ is a smooth simple regular parametrization. Let $P = \{t_0, \ldots, t_N\}$ be a partition of $[a,b]$. By the fundamental theorem of calculus for each component of γ,

$$\gamma(t_i) - \gamma(t_{i-1}) = \int_{t_{i-1}}^{t_i} \gamma'(t)dt.$$

Therefore, by the triangle inequality for integrals (Theorem 7.7.13),

$$\ell_P(\gamma) := \sum_{i=1}^N ||\gamma(t_i) - \gamma(t_{i-1})|| = \sum_{i=1}^N \left|\left| \int_{t_{i-1}}^{t_i} \gamma'(t)dt \right|\right| \leq \sum_{i=1}^N \int_{t_{i-1}}^{t_i} ||\gamma'(t)||dt = \int_a^b ||\gamma'(t)||dt.$$

Thus, $\sup_P \ell_P(\gamma) \leq \int_a^b ||\gamma'(t)||dt$ so the supremum is bounded and hence exists. The same inequality holds on any subinterval. Now, for $a \leq t \leq b$, let $S(t)$ be the supremum of polygonal approximations of the curve γ on $[a,t]$. For $h > 0$,

$$\frac{||\gamma(t+h) - \gamma(t)||}{h} \leq \frac{S(t+h) - S(t)}{h} \leq \frac{1}{h} \int_t^{t+h} ||\gamma'(u)||du,$$

since $S(t+h) - S(t)$ overestimates the distance from $\gamma(t)$ to $\gamma(t+h)$. Hence,

$$\lim_{h \to 0^+} \frac{||\gamma(t+h) - \gamma(t)||}{h} = ||\gamma'(t)|| = \lim_{h \to 0^+} \frac{1}{h} \int_t^{t+h} ||\gamma'(u)||du.$$

By the squeeze theorem,

$$\lim_{h \to 0^+} \frac{S(t+h) - S(t)}{h} = ||\gamma'(t)||.$$

The same holds for $h \to 0^-$ so $S'(t) = ||\gamma'(t)||$. Thus,

$$S(t) = \int_a^t ||\gamma'(u)||du, \qquad a \leq t \leq b$$

so $S(b) = \displaystyle\int_a^b ||\gamma'(t)||dt$. This completes the proof. ∎

This solidifies the integral formula for arc length (Definition 11.2.1) using both a rigorous derivation and a physical principle.

11.2.4 Line integrals of scalar functions

Here you will expand your newfound ideas to a more general situation.

What is the mass of a curve in $\mathbb{R}^n$ with variable density?

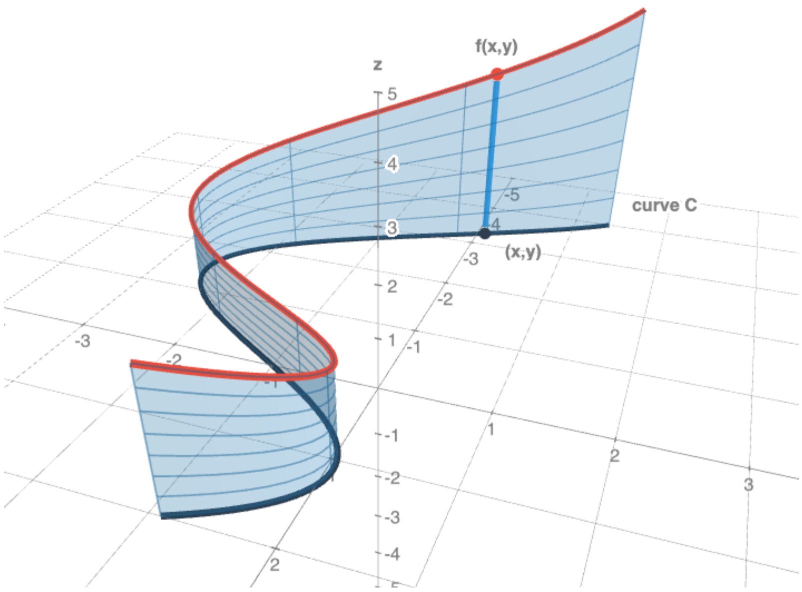
More formally, you want to integrate a scalar function along a curve. The definition of arc length naturally suggests a good definition.

Definition 11.2.12 Let C be a curve in $\mathbb{R}^n$ parametrized by $\gamma : [a,b] \to \mathbb{R}^n$. Let f be a bounded real-valued function defined on C. The **line integral of f over C** is given by

$$\int_C f\,ds := \int_a^b f(\gamma(t))||\gamma'(t)||dt.$$

If this integral exists, then f is **integrable on the curve C**.

The symbol ds is the **arclength element**. It has no formal meaning but you can view it as the infinitesimal length of the curve. For a curve C in $\mathbb{R}^2$, the line integral $\int_C f\,ds$ can be geometrically interpreted as the "net area of the surface traced out by f along C". This idea is illustrated in the Math3D graph below.



Is this well-defined? That is, does the line integral depend on the parameterization?

Theorem 11.2.13 (Invariance of line integrals) Let C be a curve in $\mathbb{R}^n$. Let f be a bounded real-valued function defined on C. Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be parametrizations of C. Assume γ_1 is a reparametrization of γ_2. The function $(f \circ \gamma_1)||\gamma_1'||$ is integrable on $[a,b]$ if and only if the function $(f \circ \gamma_2)||\gamma_2'||$ is integrable on $[c,d]$. If so,

$$\int_a^b f(\gamma_1(t))||\gamma_1'(t)||dt = \int_c^d f(\gamma_2(t))||\gamma_2'(t)||dt.$$

Proof. This is left as an exercise. Use chain rule and change of variables. ∎

Theorem 11.2.13 proves that line integrals of scalar functions are well-defined. That is, you can integrate over a curve in $\mathbb{R}^n$ *independent of the choice of your parametrization*.

Example 11.2.14 Let C in $\mathbb{R}^2$ be the piece of parabola $y = x^2$ for $0 \le x \le 1$. The line integral of $f(x, y) = x + y$ on C can be computed with the parametrization $\gamma : [0, 1] \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$. Then

$$\int_C f\,ds = \int_0^1 (t + t^2)\sqrt{1 + 4t^2}\,dt = \frac{1}{192}(-16 + 134\sqrt{5} - 3\sinh^{-1}(2)) \approx 1.4547$$

View this Math3D demo of this line integral.

Arc length (and vector calculus in general) has beautiful geometric and physical interpretations. The notation is designed to be highly suggestive of these connections. If you learn to informally use this notation, these relationships become obvious. You will conclude this section with a brief digression on this heuristic understanding.

11.2.5 Elements and infinitesimals

Recall the symbols referred to as *infinitesimals* and *elements* of "n-dimensional solids" in $\mathbb{R}^n$.

- The infinitesimal length in $\mathbb{R}$ is represented by the length element dx.
- The infinitesimal area in $\mathbb{R}^2$ is represented by the area element dA.
- The infinitesimal volume in $\mathbb{R}^n$ is represented by the volume element dV.

Again, none of these symbols have a formal meaning[7] but the basic idea is:

To find the mass of an object, integrate with respect to its corresponding element.

The change of variables theorem implies that integration with respect to the volume element dV is *independent of the choice of coordinates*. For example, in $\mathbb{R}^3$, you may write

$$dV = dx\,dy\,dz = r\,dr\,d\theta\,dz = \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$$

to explain integration with respect to dV in one of rectangular, cylindrical, or spherical coordinates. Regardless of which coordinate system you choose, the integral will be same! In other words, the volume element dV is *intrinsic* to the space $\mathbb{R}^n$ and not dependent on how the space is described.

Now, the infinitesimal arc length for a curve C in $\mathbb{R}^n$ is represented by the arc length element ds. If C is parametrized by a C^1 curve $\gamma : [a, b] \to \mathbb{R}^n$ then you may write

$$ds = ||\gamma'(t)||dt.$$

This has no formal meaning but it is useful to treat ds as the "infinitesimal length of the curve" (i.e. an arbitrary small piece of the curve). Using a specific parametrization γ is like a choice of coordinates for the curve C. The invariance property of Theorem 11.2.13 means that the symbol ds can represent integration over the curve C regardless of the parametrization. If γ_1 and γ_2 are both parametrizations of C then

$$ds = ||\gamma_1'(t)||dt = ||\gamma_2'(t)||dt$$

In other words, the arc length element ds is *intrinsic* to the curve C and not dependent on how the curve is parameterized. You will soon leverage this intuition.

In the next section, you will extend these ideas with integrals of scalar functions over curves (i.e. mass of a wire with variable density) to integrals of *vector fields* over *oriented* curves (i.e work done by a force field). The physical and geometric interpretations of integral calculus with curves will take centre stage.

[7] All of this can be formalized with differential forms; see a course on differential geometry for details.

Exercises for Section 11.2

Concepts and definitions

11.2.1 Let $\gamma : [a,b] \to \mathbb{R}^n$ be a parametrization of a curve $C \subseteq \mathbb{R}^n$. Let $s : [a,b] \to [0,\infty)$ be the arc length parameter of γ. Consider the following algebraic expressions.

$$\gamma(t) \qquad ||\gamma(t)|| \qquad \gamma'(t) \qquad ||\gamma'(t)|| \qquad \int_a^b ||\gamma'(t)||dt \qquad \int_a^t ||\gamma'(u)||du \qquad s(t) \qquad s'(t) \qquad s(b)$$

View γ as the motion of a particle along the path C, so you can interpret each expression above physically. For each physical description below, match the corresponding expression(s) above.

(a) The position of the particle at time t.

(b) The speed of the particle at time t.

(c) The total distance travelled by the particle.

(d) The distance travelled by the particle by time t.

Computations

11.2.2 Let $\gamma : [1,6] :\to \mathbb{R}^3$ be a parametrization of a curve $C \subseteq \mathbb{R}^3$ with arc length parameter $s(t) = t^2 - t$.

(a) If possible, compute the following quantities.

i) $\ell(C)$

ii) $\displaystyle\int_2^4 ||\gamma'(t)||dt$

iii) $||\gamma(4) - \gamma(2)||$

iv) $||\gamma(4)|| - ||\gamma(2)||$

v) $||\gamma'(5)||$

vi) $\gamma'(5)$

(b) Let $\phi : [A,B] :\to \mathbb{R}^3$ be the **arc length** reparametrization of γ. What is A and B?

(c) If possible, compute the following quantities.

i) $\ell(C)$

ii) $\displaystyle\int_2^4 ||\phi'(t)||dt$

iii) $||\phi(4) - \phi(2)||$

iv) $\displaystyle\int_{10}^{14} ||\phi'(t)||dt$

v) $||\phi'(5)||$

vi) $\phi'(5)$

11.2.3 Express the arc length of the curve as an integral. Evaluate it using WolframAlpha.

(a) The curve $y = x^3$ from $x = 0$ to $x = 5$

(b) The perimeter of the ellipse $\frac{x^2}{4} + y^2 = 1$.

(c) The helix $\gamma(t) = (t\cos t, t\sin t, t)$ for $0 \le t \le 10\pi$.

11.2.4 A particle travels along the path $\gamma(t) = (t^2, t^3)$ over the interval $[0, 237]$.

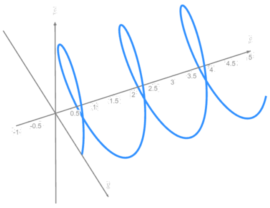
(a) Compute the arc length parameter of the particle's path.

(b) Reparametrize the path using an arc length parametrization.

11.2.5 Express the length of the curve $\gamma(t) = (t, t^2, t^3, \ldots, t^n)$ for $0 \le t \le 1$ as an integral. Evaluate this integral for $n = 5$.

11.2.6 Evaluate $\displaystyle\int_C (x^2 + y^2 + z)ds$ where C is a helix parametrized by $\gamma(t) = (\cos t, \sin t, t)$ for $0 \leq t \leq 2\pi$.

11.2.7 Find the mass of the spiral coil C parametrized by $\gamma(t) = (\cos t, t/4, \sin t)$ for $0 \leq t \leq 6\pi$ with density $\rho(x, y, z) = x + y$. Setup your integral and evaluate it with WolframAlpha.



Proofs

11.2.8 Let $\gamma_1 : [a, b] \to \mathbb{R}^n$ and $\gamma_2 : [c, d] \to \mathbb{R}^n$ be parametrizations of a curve $C \subseteq \mathbb{R}^n$. Prove that if γ_1 is a reparametrization of γ_2 then $\displaystyle\int_a^b ||\gamma_1'(t)||dt = \int_c^d ||\gamma_2'(t)||dt$.
Hint: Use multivariable chain rule and a substitution.

11.2.9 Let $C \subseteq \mathbb{R}^2$ be a curve parametrized by $\gamma : [a, b] \to \mathbb{R}^2$. Let $f : C \to [0, \infty)$ be continuous.

(a) Conjecture an expression for the average value of f on C.

(b) Justify your formula through a heuristic derivation.

(c) Prove that the average value of f on C is attained by f.

11.2.10 Prove the invariance of line integrals of scalar functions (Theorem 11.2.13).

11.2.11 Show that γ is parametrized by arc length if and only if its arc length parameter satisfies $s(t) = t - a$ for $a \leq t \leq b$.

Applications and beyond

11.2.12 You walk along a path $\gamma : [0, 4] \to \mathbb{R}^2$. Your watch records your GPS position at a few times.

t	0.0	2.3	2.7	4.0
$\gamma(t)$	$(0, 0)$	$(-0.5, 2)$	$(0.5, 2.5)$	$(0, 4.0)$

Position is measured in kilometres and time is measured in hours.

(a) Estimate the distance you travelled.

(b) Prove that your path is **not** parametrized by arc length.

11.2.13 Let $\gamma : [a, b] \to \mathbb{R}^n$ be a parametrization of a curve.

(a) Use a regular partition P_N of $[a, b]$ to construct a polygonal approximation $\ell_N(\gamma)$ of the arc length $\ell(\gamma)$.

(b) You want to show that $\displaystyle\lim_{N\to\infty} \ell_N(\gamma) = \ell(\gamma)$ but this is not easy. Here is an attempt.

1. *As $N \to \infty$, the width of the regular partition P_N tends to zero.*
2. *Since P_N is a partition of $[a,b]$, it follows that*

$$\lim_{N\to\infty} \ell_N(\gamma) = \sup_P \left\{ \sum_{i=1}^{N} ||\gamma(t_i) - \gamma(t_{i-1})|| \right\}$$

   where the supremum is over all partitions $P = \{t_0, \ldots, t_N\}$ of $[a,b]$.
3. *By Theorem 11.2.11, this supremum is equal to $\displaystyle\int_a^b ||\gamma'(t)||dt$ which is the arc length.*

Technically, none of the statements are false but this argument is missing a lot of justification. Identify what key items need to be justified. Do not justify them.

11.2.14 A thin wire in the shape of a curve C in $\mathbb{R}^n$ has a continuous mass density function ρ. You can informally derive its mass formula using infinitesimals.

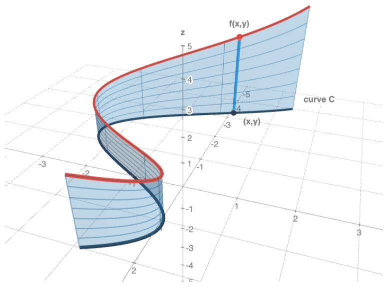
(a) Fill in the blanks of this incomplete attempt.

1. *An infinitesimal piece of the curve C has length __________.*
2. *The __________ is equal to ρds.*
3. *By __________ , the total mass of the wire is therefore equal to __________.*

(b) Line 2 should be justified with a physical principle. Add the necessary justification.

(c) Sketch and label a picture illustrating this argument with infinitesimals.

11.2.15 A fence is laid along a curve $C \subseteq \mathbb{R}^2$ parametrized by $\gamma : [a,b] \to \mathbb{R}^2$. The height of the fence varies along the curve, so it is described by a continuous function $f : C \to [0,\infty)$.



Use these quantities to write a formal mathematical expression for each physical quantity.

(a) What is the length of the fence?

(b) What is the area of the fence?

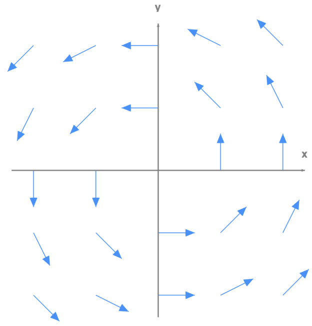
(c) What is the maximum height of the fence?

(d) What is the average height of the fence?

11.2.16 A thin wire C in $\mathbb{R}^2$ is bent into the shape of a semicircle of radius R. If the mass density $\rho(x,y)$ at a point (x,y) is directly proportional to its distance from the line through the endpoints, find the mass of the wire.

11.3. Line integrals

The physics of forces and fluids studies how vector fields affect the motion of a particle. Recall from Section 1.3 that a **vector field** F is a map from a subset of $\mathbb{R}^n$ to $\mathbb{R}^n$. You can think of it as placing a vector (e.g. a force) at each point in $\mathbb{R}^n$, such as a fluid flow.



For instance, if you are paddling along a flowing river, how much is the river helping you or pushing you? This will depend on your path C and the direction along C that you are paddling. If you paddle downstream along a path, you will be helped by the river. If you paddle upstream along the same path, you will have to paddle hard to push against the river. The choice of direction is referred to as *orientation*. This raises a core question in physics that fundamentally motivates this chapter.

*What is the work done by a vector field F along an **oriented** curve C?*

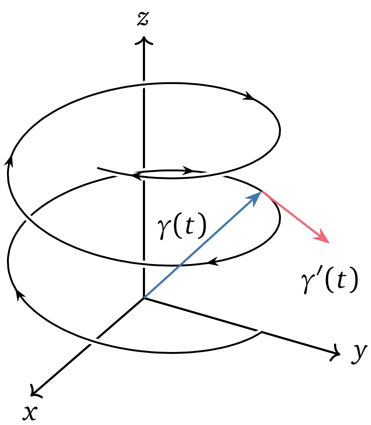
The basic idea will be to "add up" the work done along the curve by integrating along the curve with the arc length element ds. This strategy presents two obstacles. First, you will need to take into account the direction of motion. Second, your definition must be independent of parametrization. These issues therefore push you to address a preliminary problem.

How can you define the orientation of a curve?

This will require some thoughtful definitions to be sufficiently rigorous and avoid dependence on a parametrization. Once you complete this task, you will be ready to define a line integral along an oriented curve and therefore calculate the work done. Your investigations will always be guided by physical and geometric intuition.

11.3.1 Oriented curves

Parametrizations represent motion and define curves. How can you keep track of the direction of motion? The orientation of a parametrization $\gamma : [a, b] \to \mathbb{R}^n$ is determined by the direction of its unit tangent vector $T(t) = \frac{\gamma'(t)}{||\gamma'(t)||}$, which was informally introduced in Section 1.1.



The unit tangent vector only contains information about direction, and nothing about speed. How does the unit tangent vector change under reparametrization? The answer depends only on the orientation.

Theorem 11.3.1 (Invariance of unit tangent) Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be parametrizations of a curve C.

(a) If γ_1 is a reparametrization of γ_2 with the same orientation, then

$$\forall s \in (a,b), \forall t \in (c,d), \quad \gamma_1(s) = \gamma_2(t) \implies \frac{\gamma_1'(s)}{||\gamma_1'(s)||} = \frac{\gamma_2'(t)}{||\gamma_2'(t)||}.$$

(b) If γ_1 is a reparametrization of γ_2 with the opposite orientation, then

$$\forall s \in (a,b), \forall t \in (c,d), \quad \gamma_1(s) = \gamma_2(t) \implies \frac{\gamma_1'(s)}{||\gamma_1'(s)||} = -\frac{\gamma_2'(t)}{||\gamma_2'(t)||}.$$

Proof. This is left as an exercise. It follows from the definitions, chain rule, and the identity that $\frac{x}{|x|} = 1$ if $x > 0$ and $\frac{x}{|x|} = -1$ if $x < 0$. ■

Informally speaking, this theorem can be interpreted as an invariance property.

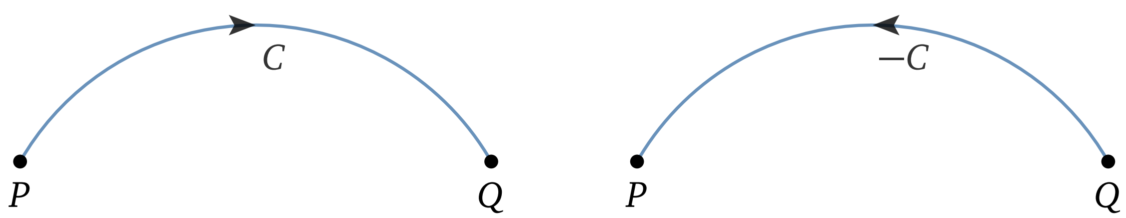
> *The unit tangent vector of two parametrizations with the same orientation remains the same at every point along the curve.*

This observation suggests a way to define an oriented curve.

Definition 11.3.2 An **oriented curve** C is a set[8] of parametrizations that are reparametrizations of each other with the same orientation.

Remark 11.3.3 Formally speaking, C is a set of maps $\gamma : [a,b] \to \mathbb{R}^n$. Informally, you may think of C as the set in $\mathbb{R}^n$ traced out by a parametrization along with a "direction". This abuse of notation is harmless since all related quantities will not depend on the parametrization.

An oriented curve has a corresponding oriented curve with opposite orientation, as illustrated in the figure below.



Definition 11.3.4 Let C be an oriented curve parametrized by $\gamma : [a,b] \to \mathbb{R}^n$. The **oppositely oriented curve** $-C$ is the set of parametrizations that are reparametrizations of γ with the opposite orientation.

Remark 11.3.5 You can prove that $-C$ is an oriented curve. This requires a variant of Lemma 11.1.26 that includes how orientation is affected by symmetry and transitivity.

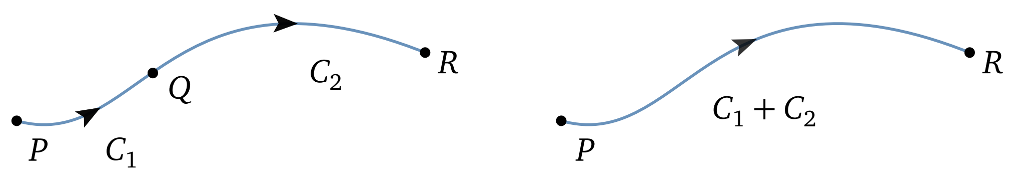
[8]The correct mathematical terminology is *equivalence class*.

Oriented curves are usually described with informal language because this is usually enough to construct a parametrization.

Example 11.3.6 Let C be the path along the parabola $y = x^2$ from the origin to the point $(1,1)$. The map $\gamma(t) = (t, t^2)$ on $0 \le t \le 1$ is a parametrization of C. Consider three other continuous maps whose image is equal to C.

- $\gamma_1(t) = (2t, 4t^2)$ for $0 \le t \le 1/2$ is a reparametrization of γ with the same orientation.
- $\gamma_2(t) = (1-t, (1-t)^2)$ for $0 \le t \le 1$ is a reparametrization of γ with the opposite orientation. Thus, γ_2 is a parametrization for $-C$.
- $\gamma_3(t) = (\sin^2(t), \sin^4(t))$ for $0 \le t \le \pi$ is not a reparametrization of γ, because it is not simple and hence not a parametrization. It traces out the same piece of the parabola but it goes from the origin to $(1,1)$ and back.

Gluing oriented curves together is a common practice and is referred to as "concatenation". If C_1 and C_2 are oriented curves such that the end of C_1 is the start of C_2, then the concatenation of C_1 and C_2 is the piecewise oriented curve $C_1 + C_2$ defined by joining them together.



The above figure illustrates this process. The definition below is a more formal version.

Definition 11.3.7 Let C_1 and C_2 be oriented curves in $\mathbb{R}^n$. The **concatenation of** C_1 **with** C_2, denoted $C = C_1 + C_2$, is the set of continuous maps $\gamma : [a,b] \to \mathbb{R}^n$ such that there exists a value $c \in (a,b)$ where $\gamma|_{[a,c]}$ and $\gamma|_{[c,b]}$ are parametrizations of C_1 and C_2 respectively.

Remark 11.3.8 This definition can be extended to concatenate any finite number of oriented curves. Moreover,the concatenation of C_1 and $-C_2$ can be denoted $C_1 + (-C_2)$ or equivalently as $C_1 - C_2$.

Example 11.3.9 Let C_1 be the path in $\mathbb{R}^2$ along the x-axis from $(-1,0)$ to $(1,0)$. Let C_2 be the path along the upper half unit circle from $(1,0)$ to $(-1,0)$. The concatenation $C_1 + C_2$ exists and is a closed piecewise curve.

Concatenation suggests a natural generalization for piecewise oriented curves.

Definition 11.3.10 A **piecewise oriented curve** in $\mathbb{R}^n$ is the concatenation of finitely many oriented curves in $\mathbb{R}^n$.

With a formal description of oriented curves, you can study how force fields affect moving particles by defining line integrals along oriented curves.

11.3.2 Line integrals of vector fields

To calculate the work done by a vector field on a moving object, the physical principle is simple:

> *The work done by a **constant** force on an object is the force in the direction of motion multiplied by the displacement of the object.*

To derive a good definition, you can do the usual "chop, estimate, and refine" procedure but that is postponed to the end of the reading. Instead, you can derive an integral formula by treating the arc length element ds like an infinitesimal.

> *Consider an infinitesimal piece of the curve with infinitesimal length ds. The tangent direction of the curve at this point is the unit tangent vector T. The amount of force F in the tangential direction is therefore $F \cdot T$. Thus, the infinitesimal work done is $(F \cdot T)ds$. Integrating this along the curve C gives the total work done $\int_C F \cdot T ds$.*

You have quickly stumbled onto the formal definition.

Definition 11.3.11 Let C be an oriented curve parametrized by $\gamma : [a,b] \to \mathbb{R}^n$ with unit tangent vector T. Let F be a vector field in $\mathbb{R}^n$ defined on C. The **line integral of F over C** is given by

$$\int_C F \cdot T ds := \int_a^b F(\gamma(t)) \cdot T(t) ||\gamma'(t)|| dt$$

provided this integral exists. Equivalently, this is the **work done by F along the curve C**.

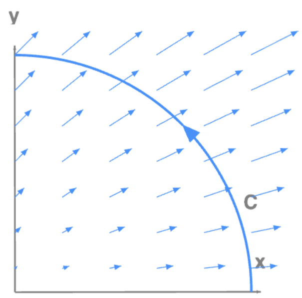
Unsurprisingly, line integrals are invariant under choice of parametrization.

Theorem 11.3.12 (Invariance of line integrals) Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be parametrizations of an oriented curve C with unit tangent vectors T_1 and T_2 respectively. Let F be a vector field in $\mathbb{R}^n$ defined on C. The function $(F \circ \gamma_1)(t) \cdot T_1(t)||\gamma_1'(t)||$ is integrable on $[a,b]$ if and only if $(F \circ \gamma_2)(t) \cdot T_2(t)||\gamma_2'(t)||$ is integrable on $[c,d]$. If so,

$$\int_a^b F(\gamma_1(t)) \cdot T_1(t)||\gamma_1'(t)||dt = \int_c^d F(\gamma_2(t)) \cdot T_2(t)||\gamma_2'(t)||dt.$$

Proof. This is left as an exercise. Use chain rule, change of variables, and the definition of the unit tangent. Remember the orientation must be used as an important step in your proof. ■

Example 11.3.13 Suppose you want to compute $\displaystyle\int_C F \cdot T ds$ where $F(x,y) = (x+y, y)$ and C is the quarter circle oriented counterclockwise as shown below.



Parametrize C by $\gamma(t) = (\cos t, \sin t)$ for $0 \le t \le \pi/2$. You can confirm that this parametrization is smooth simple and regular. Then $F(\gamma(t)) = (\cos t + \sin t, \sin t)$ so, by definition,

$$\int_C F \cdot T ds = \int_0^{\pi/2} (\cos t + \sin t, \sin t) \cdot \frac{\gamma'(t)}{||\gamma'(t)||} ||\gamma'(t)|| dt$$

Notice the scalars $||\gamma'(t)||$ cancel. Therefore, we get that

$$\int_C F \cdot T ds = \int_0^{\pi/2} (\cos t + \sin t, \sin t) \cdot \gamma'(t) dt = \int_0^{\pi/2} (\cos t + \sin t, \sin t) \cdot (-\sin t, \cos t) dt$$
$$= \int_0^{\pi/2} -\sin^2 t dt = \frac{-\pi}{4}.$$

When computing line integrals, you never need to actually compute the unit tangent vector

$$T(t) = \frac{\gamma'(t)}{||\gamma'(t)||}$$

because $||\gamma'(t)||$ will always cancel with the arc length element $ds = ||\gamma'(t)||dt$. That is,

$$\int_C F \cdot T ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt.$$

This observation leads to equivalent notation for a line integral:

$$\int_C F \cdot d\gamma.$$

Here γ is a parametrization of C and you may informally treat the symbol $d\gamma$ as $d\gamma = \gamma'(t)dt$. Notice this element is a "vector" of elements $d\gamma = (\gamma_1'(t)dt, \dots, \gamma_n'(t)dt)$.

Example 11.3.14 From the previous example, notice

$$\int_C F \cdot d\gamma = \int_0^{\pi/2} F(\gamma(t)) \cdot \gamma'(t) dt = \int_0^{\pi/2} (\cos t + \sin t, \sin t) \cdot (-\sin t, \cos t) dt = -\frac{\pi}{4}.$$

There is yet another notation for line integrals. For a vector field $F = (F_1, \dots, F_n)$ of $\mathbb{R}^n$, you can denote

$$\int_C F \cdot d\gamma = \int_C F_1 dx_1 + F_2 dx_2 + \dots + F_n dx_n$$

by thinking of $d\gamma$ as $(dx_1, \dots, dx_n)$, where

$$\int_C F_j dx_j = \int_a^b F_j(\gamma(t)) \gamma_j'(t) dt$$

for $1 \le j \le n$. Again, this meaningless symbol pushing is only for notational convenience.

Example 11.3.15 Again, from the previous example, notice for $F(x,y) = (P(x,y), Q(x,y)) = (x+y, y)$,

$$\int_C F \cdot d\gamma = \int_C P dx + Q dy = \int_C (x+y) dx + y dy.$$

As $\gamma(t) = (\cos t, \sin t)$ for $0 \le t \le \pi/2$ parametrizes the quartercircle C, the above equals

$$\int_0^{\pi/2} (\cos t + \sin t)(-\sin t) + \sin t(\cos t) dt = -\frac{\pi}{4}.$$

You will need to become comfortable with using all of this equivalent notation. There are also some basic properties of line integrals that are worth recording.

Lemma 11.3.16 Let C, C_1, C_2 be oriented curves in $\mathbb{R}^n$. Let F and G be continuous vector fields in $\mathbb{R}^n$ defined on C, C_1, and C_2. All of the following hold:

(a) $\displaystyle\int_{-C} F \cdot T\,ds = -\int_C F \cdot T\,ds$.

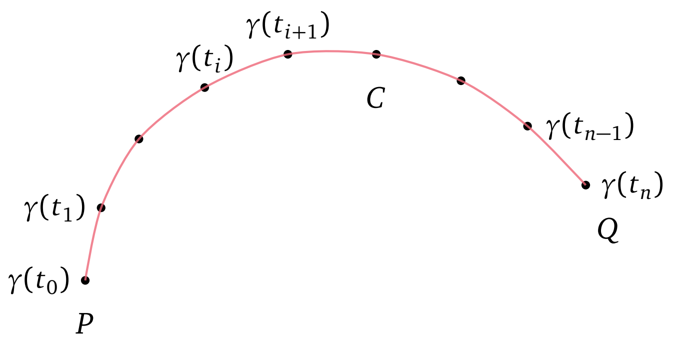
(b) $\displaystyle\int_C (F + \lambda G) \cdot T\,ds = \int_C F \cdot T\,ds + \lambda \int_C G \cdot T\,ds$ for $\lambda \in \mathbb{R}$.

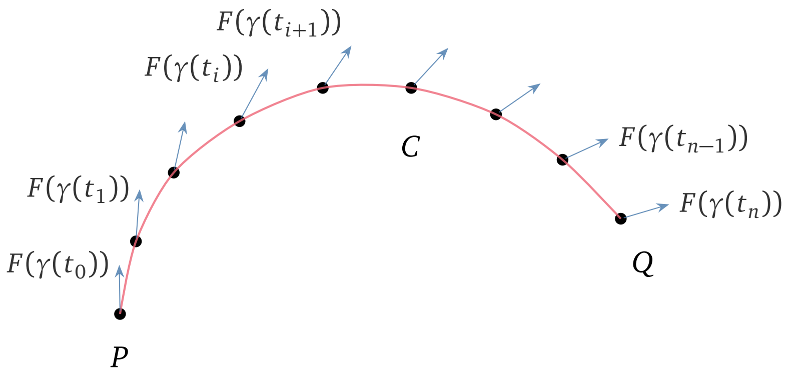
(c) If the end of C_1 is the start of C_2 then $\displaystyle\int_{C_1+C_2} F \cdot T\,ds = \int_{C_1} F \cdot T\,ds + \int_{C_2} F \cdot T\,ds$.

Proof. These are left as exercises. For (a), show that if T is a unit tangent vector for C then $-T$ is a unit tangent vector for $-C$. For (b), use the definitions and linearity of the single-variable integral. For (c), use the definitions and additivity of the single-variable integral. ■

11.3.3 Derivation of work done along a curve

Here is a more rigorous derivation of the definition of a line integral for vector fields. Start with chopping a parametrization $\gamma : [a, b] \to \mathbb{R}^n$ of the oriented curve C. Let $\{t_0, t_1, \ldots, t_N\}$ be a regular partition of $[a, b]$.





For a given segment of the curve, the work done along the segment is approximately the amount of force $F(\gamma(t_i))$ in the *tangential* direction of the curve multiplied by the displacement $\gamma(t_{i+1}) - \gamma(t_i)$. The tangential direction of the curve at $\gamma(t)$ is the unit tangent vector $T(t)$ so the amount of force of $F(\gamma(t_i))$ in this direction is given by

$$F(\gamma(t_i)) \cdot T(t_i).$$

The total work done is therefore approximately

$$\sum_{i=1}^{N} F(\gamma(t_i)) \cdot T(t_i) \|\gamma(t_{i+1}) - \gamma(t_i)\| \approx \sum_{i=1}^{N} F(\gamma(t_i)) \cdot T(t_i) \|\gamma'(t_i)\| (t_{i+1} - t_i).$$

The latter approximation follows since $||\gamma(t_{i+1}) - \gamma(t_i)|| \approx ||\gamma'(t_i)||(t_{i+1} - t_i)$. By taking $N \to \infty$, this yields

$$\int_a^b F(\gamma(t)) \cdot T(t)||\gamma'(t)||dt = \int_C F \cdot T ds.$$

You can formalize this derivation into a theorem but this level of detail will be satisfactory for your purposes. This section completes the key definitions for integral calculus with curves. You have the foundation for analyzing the work done by a vector field on a moving object. For the rest of the chapter, you will combine differential and integral calculus with curves and discover powerful generalizations of the fundamental theorem of calculus. These discoveries will have remarkable physical meanings.

Exercises for Section 11.3

Concepts and definitions

11.3.1 Let $\gamma : [a,b] \to \mathbb{R}^n$ be a parametrization of an oriented curve $C \subseteq \mathbb{R}^n$ with unit tangent vector T. Consider the following algebraic expressions. View γ as the motion of a particle along the path C.
For each physical description below, circle the corresponding equivalent expression(s).

(a) The direction of motion of the particle at time t.

$$\gamma(t) \qquad \gamma'(t) \qquad \frac{\gamma'(t)}{||\gamma'(t)||} \qquad T(t) \qquad \gamma(t)\cdot T(t)$$

(b) The amount of tangential force experienced by the particle at time t.

$$T(t) \qquad F(\gamma(t)) \qquad F(\gamma(t))\cdot T(t) \qquad F(\gamma(t))\cdot\gamma(t) \qquad F(\gamma(t))\cdot\gamma'(t)$$

(c) The infinitesimal distance travelled by the particle at time t.

$$||\gamma(t)|| \qquad ||\gamma'(t)|| \qquad \gamma'(t) \qquad ||\gamma(t)||dt \qquad ||\gamma'(t)||dt \qquad \gamma'(t)dt$$

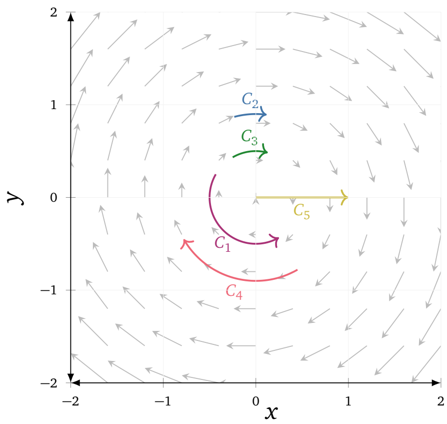
(d) The total work done by the vector field on the particle.

$$\int_a^b F(t)dt \quad \int_a^b F(\gamma(t))dt \quad \int_a^b F(\gamma(t))\cdot T(t)dt \quad \int_a^b F(\gamma(t))\cdot T(t)||\gamma'(t)||dt \quad \int_a^b F(\gamma(t))\cdot\gamma'(t)dt$$

11.3.2 A vector field F in $\mathbb{R}^2$ and oriented curves C_1, C_2, C_3, C_4 and C_5 are shown on the right. The curves C_2 and C_3 are the same length. Consider the five line integrals.

$$\int_{C_1} F\cdot T ds \qquad \int_{C_2} F\cdot T ds \qquad \int_{C_3} F\cdot T ds$$

$$\int_{C_4} F\cdot T ds \qquad \int_{C_5} F\cdot T ds$$

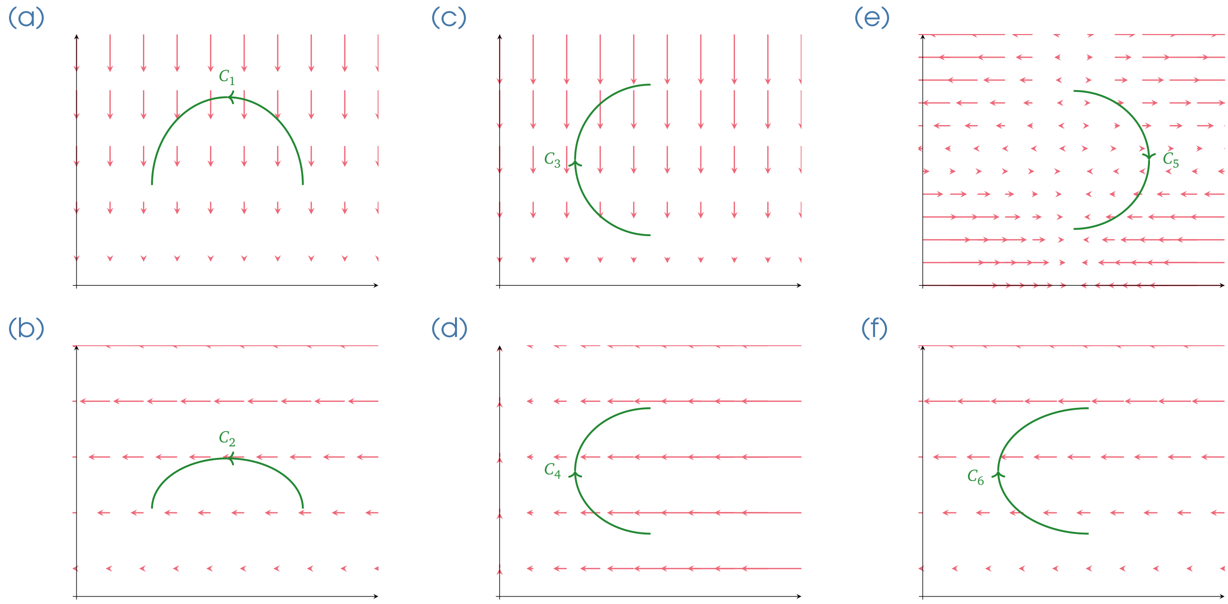


(a) Which of the line integrals are positive?

(b) Which of the line integrals are negative?

(c) Arrange the line integrals in ascending order.

11.3.3 For each vector field, determine if the line integral is likely positive, negative, or zero.



Computations

11.3.4 Find

$$\int_C F \cdot T\, ds$$

where $F(x, y) = (-2x, y)$ and C is the curve $y = 2x^3$ from $(0, 0)$ to $(2, 16)$.

11.3.5 Let $F(x, y) = (y^2, x)$ and C be the arc of the parabola $x = y^2+3$ from $(4, -1)$ to $(7, 2)$. Evaluate

$$\int_C F \cdot d\gamma.$$

11.3.6 Compute

$$\int_C y\,dx + z\,dy + x\,dz$$

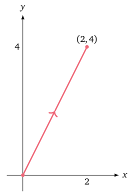
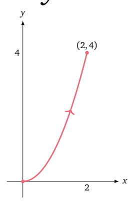
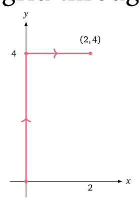
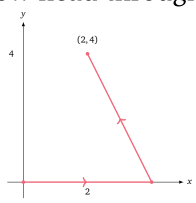
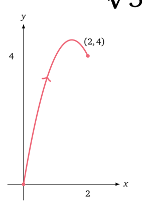
where C is the curve parametrized by $\gamma(t) = (2t, t, t^2)$ for $1 \le t \le 3$.

11.3.7 Calculate

$$\int_C xze^{xy}\,dx$$

where C is the curve given by $\gamma(t) = (t^4, 1, t^{-4})$ for $0 \le t \le 2$.

11.3.8 Let's compare two vector fields F and G. Each path C starts at $(0,0)$ and ends at $(2,4)$. You will fill in the table below with the help of your classmates.

path C from $(0,0)$ to $(2,4)$	$F(x,y)=(2x+3y,3x+5)$ $\int_C F\cdot d\gamma$	$G(x,y)=(4x+2y,3x+2)$ $\int_C G\cdot d\gamma$
(a) straight line 		
(b) parabola $y=x^2$ 		
(c) street grid through $(0,4)$ 		
(d) arrow head through $(4,0)$ 		
(e) sine curve $y=\frac{8}{\sqrt{3}}\sin\left(\frac{\pi x}{3}\right)$ 		

(a) Calculate one of the entries (or rows) in the previous table. As a class, we will fill in the entire table. Setup the integral(s) and use WolframAlpha to evaluate it.

(b) What is the difference between the vector fields F and G?

Proofs

11.3.9 Let $\gamma_1 : [a,b] \to \mathbb{R}^n$ and $\gamma_2 : [c,d] \to \mathbb{R}^n$ be parametrizations of a curve C. Prove that if γ_1 is a reparametrization of γ_2 with the opposite orientation, then

$$\forall s \in (a,b), \forall t \in (c,d), \quad \gamma_1(s) = \gamma_2(t) \implies \frac{\gamma_1'(s)}{||\gamma_1'(s)||} = -\frac{\gamma_2'(t)}{||\gamma_2'(t)||}.$$

11.3.10 Let C be an oriented curve parametrized by $\gamma : [a,b] \to \mathbb{R}^n$ with unit tangent vector T.

(a) Prove that the curve $-C$ with the opposite orientation as C has unit tangent vector $-T$. *Hint:* Use the previous problem.

(b) Let F be a vector field in $\mathbb{R}^n$ continuous on C. Show that $\displaystyle\int_{-C} F \cdot T\,ds = -\int_C F \cdot T\,ds.$

11.3.11 Let C be an oriented curve in $\mathbb{R}^n$. Let F and G be vector fields in $\mathbb{R}^n$ continuous on C. Fix $\lambda \in \mathbb{R}$. Prove that

$$\int_C (F + \lambda G) \cdot d\gamma = \int_C F \cdot d\gamma + \lambda \int_C G \cdot d\gamma.$$

Remember to justify that all three integrals must exist.

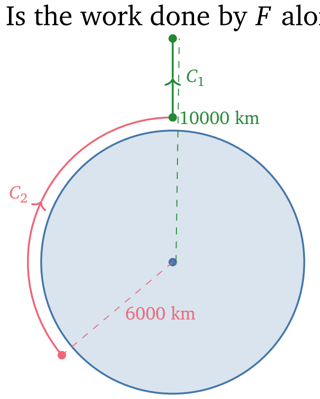
11.3.12 Let C_1 and C_2 be oriented curves in $\mathbb{R}^n$, such that the end of C_1 is the start of C_2. Let F be a vector field in $\mathbb{R}^n$ that is continuous on C_1 and on C_2. Prove that if F is integrable on C_1 and on C_2, then F is integrable on $C_1 + C_2$ and

$$\int_{C_1+C_2} F \cdot d\gamma = \int_{C_1} F \cdot d\gamma + \int_{C_2} F \cdot d\gamma.$$

Applications and beyond

11.3.13 Goku travels along a path subject to a force F, due to gravity. As a Super Saiyan, Goku walks along C_2 and then flies radially away from the center of the Earth along C_1.

(a) Is the work done by F along C_1 positive, zero, or negative?



(b) Is the work done by F along C_2 positive, zero, or negative?

(c) If Goku is at position $x \in \mathbb{R}^3$ and the center of the Earth is the origin then

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

1where G is the gravitational constant, M is the mass of the Earth, and m is the mass of Goku. Calculate the work done by gravity as Goku travels along C_1.

11.3.14 Calculate the work done by a field $F(x,y,z) = (xz, 2y, z)$ on a particle moving in a straight line from $(1,5,7)$ to $(2,3,7)$.

11.3.15 A particle moves through the field $F(x,y) = (3x^2y + 4x + y, x^3 + x)$ counterclockwise from $(0,2)$ along the circular path $x^2 + y^2 = 4$ to $(0,-2)$ and then back to $(0,2)$ along the y-axis. How much work is done by the field on the particle?

11.4. Fundamental theorem of line integrals

You have developed a strong foundation for differentiation and integration over curves and oriented curves. This calculus with curves in $\mathbb{R}^n$ has had many close parallels to calculus over $\mathbb{R}$ since curves are 1-dimensional objects. In an effort to continue generalizing single variable calculus, you can revisit the *fundamental theorem of calculus*. If F is real-valued function that is C^1 on an open set containing the interval $[a,b]$, then

$$\int_a^b F'(x)\,dx = F(b) - F(a). \tag{11.4.1}$$

In other words, differentiation and integration over $\mathbb{R}$ are dual operations. This miraculous discovery is at the heart of single-variable calculus and drives many deep applications. As your first major reward in vector calculus, you will generalize this beautiful theorem over $\mathbb{R}$ to curves in $\mathbb{R}^n$. While this study will have far-reaching implications to the physics of particles moving through vector fields, you will first focus on the mathematical details here.

11.4.1 Statement and proof

Without further ado, here is one of the fundamental theorems of vector calculus.

Theorem 11.4.1 (Fundamental theorem of line integrals) Let C be an oriented piecewise curve in $\mathbb{R}^n$ parametrized by $\gamma : [a,b] \to \mathbb{R}^n$. Let f be a real-valued function that is C^1 on an open set containing C. Then

$$\int_C \nabla f \cdot d\gamma = f(\gamma(b)) - f(\gamma(a)).$$

Remark 11.4.2 In particular, the value of the line integral of the gradient vector field ∇f depends *only* on the endpoints $\gamma(b)$ and $\gamma(a)$. It does not depend on the path.

Its proof is elegant and relies on the fundamental theorem of calculus as stated in (11.4.1).

Proof. Without loss of generality, assume C is a curve in $\mathbb{R}^n$. Hence, its parametrization $\gamma : [a,b] \to \mathbb{R}^n$ is C^1 on (a,b) and continuous on $[a,b]$. By definition,

$$\begin{aligned}
\int_C \nabla f \cdot d\gamma &= \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t) dt && \text{by definition} \\
&= \int_a^b \frac{d}{dt}\big((f \circ \gamma)(t)\big) dt && \text{by the chain rule} \\
&= (f \circ \gamma)(b) - (f \circ \gamma)(a),
\end{aligned}$$

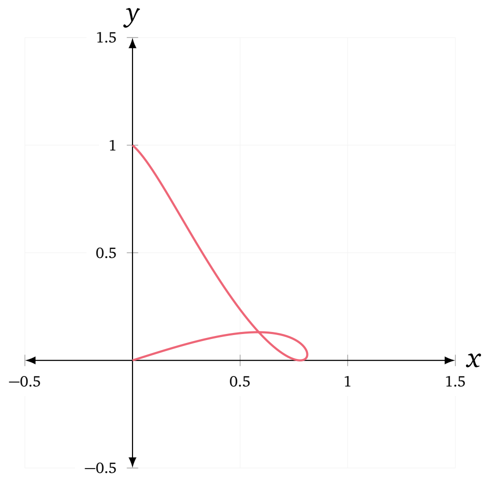
as required. The last step follows by the fundamental theorem of calculus. ∎

Many curves can have nasty parametrizations. You can swiftly apply the fundamental theorem of line integrals to calculate line integrals with a *gradient* vector field.

Example 11.4.3 Suppose $f(x,y) = x^2y - x$. Notice $\nabla f(x,y) = (2xy - 1, x^2)$. Let C be the oriented curve given by $\gamma : [0,1] \to \mathbb{R}^2$ with

$$\gamma(t) = (\sin(\pi t)e^{-t^2}, t\cos^2(\pi t)).$$

Notice C is piecewise smooth. An illustration is included below.



You can try to compute the line integral $\int_C \nabla f \cdot d\gamma$ in two different ways. First, if you try a direct parametrization, i.e. by definition, then you would find that

$$\begin{aligned}\int_C \nabla f \cdot d\gamma &= \int_0^1 \nabla f(\gamma(t)) \cdot \gamma'(t)dt \\ &= \int_0^1 \nabla f(\sin(\pi t)e^{-t^2}, t\cos^2(\pi t)) \cdot (\pi\cos(\pi t)e^{-t^2} - 2t\sin(\pi t)e^{-t^2}, \\ &\cos^2(\pi t) - 2\pi t\cos(\pi t)\sin(\pi t))dt \\ &= \int_0^1 (2te^{-t^2}\sin(\pi t)\cos^2(\pi t) - 1, \sin^2(\pi t)e^{-2t^2}) \cdot (\pi\cos(\pi t)e^{-t^2} - 2t\sin(\pi t)e^{-t^2}, \\ &\cos^2(\pi t) - 2\pi t\cos(\pi t)\sin(\pi t))dt\end{aligned}$$

Computing this integral is a hopeless task to say the least. Luckily, you have discovered a very powerful tool for evaluating such an integral. By Theorem 11.4.1,

$$\int_C \nabla f \cdot d\gamma = f(\gamma(1)) - f(\gamma(0)) = f(0,1) - f(0,0) = 0.$$

It is fun to crush calculations with a big theorem.

The line integral of *gradient* vector field over a *closed* curve will always give the same result.

Example 11.4.4 Let $f(x,y) = e^{-xy}$. Then, $\nabla f(x,y) = (-ye^{-xy}, -xe^{-xy})$. Let C be the oriented curve given by $\gamma(t) = (\cos(t), \sin(t))$ for $t \in [0, 2\pi]$. That is, C is the unit circle traversed counterclockwise. Notice $\gamma(0) = \gamma(2\pi) = (1,0)$ since the curve is closed. Thus, by Theorem 11.4.1,

$$\int_C (-ye^{-xy}, -xe^{-xy}) \cdot d\gamma = e^{-xy}\Big|_{\gamma(0)}^{\gamma(2\pi)} = e^{-xy}\Big|_{(1,0)}^{(1,0)} = 1 - 1 = 0.$$

Since the endpoints were the same in the previous example, the fundamental theorem of line integrals implies the line integral of a *gradient* vector field over a *closed* must evaluate to *zero*! This observation is reminiscent of physical laws on *conservation*.

11.4.2 Conservative vector fields and potentials

Theorem 11.4.1 and its immediate consequences suggests that this family of vector fields ought to have its own name.

Definition 11.4.5 A vector field F is **conservative** on an open set $U \subseteq \mathbb{R}^n$ if there exists a real-valued function $f : U \to \mathbb{R}$ such that $F = \nabla f$ on U. The function f is the **potential function** or **scalar potential** of F.

Remark 11.4.6 Other names include **gradient vector field**. Note physicists usually define potentials with a negative sign, so the relation $F = -\nabla f$ is often used to define a potential. You will not use this convention here.

Informally speaking, the fundamental theorem of line integrals says:

If F is a conservative vector field, then $\int_C F \cdot d\gamma$ only depends on the endpoints of C.

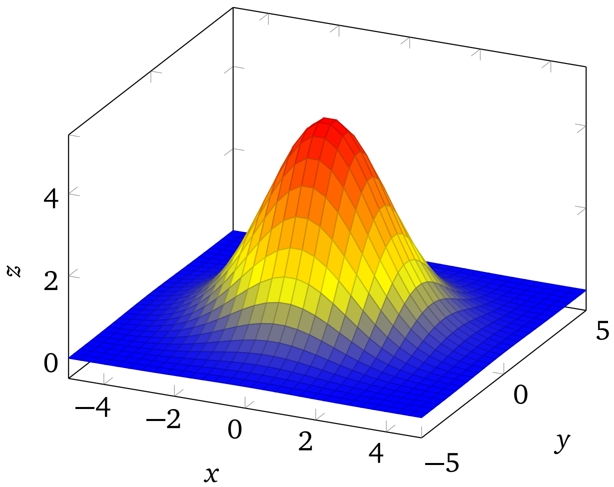
These vector fields have deep and valuable physical interpretations.

Example 11.4.7 On a small piece of Earth's surface, the gravitational field is a uniform acceleration at about 9.8 meters per second squared straight down. This induces a gravitational vector field ∇f with potential

$$f(x, y) = -9.8h(x, y)$$

where $h(x, y)$ is a function which gives the height above sea level at (x, y).

Suppose we have $h(x, y) = 5e^{-\frac{1}{5}(x^2+y^2)}$, representing some hill on Earth. The graph of this function is pictured below.



If you placed a ball on this hill, which way would it roll? The acceleration of the ball along the hill is given by $\nabla f = -9.8\nabla h$. This means that the ball runs in the direction of steepest ascent for the potential, which happens to be the direction of steepest descent along the hill; that is, straight down the hill. Thankfully, this matches what you see in reality.

Now, the fundamental line theorem of line integrals is fantastic but it only applies to gradient vector fields. This immediately creates a vital question.

How do you determine whether a vector field F is conservative? In other words, how do you find a potential f such that $F = \nabla f$?

One approach is to directly solve for the potential.

Example 11.4.8 Is the vector field

$$F(x,y) = (e^x \cos(y), 1 - e^x \sin(y))$$

conservative? Suppose F has a potential f, so $F = \nabla f$. Hence, f must satisfy the pair of partial differential equations

$$\frac{\partial f}{\partial x} = e^x \cos(y) \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 1 - e^x \sin(y) \tag{11.4.2}$$

You can integrate the first above equation with respect to x holding y fixed. This gives

$$f(x,y) = \int e^x \cos(y)dx = e^x \cos(y) + \varphi(y),$$

where φ is an arbitrary function of y. Notice you must add an arbitrary function of y instead of an arbitrary constant because applying the $\frac{\partial}{\partial x}$ operator to f treats terms involving only y as constants with respect to x. From this equation, it follows that

$$\frac{\partial f}{\partial y} = -e^x \sin(y) + \varphi'(y).$$

By (11.4.2), $\varphi'(y) = 1$. This implies $\varphi(y) = y + C$ for some arbitrary $C \in \mathbb{R}$. Then,

$$f(x,y) = e^x \cos(y) + y + C.$$

Now, you have obtained a solution assuming one exists. You can indeed verify by direct calculation that $\nabla f = F$. Notice that a conservative vector field has infinitely many potential functions, all of which differ by a constant.

You can try this same approach on a vector field which happens to be not conservative.

Example 11.4.9 Is the vector field

$$G(x,y) = (y, -x)$$

conservative? Suppose G has a potential function g, so $G = \nabla g$. Thus, g must satisfy

$$\frac{\partial g}{\partial x} = y \qquad \text{and} \qquad \frac{\partial g}{\partial y} = -x \tag{11.4.3}$$

Integrating the first above equation with respect to x, you get that

$$g(x,y) = \int y dx = xy + \psi(y)$$

where ψ is an arbitrary function of y. Then, differentiating this with respect to y,

$$\frac{\partial g}{\partial y} = x + \psi'(y)$$

By (11.4.3), you deduce that $\psi'(y) = -2x$. This is a contradiction because ψ must be a function of y alone. Therefore, the vector field G does not have a potential function and hence is not conservative.

11.4.3 Irrotational vector fields

The previous examples demonstrate a process for determining whether a vector field is conservative. You must attempt to solve some partial differential equations. For a given explicit example, this direct method is manageable but it can be quite hard to solve such equations in general. When this kind of computational barrier exists, you may want to search for equivalent (and hopefully more tractable) definitions.

How can you equivalently characterize conservative vector fields?

A solution to this problem is revolutionary for physics and complex analysis. You will more deeply explore this issue in the next section. Right now, you can make a first crucial observation using Clairaut's theorem (Theorem 6.1.8).

Definition 11.4.10 A C^1 vector field $F = (F_1, \ldots, F_n)$ is **irrotational** on an open set $U \subseteq \mathbb{R}^n$ if

$$\forall 1 \le i < j \le n, \qquad \partial_i F_j = \partial_j F_i \quad \text{on } U.$$

This easy-to-verify property is a necessary condition of conservative vector fields.

Lemma 11.4.11 (Mixed partials test) Let F be a vector field in $\mathbb{R}^n$ that is C^1 on an open set U. If F is conservative on U, then F is irrotational on U.

Proof. Since F is conservative and C^1 on U, there exists a C^2 real-valued function $f : U \to \mathbb{R}$ such that $F = \nabla f$. Thus, for $1 \le i < j \le n$,

$$\begin{aligned} \partial_i F_j &= \partial_i(\partial_j f) && \text{as } F = \nabla f, \\ &= \partial_j(\partial_i f) && \text{by Clairaut's theorem}, \\ &= \partial_j F_i && \text{as } F = \nabla f, \end{aligned}$$

as required. ■

This lemma can sometimes be used to determine whether a vector field is conservative.

Example 11.4.12 Recall the vector field $F(x, y) = (e^x \cos(y), 1 - e^x \sin(y))$ is conservative from the direct approach in Example 11.4.8. By Lemma 11.4.11, it follows that F is irrotational. Indeed, notice that

$$\frac{\partial}{\partial x}(1 - e^x \sin(y)) = -e^x \sin(y) = \frac{\partial}{\partial y}(e^x \cos(y))$$

so F is irrotational by definition. Note, however, this does not imply that F is conservative.

Example 11.4.13 You showed the vector field $G(x, y) = (y, -x)$ from Example 11.4.9 is not conservative by directly attempting to solve for a potential. Is it irrotational? Notice that

$$\frac{\partial}{\partial x}(-x) = -1 \ne 1 = \frac{\partial}{\partial y}(y)$$

Since these partials are not equal, this vector field is not irrotational. The contrapositive of Lemma 11.4.11 implies that G is not conservative. This gives an alternate quicker proof compared to the direct approach of Example 11.4.9.

The converse of Lemma 11.4.11 is false in general.

Example 11.4.14 The vector field

$$H(x,y) = \Big(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\Big)$$

defined on the open set $U = \mathbb{R}^2 \setminus \{(0,0)\}$ is irrotational but not conservative. You can verify by direct computation that H is irrotational. Moreover, if C is the unit circle in $\mathbb{R}^2$ centred at $(0,0)$, then you can verify that

$$\int_C H \cdot T \, ds = 2\pi.$$

If H were conservative, then this line integral over the *closed* curve C should be equal to zero by the fundamental theorem of line integrals; thus, H cannot be conservative. The details of these calculations are left as an exercise.

The fundamental theorem of line integrals brings the family of conservative vector fields under the spotlight, because their line integrals do not depend on the path taken! This feature is highly desirable so you will want to understand these vector fields more deeply. You can search for potential functions by attempting to solve some partial differential equations, but this is not always tractable. You learned that every conservative vector field is also irrotational, which is easy to verify; sadly, the converse is not necessarily true. As usual, remarkable progress creates insightful questions.

> *What are equivalent formulations of conservative vector fields? When are irrotational vector fields necessarily conservative? How can you generalize the fundamental theorem of line integrals to non-conservative vector fields?*

The next section will begin to unravel some of these intriguing questions.

Exercises for Section 11.4

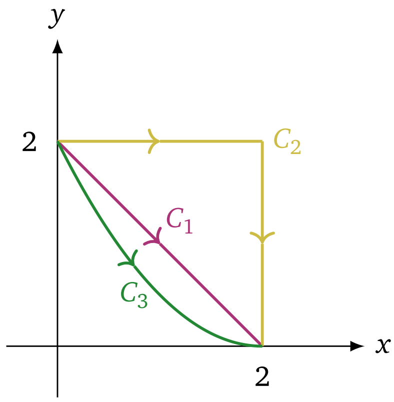
Concepts and definitions

11.4.1 Let $f(x,y) = x^2 + y^2$ and define $F = \nabla f$. Consider the three paths.

(a) Determine the work done by F along C_1, the straight line from $(0,2)$ to $(2,0)$.

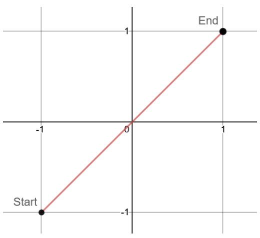
(b) Determine the work done by F along C_2, the parabola $(t, \frac{1}{2}(t-2)^2)$ from $(0,2)$ to $(2,0)$.

(c) Determine the work done by F along C_3, the street grid from $(0,2)$ to $(2,2)$ to $(2,0)$.

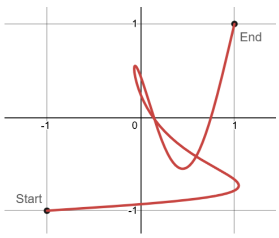


11.4.2 Let $f(x,y) = x^3 - ye^x$. For each of the curves C below, compute $\displaystyle\int_C \nabla f \cdot d\gamma$.

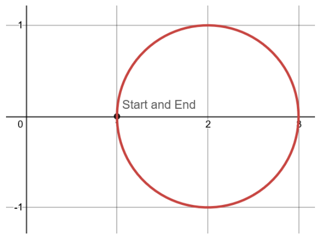
(a)



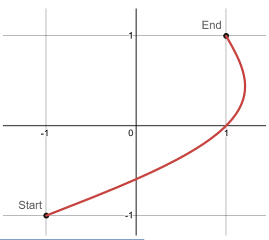
(c)



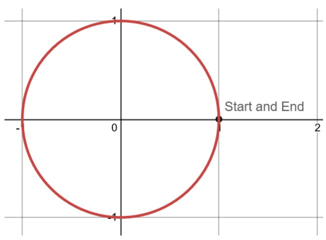
(e)



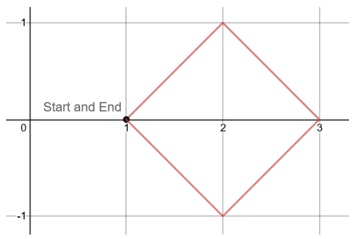
(b)



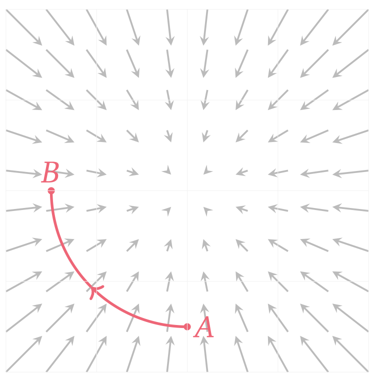
(d)



(f)



11.4.3 Consider the work done by a vector field F along a curve C from A to B, as illustrated below.



Assume $F = \nabla f$ for some scalar potential f. Explain why the diagram suggests $f(A) = f(B)$.

Computations

11.4.4 For each vector field, determine if it is irrotational.

(a) $F(x,y) = (y,x)$

(b) $G(x,y) = (-y,x)$

(c) $H(x,y) = (1+y, x+\cos y)$

(d) $J(x,y,z) = (yz^2, xz^2, 2xyz)$

11.4.5 For each vector field, determine if it is conservative. If so, find all possible potential functions.

(a) $F(x,y) = (y,x)$

(b) $G(x,y) = (-y,x)$

(c) $H(x,y) = (1+y, x+\cos y)$

(d) $J(x,y,z) = (yz^2, xz^2, 2xyz)$

11.4.6 Let $F(x,y) = (e^x\cos(y), e^x\sin(y))$.

(a) By direct calculation, find the work done by F along the boundary of the square $\partial[0,1]^2$ oriented counterclockwise.

(b) Is F conservative? Explain why or why not.

11.4.7 Joel, Ellie, and Tess are determining whether the vector field $F(x,y) = (2x+3y, 3x+5)$ is conservative and, if so, solve for all potentials. Each of them has a different argument.

(a) Joel gives the following argument.

> 1. *Assume there exists $f : \mathbb{R}^2 \to \mathbb{R}$ such that $F = \nabla f$ so*
>
> $$(1) \quad \frac{\partial f}{\partial x} = 2x+3y; \qquad (2) \quad \frac{\partial f}{\partial y} = 3x+5$$
>
> 2. *Integrating (1) with respect to x, it follows that $f(x,y) = x^2+3xy+C$ for some constant $C \in \mathbb{R}$.*
> 3. *Integrating (2) with respect to y, it follows that $f(x,y) = 3xy+5y+C$ for some constant $C \in \mathbb{R}$.*
> 4. *Since the solutions to (1) and (2) do not match, F is not conservative.*

Is Joel's argument correct and well-justified? If not, explain why not.

(b) Ellie gives the following argument.

> 1. *Assume there exists $f : \mathbb{R}^2 \to \mathbb{R}$ such that $F = \nabla f$ so*
>
> $$(1) \quad \frac{\partial f}{\partial x} = 2x+3y; \qquad (2) \quad \frac{\partial f}{\partial y} = 3x+5$$
>
> 2. *Integrating (1) with respect to x, note $f(x,y) = x^2+3xy+\phi(y)$ for some C^1 map $\phi : \mathbb{R} \to \mathbb{R}$.*
> 3. *Differentiating this expression, it follows by (2) that $3x+\phi'(y) = \frac{\partial f}{\partial y} = 3x+5$.*
> 4. *This implies $\phi'(y) = 5$ for all $y \in \mathbb{R}$ in which case $\phi(y) = 5y+C$ for some constant $C \in \mathbb{R}$.*
> 5. *Hence, a solution to (1) and (2) exists, so F is conservative.*
> 6. *Every potential of F is of the form $x^2+3xy+5y+C$ for some constant $C \in \mathbb{R}$.*

Is Ellie's argument correct and well-justified? If not, explain why not.

(c) Tess gives the following argument.

> 1. *Define* $f : \mathbb{R}^2 \to \mathbb{R}$ *by* $f(x, y) = x^2 + 3xy + 5y$.
> 2. *By direct calculation,* $\nabla f(x, y) = (2x + 3y, 3x + 5) = F(x, y)$ *for all* $(x, y) \in \mathbb{R}^2$.
> 3. *Thus,* F *is conservative.*
> 4. *Every potential of* F *is of the form* $x^2 + 3xy + 5y + C$ *for some constant* $C \in \mathbb{R}$.

Is Tess's argument correct and well-justified? If not, explain why not.

11.4.8 Jasmine, Mulan, and Tiana are deciding if the vector field $G(x, y) = (4x + 2y, 3x + 2)$ is conservative and, if so, solve for all potentials. Each of them has a different argument.

(a) Jasmine gives the following argument.

> 1. *Assume there exists* $g : \mathbb{R}^2 \to \mathbb{R}$ *such that* $G = \nabla g$ *so*
> $$(1) \quad \frac{\partial g}{\partial x} = 4x + 2y; \qquad (2) \quad \frac{\partial g}{\partial y} = 3x + 2$$
> 2. *Integrating (1) with respect to* x*, note* $g(x, y) = 2x^2 + 2xy + \phi(y)$ *for some* C^1 *map* $\phi : \mathbb{R} \to \mathbb{R}$.
> 3. *Differentiating this expression, it follows by (2) that* $2x + \phi'(y) = \frac{\partial g}{\partial y} = 3x + 2$.
> 4. *This implies* $\phi'(y) = x + 2$ *for all* $y \in \mathbb{R}$*, so* $\phi(y) = xy + 2y + C$ *for some constant* $C \in \mathbb{R}$.
> 5. *Hence, a solution to (1) and (2) exists, so* G *is conservative.*
> 6. *Every potential of* G *is of the form* $2x^2 + 3xy + 2y + C$ *for some constant* $C \in \mathbb{R}$.

Is Jasmine's argument correct and well-justified? If not, explain why not.

(b) Mulan gives the following argument.

> 1. *Assume there exists* $g : \mathbb{R}^2 \to \mathbb{R}$ *such that* $G = \nabla g$ *so*
> $$(1) \quad \frac{\partial g}{\partial x} = 4x + 2y; \qquad (2) \quad \frac{\partial g}{\partial y} = 3x + 2$$
> 2. *Integrating (1) with respect to* x*, note* $g(x, y) = 2x^2 + 2xy + \phi(y)$ *for some* C^1 *map* $\phi : \mathbb{R} \to \mathbb{R}$.
> 3. *Differentiating this expression, it follows by (2) that* $2x + \phi'(y) = \frac{\partial g}{\partial y} = 3x + 2$.
> 4. *This implies* $\phi'(y) = x + 2$ *for all* $y \in \mathbb{R}$*, which is a contradiction.*
> 5. *Hence,* G *is not conservative.*

Is Mulan's argument correct and well-justified? If not, explain why not.

(c) Tiana gives the following argument.

> 1. *Note that* $\frac{\partial}{\partial y}(4x + 2y) = 2 \neq 3 = \frac{\partial}{\partial x}(3x + 2)$.
> 2. *Thus,* $\partial_2 G_1 \neq \partial_1 G_2$ *so* G *is not irrotational.*
> 3. *Since* G *is not irrotational,* G *is not conservative.*

Is Tiana's argument correct and well-justified? If not, explain why not.

Proofs

11.4.9
(a) Prove Lemma 11.4.11 without looking at the textbook proof.

(b) Disprove the converse of Lemma 11.4.11. Search the textbook for a classic counterexample.

11.4.10 Irrotational vector fields are also called **curl-free** vector fields. The curl of a vector field in $\mathbb{R}^2$ and $\mathbb{R}^3$ has important an geometric meaning which you will explore soon. For now, you will study the algebraic definition and its relationship with irrotational vector fields.

(a) For a C^1 vector field $F = (F_1, F_2)$ in $\mathbb{R}^2$, define the **curl** of F to be the continuous real-valued function

$$\operatorname{curl}(F) = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}.$$

Let $U \subseteq \mathbb{R}^2$ be open. Verify that F is irrotational on U if and only if $\operatorname{curl}(F) = 0$.

(b) For a C^1 vector field $F = (F_1, F_2, F_3)$ in $\mathbb{R}^3$, define the **curl** of F to be the continuous vector field

$$\operatorname{curl}(F) = \Big(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\Big)$$

Let $U \subseteq \mathbb{R}^3$ be open. Verify that F is irrotational on U if and only if $\operatorname{curl}(F) = 0$ on U.

11.4.11 Asif shared a terribly written proof of the fundamental theorem of line integrals for curves.

1. *Assume C is a curve. Let $\gamma : [a, b] \to \mathbb{R}^n$ be a parametrization of C.*
2. *By assumption, $F = \nabla f$ on an open set containing C where f is C^1 and real-valued.*
3. *Then* $\displaystyle\int_C F \cdot d\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt$
4. $\displaystyle\implies \int_a^b F(\gamma(t)) \cdot \gamma'(t)dt = \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)dt$
5. $\displaystyle\implies \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)dt = \int_a^b \frac{d}{dt} f(\gamma(t))dt$
6. $\displaystyle\implies \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)dt = f(\gamma(b)) - f(\gamma(a))$

There is no error but the writing needs significant improvement.

(a) Each of Lines 3, 4, 5, and 6 should include one of the justifications (A), (B), (C), or (D) written below. Match the line to the justification.

(A) by the fundamental theorem of calculus
(B) by the chain rule
(C) by the definition of a line integral of a vector field
(D) by the assumption that $F = \nabla f$

(b) Justify the application of the chain rule.

(c) Justify the application of the fundamental theorem of calculus.

(d) Asif can greatly improve his proof writing. What tips would you give him?

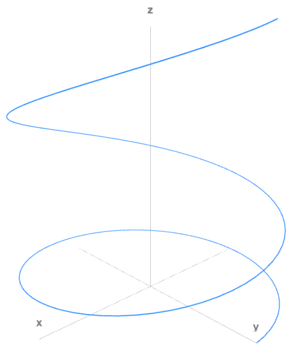
Applications and beyond

11.4.12 Consider an object of mass $M > 0$. Let G be the universal gravitational constant. By direct calculation, show that the gravitational field

$$F(x,y,z) = -\frac{GM(x,y,z)}{||(x,y,z)||^3}$$

has a potential function f given by $f(x,y,z) = \frac{GM}{||(x,y,z)||}$.

11.4.13 A ball is rolling down a helical track, following a path parametrized by $\gamma(t) = \left(3\sin(t), 3\cos(t), \frac{1}{2}\left(4 - \frac{t}{\pi}\right)^2\right)$.



(a) Express the work done by gravity on the ball as it rolls down the track as a single variable integral.

(b) Is the gravity force field F conservative? If so, find a potential function for F.

(c) If F is conservative, use the fundamental theorem of line integrals to calculate the work done by gravity.

11.4.14 The names "conservative vector field" and "potential function" have connections to physics. You will investigate one such connection: conservation of kinetic and potential energy. Suppose that $F : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field on $\mathbb{R}^3$. Suppose also that F represents a force and that this force acts on an object moving along a smooth curve C, parametrized by $\gamma : [a,b] \to \mathbb{R}^3$.

(a) The force acting on an object is equal to its mass multiplied by its acceleration. Write this in terms of m and γ.

(b) The kinetic energy of an object is equal to half its mass multiplied by the square of its speed. Write this in terms of m and γ.

(c) Calculate $\int_C F \cdot d\gamma$ from the definition and write it in terms of kinetic energy. *Hint:* If $f, g : [a,b] \to \mathbb{R}^n$ are differentiable, then $\frac{d}{dt}(f(t)\cdot g(t)) = f'(t)\cdot g(t) + f(t)\cdot g'(t)$.

(d) Let $P(x,y,z)$ be the potential energy of an object at (x,y,z). Then $-P$ is a potential function for F. Use this potential function to find another formula for $\int_C F \cdot d\gamma$ in terms of potential energy.

(e) Prove the law of conservation of energy: $K(\gamma(a)) + P(\gamma(a)) = K(\gamma(b)) + P(\gamma(b))$.

11.5. Conservative vector fields

The fundamental theorem of line integrals demonstrates that conservative vector fields $F = \nabla f$ and their potentials f possess unique and special properties. Algebraically speaking, the potential f acts like an antiderivative of the conservative vector field F. To further solidify your understanding, you can view conservative vector fields from another perspective.

How can you physically interpret conservative vector fields?

This viewpoint will connect to path-independence and conservation of work done. Informally speaking, the work done by a conservative vector field F between two points does not depend on the path taken. This means that the work done by F on a closed curve is always zero! Vector fields with these properties are referred to as *path-independent* vector fields, so conservative vector fields are necessarily path-independent. This suggests another question.

When are path-independent vector fields the same as conservative vector fields?

The answer will surprisingly be *always*! A similar question arose in the last section. You learned that conservative vector fields are always irrotational, so you may ask:

When are irrotational vector fields the same as conservative vector fields?

Irrotational vector fields are relatively easy to identify, so a positive answer would be wonderful. Unfortunately, Example 11.4.14 demonstrates that the answer is *not* always. This may be discouraging but you will see that there are common topological domains where these types of vector fields are identical. Your overall goal in this section will be to deepen your understanding of conservative vector fields and their equivalent characterizations.

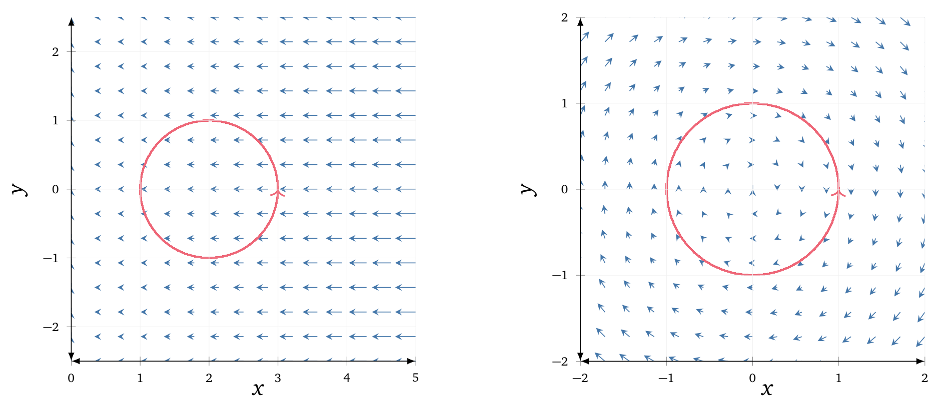
11.5.1 Physical viewpoints

Remember you have constructed an excellent model for calculating the work done by a vector field on a moving particle. Physicists study many different kinds of vector fields, such as electromagnetic or gravitational, but these can have wildly different properties. There is a central question that arises in those studies.

Does the work done by the vector field depend on the path taken by the particle?

If not, such vector fields are referred to as *path-independent*. Conservative vector fields are path-independent which gives elegant physical interpretations.

Example 11.5.1 A river is steadily flowing with velocity $F(x, y) = (-x, 0)$ at point $(x, y) \in \mathbb{R}^2$. A whirlpool is swirling with velocity $G(x, y) = (y, -x)$. These velocity fields are illustrated below with F on the left and G on the right.



For each velocity field, you swim in a counterclockwise loop in the first quadrant. What is the work done by each vector field?

First, consider the steady river and its velocity field F. When you are swimming against the flow, the river is doing negative work on you. However, if you are swimming with the flow, the river does positive work on you. Since the velocity field F is conservative with potential $f(x,y) = -\frac{x^2}{2}$, these two effects exactly cancel out after you swim in a closed loop.

Second, consider the whirlpool and its velocity field G. As you swim around in the counterclockwise circle, you will always be swimming against the river. Thus, the river will do strictly negative work on you. By the fundamental theorem of line integrals, G cannot be a conservative vector field.

This physical interpretation emphasizes the importance of path independence for conservative vector fields. You will next explore how deep this connection lies.

11.5.2 Path independence

Remarkably, a vector field is conservative if and only if it satisfies path independence.

Theorem 11.5.2 Let F be a vector field in $\mathbb{R}^n$ that is continuous on an open path-connected set $U \subseteq \mathbb{R}^n$. The following are equivalent:

(a) There exists C^1 real-valued function f such that $F = \nabla f$ on U.

(b) For any oriented piecewise curves C_1 and C_2 in U with the same start and end points,

$$\int_{C_1} F \cdot d\gamma = \int_{C_2} F \cdot d\gamma.$$

(c) For any closed piecewise curve C in U,

$$\int_C F \cdot d\gamma = 0.$$

Remark 11.5.3 Traditionally speaking, (a) defines gradient vector fields, (b) defines path-independent vector fields, and (c) defines conservative vector fields. This theorem shows that they are all equivalent, so you can interchangeably use the adjectives "gradient", "path-independent", and "conservative" to describe a vector field.

Theorem 11.5.2 is an absolute gem with brilliant applications to complex analysis, partial differential equations, fluid dynamics, and algebraic topology. While those subjects are beyond the scope of this text, you can still leverage these equivalent definitions to quickly identify conservative or non-conservative vector fields.

Example 11.5.4 The vector field

$$H(x,y) = \left(-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$$

from Example 11.4.14 is not conservative on $U = \mathbb{R}^2 \setminus \{(0,0)\}$ by Theorem 11.5.2 since $\int_C H \cdot T\, ds \neq 0$ where C is the counterclockwise unit circle in $\mathbb{R}^2$.

The proof of Theorem 11.5.2 combines several ingredients but only one implication requires some new ideas. As you shall see, (a) $\implies$ (b) or (a) $\implies$ (c) are immediate consequences of the fundamental theorem of line integrals. The proof that (b) $\iff$ (c) follows from basic properties of line integrals. The true challenge is proving (a) assuming either (b) or (c). For instance, how do you construct a potential function assuming only (b)? Here is the big idea.

Define a real-valued function using a line integral.

This construction will be satisfactory only by assuming either (b) or (c) holds.

Proof. It suffices to prove the following chain of implications:

$$(\text{a}) \implies (\text{c}) \implies (\text{b}) \implies (\text{a}).$$

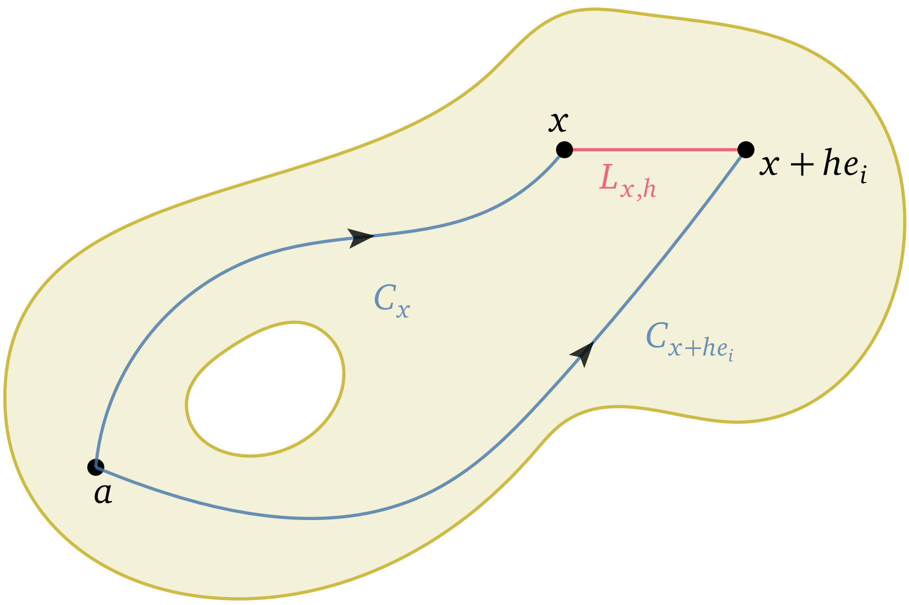
(a) $\implies$ (c): This follows from the fundamental theorem of line integrals (Theorem 11.4.1).
(c) $\implies$ (b): Assume (c). Let C_1 and C_2 be two oriented curves in U with the same endpoints. Since C_1 and C_2 start and stop at the same points, the concatenated curve $C_1+(-C_2)$ is an oriented closed piecewise curve inside U. Thus, by (c) and Lemma 11.3.16,

$$0=\int_{C_1+(-C_2)} F\cdot T\,ds=\int_{C_1} F\cdot T\,ds+\int_{-C_2} F\cdot T\,ds=\int_{C_1} F\cdot T\,ds-\int_{C_2} F\cdot T\,ds.$$

Rearranging the integrals gives (b).
(b) $\implies$ (a): Assume (b). Fix a point $a\in U$. For each $x\in U$, fix a piecewise oriented curve C_x from a to x lying inside U. Define the function $f:U\to\mathbb{R}$ by

$$f(x)=\int_{C_x} F\cdot T\,ds. \tag{11.5.1}$$

Fix $1\le i\le n$ and e_i be the ith standard basis vector in $\mathbb{R}^n$. There exists $\delta>0$ such that for all $h\in(-\delta,\delta)$, the line segment $L_{x,h}$ from x to $x+he_i$ lies inside U. The setup including the curve C_x, the curve C_{x+he_i}, and the line segment $L_{x,h}$ is illustrated below.



It follows that

$$\begin{aligned} f(x+he_i)&=\int_{C_{x+he_i}} F\cdot T\,ds\\ &=\int_{C_x} F\cdot T\,ds+\int_{L_{x,h}} F\cdot T\,ds=f(x)+\int_{L_{x,h}} F\cdot T\,ds. \end{aligned}$$

Parametrize the line segment $L_{x,h}$ with $\gamma(t) = x + te_i$ for $0 \le t \le h$. Thus, $\gamma'(t) = e_i$ and

$$\begin{aligned}\int_{L_{x,h}} F \cdot T\, ds &= \int_0^h \big(F(x+te_i)\cdot e_i\big)\, dt \\ &= \int_0^h F_i(x+te_i)dt \\ &= \int_0^h F_i(x_1,\ldots,x_{i-1},x_i+t,x_{i+1},\ldots,x_n)\, dt.\end{aligned}$$

Combining this with the previous identity, it follows by L'Hopital's rule that

$$\begin{aligned}\partial_i f(x) &= \lim_{h\to 0} \frac{f(x+he_i) - f(x)}{h} \\ &= \lim_{h\to 0}\left[\frac{1}{h}\int_0^h F_i(x+te_i)dt\right] \\ &= \lim_{h\to 0} F_i(x+he_i) = F_i(x).\end{aligned}$$

This proves that $F = \nabla f$, as required. ■

Remark 11.5.5 Notice the definition of the candidate potential function (11.5.1) was *independent* of the choice of curve C_x in U, because you are assuming the vector field is path independent. This feature is crucial to the proof and it is the big new idea.

This equivalent definition of conservative vector fields is marvelous from a theory standpoint. It builds a bridge between gradients and path independence. On the other hand, path independence is not computationally feasible to verify; how would you calculate the line integral along every closed curve? You are still searching for an approach better than solving some differential equations as in Example 11.4.8.

11.5.3 Irrotational vector fields and sets without holes

You have seen that every conservative vector field is irrotational (Lemma 11.4.11) but not every irrotational vector field is conservative (Example 11.4.14). This failure of the converse is disheartening since you could conceivably check whether a vector field is conservative by simply calculating some partial derivatives.

Example 11.5.6 Recall the vector fields

$$G(x,y) = (-y,x), \qquad H(x,y) = \Big(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\Big)$$

from Examples 11.4.9 and 11.4.14 respectively. Both G and H are not conservative, but G is not irrotational whereas H is irrotational. Their algebraic expressions do not seem that different except that the domain of G is $\mathbb{R}^2$ whereas the domain of H is $\mathbb{R}^2 \setminus \{(0,0)\}$. The domain of H has a hole! This subtle distinction will shockingly be the source of your problem.

Failure, as usual, has been quite productive. This example suggests an astounding observation.

The topology of the domain may affect the calculus of vector fields.

A lemma of Poincaré serves as evidence for this conjectural observation.

Lemma 11.5.7 (Poincaré's lemma) Let F be a vector field in $\mathbb{R}^n$ that is C^1 on the open set $U \subseteq \mathbb{R}^n$. If U is convex and F is irrotational on U, then F is conservative on U.

The proof below assumes a theorem on swapping differentiation and integration in one variable.

Proof. Without loss of generality, assume $0 \in U$. For each $x \in U$, let L_x be the line segment from 0 to x. Define the function $f : U \to \mathbb{R}$ by

$$f(x) = \int_{L_x} F \cdot T \, ds.$$

Parametrize L_x by $\gamma(t) = tx$ for $0 \le t \le 1$, so $\gamma'(t) = x$. By definition, for $x \in U$,

$$f(x) = \int_0^1 F(tx) \cdot x dt = \int_0^1 \sum_{j=1}^n F_j(tx) x_j dt$$

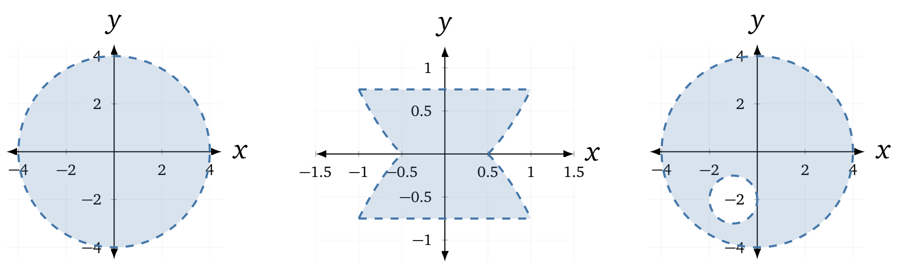
Fix $1 \le i \le n$. Taking the ith partial derivative ∂_i of both sides, it follows that

$$\begin{aligned}
\partial_i f(x) &= \int_0^1 F_i(tx) + \sum_{j=1}^n \partial_i F_j(tx) t x_j \, dt \\
&= \int_0^1 F_i(tx) + \sum_{j=1}^n \partial_j F_i(tx) t x_j \, dt && \text{since } F \text{ is irrotational} \\
&= \int_0^1 \frac{d}{dt}(t F_i(tx)) \, dt \\
&= F_i(x).
\end{aligned}$$

Hence, $\nabla f = F$ as required. ■

The assumption of *convexity* (see Section 2.7) excludes "holes" in your domain.

Example 11.5.8 Consider an open ball, an hourglass, and a washer. The open ball is convex, whereas the latter two regions are not.



The hourglass region does not have a hole, so convexity is a bit too restrictive.

Convex sets are perhaps a little too special. You would like to include more general sets.

How can you describe a "set without a hole"?

An exploration of the two-dimensional case is worthwhile.

Example 11.5.9 Intuitively speaking, the open ball and the hourglass in Example 11.5.8 do not have holes but the washer does. How can you explain this idea more precisely? Look at the hole in the washer. If you draw a simple closed curve C around the "hole", then the inside of C of your curve does *not* lie inside the washer. On the other hand, if you draw any simple closed curve C inside the hourglass (or the open ball), the inside of C must lie entirely inside the hourglass. This basic observation distinguishes these sets exactly as you want!

These observations suggest that you should define "holes" in $\mathbb{R}^2$ using the "inside" of simple closed curves. Simple closed curves in $\mathbb{R}^2$ should always have an inside and an outside. That seems obvious[9], right? Shockingly, no!

Theorem 11.5.10 (Jordan curve theorem) Let C be a simple closed curve in $\mathbb{R}^2$. Then C divides $\mathbb{R}^2$ into two regions: an open bounded region Ω (the inside) and an unbounded region $\mathbb{R}^2 \setminus \Omega$ (the outside). Moreover, the inside Ω is Jordan measurable and $\partial\Omega = C$.

Mathematicians did not even realize this statement required proof for many centuries! This deep theorem is surprisingly difficult to prove[10], and its history is filled with several incorrect attempts. You may take this theorem for granted and therefore attempt to better define two-dimensional sets "without holes".

Definition 11.5.11 A set $D \subseteq \mathbb{R}^2$ is a **simply connected domain** if D is open, path-connected, and, for every simple closed curve lying in D, its inside is a subset of D.

This intuitive definition classifies sets in $\mathbb{R}^2$ as you may want.

Example 11.5.12 Examples of simply-connected domains in $\mathbb{R}^2$ include any open disk, the upper half plane $\{(x, y) \in \mathbb{R}^2 : y > 0\}$, or the entire plane $\mathbb{R}^2$. Examples of open sets which are not simply-connected include any washer, the plane without a line, or the plane without the origin $\mathbb{R}^2 \setminus \{(0, 0)\}$.

It is substantially more difficult to define simply connected domains in higher dimensions. Instead, you will be satisfied with a basic intuition for the three-dimensional case.

*A set $D \subseteq \mathbb{R}^3$ is a **simply connected domain** if D is open, path-connected, and any simple closed curve can shrink continuously to a point while staying entirely inside D.*

This intuitive notion can be applied to explicit examples.

Example 11.5.13 Examples of simply-connected domains in $\mathbb{R}^3$ include any open ball, $\mathbb{R}^3$, $\mathbb{R}^3 \setminus B_1(0)$, or $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$. Examples of open sets which are not simply connected include 3-space minus a line $\mathbb{R}^3 \setminus \{(0, 0, z) : z \in \mathbb{R}\}$, or 3-space minus a plane $\mathbb{R}^3 \setminus \{(x, y, 0) : x, y \in \mathbb{R}\}$. Sketch these pictures and some simple closed curves to convince yourself.

After many years of persistence in mathematical education, these investigations will lead to a beautiful theorem.

Theorem 11.5.14 Let F be a vector field in $\mathbb{R}^2$ (or $\mathbb{R}^3$) that is C^1 on the open path-connected set D. If D is simply connected and F is irrotational on D, then F is conservative on D.

Proof. Omitted. This is far beyond the scope of this text. ■

Although you do not have rigorous definitions to understand the formal statement, you can at least appreciate its conclusion on a crude yet intuitive level.

*Irrotational vector fields are conservative if the vector field domain has "no **loop** holes".*

[9] That is one reason why you should not use the words "trivial" or "obvious" in proofs.

[10] Technically speaking, this formulation is not so absurdly hard, because you are always assuming the curve has a regular parametrization γ. Standard ideas from differential geometry can establish this case. However, if you assume γ is continuous (rather than C^1), then you need some algebraic topology, which is a graduate level course.

The word "loop" is emphasized because the simply connected sets cannot stop closed loops from shrinking. If you want to prove that a vector field is (or is not) conservative, then this theorem finally provides a powerful alternative to searching for potential functions.

Example 11.5.15 The vector field $F(x,y)=(x,y)$ is irrotational on $\mathbb{R}^2$. Since $\mathbb{R}^2$ is simply connected, Theorem 11.5.14 (or Lemma 11.5.7) implies that F is conservative on D.

You have wielded all of your tools with differential and integral calculus and applied them to study curves, especially in a physical context. You can calculate the length of a curve, find the mass of a wire, and determine the work done by a vector field. You have investigated conservative vector fields in $\mathbb{R}^n$ and discovered several equivalent formulations: the existence of a potential, path-independence, and work done along any closed curve is zero. Irrotational vector fields are sometimes the same as conservative vector fields, depending on the topological nature of the domain. Each of these ideas have deep physical and geometric meanings. These leaves open some significant questions.

> *What properties do non-conservative vector fields satisfy? How do they manifest geometrically? How can you physically interpret irrotational vector fields?*

A complete study of these questions in $\mathbb{R}^n$ is beautiful but geometrically daunting[11]. Since your physical reality is most naturally described in two or three dimensions, you will focus your attention on vector fields in $\mathbb{R}^2$ and $\mathbb{R}^3$ for the remainder of vector calculus. This will lead to spectacular theorems, and deepen your physical and geometric understanding of non-conservative vector fields.

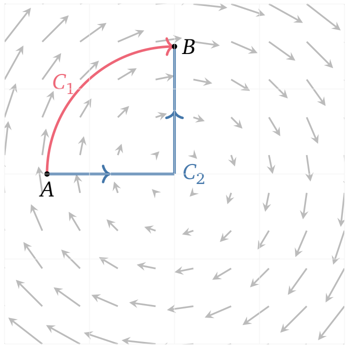
[11] See a course in differential geometry.

Exercises for Section 11.5

Concepts and definitions

11.5.1

The oriented curves C_1 and C_2 both start at A and end at B. Is the given vector field conservative? Explain why or why not using these curves.



11.5.2 Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuous. Which informal statement is equivalent to the phrase "F *is conservative on* $\mathbb{R}^n$"? If it is equivalent, identify the most similar statement in Theorem 11.5.2.

(a) F is a path independent vector field.

(b) F is a gradient vector field.

(c) F is an irrotational vector field.

(d) The work done by F along any curve is zero.

(e) The work done by F only depends on the endpoints of the path.

11.5.3 Which of these physical situations illustrate examples of a conservative vector field? *Hint:* Think about work done on a closed curve.

(a) A steady flow of lava.

(b) A whirlpool.

(c) The gravitational force field of the Earth.

11.5.4

(a) Define $F(x,y) = \left(-\dfrac{y}{x^2+y^2}, \dfrac{x}{x^2+y^2}\right)$ on $\mathbb{R}^2 \setminus \{(0,0)\}$. You have done some computations.

A) $\dfrac{\partial F_1}{\partial x} = \dfrac{2xy}{(x^2+y^2)^2} = -\dfrac{\partial F_2}{\partial y}$

B) $\dfrac{\partial F_2}{\partial x} = \dfrac{y^2-x^2}{(x^2+y^2)^2} = \dfrac{\partial F_1}{\partial y}$

C) $\displaystyle\int_C F \cdot d\gamma = 2\pi$ where C is the counterclockwise oriented unit circle centered at the origin.

D) $F(x,y) \neq \nabla f(x,y)$ for any C^1 function f on $\mathbb{R}^2 \setminus \{(0,0)\}$.

Which of these allow you to conclude whether or not F is irrotational? conservative? path-independent?

(b) Define $G(x,y) = \left(\dfrac{x}{x^2+y^2}, \dfrac{y}{x^2+y^2}\right)$ on $\mathbb{R}^2 \setminus \{(0,0)\}$. You have done some computations.

A) $\dfrac{\partial G_1}{\partial x} = \dfrac{y^2-x^2}{(x^2+y^2)^2} = -\dfrac{\partial G_2}{\partial y}$

B) $\dfrac{\partial G_2}{\partial x} = \dfrac{-2xy}{(x^2+y^2)^2} = \dfrac{\partial G_1}{\partial y}$

C) $\displaystyle\int_C G \cdot d\gamma = 0$ where C is *any* counterclockwise oriented circle centered at the origin.

D) $G(x, y) = \nabla \ln(x^2 + y^2)$

Which of these allow you to conclude whether or not G is irrotational? conservative? path-independent?

(c) Define $H(x, y) = (y, x)$ on $\mathbb{R}^2$. You have done some computations.

A) $\dfrac{\partial H_1}{\partial x} = 0 = \dfrac{\partial H_2}{\partial y}$

B) $\dfrac{\partial H_2}{\partial x} = 1 = \dfrac{\partial H_1}{\partial y}$

C) $\displaystyle\int_C H \cdot d\gamma = 0$ where C is a counterclockwise oriented unit circle centered at *any* point in $\mathbb{R}^2$.

D) $H(x, y) = \nabla(xy)$

Which of these allow you to conclude whether or not H is irrotational? conservative? path-independent?

(d) Does there exist a C^1 vector field $J : \mathbb{R}^2 \to \mathbb{R}^2$ with all of the following properties? If so, find it. If not, explain why not.

A) $\dfrac{\partial J_2}{\partial x} = 237 = \dfrac{\partial J_1}{\partial y}$

B) $\displaystyle\int_C J \cdot d\gamma = 0$ where C is any closed curve in $\mathbb{R}^2$.

C) J is not a gradient vector field.

11.5.5 Identify which sets in $\mathbb{R}^2$ are simply connected domains. Sketch them to help you guess.

(a) $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 9\}$

(b) $\{(x, y) \in \mathbb{R}^2 : 4 < x^2 + y^2 < 9\}$

(c) $\{(x, y) \in \mathbb{R}^2 : 4 < x^2 + y^2 < 9,\ y > 0\}$

(d) $\{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < 9\}$

(e) $\{(x, y) \in \mathbb{R}^2 : y > 0\}$

(f) $\{(x, y) \in \mathbb{R}^2 : y \neq 0\}$

(g) $\mathbb{R}^2$

(h) $\mathbb{R}^2 \setminus \{(0, 0)\}$

11.5.6 Identify which sets in $\mathbb{R}^3$ are simply connected domains. Sketch them to help you guess.

(a) $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < 9\}$

(b) $\{(x, y, z) \in \mathbb{R}^3 : 4 < x^2 + y^2 + z^2 < 9\}$

(c) $\{(x, y, z) \in \mathbb{R}^3 : 4 < x^2 + y^2 + z^2 < 9,\ z > 0\}$

(d) $\{(x, y, z) \in \mathbb{R}^3 : 0 < x^2 + y^2 + z^2 < 9\}$

(e) $\{(x, y, z) \in \mathbb{R}^3 : z > 0\}$

(f) $\{(x, y, z) \in \mathbb{R}^3 : z \neq 0\}$

(g) $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$

(h) $\mathbb{R}^3 \setminus \operatorname{span}\{(0, 0, 1)\}$

Computations

11.5.7 For each vector field below, do all of the following:

i) Determine if it is irrotational.

ii) Determine if you can apply Poincaré's lemma to conclude it is conservative.

iii) If possible, search for a potential function and decide if the vector field is conservative.

(a) $F(x, y, z) = \nabla(xyz)$

(b) $F(x, y, z) = \nabla \ln(xyz)$

(c) $F : \mathbb{R}^3 \to \mathbb{R}^3$ where $\operatorname{curl}(F) = (2, 3, 7)$.

(d) $F:\mathbb{R}^3 \to \mathbb{R}^3$ where $\operatorname{curl}(F) = (0,0,0)$.

(e) $F:U \to \mathbb{R}^3$ where $\operatorname{curl}(F) = (0,0,0)$ and $U = \{(x,y,z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\}$

(f) $F:U \to \mathbb{R}^3$ where $\operatorname{curl}(F) = (0,0,0)$ and $U = \mathbb{R}^3 \setminus \operatorname{span}\{(0,0,1)\}$

(g) $F(x,y,z) = (x,y,z)$

(h) $F(x,y,z) = (z,x,y)$

(i) $F(x,y,z) = \left(\frac{1}{x}, \frac{1}{y}, \frac{1}{z}\right)$

(j) $F(x,y,z) = \left(\frac{1}{x}, \frac{1}{y}, \frac{1}{xy}\right)$

Proofs

11.5.8 The textbook proves Theorem 11.5.2 by checking a minimal number of implications. Although it is not necessary, it is good practice to try and prove other implications.

(a) Prove that (a) implies (b) in Theorem 11.5.2.

(b) Prove that (b) implies (c) in Theorem 11.5.2.

11.5.9 Below is a proof that (b) implies (a) in Theorem 11.5.2. In other words, path-independent vector fields are gradient vector fields. As usual, important details are missing.

1. *Assume $F = (F_1, \ldots, F_n)$ is path-independent on the open path-connected set $U \subseteq \mathbb{R}^n$.*
2. *Fix $a \in U$. For each $x \in U$, let C_x be a curve from a to x lying inside U.*
3. *Define $f : U \to \mathbb{R}$ by $f(x) = \int_{C_x} F \cdot d\gamma$.*
4. *Fix $1 \le i \le n$ and let e_i be the ith standard basis vector in $\mathbb{R}^n$.*
5. *For $|h| > 0$ sufficiently small, the straight line $L_{x,h}$ from x to $x + he_i$ lies inside U.*
6. *Then $f(x+he_i) = \int_{C_{x+he_i}} F \cdot d\gamma = \int_{C_x} F \cdot d\gamma + \int_{L_{x,h}} F \cdot d\gamma = f(x) + \int_{L_{x,h}} F \cdot d\gamma$.*
7. *By a direct parametrization of the line, $\int_{L_{x,h}} F \cdot d\gamma = \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n)\,dt$.*
8. *It follows that*

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(x+he_i) - f(x)}{h} \tag{11.5.2}$$

$$= \lim_{h\to 0} \frac{1}{h} \int_{L_{x,h}} F \cdot d\gamma \tag{11.5.3}$$

$$= \lim_{h\to 0} \frac{1}{h} \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n)\,dt \tag{11.5.4}$$

$$= F_i(x). \tag{11.5.5}$$

9. *Therefore, $F = (F_1, \ldots, F_n) = (\partial_1 f, \ldots, \partial_n f) = \nabla f$ so F is conservative.*

(a) How and where does the proof use that U is path-connected?

(b) How and where does the proof use that U is open?

(c) How and where does the proof use that F is path-independent? *Hint:* There are 2 lines.

(d) Line 8 has justifications missing for each of the 4 numbered equations. Add them.

11.5.10 Irrotational vector fields are conservative provided the domain of your vector field has "no holes". One description for a domain with "no holes" is a convex domain. This special case is verified by Poincaré's lemma (Lemma 11.5.7). Below is an attempted proof.

1. *Without loss, assume U is star-shaped about the origin and F is irrotational on U.*
2. *For each $x \in U$ the straight line C_x from the origin to x lies inside U.*
3. *Define $f : U \to \mathbb{R}$ by $f(x) = \int_{C_x} F \cdot d\gamma$.*
4. *By a direct parametrization, $f(x) = \int_0^1 F(tx) \cdot x\,dt = \int_0^1 \sum_{i=1}^n F_i(tx)x_i dt$.*
5. *Fix $1 \le k \le n$. Taking the kth partial derivative, it follows that*

$$\frac{\partial f}{\partial x_k} = \int_0^1 \Big[\sum_{i=1}^n \frac{\partial F_i}{\partial x_k}(tx)tx_i + F_k(tx)\Big]dt \tag{11.5.6}$$

$$= \int_0^1 \Big[\sum_{i=1}^n \frac{\partial F_k}{\partial x_i}(tx)tx_i + F_k(tx)\Big]dt \tag{11.5.7}$$

$$= \int_0^1 \frac{d}{dt}\big[tF_k(tx)\big]dt \tag{11.5.8}$$

$$= F_k(x). \tag{11.5.9}$$

6. *Therefore, $F = (F_1, \ldots, F_n) = (\partial_1 f, \ldots, \partial_n f) = \nabla f$ so F is conservative.*

There are no serious errors, but many details and key justifications are missing.

(a) One step is missing some seriously important justification. Identify that step and explain what needs to be justified.

(b) Now, you can fill in smaller details. For line 3, why is f well-defined?

(c) For line 4, explicitly define the parametrization and check it gives the desired integral.

(d) For line 5, each numbered equation could use several brief details or justifications. Add them.

Applications and beyond

11.5.11 Pick one of the open path-connected sets from Exercise 11.5.5 which is not a simply connected domain. Draw a picture proof that the set is not simply connected.

11.5.12 Pick one of the open path-connected sets $S \subseteq \mathbb{R}^3$ from Exercise 11.5.6 which is not a simply connected domain. Give an example of a vector field on S which is irrotational but not conservative.

12. Fundamental theorems in 2D

In $\mathbb{R}^2$, to find the work done along a curve C by a vector field F, you have two options so far. First, you can directly parametrize $C \subseteq \mathbb{R}^2$ by $\gamma : [a, b] \to \mathbb{R}^2$ and calculate the line integral

$$\int_C F \cdot d\gamma = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt.$$

This almost always works in explicit examples but, as a mathematician, you are not only concerned with calculating. You want to search for deeper relationships about topics and, in this case, about line integrals.

Second, if your vector field is *conservative*, then you can apply the fundamental theorem of line integrals. That is, if the curve C starts at $p \in \mathbb{R}^2$ and ends at $q \in \mathbb{R}^2$ and $F = \nabla f$ on an open set containing C, then

$$\int_C \nabla f \cdot d\gamma = f(q) - f(p).$$

This miraculous theorem gives a fast calculation, has a deep physical meaning, and also generalizes the fundamental theorem of calculus in $\mathbb{R}$. Indeed, the scalar function f acts like an "antiderivative" for the vector F since $F = \nabla f$.

There is, however, a drawback to this second option: not all vector fields are conservative! This prompts an immediate question for your mathematical curiosity.

> *Can you generalize the fundamental theorem of line integrals in $\mathbb{R}^2$ to any vector field, including those which are not conservative?*

Green's theorem will be your incredible solution to this core question, and therefore be a flexible *third* option for calculating line integrals. Its discovery will unearth vital physical quantities describing 2D fluids. Most significantly, you will complete the mathematical story in $\mathbb{R}^2$ about the intimate relationship between integration and differentiation.

12.1. Circulation and curl in 2D

Before you attempt to generalize the fundamental theorem of line integrals, you need to search for more precise physical descriptions of vector fields $\mathbb{R}^2$. This will solidify and broaden your understanding of non-conservative vector fields. Think of a flowing river viewed from above. Below is a natural physical question.

How much does the water swirl around a loop (or at a point) and in what direction?

Your goal is to translate this heuristic physical notion into formal mathematical definitions with vector fields in $\mathbb{R}^2$. You will introduce *circulation* and *curl* for measuring swirly-ness.

12.1.1 Circulation in 2D

To measure swirly-ness along a curve, you can ask a more formal question.

How much does the vector field ***tangentially*** *flow along a closed curve?*

This returns you to the original concept of work done.

Definition 12.1.1 (Circulation) Let F be a vector field in $\mathbb{R}^2$ defined on an oriented piecewise curve C in $\mathbb{R}^2$. Assume C is closed. The **circulation of F along C** is the line integral

$$\int_C F \cdot T ds.$$

Remark 12.1.2 The integral symbol

$$\oint_C \quad \text{instead of} \quad \int_C$$

is often used to indicate that the curve C is closed. Both are acceptable.

In other words, the work done by F along a *closed* curve C is identical to the circulation of F along C. Computing circulation is the same as computing work done.

Example 12.1.3 To calculate circulation of $F(x,y) = (y,-x)$ along the unit circle C oriented counter clockwise, you can parametrize C by $\gamma(t) = (\cos(t),\sin(t))$ for $t \in [0,2\pi]$. If you plot the vector field and curve, you may guess that the circulation is negative because the particle always moves in the opposite direction of the vector field; this can be visually confirmed by viewing this Math3D demo. Alternatively, you can directly calculate the circulation as

$$\begin{aligned}\int_C F \cdot T ds &= \int_0^{2\pi} F(\gamma(t)) \cdot \gamma'(t) dt \\ &= \int_0^{2\pi} F(\cos(t),\sin(t)) \cdot (-\sin(t),\cos(t)) dt \\ &= \int_0^{2\pi} (\sin(t),-\cos(t)) \cdot (-\sin(t),\cos(t)) dt \\ &= -\int_0^{2\pi} 1 dt = -2\pi.\end{aligned}$$

12.1.2 Curl in 2D

To measure swirly-ness at a point, you can use two-dimensional curl.

Definition 12.1.4 Let $F = (F_1, F_2)$ be a C^1 vector field in $\mathbb{R}^2$. The **curl** of F is the continuous real-valued function

$$\operatorname{curl}(F) = \partial_1 F_2 - \partial_2 F_1$$

Remark 12.1.5 Curl will not be the same in three dimensions. The concept of two-dimensional curl is not standard; most sources will only define curl in three dimensions. You will utilize both and the notation will overlap, so pay attention to the context.

Remark 12.1.6 Notice that F is irrotational if and only if $\operatorname{curl}(F) = 0$. You can therefore equivalently refer to such vector fields as **curl-free**.

This definition seems totally mysterious. How does this combination of derivatives have any relation to swirly-ness? The true origin of curl lies in the following breakthrough.

Lemma 12.1.7 Let F be a vector field in $\mathbb{R}^2$. Fix $p \in \mathbb{R}^2$ in its domain. If F is C^1 on an open set containing p then

$$(\operatorname{curl} F)(p) = \lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{area}(B_\varepsilon(p))} \oint_{\partial B_\varepsilon(p)} (F \cdot T)\, ds,$$

where $\partial B_\varepsilon(p)$ is the circle of radius ε centered at p, oriented counterclockwise.

Proof. This is left as an exercise. There are several proofs. A challenging approach is to directly parametrize the line integrals and use a linear approximation for the vector field; this method will illuminate how curl arises. An easier approach is to apply Green's theorem which will be introduced in the next section; however, this method will not be as illuminating. ■

Informally speaking, this lemma says:

> *Curl is infinitesimal circulation (or circulation density).*

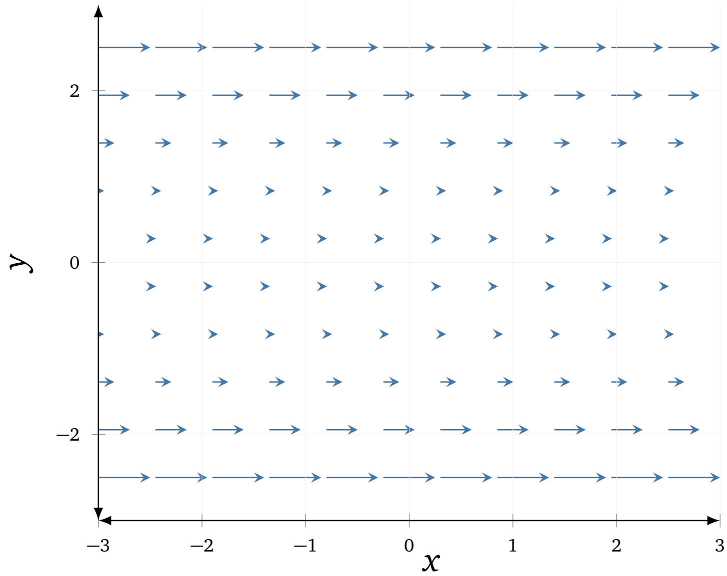
In other words, curl at a point p is a shrinking limit of counterclockwise circulation around p. You can view Lemma 12.1.7 as a variant of the fundamental theorem of calculus: a shrinking limit of integrals is a derivative. While the definition of curl has computational utility, this lemma provides a geometrically useful viewpoint of curl.

To find the signs of $(\operatorname{curl} F)(p)$ for some point p, imagine the vector field F represents the flow of some fluid, such as water. Lemma 12.1.7 suggests that the sign of $(\operatorname{curl} F)(p)$ should match the sign of the circulation for a small closed curve centred at p. This small closed curve can be viewed as a paddle dipped into the water.

> *If you hold a small paddle into the water at p, will the paddle turn counterclockwise or clockwise?*

If it turns counterclockwise, $(\operatorname{curl} F)(p) > 0$. If it turns clockwise, $(\operatorname{curl} F)(p) < 0$. If the paddle does not turn at all, $(\operatorname{curl} F)(p) = 0$. You can practice this idea in an example.

Example 12.1.8 Below is the vector field $F(x, y) = (y^2, 0)$



You can use this plot to guess the signs of curl(F) at some points.

- *Curl at* $(1,2)$ *is negative.* If you place a paddle at $(1,2)$, then the flow both above and below this point pushes rightward. However, the flow above this point is much stronger than below the point. Thus, your paddle must turn clockwise, so $(\operatorname{curl} F)(1,2) < 0$.
- *Curl at* $(1,-2)$ *is positive.* If you place a paddle at $(1,-2)$, then the flow both above and below this point pushes rightward. However, the flow below this point is much stronger than above the point. Thus, your paddle must turn counterclockwise so $(\operatorname{curl} F)(1,-2) > 0$.
- *Curl at* $(1,0)$ *is zero.* If you place a paddle at $(1,0)$, then the flow both above and below this point pushes rightward, and with roughly the same strength. The paddle is not inclined to turn, so $(\operatorname{curl} F)(1,0) \approx 0$.

Indeed, by direct computation, you can see that

$$(\operatorname{curl} F)(x,y) = \frac{\partial}{\partial x} 0 - \frac{\partial}{\partial y} y^2 = 0 - 2y = -2y$$

in which case $(\operatorname{curl} F)(1,2) = -4 < 0$, and $(\operatorname{curl} F)(1,-2) = 4 > 0$ and $(\operatorname{curl} F)(1,0) = 0$. This numerical analysis matches your heuristic analysis above.

This summarizes how to measure swirly-ness along a curve (circulation) and at a point (curl) in a mathematically rigorous manner. You will extend these notions in Section 12.3 to measure flow rate of a two-dimensional fluid. Before doing so, you can already apply this fresh intuition to discover a new fundamental theorem of vector calculus in $\mathbb{R}^2$, which will reveal a deep insight about non-conservative vector fields.

Exercises for Section 12.1

Concepts and definitions

12.1.1 Let F be a C^1 vector field in $\mathbb{R}^2$ defined on an open set $U \subseteq \mathbb{R}^2$. Which of the following is true?

(a) If F is curl-free, then F is irrotational.

(b) If F is irrotational, then F is curl-free.

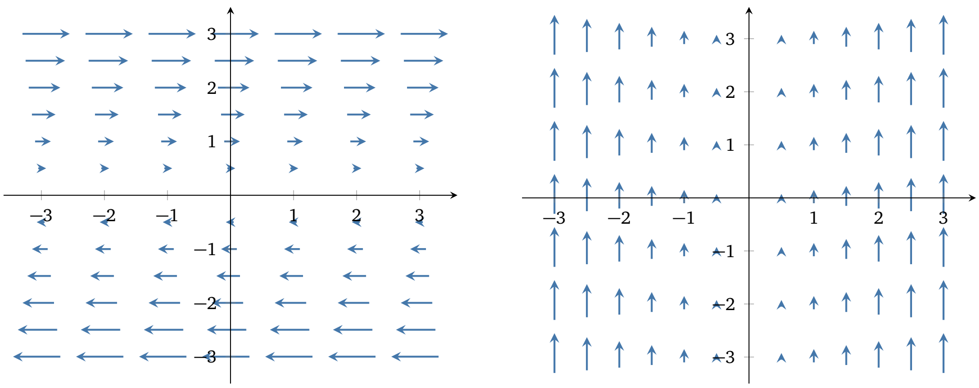
(c) If F is conservative, then F is curl-free.

(d) If F is curl-free, then F is conservative.

(e) If the circulation of F along every closed piecewise curve is zero, then F is conservative.

(f) If F is conservative, then the circulation of F along every closed piecewise curve is zero.

12.1.2 How do you interpret curl? Think of dipping a paddlewheel in a moving fluid. Does it spin clockwise or counterclockwise? That's the sign of curl. How fast does it spin? That is the magnitude of curl. Consider both vector fields $F : \mathbb{R}^2 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$.



$F(x,y)$ $G(x,y)$

For each quantity below, determine whether it is positive, negative, or zero.

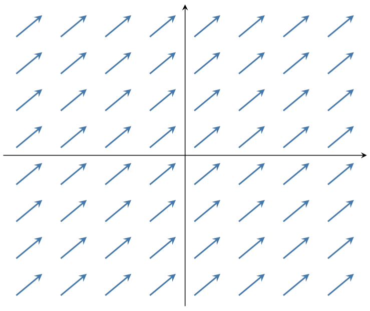
(a) The circulation along the counterclockwise curve $C = \partial[-2,2]^2$ for each vector field.

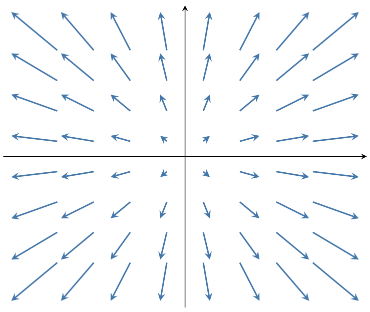
(b) The curl at $(-1,-1)$ for each vector field.

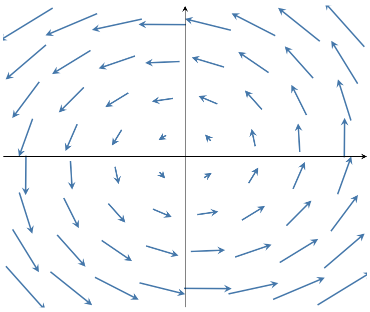
(c) The curl at $(1,1)$ for each vector field.

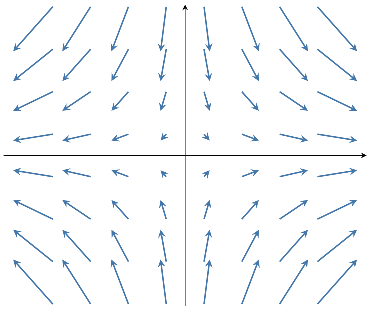
(d) The curl at $(1,0)$ for each vector field.

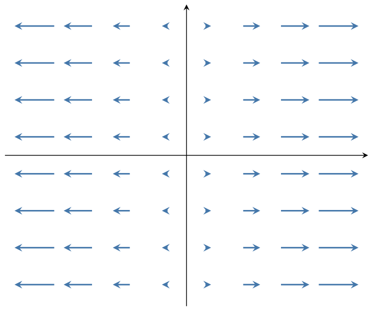
12.1.3 Identify which vector fields in $\mathbb{R}^2$ appear to be irrotational.

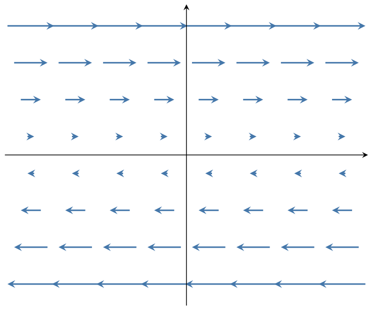
(a)


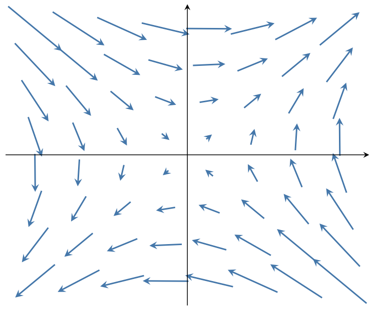
(b)


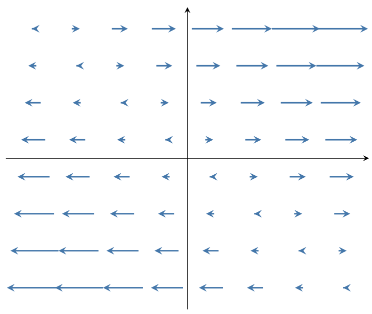
(c)


(d)


(e)


(f)


(g)


(h)


Computations

12.1.4 Identify which of the following vector fields in $\mathbb{R}^2$ are irrotational.

(a) $F_1(x,y) = (x,0)$

(b) $F_2(x,y) = (y,0)$

(c) $F_3(x,y) = (x,-y)$

(d) $F_4(x,y) = (x,y)$

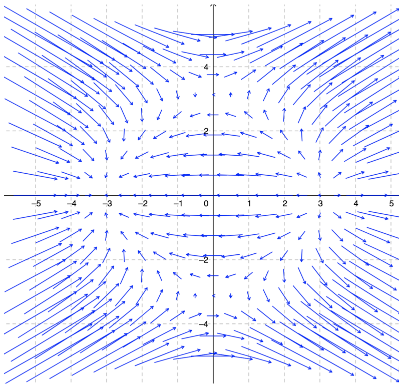
(e) $F_5(x,y) = (-y,x)$

(f) $F_6(x,y) = (y,x)$

(g) $F_7(x,y) = (1,1)$

(h) $F_8(x,y) = (x+y,0)$

12.1.5 Consider the $\mathbb{R}^2$ vector field $F(x,y) = (x^2 + y^2 - 10, xy)$.



(a) Without calculating, what is the sign of the curl at each point?

i) $(-2,-2)$ ii) $(-2,3)$ iii) $(4,0)$ iv) $(0,5)$

(b) Confirm your answers in (a) by computing $\text{curl}(F)$ at each point listed in (a) to (b).

(c) Is the work done by F along the straight line from $(3,-2)$ to $(3,-4)$ positive or negative? Do not perform any calculations.

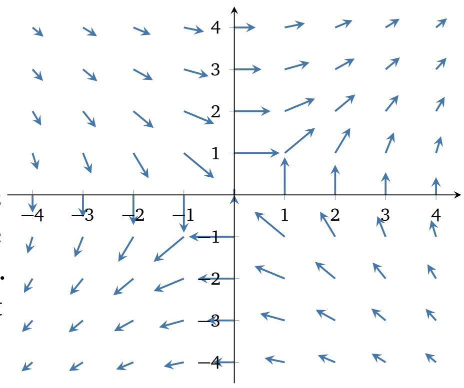
(d) Let

$$R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, y \geq 0\}$$

be the solid upper semicircle of radius 2. Let $C = \partial R$ be the counterclockwise oriented boundary of the region R. Calculate the circulation of F along C.

12.1.6 Consider a fluid flow in $\mathbb{R}^2$ defined by the vector field

$$F(x,y) = \left(\frac{y}{x^2+y^2+1}, \frac{x}{x^2+y^2+1}\right)$$



Let R be the solid rectangle with vertices $(0,0), (0,2), (3,0)$ and $(3,2)$. Let $C = \partial R$ be the counterclockwise oriented boundary of the rectangle. Use WolframAlpha to numerically evaluate any explicit integrals.

(a) Determine if the fluid is swirling clockwise or counterclockwise near $(-3,1)$.

(b) Compute the circulation of the fluid along C by direct calculation[1].

(c) Evaluate $\displaystyle\iint_R \text{curl}(F)\,dA$ by direct calculation[2]. How is this related to the previous part?

[1]In other words, don't use Green's theorem. Spoilers!

[2]Computer algebra systems may have a difficult time finding the exact value of this iterated double integral; if need be, use a numerical evaluation to get a decimal approximation.

12.2. Green's theorem and curl

You can now return to some core questions.

> *What properties do non-conservative vector fields in $\mathbb{R}^2$ satisfy? Can you generalize the fundamental theorem of line integrals for such vector fields?*

These motivations are phrased mathematically but any answers will have important physical ramifications. Remember the fundamental theorem of line integrals implies that conservative vector fields are path independent. By Theorem 11.5.2, the same cannot be true for non-conservative vector fields. This suggests you can try to better understand path *dependence*.

> *Can you relate the work done by two different paths C_1 and C_2 of a vector field in $\mathbb{R}^2$?*

Assuming C_1 and C_2 start and end at the same points, you can reformulate this question in terms of a closed curve $C = C_1 - C_2$.

> *Is there another way to calculate circulation of vector field along C in $\mathbb{R}^2$?*

This section is dedicated to the remarkable resolution of these problems: Green's theorem! This theorem holds for all vector fields, including non-conservative ones, so it is fundamental to vector calculus in $\mathbb{R}^2$ and hence the physical dynamics of fluid flows. The discovery of Green's theorem and its proof will be rooted in your developed physical intuition for circulation and work done in $\mathbb{R}^2$.

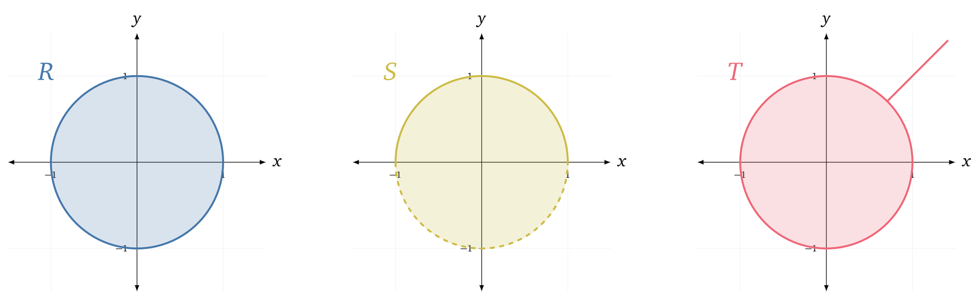
12.2.1 Regular regions and orienting the boundary

Before you can state Green's theorem, you will need a couple of key definitions. First, you will want to study circulation of curves in $\mathbb{R}^2$ that are the boundary of regions in $\mathbb{R}^2$. These are probably the types of regions that naturally come to mind, but you will need to be more topologically precise.

Definition 12.2.1 A set $R \subseteq \mathbb{R}^n$ is a **regular region** if R is a compact Jordan measurable set, and the closure of the interior of R is equal to R, i.e. $\overline{R^o} = R$.

A few examples will help clarify the purpose of this definition.

Example 12.2.2 Consider the three Jordan measurable sets R, S, T in $\mathbb{R}^2$ below.



The set R is regular since the closure of its interior, the open unit ball, is equal to R, the closed unit ball. The set S is not regular since it is not closed and hence not compact. The set T is compact but not regular; the closure of the interior of T, is equal to the closed unit ball R, i.e. $\overline{T^o} = R \neq T$. Geometrically, T has some "extra" boundary that does not touch its interior.

The notion of a regular region has a heuristic understanding.

A regular region is a compact set whose boundary always touches its interior.

These kinds of regions occur in the Jordan curve theorem (Theorem 11.5.10), namely the "inside" of a closed curve in $\mathbb{R}^2$. This notion of "inside" allows you to describe a consistent orientation for the boundary.

Definition 12.2.3 Let $R \subseteq \mathbb{R}^2$ be a regular region whose boundary ∂R is a finite disjoint union of piecewise curves. The boundary ∂R is **positively oriented** (resp.[3] **negatively oriented**) if the unit normal along each piecewise curve points outward away from R (resp. inward towards R). That is, the region always stays to the left (resp. right) as you traverse the boundary.

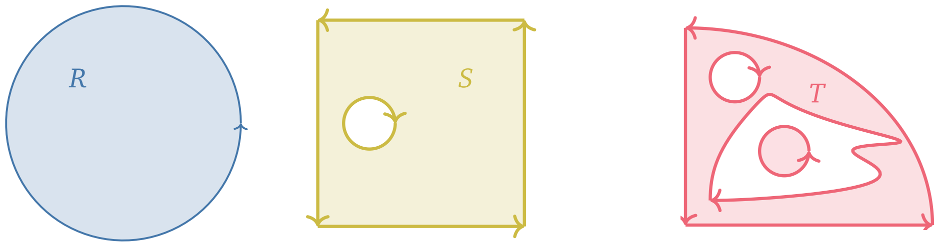
Remark 12.2.4 This definition is not completely formal[4], but this will suffice for your purposes. Your goal is to identify whether a boundary is positively oriented in explicit examples.

Notice the boundary must be positively oriented with respect to the region. Informally speaking,

*The boundary of a regular region is **positively oriented** if the region always stays to the left as you traverse the boundary.*

This informal description of the definition can applied in some examples.

Example 12.2.5 Illustrated below are three regular regions R, S, T in $\mathbb{R}^2$. Each boundary is a finite disjoint union of piecewise curves which have been oriented according to the indicated arrows. All of them are *positively oriented*.



For the disk R, its interior is to the left if you travel the boundary in the indicated direction. Hence, this boundary is positively oriented.

For the set S, its boundary is a disjoint union of two piecewise curves. For the interior of the set to be kept to the left as you travel the outer square, you must travel *counterclockwise*. To keep the interior to the left as you travel the inner circle, you must travel *clockwise*.

For the set T, travelling about the outer boundary, you have to move counterclockwise to keep the interior to the left. For both the upper circle and the complicated curve, you must travel clockwise to keep the boundary to the left. Finally, for the circle bounded inside the complicated curve, you must once again travel counterclockwise.

Equipped with regular regions and oriented boundaries, you can unravel Green's theorem.

12.2.2 Statement and proof

One of your main goals is to find an alternate method for calculating circulation. The critical insight is Lemma 12.1.7, namely that curl is *infinitesimal* circulation.

[3] The abbreviation "resp." stands for "respectively". It is commonly used in mathematics to define two very similar concepts in parallel rather than repeating the same sentences twice.

[4] See a course on curves and surfaces, or a course on differential geometry for a more rigorous definition.

If F is a C^1 vector field in $\mathbb{R}^2$, then for any point $p \in \mathbb{R}^2$ in its domain

$$(\operatorname{curl} F)(p) = \lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{area}(B_\varepsilon(p))} \oint_{\partial B_\varepsilon(p)} F \cdot T \, ds,$$

where $\partial B_\varepsilon(p)$ is the boundary of ε-disk centred at p oriented counterclockwise.

With this golden nugget and some other ingenious ideas, you can discover Green's theorem.

Theorem 12.2.6 (Green's theorem – curl form) Let F be a vector field in $\mathbb{R}^2$ that is C^1 on a regular region $R \subseteq \mathbb{R}^2$. If the boundary ∂R is a finite disjoint union of positively oriented piecewise curves, then

$$\oint_{\partial R} (F \cdot T) \, ds = \iint_R \operatorname{curl}(F) \, dA.$$

Remark 12.2.7 Be careful to apply Green's theorem only with a positive orientation of the boundary. If $C = \partial R$ is negatively oriented, then you can instead apply Green's theorem using its positively oriented counterpart $-C$.

The curl form of Green's theorem can be succinctly summarized in a heuristic manner.

The total infinitesimal circulation over R is the circulation along its boundary ∂R.

This statement can be translated more informally.

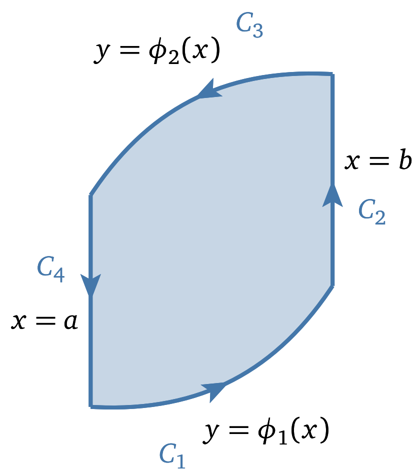
The total amount of swirly-ness inside is the amount of swirl along the edge.

Before reading the proof, you are encouraged to watch this 8 minute video on deriving the curl form of Green's theorem; it will give a flavour of the main ideas. The proof strategy has two main stages. First, you will prove Green's theorem for special types of regular regions R, namely regular regions $R \subseteq \mathbb{R}^2$ which are both x-simple and y-simple . Second, you will reduce the general case to this special case.

Proof. (Sketch) First, assume the regular region R is both x-simple and y-simple (see Definition 9.2.5). That is, there exists continuous functions $\phi_1 : [a,b] \to \mathbb{R}, \phi_2 : [a,b] \to \mathbb{R}, \psi_1 : [c,d] \to \mathbb{R}, \psi_2 : [c,d] \to \mathbb{R}$ such that

$$\begin{aligned} R &= \{(x,y) \in \mathbb{R}^2 : a \le x \le b, \phi_1(x) \le y \le \phi_2(x)\} \\ &= \{(x,y) \in \mathbb{R}^2 : c \le y \le d, \psi_1(y) \le x \le \psi_2(y)\}. \end{aligned}$$

Begin with the x-simple description, as illustrated below.



Labelling the oriented curves C_1, C_2, C_3, C_4 as in the above diagram, the positively oriented boundary of R satisfies $\partial R = C_1 + C_2 + C_3 + C_4$. By directly parametrizing each curve,

$$\begin{aligned}\oint_{\partial R}(F_1,0)\cdot T\,ds &= \int_{C_1}(F_1,0)\cdot T\,ds + \int_{C_2}(F_1,0)\cdot T\,ds + \int_{C_3}(F_1,0)\cdot T\,ds + \int_{C_4}(F_1,0)\cdot T\,ds \\ &= \int_a^b F_1(x,\phi_1(x))dx + 0 - \int_a^b F_1(x,\phi_2(x))dx - 0 \\ &= \int_a^b F_1(x,\phi_1(x)) - F_1(x,\phi_2(x))dx.\end{aligned}$$

Notice that the line integrals of $(F_1, 0)$ along C_2 and C_4 were both zero because they have unit tangents $T = (0,1)$ along the entire curve. By the fundamental theorem of calculus,

$$F_1(x,\phi_1(x)) - F_1(x,\phi_2(x)) = \int_{\phi_1(x)}^{\phi_2(x)} -\partial_2 F_1(x,y)\,dy$$

and hence

$$\oint_{\partial R}(F_1,0)\cdot T\,ds = \int_a^b\int_{\phi_1(x)}^{\phi_2(x)} -\partial_2 F_1(x,y)\,dy\,dx = \iint_R -\partial_2 F_1\,dA.$$

Similarly, by viewing ∂R using the y-simple form, it follows that

$$\begin{aligned}\oint_{\partial R}(0,F_2)\cdot T\,ds &= \int_c^d F_2(\psi_2(y),y) - F_2(\psi_1(y),y)dy \\ &= \int_c^d\int_{\psi_1(y)}^{\psi_2(y)} \partial_1 F_2(x,y)\,dxdy = \iint_R \partial_1 F_2\,dA\end{aligned}$$

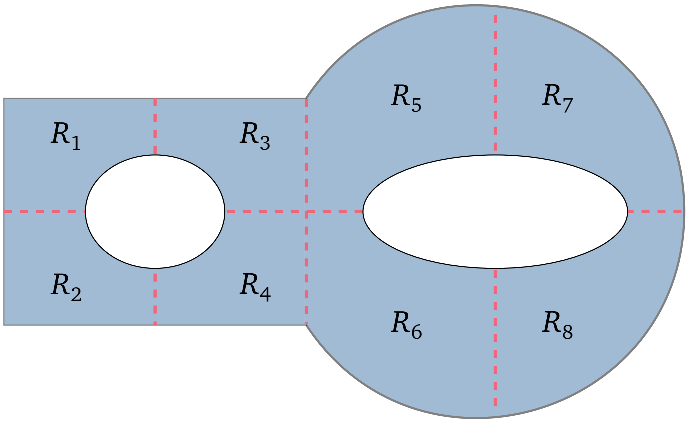
by the fundamental theorem of calculus. Adding both of these identities, you get that

$$\begin{aligned}\oint_{\partial R} F\cdot T\,ds &= \oint_{\partial R}(F_1,0)\cdot T\,ds + \oint_{\partial R}(0,F_2)\cdot T\,ds \\ &= \iint_R \partial_1 F_2 - \partial_2 F_1\,dA \\ &= \iint_R (\operatorname{curl} F)\,dA\end{aligned}$$

as required. This proves Green's theorem assuming R is both x-simple and y-simple.

To prove Green's theorem for any regular region R, you can decompose R into a finite union of regular regions $R_1, \ldots, R_N$ where each R_i is both x-simple and y-simple. An illustration[5] of this decomposition is below.

[5] Based on a figure in Holden [11].



You can then apply Green's theorem to each regular region R_i and add up the circulations. From the diagram, you can see that the circulation along the *interior* edges will all cancel! This occurs because the work done along the edge is counted twice: once in each direction. The only remaining pieces will be the edges counted exactly once, which necessarily lie on the boundary of R. This strategy is emulated in the following sequence of equalities.

$$\iint_R \operatorname{curl} F\, dA = \sum_{i=1}^{N} \iint_{R_i} \operatorname{curl} F\, dA = \sum_{i=1}^{N} \oint_{\partial R_i} F \cdot T\, ds = \oint_{\partial R} F \cdot T\, ds.$$

This conveys the main steps to the complete the proof of Green's theorem. ■

Phew! Now that was a serious proof with a magical finale of cancellation. Next, you can explore some calculations with Green's theorem to reveal its power.

12.2.3 Examples with Green's theorem in curl form

First, you can calculate circulation of non-conservative vector fields.

Example 12.2.8 Consider the vector field $F(x, y) = (y\sin(2x)e^{\sin^2(x)}, 3x^2 + e^{\sin^2(x)})$. Suppose you want to compute the circulation

$$\oint_{\partial R} (F \cdot T) ds$$

where ∂R is the positively oriented boundary of the rectangle $R = [0,1] \times [1,2]$. Directly parametrizing this curve will result in a nightmarish line integral that will appear impossible to compute. Moreover, you cannot use the fundamental theorem of line integrals since F is not conservative. Indeed, you can verify by direct calculation that

$$(\operatorname{curl} F)(x, y) = \frac{\partial}{\partial x}(3x^2 + e^{\sin^2(x)}) - \frac{\partial}{\partial y}(y\sin(2x)e^{\sin^2(x)}) = 6x,$$

so F is not irrotational and hence not conservative. Despite these setbacks, you now have a third option. Since R is a rectangle and hence a regular region and its boundary ∂R is a positively oriented piecewise curve, it follows by Green's theorem and some basic integral calculations that

$$\oint_{\partial R} (F \cdot T)ds = \iint_R \operatorname{curl}(F)dA = \iint_R 6x\ dA = \int_1^2 \int_0^1 6x\ dxdy = 3.$$

Second, you can calculate circulation for conservative vector fields in three different ways.

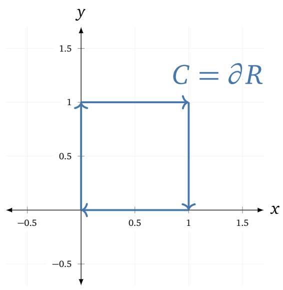
Example 12.2.9 (Three conservative methods) You have accumulated a number of methods for evaluating line integrals of conservative vector fields. Here, you will apply each of these methods to evaluate the same integral. Define the vector field

$$F(x,y) = (e^x \cos(y), 1 - e^x \sin(y))$$

Let $C = \partial R$ be the *negatively* oriented boundary of the rectangle $R = [0,1]^2$. You can calculate the circulation of F along C in a few different ways.

First, you can calculate a line integral directly from the definition. Notice that ∂R is parametrized by the following map

$$\gamma(t) = \begin{cases} (t,1) \text{ for } t \in [0,1] \\ (1,2-t) \text{ for } t \in [1,2] \\ (3-t,0) \text{ for } t \in [2,3] \\ (0,t-3) \text{ for } t \in [3,4] \end{cases}$$



By definition of line integral, it follows that

$$\begin{aligned}
\oint_{\partial R} F \cdot T\ ds &= \int_0^4 F(\gamma(t)) \cdot \gamma'(t)dt \\
&= \int_0^1 F(t,1) \cdot (1,0)dt + \int_1^2 F(1,2-t) \cdot (0,-1)dt \\
&\quad + \int_2^3 F(3-t,0) \cdot (-1,0)dt + \int_3^4 F(0,t-3) \cdot (0,1)dt \\
&= \int_0^1 e^t \cos(1)dt + \int_1^2 -1 - e\sin(t-2)dt + \int_2^3 -e^{3-t}dt + \int_3^4 1 - \sin(t-3)dt \\
&= (e\cos(1) - \cos(1)) + (e - e\cos(-1) - 1) + (1-e) + (1 + \cos(1) - 1) \\
&= 0.
\end{aligned}$$

Hence, the circulation of F along C is zero. That was a tedious yet straightforward calculation.

Second, you can apply the fundamental theorem of line integrals (Theorem 11.4.1) provided F is conservative. You verified in Example 11.4.8 that $F(x,y) = \nabla f(x,y)$ where $f(x,y) = (y + e^x \cos(y))$. Since C is a piecewise curve, it follows by the fundamental theorem of line integrals that

$$\oint_{\partial R} F \cdot d\gamma = f(\gamma(4)) - f(\gamma(0)) = f(0,1) - f(0,1) = 0.$$

This outcome is expected since F is conservative and C is closed.

Third, you can apply your newly minted Green's Theorem to solve this problem. However, notice that ∂R is actually negatively oriented, as our parametrization travels about it clockwise (thus unit normal must point into R). This implies that $-\partial R$ must be positively oriented. Thus, by Green's theorem, it follows that

$$\begin{aligned}\oint_{\partial R} F \cdot T\, ds = -\oint_{-\partial R} F \cdot T\, ds &= -\iint_R \operatorname{curl} F\, dA \\ &= -\iint_R \partial_1(1 - e^x \sin(y)) - \partial_2(e^x \cos(y))\, dA \\ &= -\iint_R 0\, dA = 0.\end{aligned}$$

Unsurprisingly, all three methods yield the same answer: zero.

Finally, you can address one of this section's motivating questions.

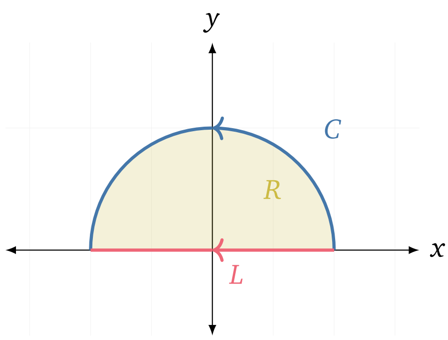
Can you relate the work done by two different paths in $\mathbb{R}^2$?

Indeed you can! Green's theorem is the bridge.

Example 12.2.10 (Moving the curve) Let C be the oriented semicircular curve in $\mathbb{R}^2$ parametrized by $\gamma(t) = (\cos t, \sin t)$ for $0 \le t \le \pi$. You want to compute the work done by the vector field $F(x, y) = (y - 2, -x + 2\sin(y^2))$ along C, that is, you want to evaluate the line integral $\int_C F \cdot T\, ds$. What methods can you try?

You can try computing this line integral directly, but it will create a messy integral that is impossible to evaluate. Moreover, F is not conservative as $\operatorname{curl} F(x, y) = -2$, so you cannot use the fundamental theorem of line integrals. Also, C is not closed so you cannot directly use Green's theorem. This eliminates some natural choices.

There is one more option! You can *move the curve*. Although C is not closed, you can "close it off" with another well-chosen curve. For instance, let $L \subseteq \mathbb{R}^2$ be the straight line segment from $(1, 0)$ to $(-1, 0)$, then the region R between the curves has positively oriented boundary $\partial R = C - L$.



Thus, the piecewise curve $C - L = \partial R$ is a positively oriented boundary enclosing the regular region $R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1, y \ge 0\}$. Thus, by Green's Theorem and properties of line integrals,

$$\int_C F \cdot T\, ds - \int_L F \cdot T\, ds = \oint_{C-L} F \cdot T\, ds = \iint_R (\operatorname{curl} F)\, dA.$$

This relates the work done along C and the work done along L! Parametrizing L by $\varphi(t) =$

$(-t, 0)$ for $-1 \leq t \leq 1$, you can verify that

$$\begin{aligned}\int_C F \cdot T\,ds &= \int_L F \cdot T\,ds + \iint_R (\operatorname{curl} F)\,dA \\ &= \int_{-1}^{1} F(\varphi(t)) \cdot \varphi'(t)dt + \iint_R -2\,dA \\ &= \int_{-1}^{1} (-2, t) \cdot (-1, 0)dt - 2 \cdot \operatorname{area}(R) \\ &= \int_{-1}^{1} 2dt - \pi = 4 - \pi.\end{aligned}$$

The concept of *moving a curve* via Green's theorem is a new and brilliant idea! It allows you to relate the work done between two paths by a double integral over the curl. This deep idea will have many applications beyond simplifying some calculations. Green's theorem in curl form has provided an incredible property to relate work done by non-conservative vector fields. This theorem is so miraculous that you will be able to apply its philosophy and ideas in many different mathematical and physical contexts. Indeed, these methods for describing the swirly-ness of a fluid will next be extended to describe the flow rate of a fluid.

Exercises for Section 12.2

Concepts and definitions

12.2.1 Which of the following are regular regions in $\mathbb{R}^2$?

(a) $\{(x,y)\in\mathbb{R}^2 : x^2+y^2\le 2\}$

(b) $\{(x,y)\in\mathbb{R}^2 : x^2+y^2< 2\}$

(c) $\{(x,y)\in\mathbb{R}^2 : 1\le x^2+y^2\le 2\}$

(d) $\mathbb{R}^2$

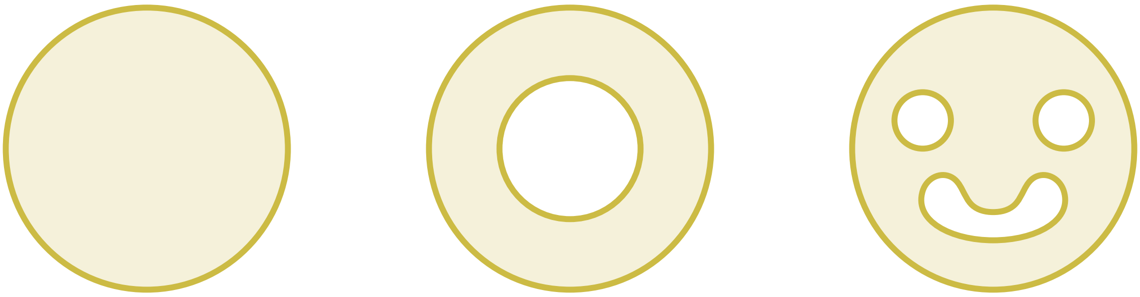
(e) $[-1,1]^2\cup[2,4]^2$

(f) $[-1,1]^2\cup\{(2,0)\}$

(g) $[-1,1]^2\cup\{(x,0): 1\le x\le 2\}$

(h) $[-1,1]^2\cup\overline{B_1(2,0)}$

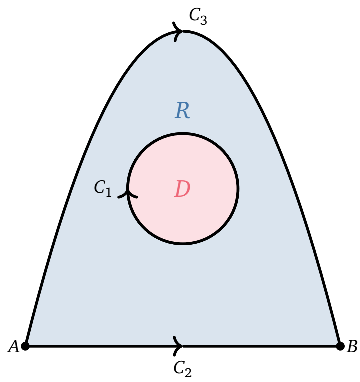
12.2.2 Below are three regular regions. Orient their boundaries with the positive orientation.



12.2.3 Consider the regular regions R and D in $\mathbb{R}^2$, the *oriented* curves C_1, C_2, C_3, and the points A and B in $\mathbb{R}^2$. Note the interiors of R and D are disjoint, that is,

$$R^o\cap D^o=\varnothing.$$

Let $F:\mathbb{R}^2\to\mathbb{R}^2$ be a C^1 vector field.



(a) Express each **positively oriented** boundary below in terms of the curves C_1, C_2, C_3.
i) ∂D ii) $\partial(R\cup D)$ iii) ∂R

(b) If $\displaystyle\iint_R \operatorname{curl}(F)dA = 20$ and $\displaystyle\iint_D \operatorname{curl}(F)dA=23$, then compute (if possible) the following:

i) $\displaystyle\int_{C_1} F\cdot T\,ds$ ii) $\displaystyle\int_{C_2} F\cdot T\,ds$ iii) $\displaystyle\int_{C_2} F\cdot T\,ds+\int_{C_3} F\cdot T\,ds$ iv) $\displaystyle\int_{C_2} F\cdot T\,ds-\int_{C_3} F\cdot T\,ds$

(c) If the work done by F along C_3 equals 3 and the total infinitesimal circulation of F on $R\cup D$ equals 7, then (if possible) find the work done by F along C_2.

(d) If $\displaystyle\iint_R \operatorname{curl}(F)dA=5$ and $\displaystyle\oint_{C_1} F\cdot T\ ds=11$ then compute (if possible) $\displaystyle\oint_{\partial(R\cup D)} F\cdot T\ ds.$

Computations

12.2.4 Let $F(x,y)=(-y,x)$ and C be the positively oriented circle of radius 2 at the origin. Calculate the work done by F to push an object along C in two different ways.

(a) Directly parametrize C.

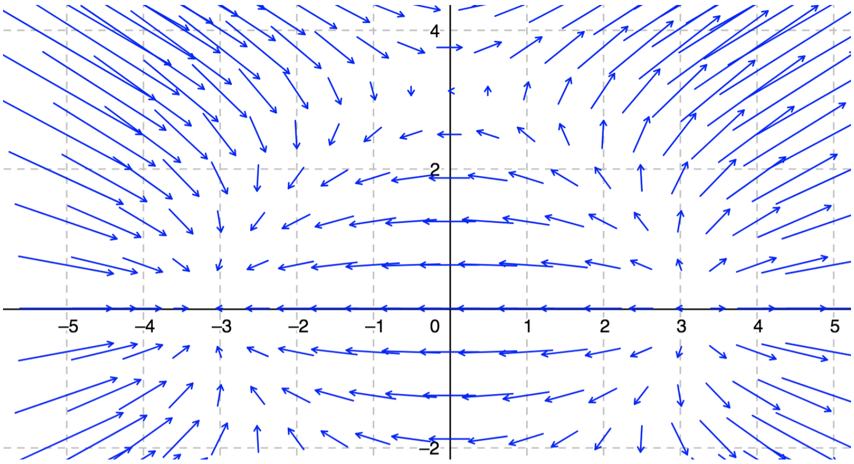
(b) Use Green's theorem.

12.2.5 Recall the vector field $F(x,y) = (x^2 + y^2 - 10, xy)$ in $\mathbb{R}^2$ from worksheet H6 Question 3. Let

$$R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, y \geq 0\}$$

and let $C = \partial R$ be the positively oriented boundary of this regular region R.

(a) Sketch the curve C and region R on the vector field. Make the orientation of C clear.



(b) Calculate the circulation of F along C using Green's theorem. *Hint:* This should be short.

(c) Calculate the work done by F to move a particle from $(-2,0)$ to $(2,0)$ along the x-axis.

(d) Use the previous two parts to find the work done by F to move a particle from $(-2,0)$ to $(2,0)$ along the upper semicircle $x^2 + y^2 = 4, y \geq 0$. Do not parametrize. *Hint:* This is short.

12.2.6 Let R be the region bounded by $y = 2x$ and $y = x^2/2$ and positively orient its boundary $C = \partial R$. Calculate

$$\oint_C (\cos x + y)dx + (x^2 + 2e^y)dy.$$

12.2.7 Compute

$$\oint_C (e^{\arctan x} - 3y)dx + (\sqrt{4+y^2} - x)dy$$

where $C = \partial[-6,-2]^2$ is positively oriented.

12.2.8 Use Green's theorem to determine the work done by $F(x,y) = (x - y, y)$ to push a particle along the curve $y = \sin(x)$ from $(\pi, 0)$ to $(0,0)$. *Hint:* Move the curve to the x-axis.

12.2.9 Let $R \subseteq \mathbb{R}^2$ be a regular region whose boundary ∂R is a finite disjoint union of positively oriented piecewise curves. Express the area of R as

$$\oint_{\partial R} P\,dx + Q\,dy$$

for at least 3 different choices of vector fields $F = (P, Q)$.

12.2.10 Find the area of the region bounded by the epicycloid $\gamma(t) = (5\cos(t) - \cos(5t), 5\sin(t) - \sin(5t))$ for $0 \leq t \leq 2\pi$.

Proofs

12.2.11 Use Green's theorem to prove that if F is conservative on an open set $U \subseteq \mathbb{R}^2$ and R is a regular region inside U then F has zero circulation along ∂R.

12.2.12 Curl is the "circulation density", i.e. the local swirl of a fluid at a point. This concept is formalized by Lemma 12.1.7, which can be proved with or without Green's theorem. Below is an incomplete proof with Green's theorem.

1. *For* $\varepsilon > 0$*, Green's theorem implies that* $\oint_{\partial B_\varepsilon(p)} (F \cdot T)\, ds = \iint_{\overline{B_\varepsilon(p)}} (\operatorname{curl} F) dA.$
2. *Dividing both sides by* $\operatorname{area}(B_\varepsilon(p)) = \pi\varepsilon^2$ *and taking* $\varepsilon \to 0^+$*, it follows that*

$$\lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{area}(B_\varepsilon(p))} \oint_{\partial B_\varepsilon(p)} (F \cdot T)\, ds = \lim_{\varepsilon \to 0^+} \frac{1}{\pi\varepsilon^2} \iint_{\overline{B_\varepsilon(p)}} (\operatorname{curl} F) dA.$$

3. *The quantity in the righthand limit is the average value of* $\operatorname{curl} F$ *over the closed ball* $\overline{B_\varepsilon(p)}$.
4. *As* $\varepsilon \to 0^+$*, this average value approaches* $(\operatorname{curl} F)(p)$.

One line requires more justification than the others. Identify the line and add the details.

12.2.13 The proof of Green's theorem in your readings was incomplete so, alas, your heart must be broken. It only proves the case for regions which are **both** x-simple and y-simple. The remaining ideas are illustrated in this example with your heart $R \subseteq \mathbb{R}^2$ which has been broken into 4 pieces. Putting together the pieces will complete both your heart and the remaining ideas behind the proof.

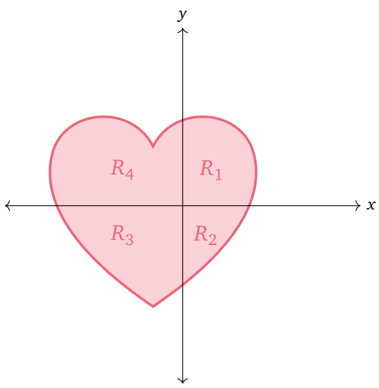
(a) Which of the 4 pieces R_1, R_2, R_3, R_4 are x-simple?

(b) Which of the 4 pieces R_1, R_2, R_3, R_4 are y-simple?

(c) Describe how your heart has been broken into 4 pieces. Be mathematical, not emotional.

(d) How can you break your heart so all of the regions are both x-simple and y-simple?

(e) If your heart were broken into pieces which are both x-simple and y-simple, you can complete the proof of Green's theorem and mend your soul. Explain why.



12.3. Flux and divergence in 2D

As you did with swirly-ness in Section 12.1, you can continue to search for more precise physical descriptions of vector fields $\mathbb{R}^2$. Think again of a flowing river viewed from above.

How much water flows outside a loop (or through a point)?

Your goal is again to translate these heuristic physical notions into formal mathematical definitions with vector fields in $\mathbb{R}^2$. You will introduce *flux* and *divergence* for measuring flow rate. The story will closely parallel circulation and curl, so pay attention to the similarities. Before you begin, you will need to formally describe what it means to "flow outward".

12.3.1 Unit normal in 2D

A vector "crosses" an oriented closed curve C in $\mathbb{R}^2$ if it points orthogonally to the direction of motion, namely the unit tangent T. This orthogonal direction will be referred to as the unit normal of C. In $\mathbb{R}^2$, you have two different choices of unit vectors that will be orthogonal to T, so you will want to pick a consistent direction that *depends only on the orientation of the curve* C. Ordered bases from linear algebra permit a standard choice.

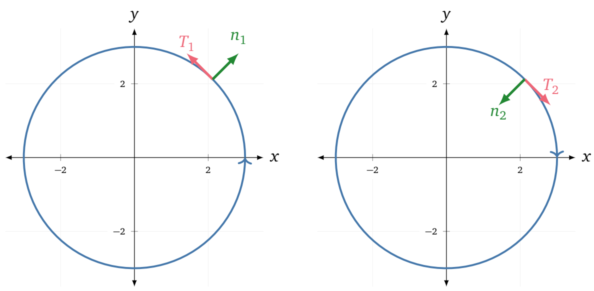
Definition 12.3.1 Let C be an oriented closed curve in $\mathbb{R}^2$ parametrized by $\gamma : [a, b] \to \mathbb{R}^2$ with unit tangent vector T. The **unit normal** of C is the continuous function $n : [a, b] \to \mathbb{R}^2$ such that for every $t \in (a, b)$, the output $n(t)$ is a unit vector orthogonal to $T(t)$ and $\{n(t), T(t)\}$ is a positively-oriented ordered basis for $\mathbb{R}^2$.

Remark 12.3.2 Recall $\{u, v\}$ is a positively-oriented ordered basis of $\mathbb{R}^2$ if it can be mapped to the standard basis $\{e_1, e_2\}$ by a linear transformation with positive determinant.

Remark 12.3.3 Section 1.1 introduces a different definition of unit normal N for parametrized curves in $\mathbb{R}^n$. This unit normal N is defined using a derivative of the unit tangent T, and it is *not* the same as the definition of the unit normal n for oriented curves in $\mathbb{R}^2$. This should not usually cause confusion, but you will need to pay attention to the context.

You can verify that the direction of the unit normal, the direction of the unit tangent, and the orientation of C are interlinked; if you specify one of them, then the rest are fixed. The unit circle in $\mathbb{R}^2$ will be the classic example for remembering these relationships.

Example 12.3.4 Let C_1 be a radius 3 circle centered at the origin oriented counterclockwise. Let C_2 be the same circle oriented clockwise, so $C_2 = -C_1$. The lefthand figure illustrates C_1 and its unit tangent T_1 and unit normal n_1. The righthand figure below illustrates C_2 and its unit tangent T_2 and unit normal n_2.



Observe that $\{n_1, T_1\}$ and $\{n_2, T_2\}$ are both positively-oriented ordered bases in $\mathbb{R}^2$. You can confirm these visuals with direct calculations.

Notice C_1 is parametrized by $\gamma_1(t) = (3\cos(t), 3\sin(t))$ for $t \in [0, 2\pi]$. You can find the unit tangent and unit normal to C. By definition, the unit tangent is given by

$$T_1(t) = \frac{\gamma_1'(t)}{\|\gamma_1'(t)\|} = \frac{(-3\sin(t), 3\cos(t))}{\sqrt{9\cos^2(t) + 9\sin^2(t)}} = \frac{1}{3}(-3\sin(t), 3\cos(t)) = (-\sin(t), \cos(t))$$

By definition, the unit normal $n_1 : [0, 2\pi] \to \mathbb{R}^2$ satisfies three key properties:

- $T_1(t) \cdot n_1(t) = 0$ for all $t \in [0, 2\pi]$.
- $\|n(t)\| = 1$ for all $t \in [0, 2\pi]$.
- $\{n_1(t), T_1(t)\}$ is a positively oriented ordered basis of $\mathbb{R}^2$.

Since $T_1(t) = (-\sin t, \cos t)$, the first relation implies that there exists $\alpha \in \mathbb{R}$ such that $n_1(t) = \alpha(\cos(t), \sin(t))$. Since $\|n(t)\| = |\alpha| = 1$, it must be that either $\alpha = \pm 1$. Since the basis must be positively oriented, it follows that $\alpha = 1$. Overall, the unit normal is

$$n_1(t) = (\cos(t), \sin(t)).$$

This confirms exactly what the lefthand figure above illustrates.

On the other hand, the clockwise oriented circle C_2 can be parametrized by $\gamma_2(t) = (3\cos(t), -3\sin(t))$ for $t \in [0, 2\pi]$. You can similarly verify that $T_2(t) = (-\sin(t), -\cos(t))$ and $n_2(t) = (-\cos(t), \sin(t))$. This also confirms exactly what the righthand figure above illustrates; the computational details are left as an exercise.

The process for computing the unit normal has a standard formula.

Lemma 12.3.5 Let C be an oriented curve in $\mathbb{R}^2$ parametrized by $\gamma : [a, b] \to \mathbb{R}^2$. Write $\gamma(t) = (x(t), y(t))$ for $a \le t \le b$. The unit tangent T and unit normal n of C are given by

$$T(t) = \frac{(x'(t), y'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \quad \text{and} \quad n(t) = \frac{(y'(t), -x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \quad \text{for } a < t < b.$$

Proof. This is left as an exercise. Follow the same ideas as in Example 12.3.4. ■

Equipped with the unit normal of a curve in $\mathbb{R}^2$, you can resume your study of flow rate.

12.3.2 Flux in 2D

Line integrals in $\mathbb{R}^2$ measure the *tangential flow* along a curve, so they are sometimes referred to as *tangential* line integrals. To measure the flow rate across a curve in $\mathbb{R}^2$, you cannot use tangential line integrals because the question fundamentally differs.

*How much does the vector field **orthogonally** flow across a curve?*

You require a new kind of line integral which is unique to $\mathbb{R}^2$.

Definition 12.3.6 (Normal line integral) Let F be a vector field in $\mathbb{R}^2$ defined on an oriented curve C in $\mathbb{R}^2$. If C is parametrized by $\gamma : [a, b] \to \mathbb{R}^2$, then the **normal line integral of F along C** is given by

$$\int_C F \cdot n \, ds := \int_a^b F(\gamma(t)) \cdot n(t) \|\gamma'(t)\| dt.$$

Remark 12.3.7 The letter n in the integrand is always the unit normal of C. Remember that the direction of n is automatically specified from the orientation of C.

Normal line integrals enjoy the same properties as tangential line integrals, such as invariance under parametrization (Theorem 11.3.1) and orientation reversal, linearity, and additivity (Lemma 11.3.16). For the sake of brevity, these properties are not formally stated here but you may freely use them.

Now, recall circulation is the same as the tangential line integral along closed curves. Correspondingly, flux is the same as the normal line integral along closed curves.

Definition 12.3.8 (Flux) Let F be a vector field in $\mathbb{R}^2$ defined on an oriented curve C in $\mathbb{R}^2$. Assume C is closed. The **flux of F across** C is the normal line integral of F along C, namely,

$$\oint_C F \cdot n\,ds.$$

Remark 12.3.9 Note that **outward flux** corresponds to a curve C oriented counterclockwise; similarly, **inward flux** corresponds to a curve C oriented clockwise.

Notice circulation measures the total amount of tangential force whereas flux measures the total amount of a normal force. Computing flux is similar to work done.

Example 12.3.10 Let C be the positively oriented boundary of the upper half disk

$$\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9, y \geq 0\}.$$

You can decompose C into the two parts. Let C_1 be the top half of the radius 3 circle oriented from right to left. Then C_1 is parametrized by $\gamma_1(t) = (3\cos(t), 3\sin(t))$ for $t \in [0,\pi]$. Let C_2 be the straight line which closes this semi-circle, so C_2 is parametrized by $\gamma_2(t) = (t,0)$ for $t \in [-3,3]$. The tangent vector along $C_1 + C_2$ is always pointing counterclockwise, so $C = C_1 + C_2$ is a positively oriented closed piecewise curve.

You can calculate the flux of $F(x,y) = (e^y, 2y - x^2)$ across $C = C_1 + C_2$ because

$$\oint_{C_1+C_2} F \cdot n ds = \int_{C_1} F \cdot n_1 ds + \int_{C_2} F \cdot n_2 ds.$$

Since C is positively oriented, the line integral $\oint_C F \cdot n\,ds$ is the outward flux across C. Thus, the unit normal for the semicircle C_1 will have a positive y component and the unit normal of the line segment C_2 will have a negative y component.

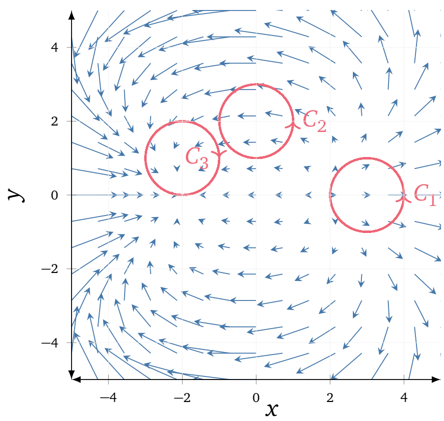
Following Example 12.3.4, you can verify that C_1 has unit normal $n_1(t) = (\cos(t), \sin(t))$ with parametrization $\gamma_1(t) = (3\cos t, 3\sin t)$ for $0 \leq t \leq \pi$. Note that $\|\gamma_1'(t)\| = 3$. Similarly, since C_2 is just a straight line along the horizontal axis, the unit normal is given by $n_2(t) = (0,-1)$ with parametrization $\gamma_2(t) = (t,0)$ for $-3 \leq t \leq 3$. Additionally, you can verify that $\|\gamma_2'(t)\| = 1$ as $\gamma_2'(t) = (1,0)$.

Now, you have everything to set up the flux integral. You can verify it follows that

$$\begin{aligned}
\oint_C F \cdot n ds &= \int_0^\pi F(\gamma_1(t)) \cdot n_1(t) \|\gamma_1'(t)\| dt + \int_{-3}^3 F(\gamma_2(t)) \cdot n_2(t) \|\gamma_2'(t)\| dt \\
&= \int_0^\pi F(3\cos(t), 3\sin(t)) \cdot (\cos(t), \sin(t))(3) dt + \int_{-3}^3 F(t,0) \cdot (0,-1)(1) dt \\
&= 3\int_0^\pi e^{3\sin(t)}\cos(t) + 6\sin^2(t) - 9\cos^2(t)\sin(t) dt + \int_{-3}^3 t^2 dt = 9\pi.
\end{aligned}$$

The flux of a vector field F measures how much F aligns with the unit normal n of a curve C. If C is oriented counterclockwise (i.e. outward unit normal), flux is positive when F has a net flow outward, and negative when F has a net flow inward.

Example 12.3.11 What is the sign of the flux of the vector field F across C_1 or C_2 or C_3? You can analyze each curve individually.



- Since C_1 is oriented counterclockwise, the unit normal points outward, so $\int_{C_1} F \cdot n \, ds$ measures outward flux. The flux of F across C_1 is therefore positive, because the vectors entering the curve are smaller in magnitude than those exiting.
- Since C_2 is oriented counterclockwise, $\int_{C_2} F \cdot n \, ds$ again measures outward flux. The flux of F across C_2 is roughly zero, because the vectors entering this curve are roughly the same magnitude as those exiting.
- Since C_3 is oriented clockwise, $\int_{C_3} F \cdot n \, ds$ measures inward flux. The flux of F across C_3 is therefore positive, because any non-tangential vectors are always pointing inward.

Overall, this implies that $\int_{C_1} F \cdot n \, ds > 0$, $\int_{C_2} F \cdot n \, ds \approx 0$, and $\int_{C_3} F \cdot n \, ds > 0$.

12.3.3 Divergence in 2D

To measure flow through a point, you can use two-dimensional divergence.

Definition 12.3.12 Let $F = (F_1, F_2)$ be a C^1 vector field in $\mathbb{R}^2$. The **divergence** of F is the continuous real-valued function

$$\operatorname{div}(F) = \partial_1 F_1 + \partial_2 F_2.$$

Remark 12.3.13 Common equivalent notation includes $\nabla \cdot F$. Unlike curl, divergence has essentially the same definition in any dimension. Namely, if G is a C^1 vector field in $\mathbb{R}^n$, then the **divergence of** G is the continuous real-valued function

$$\operatorname{div}(G) = \partial_1 G_1 + \cdots + \partial_n G_n.$$

Remark 12.3.14 A point $p \in \mathbb{R}^2$ is a **source** if $\operatorname{div}(F)(p) > 0$ and a **sink** if $\operatorname{div}(F)(p) < 0$.

Remark 12.3.15 A vector field F is **sourceless** if $\operatorname{div}(F) = 0$.

As with curl (Definition 12.1.4), this formal definition appears unmotivated. The true origin of divergence is similar to curl (Lemma 12.1.7).

Lemma 12.3.16 Let F be a vector field in $\mathbb{R}^2$. Fix $p \in \mathbb{R}^2$ in its domain. If F is C^1 on an open set containing p then

$$(\operatorname{div} F)(p) = \lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{area}(B_\varepsilon(p))} \oint_{\partial B_\varepsilon(p)} (F \cdot n) \, ds,$$

where $\partial B_\varepsilon(p)$ is the positively oriented circle of radius ε centered at p.

Proof. This is left as an exercise. The possible proofs are similar to Lemma 12.1.7. ■

Informally speaking, this lemma says:

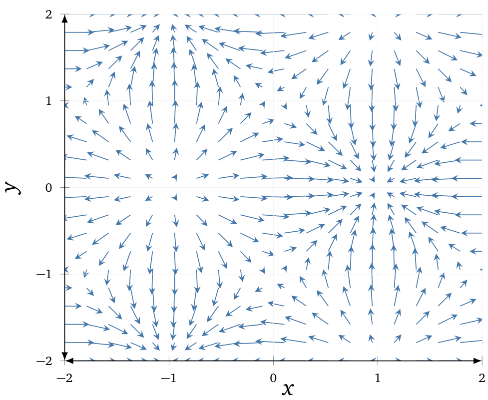
Divergence is infinitesimal flux (or flux density).

In other words, divergence at a point p is a shrinking limit of *outward* flux around p. The definition of divergence has computational utility, whereas this lemma provides a physically useful viewpoint of divergence.

Imagine the vector field represents a flowing body of water. For any point p in this body, $(\operatorname{div} F)(p) > 0$ if p is a source. In other words, there seems to be more water emerging from p than draining into it. Similarly, $(\operatorname{div} F)(p) < 0$ if p is a sink. In other words, there seems to be more water draining into p than emerging from it. Otherwise, $(\operatorname{div} F)(p) = 0$.

Example 12.3.17 Below is the plot of the vector field $F(x,y) = (\cos(\frac{\pi}{2}x)\cos(\frac{\pi}{2}y), -\sin(\frac{\pi}{2}x)\sin(\frac{\pi}{2}y))$. You can use this plot to guess the signs of divergence at various points.

- *Divergence at* $(-1,0)$ *is positive*. The point $(-1,0)$ has water flowing out from it in all directions. Then it must be a source of water so $(\operatorname{div} F)(-1,0) > 0$.
- *Divergence at* $(1,0)$ *is negative*. The point $(1,0)$ has water flowing into it from all directions. Then, it must be a sink of water with $(\operatorname{div} F)(1,0) < 0$.
- *Divergence at* $(-1,1)$ *is zero*. The point $(-1,1)$ has vectors of the same magnitude flowing out as flowing in. Also, the direction of these vectors is the same while passing through $(-1,1)$. Thus, this point cannot be a source or sink, so $(\operatorname{div} F)(-1,1) \approx 0$.
- *Divergence at* $(1,\frac{1}{2})$ *is negative*. The vectors flowing into this point have a slightly greater magnitude than those flowing out. Hence, $(1,\frac{1}{2})$ is a sink so $(\operatorname{div} F)(1,\frac{1}{2}) < 0$.



You can verify this heuristic approach numerically. By definition of divergence,

$$(\operatorname{div} F)(x,y) = \frac{\partial}{\partial x}\cos(\tfrac{\pi}{2}x)\cos(\tfrac{\pi}{2}y) + \frac{\partial}{\partial y}\left(-\sin(\tfrac{\pi}{2}x)\sin(\tfrac{\pi}{2}y)\right) = -\pi\sin(\tfrac{\pi}{2}x)\cos(\tfrac{\pi}{2}y)$$

so $(\operatorname{div} F)(-1,0) = \pi$, $(\operatorname{div} F)(1,0) = -\pi$, $(\operatorname{div} F)(-1,1) = 0$, and $(\operatorname{div} F)(1,\frac{1}{2}) = -\pi\frac{\sqrt{2}}{2}$.

This completes a thorough analysis of the physical and geometric properties of vector fields in $\mathbb{R}^2$. Combined with Section 12.1, you can rigorously describe swirly-ness and flow rate, whether along a curve or at a point. This formal mathematical language is incredibly useful in applications such as fluid dynamics and for contextualizing this enormous theory. As your concluding application of the chapter, you will reformulate Green's theorem in terms of divergence and flux.

Exercises for Section 12.3

Concepts and definitions

12.3.1 Let F be a C^1 vector field in $\mathbb{R}^2$ defined on an open set $U \subseteq \mathbb{R}^2$. Which of the following is true?

(a) If F is divergence-free, then F is irrotational.

(b) If F is irrotational, then F is divergence-free.

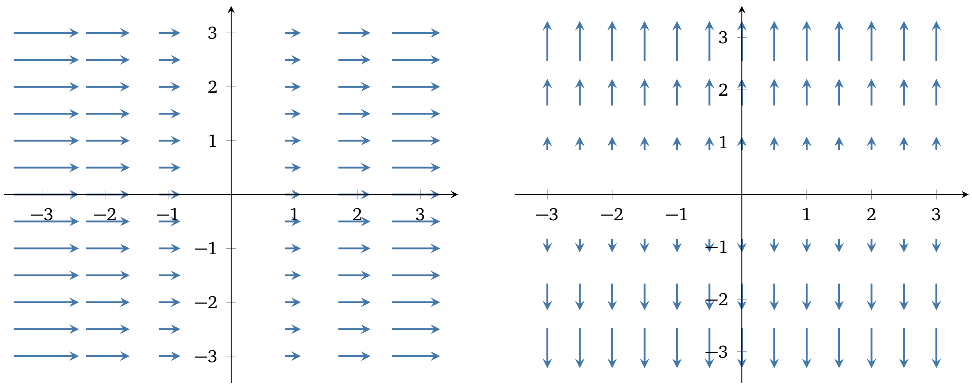
(c) If F is divergence-free, then F is curl-free.

(d) If F is curl-free, then F is divergence-free.

(e) If the flux of F along every closed piecewise curve is zero, then F is conservative.

(f) If F is conservative, then the flux of F across every closed piecewise curve is zero.

12.3.2 How do you interpret divergence? Think of a hose or a drain in a pool. Does water flow in or flow out? That's the sign of divergence. How fast does it flow? That is the size of divergence. A **source** is a point with positive divergence. A **sink** is a point with negative divergence. Consider both vector fields $F : \mathbb{R}^2 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$.



$F(x, y)$ $G(x, y)$

For each quantity below, determine whether it is positive, negative, or zero for each vector field.

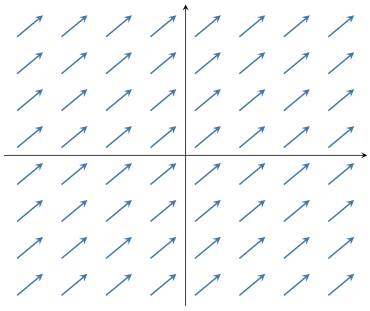
(a) The flux through $C = \partial[1, 3]^2$ traversed counterclockwise for each vector field.

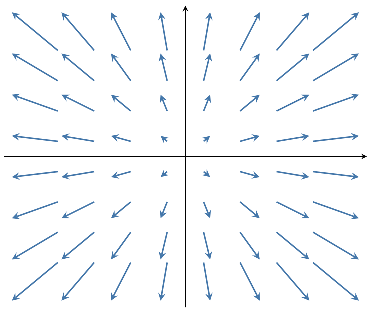
(b) The divergence at $(2, 2)$ for each vector field. Is it a source or a sink or neither?

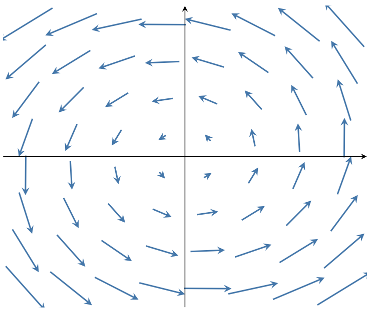
(c) The divergence at $(-2, -2)$ for each vector field. Is it a source or a sink or neither?

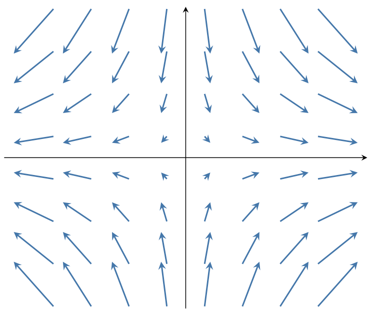
(d) The divergence at $(2, 0)$ for each vector field. Is it a source or a sink or neither?

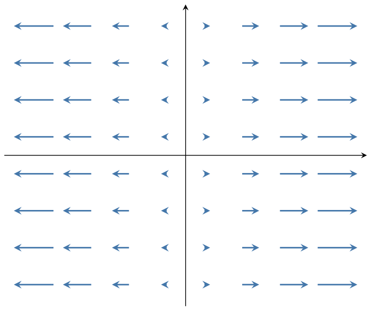
12.3.3 Identify which vector fields in $\mathbb{R}^2$ are sourceless. Do not guess an equation.

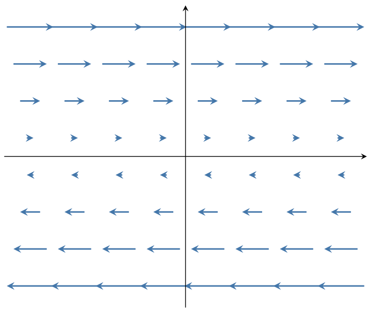

(a)

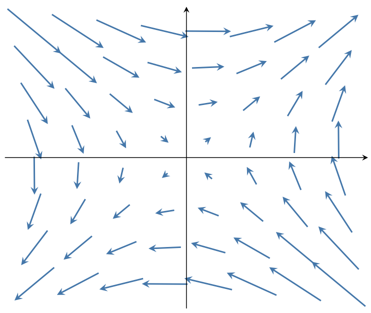

(b)

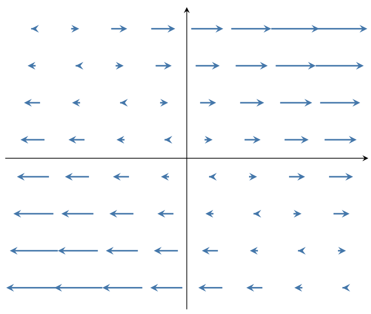

(c)


(d)

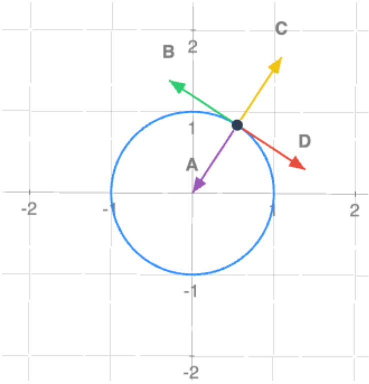

(e)


(f)


(g)


(h)

12.3.4 Calculating unit normals can be confusing with all the sign changes and swaps. To keep everything straight, the unit circle is your best memory tool.



Parametrize the unit circle S with

$$\gamma(t) = (x(t), y(t)), \qquad a \leq t \leq b,$$

for C^1 functions $x : [a, b] \to \mathbb{R}$ and $y : [a, b] \to \mathbb{R}$.

The actual choice of functions does not matter but you can keep in mind the usual options for a unit circle.

The vectors A, B, C, D illustrate **unit** vectors (either tangent or normal to the circle) at an arbitrary time t.

(a) If S is oriented counterclockwise, then which of A, B, C, D are its unit tangent T?

(b) If S is oriented counterclockwise, then which of A, B, C, D are its unit normal n?

(c) You keep forgetting the formula for the unit normal. Is it

$$n(t) = \frac{(-y'(t), x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \qquad \text{or} \qquad n(t) = \frac{(y'(t), -x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}}\ ?$$

Use the unit circle to help you decide.

Computations

12.3.5 Identify which of the following vector fields in $\mathbb{R}^2$ are sourceless.

(a) $F_1(x, y) = (x, 0)$

(b) $F_2(x, y) = (y, 0)$

(c) $F_3(x, y) = (x, -y)$

(d) $F_4(x, y) = (x, y)$

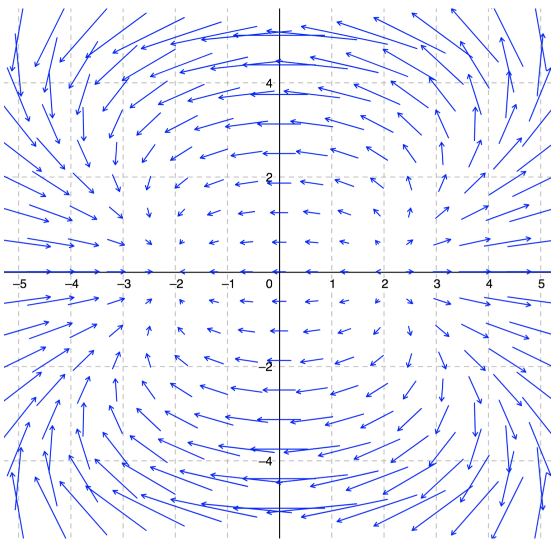
(e) $F_5(x, y) = (-y, x)$

(f) $F_6(x, y) = (y, x)$

(g) $F_7(x, y) = (1, 1)$

(h) $F_8(x, y) = (x + y, 0)$

12.3.6 Consider the $\mathbb{R}^2$ vector field $F(x, y) = (x^2 - y^2 - 4, xy)$



(a) Without calculating, what is the sign of the flux across each given closed curve?

i) C_1 is the counterclockwise oriented circle of radius 2 centered at $(2,0)$.
ii) C_2 is the counterclockwise oriented circle of radius 2 centered at $(-2,0)$.
iii) C_3 is the counterclockwise oriented square with vertices $(1,1),(1,3),(3,3),(3,1)$

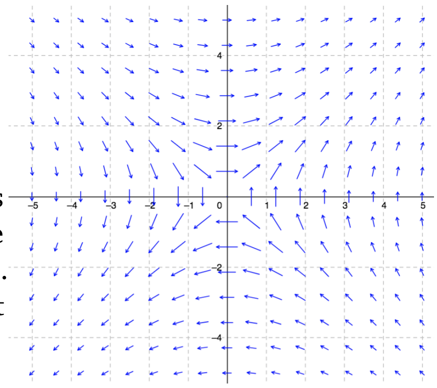
(b) Determine whether $(2,2)$ is a source, a sink, or neither. Explain with and without calculation.

(c) Let

$$R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, x \geq 0\}$$

be the solid right half semicircle of radius 2. Let $C = \partial R$ be the counterclockwise oriented boundary. Thus, C is a closed curve. Calculate the flux of F across C.

12.3.7 Consider a fluid flow in $\mathbb{R}^2$ defined by the vector field



$$F(x,y) = \left(\frac{y}{x^2+y^2+1}, \frac{x}{x^2+y^2+1}\right)$$

Let R be the solid rectangle with vertices $(0,0),(0,2),(3,0)$ and $(3,2)$. Let $C = \partial R$ be the counterclockwise oriented boundary of the rectangle. Use WolframAlpha to numerically evaluate any explicit integrals.

(a) Determine if the point $(2,2)$ is a source or a sink.

(b) Compute the net flux across C by direct calculation.

(c) Evaluate $\displaystyle\iint_R \operatorname{div}(F)\,dA$ by direct calculation. How is this related to the previous part?

Proofs

12.3.8 Let C be an oriented curve in $\mathbb{R}^2$ parametrized by $\gamma : [a,b] \to \mathbb{R}^2$. Write $\gamma(t) = (x(t), y(t))$.

(a) Show that the unit tangent T and unit normal n satisfy

$$T(t) = \frac{(x'(t), y'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}}, \qquad n(t) = \frac{(y'(t), -x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \qquad \text{for } a < t < b.$$

(b) Express the tangential line integral using an equivalent form of notation: if $F = (P,Q)$ is a vector field in $\mathbb{R}^2$ which is C^1 on an open set containing C then

$$\int_C F \cdot T\,ds = \int_C P\,dx + Q\,dy.$$

(c) Express the normal line integral using an equivalent form of notation: if $F = (P,Q)$ is a vector field in $\mathbb{R}^2$ which is C^1 on an open set containing C then

$$\int_C F \cdot n\,ds = \int_C P dy - Q dx.$$

12.3.9 Show that normal line integrals are invariant under reparametrization.

12.4. Green's theorem and divergence

Green's theorem in curl form establishes a beautiful relationship between circulation and curl. This constitutes a powerful tool for analyzing the *tangential* flow of a vector field, whether or not it is conservative. The arguments were elegant and flexible, so you can investigate the same questions about the *normal* flow of a vector field.

Can you relate the normal flow across two paths C_1 and C_2 of a vector field in $\mathbb{R}^2$?

Assuming C_1 and C_2 start and end at the same points, you can again reformulate this question in terms of a closed curve $C = C_1 - C_2$.

Is there another way to calculate flux of vector field across C in $\mathbb{R}^2$?

You will therefore want to establish an equivalent relationship between flux and divergence. Indeed, you will apply the same ideas for circulation and curl to develop this equivalent version of Green's theorem, and the similarities in arguments will be significant. Both of them are fundamental to vector calculus in $\mathbb{R}^2$ and the physical dynamics of fluid flows.

12.4.1 Regular regions and orienting the boundary

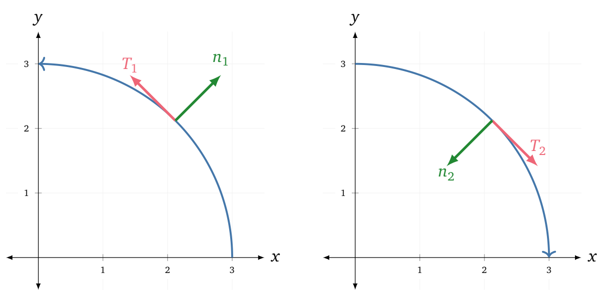
Now, you have already defined regular regions R and the orientation of their boundaries $C = \partial R$ in Section 12.2.1. These same definitions are still necessary for stating Green's theorem in divergence form, but your perspective on orientation changes. For circulation and curl, you view the boundary's orientation using the direction of *unit tangent T* of C, namely whether the region R stays on your left as you traverse its boundary C. For divergence and flux, you will view the boundary's orientation using the direction of the *unit normal n* of C.

As discussed in Section 12.3.1, if C is an oriented curve in $\mathbb{R}^2$, then its unit tangent T and unit normal n are automatically decided. The unit tangent T is defined by normalizing the velocity of a parametrization, and the unit normal n of C is chosen so that the orthogonal ordered basis $\{n, T\}$ is positively oriented at all points on the curve. That is, this ordered basis can be mapped to the ordered standard basis $\{e_1, e_2\}$ by a linear transformation with positive determinant. Informally, this description is a righthand rule.

If the unit tangent T points north (up) then the unit normal n points east (right). Equivalently, using your right hand, if T is your middle finger and n is your index finger then your thumb points out of the page.

The unit normal is therefore also defined for non-closed curves.

Example 12.4.1 Illustrated below are two oriented curves C_1 and C_2 in $\mathbb{R}^2$, where $C_2 = -C_1$, along with their unit tangents and unit normals.



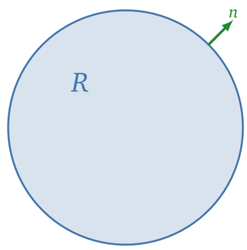
Notice the ordered bases $\{n_1, T_1\}$ and $\{n_2, T_2\}$ are both positively oriented. You can convince yourself using the righthand rule description.

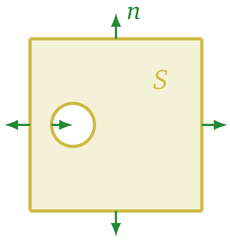
The orientation of a curve C in $\mathbb{R}^2$ can thus be described by either specifying the unit tangent or by specifying the unit normal. You only need to provide one of them. This gives another informal way of orienting the boundary of a regular region by Definition 12.2.3.

> *The boundary of a regular region is positively oriented if its unit normal always points outward as you traverse the boundary.*

This geometric description is straightforward to identify in explicit examples.

Example 12.4.2 Illustrated below are two regular regions R and S in $\mathbb{R}^2$. Each boundary is a finite disjoint union of piecewise curves which has been oriented according to the indicated unit normals. Both of them are positively oriented.





To specify orientation, remember you do not need to also include arrows along the curves themselves, because the unit normals have a unique corresponding unit tangent.

12.4.2 Statement and proof

Similar to curl and circulation, divergence is *infinitesimal* flux, as formulated in Lemma 12.3.16.

> *If F is a C^1 vector field in $\mathbb{R}^2$, then for any point $p \in \mathbb{R}^2$ in its domain*
>
> $$(\operatorname{div} F)(p) = \lim_{\varepsilon\to 0^+} \frac{1}{\operatorname{area}(B_\varepsilon(p))} \oint_{\partial B_\varepsilon(p)} F \cdot n \, ds,$$
>
> *where $\partial B_\varepsilon(p)$ is the boundary of the ε-disk centred at p oriented counterclockwise.*

With this gem, you can discover the divergence form of Green's theorem.

Theorem 12.4.3 (Green's theorem – divergence form) Let F be a vector field in $\mathbb{R}^2$ that is C^1 on a regular region $R \subseteq \mathbb{R}^2$. If the boundary ∂R is a finite disjoint union of positively oriented piecewise curves, then

$$\oint_{\partial R} (F \cdot n)\, ds = \iint_R \operatorname{div}(F)\, dA.$$

Remark 12.4.4 Since ∂R is assumed to be positively oriented, the unit normal n necessarily points outward from R. If you negatively orient ∂R, then this will introduce a negative sign in the above identity.

The divergence form of Green's theorem can again be heuristically summarized.

> *The total infinitesimal flux over R is the flux across its boundary ∂R.*

This statement can be translated more informally.

> *The total flow inside is the net flow across the edge.*

Watch this 5 minute video on deriving the divergence form of Green's theorem. You can follow the same proof strategy as the curl version of Green's theorem. The details of this argument are entirely analogous, but the calculations are messier. Instead, you will give an alternate proof of Green's theorem in divergence form by applying the curl form!

Proof. If $F = (P,Q)$, then define the corresponding vector field $G = (-Q, P)$ so G is a vector field in $\mathbb{R}^2$ that is C^1 on the regular region R and

$$\operatorname{curl} G = \partial_1 P - \partial_2(-Q) = \partial_1 P + \partial_2 Q = \operatorname{div} F.$$

Moreover, if $C = \partial R$ is parametrized by $\gamma(t) = (x(t), y(t))$ for $a \le t \le b$, then its unit tangent and unit normal are given by

$$T(t) = \frac{(x'(t), y'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}} \qquad n(t) = \frac{(y'(t), -x'(t))}{\sqrt{[x'(t)]^2 + [y'(t)]^2}}$$

since C is positively oriented. Since $ds = \sqrt{[x'(t)]^2 + [y'(t)]^2}dt$, this implies that

$$\begin{aligned}\oint_C F \cdot n\, ds &= \int_a^b F(x(t), y(t)) \cdot (y'(t), -x'(t))dt \\ &= \int_a^b P(x(t), y(t))y'(t) + Q(x(t), y(t))(-x'(t))dt \\ &= \int_a^b -Q(x(t), y(t))x'(t) + P(x(t), y(t))y'(t)dt \\ &= \int_a^b G(x(t), y(t)) \cdot (x'(t), y'(t))dt \\ &= \oint_C G \cdot T\, ds.\end{aligned}$$

Since the boundary ∂R is a finite disjoint union of positively oriented piecewise curves, it follows by Green's theorem in curl form (Theorem 12.2.6) that

$$\oint_C F \cdot n\, ds = \oint_C G \cdot T\, ds = \iint_R \operatorname{curl}(G)\, dA = \iint_R \operatorname{div}(F) dA$$

as required. ■

This slick proof will unfortunately not generalize to the three dimensional version of flux and circulation. It is a miraculous coincidence of $\mathbb{R}^2$. Nonetheless, you can at least appreciate that flux and divergence enjoy the same relationship as circulation and curl.

12.4.3 Examples with Green's theorem in divergence form

You can study some routine calculations with Green's theorem in divergence form.

Example 12.4.5 Let C be the positively oriented boundary of the upper half disk

$$R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 9, y \ge 0\},$$

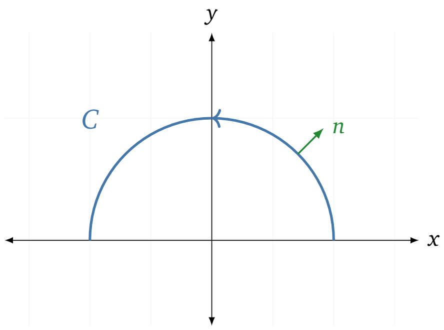
so $C = \partial R$. Recall in Example 12.3.10 that you calculated the flux of $F(x, y) = (e^y, 2y - x^2)$ across C by the definition of the line integral. It was a rather involved calculation. You can instead apply Green's theorem. Since F is a C^1 vector field on the regular region R and its boundary $C = \partial R$ is a positively oriented piecewise curve enclosing the regular region R, it follows by Green's theorem in divergence form that

$$\oint_C F \cdot n ds = \iint_R \operatorname{div}(F) dA$$
$$= \iint_R \partial_1(e^y) + \partial_2(2y - x^2) dA$$
$$= \iint_R 2dA = 2\operatorname{area}(R) = 9\pi.$$

This matches the answer in Example 12.3.10 but at a fraction of the effort.

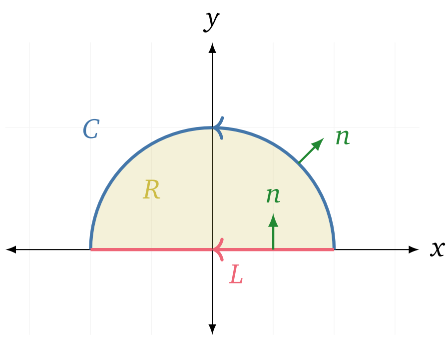
Green's theorem again allows you to relate the flux across two curves.

Example 12.4.6 (Moving the curve) Suppose you want to compute the *upward* flux of the vector field $F(x,y) = (\sin(y^2) - 2, -x + 5y)$ across the semicircle $C = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1, y \geq 0\}$. To orient C with an upward unit normal (i.e. a unit normal with non-negative y-component), you must traverse C from right to left as illustrated below.



With this orientation of C, you are seeking to evaluate the line integral $\int_C F \cdot n ds$. What methods can you try?

You can try computing this line integral directly, but it will create a messy integral. Moreover, C is not closed so you cannot directly use Green's theorem. There is again one more option: *move the curve*. Orient the straight line segment $L \subseteq \mathbb{R}^2$ from $(1,0)$ to $(-1,0)$ with an upward normal. The region R between the curves has positively oriented boundary $\partial R = C - L$.



Thus, the piecewise curve $C - L = \partial R$ is a positively oriented boundary enclosing the regular region $R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, y \geq 0\}$. Thus, by Green's theorem,

$$\int_C F \cdot n\, ds - \int_L F \cdot n\, ds = \oint_{C-L} F \cdot n\, ds = \iint_R (\operatorname{div} F) dA.$$

This relates the upward flux across C and the upward flux across L! Parametrizing L by $\gamma(t) = (-t, 0)$ for $-1 \leq t \leq 1$, you can verify that $n\, ds = n(t)||\gamma'(t)||dt = (0,1)dt$ so

$$\begin{aligned}
\int_C F \cdot n\, ds &= \int_L F \cdot n\, ds + \iint_R (\operatorname{div} F)\, dA \\
&= \int_{-1}^{1} F(-t,0) \cdot (0,1) dt + \iint_R 5\, dA \\
&= \int_{-1}^{1} (-2,t) \cdot (0,1) dt + 5\, \operatorname{area}(R) \\
&= \int_{-1}^{1} t dt + \frac{5\pi}{2} = \frac{5\pi}{2}.
\end{aligned}$$

Notice how these arguments almost exactly parallel Example 12.2.10.

This concludes a remarkable chapter on vector calculus in $\mathbb{R}^2$. The physics of force fields were central to your investigations and this focus has revealed two marvelous generalizations of the fundamental theorem of calculus. First, you have the fundamental theorem of line integrals in $\mathbb{R}^2$, which you have previously discovered. This provides a deep statement about conservative vector fields and path independence. Second, you have Green's theorem in $\mathbb{R}^2$, which comes in two equivalent flavours. This has generated fantastic insights on work done by non-conservative vector fields and the relationships between curl and circulation as well as divergence and flux.

Vector calculus in $\mathbb{R}^2$ has proved to be unreasonably useful to analyze the dynamics of fluids and flows along curves. You might already be wondering how to generalize these ideas.

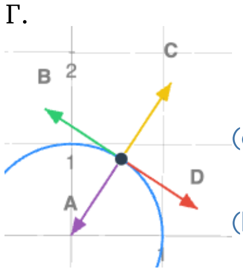
How can you describe the physics of fluids and flows in $\mathbb{R}^3$?

Vector calculus in $\mathbb{R}^3$ will require a thorough study of two-dimensional surfaces and their dynamics with three dimensional vector fields. This subject will be the focus of the next chapter in preparation for the fundamental theorems in $\mathbb{R}^3$.

Exercises for Section 12.4

Concepts and definitions

12.4.1 The vectors A, B, C, D below are unit vectors which are normal or tangent to the oriented curve Γ.



(a) If Γ is oriented so that B is its unit tangent vector then which vector is its unit normal?

(b) If Γ is oriented so that D is its unit tangent vector then which vector is its unit normal?

12.4.2 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 vector field and let $R \subseteq \mathbb{R}^2$ be a regular region whose boundary $C = \partial R$ is a finite disjoint union of positively oriented piecewise curves. Each expression either equals the **outward** flux of F through C, or the **inward** flux of F through C, or there is not enough information to decide. Identify each of them.

(a) $\displaystyle\oint_C (F \cdot n)ds$

(b) $\displaystyle-\oint_C (F \cdot n)ds$

(c) $\displaystyle\oint_{-C} (F \cdot n)ds$

(d) $\displaystyle\iint_R \operatorname{div}(F)dA$

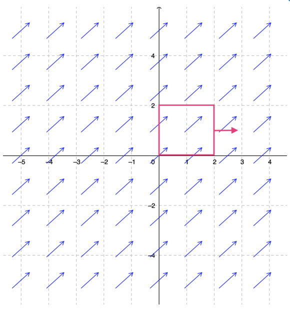
(e) $\displaystyle-\iint_R \operatorname{div}(F)dA$

(f) $\displaystyle-\oint_{-C} (F \cdot n)ds$

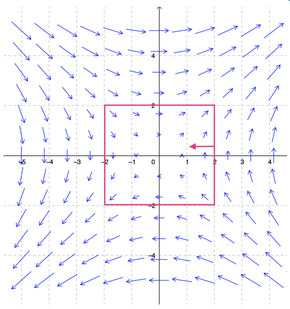
12.4.3 Each closed curve C is oriented by specifying a unit normal (and hence a unit tangent). Each curve encloses a regular region R. For each vector field F and curve C below, determine the sign of

$$\text{i) } \oint_C (F \cdot n)ds \qquad \text{and} \qquad \text{ii) } \iint_R \operatorname{div}(F)dA.$$

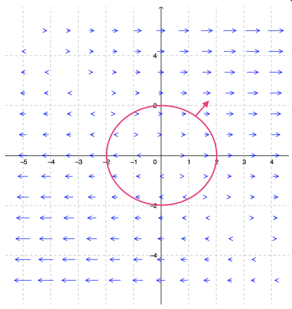
(a)



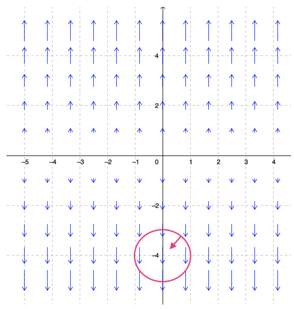
(b)



(c)



(d)



Computations

12.4.4 Expressing the unit normal in terms of a parametrization requires some care with signs. Let $C \subseteq \mathbb{R}^2$ be an oriented curve parametrized by $\gamma(t) = (x(t), y(t))$ for $a \le t \le b$.

(a) Express the unit normal n of C in terms of x and y. *Hint:* Use a unit circle to check signs.

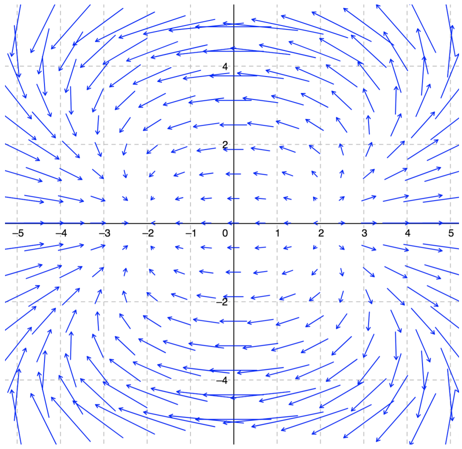
(b) Parametrize the negatively oriented curve $-C$ using x and y.

(c) Express the unit normal n of $-C$ in terms of x and y. *Hint:* Use a unit circle to check signs.

12.4.5 Let $F(x,y) = (x^2 - y^2 - 4, xy)$ be a vector field in $\mathbb{R}^2$. Let

$$R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 4, x \ge 0\}$$

be the solid right semicircle of radius 2. Let $C = \partial R$ be the positively oriented boundary of R. Remember you directly calculated the outward flux of F through C in Exercise 12.3.6.

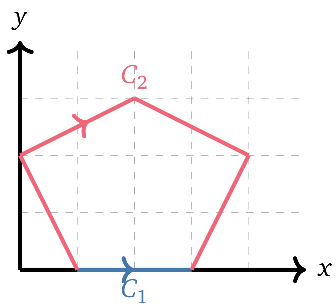


(a) Sketch the curve C and region R on the vector field. Make the orientation of C clear.

(b) Calculate the outward flux of F across C using Green's theorem. *Hint:* This should be short!

12.4.6 Let R be the pentagon with vertices $(1,0), (3,0), (4,2), (2,3)$ and $(0,2)$. Let C_1 and C_2 be two oriented curves starting at $(1,0)$ and ending at $(3,0)$ as shown below.

(a) Sketch the unit normal on C_2 and the unit normal on C_1. Let $F(x,y) = (-y,x)$ be a vector field in $\mathbb{R}^2$. Then express the outward flux of F through ∂R as a sum of line integrals over C_1 and C_2.



(b) Compute

$$I = \int_{C_2} (F \cdot n) ds$$

by moving the line integral to C_1. Do not attempt a direct parameterization; it is a long mess.

12.4.7 Let R be the rhombus with vertices $(-1,0), (0,-1), (1,0)$ and $(0,1)$ so R is a regular region in $\mathbb{R}^2$. Compute the outward flux of $F(x,y) = (x^2 + y^2 - 5, -2xy)$ through ∂R in two ways:

(a) Directly as a line integral. This will not be fun.

(b) Using Green's theorem. This will be fun.

12.4.8 Let $F(x,y) = (2x + \sin(y)\cos(y), -y + \cos(x)e^x)$ and C be the boundary of a triangle with vertices $(0,0), (1,0)$ and $(0,1)$. Compute the outward flux of F through C. *Hint*: This is fast.

12.4.9 Let $R \subseteq \mathbb{R}^2$ be a regular region whose boundary ∂R is a finite disjoint union of positively oriented piecewise curves. Express the area of R as the flux

$$\oint_{\partial R} (F \cdot n)\, ds$$

for at least 3 different choices of vector fields F.

Proofs

12.4.10 Let F be a vector field in $\mathbb{R}^2$ that is C^1 on an open set $U \subseteq \mathbb{R}^2$. Let R be a regular region inside U whose boundary ∂R is a finite disjoint union of positively oriented piecewise curves. Prove that if F is sourceless on U, then F has zero outward flux through ∂R.

12.4.11 Let $C \subseteq \mathbb{R}^2$ be an oriented curve and let F be a vector field in $\mathbb{R}^2$ that is continuous on C. Use Exercise 12.4.4 to prove that

$$\int_{-C} (F \cdot n)\, ds = -\int_{C} (F \cdot n)\, ds.$$

12.4.12 Prove that Green's theorem in divergence form implies the curl form.

12.4.13 By following the proof of Green's theorem in curl form, establish Green's theorem in divergence form for regular regions that are both x-simple and y-simple.

13. Integration on surfaces

You have developed a sophisticated theory for calculus with curves, which serves as an excellent model for the physics of a particle moving through a vector field. Physics, however, demands a much deeper understanding of vector fields in $\mathbb{R}^3$ beyond a single particle. For instance, you may want to measure the behaviour of a magnetic field with a coil, or study water flowing through a steady river through a net. This creates some serious physical questions.

> *Does the net form a smooth shape? What is the area of the net? How much does the water swirl along the edge of the net? How much water flows across the net?*

The physics of fluids and force fields in $\mathbb{R}^3$ need (two-dimensional) *surfaces* in addition to curves. You will therefore need to develop a theory for calculus with surfaces.

Differentiation with surfaces in $\mathbb{R}^3$ has already been carefully constructed with tangent spaces (Section 4.5) and the definition of smooth manifolds (Section 4.6). This theory is quite developed for implicit manifolds (Section 5.5) but, as with curves, you will need a few more tools for differentiation on surfaces in parametric form (Section 1.5.1). That said, you have a good foundation for *differential* calculus with surfaces.

Integration with surfaces in $\mathbb{R}^3$ has little visible progress. Intuitively, every two-dimensional smooth manifold (and hence smooth surface) in $\mathbb{R}^3$ should have zero volume. That is consistent with your definition of three-dimensional volume, but it seems a surface should have two-dimensional area.

> *What is the area of a surface? How can you integrate along a surface? How can you define flux across a surface?*

These motivations will form the basis of *integral* calculus with surfaces.

13.1. Surfaces

Surfaces in $\mathbb{R}^3$ act as a two-dimensional generalization of curves in $\mathbb{R}^3$. To develop integration on surfaces, it is natural to closely parallel the development of integration on curves. You will want to generalize the ideas for curves and thus consider *parametrically* described sets with *two* variables, so unfortunately the deep tools in Chapter 5 for 2-dimensional manifolds written in implicit form will not be sufficient. Nonetheless, you can ask many of the same questions.

> *What is a surface? What is a parametrized surface? Will "equivalent" parametrizations of the same surfaces yield different answers? How are two parametrizations of the same surfaces considered equivalent?*

You will resolve these questions by relying on your experience with curves. By the end of this section, you will have built the foundation for calculus with surfaces.

13.1.1 Simple smooth regular parametrizations

As a first step towards defining surfaces, you may be inspired by the informal study of Section 1.5 and Definition 11.1.1 to conjecture the following.

Definition 13.1.1 Let $S \subseteq \mathbb{R}^3$ be a set. Let $U \subseteq \mathbb{R}^2$ be a regular region with path-connected interior. A map $G : U \to \mathbb{R}^3$ is a **(2-variable) parametrization of** S if $\operatorname{im}(G) = S$, and G is continuous.

Remark 13.1.2 The requirement for the domain to be a regular region with path-connected interior is probably somewhat mysterious. It has been chosen to exclude some annoying situations, which you do not need to explore at such an early stage. The theory of surfaces is really quite delicate; the definitions in this book are *not* the modern standard[1] but they will be a good starting point for learning vector calculus in $\mathbb{R}^3$.

Example 13.1.3 Any rectangle in $\mathbb{R}^2$ and any closed disk in $\mathbb{R}^2$ is a regular region with path-connected interior. This fact will be repeatedly used without mention.

Example 13.1.4 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \le z \le 3\}$ be a cylinder of radius 2 and height 3. You can parametrize S using restrictions of the cylindrical coordinates map. For instance, the map $G : [0, 2\pi] \times [0, 3] \to \mathbb{R}^3$ given by

$$G(s, t) = (2\cos s, 2\sin s, t)$$

is continuous. You can verify that $[0, 2\pi] \times [0, 3]$ is a regular region with path-connected interior and $\operatorname{im}(G) = S$, so G is a parametrization of S.

However, this definition immediately has a serious problem. It also permits curves!

Example 13.1.5 Define the set $C = \{(x, x^2, x^3) : 0 \le x \le 1\}$. The set C is a curve, but you can parametrize C using maps with *two-dimensional* domains. For instance, $G : [0, 1] \times [\pi, 237] \to \mathbb{R}^3$ given by $G(s, t) = (s, s^2, s^3)$ parametrizes C. This choice is rather silly since the second variable has no impact on the image. How can you detect this lack of dependence on the second variable? Derivatives detect change with respect to variables, so you can investigate

[1] For a complete introduction, see a course in differential geometry such as one dedicated to curves and surfaces.

them. Notice that

$$\partial_1 G(s,t) = (1, 2s, 3s^2) \qquad \partial_2 G(s,t) = (0,0,0).$$

The vanishing second partial indicates that G does not depend on the second variable. You do not want 1-dimensional curves to be considered 2-dimensional surfaces.

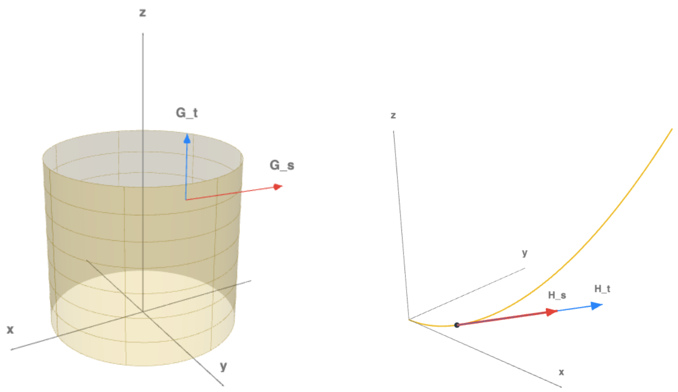
What if a parametrization with domain in $\mathbb{R}^2$ has non-vanishing partials?

This request is still not enough.

Example 13.1.6 The curve $C = \{(x, x^2, x^3) : 0 \leq x \leq 1\}$ can be parametrized by the map $H : [0,1] \times [0,1] \to \mathbb{R}^3$ given by $H(s,t) = (st, (st)^2, (st)^3)$. This situation is rather annoying. How can you disallow curves? The insight will arise from comparing the parametrization G of the 2-dimensional cylinder S from Example 13.1.4 with this parametrization H of the 1-dimensional curve C. You can calculate their partials.

$$\partial_1 G = \begin{bmatrix} -2\sin s \\ 2\cos s \\ 0 \end{bmatrix} \quad \partial_2 G = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \partial_1 H = \begin{bmatrix} t \\ 2t(st) \\ 3t(st)^2 \end{bmatrix} \quad \partial_2 H = \begin{bmatrix} s \\ 2s(st) \\ 3s(st)^2 \end{bmatrix}$$

Both of these never vanish on the interior of their domains. On the other hand, notice that $\{\partial_1 G, \partial_2 G\}$ are always linearly *independent* whereas $\{\partial_1 H, \partial_2 H\}$ are always linearly *dependent*. Informally speaking, you can always move in *two* directions on S via G, whereas you can only move in *one* direction on C via H. This matches your intuitive distinction between a surface and a curve! See the illustration below for a visual.



Play with this Math3D cylinder demo and this Math3D curve demo for better visuals.

This additional obstacle gives a new suggestion.

*What if a parametrization with domain in $\mathbb{R}^2$ has **linearly independent** partials?*

This critical observation suggests two key definitions.

Definition 13.1.7 Let $U \subseteq \mathbb{R}^2$. A map $G : U \to \mathbb{R}^3$ is **smooth**[2] if G is C^1 on the interior of U, and its partials $\partial_1 G$ and $\partial_2 G$ are bounded on the interior of U.

[2]As with Definition 11.1.5 for smooth 1-variable maps, the notion of "smooth" depends on context so no standard definition exists. Our choice is good enough for most examples. It is designed to avoid improper integrals for simplicity's sake and to be compatible with our version of change of variables (Theorem 9.8.2).

Definition 13.1.8 Let $U \subseteq \mathbb{R}^2$. A map $G : U \to \mathbb{R}^3$ is **regular** if G is differentiable on the interior of U and, for every interior point p of U, the set $\{\partial_1 G(p), \partial_2 G(p)\}$ is linearly independent.

Example 13.1.9 The map $G : [0, 2\pi] \times [0, 3] \to \mathbb{R}^4$ given by $G(s,t) = (2\cos s, 2\sin s, t)$ in Example 13.1.4 is smooth and regular. These are left as straightforward exercise to verify.

Example 13.1.10 The 2-variable parametrizations of the curve C in Examples 13.1.5 and 13.1.6 are smooth but not regular. This is left as a straightforward exercise to verify.

While smooth regular parametrizations are better, they are not perfect.

Example 13.1.11 The map $G : [0, 2\pi] \times [0, 3] \to \mathbb{R}^3$ given by

$$G(s,t) = (2\sin s, 2\sin s \cos s, t)$$

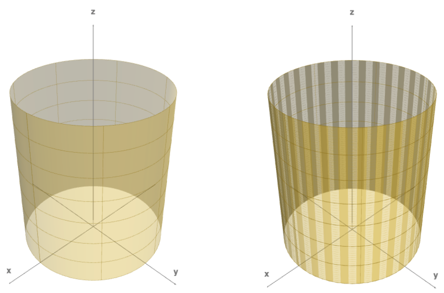
parametrizes the figure eight cylinder $S = \{(x, y, z) \in \mathbb{R}^3 : x^4 = x^2 - y^2, 0 \le z \le 3\}$. You can verify that the map G is smooth and regular, but this set S will not be a 2-dimensional smooth manifold (Definition 4.6.4) along the z-axis! View this Math3D demo for an illustration. Notice this issue occurs precisely when $s \in \{0, \pi, 2\pi\}$ and $t \in [0, 3]$ because G fails to be injective on the interior of its domain.

The above example demonstrates that injectivity will be desirable. As with curves, you want to require injectivity on the *interior* of the parametrization's domain.

Example 13.1.12 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \le z \le 3\}$ be a cylinder of radius 2 and height 3. A cylinder is certainly a 2-dimensional smooth manifold. You can parametrize S using restrictions of the cylindrical coordinates map. For instance, the maps $G : [0, 2\pi] \times [0, 3] \to \mathbb{R}^3$ and $H : [0, 4\pi] \times [0, 3] \to \mathbb{R}^3$ given by

$$G(s,t) = (2\cos s, 2\sin s, t) \qquad H(s,t) = (2\cos s, 2\sin s, t)$$

are continuous and you can verify that $\mathrm{im}(G) = S = \mathrm{im}(H)$, so both maps G and H parametrize S. You can also verify that both maps are smooth and regular, but neither G nor H is injective. This Math3D demo illustrates this concern.



This may seem alarming but there is a key distinction. The map H wraps around the cylinder *twice*, so H fails to be injective on its entire domain, a set of positive Jordan measure. On the other hand, the map G fails to be injective *on the boundary of its domain* in $\mathbb{R}^2$, namely $G(s,t) = G(s',t')$ when $t = t'$ and either $s = s'$ or $\{s, s'\} = \{0, 2\pi\}$. The image of this subset under G corresponds to a single edge along the cylinder in $\mathbb{R}^3$. Informally speaking, this edge is created by "gluing together" the boundary of the rectangle to make a cylinder.

These examples suggest an additional property for parametrizations.

Definition 13.1.13 Let $U \subseteq \mathbb{R}^2$ be a set. A map $G : U \to \mathbb{R}^3$ is **simple** if G is injective on U except possibly along the boundary. That is,

$$\forall x, y \in U, G(x) = G(y) \implies x = y \quad \text{or} \quad x, y \in \partial U.$$

This definition eliminates the remaining bad examples and keeps the good ones.

Example 13.1.14 For the figure eight cylinder in Example 13.1.11, the map G is not simple.

Example 13.1.15 For the cylinder in Example 13.1.12, the map G is simple whereas H is not.

Your efforts are confirmed with a major theorem on 2-variable parametrizations in $\mathbb{R}^3$.

Theorem 13.1.16 If a map $G : U \to \mathbb{R}^3$ is a simple smooth regular 2-variable parametrization of a set $S \subseteq \mathbb{R}^3$, then the set S is a 2-dimensional smooth manifold at $G(c)$ for every $c \in U^o$.

Remark 13.1.17 This theorem constitutes the 2-dimensional analogue of curves (Theorem 11.1.11), and the parametric analogue of the relationship between the kernel and implicit manifolds (Theorem 5.5.7).

Proof. Omitted. This follows from the implicit function theorem. See a key step, for example, in Theorem 2 of [13, Section 3.2] or Theorem 3.13 of [11]. ■

Although there are many details to check, you can apply this theorem with relative ease.

Example 13.1.18 The cylinder $S = \{(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 = 4, 0 \leq z \leq 3\}$ can be parametrized in many ways. For instance, $G(s,t) = (2\cos s, 2\sin s, t)$ for $0 \leq s \leq 2\pi, 0 \leq t \leq 3$ is a simple smooth regular parametrization so, by Theorem 13.1.16, S is a 2-dimensional smooth manifold at $G(s,t)$ for $0 < s < 2\pi, 0 < t < 3$ and hence at every point of S except possibly the the top and bottom circles as well as the edge $(2, 0, z)$ for $0 \leq z \leq 3$. By choosing another simple smooth regular parametrization, you can prove the set S is a 2-dimensional smooth manifold at $(2, 0, z)$ for $0 \leq z \leq 3$ using Theorem 13.1.16. This is left as an exercise.

This wraps up your first objective of determining valid parametrizations.

13.1.2 Surfaces and piecewise surfaces

Theorem 11.1.11 confirms a good definition for parametrized surfaces in $\mathbb{R}^3$.

Definition 13.1.19 A set $S \subseteq \mathbb{R}^3$ is a **surface** in $\mathbb{R}^3$ if there exists a simple smooth regular two-variable parametrization of S.

Remark 13.1.20 The word "surface" is used in many contexts, so beware when you venture into other sources.[3]

Example 13.1.21 The cylinder $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \leq z \leq 3\}$ in Example 13.1.4 is a parametrized simple smooth manifold, because you can verify that the map $G(s,t) =$

[3] Definition 13.1.19 will suffice for your purposes but it is not the modern standard. Namely, there are entire classes of sets that are "surfaces" based on a more universal definition commonly used today, but that do not meet the more restrictive definition given above. A thorough exploration of surfaces belongs to the study of differential geometry, such as a dedicated course on curves and surfaces. This first attempt can be viewed as progress.

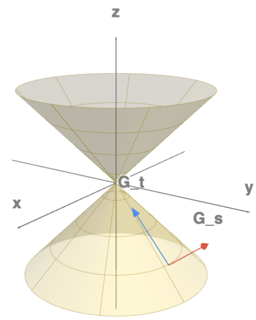
$(2\cos s, 2\sin s, t)$ for $0 \le s \le 2\pi, 0 \le t \le 3$ is a simple smooth regular parametrization of S.

Example 13.1.22 Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be C^1. You can verify that the graph of f is a surface in $\mathbb{R}^3$. This is left as a routine exercise.

Sometimes you may want to consider sets which can be broken up into pieces.

Example 13.1.23 The two-sided cone $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = z^2, -2 \le z \le 2\}$ is parametrized by the cylindrical coordinates map $G : [0,2\pi] \times [-2,2] \to \mathbb{R}^3$ given by

$$G(s,t) = (t\cos s, t\sin s, t).$$



Notice that

$$\partial_1 G(s,t) = (-t\sin s, t\cos s, 0),$$
$$\partial_2 G(s,t) = (\cos s, \sin s, 1).$$

You can verify that the map G is smooth, and $\{\partial_1 G(s,t), \partial_2 G(s,t)\}$ is linearly dependent if and only if $s \in (0,2\pi)$ and $t = 0$. Play with this Math3D demo to view this phenomenon.

Thus, G is not regular so you cannot conclude whether S is a surface. In fact, S is not a surface but it is not easy to formally prove as you need to exclude all possible parametrizations. You can, however, write S as a union of two surfaces.

- Define the upper cone
$$S_1 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = z^2, 0 \le z \le 2\}.$$
The map $G_1 : U_1 \to \mathbb{R}^3$ given by $U_1 = [0,2\pi] \times [0,2]$ and $G_1(s,t) = (t\cos s, t\sin s, t)$ is a simple smooth regular parametrization of S_1 and hence S_1 is a surface.
- Define the lower cone
$$S_2 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = z^2, -2 \le z \le 0\}.$$
The map $G_2 : U_2 \to \mathbb{R}^3$ given by $U_2 = [0,2\pi] \times [-2,0]$ and $G_2(s,t) = (t\cos s, t\sin s, t)$ is a simple smooth regular parametrization of S_2 and hence S_2 is a surface.

Thus, $S = S_1 \cup S_2$ and, more importantly, the intersection $S_1 \cap S_2$ is equal to *the image of* the domain boundaries of G_1 and G_2, namely

$$S_1 \cap S_2 = G_1(\partial U_1) \cap G_2(\partial U_2).$$

The surfaces S_1 and S_2 are "glued along their parametrization's domain boundaries".

You will want to introduce language that captures this kind of example.

Definition 13.1.24 A set $S \subseteq \mathbb{R}^3$ is a **piecewise surface** if S can be constructed by gluing together finitely many surfaces along their parametrization's domain boundaries.

Remark 13.1.25 This definition is not at all rigorous. It is a complicated process to describe formally. Instead, you will only need to intuitively identify "gluing" in explicit examples.

This expands the class of surfaces under consideration to many natural possibilities.

Example 13.1.26 Examples of piecewise surfaces include the two-sided cone in Example 13.1.23, the figure eight cylinder in Example 13.1.11, and the unit cube $\partial[0,1]^3$.

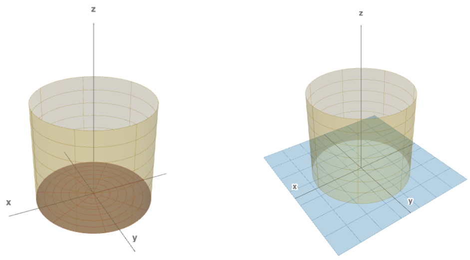
Example 13.1.27 The cylinder $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \le z \le 3\}$ is a surface. You can also verify that the disk

$$D = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \le 4, z = 0\}$$

and the bounded plane

$$P = \{(x,y,z) \in \mathbb{R}^3 : -2 \le x \le 2, -2 \le y \le 2, z = 0\}$$

are both surfaces. The set $S \cup D$ is a piecewise surface but the set $S \cup P$ is not a piecewise surface. This intuition can be viewed in this Math3D demo.



You can parametrize S and D such that the boundaries are "glued" together, so the union $S \cup D$ is a piecewise surface. However, you cannot parametrize S and P such that the boundaries are "glued" together, so the union $S \cup P$ is not a piecewise surface.

These definitions are an excellent starting point for performing calculus with surfaces.

13.1.3 Reparametrizations and orientation

As with curves, you will not want different choices of parametrizations to lead to different conclusions. This leaves one last question.

How are two parametrizations considered equivalent?

Motivated by Definition 11.1.23, you can construct a sensible definition.

Definition 13.1.28 Let $G : U \to \mathbb{R}^3$ and $H : V \to \mathbb{R}^3$ be simple smooth regular 2-variable parametrizations of a set $S \subseteq \mathbb{R}^3$. Define G to be a **reparametrization** of H if there exists a continuous invertible $\varphi : U \to V$ such that φ is C^1 on the interior of U, its Jacobian determinant $\det D\varphi$ is bounded and never zero on the interior of U, and $G = H \circ \varphi$.

- If $\det D\varphi > 0$ on the interior of U, then G has the **same orientation** as H.
- If $\det D\varphi < 0$ on the interior of U, then G has the **opposite orientation** as H.

Remark 13.1.29 Lemma 9.8.8 implies that the restriction $\varphi|_{U^\circ} : U^\circ \to V^\circ$ is a diffeomorphism.

Example 13.1.30 The upper hemisphere S of radius 3 can be parametrized by the map $G : [0, 2\pi] \times [0, \pi/2] \to \mathbb{R}^3$ defined as

$$G(u,v) = (3\cos u \cos v, 3\sin u \cos v, 3\sin v)$$

It is left as an exercise to show that G is a simple smooth regular parametrization of S. You can also verify these other possibilities.

- $H(s,t) = (3\cos t\cos s, 3\sin t\cos s, 3\sin s)$ for $0 \leq s \leq \pi/2$ and $0 \leq t \leq 2\pi$ is a reparametrization of G with the opposite orientation.
- $J(s,t) = (\sqrt{9-t^2}\cos s, \sqrt{9-t^2}\sin s, t)$ for $0 \leq s \leq 2\pi$ and $0 \leq t \leq 3$ is a reparametrization of G with the same orientation.
- $K(s,t) = (\sqrt{6t-t^2}\cos s, \sqrt{6t-t^2}\sin s, 3-t)$ for $0 \leq s \leq 2\pi$ and $0 \leq t \leq 3$ is a reparametrization of G with the opposite orientation.

The key challenge is to find the reparametrizing map φ in Definition 13.1.28; once you do, there are quite a few details to check and they are all rather tedious.

The property of reparametrization is symmetric, transitive, and reflexive.

Lemma 13.1.31 Let $G_1 : U_1 \to \mathbb{R}^3, G_2 : U_2 \to \mathbb{R}^3, G_3 : U_3 \to \mathbb{R}^3$ be simple smooth regular 2-variable parametrizations of a set $S \subseteq \mathbb{R}^3$. All of the following hold:

(a) (**Reflexive**) G_1 is a reparametrization of itself.

(b) (**Symmetry**) If G_1 is a reparametrization of G_2, then G_2 is a reparametrization of G_1.

(c) (**Transitive**) If G_1 is a reparametrization of G_2 and G_2 is a reparametrization of G_3, then G_1 is a reparametrization of G_3.

Proof. These are left as exercises. Use properties of inverse functions and the chain rule. ■

These three properties allow you to think of a surface using any "equivalent" simple smooth regular parametrization.[4] Orientation will be studied more closely in a later section.

By exploring many examples and relying on your experience with curves, you have created simple smooth regular 2-variable parametrizations and thus defined surfaces. You will omit the adjectives "simple smooth regular" because you will exclusively study these parametrizations for all of vector calculus.

*A **parametrization** G will henceforth be a simple smooth regular parametrization.*

These maps act as the perfect descriptions for surfaces in $\mathbb{R}^3$, because Theorem 13.1.16 guarantees that the tangent space of these surfaces are 2-dimensional subspaces of $\mathbb{R}^3$. This creates a gateway for approximating *surfaces* by *parallelograms*, and that is precisely how all of your existing calculus tools will once again thrive.

[4]The formal terminology is *equivalence class* of simple smooth regular parametrizations. For more on equivalence classes, see a course on mathematical proofs. These same ideas often appear in algebra or number theory.

Exercises for Section 13.1

Concepts and definitions

13.1.1 The unit sphere S in $\mathbb{R}^3$ is parametrized by map $G : [0,2\pi] \times [0,\pi] \to \mathbb{R}^3$, where

$$G(u,v) = \begin{bmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{bmatrix}, \qquad \partial_1 G(u,v) = \begin{bmatrix} -\sin u \sin v \\ \cos u \sin v \\ 0 \end{bmatrix} \qquad \partial_2 G(u,v) = \begin{bmatrix} \cos u \cos v \\ \sin u \cos v \\ -\sin v \end{bmatrix}.$$

(a) Is G smooth? Briefly explain but do not include all the details.

(b) Is G regular? Briefly explain but do not include all details.

(c) Is G simple? Briefly explain but do not include all details.

13.1.2 Define the following five maps.

- $G_A(s,t) = (s,t,st)$ for $-1 \leq s \leq 1, -1 \leq t \leq 1$.
- $G_B(s,t) = (s,t,st)$ for $s^2 + t^2 \leq 1$
- $G_C(s,t) = (st, st, (st)^2)$ for $s^2 + t^2 \leq 1$.
- $G_D(s,t) = (s^3, t^3, (st)^3)$ for $-1 \leq s \leq 1, -1 \leq t \leq 1$.
- $G_E(s,t) = (s\cos t, s\sin t, s^2 \cos t \sin t)$ for $0 \leq s \leq 1, 0 \leq t \leq 4\pi$.
- $G_F(s,t) = (s,t,(st)^{1/3})$ for $-1 \leq s \leq 0, 0 \leq t \leq 1$.

Define the set $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = xy\}$.

(a) Which of these maps are parametrizations of S?

(b) Which of these maps are smooth?

(c) Which of these maps are regular?

(d) Which of these maps are simple?

13.1.3 Define the following five maps.

- $G_A(s,t) = (t,s,st)$ for $0 \leq s \leq 4, -1 \leq t \leq 1$.
- $G_B(s,t) = (s^3, t, s^3 t)$ for $-1 \leq s \leq 1, 0 \leq t \leq 4$.
- $G_C(s,t) = (s, 4s^2, 4s^3)$ for $-1 \leq s \leq 1, 0 \leq t \leq 4$.
- $G_D(s,t) = (s, t^2, st^2)$ for $0 \leq s \leq 1, -2 \leq t \leq 2$.
- $G_E(s,t) = (\cos s, 4\sin t, 4\cos s \sin t)$ for $0 \leq s \leq \pi, 0 \leq t \leq \pi$.

Define $S = \{(x,y,z) \in \mathbb{R}^3 : -1 \leq x \leq 1, 0 \leq y \leq 4, z = xy\}$.

(a) Which of these maps are parametrizations of S?

(b) Which of these maps are smooth?

(c) Which of these maps are regular?

(d) Which of these maps are simple?

13.1.4 Let $S \subseteq \mathbb{R}^3$. Determine which of the following are true or false. If true, briefly justify. If false, state a counterexample.

(a) If there exists a simple smooth regular 2-variable parametrization of S, then S is a surface.

(b) If S is a surface, then S is a piecewise surface.

(c) If S is a piecewise surface, then S is a surface.

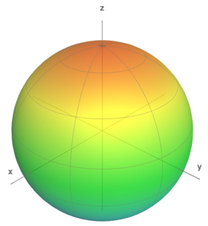
(d) If S is a piecewise surface, then S is compact.

(e) If S is a surface, then S is a 2-dimensional smooth manifold.

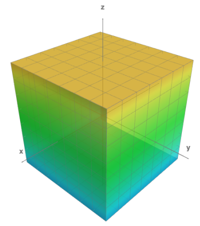
(f) If S is a 2-dimensional smooth manifold, then S is a surface.

13.1.5 For each set $S \subseteq \mathbb{R}^3$, identify whether S is a surface, a piecewise surface, both, or neither.
Hint: Can you intuitively describe the entire surface "smoothly" with only two parameters? Think of folding and gluing a piece of paper to form the shape.

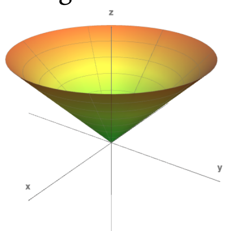
(a) A sphere



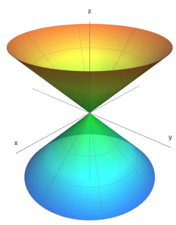
(c) A cube



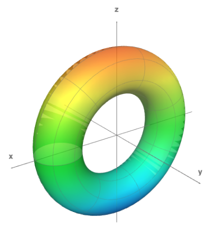
(e) A single cone



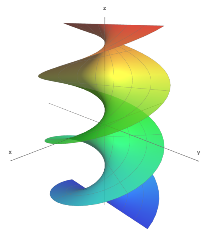
(b) A double cone



(d) A torus



(f) A helicoid



Computations

13.1.6 Let $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 8, z > 2\}$ which is parametrized by either of:

- $G(u,v) = (\sqrt{8}\sin u\cos v, \sqrt{8}\sin u\sin v, \sqrt{8}\cos u)$ for $0 \le u \le \pi/4$ and $0 \le v \le 2\pi$.
- $H(s,t) = (s,t,\sqrt{8-s^2-t^2})$ for $0 \le s^2+t^2 \le 4$.

Assume G and H are simple smooth regular parametrizations of S. Verify by definition that G is a reparametrization of H. Determine whether it is the same orientation or the opposite orientation.

13.1.7 The map $G(u,v) = (u,v,u^2+v^3)$ for $0 \le u \le 2, 0 \le v \le 3$ parametrizes part of a graph. Each map below is simple, smooth, and regular. For each map, is it a reparametrization of G? If so, determine whether it has the same orientation.

(a) $H_1(u,v) = (2u, 3v, 4u^2+27v^3)$ for $0 \le u \le 1, 0 \le v \le 1$.

(b) $H_2(u,v) = (v,u,v^2+u^3)$ for $0 \le u \le 3, 0 \le v \le 2$.

(c) $H_3(u,v) = (u,v,u^2+v^3)$ for $-2 \le u \le 0, -3 \le v \le 0$.

(d) $H_4(u,v) = (2-u, v, (2-u)^2+v^3)$ for $0 \le u \le 2, 0 \le v \le 3$.

(e) $H_5(u,v) = (2-u, 3-v, (2-u)^2+(3-v)^3)$ for $0 \le u \le 2, 0 \le v \le 3$.

13.1.8 Parametrize the upper hemisphere $H = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 4, z \ge 0\}$ using three different coordinate systems: rectangular, cylindrical, and spherical. Which of your parametrizations are simple, smooth, and regular?

13.1.9 Show that the map $G : [0,2\pi] \times [0,3] \to \mathbb{R}^4$ given by $G(s,t) = (2\cos s, 2\sin s, t)$ in Example 13.1.4 is regular.

13.1.10 Show that the parametrizations of the curve C in Examples 13.1.5 and 13.1.6 are not regular.

13.1.11 For the figure eight cylinder in Example 13.1.11, show that G is not simple.

13.1.12 For the cylinder in Example 13.1.12, show that G is simple and H is not simple.

Proofs

13.1.13 Let $f : [a,b] \times [c,d] \to \mathbb{R}$ be C^1. Prove that the graph of f is a surface in $\mathbb{R}^3$.

13.1.14 Define $S = \partial\{(x,y,z) \in \mathbb{R}^3 : 0 \le z \le 9 - x^2 - y^2\}$. Show that S is a piecewise surface.

13.1.15 Show that the figure eight cylinder in Example 13.1.11 is a piecewise surface.

13.1.16 Prove that the boundary of an ice cream cone

$$S = \partial\{(x,y,z) \in \mathbb{R}^3 : \sqrt{x^2+y^2} \le z \le \sqrt{4-x^2-y^2}\}$$

is a piecewise surface.

13.1.17 Show that the unit cube $\partial[0,1]^3$ is a piecewise surface.

13.1.18 Simple smooth regular parametrizations are brilliant because Theorem 13.1.16 shows that a set is locally the graph of a 2-variable function. By analyzing some incorrect arguments, you will discover some intricacies with Theorem 13.1.16.

(a) Semeon attempts to incorrectly prove that a helix in $\mathbb{R}^3$ is not a 2-dimensional smooth manifold.

1. *Let $C = \{(\cos t, \sin t, t) \in \mathbb{R}^3 : 0 \le t \le 1\}$.*
2. *Define $G : [0,1]^2 \to \mathbb{R}^2$ by $G(s,t) = (\cos t, \sin t, t)$.*
3. *Notice G is a 2-variable parametrization of C since $G([0,1]^2) = C$ and G is continuous.*
4. *However, G is not regular since $\partial_1 G(s,t) = (0,0,0)$ implies $\{\partial_1 G, \partial_2 G\}$ is not linearly independent.*
5. *Thus, by Theorem 13.1.16, C is not a 2-dimensional smooth manifold.*

Aside from missing details, Semeon makes a mistake in one line. Identify the flaw and briefly explain how to fix the argument. Do not fix it.

(b) Fabian attempts to show that the half cone in $\mathbb{R}^3$ is a 2-dimensional smooth manifold.

1. *Let $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = z^2, 0 \le z \le 3\}$.*
2. *Define $U = [0,2\pi] \times [0,3]$ and $G : U \to \mathbb{R}^3$ by $G(u,v) = (v\cos u, v\sin u, v)$.*
3. *Notice $G(U) = S$ and G is continuous.*
4. *Also $\partial_1 G(u,v) = (-v\sin u, v\cos u, 0)$ and $\partial_2 G(u,v) = (\cos u, \sin u, 1)$ are continuous and bounded on the interior of U, so G is smooth.*
5. *The partials are also linearly independent at every interior point in U, so G is regular.*
6. *Finally, G is injective except on $\{(u,v) \in [0,2\pi] \times [0,3] : v = 0$ or $u = 0$ or $u = 2\pi\} \subseteq \partial U$.*
7. *Thus, G is a simple smooth regular parametrization of S.*
8. *By Theorem 13.1.16, S is a 2-dimensional smooth manifold.*

Aside from missing details, Fabian makes a subtle mistake in one line. Identify the flaw and briefly explain how to fix the argument. Do not fix it.

13.1.19 Let $G : U \to \mathbb{R}^3$ be a simple smooth regular two-variable parametrization of a set $S \subseteq \mathbb{R}^3$, where $U = [a,b] \times [c,d] \subseteq \mathbb{R}^2$ is a rectangle. Let $V = [c,d] \times [a,b]$ and define $H : V \to \mathbb{R}^3$ by $H(s,t) = G(t,s)$ for $(s,t) \in V$.

(a) Prove H is a simple smooth regular two-variable parametrization of S .

(b) Prove that G is a reparametrization of H with the opposite orientation.

13.1.20 Let $G_1 : U_1 \to \mathbb{R}^3, G_2 : U_2 \to \mathbb{R}^3$, and $G_3 : U_3 \to \mathbb{R}^3$ be simple smooth regular 2-variable parametrizations. Prove that if G_1 is a reparametrization of G_2 and G_2 is a reparametrization of G_3, then G_1 is a reparametrization of G_3.

13.2. Surface area

As you did with curves in Section 11.2, you can begin to develop integral calculus for surfaces with a classical question.

What is the area of a surface in $\mathbb{R}^3$?

Surprisingly, this question is much more subtle and its historical development was fraught with pathological examples like the Schwarz lantern. Infinitesimals and your geometric intuition with the derivatives will be your key to deriving a sensible integral formula. You will again later heuristically derive its integral definition again by chopping, estimating, and refining. This central question can also be generalized to permit a surface with varying density.

What is the mass of a sheet in $\mathbb{R}^n$?

This physical problem requires you to define integrals of scalar functions over surfaces. You will see that many of the same ideas for curves will translate nicely, but the computations and derivations will be more intricate.

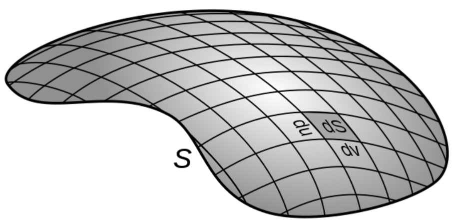
13.2.1 Surface area

A rigorous derivation of surface area involves discrete approximations to the surface; this is postponed to the next subsection. You can heuristically derive surface area using infinitesimals in a pinch. Let $S \subseteq \mathbb{R}^3$ be a surface parametrized by $G : U \to \mathbb{R}^3$.

A du by dv infinitesimal rectangle in $U \subseteq \mathbb{R}^2$ has area $du dv$. The linear approximation of G transforms this rectangle to a parallelogram with vectors

$$G(u+du,v)-G(u,v)=\partial_1 G du \qquad G(u,v+dv)-G(u,v)=\partial_2 G dv$$

in $\mathbb{R}^3$. This parallelogram has area $||\partial_1 G \times \partial_2 G|| du dv$. Integrating over U gives the surface area $\iint_U ||\partial_1 G \times \partial_2 G|| dA$.[5]



This heuristic motivates the definition of surface area.

Definition 13.2.1 Let $S \subseteq \mathbb{R}^3$ be a surface parametrized by $G : U \to \mathbb{R}^3$. The **surface area** of S is defined as

$$A(S) = \iint_U ||\partial_1 G \times \partial_2 G|| dA.$$

Remark 13.2.2 Note the integrand is technically not defined on ∂U but this issue is not serious. Indeed, since G is smooth and U is Jordan measurable, the integral is not improper and necessarily exists by Theorem 7.7.4. In other words, the surface area exists for any surface.

[5] Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.

Remark 13.2.3 Recall that $a \times b$ is the cross product of two vectors $a, b \in \mathbb{R}^3$. The norm $||a \times b||$ represents the area of the parallelogram defined by a and b. Other properties of the cross product include: for any $a, b, c \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$,

$$\begin{aligned} a \times b &= -b \times a, \\ a \times (\lambda b) &= (\lambda a) \times b = \lambda (a \times b), \\ a \times (b + c) &= (a \times b) + (a \times c). \end{aligned}$$

Each of these has a geometric interpretation; see this 3Blue1Brown video for details.

Before worrying about the choice of parametrization in Definition 13.2.1, you can investigate this integral formula with a routine computation.

Example 13.2.4 Let S be the surface in $\mathbb{R}^3$ defined by $x^2 + y^2 + z = 25$ lying above the xy-plane. It can be parametrized by $G : U \to \mathbb{R}^3$ where

$$G(u, v) = (u, v, 25 - u^2 - v^2)$$

and $U = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 \leq 25\}$. By direct calculation, $\partial_1 G(u, v) = (1, 0, -2u)$ and $\partial_2 G(u, v) = (0, 1, -2v)$ in which case

$$(\partial_1 G \times \partial_2 G)(u, v) = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ 1 & 0 & -2u \\ 0 & 1 & -2v \end{bmatrix} = (2u, 2v, 1).$$

This implies $||(\partial_1 G \times \partial_2 G)(u, v)|| = \sqrt{1 + 4u^2 + 4v^2}$ so

$$A(S) = \iint_U ||\partial_1 G \times \partial_2 G|| dA = \iint_{u^2+v^2 \leq 25} \sqrt{1 + 4u^2 + 4v^2} du dv.$$

Converting to polar coordinates with $u = r \cos\theta$ and $v = r \sin\theta$, you can check by a substitution $w = 1 + 4r^2$ that

$$A(S) = \int_0^{2\pi} \int_0^5 r\sqrt{1 + 4r^2} dr d\theta = \frac{\pi}{4} \int_1^{101} \sqrt{w} dw = \frac{\pi}{6}(101^{3/2} - 1).$$

Now, the definition of surface area does not depend on the parametrization.

Theorem 13.2.5 (Invariance of surface area) Let $G : U \to S$ and $H : V \to S$ be parametrizations of the surface $S \subseteq \mathbb{R}^3$. Assume G is a reparametrization of H. The function $||\partial_1 G \times \partial_2 G||$ is integrable on U if and only if $||\partial_1 H \times \partial_2 H||$ is integrable on V. If so,

$$\iint_U ||\partial_1 G \times \partial_2 G|| dA = \iint_V ||\partial_1 H \times \partial_2 H|| dA.$$

Proof. This is left as a challenging exercise. Write $G = H \circ \varphi$ where $\varphi : U \to V$ is the map from Definition 13.1.28. First, using the chain rule and properties of the cross product, prove

$$\forall u \in U^o, \qquad (\partial_1 G \times \partial_2 G)(u) = (\partial_1 H \times \partial_2 H)(\varphi(u)) \det D\varphi(u).$$

It is a messy calculation. Afterwards, apply a change of variables (Theorem 9.8.2); the assumptions can be verified using the many properties of H and φ. ∎

Thus, surface area is well-defined so you can define the **surface element** to be

$$dS = ||\partial_1 G \times \partial_2 G|| dA$$

where $G : U \to \mathbb{R}^3$ is a parametrization of a surface $S \subseteq \mathbb{R}^3$. Informally speaking, dS represents the area of a small piece of a surface S. This notation allows you to write the simple identity

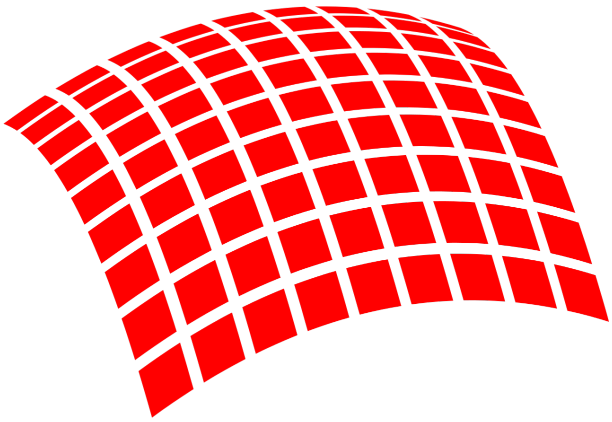
$$A(S) = \iint_S 1 dS.$$

Notice the abuse of notation to write S for the surface as a set in $\mathbb{R}^3$ and dS for the surface element. The letter "S" is used for both, but they have no relation.

13.2.2 Derivation of surface area

You will now chop, estimate, and refine your way to a more rigorous derivation of surface area. The area of a parallelogram in $\mathbb{R}^3$ defined by $a, b \in \mathbb{R}^3$ is equal to $||a \times b||$. Finding the length of a "wavy" surface will rely on this fact[6].

Let $G : R \to \mathbb{R}^3$ be a (simple smooth regular) parametrization of a surface $S \subseteq \mathbb{R}^3$. For simplicity, assume R is a rectangle in $\mathbb{R}^2$. A discrete approximation of the surface is a bunch of small parallelograms so we begin by chopping up the surface. Fix $N \in \mathbb{N}^+$. Let P_N be a regular partition of the rectangle R with N subintervals on both sides. Let $\{R_{ij}\}_{i,j}$ be the subrectangles of P_N. [7]



Each R_{ij} has width Δu and height Δv. Let (u_i, v_j) be the top right corner of R_{ij}. The piece of the surface $G(R_{ij})$ is linearly approximated by the parallelogram $dG_{(u_i,v_j)}(R_{ij})$. The area of the approximation to the surface $G(U)$ using the partition P is therefore

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \text{area}(dG_{(u_i,v_j)}(R_{ij})). \tag{13.2.1}$$

If you translate the rectangle R_{ij} so that its bottom left corner is at the origin, then the vectors $\begin{bmatrix} \Delta u \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ \Delta v \end{bmatrix}$ correspond to its sides. Mapped under the differential $dG_{(u_i,v_j)}$, you obtain a parallelogram defined by the vectors

$$dG_{(u_i,v_j)}\left(\begin{bmatrix} \Delta u \\ 0 \end{bmatrix}\right) = \partial_1 G(u_i, v_j)\Delta u \quad \text{and} \quad dG_{(u_i,v_j)}\left(\begin{bmatrix} 0 \\ \Delta v \end{bmatrix}\right) = \partial_2 G(u_i, v_j)\Delta v$$

[6]Watch for geometric and visual intuition on the cross product.
[7]Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.

The area of this parallelogram is

$$||\partial_1 G(u_i, v_j)\Delta u \times \partial_2 G(u_i, v_j)\Delta v|| = ||\partial_1 G(u_i, v_j) \times \partial_2 G(u_i, v_j)||\Delta u \Delta v.$$

This must be equal to area $\left(dG_{(u_i,v_j)}(R_{ij})\right)$ yielding a discrete approximation of surface area via (13.2.1). Taking a limit of these discrete approximation should presumably yield

$$\lim_{N\to\infty} \sum_{i=1}^{N}\sum_{j=1}^{N} ||(\partial_1 G \times \partial_2 G)(u_i, v_j)||\Delta u \Delta v = \iint_U ||\partial_1 G \times \partial_2 G||dA.$$

This better illustrates that our integral definition should be sensible.

13.2.3 Surface integrals of scalar functions

Next you will slightly generalize your new ideas for surface area.

> *What is the mass of a surface in $\mathbb{R}^3$ with variable density?*

More formally, you want to integrate a scalar function over a surface. Infinitesimals make this a straightforward informal process. Recall the symbol dS is the surface area element. As before, it has no formal meaning.

> *An infinitesimal piece of the surface S parametrized by $G : U \to \mathbb{R}^3$ has area*
>
> $$dS = ||\partial_1 G \times \partial_2 G||dA.$$
>
> *The contribution of f on this piece is therefore given by*
>
> $$f\,dS = (f \circ G)||\partial_1 G \times \partial_2 G||dA$$
>
> *Integrating over U gives the surface integral*
>
> $$\iint_S f\,dS = \iint_U (f \circ G)||\partial_1 G \times \partial_2 G||dA.$$

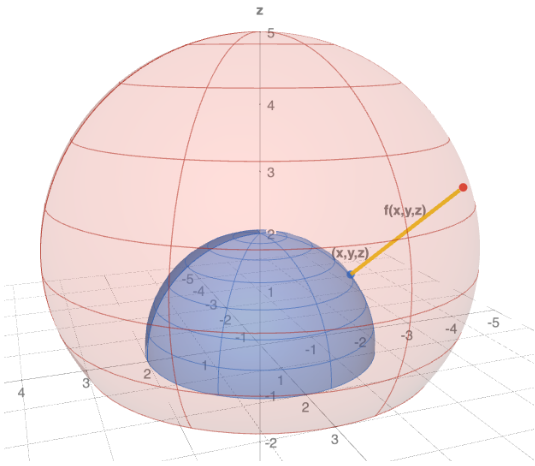
These short few sentences stumble quickly onto a formal definition.

Definition 13.2.6 Let $S \subseteq \mathbb{R}^3$ be a surface parametrized by $G : U \to \mathbb{R}^3$. Let f be a bounded real-valued function on S. The **scalar surface integral of f over S** is given by

$$\iint_S f\,dS := \iint_U (f \circ G)||\partial_1 G \times \partial_2 G||dA.$$

If this integral exists, then f is **integrable on the surface S**.

For a surface S in $\mathbb{R}^3$, this scalar surface integral can be interpreted as mass if f is non-negative. It also has a (not commonly used) geometric interpretation as the "net volume of the solid traced out by lines of height f orthogonal to the surface S", as illustrated in this Math3D demo.



As you might expect, the scalar surface integral does not depend on the parametrization.

Theorem 13.2.7 (Invariance of scalar surface integrals) Let S be surface in $\mathbb{R}^3$. Let $G : U \to \mathbb{R}^3$ and $H : V \to \mathbb{R}^3$ be parametrizations of S. Let f be a bounded real-valued function defined on S. Assume G is a reparametrization of H. The function $(f \circ G)||\partial_1 G \times \partial_2 G||$ is integrable on U if and only if $(f \circ H)||\partial_1 H \times \partial_2 H||$ is integrable on V. If so,

$$\iint_U (f \circ G)||\partial_1 G \times \partial_2 G||dA = \iint_V (f \circ H)||\partial_1 H \times \partial_2 H||dA.$$

Proof. Omitted. The strategy is identical to Theorem 13.2.5. ■

Calculations of scalar surface integrals are quite similar to surface area.

Example 13.2.8 You can calculate the scalar surface integral

$$\iint_S (z+1)dS,$$

which was illustrated above where S is the (blue) upper hemisphere of radius 2 and $f(x,y,z) = z+1$. Since $f \geq 0$ on S, you can also intrepret this integral as the mass of S with density f. Either way, you can calculate this quantity. Note S is parametrized by $G : U \to \mathbb{R}^3$, where

$$G(u,v) = (2\sin u \cos v, 2\sin u \sin v, 2\cos u)$$

and $U = \{(u,v) \in \mathbb{R}^2 : 0 \leq u \leq \frac{\pi}{2}, 0 \leq v \leq 2\pi\}$. Observe that

$$\partial_1 G(u,v) = (2\cos u \cos v, 2\cos u \sin v, -2\sin u) \qquad \partial_2 G(u,v) = (-2\sin u \sin v, 2\sin u \cos v, 0)$$

so, by direct calculation, you can check that

$$\begin{aligned}(\partial_1 G \times \partial_2 G)(u,v) &= \det\begin{bmatrix} e_1 & e_2 & e_3 \\ 2\cos u \cos v & 2\cos u \sin v & -2\sin u \\ -2\sin u \sin v & 2\sin u \cos v & 0 \end{bmatrix} \\ &= (4\sin^2 u \cos v, 4\sin^2 u \sin v, 4\sin u \cos u).\end{aligned}$$

Using trigonometric identities, $||\partial_1 G \times \partial_2 G|| = 4|\sin u| = 4\sin u$ for $0 \leq u \leq \frac{\pi}{2}$. Hence,

$$\begin{aligned}\iint_S (z+1)dS &= \int_0^{2\pi}\int_0^{\pi/2}(2\cos u + 1)||(\partial_1 G \times \partial_2 G)(u,v)||du dv \\ &= \int_0^{2\pi}\int_0^{\pi/2}(2\cos u + 1)\cdot 4\sin u du dv = 16\pi\end{aligned}$$

Thus, the net volume traced out by f along S (or the mass of S with density f) is 16π.

This investigation of surface area and scalar surface integrals closely paralleled arc length and scalar line integrals. Your slow and careful process of generalizing concepts allowed you to design the key components without too many complications. Over the next few sections, you will further extend these ideas with integrals of scalar functions over surfaces (i.e. mass of a sheet with variable density) to integrals of *vector fields* over *oriented* surfaces (i.e flux across a sheet). Oriented surfaces are shockingly more subtle than oriented curves, so you will need to dedicate special efforts to describing them.

Exercises for Section 13.2

Concepts and definitions

13.2.1 Let $a, b, c \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$. Let e_1, e_2, e_3 denote the standard basis vectors in $\mathbb{R}^3$. Which of the following are true or false?

(a) $a \times b$ is the area of the parallelogram in $\mathbb{R}^3$ with sides a and b.

(b) $\frac{1}{2}||a \times b||$ is the area of the triangle in $\mathbb{R}^3$ with sides a and b.

(c) $a \times a = 0$

(d) $a \times \lambda = 0$

(e) $a \times b = -b \times a$

(f) $\lambda(a \times b) = (\lambda a) \times (\lambda b)$

(g) $e_1 \times e_2 = e_3$

(h) $e_1 \times e_3 = e_2$

13.2.2 Let $S \subseteq \mathbb{R}^3$ be a surface. Let $G : U \to \mathbb{R}^3$ and $H : V \to \mathbb{R}^3$ be parametrizations of S. Which of the following quantities are equal to the surface area of S?

(a) $\displaystyle\iint_S 1 dS$

(b) $\displaystyle\iint_S 1 dA$

(c) $\displaystyle\iint_U 1 du dv$

(d) $\displaystyle\iint_U ||\partial_1 G \times \partial_2 G|| dA$

(e) $\displaystyle\iint_U ||\partial_2 G \times \partial_1 G|| dA$

(f) $\displaystyle\iint_V ||\partial_1 H \times \partial_2 H|| dA$

13.2.3 Let $f : \mathbb{R}^3 \to [0, \infty)$ be continuous and S be a surface in $\mathbb{R}^3$ parametrized by $G : U \to \mathbb{R}^3$. Let V be the volume between the two illustrated surfaces lying above the $z = 0$ plane. Which quantities represent V? *Hint*: Some of these are nonsense.

(a) $\displaystyle\iint_S f \times dS$

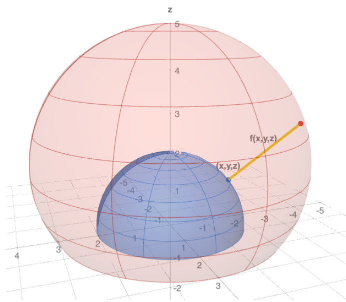
(b) $\displaystyle\iint_S f \cdot T ds$

(c) $\displaystyle\iint_S f dS$

(d) $\displaystyle\iint_S f dx dy$

(e) $\displaystyle\iint_U f(G(s,t))||\partial_1 G \times \partial_2 G|| dA$

(f) $\displaystyle\iint_U f(G(s,t))(\partial_1 G \times \partial_2 G) dA$



Computations

13.2.4 Define the paraboloid $S = \{(x, y, z) \in \mathbb{R}^3 : z = y^2, 0 \le x \le 2, -1 \le y \le 1\}$. Note S is parametrized by

$$P(s,t) = (s, t, t^2), \qquad 0 \le s \le 2, -1 \le t \le 1.$$

(a) Compute the partials $\partial_1 P$ and $\partial_2 P$.

(b) Compute $||\partial_1 P \times \partial_2 P||$.

(c) Express $A(S)$, the surface area of S, as an iterated double integral and then evaluate it.

13.2.5 Use cylindrical coordinates to show that the surface area of sphere of radius R is $4\pi R^2$.

13.2.6 Use spherical coordinates to show that the surface area of a sphere of radius R is $4\pi R^2$.

13.2.7 Let S be a surface parametrized by $G : U \to \mathbb{R}^3$ where $U = [1,3] \times [2,7]$.

(a) Estimate the surface area of S given that $DG(1,2) = \begin{bmatrix} -1 & 9 \\ 8 & 6 \\ 2 & 3 \end{bmatrix}$.

(b) Improve your estimate if you also learn that $DG(1,4) = \begin{bmatrix} -1 & 7 \\ 6 & 3 \\ 1 & 2 \end{bmatrix}$.

13.2.8 Find the surface area of $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z^2, 1 \leq z \leq 2\}$.

13.2.9 Find the surface area of $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 4, (x-1)^2 + y^2 \leq 1\}$.

13.2.10 Evaluate $\displaystyle\iint_S \frac{z}{\sqrt{x^2+y^2}} dS$ where S is the upper hemisphere $x^2 + y^2 + z^2 = 9$.

13.2.11 Evalute $\displaystyle\iint_S (x^2 + xy - z) dS$ where S is the surface defined by $R(u, v) = (u, 2v, u^2)$ for $0 \leq u, v \leq 1$.

Proofs

13.2.12 Let S be a surface in $\mathbb{R}^3$. Let f and g be continuous real-valued functions on S. Fix $\lambda \in \mathbb{R}$. Prove that

$$\iint_S (f + \lambda g) dS = \iint_S f\, dS + \lambda \iint_S g dS.$$

13.2.13 Let $G : U \to S$ and $H : V \to S$ be parametrizations of a surface $S \subseteq \mathbb{R}^3$. Assume G is a reparametrization of H. Prove that

$$\forall u \in U^o, \qquad (\partial_1 G \times \partial_2 G)(u) = (\partial_1 H \times \partial_2 H)(\varphi(u)) \det D\varphi(u).$$

13.2.14 Let $G : U \to S$ and $H : V \to S$ be parametrizations of the surface $S \subseteq \mathbb{R}^3$. Assume G is a reparametrization of H. Prove that the function $||\partial_1 G \times \partial_2 G||$ is integrable on U if and only if $||\partial_1 H \times \partial_2 H||$ is integrable on V. If so, show that

$$\iint_U ||\partial_1 G \times \partial_2 G|| dA = \iint_V ||\partial_1 H \times \partial_2 H|| dA.$$

Applications and beyond

13.2.15 A thin sheet in the shape of a surface S in $\mathbb{R}^3$ has a continuous mass density function $\rho(x, y, z)$.

(a) Conjecture a formula for the mass of the sheet.

(b) Heuristically justify your formula using infinitesimals in 2 or 3 sentences.

(c) Find the mass of the spiral sheet S parametrized by $G(s, t) = (3t-3, (s+1)\cos(2t), (s+1)\sin(2t))$ for $0 \leq t \leq 6\pi$ and $0 \leq s \leq 1$ with density $\rho(x, y, z) = y^2+z^2$. Setup your integral and evaluate it with WolframAlpha.

13.2.16 Recall the derivation of surface area by infinitesimals.

1. *A du by dv infinitesimal rectangle in $U \subseteq \mathbb{R}^2$ has area $dudv$.*
2. *The linear approximation of G transforms this rectangle to a parallelogram with vectors*

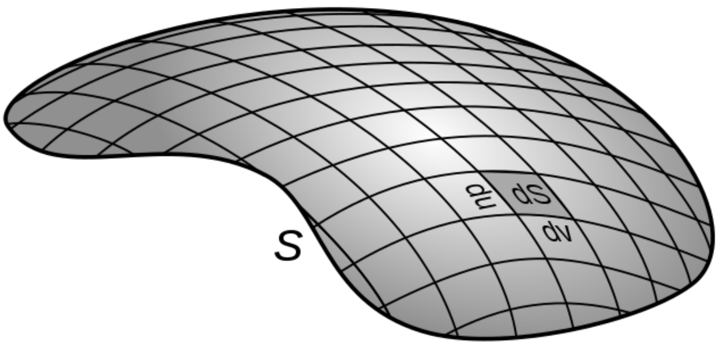
$$G(u+du,v)-G(u,v)=\partial_1 G(u,v)du \in \mathbb{R}^3 \qquad G(u,v+dv)-G(u,v)=\partial_2 G(u,v)dv \in \mathbb{R}^3$$

3. *This parallelogram has area $||(\partial_1 G \times \partial_2 G)(u,v)||dudv$.*
4. *Integrating over U gives the surface area $\iint_U ||\partial_1 G \times \partial_2 G||dA$.*

Line 2 could use some better visuals. Label the diagram below with the 5 infinitesimal quantities:

$$G(u,v) \qquad G(u+du,v) \qquad G(u,v+dv) \qquad \partial_1 G(u,v)du \qquad \partial_2 G(u,v)dv$$

Three of these are points, and two of them are vectors. You will need to draw these two vectors.[8]



13.2.17 Let S be the surface formed by rotating the graph of $y = f(x)$ around the x axis between $x = a$ and $x = b$. Assume that $f(x) \geq 0$ for $a \leq x \leq b$. Show that the surface area of S is

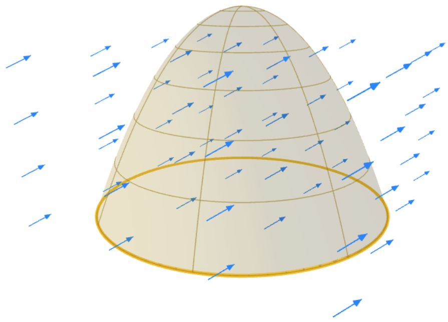
$$2\pi \int_a^b f(x)\sqrt{1+f'(x)^2}dx.$$

13.2.18 A thin sheet S in $\mathbb{R}^3$ is bent into the shape of the paraboloid $z = x^2 + y^2$ with $z \leq 4$. If the mass density $\rho(x,y,z)$ at a point (x,y,z) is directly proportional to its distance from the xy-plane, find the mass of the sheet.

[8] Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.

13.3. Orientation and relative boundary

The physics of vector fields in $\mathbb{R}^3$ continues to model forces and fluid flows. You have studied one aspect of this situation with work done along curves, but the geometry of three dimensions introduces new questions with surfaces. For instance, how much water flows across a surface? Or how much water flows along the edge of a surface? As with line integrals and oriented curves (Section 11.3), you will encounter some preliminary problems about surfaces before resolving these physical questions.



The above diagram shows a velocity field F traveling through a surface S. On one hand, you can say that the *upward* flow of F through S is positive since all of the arrows point upward as they cross S. On the other hand, you can also say that the *downward* flow of F through S is negative for the same reason. You will therefore need to take into account the direction of motion across the surface S. This suggests your first preliminary problem.

How can you orient a surface?

The above diagram also highlights the circular "edge" of the surface S. Is this the topological boundary of S? Unfortunately not. Since S is closed, the topological boundary ∂S is equal to S itself. To discuss the amount of water flowing along this edge, you will need a more subtle topological description that identifies this "edge". This poses your second preliminary problem.

How can you define the edge of a surface?

Before attempting to describe the amount of flow across a surface, you are therefore compelled to handle these intriguing geometric and topological questions about surfaces. Surfaces are more subtle than curves, so this will require very careful definitions to be sufficiently rigorous and avoid dependence on a parametrization. This section will be dedicated to these tasks and, in the next section, you will be ready to describe water flow across a surface.

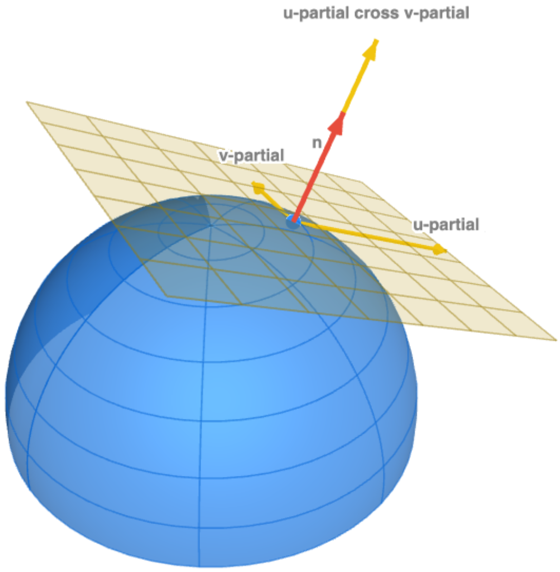
13.3.1 Unit normal

To orient surfaces in $\mathbb{R}^3$, you can try to draw from your experience with curves in $\mathbb{R}^n$. Curves in $\mathbb{R}^n$ are oriented using their unit tangent vectors (Section 11.3.1); this was acceptable since there are only two directions to move along a curve. For a surface in $\mathbb{R}^3$, tangent vectors alone will be insufficient; there are infinitely many directions to move along a (2-dimensional) surface.

More seriously, tangent vectors do not inherently describe "moving across" a surface. This requires a notion of orthogonality. As you may recall from Section 12.3.1, you can also orient curves in $\mathbb{R}^2$ with a *unit normal*! This spawns your central aim to define orientation of a surface.

Assign a consistent choice of normal vector to the surface.

Let $S \subseteq \mathbb{R}^3$ be a surface parametrized by the map $G : U \to \mathbb{R}^3$. Since S has two linearly independent tangent vectors $\partial_1 G(u,v)$ and $\partial_2 G(u,v)$ at all points $(u,v) \in U^o$, their cross product $(\partial_1 G \times \partial_2 G)(u,v)$ gives a natural way to describe a unit normal at that point.



The unit normal satisfies the righthand rule: if your index finger is $\partial_1 G$ and your middle finger is $\partial_2 G$ then your thumb is $\partial_1 G \times \partial_2 G$ which gives the unit normal n. This suggests one way to assign a unit normal.

Definition 13.3.1 Let $G : U \to \mathbb{R}^3$ be a parametrization of a surface in $\mathbb{R}^3$. The **unit normal (of the parametrization** G**)** is given by

$$\frac{\partial_1 G \times \partial_2 G}{||\partial_1 G \times \partial_2 G||},$$

which is a C^1 function defined on the interior of $U \subseteq \mathbb{R}^2$.

For a given parametrization, this can be calculated in a straightforward manner.

Example 13.3.2 Let S be the upper hemisphere of radius 2 in $\mathbb{R}^3$ parametrized by G. Note S can be parametrized by $G : U \to \mathbb{R}^3$ where

$$G(u,v) = (2\sin u \cos v, 2\sin u \sin v, 2\cos u)$$

and $U = \{(u,v) \in \mathbb{R}^2 : 0 \le u \le \frac{\pi}{2}, 0 \le v \le 2\pi\}$. For $(u,v) \in U^o = (0, \frac{\pi}{2}) \times (0, 2\pi)$, observe that

$$\partial_1 G(u,v) = (2\cos u \cos v, 2\cos u \sin v, -2\sin u) \qquad \partial_2 G(u,v) = (-2\sin u \sin v, 2\sin u \cos v, 0)$$

so, by direct calculation, you can check that

$$\begin{aligned}(\partial_1 G \times \partial_2 G)(u,v) &= \det \begin{bmatrix} e_1 & e_2 & e_3 \\ 2\cos u \cos v & 2\cos u \sin v & -2\sin u \\ -2\sin u \sin v & 2\sin u \cos v & 0 \end{bmatrix} \\ &= (4\sin^2 u \cos v, 4\sin^2 u \sin v, 4\sin u \cos u).\end{aligned}$$

Notice $||(\partial_1 G \times \partial_2 G)(u,v)|| = 4|\sin u| = 4\sin u$ as $0 < u < \frac{\pi}{2}$. Therefore, for $0 < u < \frac{\pi}{2}$ and $0 < v < 2\pi$, the unit normal of the parametrization G is given by

$$\frac{(\partial_1 G \times \partial_2 G)(u,v)}{||(\partial_1 G \times \partial_2 G)(u,v)||} = (\sin u \cos v, \sin u \sin v, \cos u).$$

This means the unit normal on a sphere points in the same direction as the position vector directly away from the origin. View this geometrically intuitive fact in this Math3D demo.

As with all definitions for surfaces, you must investigate how the unit normal depends on the choice of parametrization. Fortunately, your definition of reparametrizations was designed precisely for this moment.

Lemma 13.3.3 Let $G : U \to \mathbb{R}^3$ and $H : V \to \mathbb{R}^3$ be parametrizations of a surface $S \subseteq \mathbb{R}^3$. Assume G is a reparametrization of H with $\varphi : U \to V$ satisfying $G = H \circ \varphi$.

(a) If G is a reparametrization of H with the same orientation, then

$$\frac{(\partial_1 G \times \partial_2 G)(u,v)}{||(\partial_1 G \times \partial_2 G)(u,v)||} = \frac{(\partial_1 H \times \partial_2 H)(s,t)}{||(\partial_1 H \times \partial_2 H)(s,t)||}$$

for $(u,v) \in U^o$ and $(s,t) = \varphi(u,v) \in V^o$.

(b) If G is a reparametrization of H with the opposite orientation, then

$$\frac{(\partial_1 G \times \partial_2 G)(u,v)}{||(\partial_1 G \times \partial_2 G)(u,v)||} = -\frac{(\partial_1 H \times \partial_2 H)(s,t)}{||(\partial_1 H \times \partial_2 H)(s,t)||}$$

for $(u,v) \in U^o$ and $(s,t) = \varphi(u,v) \in V^o$.

Remark 13.3.4 This closely parallels Theorem 11.3.1 for oriented curves.

Proof. This is left as an exercise. Use Definition 13.1.28, chain rule, and cross product properties in Remark 13.2.3. ∎

You can study an explicit example of this lemma in action.

Example 13.3.5 Let S be the upper hemisphere of radius 2. Consider another parametrization $H : V \to \mathbb{R}^3$ given by

$$H(s,t) = (2\sin t \cos s, 2\sin t \sin s, 2\cos t)$$

where $V = \{(s,t) \in \mathbb{R}^2 : 0 \le s \le 2\pi, 0 \le t \le \pi/2\}$. Notice $H(s,t) = G(t,s)$ where G is the parametrization from Example 13.3.2. Therefore, by a cross product property, for $(s,t) \in V$,

$$(\partial_1 H \times \partial_2 H)(s,t) = (\partial_2 G \times \partial_1 G)(t,s) = -(\partial_1 G \times \partial_2 G)(t,s).$$

This occurs because H has the opposite orientation as G. You can verify this directly by Definition 13.1.28, because $H = G \circ \varphi$ where the continuous invertible map $\varphi : V \to U$ satisfies $\varphi(s,t) = (t,s)$. Note φ is C^1 on the interior of U and

$$\det D\varphi = \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1 < 0.$$

There are infinitely many ways to parametrize S and get the same (or opposite) orientation.

Lemma 13.3.3 confirms that reparametrization does not affect the unit normal provided the parametrizations have the same orientation. This outcome paves the way for oriented surfaces.

13.3.2 Oriented surfaces

With your definitions of parametrizations and surfaces thus far, it is natural to define oriented surfaces similar to how you defined oriented curves in Definition 11.3.2.

Definition 13.3.6 An **oriented surface** S is a set of (2-variable simple smooth regular) parametrizations that are reparametrizations of each other with the same orientation.

Remark 13.3.7 It is a common abuse of notation for S to also denote the set of points in $\mathbb{R}^3$.

Although this definition and language are natural, it poses a serious question.

Can you assign a consistent choice of unit normal to an oriented surface?

In other words, the unit normal should *continuously* vary on the oriented surface. Example 13.3.2 with the hemisphere illustrates this phenomenon as the unit normal always points away from the origin. Of course, there are many equivalent parametrizations for the same surface. You will want to consider these parametrizations as the same provided they have the same consistent choice of orientation.

Definition 13.3.8 Let S be an oriented surface in $\mathbb{R}^3$. A **unit normal** $n : S \to S^2$ is a *continuous* function from the set $S \subseteq \mathbb{R}^3$ to the set of unit vectors $S^2 = \{w \in \mathbb{R}^3 : ||w|| = 1\} \subseteq \mathbb{R}^3$ which, for any parametrization $G : U \to \mathbb{R}^3$ of the oriented surface S, is defined by

$$n(G(u,v)) = \frac{(\partial_1 G \times \partial_2 G)(u,v)}{||(\partial_1 G \times \partial_2 G)(u,v)||}$$

for all $(u,v) \in U^o$.

Remark 13.3.9 You may write $n(u,v)$ or simply n instead of $n(G(u,v))$. The parametrization is usually understood from context so the abuse of notation $n = n(u,v) = n(G(u,v))$ is acceptable.

Lemma 13.3.3 shows that a unit normal of an oriented surface is independent of its parametrization and hence well-defined. Since the unit normal is a continuous function, you can “fill in" the remaining undefined points of the unit normal.

Example 13.3.10 Consider the upper hemisphere S from Example 13.3.2 and the same parametrization G. Let $n : S \to S^2$ be the unit normal pointing away from the origin. For $0 \le u \le \pi/2$ and $0 \le v \le 2\pi$,

$$(\partial_1 G \times \partial_2 G)(u,v) = 0 \iff ||(\partial_1 G \times \partial_2 G)(u,v)|| = 4 \sin u = 0 \iff u = 0.$$

Therefore, by definition of the unit normal, for all $0 < u \le \pi/2, 0 \le v \le 2\pi$,

$$n = n(G(u,v)) = (\sin u \cos v, \sin u \sin v, \cos u).$$

This can be continuously extended to include $u = 0$, but that is not obvious beforehand.

Now, *once you choose* a unit normal on an oriented surface, it is independent of any parametrization with the same orientation, but the choice of unit normal (i.e. choice of orientation) is not unique. There are usually exactly two choices.

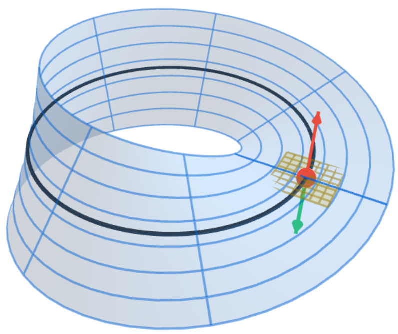
Example 13.3.11 Let S be the upper hemisphere of radius 2 again. Let n be the unit normal of S parametrized by G and let n' be the unit normal of S parametrized by H. Example 13.3.5 shows that G and H have opposite orientations in which case

$$n' = -n.$$

These constitute the two unique choices of orientation.

Remark 13.3.12 In many scenarios, you are given a non-oriented surface S and must create an oriented surface from S by defining a unit normal $n : S \to S^2$. This process is referred to as "*choosing a unit normal*" or "*specifying an orientation*" or "*orienting the surface*", because there are two possible choices.

Surprisingly, there are *non-orientable* surfaces! The Mobius strip is the quintessential example. Play with this Math3D demo to watch an invalid "unit normal" *continuously* traverse a path and switch orientation.



Such non-orientable examples are more closely studied in differential geometry, such as a dedicated course on curves and surfaces. In this text, you will only handle orientable surfaces, i.e. surfaces with a unit normal, so any surface you consider will always have exactly two choices of orientation. This completes your study of oriented surfaces.

13.3.3 Relative boundary of a surface

Now, you must address the remaining motivational question of this section.

> *How can you define the "edge" of a surface?*

You have studied the *topological* boundary for sets S in $\mathbb{R}^n$. If S is a closed set with no interior points in $\mathbb{R}^3$ (like any surface) then $\partial S = S$. This is not helpful if you want to distinguish the "edge" of a surface. You want to define the boundary *relative to the surface*.

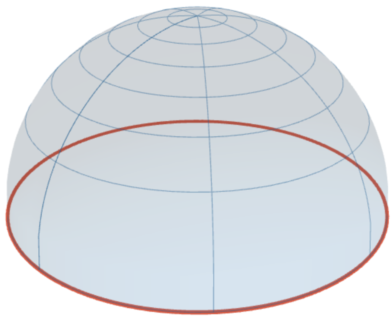
Example 13.3.13 For most surfaces, you can intuitively guess the *relative boundary*.

- The relative boundary of the upper hemisphere

  $$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 2, z \geq 0\}$$

  should be the circle lying in the $z = 0$ plane

  $$C = \{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 = 2\}.$$

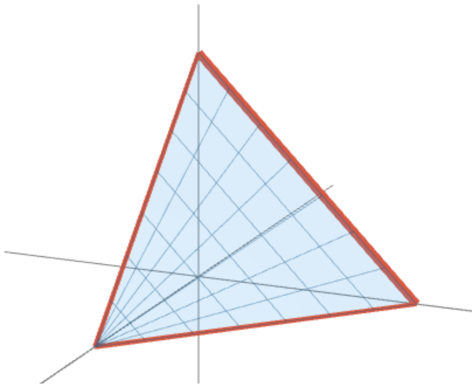


- The relative boundary of the triangle

$$S = \{(x,y,z) \in \mathbb{R}^3 : x+y+z = 1, x \geq 0, y \geq 0, z \geq 0\}$$

should be the three edges $C_1 \cup C_2 \cup C_3$ given by

$C_1 = \{(t, 1-t, 0) : 0 \leq t \leq 1\}$, $C_2 = \{(0, t, 1-t) : 0 \leq t \leq 1\}$, $C_3 = \{(1-t, 0, t) : 0 \leq t \leq 1\}$.

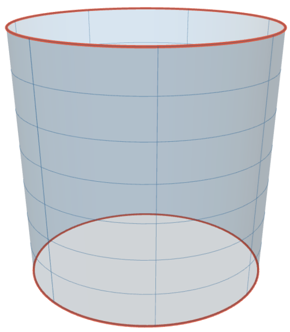


- The relative boundary of the cylinder

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \leq z \leq 3\}$$

should be the two circles $C_1 \cup C_2$ given by

$$C_1 = \{(x,y,0) \in \mathbb{R}^3 : x^2 + y^2 = 4\} \quad C_2 = \{(x,y,3) \in \mathbb{R}^3 : x^2 + y^2 = 4\}.$$



In future sections, you need only to produce these guesses without proof, but the notion deserves a formal definition to provide some geometric intuition.

Definition 13.3.14 Let S be a piecewise surface in $\mathbb{R}^3$.

- A point $p \in S$ is a **(relative) boundary point**[9] of S if there exists an open set $V \subseteq \mathbb{R}^3$ containing p, an open set $U \subseteq \mathbb{R}^2$, a continuous invertible map

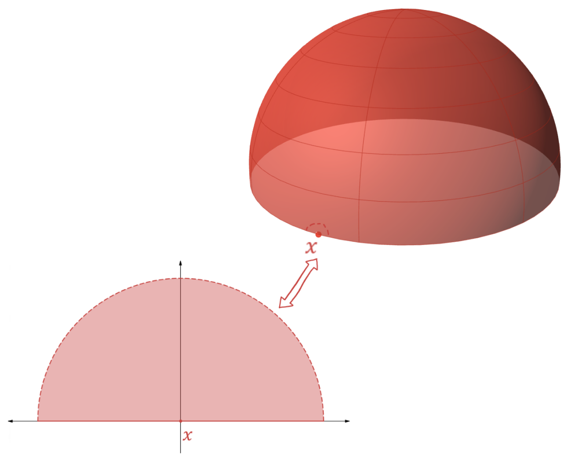
$$\varphi : U \cap \{(x, y) \in \mathbb{R}^2 : y \geq 0\} \to V \cap S$$

such that the inverse φ^{-1} is continuous and $\varphi^{-1}(p)$ lies on the x-axis.
- The **(relative) boundary** of S, denoted ∂S, is the set of its (relative) boundary points.

Remark 13.3.15 The abuse of notation with ∂S to denote both the topological boundary and the relative boundary can be confusing. These are also confusingly both referred to as "the boundary", so you must rely on context. It is usually clear which one is intended. For the remainder of this chapter, "boundary" of a surface will mean "relative boundary".

This complicated formal definition captures a simple idea in the figure below.

The relative boundary of a surface in $\mathbb{R}^3$ looks nearby like the edge of a plane in $\mathbb{R}^2$.



Checking this definition formally in an explicit example can be quite technical. An example is included below if you are interested.

Example 13.3.16 Define the surface $S = \{(x, y, z) \in \mathbb{R}^3 : z = \sin(xy), 0 \leq x \leq 3, 7 \leq y \leq 12\}$ parameterized by a restriction of the graph function $G : \mathbb{R}^2 \to \mathbb{R}^3$ defined as

$$G(x, y) = (x, y, \sin(xy)).$$

You can show $p = (2, 7, \sin 14)$ is a relative boundary point of S. Define $\psi : \mathbb{R}^2 \to G(\mathbb{R}^2)$ by

$$\psi(u, v) = (u, v + 7, \sin(u(v + 7))), \qquad \forall (u, v) \in \mathbb{R}^2,$$

which has inverse

$$\psi^{-1}(x, y, z) = (x, y - 7), \qquad \forall (x, y, z) \in G(\mathbb{R}^2).$$

[9]This definition is not quite the modern standard; see a course in differential geometry.

You can see that both ψ and ψ^{-1} are continuous. Moreover, $\psi^{-1}(p) = (2,0)$ which lies on the x-axis. Define the open sets

$$V = (1.5, 2.5) \times (6.5, 7.5) \times (-2, 2) \qquad \text{and} \qquad U = (1.5, 2.5) \times (-0.5, 0.5)$$

so U is the projection of V onto the xy-plane translated down by 7 units. Since $|\sin| \leq 1$, you can verify directly that $\psi(U) = V \cap S$. Thus, $\varphi = \psi|_U : U \to V$ will satisfy all the desired properties in Definition 13.3.14. Hence, $p = (2, 7, 14)$ is a relative boundary point of S.

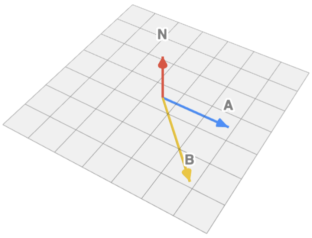
You will write down the relative boundary ∂S of a surface S without rigorous justification, so do not worry about the technical details in the above example. Your primary goal with the formal definition is to interpret it geometrically and visually in examples.

Overall, you have achieved your key preliminary goals this section. You can orient (most) surfaces with a unit normal and this definition is independent of parametrization. You can also identify the "edge" of a surface with the intuitive notion of relative boundary; this will re-appear in the following chapter to analyze the swirly-ness along the edge of a surface. More immediately, the development of surfaces, surface area, and orientation allow you to conclude this chapter by finally defining the flux of a vector field across a surface.

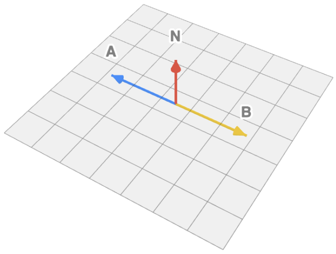
Exercises for Section 13.3

Concepts and definitions

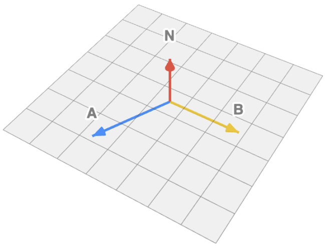
13.3.1 It is time to practice the righthand rule! Assume the red vector N is orthogonal to both the blue vector A and the yellow vector B. Is $A \times B$ a positive scalar multiple of N?



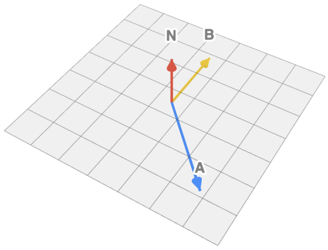
(a)



(c)

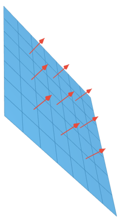


(b)

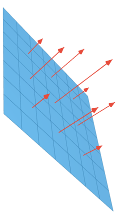


(d)

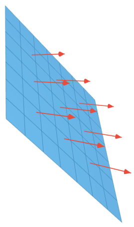
13.3.2 Which figures below show an orientation of the plane? If not, explain why.



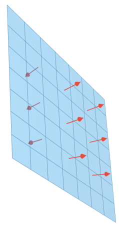
(a)



(b)



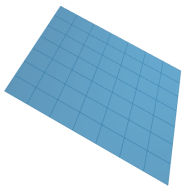
(c)



(d)

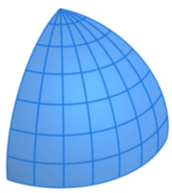
13.3.3 The boundary of a surface is usually easy to distinguish visually. For each of the surfaces below, draw the boundary of the surface.

(a) Parallelogram

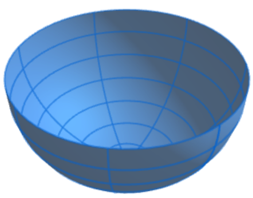


(c) Sphere

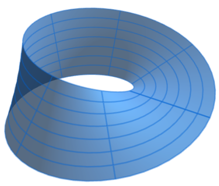
(e) First octant of a sphere



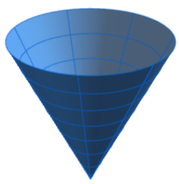
(b) Lower hemisphere



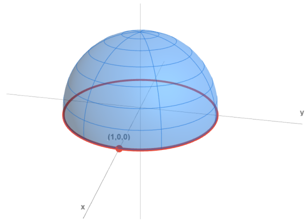
(d) Mobius strip



(f) Cone



13.3.4 The (relative) boundary of a surface is often easy to recognize visually but proving that a point is on the boundary is technical.



(a) The circle $\{(x,y,0) \in \mathbb{R}^3 : x^2+y^2=1\}$ is the boundary of the upper hemisphere. Draw a picture proof illustrating that the point $(1,0,0)$ is a boundary point.

(b) Gina and Hassan are confused about the definition of a boundary point of the surface. They know that the hemisphere S is parameterized by the function

$$F(s,t) = (\cos(s)\sin(t), \sin(s)\sin(t), \cos(t)) \quad 0 \le s \le 2\pi, 0 \le t \le \pi/2$$

and that $F(0,0) = (0,0,1)$.

> Hassan asserts: *"Since the point $(0,0)$ is on the boundary of the domain of F, then $(1,0,0)$ has to be a boundary point of the hemisphere."*
>
> Gina asserts: *"$(0,0,1)$ isn't a boundary point, because it is not on the circle!"*

Who is correct? Clarify the relationship between the boundary of the domain of F and the boundary of the surface of S for Gina and Hassan.

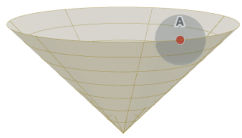
13.3.5 The formal definition of relative boundary points for a piecewise surface $S \subseteq \mathbb{R}^3$ is rather technical; review Definition 13.3.14 again. Abed, Brita, and Chang are deciding which points belong to the relative boundary of the cone

$$S = \{(x,y,z) \in \mathbb{R}^3 : z = \sqrt{x^2+y^2}, 0 \le z \le 1\}.$$

Instead of formally applying the definition, they give informal explanations that attempt to match the idea behind the definition. Use this Math3D demo to help you visualize.

(a) Abed argues the point $A \in \mathbb{R}^3$ is a relative boundary point of S.

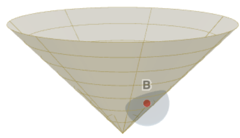
"The map $G : [0,2\pi] \times [0,1] \to \mathbb{R}^3$ given by $G(\theta, z) = (z\cos\theta, z\sin\theta, z)$ parametrizes S. Notice G maps a point from the boundary $\partial([0,2\pi]\times[0,1])$ to the point A, so A belongs to the relative boundary ∂S."



Does Abed's explanation appear consistent with the idea behind the definition? If not, explain.

(b) Brita argues the point $B \in \mathbb{R}^3$ is a relative boundary point of S.

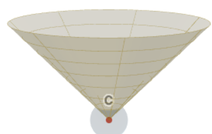
"Take V to be a small open ball around B. Cut out $V \cap S$, which is a piece of the cone. Fold this piece $V \cap S$ in half along the point B and then flatten it to $\mathbb{R}^2$. The image of B can be made to lie on the x-axis. Thus, B belongs to the relative boundary ∂S."



Does Brita's explanation appear consistent with the idea behind the definition? If not, explain.

(c) Chang argues the point $B \in \mathbb{R}^3$ is a relative boundary point of S.

"Take V to be a small open ball around C. Cut out $V \cap S$, which a piece of the cone. Flatten $V \cap S$ to $\mathbb{R}^2$ and obtain a disk. Cut this disk from the edge to the centre C. Open up the disk along this cut to get a semi-disk with C along the x-axis. Thus, $C \in \partial S$."



Does Chang's explanation appear consistent with the idea behind the definition? If not, explain.

(d) Now, decide for yourself: which of A, B, C are relative boundary points of the cone?

Computations

13.3.6 There is a shortcut to identifying orientations of a surface. Let S be the lower hemisphere of radius 1 centered at the origin. Orient S so that its unit normal n points *away from the origin*.

Each map below is a parametrization[10] of the *non-oriented* surface S. Which of them are parametrizations of the *oriented* surface S? That is, which ones have the same unit normal?

(a) $G(s,t) = (s,t,-\sqrt{1-s^2-t^2})$ for (s,t) with $s^2+t^2 \leq 1$.

(b) $H(s,t) = (\sqrt{1-s^2}\cos(t), \sqrt{1-s^2}\sin(t), s)$ for $(s,t) \in [-1,0] \times [0,2\pi]$.

(c) $J(s,t) = (s+t, s-t, -\sqrt{1-2s^2-2t^2})$ for (s,t) with $s^2+t^2 \leq \frac{1}{2}$.

(d) $K(s,t) = (\cos(s)\cos(t), \sin(s)\cos(t), \sin(t))$ for $(s,t) \in [0,2\pi] \times [\pi, 3\pi/2]$.

13.3.7 Parametrize the (relative) boundary of the following surfaces $S \subseteq \mathbb{R}^3$.

(a) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 = 9, 0 \leq z \leq 9\}$

(b) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 = 9, 0 \leq z \leq x+y\}$

(c) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 9, (x-1)^2+y^2 \leq 1\}$

(d) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 9\}$

(e) $S = \partial[0,1]^3$

Proofs

13.3.8 Let $S \subseteq \mathbb{R}^3$ be the graph of $f : U \to \mathbb{R}$ where $U \subseteq \mathbb{R}^2$ is a path-connected regular region. Assume f is C^1 on the interior of its domain. Orient S with an upward unit normal n. Let $G : U \to \mathbb{R}^3$ be the graph function which parametrizes S. Prove that

$$n = \frac{(-\partial_1 f, -\partial_2 f, 1)}{\sqrt{1+(\partial_1 f)^2+(\partial_2 f)^2}}.$$

More formally, you must show

$$n(G(x,y)) = \frac{(-\partial_1 f(x,y), -\partial_2 f(x,y), 1)}{\sqrt{1+(\partial_1 f(x,y))^2+(\partial_2 f(x,y))^2}}, \qquad \text{for } (x,y) \in U^\circ.$$

13.3.9 Prove Lemma 13.3.3 using Exercise Section 13.2.3.

Applications and beyond

13.3.10 The track of a roller coaster can be thought of as a parametric surface, like a long strip of paper. On most roller coasters, a rider gives an orientation of the surface: at each point, the rider's head is "up". Roller coasters can have a lot of different elements. Let's look at two hypothetical coasters built out of two different features: vertical loops and half-inversions.

[10] Note G, H, and J are technically not smooth maps but this will not cause any issue for these calculations. This technical issue later arises with surface integrals and whether or not they are improper.

(a) vertical loop[11]

(b) half inversion [12]

(a) The Leviathan roller coaster has two vertical loops, followed by a half inversion, then another vertical loop, and then one last half inversion before returning to the station. Is this track an orientable surface?

(b) The Behemoth roller coaster has two vertical loops, followed by a half inversion, then another vertical loop, and then two half inversions before returning to the station, and then launching again for another lap around the track. Is this track an orientable surface?

13.3.11 Define the map $G : [0, 2\pi] \times [-1, 1] \to \mathbb{R}^3$ by

$$G(s,t) = \left(\left(1 + \frac{t}{2}\cos\left(\frac{s}{2}\right)\right)\cos(s), \left(1 + \frac{t}{2}\cos\left(\frac{s}{2}\right)\right)\sin(s), \frac{t}{2}\sin\left(\frac{s}{2}\right)\right)$$

The Mobius strip is defined as $M = \mathrm{im}(G)$. You may assume without proof that G is simple.

(a) Prove that M is a surface.

(b) Calculate the surface area of M to 3 decimal places using computer algebra software.

(c) Prove that the function $n : (0, 2\pi) \times (-1, 1) \to \mathbb{R}^3$ given by

$$n(u,v) = \frac{(\partial_1 G \times \partial_2 G)(u,v)}{||(\partial_1 G \times \partial_2 G)(u,v)||}$$

does *not* define a unit normal on M.

[11]Image retrieved from Wikimedia Commons on 2024-08-07, licensed under PD.
[12]Image by Davil Fulmer, retrieved from Wikimedia Commons on 2024-08-07 licensed under CC BY 2.0.

13.4. Surface integrals

With a better description for orientation of surfaces, you can explain what it means to "cross" the surface by introducing the unit normal n of a surface S in $\mathbb{R}^3$. You are therefore ready to return to the physics of forces and fluids in $\mathbb{R}^3$, and measure the amount of water flowing across a surface. This parallels the unit normal and flux for curves in $\mathbb{R}^2$ in Section 12.3.1, and raises a core question in physics that fundamentally motivates this chapter.

What is the flux of a vector field F across an oriented surface S?

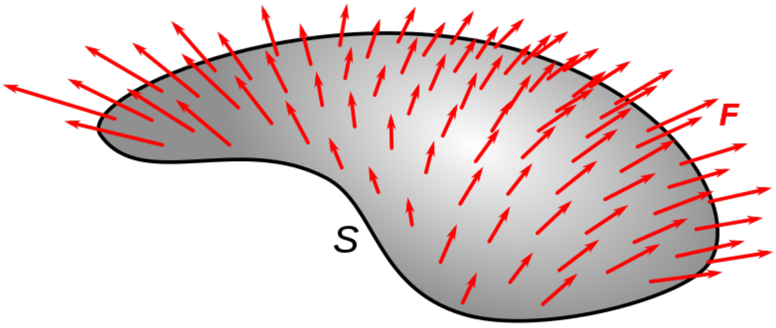
As with curves in $\mathbb{R}^2$, the basic idea will be to "add up" the flux across the surface by integrating along the surface with the surface area element dS. A good definition must be independent of parametrization. Once you have defined flux, you will be able to formulate some standard properties that parallels work done along curves. Your investigations will always be guided by physical and geometric intuition.

13.4.1 Derivation of surface integrals

Surface integrals of vector fields have a beautiful physical interpretation.

What is the amount of fluid flowing across a surface?

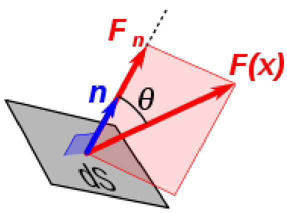
This quantity is measured by the *flux* of a vector field F across an orientable surface S.[13]



The physical principle is simple:

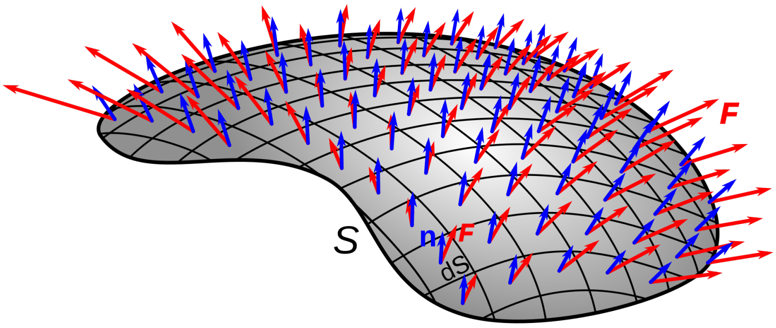
*The flux of a **constant** flow across a rectangle is the rate of the flow orthogonal to the surface multiplied by the area of the rectangle.*

This suggests an integral formula by treating the surface area element dS like an infinitesimal.[14]



Consider an infinitesimal piece of the orientable surface S with infinitesimal area dS. The direction across the surface is the unit normal n. The amount of flow F across S is therefore $F \cdot n$. Thus, the infinitesimal flux is $(F \cdot n)dS$. Integrating this over the surface S gives the total flux $\iint_S F \cdot n \, dS$.

[13]Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.
[14]Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.



If $G : U \to \mathbb{R}^3$ parametrizes S, then the unit normal and surface element are given by

$$n = \frac{\partial_1 G \times \partial_2 G}{||\partial_1 G \times \partial_2 G||}, \qquad dS = ||\partial_1 G \times \partial_2 G|| dA$$

Formal symbol manipulation suggests that you can write the vector-valued element

$$n\, dS = (\partial_1 G \times \partial_2 G)\, dA.$$

Again, this expression is purely symbolic and has no formal meaning but it is a helpful mnemonic and explains how you can calculate the suggested integral $\int\int_S F \cdot n\, dS$.[15]

13.4.2 Surface integrals of vector fields

You have stumbled upon a natural formal definition.

Definition 13.4.1 Let S be an oriented surface in $\mathbb{R}^3$ parametrized by $G : U \to \mathbb{R}^3$ with unit normal n. Let F be a vector field defined on S. The **surface integral of** F **over** S is given by

$$\int\int_S F \cdot n\, dS := \int\int_U (F \circ G) \cdot (\partial_1 G \times \partial_2 G)\, dA$$

provided it exists. Equivalently, this is the **flux of** F **across the surface** S **(in the** n **direction)**.

You can also "chop, estimate, and refine" to more rigorously derive the same formal definition. Now, you can digest this new concept in an explicit computational example.

Example 13.4.2 Let S be the upper hemisphere of radius 2 as in the previous example. Let $F(x, y, z) = (0, 0, z)$ so F is flowing directly up. You want to compute the total flux of F across S, that is,

$$\int\int_S F \cdot n dS.$$

As with the previous example, $G : U \to \mathbb{R}^3$ parametrizes S where

$$G(u, v) = (2 \sin u \cos v, 2 \sin u \sin v, 2 \cos u)$$

and $U = \{(u, v) \in \mathbb{R}^2 : 0 \le u \le \frac{\pi}{2}, 0 \le v \le 2\pi\}$. By definition of F, for $(u, v) \in U$,

$$F(G(u, v)) = (0, 0, 2 \cos u).$$

[15] Image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.

From the earlier calculations, for $0 < u < \frac{\pi}{2}, 0 < v < 2\pi$,

$$(\partial_1 G \times \partial_2 G)(u,v) = (4\sin^2 u \cos v, 4\sin^2 u \sin v, 4\sin u \cos u)$$

so

$$F(G(u,v)) \cdot (\partial_1 G \times \partial_2 G)(u,v) = 8 \sin u \cos^2 u.$$

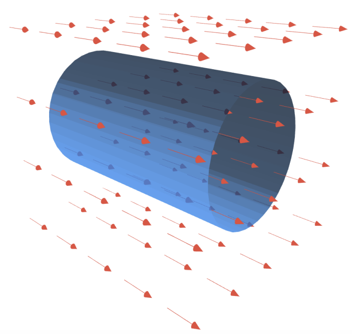
Therefore, the total flux of F across S is equal to

$$\begin{aligned}\iint_S F \cdot n dS &= \iint_U (F \circ G) \cdot (\partial_1 G \times \partial_2 G) dA \\ &= \int_0^{2\pi} \int_0^{\pi/2} F(G(u,v)) \cdot (\partial_1 G \times \partial_2 G)(u,v) du dv \\ &= \int_0^{2\pi} \int_0^{\pi/2} 8 \sin u \cos^2 u \, du dv \\ &= \frac{16\pi}{3}.\end{aligned}$$

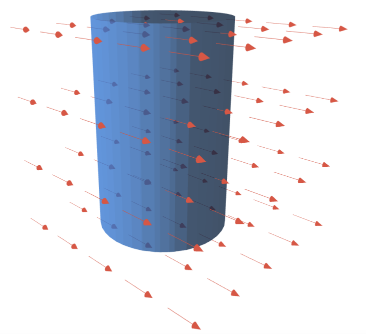
Even with many details compressed, that was a doozy. A single calculation takes a long time.

As already explained, these integrals also have natural physical meanings.

Example 13.4.3 Consider the cylinders A and B in $\mathbb{R}^3$ with the same (constant) vector field F.



A



B

Assume both cylinders are oriented so the unit normal points away from their axis of rotation. What is the flux of F across A and across B? That is, what are the values of

$$\iint_A F \cdot n \, dS \quad \text{and} \quad \iint_B F \cdot n \, dS?$$

Both surface integrals are equal to zero! However, the reasons are entirely different.

For the lefthand cylinder A, notice the vector field F is always orthogonal to the unit normal n of A, so $F \cdot n = 0$ everywhere along the cylinder A implying $\iint_A F \cdot n \, dS = 0$. For the righthand cylinder B, you can use symmetry. On the left side of B, the vector field F flows "against" the unit normal n of B, that is $F \cdot n < 0$ on that side. On the right side of B, the vector field F flows "with" the unit normal n of B, that is $F \cdot n > 0$. These positive and negative contributions appear symmetric if you cut the cylinder in halves, so you can expect overall that $\iint_B F \cdot n \, dS = 0$.

13.4.3 Basic properties

From your experience with line integrals and oriented curves, there are some routine properties that you may want to verify for surface integrals. First, the flux is invariant under parametrization and hence surface integrals are well-defined.

Theorem 13.4.4 (Invariance of flux) Let S be an oriented surface in $\mathbb{R}^3$ with unit normal n. Let F be a vector field defined on S. Let $G : U \to \mathbb{R}^3$ and $H : V \to \mathbb{R}^3$ be parametrizations of S with the same orientation. The function $(F \circ G) \cdot (\partial_1 G \times \partial_2 G)$ is integrable on U if and only if $(F \circ H) \cdot (\partial_1 H \times \partial_2 H)$ is integrable on V. If so,

$$\iint_U (F \circ G) \cdot (\partial_1 G \times \partial_2 G) dA = \iint_V (F \circ H) \cdot (\partial_1 H \times \partial_2 H) dA.$$

Proof. As with Theorem 13.2.5, this is left as a challenging exercise. The ideas and steps are nearly identical except you will need to take into the orientation at some step in your proof. ■

There are also some natural operations on surfaces that you may want to apply to surface integrals. Using Definition 13.1.28 as a benchmark, you can define an oriented surface with the opposite orientation of another.

Definition 13.4.5 Let S be an oriented surface in $\mathbb{R}^3$. Its **oppositely oriented surface** $-S$ is the reparametrization of S with the opposite orientation.

Example 13.4.6 Let S be the upper unit hemisphere in $\mathbb{R}^3$ with upward unit normal. Then $-S$ is the upper unit hemisphere in $\mathbb{R}^3$ with downward unit normal.

Lemma 13.4.7 Let $S \subseteq \mathbb{R}^3$ be an oriented surface with unit normal n. The oppositely oriented surface $-S$ has unit normal $-n$.

Proof. This is left as an exercise. Use a parametrization of S and swap the variables to parametrize $-S$. ■

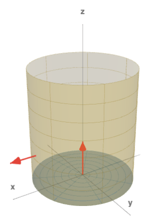
As you did with curves, you can also concatenate surfaces. A formal definition is too complicated to describe here, so an informal description will suffice.

> *Let S and T be oriented surfaces in $\mathbb{R}^3$. The* ***concatenation of** S **and** T* *is the piecewise oriented surface $S + T$ which may be formed by gluing together all or some of their relative boundaries.*

This very loose yet intuitive description is better understood through an example.

Example 13.4.8 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, 0 \le z \le 3\}$ be the cylinder with unit normal pointing away from the z-axis. Let $T = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 4, z = 0\}$ be the disk lying in the $z = 0$ plane with upward unit normal. The concatenation $S + T$ is the piecewise oriented surface that looks like a can without a lid. See this Math3D demo.

For a piecewise oriented surface $S + T$ to be defined, notice the unit normals do not need to appear "consistent"; the unit normal of S points outside of the can whereas the unit normal of T points inside the can.



Comparing with Lemma 11.3.16 and utilizing your geometric intuition, you can conjecture some standard properties about these operations on surfaces.

Lemma 13.4.9 Let S, T be oriented surfaces in $\mathbb{R}^3$. Let F and G be continuous vector fields in $\mathbb{R}^3$ defined on S and T. All of the following hold:

(a) If $-S$ is the oppositely oriented surface of S, then

$$\iint_{-S} F \cdot n \, dS = -\iint_S F \cdot n dS.$$

(b) For $\lambda \in \mathbb{R}$,

$$\int_S (F + \lambda G) \cdot n \, dS = \iint_S F \cdot n \, dS + \lambda \iint_S G \cdot n \, dS$$

(c) If $S + T$ is an oriented surface in $\mathbb{R}^3$, then

$$\iint_{S+T} F \cdot n \, dS = \iint_S F \cdot n \, dS + \iint_T F \cdot n \, dS.$$

Proof. Some of these are left as exercises. For (a), show that if n is a unit normal for S, then $-n$ is a unit normal for $-S$. For (b), use the definition of surface integrals and linearity of double integrals. Item (c) cannot be proven since concatenation is not formally defined, but you can take this property for granted. ■

This completes the key definitions for integral calculus on surfaces, answering all of this chapter's motivational questions. You have carefully defined surfaces and surface area. You have also introduced the concepts of orientation and relative boundary for surfaces. This laid the foundation for analyzing the flux of a vector field across a surface. In the last chapter of this book, you will combine differential and integral calculus with curves and surfaces and discover amazing generalizations of the fundamental theorem of calculus for $\mathbb{R}^3$ exactly as you did for $\mathbb{R}^2$. This grand finale will weave a rich tapestry of all your multivariable calculus and have remarkable physical ramifications for fluids, forces, and flows in $\mathbb{R}^3$.

Exercises for Section 13.4

Concepts and definitions

13.4.1 Let S be an oriented surface in $\mathbb{R}^3$ parametrized by $G : U \to \mathbb{R}^3$. Let F be a vector field in $\mathbb{R}^3$ that is continuous on S. Which of these quantities represents the flux of F across S?

(a) $\displaystyle\iint_S F\,dS$

(b) $\displaystyle\iint_S F \cdot dS$

(c) $\displaystyle\iint_S F \cdot n\,dS$

(d) $\displaystyle\iint_S F \cdot n\,dA$

(e) $\displaystyle\iint_U (F \circ G)(s,t) \cdot n(s,t)\|(\partial_1 G \times \partial_2 G)(s,t)\|dA$

(f) $\displaystyle\iint_U (F \circ G)(s,t) \cdot (\partial_1 G \times \partial_2 G)(s,t)dA$

13.4.2 An oriented surface S in $\mathbb{R}^3$ and vector fields F_1, F_2, F_3, and F_4 on S are shown on the right. Consider the four flux integrals.

$$I_1 = \int_S F_1 \cdot n\,dS \qquad I_2 = \int_S F_2 \cdot n\,dS$$

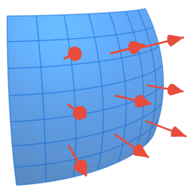
$$I_3 = \int_S F_3 \cdot n\,dS \qquad I_4 = \int_S F_4 \cdot n\,dS$$

(a) Which of the flux integrals are positive?

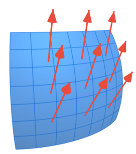
(b) Which of the flux integrals are negative?

(c) Assume all vectors have the same magnitude. Arrange the flux integrals in ascending order.

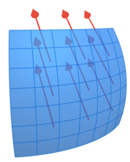
(d) How might your previous answer change if the vectors of F_2 and F_3 have much larger magnitude compared to F_1 and F_4?



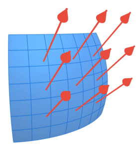
Oriented surface S (above)



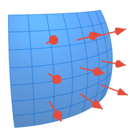
F_1



F_3



F_2



F_4

Computations

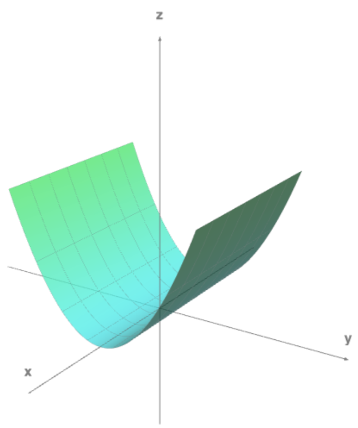
13.4.3 Computing a surface integral involves a lot of setup and attention to detail. Pierce, unfortunately, lacks attention to detail. Pierce's paraboloid

$$S = \{(x,y,z) \in \mathbb{R}^3 : z = y^2, -1 \le x \le 1, -1 \le y \le 1\}$$

is parametrized by $P : [-1,1]^2 \to \mathbb{R}^3$ where

$$P(s,t) = (s,t,t^2)$$

for $-1 \le s,t \le 1$. Orient S with upward unit normal.



He wants to calculate the flux of

$$F(x,y,z) = (x^2 - 2y, z - y, xy)$$

through S. Pierce attempts to setup the flux integral below.

1. *Note* $\partial_1 P(s,t) = (1,0,0)$ *and* $\partial_2 P(s,t) = (0,1,2t)$.
2. *By direct computation, for* $-1 < s,t < 1$,

$$\begin{aligned}(\partial_1 P \times \partial_2 P)(s,t) &= \det\begin{bmatrix} e_1 & e_2 & e_3 \\ 1 & 0 & 0 \\ 0 & 1 & 2t \end{bmatrix} \\ &= \det\begin{bmatrix} 0 & 0 \\ 1 & 2t \end{bmatrix} e_1 - \det\begin{bmatrix} 1 & 0 \\ 0 & 2t \end{bmatrix} e_2 + \det\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} e_3 \\ &= (0, -2t, 1)\end{aligned}$$

3. *Thus, the unit normal is* $n(s,t) = \dfrac{(\partial_1 P \times \partial_2 P)(s,t)}{||(\partial_1 P \times \partial_2 P)(s,t)||} = \dfrac{(0,-2t,1)}{\sqrt{1+4t^2}}$ *for* $-1 < s,t < 1$.
4. *It follows by definition that*

$$\begin{aligned}\iint_S F \cdot n \, dS &= \int_{-1}^{1}\int_{-1}^{1} F(P(s,t)) \cdot n(s,t) ds dt \\ &= \int_{-1}^{1}\int_{-1}^{1} \frac{F(0,-2t,1)\cdot(0,-2t,1)}{\sqrt{1+4t^2}} ds dt\end{aligned}$$

(a) One line has a commonly accepted nonsense expression. Identify the acceptable nonsense.

(b) One line has a computational error. Identify it and fix the error.

(c) One line requires more justification. Identify it and add the justification.

(d) Finish the calculation after making all of the above corrections.

13.4.4 Let S be the portion of the plane $8x - 6y + 2z = 10$ with upwards facing normal where $0 \le x \le 1$ and $0 \le y \le 1$ and let $F(x,y,z) = (x, y, 3z)$. Calculate

$$\iint_S F \cdot n \, dS.$$

13.4.5 Let S be five of the faces of the unit cube except for the top ($z = 1$) removed. Compute

$$\iint_S (2xyz, 3xyz, xyz) \cdot n \, dS$$

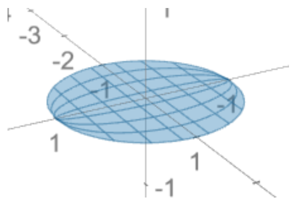
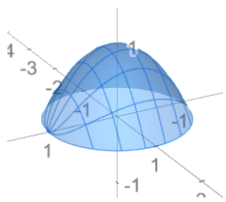
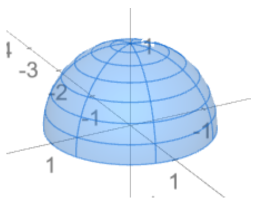
where S is oriented with normal facing away from the cube.

13.4.6 Use geometric reasoning to evaluate

$$\iint_S (x, y, 0) \cdot n \, dS$$

where S is the surface $x^2 + y^2 = 9$ and $0 \le z \le 4$, oriented with normal facing radially away from the z-axis.

13.4.7 Each surface S has the same boundary, $C = \partial S = \{(x,y,0) \in \mathbb{R}^3 : x^2+y^2=1\}$. Orient S with **upward unit normal**. Let's compare the flux of two vector fields F and G through each surface.

Surface S with boundary $\partial S = \{(x,y,0) \in \mathbb{R}^3 : x^2+y^2=1\}$	$F(x,y,z) = (x, y+z, 3z-2)$ $\iint_S F \cdot n \, dS$	$G(x,y,z) = (3y-2z, z, x-1)$ $\iint_S G \cdot n \, dS$
i) A disk  $D(s,t) = (s,t,0)$ for $s^2+t^2 \leq 1$.		
ii) A paraboloid  $P(s,t) = (s,t,1-s^2-t^2)$ for $s^2+t^2 \leq 1$ • $\partial_1 P = (1,0,-2s)$ • $\partial_2 P = (0,1,-2t)$		
iii) A hemisphere  $H(s,t) = (\cos(s)\sin(t), \sin(s)\sin(t), \cos(t))$ • $\partial_1 H = (-\sin(s)\sin(t), \cos(s)\cos(t), 0)$ • $\partial_2 H = (\cos(s)\cos(t), \sin(s)\cos(t), -\sin(t))$		

(a) Calculate each entry in the above table. We recommend working with others to split up the tasks. Setup the integral(s) and use WolframAlpha to evaluate it.

(b) Look over the table. What is the difference between the vector fields F and G?

Proofs

13.4.8 Let S be an oriented surface in $\mathbb{R}^3$. Let F and G be vector fields in $\mathbb{R}^3$ continuous on S. Fix $\lambda \in \mathbb{R}$. Prove that

$$\iint_S (F + \lambda G) \cdot n \, dS = \iint_S F \cdot n \, dS + \lambda \iint_S G \cdot n \, dS.$$

13.4.9 Let S be an oriented surface in $\mathbb{R}^3$ with unit normal n. Prove that if $-S$ is the oppositely oriented surface of S, then $-n$ is the unit normal of $-S$.

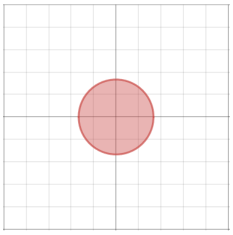
13.4.10 Let S be an oriented surface in $\mathbb{R}^3$ with unit normal n. Let F be a vector field in $\mathbb{R}^3$ continuous on S. Prove that if $-S$ is the oppositely oriented surface of S, then

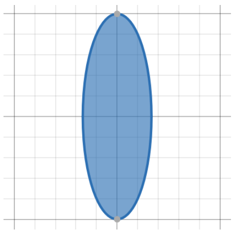
$$\iint_{-S} F \cdot n \, dS = -\iint_{S} F \cdot n \, dS.$$

13.4.11 Prove the invariance of flux integrals (Theorem 13.4.4) using Exercise 13.2.13.

Applications and beyond

13.4.12 Vegeta is watering his garden. He zones out for a second, and sprays the side of his house. The opening of the hose is a small circle and after arcing through the air it hits the wall in an ellipse, as shown below.





(a) If the vector field $F(x, y, z)$ represents the velocity of the water at a point along the stream, what does the flux integral $\iint_S F \cdot n \, dS$ represent?

(b) What can you say about the flux integrals of F through the opening of the hose and at the wall?

(c) The impact zone has a higher area than the opening of the hose. Combined with your answer to part (b), what does this tell you about F as it passes these two surfaces?

14. Fundamental theorems in 3D

In Chapters 11 and 12, you formed rigorous definitions in $\mathbb{R}^2$ for key physical concepts about 2-dimensional vector fields: work done along a *curve*, and flux across a *curve*. These led to surprisingly deep relationships with the gradient (fundamental theorem of line integrals), 2-dimensional curl (Green's theorem), and 2-dimensional divergence (Green's theorem again).

This study will generalize to $\mathbb{R}^3$ where physical applications abound. You have already constructed rigorous definitions for key physical concepts about 3-dimensional vector fields: work done along a *curve*, and flux across a *surface*. More precisely, given a vector field F in $\mathbb{R}^3$, a curve $C \subseteq \mathbb{R}^3$ parametrized by $\gamma : [a,b] \to \mathbb{R}^3$, and a surface $S \subseteq \mathbb{R}^3$ parametrized by $G : U \to \mathbb{R}^3$, the work done by F along C and the flux of F across S are respectively defined as

$$\int_C F \cdot T\, ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt, \quad \text{and} \quad \int_S F \cdot n\, dS = \iint_U (F \circ G) \cdot (\partial_1 G \times \partial_2 G) dA.$$

These capture physical notions about tangential flow and normal flow of fluids in $\mathbb{R}^3$, so you will want to discover deeper theorems about them.

Luckily, you already have one core theorem in $\mathbb{R}^3$, namely the fundamental theorem of line integrals. This only applies, however, if F is conservative. That is, if the curve C starts at $p \in \mathbb{R}^3$ and ends at $q \in \mathbb{R}^3$ and $F = \nabla f$ on an open set containing C, then

$$\int_C \nabla f \cdot d\gamma = f(q) - f(p).$$

This state of affairs creates many possible branches for generalizations from $\mathbb{R}^2$ to $\mathbb{R}^3$.

Can you generalize the fundamental theorem of line integrals to surface integrals?
Can you extend Green's theorem in curl form to $\mathbb{R}^3$? What will be 3D curl?
Can you extend Green's theorem in divergence form to $\mathbb{R}^3$? What will be 3D divergence?

This final frontier in multivariable calculus will have a climactic finish with two grand theorems: the divergence theorem and Stokes' theorem. These spectacular results have deep physical consequences and will ultimately form a mathematical masterpiece.

14.1. Flux and divergence in 3D

Vector fields in $\mathbb{R}^3$ have deep physical meanings with fluids and forces, so surface integrals can be used to describe interesting phenomenon. As you have already seen, surface integrals are physically interpreted as “flux across a surface". If you think of a flowing fluid in space, then you may have some natural physical questions.

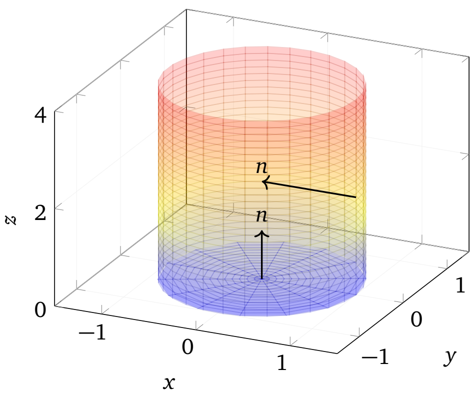
How much water flows across a "closed" surface (or through a point)?

Your goal in this section will be to translate these heuristic physical notions into formal mathematical definitions, especially an infinitesimal version of flux. You will introduce *divergence* for measuring flow rate through a point. The parallels with flux and divergence in $\mathbb{R}^2$ will be abundant so you can rely on your experience there to assist you in $\mathbb{R}^3$.

14.1.1 Flux over closed surfaces

Before introducing closed surfaces, it will be helpful to investigate a more complicated example of calculating flux, namely a piecewise surface.

Example 14.1.1 Let S be the radius 1 cylinder, extending from $z=0$ to $z=4$, centered on the z-axis. Suppose S is oriented with unit normal pointing toward the z axis. Let D be the unit circle centered on the origin in the $x-y$ plane, oriented with unit normal pointing upward.



The piecewise surface $S+D$ looks like a can without a top lid. The velocity field of a fluid is given by $F(x,y,z)=(y,-x,z+1)$ and you want to know:

How much fluid is flowing through this oriented opened can?

In other words, you want to calculate the flux through the oriented surface $S\cup D$ and hence compute the surface integral

$$\iint_{S+D} F\cdot n\, dS = \iint_S F\cdot n\, dS + \iint_D F\cdot n\, dS.$$

First, you can calculate the flux through the disk. You can compute everything as you did before, but you can also use geometry rather nicely here. For every point in D, notice that $z=0$ and its upward unit normal must therefore be $n=(0,0,1)$. Hence, everywhere on D, it must be that $F\cdot n=(y,-x,z+1)\cdot(0,0,1)=z+1=1$ for $(x,y,z)\in D$. Therefore,

$$\iint_D F\cdot n\, dS = \iint_D 1 dS = A(D)=\pi$$

since the surface area of the disk D is equal to π.

Next, you can calculate the flux through the S. Here you will execute the usual calculations. This cylinder is parametrized by the map $G : [0, 2\pi] \times [0, 4] \to \mathbb{R}^3$ where $G(\theta, z) = (\cos(\theta), \sin(\theta), z)$. You can verify that

$$\partial_1 G \times \partial_2 G = (-\sin(\theta), \cos(\theta), 0) \times (0, 0, 1) = (\cos(\theta), \sin(\theta), 0)$$

This vector is orthogonal to the surface but it points in the wrong direction! To see this, notice that when $\theta = 0$, this vector equals $(1, 0, 0)$ which points away from the z-axis instead of towards it. You can instead flip this vector and calculate the flux with

$$\partial_2 G \times \partial_1 G = -\partial_1 G \times \partial_2 G = (-\cos(\theta), -\sin(\theta), 0).$$

It follows that

$$\begin{aligned}
\iint_S F \cdot n \, dS &= \int_0^{2\pi} \int_0^4 F(G(\theta, z)) \cdot (\partial_2 G \times \partial_1 G)(\theta, z) dz d\theta \\
&= \int_0^{2\pi} \int_0^4 F(\cos(\theta), \sin(\theta), z) \cdot (-\cos(\theta), -\sin(\theta), 0) \, dz \, d\theta \\
&= \int_0^{2\pi} \int_0^4 (\sin(\theta), -\cos(\theta), z+1) \cdot (-\cos(\theta), -\sin(\theta), 0) \, dz \, d\theta \\
&= \int_0^{2\pi} \int_0^4 -\cos(\theta)\sin(\theta) + \cos(\theta)\sin(\theta) \, dz \, d\theta = 0
\end{aligned}$$

Overall, the flux through this oriented surface $S \cup D$ is

$$\iint_{S+D} F \cdot n dS = \pi + 0 = \pi.$$

This concludes your calculation, but does this value represent the amount of fluid flowing through the can? What if water is pouring into the can from the opened lid? This suggests you are not capturing the entire physical situation. You may want to ensure you have accounted for all possible entryways. Informally speaking, you want to *close* the surface $S \cup D$ with the lid and calculate the flux through the *closed* surface.

Remark 14.1.2 You must be careful when applying geometry to calculate flux, as you did for the disk D above. It was a lucky coincidence that $F \cdot n$ was constant on D so the flux integral became the surface area of D. In general, you usually still need to calculate the surface integral by parametrizing D.

The conclusion of the last example suggests the need to define a *closed* surface, so fluid cannot escape your flux measurements.

Definition 14.1.3 A piecewise surface S in $\mathbb{R}^3$ is **closed** if its relative boundary ∂S is empty.

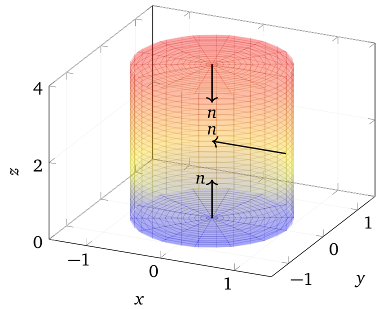
Remark 14.1.4 A surface integral over a closed surface is also denoted $\oiint_S$. You can still write $\iint_S$ but the oval around the integral signs is used to emphasize that the surface is closed.

Now, you can return to the last example.

Example 14.1.5 The surface $S \cup D$ in Example 14.1.1 is not closed since its relative boundary $\partial(S \cup D)$ is the circle $\{(x,y,z) \in \mathbb{R}^3 : x^2+y^2=1, z=4\}$ and hence non-empty. You can close this surface with the unit disk T centered on the z-axis in the $z=4$ plane., that is,

$$T = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 \leq 1, z=4\}.$$

Orient T to have a downward facing unit normal. Thus, the concatenated surface $S+D+T$ is a closed surface with unit normal oriented *inward* relative to the (solid) regular region $R = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 \leq 1, 0 \leq z \leq 4\}$.



Thus, the flux of $F(x,y,z) = (y,-x,z+1)$ across $S+D+T$ given by

$$\oiint_{S+D+T} F \cdot n \, dS$$

actually measures the net flow of the fluid into the can! To calculate it, you can first divide-and-conquer and see that

$$\oiint_{S+D+T} F \cdot n dS = \iint_{S+D} F \cdot n dS + \iint_{T} F \cdot n dS = \pi + \iint_{T} F \cdot n dS$$

from Example 14.1.1. Notice $z=4$ everywhere on T and the disk T has downard unit normal $n=(0,0,-1)$. By definition of F, it follows that $F \cdot n = -(z+1) = -5$ everywhere on T. Since n is a unit vector and T is a disk with surface area $A(T) = \pi$,

$$\int_T F \cdot n dS = -5 \int_T 1 \, dS = -5\, A(T) = -5\pi.$$

Overall, the flux of F across the closed surface $S+D+T$ is equal to

$$\oiint_{S+D+T} F \cdot n dS = -4\pi.$$

The negative value means there is a net flow of water out of the can.

With the new idea of closed surfaces, fluids cannot "escape" your flux measurements. More formally, if S is a closed piecewise surface in $\mathbb{R}^3$ enclosing a regular region $R \subseteq \mathbb{R}^3$ such that $S = \partial R$, then you can orient S with an **outward** unit normal (with respect to R). This parallels the ideas of line integrals along closed curves in $\mathbb{R}^2$ from Section 12.3 and the concept of outward flux in $\mathbb{R}^2$ in Section 12.4. This concept opens up the possibility of quantifying *infinitesimal flux*.

14.1.2 Divergence

Divergence of a vector field in $\mathbb{R}^3$ naturally generalizes from $\mathbb{R}^2$.

Definition 14.1.6 Let $F = (F_1, F_2, F_3)$ be a C^1 vector field in $\mathbb{R}^3$. The **divergence** of F is the continuous real-valued function

$$\operatorname{div}(F) = \partial_1 F_1 + \partial_2 F_2 + \partial_3 F_3.$$

Remark 14.1.7 Equivalent notation for $\operatorname{div}(F)$ includes $\nabla \cdot F$ with the gradient operator ∇, because these symbols can be suggestively written as

$$\nabla \cdot F = (\partial_1, \partial_2, \partial_3) \cdot (F_1, F_2, F_3).$$

This has no formal meaning but it is a rather pretty mnemonic.

Again, this formal definition appears unmotivated. How does this combination of derivatives have any relation to flow rate? The true origin mirrors Lemma 12.3.16 in $\mathbb{R}^2$.

Lemma 14.1.8 Let F be a vector field in $\mathbb{R}^3$. Fix $p \in \mathbb{R}^3$ in its domain. If F is C^1 on a open set containing p, then

$$(\operatorname{div} F)(p) = \lim_{\varepsilon \to 0^+} \frac{1}{\operatorname{vol}(B_\varepsilon(p))} \oiint_{\partial B_\varepsilon(p)} (F \cdot n)\, dS$$

where the sphere $\partial B_\varepsilon(p)$ is oriented with outward unit normal relative to $\overline{B_\varepsilon(p)}$.

Proof. This is left as an exercise. As with Lemma 12.1.7, there are a couple of proofs. A challenging approach is to directly parametrize the surface integrals and use a linear approximation for the vector field; this method will illuminate how divergence arises. An easier approach is to apply the divergence theorem which will be introduced in the next section; however, this method will not be as illuminating. ■

Informally speaking, this lemma says:

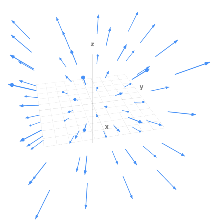
Divergence is infinitesimal flux (or flux density).

That is, divergence at a point p is a limit of *outward* flux across shrinking ε-spheres around p. The definition of divergence has computational utility, whereas this lemma provides a physically useful viewpoint of divergence.

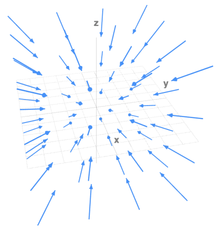
Definition 14.1.9 Let F be a C^1 vector field in $\mathbb{R}^3$. A point $p \in \mathbb{R}^3$ is a **source** of F if $(\operatorname{div} F)(p) > 0$ and a **sink** of F if $(\operatorname{div} F)(p) < 0$. A vector field F is **sourceless** if $\operatorname{div}(F) = 0$ everywhere.

Imagine the vector field represents a flowing body of water. For any point p in this body, $(\operatorname{div} F)(p) > 0$ if p is a source. In other words, there seems to be more water emerging from p than draining into it. Similarly, $(\operatorname{div} F)(p) < 0$ if p is a sink. In other words, there seems to be more water draining into p than emerging from it. Otherwise, $(\operatorname{div} F)(p) = 0$. In addition to your intuition with divergence in $\mathbb{R}^2$, some classic examples in $\mathbb{R}^3$ will help.

Example 14.1.10 The vector field $F(x,y,z) = (x,y,z)$ has everywhere positive divergence, and the vector field $G(x,y,z) = (-x,-y,-z)$ has everywhere negative divergence. You can verify this by direct computation but it also helps to compare with your physical intuition.

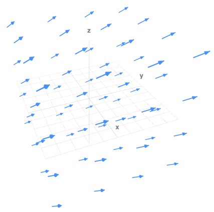


$F(x,y,z)=(x,y,z)$

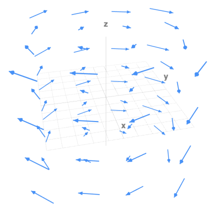


$G(x,y,z)=(-x,-y,-z)$

The flux of F across any small sphere should be positive because, heuristically speaking, "larger vectors exit the sphere rather than enter". The opposite is true for G. On the other hand, the vector fields $H(x,y,z) = (2,2,2)$ and $J(x,y,z) = (y,-x,0)$ both have zero divergence everywhere and are hence sourceless, but their behaviours are quite different.



$H(x,y,z)=(2,2,2)$



$J(x,y,z)=(y,-x,0)$

The constant vector field H is sourceless since the flux of any small sphere will be zero by symmetry. In general, it is much harder to see that a vector field is sourceless; for instance, J is sourceless but its plot does not make that easy to verify. Play with this Math3D demo to get a better visual of each vector field.

The divergence operator div on vector fields in $\mathbb{R}^3$ has some standard derivative identities.

Lemma 14.1.11 Let F and G be C^1 vector fields in $\mathbb{R}^3$ with domain $U \subseteq \mathbb{R}^3$. Fix a C^1 real-valued function f on U and fix $\lambda \in \mathbb{R}$. All of the following hold everywhere on U.

(a) $\operatorname{div}(F+\lambda G) = \operatorname{div}(F) + \lambda \operatorname{div}(G)$

(b) $\operatorname{div}(fF) = (\nabla f)\cdot F + f \operatorname{div}(F)$

(c) $\operatorname{div}(\nabla f) = \partial_1^2 f + \partial_2^2 f + \partial_3^2 f$ provided f is C^2.

Remark 14.1.12 This lemma introduces a remarkable differential operator on scalar fields f, namely the **Laplacian** $\Delta f := \partial_1^2 f + \partial_2^2 f + \partial_3^2 f$. Item (c) can therefore be equivalently written as $\nabla\cdot(\nabla f) = \Delta f$. In future studies of mathematics or physics, you will find that the Laplacian arises in many physical contexts related to energy and waves.

Remark 14.1.13 There are additional properties involving the curl of a vector field in $\mathbb{R}^3$, which has not yet been introduced. Those are postponed to a later section.

Proof. All of these identities are left as direct brute-force calculations. They are not inspiring or illuminating, but it is good computational practice with differentiation. ∎

You have carefully analyzed the physical and geometric consequences of flux and divergence for vector fields in $\mathbb{R}^3$. You can rigorously describe net flow rate, whether across a surface or through a point. This formal mathematical language is prevalent in fluid dynamics and is foundational to the theory of vector calculus in $\mathbb{R}^3$. Next, you will apply this fresh intuition to generalize the divergence form of Green's theorem in $\mathbb{R}^2$ to three dimensional space.

Exercises for Section 14.1

Concepts and definitions

14.1.1 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ and $G : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^2 vector fields with $F = (F_1, F_2, F_3)$ and $G = (G_1, G_2, G_3)$. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^2 scalar function. Which statements are true, false, or nonsense?

(a) $\operatorname{div}(F) = \partial_1 F_1 + \partial_2 F_2 + \partial_3 F_3$

(b) $\operatorname{div}(F) = \nabla \cdot F$

(c) $\operatorname{div}(\nabla f) = 0$

(d) $\operatorname{div}(\nabla f) = \partial_1^2 f + \partial_2^2 f + \partial_3^2 f$

(e) $\operatorname{div}(\nabla F) = \nabla \cdot \nabla F$

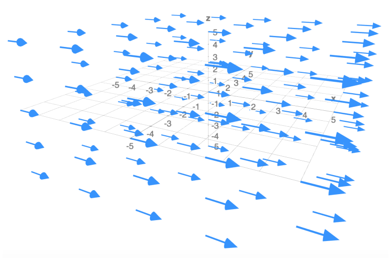
(f) $\nabla(\operatorname{div}(F)) = F$

(g) $\nabla(fg) = f\nabla g + g\nabla f$

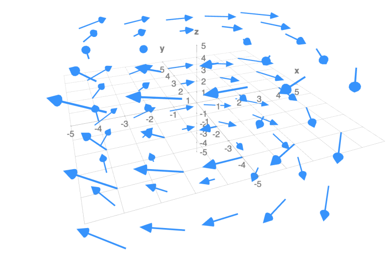
(h) $\operatorname{div}(FG) = \operatorname{div}(F)G + F\operatorname{div}(G)$

14.1.2 Three of the following vector fields are sourceless. Identify them.

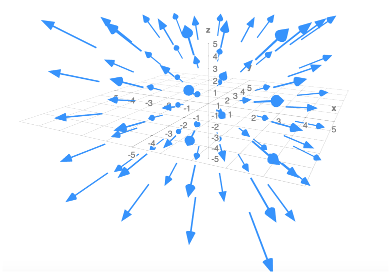
(a)



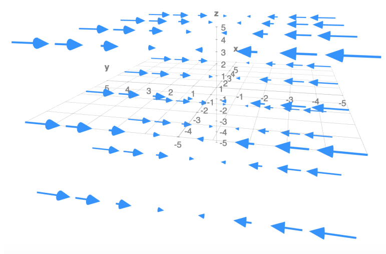
(d)



(b)



(e)

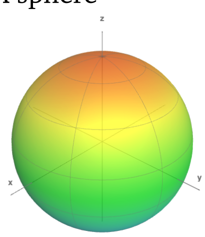


(c) $F(x, y, z) = (xy, xz, -yz)$

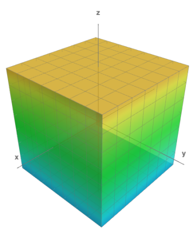
(f) $G(x, y, z) = (xz, xy, -yz)$

14.1.3 For each surface, determine if it is closed.

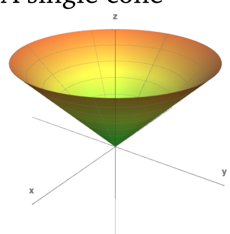
(a) A sphere



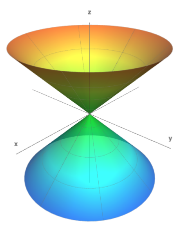
(c) A cube



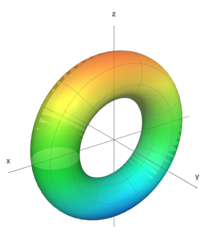
(e) A single cone



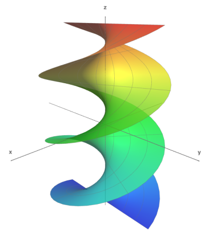
(b) A double cone



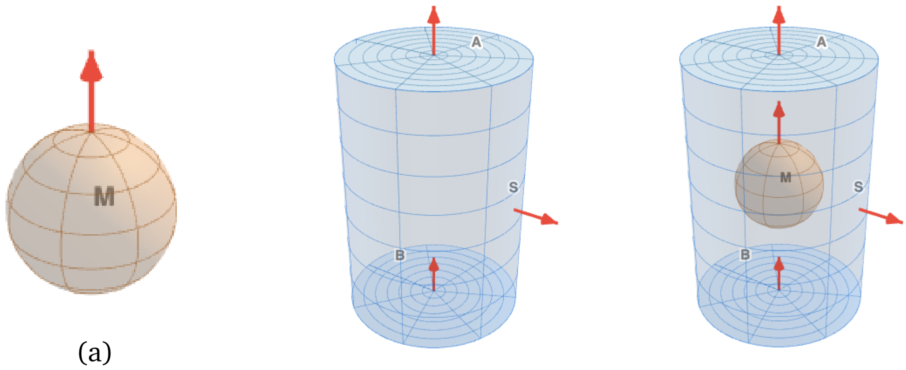
(d) A torus



(f) A helicoid



14.1.4 Remember that the concept of "outwards" for a closed surface only makes sense with respect to a region. Assume all boundaries of regions are closed piecewise surfaces.



(a) (b) (c)

Each figure above corresponds to one of the parts below.

(a) The surface M encloses a solid ball $P \subseteq \mathbb{R}^3$. The surface is oriented as illustrated below. If ∂P is oriented with outward normal (with respect to P), then express ∂P in terms of M.

(b) A solid tube $T \subseteq \mathbb{R}^3$ has a boundary ∂T with a top disk A, bottom disk B, and round side S. Each surface is oriented as illustrated below. If ∂T is oriented with outward normal (with respect to T), express ∂T in terms of A, B, and S.

(c) The solid pumpkin P is inside the solid tube T so $R = \overline{T \setminus P} \subseteq \mathbb{R}^3$ is the regular region in between them. If ∂R is oriented with outward normal (with respect to R), then express ∂R in terms of A, B, M, and S.

Computations

14.1.5 Computing flux of a vector field F through a surface S takes a few steps: parametrize, orient the unit normal, and compute. Here are partially finished examples. Use WolframAlpha to evaluate any iterated double integrals.

(a) You want to calculate the outward flux of $F(x, y, z) = (y, z, 0)$ through the boundary of the unit ball $A = \overline{B_1(0,0,0)}$ in $\mathbb{R}^3$. You have:

- Parametrized ∂A with $H(s,t) = (\cos(s)\sin(t), \sin(s)\sin(t), \cos(t))$.
- Computed $(\partial_1 H \times \partial_2 H)(s,t) = \big(-\cos(s)\sin^2(t), -\sin(s)\sin^2(t), -\cos(t)\sin(t)\big)$

Finish the calculation. Illustrate your work with a sketch.

(b) You want to calculate the outward flux of $F(x, y, z) = (x, x+y, 3z-2)$ through the boundary of the solid paraboloid $B = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}$. You have:

- Parametrized a piece of ∂B with $P(s,t) = (s, t, s^2 + t^2)$.
- Computed $(\partial_1 P \times \partial_2 P)(s,t) = (1, 0, 2s) \times (0, 1, 2t) = (-2s, -2t, 1)$.
- Parametrized a piece of ∂B with $Q(u, v) = (u, v, 1)$.

Finish the calculation. Illustrate your work with a sketch.

(c) You want to calculate the outward flux of $F(x, y, z) = (x, y, z)$ through the boundary of the solid hollowed out ball $C = \{(x, y, z) \in \mathbb{R}^3 : 137 \leq x^2 + y^2 + z^2 \leq 237\}$. You have:

- Parametrized a piece of ∂R with $G(s,t) = (237\cos(s)\sin(t), 237\sin(s)\sin(t), 237\cos(t))$.
- Computed $(\partial_1 G \times \partial_2 G)(s,t) = \big(-237^2\cos(s)\sin^2(t), -237^2\sin(s)\sin^2(t), -237^2 \cdot \cos(t)\sin(t)\big)$

Finish the calculation. Illustrate your work with a sketch.

14.1.6 Let $F(x,y,z) = (x^2, y^2, z^2)$. Is F sourceless? If not, identify the sources and sinks of F.

14.1.7 Compute the divergence of $F(x,y,z) = (3xyz^2, y^2 \sin z, xe^{2z})$. Determine if F is sourceless.

14.1.8 Give an example of a vector field in $\mathbb{R}^3$ that has only sources and no sinks.

14.1.9 Give an example of a vector field in $\mathbb{R}^3$ with only sources for $z > 0$ and only sinks for $z < 0$.

14.1.10 Let $F(x,y,z) = (x, z, x+y)$. Orient the triangle $S = \{(x,y,z) \in \mathbb{R}^3 : x+y+z = 1, x \geq 0, y \geq 0, z \geq 0\}$ with an upward unit normal. Compute the flux of F through S.

14.1.11 Even the simplest of solids have annoyingly complicated boundaries to parametrize. Let $T \subseteq \mathbb{R}^3$ be the solid tetrahedron bounded by the coordinate planes and the plane $x+y+z = 1$.

(a) Parametrize ∂T. Identify the outward unit normals on each side.

(b) Calculate the outward flux of $F(x,y,z) = (x,y,z)$ through ∂T using (a).

14.1.12 Let $F(x,y,z) = (xyz, xyz, xyz)$ and S be the surface formed by the five sides of the cube $0 \leq x,y,z \leq 1$ without side $z = 0$ and oriented with outward unit normal relative to $[0,1]^3$. Compute the flux of F through S.

Proofs

14.1.13 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field and let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^1 scalar function. Prove that

$$\operatorname{div}(fF) = (\nabla f) \cdot F + f \operatorname{div}(F).$$

14.1.14 Divergence in $\mathbb{R}^3$ is also flux density, but what does this really mean? The story parallels $\mathbb{R}^2$.

(a) Formulate this idea precisely using a C^1 vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ and a point $p \in \mathbb{R}^3$. State a conjectured lemma with shrinking *rectangles*. (Try to create it without searching the book.)

(b) Do not try to prove your conjecture but suggest a strategy for how you could prove it by using linear approximations for the scalar-valued components of $F = (f, g, h)$.

14.2. Divergence theorem

Take a moment to reflect on your knowledge of vector calculus in $\mathbb{R}^2$. The fundamental theorem of line integrals in two dimensions shows that the work done by conservative vector fields is independent of path. Green's theorem in divergence form related the normal flow across two paths or, equivalently, gave an alternate method for calculating flux across a closed curve. These fantastic results are generalizations of the fundamental theorem of calculus and provide beautiful insights into the physics of vector fields in $\mathbb{R}^2$.

This chapter's ultimate goal is to uncover similar insights for vector calculus in $\mathbb{R}^3$. Thus far, you have only collected one major result, namely the fundamental theorem of line integrals. It is the same in any dimension; the work done by a conservative vector field is independent of path. You still have questions about work done along non-conservative vector fields and questions about flux across surfaces. Here you will investigate the *normal* flow of a vector field in three dimensions but these now involve surfaces in $\mathbb{R}^3$ instead of curves.

Can you relate the normal flow across two surfaces S_1 and S_2 of a vector field?

Assuming $S = S_1 - S_2$ is a closed oriented surface, you can reformulate this question.

Is there another way to calculate flux of vector field across a closed oriented surface S?

In this section, you will discover their spectacular resolution and your second fundamental theorem for vector calculus in $\mathbb{R}^3$: the *divergence theorem*! This fantastic and widely applicable result also goes by *Gauss' theorem* because it was discovered by Karl Friedrich Gauss in 1813[1] as he studied the laws of electromagnetism. The divergence theorem forms a bridge between flux and divergence exactly as the divergence form of Green's theorem in $\mathbb{R}^2$ did. Indeed, the similarities in arguments will be significant beginning with regular regions.

14.2.1 Regular regions and orienting the boundary

Recall that you have already defined regular regions R in Section 12.2.1 but only studied them in $\mathbb{R}^2$. These regions in $\mathbb{R}^3$ also have boundaries which always touch their interiors, but the boundaries should be surfaces rather than curves. Assuming these boundaries are closed surfaces, you need to assign an orientation with respect to the region R.

Definition 14.2.1 Let $R \subseteq \mathbb{R}^3$ be a regular region whose boundary ∂R is a finite disjoint union of closed piecewise surfaces. The boundary ∂R is **positively oriented** (resp. **negatively oriented**) if the unit normal along each piecewise surface points outward (resp. inward) with respect to R.

Remark 14.2.2 Since the topological boundary ∂R is assumed to be a closed piecewise surface, its relative boundary $\partial(\partial R)$ is empty by definition. By abusing notation with the ∂ symbol and its use in higher order derivatives, this observation can be rewritten rather elegantly as

$$\partial^2 R = \partial(\partial R) = \varnothing.$$

Note, however, that the boundary of a regular region is not necessarily a closed piecewise surface, so this assumption about ∂R is always necessary. That said, you will focus on typical examples and ignore pathological situations, so this issue will not really concern you.

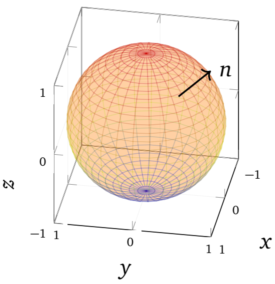
[1] Gauss did not publish his work until 1833 and 1839. A proof of the divergence theorem in $\mathbb{R}^3$ was first published by Mikhail Vasilyevich Ostrogradsky in 1831, so it is also referred to as the *Gauss-Ostrogradsky theorem*. See Stolze [19] for a more detailed history of the divergence theorem.

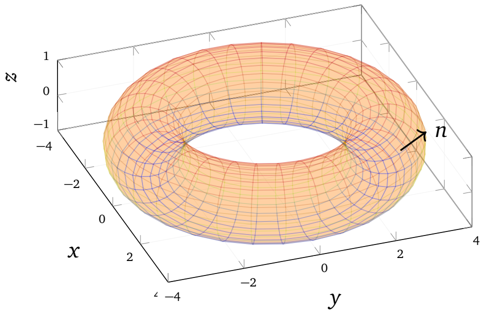
This definition is not really rigorous[2], but an intuitive understanding will be sufficient.

Example 14.2.3 The solid ball R_1 and solid torus R_2 given by

$$R_1 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}, \quad R_2 = \{(x,y,z) \in \mathbb{R}^3 : (3 - \sqrt{x^2+y^2})^2 + z^2 \leq 1\}$$

are regular regions in $\mathbb{R}^3$. The boundary ∂R_1 is a sphere and ∂R_2 is a torus, so they are both closed surfaces. These surfaces can be positively oriented with an outward unit normal as illustrated below.





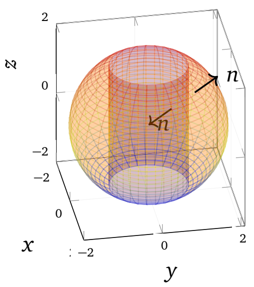
The unit normal on the sphere ∂R_1 always points outward from R_1. The same is true for the torus ∂R_2 and the corresponding solid R_2. Play with this Math3D sphere demo and this Math3D torus demo for better visuals.

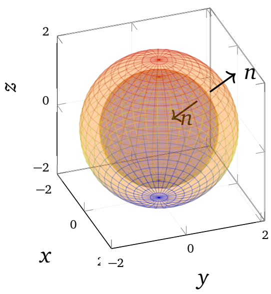
Remember that the notion of "outward" is always *with respect to a region*. This means, for instance, that the phrase "orient the unit sphere with outward unit normal" is meaningless without referring to a region. This ambiguity with language can also be incorrect, as the following examples demonstrate.

Example 14.2.4 The solid ball hollowed by a cylinder R_3 and the solid ball hollowed by another ball R_4 given by

$$R_3 = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 4, x^2 + y^2 \geq 1\}$$
$$R_4 = \{(x,y,z) \in \mathbb{R}^3 : 1 \leq x^2 + y^2 + z^2 \leq 4\}$$

are regular regions in $\mathbb{R}^3$. The boundary ∂R_3 is part of a sphere glued to a cylinder, and the boundary ∂R_4 are two disjoint spheres. These surfaces are positively oriented with outward unit normals as illustrated below.





The unit normal for ∂R_3 always points outward from R_3 whether along the sphere or along the cylinder. The unit normal for ∂R_4 always points outward from R_4 whether along the outer sphere or along the inner sphere. Notice, in particular, the inner sphere has a unit

[2]See a course in differential geometry for a more careful study of orientation.

normal that appears "inward" with respect to the ball $B_1(0,0,0)$ but it is actually outward respect to R_4. Play with this Math3D demo for R_3 and this Math3D demo for R_4.

You are now ready to explore the divergence theorem.

14.2.2 Statement and proof sketch

The divergence theorem is a spectacular generalization of the fundamental theorem of calculus to three dimensional space. It closely parallels Green's theorem in divergence form (Theorem 12.4.3) by allowing you to relate flux across the boundary of a closed surface and the volume integral of a derivative.

Theorem 14.2.5 (Divergence theorem) Let F be a vector field in $\mathbb{R}^3$ that is C^1 on a regular region $R \subseteq \mathbb{R}^3$. If its boundary ∂R is a finite disjoint union of closed piecewise surfaces and is positively oriented, then

$$\oiint_{\partial R} (F \cdot n)\, dS = \iiint_R \operatorname{div}(F)\, dV.$$

The divergence theorem can be heuristically summarized exactly as Green's theorem in $\mathbb{R}^2$.

> *The total infinitesimal flux over R is the flux across its boundary ∂R.*

This statement can be translated more informally to capture a natural physical principle.

> *The total flow inside is the net flow across the edge.*

The proof strategy is nearly identical to Green's theorem (Theorem 12.2.6) with the same two main stages but first, you prove the divergence theorem for special types of regular regions R. Second, you will reduce the general case to this special case by chopping. The complexity of higher dimensions and your current definitions of piecewise surfaces makes writing the details much messier, so only a very crude sketch is provided.

Sketch. First, assume the regular region R can be written in 6 equivalent ways[3], namely

$$\begin{aligned}
R &= \{(x,y,z) \in \mathbb{R}^3 : a \le x \le b, \alpha_1(x) \le y \le \alpha_2(x), \beta_1(x,y) \le z \le \beta_2(x,y)\}, \\
&= \{(x,y,z) \in \mathbb{R}^3 : a \le x \le b, \delta_1(x) \le z \le \delta_2(x), \gamma_1(x,z) \le y \le \gamma_2(x,z)\}, \\
&= \{(x,y,z) \in \mathbb{R}^3 : c \le y \le d, \varepsilon_1(y) \le x \le \varepsilon_2(y), \eta_1(x,y) \le z \le \eta_2(x,y)\}, \\
&= \{(x,y,z) \in \mathbb{R}^3 : c \le y \le d, \sigma_1(y) \le z \le \sigma_2(y), \tau_1(y,z) \le x \le \tau_2(y,z)\}, \\
&= \{(x,y,z) \in \mathbb{R}^3 : e \le z \le f, \xi_1(z) \le x \le \xi_2(z), \zeta_1(x,z) \le y \le \zeta_2(x,z)\}, \\
&= \{(x,y,z) \in \mathbb{R}^3 : e \le z \le f, \mu_1(z) \le y \le \mu_2(y), \nu_1(y,z) \le x \le \nu_2(y,z)\},
\end{aligned} \tag{14.2.1}$$

where all functions are continuous on their respective domains. By following the proof of Green's theorem (Theorem 12.2.6), you can verify the divergence theorem directly for this special case by dividing into components. For instance, you can prove that

$$\oiint_{\partial R} (F_1, 0, 0) \cdot n\, dS = \iiint_R \partial_1 F_1\, dV$$

[3] If you want to learn the Greek alphabet, now is your chance.

by using one of the first two forms of R stated above. This involves directly parametrizing ∂R and the single variable fundamental theorem of calculus. By symmetry, the same is true for the 2^{nd} and 3^{rd} components so this proves the divergence theorem in a special case.

Now, assume R is any regular region. Let $R' \subseteq \mathbb{R}^3$ be a rectangle containing R. Since the topological boundary ∂R is a closed piecewise surface, you can partition R' into small enough subrectangles $P = \{R'_i\}_i$ such that each subregion $R_i = R \cap R'_i$ satisfies (14.2.1)[4] You can therefore apply the divergence theorem on each regular region R_i. As with Green's theorem, the flux across all the *inside* surfaces will all cancel! This again occurs because the flux across the surface is counted twice: once in each direction. The only remaining pieces will be the surfaces counted exactly once, which necessarily lie on the topological boundary of R. This strategy is emulated in the following sequence of equalities.

$$\iiint_R \operatorname{div}(F) dV = \sum_i \iiint_{R_i} \operatorname{div}(F)\, dV = \sum_i \oint\!\!\!\!\oint_{\partial R_i} F \cdot n\, dS = \oint\!\!\!\!\oint_{\partial R} F \cdot n\, dS.$$

This crude sketch outlines the main ideas to prove the divergence theorem. ■

Its proof is a beautiful extension of the ideas leading to Green's theorem. Next, you will wield the divergence theorem to crush some calculations.

14.2.3 Examples with the divergence theorem

First, you can calculate the flux across a closed surface.

Example 14.2.6 (Six versus one) Let $R = [0,1]^3$ be the solid cube and assume its boundary ∂R is positively oriented. You want to find the flux of $F(x,y,z) = (x^2 + e^{z+x}, 137 - ye^{z+x}, \cos^{237}(xy))$ across ∂R. On one hand, you could break up the cube into six square faces and calculate six surface integrals. Yuck! On the other hand, you can throw down your new hammer. Notice that R is a regular region and F is C^1 with

$$\begin{aligned}(\operatorname{div} F)(x,y,z) &= \partial_1(x^2 + e^{z+x}) + \partial_2(137 - ye^{z+x}) + \partial_3(\cos^{237}(xy)) \\ &= 2x + e^{z+x} - e^{z+x} + 0 = 2x.\end{aligned}$$

By the divergence theorem, the outward flux of F through ∂R therefore satisfies

$$\oint\!\!\!\!\oint_{\partial R} F \cdot n dS = \iiint_R \operatorname{div}(F) dV = \iiint_{[0,1]^3} 2x dV = \int_0^1 \int_0^1 \int_0^1 2x dx dy dz = 1.$$

It was fun to smash things with a hammer, so you might as well do it again.

Example 14.2.7 (Two methods, one surface integral) You want to calculate the flux of $F(x,y,z) = (x,y,-2z)$ across the unit sphere $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ oriented with unit normal pointing away from the origin. There are two methods you can try.

One method is to calculate this surface integral directly. The unit sphere is parametrized by $G : [0,2\pi] \times [0,\pi] \to \mathbb{R}^3$ with

$$G(\theta,\phi) = (\sin(\phi)\cos(\theta), \sin(\phi)\sin(\theta), \cos(\phi))$$

[4] This step is the most challenging, but you can convince yourself of its truth with some classic solids.

You can verify that

$$\partial_1 G \times \partial_2 G = (\sin^2(\phi)\cos(\theta), \sin^2(\phi)\sin(\theta), \sin(\phi)\cos(\phi))$$

Since you want S to have a unit normal pointing away from the origin, you can verify whether this cross product of partials is a valid choice by testing a point. In particular, $G(0, \frac{\pi}{2}) = (1,0,0)$ so the normal should be equal to $n = (1,0,0)$ at this point; you can check that the cross product above is pointing in this direction (in fact, it is equal to $(1,0,0)$ but that is a coincidence). Thus, the flux across S is equal to

$$\begin{aligned}\oiint_S F \cdot n dS &= \int_0^{\pi}\int_0^{2\pi} F(G(\theta,\phi)) \cdot (\partial_\theta G \times \partial_\phi G)(\theta,\phi) d\theta d\phi \\ &= \int_0^{\pi}\int_0^{2\pi} (\sin(\phi)\cos(\theta), \sin(\phi)\sin(\theta), -2\cos(\phi)) \\ &\cdot (\sin^2(\phi)\cos(\theta), \sin^2(\phi)\sin(\theta), \sin(\phi)\cos(\phi)) d\theta d\phi \\ &= \int_0^{\pi}\int_0^{2\pi} \sin(\phi)(1-3\cos^2(\phi)) d\theta d\phi \\ &= 0.\end{aligned}$$

The last couple of steps follows from some trigonometric identities and routine single variable calculus integration techniques.

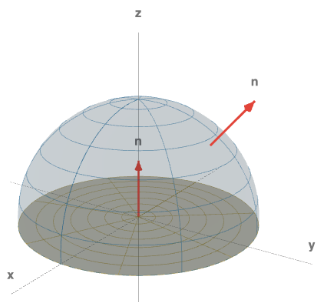
Another method is to swing your hammer again. The surface S is the topological boundary of the regular region $R = \overline{B_1(0,0,0)}$. Moreover, $S = \partial R$ is a closed surface and is positively oriented. By the divergence theorem, it follows that

$$\oiint_S F \cdot n dS = \iiint_R \operatorname{div}(F) dV = \iiint_R \big(\partial_1(x) + \partial_2(y) + \partial_3(-2z)\big) dV = \iiint_R 0 dV = 0.$$

Most importantly, the divergence theorem allows you to relate the flux across two surfaces.

Example 14.2.8 (Moving the surface) Let $S = \{(x,y,z) \in \mathbb{R}^2 : x^2+y^2+z^2 = 4, z \geq 0\}$ be the upper hemisphere of radius 2 centred at the origin. Orient S with upward unit normal. Suppose you want to calculate the flux of $F(x,y,z) = (\cos(zy)e^{y^3} + 2xz, -2yz, x^2 + 7z)$ across S. One method is to directly parametrize the surface and calculate it, but the calculations will be impossible. The surface is not closed, so you cannot directly use the divergence theorem. There is one remaining option. You can *move the surface* by closing it off.

For instance, the unit disk $D = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 \leq 4, z = 0\}$ combined with S encloses the regular region $R = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 \leq 4, z \geq 0\}$.



If you orient D with an upward unit normal, then the positively oriented boundary of R is given by $\partial R = S - D$, which is a closed piecewise surface. View this Math3D demo for better visuals. Thus, by the divergence theorem,

$$\oiint_{S-D} F \cdot n dS = \iiint_R \mathrm{div}(F) dV \implies \iint_S F \cdot n dS = \iiint_R \mathrm{div}(F) dV + \iint_D F \cdot n dS \quad (14.2.2)$$

It suffices to compute the remaining two integrals.

First, by definition of divergence,

$$(\mathrm{div}\, F)(x,y,z) = \partial_1(\cos(zy)e^{y^3} + 2xz) + \partial_2(-2yz) + \partial_3(x^2 + 7z) = 2z - 2z + 7 = 7$$

so by using the classic formula for the volume of a solid ball,

$$\iiint_R \mathrm{div}(F) dV = \iiint_R 7 dV = 7\,\mathrm{vol}(R) = 7(\frac{2\pi \cdot 2^3}{3}) = \frac{112\pi}{3}.$$

Second, since D has upward unit normal $n = (0,0,1)$ and $z = 0$ everywhere on D, it follows that $F \cdot n = x^2 + 7z = x^2$ everywhere on D. These observations imply that

$$\iint_D F \cdot n dS = \iint_D x^2\, dS.$$

Parametrizing D with polar coordinate map $G : [0,2] \times [0,2\pi] \to \mathbb{R}^2$ given by $G(r,\theta) = (r\cos\theta, r\sin\theta, 0)$, you can verify that $||\partial_1 G \times \partial_2 G(r,\theta)|| = r$ so that

$$\iint_D x^2\, dS = \int_0^2 \int_0^{2\pi} (r\cos(\theta))^2 r d\theta dr = \int_0^2 r^3 dr \int_0^{2\pi} \cos^2(\theta) d\theta = 4\pi.$$

Combining all of your calculations with (14.2.2), it follows that

$$\iint_S F \cdot n\, dS = \frac{112\pi}{3} + 4\pi = \frac{124\pi}{3}.$$

The concept of *moving a surface* via the divergence theorem is the perfect generalization to *moving a curve* with Green's theorem. It allows you to relate the flux across two surfaces by a triple integral over the divergence. This insight has many deep mathematical and physical applications, such as Gauss' law on electromagnetism.

Overall, the divergence theorem provides a relationship between flux and 3-dimensional divergence. It is a spectacular triumph towards understanding the *normal* flow of a fluid in $\mathbb{R}^3$. In the next section, you will make progress towards the final trophy of vector calculus in $\mathbb{R}^3$ by extending these ideas to the *tangential* flow of a fluid and introducing the 3-dimensional curl.

Exercises for Section 14.2

Concepts and definitions

14.2.1 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field and let $R \subseteq \mathbb{R}^3$ be a regular region whose boundary is a closed piecewise surface. Assume $S = \partial R$ is positively oriented with respect to R. Let $-S$ denote the same surface S with the **opposite orientation**. Each expression either equals the **outward** flux of F through S or the **inward** flux of F through S. Identify each of them.

(a) $\displaystyle\oiint_S (F \cdot n)\, dS$

(b) $\displaystyle -\iint_S (F \cdot n)\, dS$

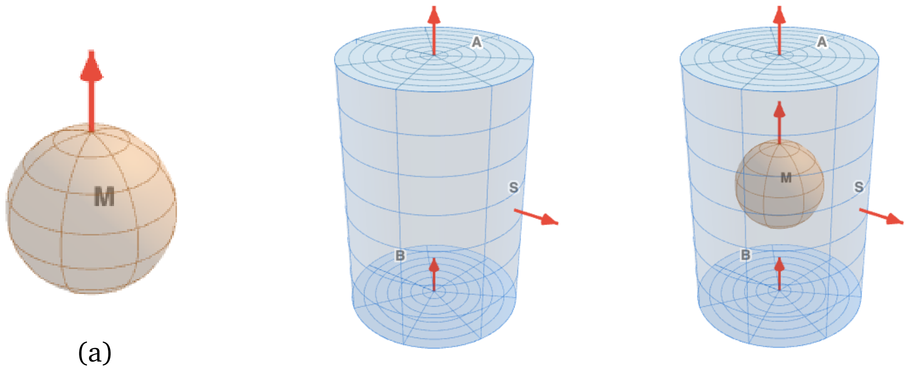
(c) $\displaystyle\oiint_{-S} (F \cdot n)\, dS$

(d) $\displaystyle\iiint_R \operatorname{div}(F)\, dV$

(e) $\displaystyle -\iiint_R \operatorname{div}(F)\, dV$

(f) $\displaystyle -\iint_{-S} (F \cdot n)\, dS$

14.2.2 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field. Assume all boundaries are closed piecewise surfaces.



(a) (b) (c)

Each figure above corresponds to one of the parts below.

(a) A solid ball $P \subseteq \mathbb{R}^3$ has boundary $M = \partial P$ oriented with outward normal. Express $\displaystyle\oiint_M (F \cdot n)\, dS$ as a triple integral.

(b) A solid tube $T \subseteq \mathbb{R}^3$ has a boundary ∂T with a top disk A, bottom disk B, and round side S. Each surface is oriented as illustrated. Relate the upward flux of F through A to the upward flux of F through B.

(c) The solid pumpkin P is inside the solid tube T so $R = \overline{T \setminus P} \subseteq \mathbb{R}^3$ is the regular region in between them. Express the upward flux through A in at least three different ways.

Computations

14.2.3 Here's two rapid calculations with the divergence theorem.

(a) Compute the outward flux of $F(x, y, z) = (y, z, 0)$ through $\partial B_1(0, 0, 0)$.

(b) Calculate the outward flux of $F(x, y, z) = (x, y, z)$ through $\partial [0, 1]^3$.

14.2.4 Armed with the divergence theorem, you can revisit some nasty flux computations.

(a) Use the divergence theorem to compute the outward flux of $F(x, y, z) = (x^3, x + y^3, 3z - 2)$ through the boundary of the solid paraboloid $R = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}$.

(b) Use the divergence theorem to compute the outward flux of $F(x, y, z) = (x, y, z)$ through the boundary of the solid hollowed out ball $\{(x, y, z) \in \mathbb{R}^3 : 137^2 \leq x^2 + y^2 + z^2 \leq 237^2\}$.

14.2.5 Computing the outward flux of $F(x,y,z)=(xyz,xyz,xyz)$ over five sides of the cube $0 \le x,y,z \le 1$ excluding the $z=0$ side is a long computation performed in Exercise 14.1.12. Now compute the same outward flux by "moving the surface integral" to the $z=0$ side of the cube.

14.2.6 Let $F(x,y,z)=(x,y,z)$. Let $R=\{(x,y,z)\in\mathbb{R}^3 : x^2+y^2 \le z \le 9\}$ be the solid paraboloid with a disk glued on top of it. Let $S=\partial R$ with **inward** normal. Compute the flux of F through S.

14.2.7 Let $F(x,y,z)=(2x,-y,xy)$ and S be the oriented boundary of a triangular prism formed by the coordinate axes, the plane $y+z=1$, and the plane $x=1$. Compute the outward flux of F through C. *Hint*: Almost no computation necessary!

14.2.8 Let $F(x,y)=(x+\sin(y)\cos(z),\cos(x)e^x+y,3z+e^x\sin(y))$ and S the same as above. Compute the outward flux of F through C. *Hint*: Almost no computation necessary!

Proofs

14.2.9 Sourceless vector fields and flux are related via the divergence theorem. Prove this lemma.

> **Lemma A.** *Let F be a vector field in $\mathbb{R}^3$ that is C^1 on an open set $U\subseteq\mathbb{R}^3$. Let R be a regular region inside U whose boundary is a closed piecewise surface. If F is sourceless on U, then F has zero outward flux through ∂R.*

14.2.10 Recall that for a C^1 vector field $G=(G_1,G_2,G_3)$ in $\mathbb{R}^3$, the **curl** of F is the vector field

$$\operatorname{curl}(G)=\left(\frac{\partial G_3}{\partial y}-\frac{\partial G_2}{\partial z},\frac{\partial G_1}{\partial z}-\frac{\partial G_3}{\partial x},\frac{\partial G_2}{\partial x}-\frac{\partial G_1}{\partial y}\right)$$

(a) Show by direct computation that if G is C^2, then $\operatorname{curl}(G)$ is a sourceless vector field.

(b) Conclude the following lemma from Lemma A.

> **Lemma B.** *Let G be a vector field in $\mathbb{R}^3$ that is C^2 on an open set $U\subseteq\mathbb{R}^3$. Let R be a regular region inside U whose boundary is a closed piecewise surface. Then* $\operatorname{curl}(G)$ *has zero outward flux through ∂R.*

14.2.11 Let F be a C^1 vector field in $\mathbb{R}^3$ on an open ball containing $p\in\mathbb{R}^3$. Orient the boundary $\partial B_\varepsilon(p)$ with outward unit normal for every $\varepsilon>0$. Prove that divergence is outward flux density:

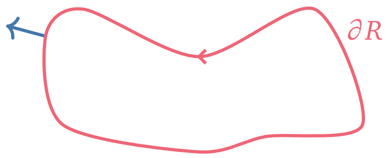
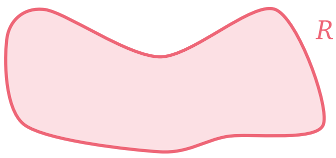
$$(\operatorname{div} F)(p)=\lim_{\varepsilon\to 0^+}\left[\frac{1}{\operatorname{vol}(B_\varepsilon(p))}\oiint_{\partial B_\varepsilon(p)}(F\cdot n)\,dS\right].$$

14.2.12 Let $F=(f,g,h)$ be a vector field in $\mathbb{R}^3$ which is C^1 on $R=[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]$. Let ∂R be oriented with an outward unit normal. Prove the divergence theorem for R:

$$\oiint_{\partial R}(F\cdot n)\,dS=\iiint_R(\operatorname{div}F)\,dV.$$

Applications and beyond

14.2.13 The divergence theorem is the $\mathbb{R}^3$ analogue of Green's theorem in flux form for $\mathbb{R}^2$. You should lean on this analogy to help build connections between theory. Fill in the table below.

Green's theorem (flux form)	Divergence theorem
$\displaystyle\oint_{\partial R} (F \cdot n)\, ds$	
$\displaystyle\iint_R \operatorname{div}(F) dA$	
	
	

14.2.14 Let $R \subseteq \mathbb{R}^3$ be a regular region with positively oriented boundary ∂R that is a finite disjoint union of closed piecewise surfaces. Express the volume of R as $\displaystyle\oiint_{\partial R} (F \cdot n)\, dS$ for at least 3 different choices of vector fields $F : \mathbb{R}^3 \to \mathbb{R}^3$.

14.3. Circulation and curl in 3D

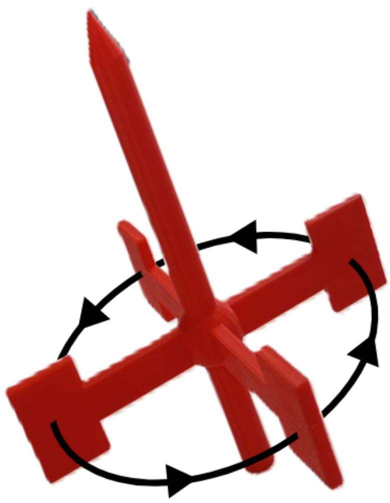
With a firm grasp on flux, you can continue to explore the physics of vector fields in $\mathbb{R}^3$, and return to the concept of work done and circulation. As you have already seen in Section 12.1 for $\mathbb{R}^2$, the line integral along a closed curve (circulation) and its infinitesimal counterpart (curl) describe the swirly-ness of a two-dimensional vector field. Inspired by this success, you can ask the same questions in three dimensional space.

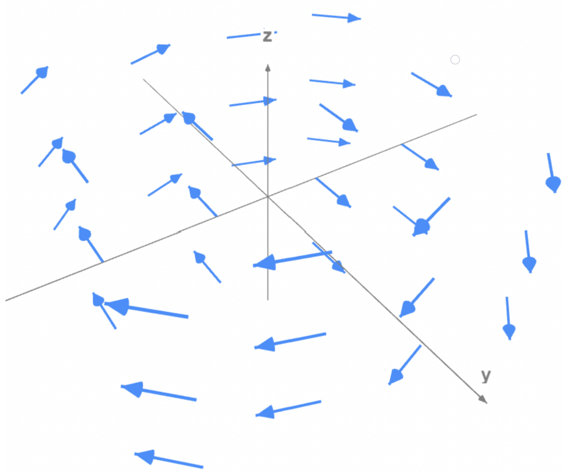
How much does the water swirl along a loop (or at a point) and in what direction?

Your goal in this section will be to translate these heuristic physical notions into formal mathematical definitions, especially an infinitesimal version of circulation. The parallels with circulation and curl in $\mathbb{R}^2$ will be frequent, so you can rely on your experience there to assist you in $\mathbb{R}^3$. The idea of circulation will generalize effortlessly in higher dimensions due to your robust theory for calculus with curves from Chapter 11. However, curl will be dramatically more delicate in $\mathbb{R}^3$.

Curl in $\mathbb{R}^2$ is measured by a single real number because, crudely speaking, fluid in $\mathbb{R}^2$ will spin either clockwise or counterclockwise at a point; there are only two choices. In other words, if you place a paddlewheel in a two-dimensional fluid, then the speed and direction it spins depend on only one factor: the position of the paddlewheel.

Circulation and curl of vector fields in $\mathbb{R}^3$ share much of the same story as in $\mathbb{R}^2$ but the geometry and linear algebra becomes a bit more sophisticated. In particular, any sensible definition of curl in $\mathbb{R}^3$ cannot possibly be measured by a single real number because there are *infinitely many* axes of rotation! This complication can be imagined with the physical example of a paddlewheel in fluid swirling in three dimensional space, as illustrated below.





If you place the paddlewheel in the fluid, the speed and direction it spins depends on *two* factors: the position of the paddlewheel *and the axis of rotation*! This geometric conundrum implies that a single real number cannot capture the physical complexity of swirling in three dimensions; your definition of curl in $\mathbb{R}^3$ will need to be sensitive to the axis of rotation. As you shall see, the curl of a vector field will itself be defined as a vector field. This choice will be geometrically natural and appropriately describe the physics of swirling in three dimensions.

14.3.1 Circulation

The definition of circulation is described by the same line integral for any dimension; it is repeated here for the sake of clarity.

Definition 14.3.1 (Circulation) Let F be a vector field in $\mathbb{R}^3$ defined on an oriented piecewise curve C in $\mathbb{R}^3$. Assume C is closed. The **circulation of F along** C is the line integral

$$\oint_C F \cdot T ds.$$

As before, circulation represents the tangential flow along a closed curve (or equivalently, work done). You can calculate it in the same way.

Example 14.3.2 Let $C = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z = 0\}$ so C is the unit circle in the xy-plane. Note that C is parametrized by $\gamma(t) = (\cos(t), \sin(t), 0)$ for $t \in [0, 2\pi]$. The circulation of $F(x, y, z) = (y, z, x)$ about C is therefore

$$\begin{aligned}\oint_C F \cdot T ds = \int_0^{2\pi} F(\gamma(t)) \cdot \gamma'(t) dt &= \int_0^{2\pi} F(\cos(t), \sin(t), 0) \cdot (-\sin(t), \cos(t), 0) dt \\ &= \int_0^{2\pi} (\sin(t), 0, \cos(t)) \cdot (-\sin(t), \cos(t), 0) dt \\ &= \int_0^{2\pi} -\sin^2(t) dt \\ &= \int_0^{2\pi} \frac{1}{2}\cos(2t) - \frac{1}{2} dt = -\pi.\end{aligned}$$

In other words, the work done along C is equal to $-\pi$.

As this example demonstrates, circulation is essentially unchanged in higher dimensions.

14.3.2 Definition of curl

Curl of a vector field in $\mathbb{R}^3$ generalizes in an unexpected way. It is itself a vector field!

Definition 14.3.3 Let F be a C^1 vector field in $\mathbb{R}^3$. The **curl** of F is the continuous $\mathbb{R}^3$-valued function given by

$$\operatorname{curl}(F) = \Big(\partial_2 F_3 - \partial_3 F_2, \partial_3 F_1 - \partial_1 F_3, \partial_1 F_2 - \partial_2 F_1\Big)$$

Remark 14.3.4 Equivalent notation for $\operatorname{curl}(F)$ includes $\nabla \times F$ where ∇ is the gradient operator and $\times$ is the cross product operator, because these symbols can be suggestively written as

$$\nabla \times F = (\partial_1, \partial_2, \partial_3) \times (F_1, F_2, F_3) = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ \partial_1 & \partial_2 & \partial_3 \\ F_1 & F_2 & F_3 \end{bmatrix}.$$

This has no formal meaning, but it acts as a helpful computational memory trick.

This algebraic formula is entirely unmotivated and rather mysterious. You will return to this mystery shortly, so pause this confusion for a moment. For now, you can at least appreciate that the curl is straightforward to compute.

Example 14.3.5 Let $F(x,y,z)=(x,2z^2,\cos y)$. The curl of F is therefore given by

$$\begin{aligned}\nabla\times F &= \det\begin{bmatrix} e_1 & e_2 & e_3\\ \partial_1 & \partial_2 & \partial_3 \\ x & 2z^2 & \cos y\end{bmatrix}\\ &= \big(\partial_2(\cos y)-\partial_3(2z^2), \partial_3(x)-\partial_1(\cos y), \partial_1(2z^2)-\partial_2(x)\big)\\ &= \big(-\sin y - 4z, 0, 0\big).\end{aligned}$$

The formal definition of curl has an immediate connection to irrotational vector fields as defined by Definition 11.4.10.

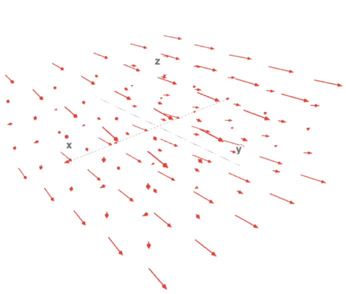
Lemma 14.3.6 A C^1 vector field F in $\mathbb{R}^3$ is irrotational if and only if $\operatorname{curl}(F)=0$ everywhere on its domain.

Proof. This is left as a straightforward exercise. You only need to compare the definitions. ■

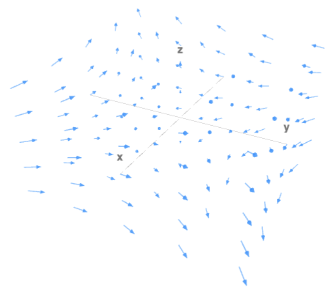
Remark 14.3.7 Irrotational vector fields in $\mathbb{R}^3$ are thus also called **curl-free** vector fields.

Notice Lemmas 11.4.11 and 14.3.6 imply that every conservative vector field F in $\mathbb{R}^3$ satisfies $\operatorname{curl}(F)=0$. Detecting whether a vector field is curl-free is a straightforward algebraic computation, but it is not usually easy to see visually.

Example 14.3.8 The vector field $F(x,y,z)=(x,2z^2,\cos y)$ in the previous example is not curl-free since $\operatorname{curl}(F)$ is not always zero. You can verify by direct computation that the vector field $G(x,y,z)=(y,x,z)$ is curl-free. View their Math3D plots below.



$F(x,y,z)$



$G(x,y,z)$

It is surprisingly difficult to tell which one is irrotational.

To truly understand curl, you will need a splash of linear algebra and geometry.

14.3.3 Geometry of curl

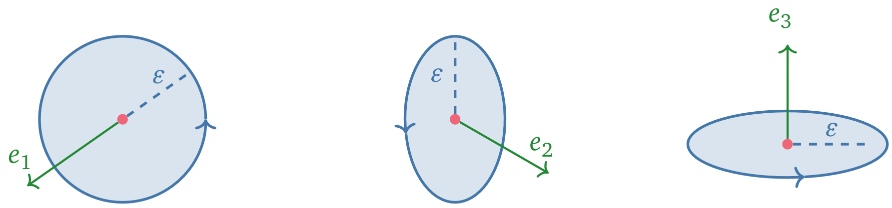
The geometric origin of three-dimensional curl can be inferred from the third component; it matches the expression for the two-dimensional curl in Definition 12.1.4, namely

$$(\operatorname{curl} F)\cdot e_3 = \partial_1 F_2 - \partial_2 F_1.$$

This suggests if you view the vector field from above, then $(\operatorname{curl} F)\cdot e_3$ will look like the usual 2-dimensional curl. This gives the crucial insight:

> *The quantity $(\operatorname{curl} F)\cdot e_3$ should be the infinitesimal circulation of shrinking positively oriented ε-circles* ***lying in the two-dimensional plane orthogonal to*** e_3.

Repeating this for each standard basis direction $\{e_1, e_2, e_3\}$, you can visualize this idea nicely.



$(\operatorname{curl} F)\cdot e_1 = \partial_2 F_3 - \partial_3 F_2$, $\quad(\operatorname{curl} F)\cdot e_2 = \partial_3 F_1 - \partial_1 F_3$, $\quad(\operatorname{curl} F)\cdot e_3 = \partial_1 F_2 - \partial_2 F_1$

Similar to Lemma 12.1.7 in $\mathbb{R}^2$, you can express each of these 3 relationships formally. Namely, for each $i = 1, 2, 3$, you can formally state that the value $(\operatorname{curl} F)(p)\cdot e_i$ is the infinitesimal circulation at $p \in \mathbb{R}^3$ of shrinking positively oriented ε-circles lying in the two-dimensional plane orthogonal to e_i; see Exercise 14.3.12 for details.

There is nothing special about the standard basis vectors. You can generalize this geometric perspective to any axis of rotation, i.e. any unit vector.

> *Given a unit vector n and a point $p \in \mathbb{R}^3$, the quantity*
> $$(\operatorname{curl} F)(p)\cdot n$$
> *measures the infinitesimal circulation around the axis defined by n according to the righthand rule.*

This notion and the righthand rule are illustrated below.[5]

Importing your intuition from $\mathbb{R}^2$, this viewpoint therefore yields two physical meanings.

> *The magnitude $|(\operatorname{curl} F)(p)\cdot n|$ is the speed at which the fluid swirls at p around n.*
> *The sign of $(\operatorname{curl} F)(p)\cdot n$ is positive if the fluid swirls counterclockwise at p around n and negative if clockwise.*

These meanings can be elegantly applied in examples.

Example 14.3.9 You want to measure how much the vector field $F(x, y, z) = (z, -x, y^2)$ swirls at the origin $(0, 0, 0)$ around the axis of rotation $n = \frac{1}{\sqrt{3}}(1, 1, -1)$. In other words, you want to calculate and interpret $(\operatorname{curl} F)(0, 0, 0)\cdot n$. View this Math3D demo for a visual. Note

$$(\operatorname{curl} F)(x, y, z) = (\partial_2(y^2) - \partial_3(-x), \partial_3(z) - \partial_1(y^2), \partial_1(-x) - \partial_2(z)) = (2y, 1, -1)$$

so $(\operatorname{curl} F)(0, 0, 0) = (0, 1, -1)$. This implies that

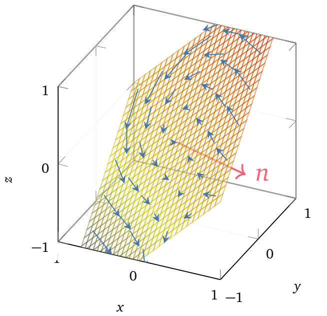
$$(\operatorname{curl} F)(0, 0, 0)\cdot n = \frac{2}{\sqrt{3}}.$$

Thus, F is swirling counterclockwise at $(0, 0, 0)$ about the n-axis at speed $\frac{2}{\sqrt{3}}$.

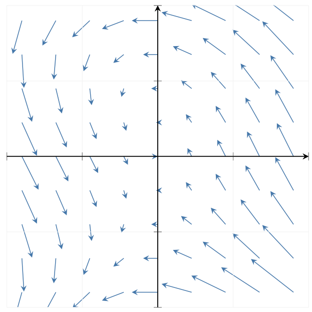
How can you visualize $(\operatorname{curl} F)(0, 0, 0)\cdot n$ with the vector field F?

[5] Right-side image retrieved from Wikimedia Commons on 2024-07-21 licensed under PD.

You can project the vector field F onto the plane $P \subseteq \mathbb{R}^3$ normal to n. The lefthand plot below is this projection and the plane P plotted in $\mathbb{R}^3$. The righthand plot is the plane P itself viewed as a two-dimensional object; this picture implies that the unit normal n is coming out of the page at the origin $(0,0)$.



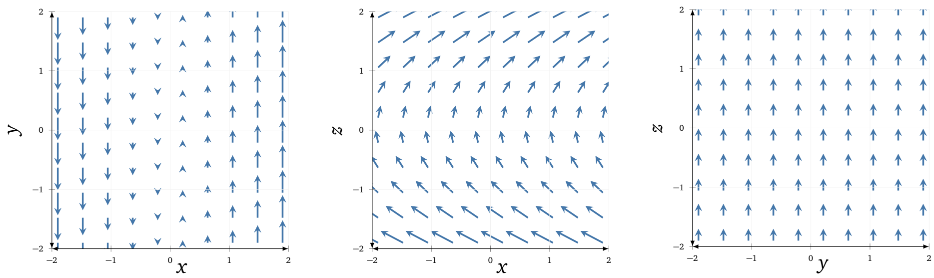
$\text{proj}_P(F)$ in $\mathbb{R}^3$



Plane P

The righthand plot shows that $\text{proj}_P(F)$ is swirling counterclockwise about the origin $(0,0)$ inside the two-dimensional plane, so F is swirling counterclockwise about the axis of rotation n in $\mathbb{R}^3$. This matches our earlier calculation.

Example 14.3.10 Let $F(x,y,z) = (z,x,1)$. You can verify that $(\operatorname{curl} F)(0,0,0) = (0,1,1)$ by direct computation. How can you visualize each component of this vector? Below are plots of F projected into each coordinate plane, namely $z=0$, $y=0$, and $x=0$ from left to right.



These plots tell you a lot about $(\operatorname{curl} F)(0,0,0)$ by thinking of a paddlewheel in each plane.

First, consider the xy-plane. You place a small paddle at the origin with the handle pointing in the direction of the unit vector $e_3 = (0,0,1)$. Based on the horizontal and vertical axes, the vector $e_3 = e_1 \times e_2$ points *out of the page*. Since the vector field flows counterclockwise about this unit vector, the quantity $(\operatorname{curl} F)(0,0,0) \cdot e_3$ should be positive, which matches the computation above.

Second, consider the xz-plane. The two-dimensional plot suggests the vector field is swirling clockwise, yet you already know that $(\operatorname{curl} F)(0,0,0) \cdot e_2 = 1 > 0$. Why does this mismatch happen? Notice that the horizontal axes correspond to $e_1 \times e_3 = -e_2$ so $e_2 = (0,1,0)$ actually points *into the page*. In other words, if you place the paddlewheel at the origin with the handle point in the e_2, then it will indeed spin counterclockwise.

You can use the same reasoning for the yz-plane to deduce that $(\operatorname{curl} F)(0,0,0) \cdot e_1 = 0$. That is, the paddlewheel will not spin at all around e_1.

This geometric exploration of curl gives a satisfying explanation for its formal definition.

14.3.4 Properties of curl

Since curl captures the spin of a fluid, you can ask a natural question.

About which axis will a fluid spin counterclockwise the fastest?

Your investigations give yet another meaning to the curl of a vector field.

Lemma 14.3.11 Let F be a C^1 vector field in $\mathbb{R}^3$. Both of the following hold:

- The maximum of $\operatorname{curl} F(p) \cdot n$ over all unit vectors $n \in \mathbb{R}^3$ occurs when $n = +\frac{\operatorname{curl} F(p)}{||\operatorname{curl} F(p)||}$ and the maximum value is $||\operatorname{curl} F(p)||$.
- The minimum of $\operatorname{curl} F(p) \cdot n$ over all unit vectors $n \in \mathbb{R}^3$ occurs when $n = -\frac{\operatorname{curl} F(p)}{||\operatorname{curl} F(p)||}$ and the minimum value is $-||\operatorname{curl} F(p)||$.

Proof. This is a consequence of the dot product. Notice how this property of the curl matches the property of the gradient as the direction of steepest ascent; indeed, see Lemma 3.4.9. ■

This lemma formalizes two key notions.

The vector $\operatorname{curl} F(p)$ *points in the direction of fastest counterclockwise spin of* F *at* p*, and its norm* $||\operatorname{curl} F(p)||$ *is the speed of the spin of* F *in this direction.*

This can be swiftly applied in computations.

Example 14.3.12 Consider the vector field $F(x, y, z) = (z, -x, y^2)$ from Example 14.3.9. By Lemma 14.3.11, the vector field F is spinning fastest counterclockwise about the axis

$$n = \frac{(\operatorname{curl} F)(0,0,0)}{||(\operatorname{curl} F)(0,0,0)||} = \frac{(0,1,-1)}{\sqrt{2}} = \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

and with speed $||(\operatorname{curl} F)(0,0,0)|| = \sqrt{2}$.

Similar to the gradient and divergence, you also have standard properties of curl.

Lemma 14.3.13 Let F and G be C^1 vector fields in $\mathbb{R}^3$ with domain $U \subseteq \mathbb{R}^3$. Fix a C^1 real-valued function f on U and fix $\lambda \in \mathbb{R}$. All of the following hold everywhere on U.

(a) $\operatorname{curl}(F + \lambda G) = \operatorname{curl}(F) + \lambda \operatorname{curl}(G)$
(b) $\operatorname{curl}(f G) = f \operatorname{curl}(G) + (\nabla f) \times G$
(c) $\operatorname{curl}(F \times G) = (G \cdot \nabla)F + (\operatorname{div} G)F + (F \cdot \nabla)G + (\operatorname{div} F)G$

Here $(G \cdot \nabla)F = \sum_{j=1}^{3} G_j \partial_j F$ and $(F \cdot \nabla)G = \sum_{j=1}^{3} F_j \partial_j G$.

Proof. These are left as boring exercises. It is nothing more than brute force algebraic manipulation. There are actually many more properties not listed here. ■

Finally, there are two key relationships between curl and the other differential operators.

Lemma 14.3.14 If F is a C^2 vector field in $\mathbb{R}^3$ and f is a C^2 real-valued function, then

$$\operatorname{curl}(\nabla f) = (0,0,0), \qquad \operatorname{div}(\operatorname{curl}(F)) = 0.$$

Proof. These are both left as exercises. You will need Clairaut's theorem for both. ■

This concludes your investigation of the physical and geometric consequences of circulation and curl for vector fields in $\mathbb{R}^3$. You can rigorously describe the swirl of a vector field, whether along a closed curve or at a point around any axis. This formal mathematical language is vital to vector calculus and fluid dynamics in $\mathbb{R}^3$. In the next section, you will apply this newfound knowledge to discover the last fundamental theorem in vector calculus by generalizing the curl form of Green's theorem in $\mathbb{R}^2$ to three dimensional space.

Exercises for Section 14.3

Concepts and definitions

14.3.1 Let $F:\mathbb{R}^3\to\mathbb{R}^3$ and $G:\mathbb{R}^3\to\mathbb{R}^3$ be C^2 vector fields with $F=(F_1,F_2,F_3)$ and $G=(G_1,G_2,G_3)$. Let $f:\mathbb{R}^3\to\mathbb{R}$ and $g:\mathbb{R}^3\to\mathbb{R}$ be C^2 scalar fields. Which statements are true, false, or nonsense?

(a) $\operatorname{curl}(F)=\left(\partial_2F_3-\partial_3F_2,\partial_3F_1-\partial_1F_3,\partial_1F_2-\partial_2F_1\right)$

(b) $\operatorname{curl}(F)=\nabla\times F$

(c) $\operatorname{curl}(\nabla f)=0$

(d) $\nabla(\operatorname{curl}(F))=F$

(e) $\operatorname{curl}(\operatorname{div}(F))=0$

(f) $\operatorname{div}(\operatorname{curl}(F))=0$

(g) $\operatorname{curl}(F\times G)=\operatorname{curl}(F)G+F\operatorname{curl}(G)$

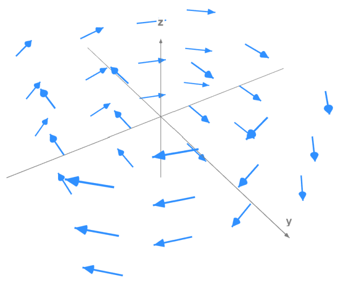
(h) $\operatorname{curl}(F\times F)=0$

14.3.2 Curl in $\mathbb{R}^3$ is a vector field. Why is it "circulation density"? Use your four viewpoints to demystify.

(a) Start with a **physical** viewpoint. Place a paddle wheel in a swirling fluid $F:\mathbb{R}^3\to\mathbb{R}^3$ at a point.

i. If you place the paddlewheel at the origin and hold the handle along the positive z-axis, which way will it spin? Be clear about your perspective of the paddlewheel.

ii. If you place the paddle wheel at the origin and hold the handle along the negative z-axis, which way will it spin? Be clear about your perspective.



(b) Next, consider a **geometric** perspective. Given a unit vector n and a point $p\in\mathbb{R}^3$,

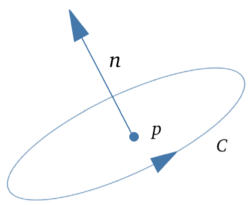
$$(\operatorname{curl}F)(p)\cdot n$$

measures the infinitesimal circulation around the axis defined by n via the righthand rule.

The sign of $(\operatorname{curl}F)(p)\cdot n$ is positive if the fluid swirls counterclockwise and negative if clockwise.

The magnitude $|(\operatorname{curl}F)(p)\cdot n|$ is the speed at which the fluid swirls around n.

Write the expressions for what you were measuring in part (a).



(c) Next, look at the **algebraic** definition of curl in Definition 14.3.3. Fix $p\in\mathbb{R}^3$ and three vectors $v_1,v_2,v_3\in\mathbb{R}^3$. If you know the scalars

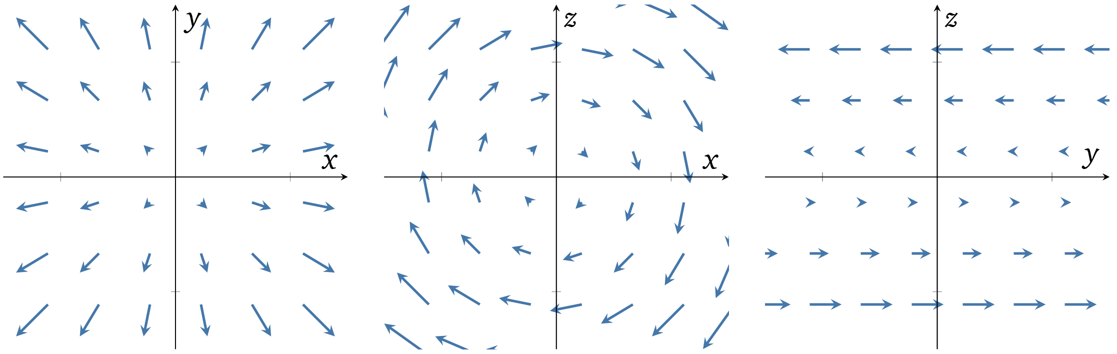
$$(\operatorname{curl}F)(p)\cdot v_1,\qquad(\operatorname{curl}F)(p)\cdot v_2,\qquad(\operatorname{curl}F)(p)\cdot v_3,$$

then can you determine the vector $(\operatorname{curl}F)(p)$?

(d) Continuing with the algebraic viewpoint, fix $p\in\mathbb{R}^3$. For which unit vector n is $(\operatorname{curl}F)(p)\cdot n$ maximum and what is its magnitude? Justify your assertion.

(See Exercise 14.3.12 for the analytic perspective.)

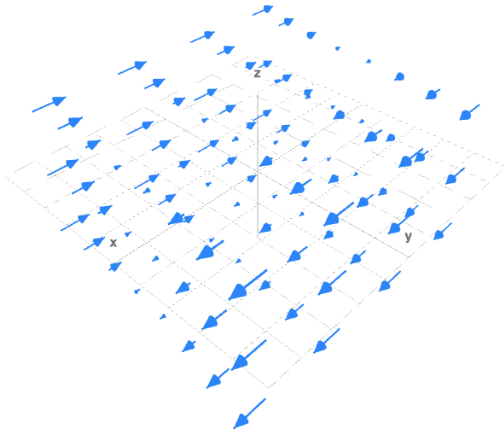
14.3.3 You have sketched the C^1 vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ projected onto each of the coordinate planes.



Write $(\operatorname{curl} F)(0,0,0) = (a,b,c)$ for some constants $a,b,c \in \mathbb{R}$. Determine whether a,b,c are positive, negative, or zero.

14.3.4

Consider the plotted vector field $H : \mathbb{R}^3 \to \mathbb{R}^3$.

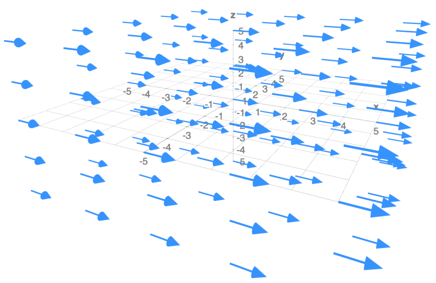


(a) Does the sign of $(\operatorname{curl} H) \cdot e_3$ appear always positive, always negative, always zero, or none of these?

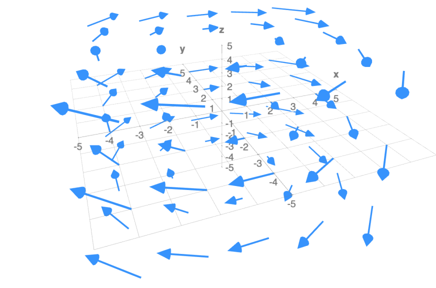
(b) Does the sign of $(\operatorname{curl} H) \cdot e_2$ appear always positive, always negative, always zero, or none of these?

(c) Does the sign of $(\operatorname{curl} H) \cdot e_1$ appear always positive, always negative, always zero, or none of these?

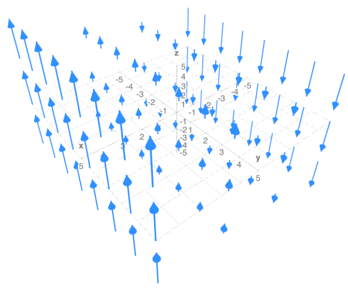
14.3.5 A vector field F in $\mathbb{R}^3$ is **irrotational** (or **curl-free**) on $U \subseteq \mathbb{R}^3$ if and only if $\operatorname{curl}(F) = 0$ on U. Four of the following vector fields are curl-free. Identify them.



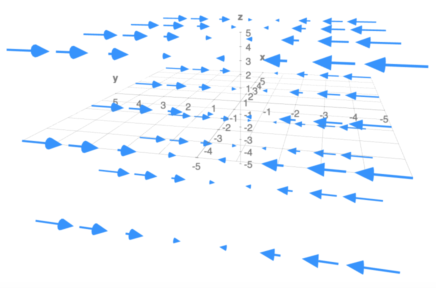
(a)



(d)



(b)



(e)

(c) $F(x,y,z) = (x,y,z)$

(f) $G(x,y,z) = (y,x,z)$

Computations

14.3.6 A fluid has velocity vector field $G : \mathbb{R}^3 \to \mathbb{R}^3$ given by $G(x,y,z) = (xy, yz, z^2)$. How fast and which way is the fluid spinning at the point $p = (1,-1,1)$ around the axis defined by $n = (0, 1/\sqrt{2}, 1/\sqrt{2})$?

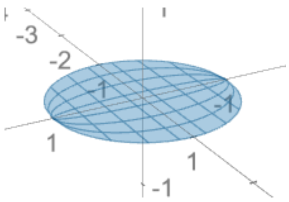
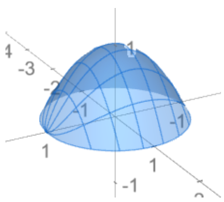
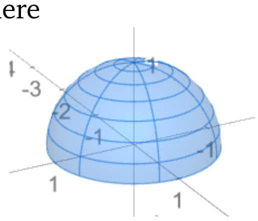
14.3.7 Each surface S is oriented with upward normal. They all have the same boundary,

$$C = \partial S = \{(x,y,0) \in \mathbb{R}^3 : x^2 + y^2 = 1\}$$

Orient C so that it is counterclockwise viewed from above. Let's compare two measurements:

- **circulation of F along the curve C**
- **total circulation of F on the surface S**

for the vector field $F(x,y,z) = (x-y, y+z, 3z-2)$. Fill in the table below with your classmates.

Surface S with boundary $\partial S = \{(x,y,0) \in \mathbb{R}^3 : x^2+y^2=1\}$	$\oint_C F \cdot T\,ds$	$\iint_S (\operatorname{curl} F) \cdot n\,dS$
(a) A disk  $D(s,t) = (s,t,0)$ for $s^2+t^2 \le 1$ $(\partial_1 D \times \partial_2 D)(s,t) = (0,0,1)$		
(b) A paraboloid  $P(s,t) = (s,t,1-s^2-t^2)$ for $s^2+t^2 \le 1$ $(\partial_1 P \times \partial_2 P)(s,t) = (2s, 2t, 1)$		
(c) A hemisphere  $H(s,t) = (\cos(s)\sin(t), \sin(s)\sin(t), \cos(t))$ $(\partial_1 H \times \partial_2 H)(s,t) =$ $\big(-\cos(s)\sin^2(t), -\sin(s)\sin^2(t), -\cos(t)\sin(t)\big)$		

(a) Calculate one of the rows in the above table. Fill in the entire table with classmates. Setup the integral(s) and use WolframAlpha to evaluate any single integrals or double integrals.

(b) Record your observations from the full table.

14.3.8 Compute the curl of $F(x,y,z) = (e^x, \cos y, e^{z^2})$. Determine whether it is irrotational.

14.3.9 Let $F(x,y,z) = (x+yz, y^2+xzy, zx^3y^2+x^7y^6)$ be the velocity vector field of a fluid.

(a) Compute the curl of F.

(b) Determine whether F is curl-free.

(c) At the point $(1,1,1)$, around which direction is the fluid spinning fastest counterclockwise? And how fast is it spinning in that direction?

14.3.10 Give an example of a vector field in $\mathbb{R}^3$ that has constant curl $(2,3,7)$.

Proofs

14.3.11 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field and let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^1 scalar function. Prove the identity

$$\operatorname{curl}(fF) = f\operatorname{curl}(F) + (\nabla f) \times F$$

Applications and beyond

14.3.12 Finally, there is the **analytic** perspective of curl in $\mathbb{R}^3$. For each unit vector n and point p, the quantity $(\operatorname{curl} F)(p) \cdot n$ should be the circulation density at p around the axis n. This is easiest to describe "component-by-component" using the standard basis vectors $n \in \{e_1, e_2, e_3\}$.

(a) Formulate this idea precisely using a C^1 vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ and a point $p = (x,y,z) \in \mathbb{R}^3$. You can prove your conjecture after the next section.

- For $\varepsilon > 0$, let $C_\varepsilon(p)$ be the closed curve parametrized by (______ , ______ , ______) for $0 \le t \le 2\pi$. Then

$$(\operatorname{curl} F)(p) \cdot e_3 = \lim_{\varepsilon \to 0^+} \left[\frac{1}{\pi\varepsilon^2} \oint_{C_\varepsilon(p)} F \cdot T\, ds \right]$$

- For $\varepsilon > 0$, let $C_\varepsilon(p)$ be the closed curve parametrized by (______ , ______ , ______) for $0 \le t \le 2\pi$. Then

$$(\operatorname{curl} F)(p) \cdot e_2 = \lim_{\varepsilon \to 0^+} \left[\frac{1}{\pi\varepsilon^2} \oint_{C_\varepsilon(p)} F \cdot T\, ds \right]$$

- For $\varepsilon > 0$, let $C_\varepsilon(p)$ be the closed curve parametrized by (______ , ______ , ______) for $0 \le t \le 2\pi$. Then

$$(\operatorname{curl} F)(p) \cdot e_1 = \lim_{\varepsilon \to 0^+} \left[\frac{1}{\pi\varepsilon^2} \oint_{C_\varepsilon(p)} F \cdot T\, ds \right]$$

(b) For each of your three conjectures, sketch a picture illustrating it.

14.3.13 Let S be a surface in $\mathbb{R}^3$ oriented with unit normal n. Let F be a vector field in $\mathbb{R}^3$ that is C^1 on S. Since curl is circulation density at a point (in any given direction), you can construct a measure for the **total circulation of F on the surface S**, namely

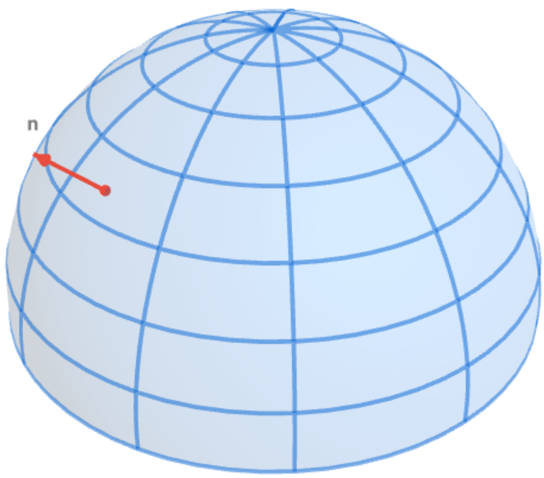
$$\iint_S (\operatorname{curl} F) \cdot n \, dS.$$

Imagine the ocean swirling over the globe!

(a) Fill in the blanks for the heuristic justification of this formula.

- *Consider an infinitesimal piece of the orientable surface S with infinitesimal area dS.*
- *The direction orthogonal to the surface is the unit normal n.*
- *The* ____ *is therefore* $(\operatorname{curl} F) \cdot n$.
- *Thus, the infinitesimal circulation is* ____________________________.
- ____________________________ *gives the total circulation* $\iint_S (\operatorname{curl} F) \cdot n \, dS$.

(b) The "chop, estimate, refine" method can make this rigorous but the details are quite technical. To keep things simple, you will illustrate a sketch. Here is an upper hemisphere which has been "chopped" into pieces. On each piece of the surface, sketch a "swirly" arrow **according to the righthand rule** (with respect to the unit normal).



What does the quantity $(\operatorname{curl} F) \cdot n$ represent about the fluid on each piece of the surface?

14.3.14 Let S be the first octant part of the sphere $x^2 + y^2 + z^2 = a^2$ with normal vector pointing away from the origin and $F(x, y, z) = (zx,\ zy,\ z^2)$.

(a) Calculate the circulation $\displaystyle\int_C F \cdot T \, ds$ along $C = \partial S$ which is oriented so that it goes from $(a, 0, 0)$ to $(0, a, 0)$ to $(0, 0, a)$ back to $(a, 0, 0)$.

(b) Find the total amount of swirl on the surface by calculating $\displaystyle\iint_S (\operatorname{curl} F) \cdot n \, dS$.

14.4. Stokes' theorem

Now that you have formalized curl as infinitesimal circulation, you can return to a central question on the physics of vector fields in $\mathbb{R}^3$ and their tangential flows.

Can you relate the work done by two different paths C_1 and C_2 of a vector field in $\mathbb{R}^3$?

Assuming C_1 and C_2 start and end at the same points, you can reformulate this question in terms of a closed curve $C = C_1 - C_2$.

Is there another way to calculate the circulation of a vector field along C in $\mathbb{R}^3$?

Over $\mathbb{R}^2$, these questions were resolved by Green's theorem and you will spectacularly generalize this idea to $\mathbb{R}^3$ with *Stokes' theorem*! This theorem holds for all vector fields, including non-conservative ones, and forms a bridge between circulation and curl. It is your third and final fundamental theorem for vector calculus in $\mathbb{R}^3$ and hence the physical dynamics of fluid flows.

14.4.1 Stokes orientation

To state Green's theorem in $\mathbb{R}^2$, you needed to orient the *topological* boundary ∂R of a two-dimensional regular region $R \subseteq \mathbb{R}^2$ in a consistent way; see Definition 12.2.3 for details. For Stokes' theorem in $\mathbb{R}^3$, you will need to orient the *relative* boundary ∂S of a two-dimensional surface $S \subseteq \mathbb{R}^3$ in a consistent way. Unlike Green's theorem, there are therefore actually *two* orientations to consider: the relative boundary ∂S and the surface S. The Stokes orientation ensures that these two orientations are "consistent" with respect to a righthand rule.

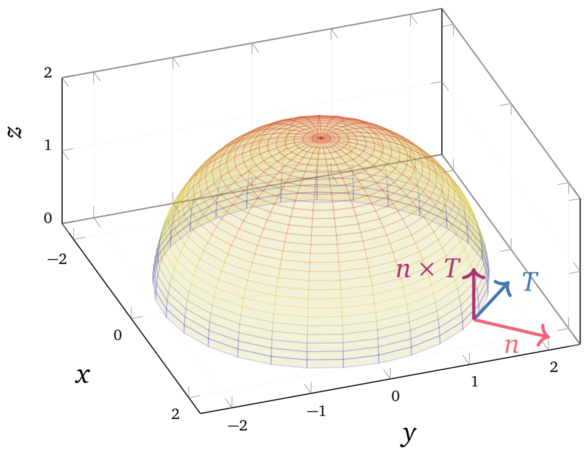
Given an oriented surface S, its relative boundary ∂S has the ***Stokes orientation*** *if S is always on the left as you traverse the boundary ∂S with your head pointing in the unit normal direction.*

This heuristic understanding will be sufficient for your purposes. A fully rigorous phrasing will be unnecessarily technical, but you can still describe it somewhat more formally.

If n is the unit normal of the oriented surface S, and T is the unit tangent of the oriented boundary ∂S, then ∂S has the Stokes orientation provided $n \times T$ points towards the surface S.

These informal descriptions will not make any sense until you explore some examples.

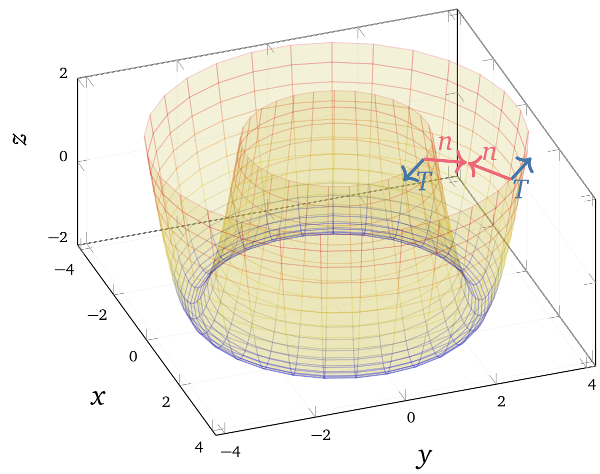
Example 14.4.1 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 4, z \geq 0\}$ be the upper hemisphere oriented with *upward* unit normal n. How can you orient its boundary $\partial S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, z = 0\}$ so that ∂S has the Stokes orientation?



The above diagram demonstrates the correct choice of unit tangent T for the curve ∂S. As you walk along the boundary with unit tangent T and your head pointing in the unit normal n direction, the surface always stays to the left. Play with this Math3D demo.

It can be quite confusing to visualize this "righthand rule" for the Stokes orientation.

Example 14.4.2 The diagram below illustrates a surface S oriented with unit normal n and its boundary ∂S oriented with unit tangent T.



Play with this Math3D demo for a better visual. On the "outer" boundary, the unit normal n of the surface S is pointing towards the z-axis. Thus, to keep the surface to your left as you traverse this piece of the boundary ∂S with your head in the unit normal direction, you must travel counter-clockwise as viewed from above; this choice is indicated by the unit tangent. On the "inner" boundary, the unit normal points away from the z-axis. To keep the surface to your left as you traverse this piece of the boundary ∂S with your head in the unit normal direction, you instead need to walk along this curve clockwise as viewed from above.

The need for the Stokes orientation comes from the ideas behind Stokes' theorem.

14.4.2 Statement and proof sketch

Stokes' theorem is your last version of the fundamental theorem of calculus in $\mathbb{R}^3$, adding to the fundamental theorem of line integrals and the divergence theorem. You can relate circulation along the relative boundary of a surface (a line integral of a vector field) to the total circulation over the surface (a surface integral of a vector field derivative).

Theorem 14.4.3 (Stokes' theorem) Let $S \subseteq \mathbb{R}^3$ be a surface oriented with unit normal n and whose boundary ∂S is a finite disjoint union of closed piecewise curves. Let F be a vector field in $\mathbb{R}^3$ that is C^1 on an open set containing S. If ∂S is endowed with the Stokes orientation, then

$$\oint_{\partial S} (F \cdot T)\, ds = \iint_S (\operatorname{curl} F) \cdot n\, dS.$$

Remark 14.4.4 The surface integral over the curl is the **total circulation of F over S**.

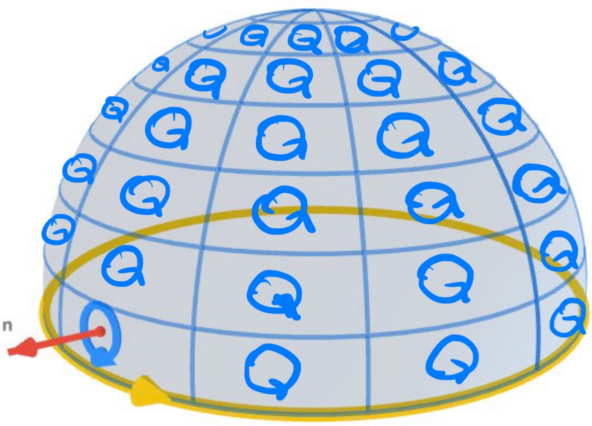
Exactly as with Green's theorem, this can be succinctly summarized in a heuristic manner.

The total infinitesimal circulation over S is the circulation along its boundary ∂S.

This can again be translated more informally[6].

> *The total amount of swirly-ness on a surface is the amount of swirl along its edge.*

These heuristic explanations can be more formally seen in the following Math3D demo.



Notice the $(\operatorname{curl} F) \cdot n$ measures the infinitesimal circulation around the unit normal n on S with respect to the righthand rule. If the boundary ∂S has an orientation consistent with this righthand rule, then ∂S has the Stokes orientation. This diagram also represents a picture proof of Stokes' theorem with the following informal explanation.

> *Each infinitesimal piece dS of the surface S has infinitesimal circulation $(\operatorname{curl} F) \cdot n\, dS$. By totaling all of these pieces, you obtain*
>
> $$\iint_S (\operatorname{curl} F) \cdot n\, dS$$
>
> *On the other hand, the circulation along the inside edges all cancel! The only remaining edges combine to give the circulation along the relative boundary ∂S endowed with the Stokes orientation. This implies that the total infinitesimal circulation also equals*
>
> $$\oint_{\partial S} F \cdot T\, ds$$

The ideas are exactly like Green's theorem (Theorem 12.2.6). Indeed, a fully rigorous proof proceeds along the same strategy but it is beyond the scope of this text. Only a crude outline is described.

Proof outline. First, you prove Stokes' theorem for a special type of surface by using Green's theorem. Assume all of the following:

- S is a surface parametrized by $G : R \to \mathbb{R}^3$.
- $R \subseteq \mathbb{R}^2$ is a regular region whose topological boundary ∂R is a finite disjoint union of piecewise closed curves.
- G maps the boundary $\partial R \subseteq \mathbb{R}^2$ continuously and bijectively to $\partial S \subseteq \mathbb{R}^3$.
- Orientation of ∂R is consistent with the orientation of ∂S under the map G.

[6]As your mathematics gets more intricate, your explanations will sound worse.

You can prove this special case of Stokes' theorem by applying Green's theorem. The main idea is to "pull back" these integrals in $\mathbb{R}^3$ to integrals in $\mathbb{R}^2$ via the map G. By linearity, it suffices to prove Stokes' theorem component-by-component. For F_1, you need to prove that

$$\oint_{\partial S} (F_1, 0, 0) \cdot T \, ds = \iint_S (0, \partial_3 F_1, -\partial_2 F_1) \cdot n \, dS.$$

You can expand the lefthand side in terms of your parametrization G, and obtain a line integral in $\mathbb{R}^2$. You can expand the righthand side in terms of your parametrization G and obtain a double integral over $\mathbb{R}^2$. This reduces the dimension to $\mathbb{R}^2$! You can use the chain rule and Green's theorem to show that these quantities are equal. The other components $(0, F_2, 0)$ and $(0, 0, F_3)$ will follow similarly. This will prove the special case.

Second, you establish the general case by chopping up the surface into these special type of surfaces. This strategy will be successful because the surface is "smooth" everywhere aside from a negligible subset of points, so a finite number of small enough pieces will satisfy all the required conditions. Upon totaling the contributions of each piece, the "common" edges will all cancel and the only remaining edges will be the boundary ∂S. ■

The proof strategy is once again a beautiful extension of Green's theorem; indeed, it relies on Green's theorem itself! Next, you can observe the power of Stokes theorem in some examples.

14.4.3 Examples with Stokes' theorem

You can begin with a simple calculation demonstrating the truth of Stokes' theorem.

Example 14.4.5 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = 0\}$ be the unit disk oriented with unit normal $n = (0, 0, 1)$. Its boundary

$$\partial S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z = 0\}$$

is a closed curve endowed with the Stokes orientation, so it is oriented counterclockwise when viewed from above. Let $F(x, y, z) = (y, z, x)$. By Stokes' theorem,

$$\oint_{\partial S} F \cdot T \, ds = \iint_S (\operatorname{curl} F) \cdot n \, dS.$$

You can verify this identity directly by evaluating both sides independently. First, the circulation of F along the boundary ∂S was directly computed in Example 14.3.2 to be $\oint_{\partial S} F \cdot T \, ds = -\pi$. Second, you can verify that

$$(\operatorname{curl} F)(x, y, z) = (\partial_2(x) - \partial_3(z), \partial_3(y) - \partial_1(x), \partial_1(z) - \partial_2(y)) = (-1, -1, -1)$$

so $\operatorname{curl}(F) \cdot n = -1$ everywhere on S. Thus, the total circulation of F over S is

$$\iint_S \operatorname{curl}(F) \cdot n dS = -\iint_S dS = -A(S) = -\pi$$

The calculations match so, indeed, Stokes' theorem holds as expected.

You can also swiftly calculate total circulation on a *closed* surface.

Example 14.4.6 Let $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ be the unit sphere with outward unit normal. The surface S is closed, so $\partial S = \varnothing$. By Stokes' theorem, the total circulation of

$F(x,y,z) = (z^2, x^2, y^2)$ over the surface S satisfies

$$\iint_S (\operatorname{curl} F) \cdot n \, dS = \oint_{\partial S} F \cdot T ds = \oint_{\varnothing} F \cdot T ds = 0.$$

There is literally nothing to calculate!

Remark 14.4.7 In the above example, you can recalculate the total circulation using the divergence theorem instead of Stokes' theorem. This is left as a short exercise.

You can continue your calculation-crushing parade with complicated vector fields.

Example 14.4.8 (Soap bubble) Let C be the closed curve in $\mathbb{R}^3$ parametrized by $\gamma(t) = (\cos t, \sin t, \cos t \sin t)$ for $0 \le t \le 2\pi$. You want to calculate the work done by

$$F(x,y,z) = (y + e^{x^2}, z - \sin(y^4), xy)$$

along C. A direct parametrization would be a nightmare since the vector field is quite complicated. Instead, you can place a "soap bubble surface" along the curve; that is, you can choose a surface S such that $\partial S = C$. For instance, define

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 \le 1, z = xy\} \tag{14.4.1}$$

so indeed $C = \partial S$. Viewed from above, the curve C is traversed counterclockwise. If you orient S with an upward unit normal, then S and its boundary ∂S will have the Stokes orientation. You can verify that $\operatorname{curl} F = (x-1, -y, -1)$ so, by Stokes' theorem, the work done by F along C satisfies

$$\oint_C F \cdot T \, ds = \iint_S (\operatorname{curl} F) \cdot n \, dS = \iint_S (x-1, -y, -1) \cdot n dS.$$

All that remains is to parametrize S and calculate the surface integral. The surface S is parametrized by $G : U \to \mathbb{R}^3$ given by $U = \{(s,t) \in \mathbb{R}^2 : s^2 + t^2 \le 1\}$ and $G(s,t) = (s,t,st)$. You can verify that

$$(\partial_1 G \times \partial_2 G)(x,y) = (-t, -s, 1)$$

Since the 3rd component is always positive, the cross product of partials points upwards, which is consistent with the orientation of S. Thus,

$$\iint_S (x-1,-y,-1) \cdot n dS = \iint_U (s-1,-t,-1) \cdot (-t,-s,1) dA = \iint_{s^2+t^2 \le 1} t - 1 ds dt.$$

You can calculate this integral directly with rectangular coordinates, with polar coordinates, or by symmetry; you will use symmetry here. Notice that the region $s^2 + t^2 \le 1$ is symmetric under the transformation $(s,t) \mapsto (s,-t)$ and the function $f(s,t) = t$ satisfies $f(s,-t) = -f(s,t)$. Thus, by symmetry the above is equal to

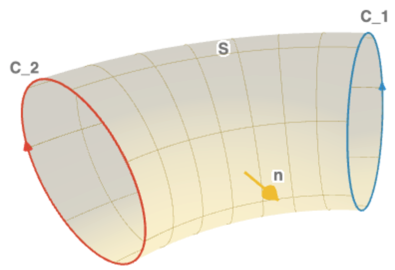
$$= \iint_{s^2+t^2 \le 1} t ds dt + \iint_{s^2+t^2 \le 1} -1 ds dt = 0 - \pi = -\pi,$$

because the latter double integral is equal to the area of the unit disk in $\mathbb{R}^2$. Overall, the work done by F along C is equal to $-\pi$.

Remark 14.4.9 Notice the choice of surface S in (14.4.1) such that $\partial S = C$ is not unique. For instance, you could instead define S by $x^2 + y^2 \leq 1$ and $z = xy + (1 - x^2 - y^2)^{237}$. Assuming you orient S with the Stokes orientation, the total circulation over S will always be the same by Stokes' theorem since this will always equal $\oint_C F \cdot T\, ds$. This invariance under the choice of surface is rather a remarkable consequence.

The true power of Stokes' theorem involves relating line integrals.

Example 14.4.10 (Moving a curve) Let C_1 and C_2 be two curves connected by a tube surface $S \subseteq \mathbb{R}^3$ as illustrated in this Math3D demo. Each curve and surface are oriented as indicated.



Stokes' theorem allows you to relate the work done by a C^1 vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ along C_1 to the work done along C_2 via the connecting surface S. Using the righthand rule, you can verify that $\partial S = -C_1 + C_2$ so by Stokes' theorem, it follows that

$$\begin{aligned} &\oint_{-C_1+C_2} F \cdot T\, ds = \iint_S \operatorname{curl}(F) \cdot n\, dS \\ \Longrightarrow &\oint_{C_2} F \cdot T\, ds = \iint_S \operatorname{curl}(F) \cdot n\, dS + \oint_{C_1} F \cdot T\, ds. \end{aligned}$$

If you are given specific equations and formulas, then this relationship can be exploited in calculations if you want to move your line integral. Sometimes you may need to come up with the connecting surface S itself.

When calculating total circulation, Stokes' theorem can also be used to move surfaces.

Example 14.4.11 (Moving a surface) Orient the upper hemisphere of radius 2 given by

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 4, z \geq 0\}$$

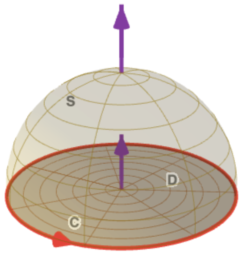
with upward normal. You can relate the total circulation of a C^1 vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ over the surface S to any other surface *with the same relative boundary*

$$C = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, z = 0\}.$$

For instance, the disk

$$D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 4, z = 0\}$$

also has relative boundary C. Orienting the surface D with upward normal and the curve C counterclockwise viewed from above, you obtain the following diagram.



Using the righthand rule, you can verify that $\partial S = C$ and $\partial D = C$, so both boundaries have the Stokes orientation. Thus, applying Stokes' theorem twice,

$$\iint_S (\operatorname{curl} F) \cdot n \, dS = \oint_C F \cdot T \, ds = \iint_D (\operatorname{curl} F) \cdot n \, dS.$$

Equivalently, you can notice that $S - D$ is a closed oriented surface with outward normal so, by Stokes' theorem,

$$\oiint_{S-D} (\operatorname{curl} F) \cdot n dS = \oint_{\varnothing} F \cdot T \, ds = 0 \implies \iint_S (\operatorname{curl} F) \cdot n \, dS = \iint_D (\operatorname{curl} F) \cdot n \, dS.$$

The strategies of *moving a surface* or *moving a line integral* via Stokes' theorem are generalizations to *moving a curve* with Green's theorem. It allows you to relate the work done across two curves by a surface integral over the curl.

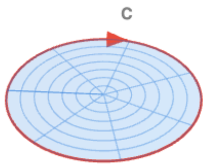
Stokes' theorem provides a deep relationship between circulation and 3-dimensional curl. This fantastic result is a beautiful combination of almost every calculus concept that you have encountered so far. It gives significant insight on the tangential flow of a fluid in $\mathbb{R}^3$ and acts as a perfect complement to the divergence theorem's implications for the normal flow of a fluid in $\mathbb{R}^3$. In the next section, you will take a step back to look at the big picture. By reviewing your achievements for vector calculus in $\mathbb{R}^3$, you can find stunning connections between the three differential operators: div, grad, and curl.

Exercises for Section 14.4

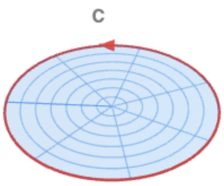
Concepts and definitions

14.4.1 The boundary of each surface below is a finite disjoint union of smooth oriented curves. If possible, orient the surface with a unit normal so its boundary has the Stokes orientation.

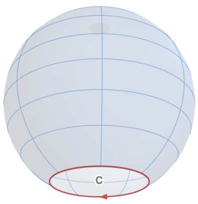
(a)



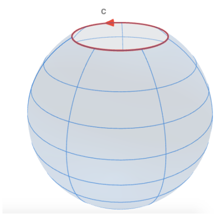
(b)



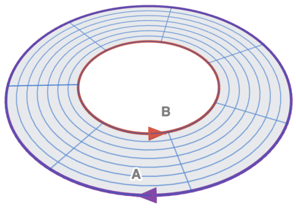
(c) (viewed from below)



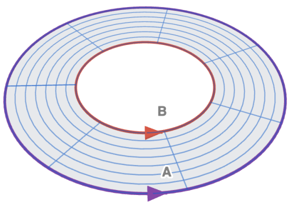
(d) (viewed from above)



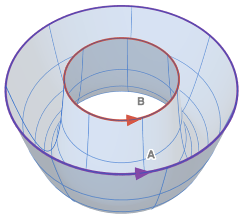
(e)



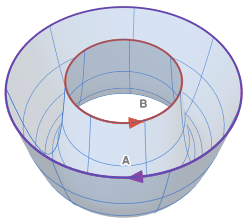
(f)



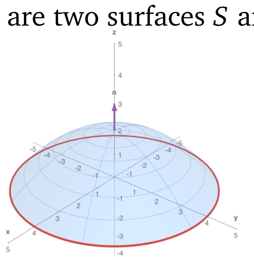
(g)



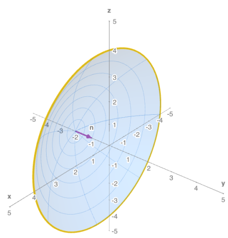
(h)



14.4.2 Illustrated below are two surfaces S and T.



surface S

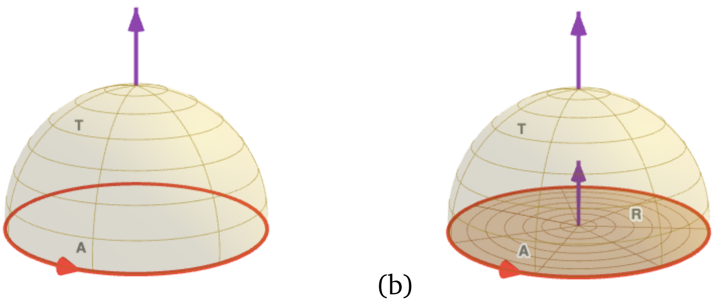


surface T

(a) The surface S is oriented with upward normal. Parametrize its boundary ∂S, the circle $x^2 + y^2 = 16$ in the xy-plane, with the Stokes orientation.

(b) The surface T is oriented with normal pointing towards the origin. Parametrize its boundary ∂T, the circle $x^2 + z^2 = 16$ in the xz-plane, with the Stokes orientation.

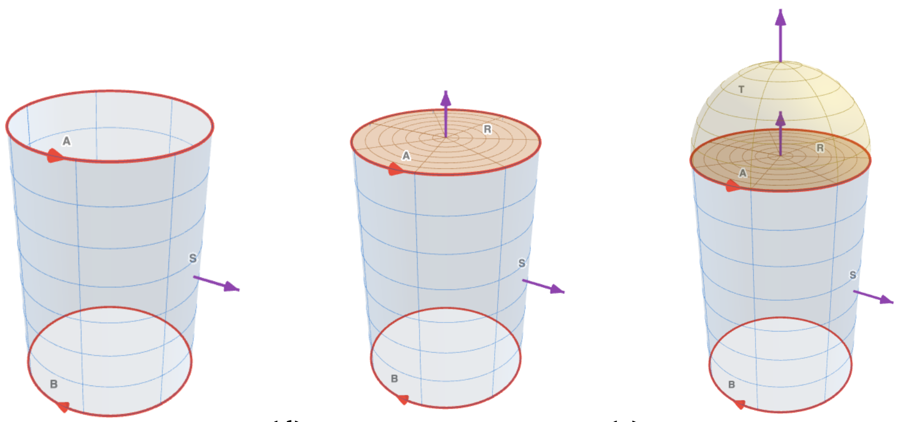
14.4.3 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field. View the surfaces and curves in this Math3D demo.



(a) (b)

(a) A hemispherical top surface $T \subseteq \mathbb{R}^3$ with boundary curve $A = \partial T$ are oriented as shown below. Express $\oint_A (F \cdot T)\,ds$ as a double integral.

(b) A disk-shaped roof R closes off this top and is oriented as illustrated. Relate the total circulation of F over the top T to the total circulation of F over the disk R.



(c) (d) (e)

(c) A silo surface $S \subseteq \mathbb{R}^3$ has a piecewise boundary ∂S with a top circle A and bottom curve B. Each piece is oriented as illustrated below. Relate the work done by F along A to the work done by F along B.

(d) The silo surface S can be capped off with the roof R from before. Give 3 different expressions for the circulation of F along B.

(e) The top T is installed onto this silo S with a roof R. Give 4 different expressions for the total circulation of F over R.

Computations

14.4.4 Here's two rapid calculations with Stokes' theorem.

(a) Compute the total circulation of $H(x, y, z) = (-y \sin x, xe^{z^2}, \cos(xyz))$ over the surface of the unit sphere S^2 with outward normal.

(b) Find the circulation of $F(x, y, z) = (x, y, z)$ over the circle $(\cos t, \sin t, 0)$ for $0 \le t \le 2\pi$.

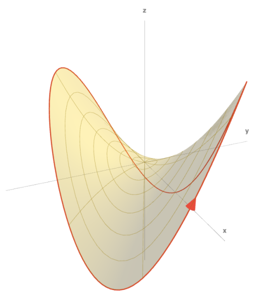
14.4.5 The previous two examples illustrate a more general phenomenon related to Stokes' theorem. The proof of the incomplete lemma below must use Stokes' theorem for both items.

> **Lemma A.** *Let $S \subseteq \mathbb{R}^3$ be a surface oriented with unit normal n whose boundary ∂S is a finite disjoint union of closed piecewise curves. Let F be a vector field in $\mathbb{R}^3$ that is C^1 on an open set containing S. Both of the following hold:*
>
> *(a) If* ________________ *then* $\displaystyle\oint_{\partial S} (F \cdot T)\, ds = 0.$
>
> *(b) If* ________________ *then* $\displaystyle\iint_S (\operatorname{curl} F) \cdot n \, dS = 0.$

Fill in the blanks.

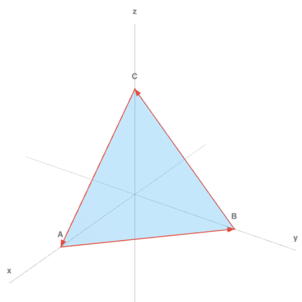
14.4.6

Let P be the Kringle surface: the piece of the saddle $z = xy$ inside the cylinder $x^2 + y^2 = 1$. Orient the boundary $C = \partial P$ so that it traverses the pringle counterclockwise when viewed from above. Calculate the work done by $F(x, y, z) = (y, z, x)$ along C using Stokes' theorem.

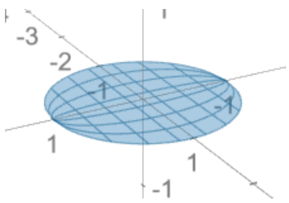


14.4.7

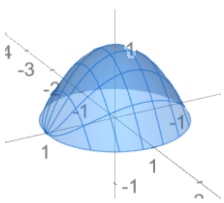
Let $T \subseteq \mathbb{R}^3$ be the triangular surface with vertices $A = (1,0,0)$, $B = (0,1,0)$ and $C = (0,0,1)$. Use Stokes' theorem to compute the work done by the force $F(x, y, z) = \big(x+z, x-y+2z, y-4x\big)$ along the perimeter of T, where the path starts at A, then goes to B, then C and returns to A.



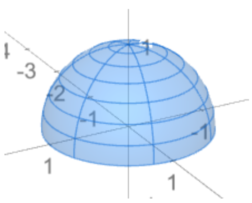
14.4.8 Let $F(x, y, z) = (x - y, y + z, 3z - 2)$. All three surfaces are oriented with **upward** unit normal.



(a)



(b)



(c)

(a) Disk $D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = 0\}$
(b) Paraboloid $P = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = 1 - x^2 - y^2\}$
(c) Hemisphere $H = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \geq 0\}$

Find the total circulation of F over each surface. *Hint:* This is a three-for-one!

14.4.9 Let M be the part of the paraboloid $z = 81 - 4x^2 - 4y^2$ above $z = 17$. Orient the surface M with unit normal n pointing out of the paraboloid. Let $F(x, y, z) = (xy, 2yz, -xz)$.

(a) Parametrize its boundary $C = \partial M$ with the Stokes orientation.

(b) Express the circulation of F along C as a single variable integral. Evaluate it.

(c) Parametrize M and determine the unit normal vector field n of M.

(d) Express the total circulation of F on M as an iterated double integral. Evaluate it.

14.4.10 Let $T \subseteq \mathbb{R}^3$ be the closed torus surface with inner radius 3 and outer radius 5. Assume the $z = 0$ plane cuts the torus in half like a bagel sliced into two pieces. Let $F(x, y, z) = (x^2 y, \sin(xyz), e^{xy^2} - xyz^4)$. Find the total circulation of F on T.

14.4.11 Here you will evaluate

$$\iint_S (\operatorname{curl} F) \cdot n \, dS$$

when $F(x, y, z) = (xy, -2, \arctan x^2)$ and S is the part of the paraboloid $z = 9 - x^2 - y^2$ above the xy—plane and with upper unit normal vector. As usual, you could directly parametrize the surface integral and calculate it. Use Stokes' theorem to solve the problem in two other ways:

(a) by computing a line integral along the boundary of S.

(b) by integrating $\operatorname{curl} F$ on the disk contained in xy-plane that has the same boundary as S.

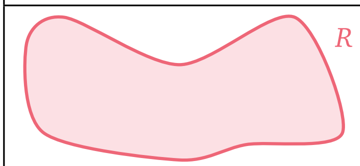
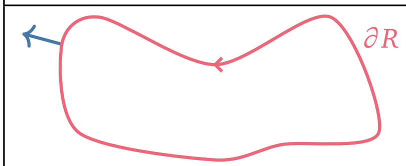
Proofs

14.4.12 Let F be a C^1 vector field in $\mathbb{R}^3$ on an open ball containing $p \in \mathbb{R}^3$. Use Stokes' theorem to prove that curl is circulation density. That is, prove your conjecture in Exercise 14.3.2 .

14.4.13 Prove that Green's theorem follows from Stokes' theorem.

Applications and beyond

14.4.14 Stokes' theorem is the $\mathbb{R}^3$ analogue of Green's theorem in circulation form for $\mathbb{R}^2$. You should lean on this analogy to help build connections between theory. Fill in the table below.

Green's theorem (curl form)	Stokes' theorem
$\oint_{\partial R} (F \cdot T) \, ds$	
$\iint_R \operatorname{curl}(F) dA$	
	
	

14.5. Div, grad, and curl

After monumental efforts, you have discovered the three grand theorems of vector calculus in $\mathbb{R}^3$: the fundamental theorem of line integrals (Theorem 11.4.1), Stokes' theorem (Theorem 14.4.3), and the divergence theorem (Theorem 14.2.5). These are informally recapped below.

If $f : \mathbb{R}^3 \to \mathbb{R}$ is a C^1 scalar field and C is an oriented curve from p to q, then

$$\int_C \operatorname{grad}(f) \cdot T \, ds = f(q) - f(p).$$

If $G : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field and S is an oriented surface whose boundary ∂S is a closed curve with the Stokes orientation, then

$$\iint_S \operatorname{curl}(G) \cdot n \, dS = \oint_{\partial S} G \cdot T \, ds.$$

If $F : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field and R is a regular region whose boundary ∂R is a closed surface with outward unit normal, then

$$\iiint_R \operatorname{div}(F) \, dV = \oiint_{\partial R} F \cdot n \, dS.$$

Each result mirrors the fundamental theorem of calculus, where the integral of a derivative returns the map itself. These three "derivatives" are the differential operators div, grad, and curl. In addition to these seemingly miraculous parallels, these operators have important relationships. You will explore these connections more deeply and uncover sufficient criterion to be a gradient vector field, i.e. $F = \nabla f$, or a curl vector field, i.e. $F = \operatorname{curl}(G)$. This final section will conclude with an elegant diagram capturing the essence of vector calculus in $\mathbb{R}^3$.

14.5.1 Gradient and curl

First, grad and curl are intimately related.

Gradient vector fields are curl-free; that is, $\operatorname{curl}(\operatorname{grad}(f)) = (0,0,0)$.

This observation was summarized long ago in Lemma 11.4.11 and Lemma 14.3.6.

When are curl-free vector fields the same as gradient vector fields?

As Example 11.4.14 shows, they are not always the same; this pattern continues for $\mathbb{R}^3$.

Example 14.5.1 You can verify that the three-dimensional vector field

$$F(x,y,z) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}, 0 \right)$$

is curl-free. On the other hand, F is not a gradient vector field. You can verify that

$$\oint_C F \cdot T \, ds = 2\pi$$

where $C = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 = 1, z = 0\}$ is the circle oriented counterclockwise viewed from above. By Theorem 11.5.2, F cannot be a gradient vector field.

Despite this setback, Poincaré's lemma and Theorem 11.5.14 demonstrates that the issue has topological origins. If the domain of the vector field does not contain any loop holes, then irrotational vector fields are the same as gradient vector fields. Poincaré's lemma is restated below to remind you of this observation.

Lemma 14.5.2 Let F be a vector field in $\mathbb{R}^3$ that is C^1 on an open set $U \subseteq \mathbb{R}^3$. Assume U is convex. If F is curl-free on U, then F is a gradient vector field. That is, there exists a scalar-valued C^2 function f on U such that $F = \operatorname{grad}(f)$ on U.

Proof. See Lemma 11.5.7. ∎

This lovely result can be used to swiftly identify gradient vector fields.

Example 14.5.3 Is the vector field

$$F(x,y,z) = (z\cos x\cos y, -z\sin x\sin y, \sin x\cos y)$$

a gradient vector field? That is, does there exist a potential f such that $F = \nabla f$? Notice that

$$\begin{aligned}
&(\operatorname{curl} F)(x,y,z)\\
=&(\partial_2(\sin x\cos y) - \partial_3(-z\sin x\sin y),\ \partial_3(z\cos x\cos y) - \partial_1(\sin x\cos y),\\
&\quad \partial_1(-z\sin x\sin y) - \partial_2(z\cos x\cos y))\\
=&(-\sin x\sin y + \sin x\sin y, \cos x\cos y - \cos x\cos y, -z\cos x\sin y + z\cos x\sin y)\\
=&(0,0,0)
\end{aligned}$$

Since F is C^1 and curl-free on its convex domain $U = \mathbb{R}^3$, it follows by Lemma 14.5.2 that F is a gradient vector field. Note this only shows the existence of a potential function, but does not actually exhibit one. You need to solve some system of partial diferential equations to find a potential function. In this case, $f(x,y,z) = z\sin x\cos y$ is one such potential.

Curl and gradient are also related via the fundamental theorem of line integrals and Stokes' theorem. Namely, they can both demonstrate path independence of gradient vector fields.

Example 14.5.4 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 scalar field. Let C_1 and C_2 be oriented curves from point $p \in \mathbb{R}^3$ to point $q \in \mathbb{R}^3$. The fundamental theorem of line integrals implies that

$$\int_{C_1} \nabla f \cdot T\, ds = f(q) - f(p) = \int_{C_2} \nabla f \cdot T\, ds$$

This proves that the gradient vector field ∇f is path independent.

Alternatively, define the closed oriented curve $C = C_1 - C_2$. Assume there exists[7] a surface $S \subseteq \mathbb{R}^3$ such that $\partial S = C$. Orienting S with the Stokes orientation, it follows that

$$\int_C \nabla f \cdot T\, ds = \iint_S \operatorname{curl}(\nabla f)\cdot n\, dS = \iint_S 0\, dS = 0$$

as $\operatorname{curl}(\nabla f) = (0,0,0)$ by Lemma 11.4.11. Since $C = C_1 - C_2$, this again implies that the gradient vector field ∇f is path-independent.

[7] See Example 14.4.8 for an example of constructing this "soap bubble" surface. However, it is not always possible to create such a surface for a given closed curve; see, for instance, the trefoil knot.

14.5.2 Curl and divergence

Second, the differential operators curl and div are also tied together.

> *Curl vector fields are divergence-free; that is,* $\operatorname{div}(\operatorname{curl}(G)) = 0$.

This was confirmed in Lemma 14.3.14 and leads to a natural question.

> *When are divergence-free vector fields the same as curl vector fields?*

As with curl and gradient, the answer is not always.

Example 14.5.5 You can verify that the vector field

$$F(x,y,z) = \frac{(x,y,z)}{(x^2+y^2+z^2)^{\frac{3}{2}}}$$

is divergence-free. On the other hand, you can also verify by direct calculation that

$$\oiint_S F\cdot n\,dS = 4\pi$$

where $S = \partial B_1(0,0,0)$ is the unit sphere centred at the origin with outward unit normal. Suppose, for a contradiction, that $F = \operatorname{curl}(G)$ for some C^1 vector field G. Stokes' theorem implies the surface integral of $\operatorname{curl}(G)$ over S must be zero since S is a closed surface and hence $\partial S = \emptyset$. This contradicts the above calculation, so F cannot be a curl vector field.

Based on your experience with gradient vector fields, you may suspect that topological subtleties explain this phenomenon with curl vector fields. Your suspicions are validated by the following analogue to Poincaré's lemma.

Lemma 14.5.6 Let F be a vector field in $\mathbb{R}^3$ that is C^1 on an open set $U \subseteq \mathbb{R}^3$. Assume U is convex. If F is divergence-free on U, then F is a curl vector field. That is, there exists a vector-valued C^2 function G on U such that $F = \operatorname{curl}(G)$ on U.

A rigorous proof is quite tricky; a better proof utilizes the more sophisticated theory of differential forms[8]. Nonetheless, if you are interested, here is a special case.

Proof. (Sketch) For simplicity, assume U is a rectangle, so $U = (a_1,a_2)\times(b_1,b_2)\times(c_1,c_2)$. This special case will allow for an ad hoc approach; the general case is notably more difficult. It suffices to show there exists C^2 vector field $G : U \to \mathbb{R}^3$ such that $G = (G_1, G_2, 0)$ and $F = \operatorname{curl}(G)$. We must therefore show there exists $G_1 : U \to \mathbb{R}$ and $G_2 : U \to \mathbb{R}$ satisfying

$$F_1 = -\partial_3 G_2, \qquad F_2 = \partial_3 G_1, \qquad F_3 = \partial_1 G_2 - \partial_2 G_1. \tag{14.5.1}$$

Fix a point $(a,b,c) \in U$. Since U is a rectangle, define for any $(x,y,z)\in U$

$$G_1(x,y,z) = \int_c^z F_2(x,y,t)\,dt + \phi_1(x,y) \qquad G_2(x,y,z) = -\int_c^z F_1(x,y,t)\,dt + \phi_2(x,y)$$

where ϕ_1 and ϕ_2 are C^1 functions yet to be chosen. By the single variable fundamental theorem of calculus, the first two conditions in (14.5.1) are automatically satisfied with these definitions.

[8]For differential forms, see a course in multivariable analysis or differential geometry.

It remains to verify the third condition. Assuming you can swap the partial derivative with the integral sign, it follows that

$$\begin{aligned}(\partial_1 G_2 - \partial_2 G_1)(x,y,z) &= \int_c^z -\partial_1 F_1(x,y,t) - \partial_2 F_2(x,y,t)dt + \partial_1\phi_2(x,y) - \partial_2\phi_1(x,y) \\ &= -\int_c^z (\partial_1 F_1 + \partial_2 F_2)(x,y,t)dt + (\partial_1\phi_2 - \partial_2\phi_1)(x,y)\end{aligned}$$

Since F is divergence-free, i.e. $\partial_1 F_1 + \partial_2 F_2 + \partial_3 F_3 = 0$, the above is equal to

$$\begin{aligned}&= \int_c^z \partial_3 F_3(x,y,t)dt + (\partial_1\phi_2 - \partial_2\phi_1)(x,y) \\ &= F_3(x,y,z) - F_3(x,y,c) + (\partial_1\phi_2 - \partial_2\phi_1)(x,y).\end{aligned}$$

This will give the desired result if $(\partial_1\phi_2 - \partial_2\phi_1)(x,y) = F_3(x,y,c)$. This can be achieved by defining $\phi_2(x,y) = \int_a^x F_3(t,y,c)dt$ and $\phi_1(x,y) = 0$ for all (x,y). ■

Like Poincaré's lemma, this result makes it much easier to identify curl vector fields.

Example 14.5.7 Is the vector field $F(x,y,z) = (-\sin y, 1, \cos x)$ a curl vector field? That is, does there exist some $G : \mathbb{R}^3 \to \mathbb{R}^3$ such that $F = \operatorname{curl}(G)$? Notice that

$$(\operatorname{div} F)(x,y,z) = \partial_x(-\sin y) + \partial_y 1 + \partial_z \cos(x) = 0$$

Since F is C^1 and divergence-free on its convex domain $U = \mathbb{R}^3$, it follows by Lemma 14.5.6 that F is a curl vector field on U. Indeed, $G(x,y,z) = (z, \sin(x), \cos(y))$ satisfies $F = \operatorname{curl}(G)$.

Divergence and curl are also related via Stokes' theorem and the divergence theorem. Namely, they can both relate flux integrals of curl vector fields.

Example 14.5.8 Let $G : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^2 vector field. Fix a closed oriented curve $C \subseteq \mathbb{R}^3$. Let S_1 and S_2 be oriented surfaces with the same relative boundary $C = \partial S_1 = \partial S_2$. Assuming C has the Stokes orientation for both surfaces, Stokes' theorem implies that

$$\iint_{S_1} (\operatorname{curl} G)\cdot n\, dS = \oint_C G\cdot T\, ds = \iint_{S_2} (\operatorname{curl} G)\cdot n\, dS.$$

Thus, the flux of $\operatorname{curl} G$ across any surface with boundary C is the same.

Alternatively, since S_1 and S_2 have the same relative boundary, the surface $S_1 - S_2$ should presumably[9] enclose a regular region $R \subseteq \mathbb{R}^3$. Assuming the closed piecewise surface $\partial R = S_1 - S_2$ is oriented with outward normal, the divergence theorem should imply that

$$\oiint_{S_1 - S_2} (\operatorname{curl} G)\cdot n\, dS = \iiint_R \operatorname{div}(\operatorname{curl} G)dV = \iiint_R 0 dV = 0$$

as $\operatorname{div}(\operatorname{curl} G) = 0$ by Lemma 14.3.14. Again, the flux of $\operatorname{curl} G$ across any surface with boundary C is the same.

[9] This assumption is not necessarily true, but it is heuristically reasonable.

14.5.3 A unified view

From the beginning, multivariable calculus has shown how you can do calculus with all of your linear algebra. Now, at the end of your journey, it is time to do linear algebra with all of your calculus (in three dimensions). This discussion will be informal but still meaningful.

Let $U \subseteq \mathbb{R}^3$ be an open set. Let $C^\infty(U)$ be the set of real-valued functions $f : U \to \mathbb{R}$ with infinitely many partial derivatives; that is, $\partial^\alpha f$ exists and is continuous on U for all multi-indices $\alpha \in \mathbb{N}^3$. The space of C^∞ scalar fields $V = C^\infty(U)$ and space of C^∞ vector fields $V^3 = V \times V \times V$ can each be thought of as a space of vectors. For example, the zero function belongs to V and acts like the zero vector. Moreover, any linear combination in V also belongs to V. Similar statements hold true for V^3. You can view the differential operators grad, curl, and div as linear transformations on these spaces.

- Gradient is a linear map of C^∞ scalar functions to C^∞ vector fields.
  That is, $\mathrm{grad} : V \to V^3$ is a linear map. Hence, if $f \in V$ then $\mathrm{grad}(f) \in V^3$.
- Curl is a linear map of C^∞ vector fields to C^∞ vector fields.
  That is, $\mathrm{curl} : V^3 \to V^3$ is a linear map. Hence, if $F \in V^3$ then $\mathrm{curl}(F) \in V^3$.
- Divergence is a linear map of C^∞ vector fields to C^∞ scalar functions.
  That is, $\mathrm{div} : V^3 \to V$ is a linear map. Hence, if $F \in V^3$ then $\mathrm{div}(F) \in V$.

Your observations about grad, div, and curl are beautifully encapsulated in this elegant diagram.

$$V \xrightarrow{\text{grad}} V^3 \xrightarrow{\text{curl}} V^3 \xrightarrow{\text{div}} V$$

At first glance, this appears to just be a composition of maps but if you dig a bit deeper, you will notice it actually captures the essence of vector calculus in $\mathbb{R}^3$. Take a moment to reflect on its meaning; see Exercise 14.5.18 for an outline.

Congratulations on reaching the end of your multivariable calculus journey! It has been a long and winding road. To build your foundations, you played with maps between different dimensions to model the real world; and you carefully constructed topological properties of sets in higher dimensions. During your development of derivatives, you blended linear algebra and calculus to create the differential and Jacobian; you masterfully applied these ideas to globally solve complex optimization problems, locally solve nonlinear systems, and define smooth manifolds; and you pushed all of this technology farther with higher order approximations and Taylor's theorem.

While evolving your theory of integration, you hammered the mantra of "chop, refine, estimate" to generalize the integral; you created the Jordan measure to average over any sensible shape; and you learned that expressions of mass, volume, or probability do not depend on how you describe your space. Blending all of this knowledge into vector calculus, you carefully defined curves and surfaces; you produced highly accurate models for work done by force fields and flux across surfaces; and you spectacularly discovered the fundamental theorems of calculus in two- and three-dimensions.

Many questions have been answered, and many more have been created. It has been an absolute pleasure to guide you on this stage of your mathematical journey. Thank you for your time and I wish you the best of luck in your next saga!

You must understand... it takes a long time to say anything... And we never say anything unless it is worth taking a long time to say. — Treebeard

Exercises for Section 14.5

Concepts and definitions

14.5.1 Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^2 vector field with $F = (F_1, F_2, F_3)$. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^2 scalar function. Which statements are true, false, or nonsense?

(a) $\operatorname{curl}(\nabla f) = 0$

(b) $\operatorname{curl}(\operatorname{curl}(F)) = 0$

(c) $\operatorname{div}(\nabla f) = 0$

(d) $\operatorname{div}(\operatorname{curl}(F)) = 0$

(e) $\operatorname{curl}(\operatorname{div}(F)) = 0$

(f) $\nabla(\operatorname{curl}(F)) = 0$

14.5.2 Which statements are true, false, or nonsense?

(a) Gradient vector fields are sourceless.

(b) Gradient vector fields are irrotational.

(c) Curl vector fields are sourceless.

(d) Curl vector fields are irrotational.

(e) Curl vector fields are conservative.

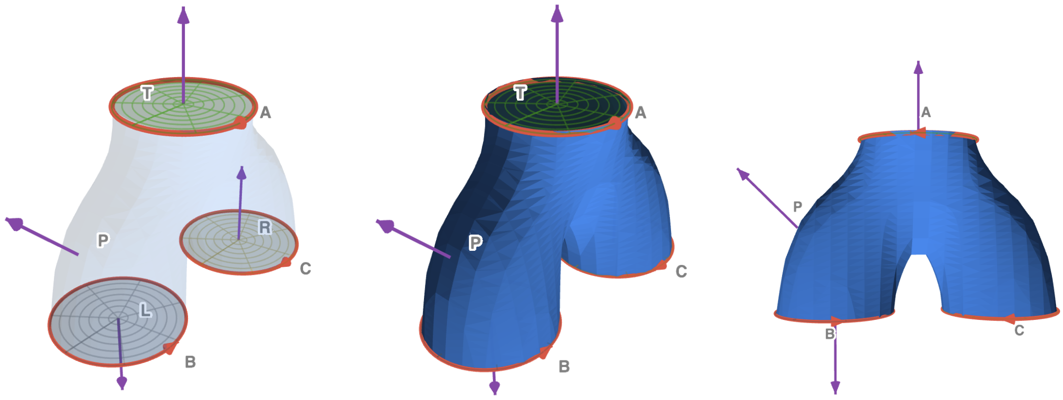
(f) Divergence vector fields are irrotational.

(g) Sourceless vector fields are conservative.

(h) Sourceless vector fields are curl vector fields.

14.5.3 Let $X \subseteq \mathbb{R}^3$ be a regular region. The topological boundary ∂X consists of 4 surfaces: the pants P, the top T, the left disk L, and the right disk R. The relative boundary ∂P consists of 3 curves: the waist hole A, left leg hole B, and right leg hole C. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^1 vector field.

All surfaces and curves are oriented as illustrated below. The leftmost figure and middle figure are **from the same perspective**, but the leftmost figure allows you to see through the pants. See this Math3D demo on Google Chrome for a better view.

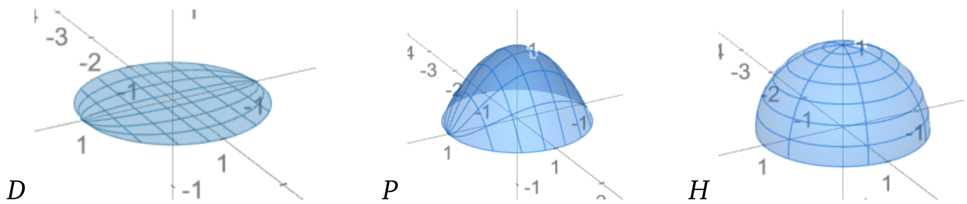


(a) Relate the flux of F through L to the flux of F through R using surface and volume integrals.

(b) Relate the work done by F along A to the work done by F along B and C via a surface integral.

(c) Relate the circulation of F along B to the circulation of F along C via a surface integral.

14.5.4 Let $F(x, y, z) = (x - y, y + z, 3z - 2)$. All three surfaces are oriented with **upward** unit normal.



- Disk $D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 1, z = 0\}$
- Paraboloid $P = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 1, z = 1 - x^2 - y^2\}$

- Hemisphere $H = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \geq 0\}$

Abbigael, Alisa, and Andy calculate all three surfaces to have the same total circulations:

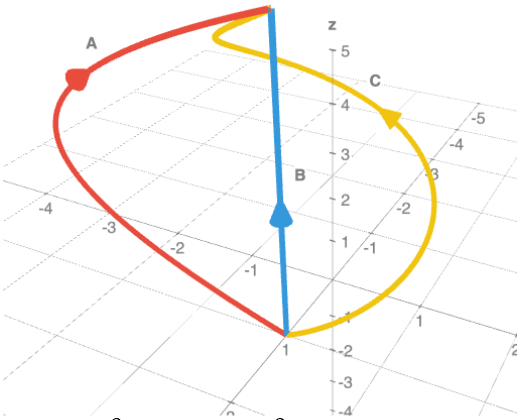
$$\iint_D (\operatorname{curl} F) \cdot n dS = \iint_P (\operatorname{curl} F) \cdot n dS = \iint_H (\operatorname{curl} F) \cdot n dS.$$

They are trying to decide whether it's a coincidence or not.

- Abbigael says "*This happens by Stokes' theorem!*"
- Alisa says "*This is because of the divergence theorem!*"
- Andy says "*This is just a silly coincidence!*"

Who is correct? Explain.

14.5.5 Three oriented curves travel from $(1,0,0)$ to $(1,0,2\pi)$. View this Math3D demo.



- Parabolic curve A parametrized by $\gamma_A(t) = \left(1, -t\frac{(2\pi-t)}{4}, t\right)$ for $0 \leq t \leq 2\pi$.
- Straight line B parametrized by $\gamma_B(t) = (1,0,t)$ for $0 \leq t \leq 2\pi$.
- Helix C parametrized by $\gamma_C(t) = (\cos t, \sin t, t)$ for $0 \leq t \leq 2\pi$.

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^2 scalar valued function. Amy, Sarah, and Fardin directly calculate all three oriented curves to have the same work done by the gradient vector field ∇f.

$$\int_A \nabla f \cdot d\gamma = \int_B \nabla f \cdot d\gamma = \int_C \nabla f \cdot d\gamma$$

They are trying to decide whether it's a coincidence or not.

- Amy says "*This happens by the fundamental theorem of line integrals!*".
- Sarah proclaims "*Our beloved Stokes' theorem explains this!*".
- Fardin exclaims "*This is truly a miraculous coincidence!*".

Who is correct? Explain.

Computations

14.5.6 Which of these are curl vector fields? Explain why or why not. *Hint:* Compute the divergence.

(a) $F(x,y,z) = (x^2, y^2, z^2)$

(b) $G(x,y,z) = (xz, -yz, xy)$

14.5.7 Is the vector field $F(x,y,z) = (-xz, yz - 4x^3, 1)$ a curl vector field? If so, find a vector field G so that $F = \operatorname{curl} G$. If not, prove it is not.

14.5.8 Let $F(x,y,z) = (x^2 + \cos(z^3), yz, 3xz)$ and $S = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \leq 0\}$ be the lower hemisphere. Calculate the downward flux of F through S by moving the surface to a disk.

14.5.9 Let T be the solid tetrahedron in $\mathbb{R}^3$ with vertices $(1,0,0),(0,1,0)$ and $(0,0,1)$. Calculate the flux of the force $F(x,y,z)=(x+z,x+y+2z,z+xy)$ outward through the three sides of T lying in the coordinate planes: xy-plane, yz-plane, and xz-plane.

14.5.10 Let S be the part of the cylinder $x^2+y^2=4$ with $1<z<3$. Notice the surface includes the round side of the can, but does not include the top or bottom. Calculate the flux of the vector field $F(x,y,z)=(x+y^3,y+zx^3,z^3)$ flowing away from the z-axis through the surface S.

14.5.11 Let $F(x,y,z)=\left(\frac{x}{x^2+y^2},\frac{y}{x^2+y^2},0\right)$ and C the pentagon with vertices $(0,1,1),(1,1,0)$, $(1,-2,-1),(-1,-3,1)$, and $(-2,0,0)$ travelled in that order. Calculate the circulation of F along C by moving the curve to a circle. Why can't we apply Stokes' theorem right away?

14.5.12 Let $F(x,y,z)=\operatorname{curl}(x^5-2zy,z^3-2,4x\sin(z))$ and let S be the upper half of the ellipsoid $x^2+\frac{y^2}{9}+4z^2=1$ with upward normal. Compute the flux of F along S by moving the integral to another surface.

14.5.13 Let S be the part of the cone $x^2+y^2=z^2$ with $1<z<3$ oriented downwards. Notice the surface does not include the tip of the cone. Calculate the circulation of $F(x,y,z)=(2z,x,x^2+y^2)$ along S.

14.5.14 Let S be the portion of the paraboloid $z=1-x^2-y^2$ with $z\geq 0$, oriented upwards. Let $F(x,y,z)=(-3xy,x^2+1,3yz)$. Compute $\iint_S F\cdot n\,dS$.

Proofs

14.5.15 Give **two proofs** of the following lemma.

> **Lemma.** *Let R be a regular region in $\mathbb{R}^3$ whose boundary $S=\partial R$ is a closed piecewise surface with outward unit normal. If $F:\mathbb{R}^3\to\mathbb{R}^3$ is a C^2 vector field, then the total circulation of F on S is zero.*

14.5.16 The moment of inertia about the z-axis for a solid $R\subseteq\mathbb{R}^3$ is defined by $I=\iiint_R (x^2+y^2)\,dV$. If R is a regular region whose boundary is a closed piecewise surface, then prove that

$$I=\frac{1}{4}\iint_{\partial R}\left(x^3+xy^2,x^2y+y^3,0\right)\cdot n\,dS.$$

Applications and beyond

14.5.17 The electric field $F:\mathbb{R}^3\setminus\{(0,0,0)\}\to\mathbb{R}^3$ generated by a point charge at the origin is given by

$$F(x,y,z)=\frac{(x,y,z)}{||(x,y,z)||^3}.$$

Gauss' law of electromagnetism captures a fundamental property of this electric field.

(a) Show F is sourceless.

(b) Let $R \subseteq \mathbb{R}^3$ be a regular region with piecewise smooth boundary ∂R. Show that if R does not contain the point charge then the outward flux of the electric field F through the surface ∂R is zero.

(c) For $\varepsilon > 0$, show by direct calculation of the surface integral that the outward flux of the electric field F through the ε-sphere $S_\varepsilon = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = \varepsilon^2\}$ is equal to 4π.

(d) Now, here's the punchline for electromagnetism.

> **Gauss' law.** *Let $R \subseteq \mathbb{R}^3$ be a regular region with piecewise smooth boundary ∂R. Show that if the interior of R contains the point charge then the outward flux of the electric field F through the surface ∂R is equal to 4π.*

Prove Gauss' law and illustrate your proof with a sketch. The idea is tremendously clever! *Hint:* Move the surface with (c).

(e) If R is a regular region in $\mathbb{R}^3$ with piecewise smooth boundary ∂R then what does the outward flux integral $\dfrac{1}{4\pi}\displaystyle\iint_{\partial R} F \cdot n dS$ represent? Explain in words. *Hint:* Combine (b) and (d).

14.5.18 Consider the diagram in Section 14.5.3. Take an arbitrary element at the leftmost V in the diagram. Map it once to V^3 and then map it again to the next V^3. The element has moved two stages to the right. What happened to this element? Repeat the same process for an arbitrary element starting at the leftmost V^3 in the diagram so it first maps to V^3 and then to V. The same thing happens. How do the three core theorems of vector calculus in $\mathbb{R}^3$ relate to this phenomenon? See Examples 14.5.4 and 14.5.8 for some insights.

Appendix

A. Hints and answers

A.1. Maps

1.1.1

(a) The domain of γ is $[0, 120]$.

(b) The codomain of γ should be $\mathbb{R}^2$.

(c) The image (or range) of γ does not change with time. It is the path of the duck.

1.1.2

(a) $||\gamma(6.1) - \gamma(6)||$

(b) $\gamma(6.1) - \gamma(6)$

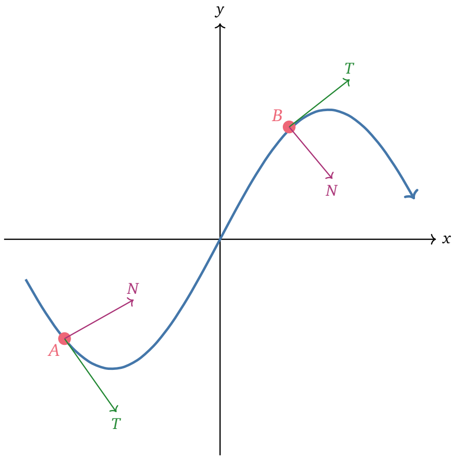
(c) $\frac{\gamma(6.1)-\gamma(6)}{0.1} = \frac{\gamma(6)-\gamma(6.1)}{-0.1}$

(d) $\frac{||\gamma(6)-\gamma(5.9)||}{0.1}$

(e) $\lim_{h\to 0} \frac{\gamma(6+h)-\gamma(6)}{h}$

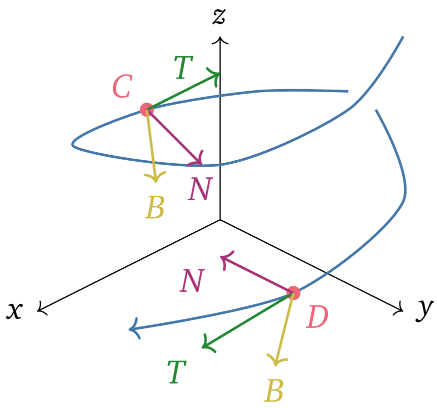
(f) $\lim_{h\to 0} \frac{||\gamma(6+h)-\gamma(6)||}{|h|} = \lim_{h\to 0} \left|\left|\frac{\gamma(6+h)-\gamma(6)}{h}\right|\right|$

1.1.3



Pay attention to the definition of the unit tangent and unit normal. They are not defined when the denominator is 0.

1.1.4



1.1.5

(a) This parametrizes the standard unit circle starting at $(1,0)$, then rotating counterclockwise exactly once. The velocity vector $\gamma_1'(t) = (-\sin t, \cos t)$ is tangent to the circle and the acceleration vector $\gamma_1''(t) = (-\cos t, -\sin t)$ points inwards towards the centre.

(b) This parametrizes the circle of radius 3 starting at $(0,3)$, then rotating clockwise exactly twice. The velocity is $\gamma_2'(t) = (6\cos(2t), -6\sin(2t))$ and the acceleration is $\gamma_2''(t) = (-12\sin(2t), -12\cos(2t))$.

(c) This parametrizes the circle of radius 3 starting at $(3,0)$, then rotating clockwise exactly once. Note the domain of γ_3 is $[-\pi, 0]$. The velocity is $\gamma_3'(t) = (-6\sin(2t), -6\cos(2t))$ and the acceleration is $\gamma_3''(t) = (-12\cos(2t), 12\sin(2t))$.

1.1.6 One example is $\gamma : [0,1] \to \mathbb{R}^2$ defined by

$$\gamma(t) = (1-t)\begin{bmatrix} 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2-3t \\ 1-t \end{bmatrix}.$$

To find another, change the time domain and rescale t accordingly.

1.1.7

(a) velocity $\gamma'(t) = (t+3, 2, 1)$
acceleration $\gamma''(t) = (1, 0, 0)$
speed $||\gamma'(t)|| = \sqrt{t^2 + 6t + 14}$

(b) The unit normal and unit tangent are given by

$$T(t) = \frac{(t+3, 2, 1)}{\sqrt{(t+3)^2 + 2^2 + 1^2}}, \qquad N(t) = \frac{(5, -2(t+3), -(t+3))}{\sqrt{5} \cdot \sqrt{(t+3)^2 + 2^2 + 1^2}}.$$

The unit binormal is given by

$$B(t) = T(t) \times N(t) = \left(0, \frac{\sqrt{5}}{5}, \frac{-2\sqrt{5}}{5}\right).$$

Use WolframAlpha to do basic derivatives and linear algebra. Even so, calculating the binormal with these expressions may seem messy, but you can simplify your computations. Notice $T(t)$ and $N(t)$ are unit vectors which are respectively positive scalar multiples of $(t+3, 2, 1)$ and $(5\sqrt{5}, -2\sqrt{5}(t+3), -\sqrt{5}(t+3))$. The cross product $B = T \times N$ is therefore a unit vector that is a positive scalar multiple of $(t+3, 2, 1) \times (5\sqrt{5}, -2\sqrt{5}(t+3), -\sqrt{5}(t+3))$. To compute B, it is easier to calculate this cross product and then divide by its norm.

1.1.8

(a) Use what you learned in Exercise 1.1.4 to help you create the circular orbit. Choose reasonable units, a radius, and a period. Explain how you came up with specific values for the radius and period.

(b) Do the same as the above but be careful to explain whether the Moon rotates clockwise or counterclockwise about the Earth.

(c) You know the position of the Earth relative to the Sun. You know the position of the Moon relative to the Earth. These are two position vectors. How do you find the position of the Moon relative to the Sun? You've done all the hard work already so the answer here should be very short.

1.1.9 Your answer should be an upward spiral curve with domain $[0, 90]$. Your function should take into account the total elevation gained, the number of loops that the car had to make, the radius of the loop, and the speed of the car. This means introducing three unknown constants.

1.2.1

(a) One of the contours is a circle. One of them is a point. Two of them are empty.

(c) This should look like a cat hiding under a blanket.

1.2.2

(a) This curve corresponds to the y-slice at $y = 1$ given by $\{(x, z) \in \mathbb{R}^2 : z = x^2 - 1\}$. Note that the curve displayed in $\mathbb{R}^3$ itself is not a slice. Remember that slices are a way of reducing dimension, and so the slices of f are subsets of $\mathbb{R}^2$.

(b) This curve corresponds to the x-slice at $x = 0$ given by $\{(y, z) \in \mathbb{R}^2 : z = -y^2\}$. The remark for the previous part applies here as well.

1.2.3

(a) The level set of h at 1000.

(b) $h(B)$ is between 5000 and 6000.

(c) As you sketch the curve, pay attention to how it crosses each contour. It is not a straight line. Notice the distance between each contour.

(d) You should identify 4 possible local extrema. Recall that extremum is a local minimum or maximum.

(e) It's on the north side of the peak closest to E.

(f) The contour on E has two different labels: 5000 and 4000. This cannot occur by the definition of a level set.

1.2.4 Graphs I, II, III, IV match with contour plots D, C, A, B respectively. This problem is based on previous MAT237 online notes which are a good secondary resource. To distinguish between C and D note that in the case D sides of the rectangular level curves are parallel to the coordinate axis, so D corresponds to I.

1.2.5 The definitions of graphs, level sets, and slices exists for functions $f : \mathbb{R}^3 \to \mathbb{R}$ of three variables. Since it's similar to the two variable case, it's your job to create them.

(a) $\{(x, y, z, w) \in \mathbb{R}^4 : w = f(x, y, z)\}$

(b) A *level set of f corresponding to value $c \in \mathbb{R}$* is $\{(x, y, z) \in \mathbb{R}^3 \ : f(x, y, z) = c\}$

(c) An *x-slice of f corresponding to value $x = a$* is defined as $\{(y, z, w) \in \mathbb{R}^3 : w = f(a, y, z)\}$.

(d) None of these can be plotted in $\mathbb{R}^2$. All of them except the graph can be plotted in $\mathbb{R}^3$. Note the graph is in $\mathbb{R}^4$. You can try to plot the graph in $\mathbb{R}^3$ by including colours or intensity but is not always straightforward.

1.2.6

(a) $\{(x_1, \dots, x_n, w) \in \mathbb{R}^{n+1} : w = f(x_1, \dots, x_n)\}$

(b) A *level set of f corresponding to value $c \in \mathbb{R}$* is $\{(x_1, \dots, x_n) \in \mathbb{R}^n \ : f(x_1, \dots, x_n) = c\}$

(c) An *x_j-slice of f corresponding to value $x_j = a$* is defined as

$$\{(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n, w) \in \mathbb{R}^n : w = f(x_1, \dots, x_{j-1}, a, x_{j+1}, \dots, x_n)\}$$

.

(d) For $n = 4$, you can try to plot the level sets and slices in $\mathbb{R}^3$ by including colours or intensity for the real-valued output but again this is not always straightforward. Nothing else has a natural choice for plotting.

1.2.7 There are many examples for both parts.

(a) One possible choice is $f(x, y) = y + x$.

(b) One possible choice is $g(x, y) = y^2 - x^2$. To prove your example works, you will need to show two sets are equal. If you are having trouble with this proof, try to prove the previous question first. One way to prove two sets are equal is to show they are subsets of each other.

1.2.8

(a) The domain of φ should be measured with distances such as metres. That is, $(x,y,z) \in A$ where x, y and z are measured in metres. The codomain of φ should be measured in mass per distance cubed such as kg/m^3.

(b) If φ is constant then Aang's bones are as dense as his skin. He's not that soft.

(c) What do high values of φ correspond to? You may need to look up a bit about the human body to make an educated guess.

(d) Use a level set.

1.2.9

(a) Model 2 is quantum. High values of f are more intense and clear. Lower values are less intense and fuzzy.

(b) The volume of the region, and the intensity of the colour.

(c) What is the volume of a point in $\mathbb{R}^3$? You do not need to know any quantum mechanics beyond the principle that we shared.

1.3.1 A, B, C, D, E match with I, IV, III, II, V respectively.

1.3.2 The quantity $-\frac{(x,y,z)}{||(x,y,z)||}$ is a unit vector. Which way does it point? The quantity $\frac{Gm_1m_2}{||(x,y,z)||^2}$ is a positive scalar. How does it vary with the magnitude of $(x,y,z) \in \mathbb{R}^3$? Both of these hints tell you something about the picture. Also pay attention to where the function is not defined.

1.4.1

(a) The area is $|\det(A)|$ and the possible shapes are a parallelogram, a line, or a point.

(b) U and V are rectangles, $f(U)$ is a sector of a disk, and $f(V)$ is a sector of a disk which looks like a cut from a donut.

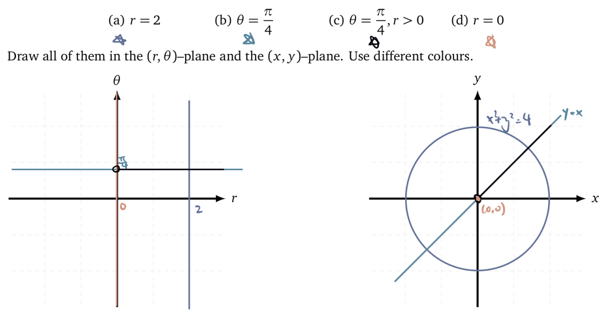
(c) The areas of U and V are the same but the area of $f(U)$ is smaller than $f(V)$. Use the area of a sector of a disk to find their values.

[0]Images retrieved from Wikimedia Commons (hydrogen eigenstate and atomic orbital cloud) on 2024-07-23 licensed under CC BY-SA.

1.4.2

(c) Are there any restrictions on θ in the (r, θ)–plane?

(d) Your picture should look vaguely like the drawing below.



(e) What does increasing your angle by $\pi/3$ look like graphically?

(f) This part is asking how you need to change your polar coordinates to scale the image by a factor of 2. Does that affect angles?

(g) Defining the inverse is tricky. Using arctan by itself isn't quite strong enough if we want the domain of our inverse to be all of $\mathbb{R}^2$. One solution is to divide the domain into three disjoint pieces: right and left half-planes and the vertical axis. We can then define the inverse function as a case function and use separate formulas for each of the three cases.

1.4.3

(b) You need to be careful because it doesn't make sense to rotate around a point in $\mathbb{R}^3$. What do you need to rotate around?

(d) Do the changes to r and z affect each other, or can you consider them separately?

(e) Part (a) is a cylinder in (x, y, z)-space, which has radius 2 and which is centered on the z-axis.
Part (b) is a half plane attached to the z-axis.
Part (c) is a solid cone with vertex at the origin and which opens in the positive z-direction.

1.4.4

(d) For part (a), the image in the (x, y, z)-space consists of all points between the two planes $y = x$ and $y = \sqrt{3}x$.

For part (b), the image in the (x, y, z)-space is a solid sphere of radius 2, with an open solid sphere of radius 1 removed.

For part (c), the image in the (x, y, z)-space is a cone. Note that this cone extends both above and below the z axis because we view $T : \mathbb{R}^3 \to \mathbb{R}^3$ as a map for all of $\mathbb{R}^3$. In particular, we allow ρ to be both positive and negative in the (ρ, θ, z)-space.

(e) How do longitude and latitude relate to ϕ and θ? Express longitude and latitude in terms of ϕ and θ.

(h) You will want to be careful with your angles here. In particular, note that (r, θ, ϕ) and $(r, \theta + \pi, -\phi)$ give the same points in (x, y, z)-space. You may want to exclude negative radii.

1.4.5

(a) You only need to draw a picture that satisfies the description. There isn't a unique way to draw it, but when you draw $f(A)$, it should be inside B. Try to make your drawings arbitrary.

(c) You should draw the sequence of points $f(x_1), f(x_2), \ldots$ in $f(A)$. Since f is continuous, what do you expect this sequence of points be approaching?

(d) Your picture should look vaguely like the drawing below. You should have one more label and the letter b will be replaced by $f(a)$.

1.5.1

(a) parametric

(b) implicit

(c) explicit

(d) none of them

1.5.2

(a) Yes, this parametrizes S. Use spherical coordinates.

(b) No, this does not parametrize S. The image of g_2 is a solid, not a surface.

(c) Yes, this parametrizes S. Does a parametrization need to be injective?

(d) Yes, this parametrizes S. Use cylindrical coordinates.

(e) Yes, this parametrizes S. Use cylindrical coordinates in a different way.

(f) No, this does not parametrize S. Can the domain be all of $\mathbb{R}^2$?

1.5.3 For each set, there are many valid choices. Only one example is provided, but your answer may look different. Remember the choice of letters for parameters do not matter.

(a) Define $f : [-1,1] \times [-1,1] \to \mathbb{R}^3$ by $f(x,y) = (x, y, \sqrt{x^2+y^2})$.

(b) Define $f : [0,2\pi] \times \mathbb{R} \to \mathbb{R}^3$ by $f(\theta, z) = (3\cos\theta, 3\sin\theta, z)$.

(c) Here are two ways:

- $f : [0,2] \times [0,2\pi] \to \mathbb{R}^3$ by $f(r,\theta) = (r^2, r\cos\theta, r\sin\theta)$
- $g : \{(y,z) \in \mathbb{R}^2 : y^2+z^2 \le 2\} \to \mathbb{R}^3$ by $g(y,z) = (y^2+z^2, y, z)$

(d) Here are two ways:

- $f : [0,2\pi] \times [-3,3] \to \mathbb{R}^3$ by $f(\theta, y) = (y\cos\theta, y, y\sin\theta)$
- $f : [0,2\pi] \times [-3,3] \to \mathbb{R}^3$ by $f(\theta, y) = (y\sin\theta, y, y\cos\theta)$

1.5.4 Note $A = D$ and $B = C$ as sets but they are each written in different forms.

(a) implicit form. Use, for example, $f(x,y,z) = xyz - 237$ to define A. Other valid choices of functions should only differ from f by a constant.

(b) parametric form. Use $g(y,z) = (\frac{237}{yz}, y, z)$ to define B.

(c) explicit form. Us $h(y,z) = \frac{237}{yz}$ to define C.

(d) none of them, not explicit since there are three equations.

1.5.5

(a) Use a map such as $h(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$.

(b) Solve for any of the variables, e.g. solve for z. This will involve a "$\pm$" sign which requires you to write the set as a union of sets in explicit form.

(c) You will need to refer to 6 different cases and why they all fail. See Example 1.5.10 for a similar case.

(d) If you have the correct map, then you it should be equal to the composition of two maps: one parametrizing the unit sphere, and a linear transformation dilating each coordinate.

(e) If you have the correct map, then you it should be equal to the composition of two maps: one parametrizing the unit sphere, and a linear transformation dilating each coordinate.

1.5.6

(a) True. $S = f^{-1}(\{0\}) = g^{-1}(\{9\})$ where $f(x,y,z) = x^2+y^2+z^2-9$ and $g(x,y,z) = x^2+y^2+z^2$.

(b) True. There are many valid choices. Here are two examples.

- Take $U = [0,2\pi]\times[0,\pi]$. Define $f : U \to \mathbb{R}^3$ by $f(\theta,\phi) = (3\sin\phi\cos\theta, 3\sin\phi\sin\theta, 3\cos\phi)$.
- Take $V = [0,2\pi]\times[-3,3]$. Define $g : V \to \mathbb{R}^3$ by $g(\theta,z) = (\sqrt{9-z^2}\cos\theta, \sqrt{9-z^2}\sin\theta, z)$.

Both $f(U)$ and $g(V)$ define the sphere of radius 3 centered at $(0,0,0)$.

(c) False. See Example 1.5.10 in the text for the beginning of the proof. While there are a total of 6 cases, you really only need to verify 2 cases because the rest are similar by symmetry.

1.5.7

(a) True. Your proof will need to define a function $g : \mathbb{R}^2 \to \mathbb{R}^3$ by $g(x,y) = (x,y,f(x,y))$. A complete proof will verify that $S = \operatorname{im}(g)$ as sets. This requires checking that each set is a subset of the other.

(b) True. Your proof will need to define a map such as $h : \mathbb{R}^3 \to \mathbb{R}$ by $h(x,y,z) = z - f(x,y)$. A complete proof will verify that $S = h^{-1}(\{0\})$ as sets. This requires checking that each set is a subset of the other.

1.5.8 (a) and (c) are true. You proved a special case in an earlier problem, and now you need to prove it for the general case.
(b) and (d) are false, and you can use the same counterexample for both.

1.5.9 This is a line which we think of as a “1-dimensional object" in linear algebra. This should not be considered a “2-dimensional manifold" in $\mathbb{R}^3$.

1.5.10 Remember a set in explicit form can always be re-written in parametric form or in implicit form. The answers below reflect how the set appears to be described.

(a) Parametric or explicit. The "dimension" appears to be 2.

(b) Implicit. The "dimension" appears to be $n-1$.

(c) Parametric or explicit. The "dimension" appears to be $n-1$.

(d) Parametric. The "dimension" appears to be n.

(e) Explicit. The "dimension" appears to be n.

(f) Implicit. The "dimension" appears to be n.

1.6.1 This retains the local directions (i.e. angles) and shapes, but it loses information about size.

1.6.2 Define $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ by $\pi(x,y,z) = (x,y)$. Then $\pi(S) = \{(x,y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$

1.6.3

(c) For $(x,y,z) \in \widetilde{S}$, $\pi(x,y,z) = \left(\frac{2x}{1-z}, \frac{2y}{1-z}\right)$.

1.6.4

(a) No. *Hint:* Its nullity cannot be too small.

(b) No. *Hint:* Its rank cannot be too big.

(c) For $(x_1,\dots,x_n) \in \mathbb{R}^n$, define $\pi(x_1,\dots,x_m,x_{m+1},\dots,x_n) = (x_1,\dots,x_m)$. The $m \times n$ matrix A looks like the $m \times m$ identity matrix glued to the $m \times (n-m)$ zero matrix.

A.2. Topology

2.1.1

(a) $p \in \overline{A}$

(b) $p \notin \overline{A}$

(c) $p \in A^o$

(d) $p \notin \partial A$

(e) $p \notin A^*$

(f) $p \notin A^o$

(g) $p \in A^*$

(h) $p \in \partial A$

2.1.2

(a)

The point a is a ...	■ interior point	☐ boundary point	■ limit point
The point b is a ...	☐ interior point	■ boundary point	■ limit point
The point c is a ...	☐ interior point	■ boundary point	■ limit point
The point d is a ...	☐ interior point	■ boundary point	☐ limit point

(b) Your drawings should look roughly like the picture below

S° $\quad$ ∂S $\quad$ $\overline{S}$

2.1.3

(a) None of them. Need to fix an arbitrary $\varepsilon > 0$ since the definition is a universal statement.

(b) $\varepsilon = 0.4$ only.

(c) $\varepsilon = 0.4$ only.

(d) None of them.

(e) None of them.

(f) $\varepsilon = 0.4$ only.

2.1.4

(a) $S^{\circ} = (137, 237)$ $\quad$ $\partial S = \{137, 237\}$ $\quad$ $S^* = [137, 237]$ $\quad$ $\overline{S} = [137, 237]$

(b) $S^{\circ} = B_r(a)$ $\quad$ $\partial S = \{x \in \mathbb{R}^n : \|x - a\| = r\}$ $\quad$ $S^* = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$
$\overline{S} = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$

(c) $S^{\circ} = B_r(a) \setminus \{a\}$ $\quad$ $\partial S = \{x \in \mathbb{R}^n : \|x - a\| = r\} \cup \{a\}$ $\quad$ $S^* = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$
$\overline{S} = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$

(d) $S^{\circ} = \mathbb{R}^n$ $\quad$ $\partial S = \varnothing$ $\quad$ $S^* = \mathbb{R}^n$ $\quad$ $\overline{S} = \mathbb{R}^n$

(e) $S^{\circ} = \varnothing$ $\quad$ $\partial S = \mathbb{Z}^n$ $\quad$ $S^* = \varnothing$ $\quad$ $\overline{S} = \mathbb{Z}^n$

(f) $S^{\circ} = \varnothing$ $\quad$ $\partial S = \mathbb{R}^n$ $\quad$ $S^* = \mathbb{R}^n$ $\quad$ $\overline{S} = \mathbb{R}^n$

2.1.5

(a) False. Consider boundary points.

(b) True

(c) True

(d) False. Consider finite sets.
Other counter-examples and justifications to all parts can be found in this section's examples and exercises.

2.1.6
(a) $\exists \varepsilon > 0$ s.t. $\forall (x, y) \in \mathbb{R}^2, ||(x, y) - (2, 0)|| < \varepsilon \implies (x, y) \in S$

(c) Focus on the structure of your proof. Did you choose an ε? Did you justify each step? Did you prove the formal open ball definition you wrote in a previous part?

2.1.7
(a) Line 5. The definition of boundary points says **every** open ball need to intersect S and S^c nontrivially.

(b) Focus on the structure of the proof. Can you fix ε here? Did you justify each step? Did you prove the formal definition of a boundary point?

2.1.9
(a) Both inclusions follow straight from the definition of closure and interior.

(b) Show that if $x \in A^o$ then $x \notin \partial A$ or vice versa.

2.1.10 Sketch a picture of two intersecting disks in $\mathbb{R}^2$ to convince yourself. This should also inspire a picture proof of both facts. Use formal definition of being an interior point of a set. Remember when proving set equality that you must show both subset inclusions.

2.1.12
(a) There are two cases for any point p, $p \in A$ or $p \notin A$. The author says *"it suffices to assume"* indicating they eliminated one of the cases. This phrase also functions to tell the reader what the author wants to show (ie a more elegant way of saying "WTS").

(b) Use line 3 and the definition of boundary points.

(c) Notice that line 4 shows that p is a limit point of A, while the set of limit points is a subset of closure.

2.1.12
(a) $\exists \varepsilon > 0$ such that $B_\varepsilon(p) \backslash \{p\} \cap S = \varnothing$

(b) C and D.

(c) Write down the set equality that you are proving and use formal definitions.

2.2.1 All these statements are equivalent.

2.2.2 $\forall p \in \mathbb{R}^n, \exists \varepsilon > 0$ s.t. $\forall K \in \mathbb{N}^+, \exists k \in \mathbb{N}$s.t.$k \geq K$ and $||x(k) - p|| \geq \varepsilon$

2.2.3 (a), (b), and (e) are equivalent

2.2.4

(a) The lemma additionally proves that $\lim_{k\to\infty} x(k) = \left(\lim_{n\to\infty} x_1(k), \ldots, \lim_{k\to\infty} x_n(k)\right)$ if the sequence $\{x(k)\}_k$ converges.

(b) It converges to $(1,0)$ by Lemma A and MAT137 (specifically it was shown using continuity in MAT137).

(c) You may assume without proof that the sequence $\{\cos k\}_k$ in $\mathbb{R}$ diverges. This is actually quite tricky to prove but we will take it for granted since it is from single-variable calculus.

2.2.5

(a) Use the first definition. Consider the sequence $\left\{(1-\frac{1}{n}, 1-\frac{1}{n})\right\}_{n=4}^{\infty}$.

(b) Negate the second definition. Consider the open ball $B_{1/2}((0,0))$.

2.2.6

(a) Many sequences can work, for example $\{(1/n, 0)\}_n$.

(b) Take, for example, the constant zero sequence (i.e $x_n = (0,0)$ for all $n \in \mathbb{N}$) in A and $\{(1/n,0)\}_n \subset A^c$.

(c) Looking at different quantifiers in the definitions may help.

2.2.7

(a) True. The proof is fairly short, but you will need to unpack several definitions. Carefully write the formal definition of convergence with quantifiers, as well as what you want to show. Then you can outline your proof structure before starting to fill in the ideas. Your proof must use that a subsequence $x(m(k))$ is defined with an *increasing* function $m : \mathbb{N}^+ \to \mathbb{N}^+$. This implies that $m(k) \to \infty$ as $k \to \infty$. In other words, for any $K \in \mathbb{N}^+$, there exists $K' \in \mathbb{N}^+$ such that [...].

(b) False. Many divergent sequences fail to satisfy this property. Consider the oscillating sequences.

2.2.8

(a) $\forall i \in \{1,\ldots,n\}, \forall \varepsilon > 0, \exists K \in \mathbb{N}^+$ s.t. $\forall k \in \mathbb{N}^+, (k \geq K \implies |x_i(k) - a_i| < \varepsilon)$.

(b) Everything from $\forall \varepsilon > 0$ onwards.

(c) This refers to the universal quantifiers $\forall i$ and $\forall \varepsilon$ in the WTS.

(d) Without already knowing the proof, it would be nearly impossible to understand why the author is doing each step. These additional phrases show the reader what to expect as the next step in the argument.

2.2.9 Sentence I corresponds to Line 4. Sentence II roughly corresponds to Lines 5 and 6.

2.2.10

(a) Since p is a limit point of A, for each $n \in \mathbb{N}^+$, pick a point in $A \setminus \{p\}$ from opens balls centered at p with radius $1/n$. This sequence of points must converge to p since the radii of the balls are shrinking to zero.

(b) Since p has a sequence in $A \setminus \{p\}$ converging to it, an open ball centered at p must eventually have a sequence of points in $A \setminus \{p\}$ enter the ball.

2.3.1 A is open. B is closed. C and D are neither open nor closed.

2.3.2

(a) Open but not closed

(b) Both(Clopen)

(c) Both(Clopen). There are exactly two clopen subsets of $\mathbb{R}^n$.

(d) Open but not closed. Recall that unions of open sets are open.

(e) Closed but not open

(f) Closed but not open. There are many ways to prove it. For example, show $\mathbb{Z}, \mathbb{R}$ are closed or $(\mathbb{Z} \times \mathbb{R})^c$ is open.

(g) Neither. Draw this rectangle and look at the edges.

(h) Closed but not open

2.3.3

(a) Open

(b) Closed

(c) Closed

(d) Neither

(e) Open

(f) Neither

(g) Closed

(h) Neither

(i) Open

(j) Neither

(k) Open

(l) Neither

2.3.4

(c) For any (x, y) with $x < 0$, you need to find a ball $B_\varepsilon((x, y))$ contained in the left half plane. This is where the picture comes into play.

To prove that your ε works, it will be helpful to remember that for any two vectors $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$, that $|v_i - w_i| \le \|v - w\|$.

2.3.5

(a) The options are:

- Show that P^c is open.
- Show that P contains all its limit points using the open ball definition of a limit point.
- Show that P contains all its limit points using the sequence definition of a limit point

(b) Suppose you have a sequence $\{a_n\}_1^\infty$ in P. What does being in P tell you about each a_n? When you take the limit, remember that you should take it coordinate by coordinate.

2.3.6

(a) A set B is open if for every point $a \in B$ there exists $\varepsilon > 0$ such that $B_\varepsilon(a) \subseteq B$.

(b) A set B is closed if for every limit point of B belongs to B.

2.3.7

(a) Notice that a point in the union is contained in one of the two sets. What does the openness of that set give you?

(b) Take the smaller of the two radii.

(c) Same idea as 5.1: it's in one of the open sets.

(d) Consider the sets $B_{\frac{1}{n}}(0)$. These are all open. What's their intersection? Why doesn't it make sense to "take the smallest radius" in this case?

2.3.8

(a) Recall the sequential definition of a limit point.

(b) There are two options here. Either $x(k)$ is in A or it must be a limit point of A. How do we choose $y(k)$ in both of these cases?

(c) Recall the definition of the closure of A. We must show that either p is in A or p is a limit point of A.

2.4.1

(a) True

(b) False. Notice the set is unbounded.

(c) False. Notice the set is not closed

(d) True

(e) False. Notice the set is not closed

(f) True

2.4.2

(a) False. This is equivalent to the definition of closed.

(b) False. This is equivalent to the definition of bounded.

(c) False. No set satisfies this statement.

(d) False. This is only true for singletons $\{a\}$ and the empty set

(e) True by definition

(f) True by Bolzano-Weierstrass

(g) False, S can be neither open or closed

(h) False. Counterexample $\mathbb{Z} \times \mathbb{R}$

Only two of these are equivalent to compactness. Amongst the six which are not equivalent to compactness, you will find:

- a statement which is never true for any set
- a statement which is only true for singletons $\{a\}$ and the empty set
- the definition of closed
- the definition of bounded

2.4.3
(a) False. There are many counterexamples.
(b) True
(c) False
(d) False
(e) True
(f) True. The proof of this statement is a good exercise.
(g) True. The Cartesian product of compact sets is compact.

2.4.4
(a) There is a particularly simple sequence you can choose here.
(b) Compactness will be useful here.

2.4.5 Prove this directly from the definition of subsequences and convergence. Monotonicity will come in handy.

2.4.6 One of these sets is easily expressed as a union of two compact sets and the other is easily expressed as an intersection of a compact set and a closed set. Think carefully about which is which and pay attention to AND and OR.

2.4.7
(a) Subsets of bounded subsets are bounded. Can you see why?
(b) You will need to apply Bolzano-Weierstrass twice for this proof.

2.4.8
(a) Which statement is more general: there exists a sequence lying in A converging to p or there exists a sequence lying in $A \setminus \{p\}$ converging to p?
(b) What does the definition of compactness tell us about subsequences of convergent sequences?
(c) Go back to the first textbook definition of closed sets.

2.4.9
(a) Suppose Line 2 is not true. Can you show that Line 1 is also not true? (this is a contrapositive argument within a larger contrapositive proof).
(b) Monotonicty is important here.
(c) Try to negate the definiton of compactness.

2.5.1
(a) $A^* = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\}$ and $A \setminus A^* = \varnothing$ $S' = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\}$.
(b) $B^* = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\}$ and $B \setminus B^* = \{(0,0)\}$ $S' = \{(x,y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\}$.
(c) $C^* = \mathbb{R}^2$ and $C \setminus C^* = \varnothing$ $S' = \mathbb{R}^2$
(d) $D^* = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 2\}$ and $D \setminus D^* = \varnothing$ $S' = \{(x,y) \in \mathbb{R}^2 : ||x - a|| \le \varepsilon\}$
(e) $E^* = \mathbb{R}^2$ and $E \setminus E^* = \varnothing$ $S' = \mathbb{R}^2$
(f) $F^* = \varnothing$ and $F \setminus F^* = \mathbb{Z} \times \mathbb{Z}$. $S' = \varnothing$

2.5.2 Only part (c) and part (f) are equivalent, others are either inequivalent or formulated incorrectly. Double check the textbook definition of the limit if we test against the value of the function $x = a$. Also, note that unless x is an element of the domain, the expression $f(x)$ doesn't make sense (check $x \in A$).

2.5.3

(a) What are $f_1(t)$ and $f_2(t)$? Are those functions continuous?

(b) Yes, but it's not useful (yet). You can apply the above theorem but, without additional theorems, you cannot evaluate the resulting limits.

2.5.4

(a) Aloy is closer to the truth. It does not follow by definition, but you need a lemma. There are many possible versions. One such lemma could be:

> **Lemma.** *Let $F : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^2$. If F is continuous at $g(0)$ and g is continuous at 0 then $\lim_{t\to 0} F(g(t)) = F(g(0))$.*

Notice this requires the definition of continuity which has not yet been revealed. Even if you attempt to the correct the proof by this lemma, you must proceed by contrapositive and assume the limit exists and equals, say $L \in \mathbb{R}$. Define $F(0,0) = L$ and $F(x,y) = f(x,y)$ for $(x,y) \neq (0,0)$, and use the above lemma with two different choices of g. This strategy will succeed but it is rather annoying to write rigorously. The sequential method is much easier to apply.

(b) The incomplete proof above can still give you a hint at what sequences to use.

2.5.5

(a) $f(2,y) = 2y$

(b) One mini-idea is the triangle inequality. The other mini-idea is adding zero.

(c) Line 2 can be informally described as saying $x \approx 2$ and $y \approx 3$. If $x \approx 2$, then what can you say about $f(x,y)$? And if $y \approx 3$, then what can you say about $f(2,y)$?

2.5.6 You will have to express $3x - 5y$ in such a way that it can be approximated by $|x-1|$ and $|y-2|$. Hint: $3x - 5y + 7 = 3(x-1) - 5(y-2)$. The triangle inequality will be useful.

2.5.7

(a) At what point in the example, was δ defined? We need to define δ at the beginning of the proof.

(b) Remember to define your variables in the correct order. Your proof should start with something like "Let $\varepsilon > 0$ be fixed." Then define δ. Then fix $(x,y) \in \mathbb{R}^2$. Assume (x,y) satisfies an inequality. If you do not do these steps, then your proof is missing something.

2.5.8 Use the fact that the function $f(t) = 2^t$ is continuous. Use an intermediate approximation inspired by the idea of "fixing a variable".

2.5.9 Use the fact that cosine is continuous

2.5.10 The triangle inequality is pivotal.

2.5.11 Modify this proof of the 1D squeeze theorem.

2.5.12

(a) Your proof should start with something like "Fix $\varepsilon > 0$. Then define M.

(b) $\lim_{k\to\infty} f(x(k)) = b$ for every sequence $\{x(k)\}_k \subseteq A$ satisfying $||x(k)|| \to \infty$ as $k \to \infty$.

(c) What happens along the line $x = 1$?

2.5.13

(a) Consider the lines $x = 0$ and $x = y$. Your final proof should involve sequences.

(b) Use the fact that $|x| \leq \sqrt{x^2+y^2}$ and $|y| \leq \sqrt{x^2+y^2}$ in combination with the squeeze theorem to get the result you need.

2.5.14 In both directions, you need to ensure f is always defined for the possible choices of x. In one direction, take $\delta_2 = \min\{\delta_1, \delta_0\}$ where δ_0 is well-chosen. In the other direction, take $\delta_1 = \delta_2$ since f will automatically be defined.

2.5.15 A good choice of δ depends only on a and A, and does not depend on f or b.

2.5.16 One direction is a straightforward application of epsilon-delta definitions. In the other direction the contrapositive will be useful.

2.6.1

(a) False. What if x is not a limit point of the domain A? Would the expression on the l.h.s. make sense in this case?

(b) Won't be true even for the one-dimensional case without extra assumptions on $f(x)$.

(c) True. Can you estimate $||f(x) - f(a)||$ using the determinant of the matrix M and the norm of vector x?

(d) True. Can we use the fact that each of the polynomials is continuous function to construct an ε–δ–proof?

2.6.2 Start by identifying the natural domain of each function. Are they continuous at every point of their domain?

2.6.3

(a) A, B, D, F are equivalent to continuity for limit points.

(b) C, D, F are equivalent to continuity for isolated points. Is it always true that function is continuous at an isolated point of its domain?

2.6.4 Only part (c) and part (e) are true. A constant function or an exponential function provide counterexamples to the others.

2.6.5 Some of the key facts you will need include: polynomials are continuous; compositions of continuous maps are continuous; trigonometric functions are continuous; a function is continuous if and only if every component function is continuous; and the function $F(x) = 1/x$ is continuous on $\mathbb{R} \setminus \{0\}$.

(a) $\mathbb{R}^2$

(b) $\mathbb{R}^n$

(c) $\{(x, y, z) \in \mathbb{R}^3 : z \neq 1\}$

(d) $\{(x, y) \in \mathbb{R}^2 : xy \neq -3\}$

(e) $\mathbb{R}^3$

2.6.6

(a) Any line through the origin are defined by $ax + by = 0$ for some $a, b \in \mathbb{R}$.

(b) Consider what the function does along parabolas.

2.6.7

(a) Try applying a theorem on openness using the continuity of $f(x, y) = y^2 - x^5 + xy$.

(b) Consider a function $f(w, x, y, z) = x^2 - y^3 - 237z^2$, can you present $S = f^{-1}(C)$ for some closed set $C \subset \mathbb{R}$?

(c) What is the preimage of $(\frac{1}{2}, \frac{3}{2})$?

2.6.8

(a) Try applying a theorem on compactness using the continuity of $f(x_1, x_2) = (x_1^2 + x_2^2, x_1^2 - x_2^2)$.

(b) Try spherical coordinates.

(c) Use that the interval $[0, 1]$ is compact.

2.6.9 Use continuity of $\frac{1}{x}$.

2.6.10 Use continuity of e^x.

2.6.11 See Example 2.5.4 for a similar argument. Note that $xy - ab = xy - xb + xb - ab$ for $x, y, a, b \in \mathbb{R}$. Remember you must have a well-written proof. Start by fixing a point $(a, b) \in \mathbb{R}^2$. Fix arbitrary $\varepsilon > 0$. Choose $\delta = [\ldots] > 0$ depending on ε, a, and b. Fix $(x, y) \in \mathbb{R}^2$ and assume $||(x, y) - (a, b)|| < \delta$. If you do not do this, then your proof is missing something.

2.6.12

(a) The theorem only applies when f, g and their product function have the same domain.

(b) Notice that the theorem produces new functions that have the same domain as the functions you start with. Thus you need to start with functions whose domain is $\mathbb{R}^2$.

(c) Induction may be helpful.

2.6.13 Your open ball proof will use the triangle inequality; the limit law proof should contain a discussion of isolated points; the proof with sequences will use Theorem 2.6.5.

2.6.14 Lines 2, 3 and 6 are referring to the location of lumpy regions. Try to figure out which is which.

2.6.15 If for some $L \in \mathbb{R}$, $g(x,y) \to L$ as $(x,y) \to (0,0)$, then define $g(0,0) = L$ to make it continuous. Now, obtain a contradiction.

2.6.16 Express the level set as a preimage. Then use Theorem 2.6.27.

2.6.17 Write the graph in parametric form. Then use Theorem 2.6.35.

2.6.18 Fix $j,k \in \mathbb{N}$. First prove that the monomial map $M_{j,k} : \mathbb{R}^2 \to \mathbb{R}$ given by $M_{j,k}(x,y) = x^j y^k$ is continuous by writing it as a composition of maps. Once you do that, remember every polynomial is a finite linear combination of monomials.

2.6.19 Use contrapositive. Consider the preimage of the sets $\{0\}$ and $\{1\}$.

2.6.20
(a) We do not know if f^{-1} exists on U. f^{-1} can be empty.
(b) What set is $f(a)$ in? What are its properties?
(c) This uses the continuity of f.
(d) Consider the relationship between preimages and images. For any $a \in f^{-1}(U), \exists \delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(U)$.
(e) Your picture should contain: U, $f^{-1}(U)$, two balls, and the image of a ball.

2.6.21 Try setting U to be an open ball.

2.6.22
(a) Try to write down both of these sets in set builder notation.
(b) Use Section 2.6.3 and the previous part.

2.7.1 A few possible examples are listed. All of them are easy to prove by the definition. Let $r > 0$ and $x \in \mathbb{R}^m$. Let $a_1, b_1, \ldots, a_m, b_m \in \mathbb{R}$ with $a_j < b_j$ for $j = 1, \ldots, m$.
(a) $\varnothing, \mathbb{R}^m, B_r(x), (a_1, b_1) \times \ldots \times (a_m, b_m)$
For the last example, any interval (a_j, b_j) could be replaced with $(-\infty, b_j)$ or (a_j, ∞)
(b) $\varnothing, \mathbb{R}^m, \overline{B_r(x)}$, $\{x\}$, any finite set of points, $[a_1, b_1] \times \ldots \times [a_m, b_m]$
For the last example, any interval $[a_j, b_j]$ could be replaced with $(-\infty, b_j]$ or $[a_j, \infty)$
(c) $\varnothing, \overline{B_r(x)}$, $\{x\}$, any finite set of points, $[a_1, b_1] \times \ldots \times [a_m, b_m]$
(d) $\varnothing, \overline{B_r(x)}$, $\{x\}$, any line, plane, graph of a continuous function, $[a_1, b_1] \times \ldots \times [a_m, b_m]$

2.7.2 A is path-connected except when $n = 1$. C is path-connected. B and D are not path-connected.

2.7.3 Intermediate value theorem will be useful. Read Example 2.7.6 for proof ideas.

2.7.4 Consider the function $F : [-5,5]^2 \to [-5,5]^2 \times \mathbb{R}$ defined by $F(x,y) = (x, y, f(x,y))$. Use Theorem 2.7.8.

2.7.5 First, do it directly from the definition of path connected sets. Second, express it as the image of a path-connected set under a continuous function. Polar coordinates will be helpful for both approaches.

2.7.6
(a) $a, b, f(a), f(b)$.
(b) At which line does the proof introduce γ?
(c) The function in question is a composition of two functions. Apply a corollary on its key property.
(d) What is $f(c)$? Replace c by a quantity.

2.8.1 The EVT applies to part (a). For the two examples where it does not apply, we cannot conclude anything so we do not know whether f has a maximum or a minimum.

2.8.2 The EVT does not apply directly to any of the given sets. There's lots of valid restrictions but it is better to choose non-trivial ones.
(a) Try a smaller ball. The closed ball $\overline{B_1(0)}$ is not a valid choice for two reasons.
(b) Try restricting the y-coordinate. The set $\{(x, y) \in \mathbb{R}^2 : 0 \leq x^2 - y^2 \leq 1\}$ is not a valid choice.
(c) Take any "large" compact set $K \subseteq \mathbb{R}^2$. However, you can actually use the EVT to prove h has a maximum on $\mathbb{R}^2$ by using K. Notice $h(x, y) \to 0$ as $|y| \to \infty$. What does this say about h inside $\mathbb{R}^2$ but outside of K? Inside K, you know the maximum of h is bounded below by what? Using these ideas, you can show h has a maximum on $\mathbb{R}^2$ but does h have a minimum on $\mathbb{R}^2$?

2.8.3
(a) The theorem shows us that $f(A)$ is compact while the lemma tells us that $f(A)$ admits a maximum and a minimum value. It takes only a couple of lines to conclude the proof.
(b) Follow the proof argument in EVT. Construct a sequence that converges to the supremum, and a sequence that converges to the infimum.

2.8.4 Notice f has a maximum if and only if $-f$ has a minimum. Your proof should be very short.

2.8.5 Use the formal definition of the limit diverging to ∞ to break $\mathbb{R}^n$ into two sections: a compact set and a non-compact set. What do we know about f on each of these sections?

2.8.6
(a) Line 2 uses the continuous image of compact sets is compact. Line 3 uses the Bolzano-Weierstrass theorem. Line 4 uses the least upper bound principle.
(b) Your argument should also include the conclusion from Line 4.
(c) Line 8 uses Lines 6, 7, and the squeeze theorem over $\mathbb{R}$. Line 9 requires Line 8 and the sequential definition of a limit point. Your justification should include all of these aspects; otherwise, you are missing something.

2.8.7
(a) Yes, it does and $f(A)$ looks roughly like the closed interval $[1.30, 2.00]$.
(b) No, it does not. What if A is a union of two disjoint rectangles in $\mathbb{R}^2$?
(c) Yes it does. It appears that $x_k \to a$ as $k \to \infty$.

(d) No, it does not. What if the two peaks in the picture were at the same height?

2.8.8

(a) What needs to happen to f far away from the origin?

(b) Similar to the previous problem.

(c) You must consider the behaviour of f far away from the origin and at the boundary and the existence of that limit. Note that there was a typo in the original question.

(d) What needs to happen to f near the origin?

A.3. Derivatives

3.1.1

(a)
 i) Derivative of f at a, an $m \times 1$ column vector
 ii) Derivative of the 1st coordinate function of f at a, a scalar
 iii) Differential of f at a evaluated at h, an $m \times 1$ column vector
 iv) Differential of f at a, a linear map $df_a : \mathbb{R} \to \mathbb{R}^m$
 v) Derivative of f, a (not necessarily linear) map $f' : A \to \mathbb{R}^m$ defined on $A = \{a \in \mathbb{R} : f \text{ is differentiable at } a\}$

(b) The first equality is a definition and the second follows from a theorem.

3.1.2

(a) Equivalent

(b) Equivalent

(c) Not equivalent. Compare with part (b).

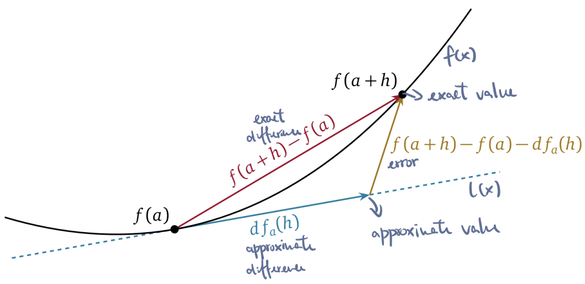
(d) Not equivalent. Note $|\cdot|$ is the absolute value function on $\mathbb{R}$ whereas $||\cdot||$ is the norm on $\mathbb{R}^n$.

(e) Equivalent

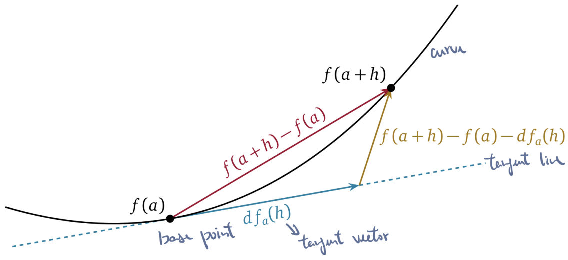
3.1.3

(a) There exists a linear map $L : \mathbb{R} \to \mathbb{R}^m$ such that $\lim\limits_{h\to 0} \dfrac{f(a+h)-f(a)-L(h)}{h} = 0$. The phrase "first order approximation" means that the difference between $f(a+h)-f(a)$ and $L(h)$ tends to zero faster than $h = h^1$ tends to zero.

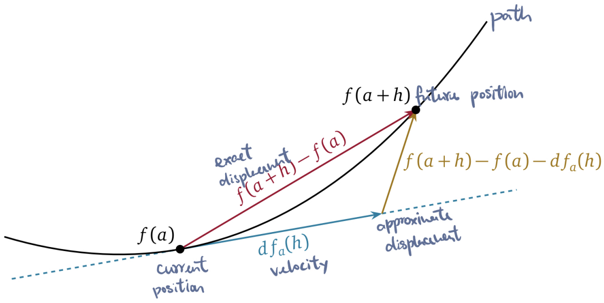
(b) Analytic



(c) Geometric



(d) Physical



3.1.4 These quantities sound very similar and indeed they are related. Two of them are identical answers but the rest are all technically different.

(a) $(1, 2\pi, 1)$

(b) $(1, 2\pi, 1)$

(c) This is a unit vector. Don't forget to adjust the magnitude.

(d) This is a linear map.

(e) This is an affine map. There are two ways to write it:

$$\begin{aligned} \forall h \in \mathbb{R}, \qquad & \ell_1(h) = (2\pi, 0, 2\pi) + (1, 2\pi, 1)h \\ \forall x \in \mathbb{R}, \qquad & \ell_2(x) = (2\pi, 0, 2\pi) + (1, 2\pi, 1)(x - a) \end{aligned}$$

Note $\ell_1(h)$ is the linear approximation to $f(a+h)$ at $h = 0$ and $\ell_2(x)$ is the linear approximation to $f(x)$ at $x = a$. Either choice is acceptable as long as you are clear.

(f) This is a line in $\mathbb{R}^n$ which contains the point $f(a)$ and has directional vector $f'(a)$.

3.1.5 Apply the theorem on limits component-by-component (Theorem 2.5.11) and the definition of differentiability for a parametric curve (Definition 3.1.1). Your proof should be quite short.

3.1.6 It will help to write down the definition of a function being differentiable at a point and use the first part of this question.

3.1.7 This proof is short but it requires careful unpacking of definitions. You must use the definition of differentials exactly as stated. You must choose the map $d(f+g)_a$ by its uniqueness. If you do not use the sum limit law, then you are missing something.

3.1.8

(a) Expand $g(t) \cdot g(t)$ into components and differentiate using single variable calculus. Use sigma notation. You will need to cite Lemma 3.1.3. Then differentiate $g(t) \cdot g(t)$ in another way; remember to use your assumption about $||g(t)||$.

(b) Write $T(t) \cdot N(t)$ and insert the definition of $N(t)$. Apply the previous part. As an aside, note the assumption that $T(t)$ exists for all $t \in I$ implies that $||\gamma'(t)||$ is never 0. This implies that T is differentiable. Moreover, the assumption that $N(t)$ exists for all $t \in I$ implies that $||T'(t)||$ is never 0.

3.2.1

(a)
(a) $\partial_j F(a)$ is a vector in $\mathbb{R}^m$.
(b) $\partial_j F$ is a vector-valued function with the domain $A°$ and codomain $\mathbb{R}^m$.
(c) $\partial_j F_i(a)$ is a scalar (element of $\mathbb{R}$).
(d) $\partial_j F_i$ is a real-valued function with the domain $A°$.

(b)
- The first equality holds by Definition 3.2.2.
- The second equality holds by Lemma 3.2.7.
- The third equality is an equivalent notation.

3.2.2 Part (b) and part (d) cannot be determined from the given graphs. Part (a) is zero and (c) is negative.

3.2.3 Of these four values, only $g_x(P)$ is positive and the other three are negative.

3.2.4

(a) False

(b) True

(c) False, think about functions involving the absolute value of one of the variables

(d) True

(e) True

(f) False

3.2.5

(a) One way should give the value of 12. Another way gives 16.

(b) One way should give the value of 2.4. Another way gives 2.6.

(c) Using the approximations $h_x(1,10) \approx 12$ and $h_y(1,10) \approx 2.4$, you should get $h(0.8,12) \approx 75.4$.

3.2.6 Two of these are zero and the other is two.

3.2.7 One of these is $(-1,0)$ and the other is $(0,-2)$.

3.2.8 The partial derivatives are $y^2z^3 \arctan(xyz) + \frac{xy^3z^4}{1+x^2y^2z^2}$, $2xyz^3 \arctan(xyz) + \frac{x^2y^2z^4}{1+x^2y^2z^2}$, and $3xy^2z^2 \arctan(xyz) + \frac{x^2y^3z^3}{1+x^2y^2z^2}$ respectively.

3.2.9 These should be constant functions.

3.2.10 $\partial_j f(x) = 2x_j$

3.2.11 $\partial_j g(x) = -\frac{x_j - a_j}{\|x-a\|^3}$

3.2.12 Write $A = \left[a_{ij}\right]_{i,j} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ to compute the partials. These partials should all be constant functions.

3.2.13

(a) Use the definition of the partial derivative and the linearity property of limits.

(b) There are two methods. One method is to follow the same argument as the previous part. The second method is to write $F + \lambda G$ in components, use Lemma 3.2.7, and the first part of this problem.

3.2.14

(a) There are many valid sentences, but they should not involve any technical mathematical terminology, e.g. "instantaneous rate of change" or "$y = 10$" or "the point $(10, 10)$". It should be a sentence that another person (who does not know any calculus) can understand. One possible description is below.

> *This quantity represents the approximate change in temperature in degrees Celsius per meter that you move in the vertical direction starting from 10 metres above, while always staying 10 metres to the right of the left side.*

You can use the word "north" and "east" instead of "vertical" and "right", because (without further specification) these are the default directions.

(b) Positive.

(c) $T_y(10, 10) \approx 0.2$

(d) Recall that $f(a, b+h) \approx f(a, b) + f_y(a, b)h$ which can be used to show that $T(10, 10.1) \approx 20.02$.

3.2.15

(a) Consider a freeze frame at the time t_0 (measured in seconds) and choose a base point on the string located x_0 meters from the origin in the positive direction of the x-axis. Quantity $u_x(x_0, t_0)$ represents the instanteneous rate of change of the displacement (also known as slope) of the string in meters per each meter walked in the positive direction along the x-axis starting from the base point.

(b) Consider a motion of a distinguished point of the string located x_0 meters away from the origin in the positive direction of the x-axis. Quantity $u_t(x_0, t_0)$ represents vertical component of the instanteneous velocity measured in meters per second of the distinguished point at the time t_0 seconds.

(c) Recall from the part 1 of the problem that $u_x(x_0, t_0)$ represents the slope. Identify the t_0-freeze frame first by choosing a picture of the string captured in appropriate color, then look at the slope of the curve at x_0 meters from the origin.

[0] Image created by Cindy Blois with permission.
[0] Image created by Cindy Blois with permission.

Similarly, recall from part 2 of the problem that $u_t(x_0, t_0)$ represent the instanteneous vertical component of the velocity of a distinguished point of the string. Identify the x-coordinate of a distinguished point on the string first. From the several freeze frames given you can determine the y-coordinate of the distinguished point at $t = 0, 0.5, 1, 1.5, 2, 3, 3.5$ seconds. Mark three positions of the distinguished point at the times $t_0 - 0.5, t_0$, and $t_0 + 0.5$. Based on these markings can you conclude what is the most plausible answer for the vertical component of the velocity of the distinguished point at the time t_0?

3.2.16 This is very similar to problem A.3.17.4. How much will the number of applicants change as the job increases the wage or the distance of its site from the city center?

3.2.17

(a) Repeat Definition 3.2.2.

(b) The exact value and approximate value are the two points you just labeled. The error is the length of a certain line segment.

(c) Your plot in $\mathbb{R}^2$ should contain a graph of the function $g(t) = f(a_1 \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$ and the tangent line to this graph at the point $t = a_j$.

(d) The answer should be a change in the revenue resulting from some change in price. Be specific about what change in price.

3.3.1

(a)
 i) $m \times 1$ column vector. This is a directional derivative at point a corresponding to directional vector v.
 ii) Vector valued function $D_v F : A^\circ \to \mathbb{R}^m$. For each point $a \in A^\circ$ it returns directional derivative corresponding to directional vector v at this point.
 iii) Scalar. It is an i'th component of part i).
 iv) Real valued function $D_v F_i : A^\circ \to \mathbb{R}$. For each point $a \in A^\circ$ it returns a scalar from part iii).

(b) First equality is Definition 3.3.1, second equality is Theorem 3.3.10. Last equality is a Lemma 3.3.5.

3.3.2

(a) Positive, $f(-2,-1) = g(0) > 0$

(b) Cannot determine. Note that $g(h)$ plotted on the graph contains information only about values of f on a line $\{a + hv : h \in \mathbb{R}\} \subset \mathbb{R}^2$. On the contrary, the definition of $\frac{\partial f}{\partial x}(-2,-1)$ involves the values of f on the horizontal line passing through $a = (-2,-1)$.

(c) Cannot determine. Same idea as in the previous part.

(d) Negative. Note that this object, in nothing but the derivative of $g(h)$ at $h = 0$.

3.3.3

(b) One is a nonlinear map and the other is an output of the map.

(c) These are nonlinear maps representing the ith partial derivative of f.

3.3.4

(a) True

(b) True

(c) False

3.3.5 $D_w f(a)$ is positive and $D_v f(a)$ is zero.

3.3.6

(a) Negative.

(b) If you assume the point $(7,7)$ satisfies $T(7,7) = 19$, then $D_{(-1,-1)}T(10,10) \approx -0.33$. If you assume the point $(13,13)$ satisfies $T(13,13) = 21$, then again $D_{(-1,-1)}T(10,10) \approx -0.33$. If you use slightly different points, then you will get slightly different values. Your point must appear to be on a nearby contour and it must be of the form (t,t) for some $t \in \mathbb{R}$.

(c) $T(9.5, 9.5) \approx 19.833$.

3.3.7

(a) $\dfrac{\partial f}{\partial x} = (2x, ye^{xy} + y)$ and $\dfrac{\partial f}{\partial y} = (-2y, xe^{xy} + x)$

(b) $D_v f = 2\dfrac{\partial f}{\partial x} - 3\dfrac{\partial f}{\partial y}$

(c) $D_v f(1,-1) = (-2, -5e^{-1} - 5)$

3.3.8 The average rate of change is $\frac{\arctan(3)-\arctan(1)-3}{\sqrt{5}} \approx -1.13$. The instantaneous rate of change is $\frac{-3}{\sqrt{5}}$. You must use a unit vector for measuring instantaneous rate of change in a direction because you are trying to measure how fast the function changes “per unit step”. This rate is approximated by the average rate of change which takes into account the distance.

Remark: This may seem confusing since you would approximate $f(a+v) - f(a)$ by $D_v f(a)$ for any vector v. But $f(a+v) - f(a)$ is not the average rate of change! The average rate of change is

$$\frac{f(a+v) - f(a)}{||v||}$$

and this approximates

$$\frac{1}{||v||} D_v f(a) = D_{v/||v||} f(a).$$

The last identity is why you must normalize your direction.

3.3.9 If $v = (v_1, v_2, v_3$ then $D_v f = v_1\left(y^2z^3 \arctan(xyz) + \frac{xy^3z^4}{1+x^2y^2z^2}\right) + v_2\left(2xyz^3 \arctan(xyz) + \frac{x^2y^2z^4}{1+x^2y^2z^2}\right) + v_3\left(3xy^2z^2 \arctan(xyz) + \frac{x^2y^3z^3}{1+x^2y^2z^2}\right)$.

3.3.10 $D_v f(x) = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} v$

3.3.11 $D_v f(x) = 2x \cdot v$.

3.3.12 $D_v f(x) = Av$.

3.3.13 This is a direct application of Definition 3.3.1.

3.3.14

(a) You want to minimize a quantity involving E. Remember that a direction is a unit vector.

(b) $q = p + hu$. Notice u is a vector which depends on E and p.

3.3.15

(a) Copy Definition 3.3.1.

(b) The error is the distance between the two labelled points.

(c) Your plot in $\mathbb{R}^2$ should contain a graph of the function $g(t) = f(a + tv)$ and the tangent line to this graph at the point $t = 0$.

(d) The answer should be a change in the revenue resulting from some change in price of PayStations and Sintendos. Be specific about what change in price.

3.3.16

(a) The four correct expressions are II, III, V, and VII. To decide which is which, notice the curve, line, and tangent line are parametrized by one variable. Also, pay attention to whether the set is a subset of $\mathbb{R}^n$ or $\mathbb{R}^m$.

(b) The three correct expressions are I, IV, and VII. To decide which is which, notice c and d are shifts from the base point $F(a)$. Also notice the vector w is a scalar multiple of $d - F(a)$. The error in the approximation corresponds to $d - c$.

3.4.1

(a)
- (a) $n \times 1$ column vector. It represents the gradient of f at the point $a \in A^\circ$.
- (b) Vector valued function $\nabla f : A^\circ \to \mathbb{R}^n$. For each point $a \in A^\circ$ of its domain, it returns a vector from part (a).
- (c) Scalar. Dot product of a vector from part (a) with the vector v.
- (d) Scalar. Same real number as in the previous part, just obtained from "matrix" multiplication of an $1 \times n$-matrix by $n \times 1$-matrix.

(b) Second equality is an immediate consequence of Definition 3.4.1. The first equality related two separate concepts which were defined independently, namely the notion of directional derivative on the left hand side and the notion of the gradient vector on the right hand side. Relation between them is a subject of Theorem 3.4.3.

(c) $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are defined at a

3.4.2 F_1, F_2, F_3, F_4 matches C, A, B, D respectively.

3.4.3

(a) Consider a few questions about the vector field, and ask how they relate to the graph. Which way are the arrows pointing? When are the vectors largest? When do the vectors vanish?

(b) Your vector field should look like arrows pointing radially inward towards the origin and getting smaller as they approach the center.

(c) Your vector field should look constant. It will look like a steady stream pointing southeast. That is, every arrow will be the same size and point southeast.

3.4.4 In all cases gradient vector fields must be perpendicular to the level curves pointing in the direction of the level curve corresponding to the next largest value.

(a) Gradient vector field is pointing outwards. The norm of the vectors *increases* as we move away from the origin.

(b) Gradient vector field is pointing outwards. The norm of the vectors *stays constant* as we move away from the origin.

(c) Gradient vector field is constant and pointing to the top left direction.

(d) Plot gradient vectors perpendicular to the level curves, pointing in the direction of the steepest ascent; the norm of the gradient vectors is bigger when two level curves are closer together, hence the norm of the gradient vectors here should increase as we move away from the origin.

(e) Amy is correct when she is saying that there is no contradiction as $\nabla G(0,0) = (0,0)$ can be a possible value of the gradient at the origin which is perpendicular to both of the level curves. At the same time, based on the picture alone we cannot make definitive conclusion whether gradient at the origin is zero or not defined.

3.4.5

(a) $\nabla f(A,B) = \Big(-\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - Ax_i - B), -\frac{2}{n}\sum_{i=1}^{n}(y_i - Ax_i - B)\Big)$

(b) $\nabla f(x) = v = (v_1, \dots, v_n)$

(c) $\nabla f(x) = -\dfrac{x-a}{||x-a||^3}$

3.4.6

(a) $\forall h \in \mathbb{R}, L_{a,v}(h) = f(a) + D_v f(a)h$

(b) $\forall h \in \mathbb{R}, L_{a,v}(h) = f(a) + (\nabla f(a) \cdot v)h$

(c) $\forall x \in \mathbb{R}^n, L_a(x) = f(a) + \nabla f(a) \cdot (x - a)$

(d) $f(0.1, 0.2) \approx 0.3$

3.4.7

(a) Consider the function $f(x,y) = y^2 + 2xy - x^3 + 4x$. Which level set of this function is the given curve?

(c) $\nabla f(a,b) = (2b - 3a^2 + 4, 2a + 2b)$ and

$$L_{a,b} = \{(x,y) \in \mathbb{R}^2 : (2b - 3a^2 + 4)(x - a) + (2a + 2b)(y - b) = 0\}. \tag{A.3.1}$$

3.4.8

(b) $P_a = \{x \in \mathbb{R}^n : \nabla f(a) \cdot (x - a) = 0\}$

3.4.9 $10x - 8y + 7z = 32$

3.4.10

(a) For $u \in S^{n-1}$, define $F_a(u) = D_u f(a)$. You cannot use the gradient.

(b) If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a and $\nabla f(a) \neq 0$ then the maximum value of $F_a : S^{n-1} \to \mathbb{R}$ is attained at the unit vector $\frac{\nabla f(a)}{||\nabla f(a)||} \in S^{n-1}$.

3.4.11 This follows quickly from the product rule for partial derivatives and the definition of the gradient.

3.4.12

(a) The coordinates correspond to Toronto. Note this does not represent heat flow.

(b) Your explanation will need to use the fact that the sun rises from East.

(c) If you spent an extra thousand dollars on one of them, what would you expect to happen?

(d) One estimate suggests they will produce about an extra 58 pens. Another estimate suggests they will produce about an extra 70 pens. Both are reasonable.

3.4.13

(a) Choose the direction $-\frac{\nabla E(p_0)}{||\nabla E(p_0)||}$ provided the denominator is not zero. Your answer must be a unit vector since it asking for a direction. However there is a critical with this choice; see below.

(b) $p_1 = p_0 - h_0 \frac{\nabla E(p_0)}{||\nabla E(p_0)||}$

(c) For $k \in \mathbb{N}$, $p_{k+1} = p_k - h_k \frac{\nabla E(p_k)}{||\nabla E(p_k)||}$.

(d) There are two cases. In one case, $\nabla E(p_k)$ is never equal to zero for any k. In another case, it may equal zero for some k. Figure out what to do in each case.

3.5.1

(a)

i) Linear map $dF_a : \mathbb{R}^n \to \mathbb{R}^m$. Recall the definition of the differential of F at the point $a \in A^\circ$.

ii) Vector in $\mathbb{R}^m$, or equivalently $m \times 1$ column vector. This is the result of the evaluation of the linear map from part i) on a vector $v \in \mathbb{R}^n$.

iii) $m \times n$ matrix. This is a Jacobian matrix at the point $a \in A^\circ$, which is precisely the matrix of a linear map from part i) written in a standard basis.

iv) Vector in $\mathbb{R}^m$, or equivalently $m \times 1$ column vector. Same vector as in ii), here it is obtained by multiplying an $m \times n$ matrix $DF(a)$ by an $n \times 1$ column vector v.

v) Linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Here, for all $j \in \{1, \dots, n\}$, the $F^j : A \to \mathbb{R}$ is a real valued function on A° given by the first component of F. Its differential is a linear map $dF^j : \mathbb{R}^n \to \mathbb{R}$. A vector constructed out of m linear functions on $\mathbb{R}^n$ is just a linear map $\mathbb{R}^n \to \mathbb{R}^m$.

vi) $m \times n$ matrix. This expression is an equivalent way of writing the Jacobian matrix at $a \in A^\circ$ and is equal to variant iii). Note that each of the partial derivatives $\partial_j F(a)$ for any $j \in \{1, \dots, n\}$ is an $m \times 1$ column vector. We then put n of them side by side to form a $m \times n$ Jacobian matrix.

vii) Scalar.

viii) $m \times n$ matrix. Yet another way to write Jacobian matrix at point $a \in A^\circ$.

(b)

- First equality is an equivalence between directional derivative and the evaluation of the differential map on a directional vector which holds by Theorem 3.5.13. The same theorem was also announced before rigorous proof as Theorem 3.3.10.
- The second equality is the subject of the Theorem 3.5.22 which proves that matrix of a differential in a standard basis is just a Jacobian matrix of f at a.
- Third equality is Definition 3.5.15 of a Jacobian matrix and the definition of matrix multiplication.
- The last equality is the subject of Lemma 3.2.7.

3.5.2 (d) is true by Theorem 3.5.22, (e) is true by Definition 3.5.15, and the rest are nonsense.

3.5.3
(a) True by Lemma 3.5.11
(b) True by Theorem 3.5.13
(c) True by Theorem 3.5.13
(d) True by Theorem 3.5.13 and Definition 3.5.15.
(e) True by Theorem 3.5.22.
(f) False. What type of mathematical objects are $d\gamma_a$ and $\gamma'(a)$?
(g) False. What type of mathematical objects are $Df(a)$ and $\nabla f(a)$?

3.5.4
(b)

$$\frac{\partial F}{\partial x} = \left(\frac{1}{1-z}, 0\right)$$
$$\frac{\partial F}{\partial y} = \left(0, \frac{1}{1-z}\right)$$
$$\frac{\partial F}{\partial z} = \left(-\frac{x}{(1-z)^2}, -\frac{y}{(1-z)^2}\right)$$

(c) $DF(2,3,7) = \begin{bmatrix} -1/6 & 0 & 1/18 \\ 0 & -1/6 & 1/12 \end{bmatrix}$
(d) The differential of F at $(2,3,7)$ is the function $dF_{(2,3,7)}\mathbb{R}^3 \to \mathbb{R}^2$ defined by $dF_{(2,3,7)}(h) = \begin{bmatrix} -1/6 & 0 & 1/18 \\ 0 & -1/6 & 1/12 \end{bmatrix} h$ for all $h \in \mathbb{R}^3$.
(e) $\begin{bmatrix} 2/9 \\ 1/12 \end{bmatrix}$.
(f) $\begin{bmatrix} 2/9 \\ 1/12 \end{bmatrix}$.
(g) $F(3,3,4) \approx \begin{bmatrix} -2/3 \\ -3/4 \end{bmatrix}$

3.5.5
(a) $dg_{(1,2)} : \mathbb{R}^2 \to \mathbb{R}^3$ is the function defined by $dg_{(1,2)}(h) = \begin{bmatrix} 2 & 1 \\ 2 & 4 \\ 2 & -4 \end{bmatrix}$ for all $h \in \mathbb{R}^2$.
(b) $g(1.1, 1.8) \approx (2, 4.4, -2)$

3.5.6 $Df(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$

3.5.7 $df_{(r,\theta,z)} : \mathbb{R}^3 \to \mathbb{R}^3$ is defined by $df_{(r,\theta,z)}(h) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} h$ for all $h \in \mathbb{R}^3$.

3.5.8
$$Df(\rho,\theta,\phi)=\begin{bmatrix}\cos\theta\sin\phi & -\rho\sin\theta\sin\phi & \rho\cos\theta\cos\phi\\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi & \rho\sin\theta\cos\phi\\ \cos\phi & 0 & -\rho\sin\phi\end{bmatrix}$$

3.5.9 Although you cannot formally apply Theorem 3.5.22, you can use the idea to help define your linear map. If F were differentiable at $(4,6)$, what would be its differential at $(4,6)$ according to Theorem 3.5.22? Once you guess this, you will know how to define your desired linear map $L:\mathbb{R}^2\to\mathbb{R}^2$. As part of your proof, you will need to use the identity $\lim_{(h_1,h_2)\to(0,0)}\frac{h_1h_2}{||(h_1,h_2)||}=0$. This requires its own ε-δ proof but you may take it for granted. State when you apply it.

3.5.10 Although you cannot formally apply Theorem 3.5.22, you can use the idea to help define your linear map. If G were differentiable at $(4,1,6)$, what would be its differential at $(4,1,6)$ according to Theorem 3.5.22? Once you find this, you will be able to define a linear map $L:\mathbb{R}^3\to\mathbb{R}^2$. As part of your proof, you will need to use the identity $\lim_{(h_1,h_2,h_3)\to(0,0,0)}\frac{h_3^2}{||(h_1,h_2,h_3)||}=0$. This requires its own ε-δ proof but you may take it for granted. State when you apply it.

3.5.11

(a) This requires a careful application of the definition of differentiability. Choose a candidate for the differential and ensure it satisfies all the requirements of the definition. Remember that $f(x+y)=f(x)+f(y)$ since f is linear.

(b) First, justify the Jacobian exists. Second, calculate the partials of f. It will help to write M as a bunch of column vectors.

3.5.12

(a) For the "only if" direction, express the differential $L=df_a$ in components $L=(L_1\ldots,L_m)$. Explain why each component must be linear. Then use these to finish the proof. For the "if" direction, define a linear map $L=(df_a^1,\ldots,df_a^m)$ and use this to prove your assertion.

(b) First, justify the Jacobian exists. Second, use Definition 3.5.15 to confirm the identity.

3.5.13 This requires a careful application of the definition. It will be easiest to do both parts of the proof concurrently by choosing a candidate for the differential and showing that it satisfies the definition.

3.5.14

(a) The function $L:\mathbb{R}^n\to\mathbb{R}^m$ given by $L(x)=F(a)+DF(a)(x-a)$ is the linear approximation of F at a. The error is $||h||\varepsilon_a(h)$.

(b) It doesn't just tend to zero. It tends to zero faster than ____.

(c) After choosing ε_a and the linear map in the definition of differentiability, the proof is very short. Be careful to justify each step and not violate any limit laws or write nonsensical expressions (e.g. quotients of vectors). Remember to define ε_a at the origin separately.

3.5.15

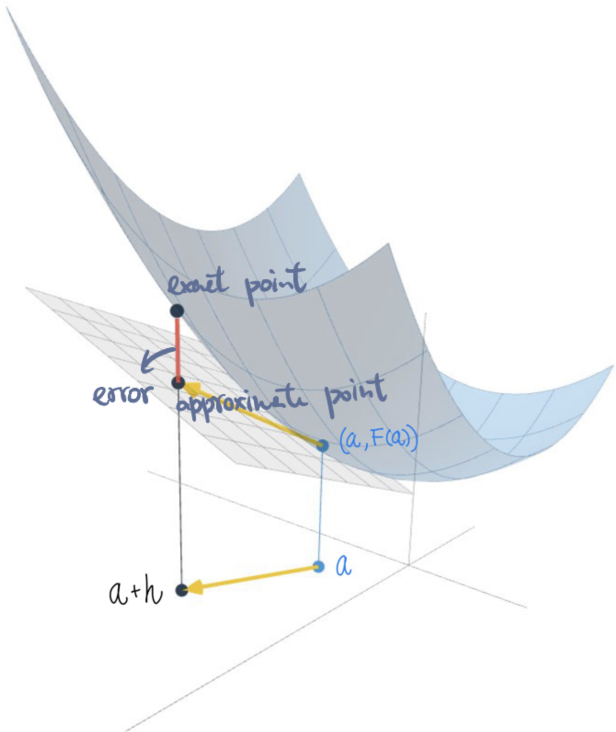
(c) $||h||$ and $||F(a+h)-F(a)-DF(a)h||$

(d) At what rate does the length of the green and red lines go to zero? Compare them.

3.5.16

(a) The picture uses $m=1$ and $n=2$.

(b) Analytic



(c) $(a+h, F(a+h))$

(d) $(a+h, F(a)+dF_a(h))$

(e) $F(a+h)-F(a)-dF_a(h)$

(f) $P=\left\{\left(x, F(a)+dF_a(x-a)\right): x \in \mathbb{R}^n\right\}$

3.5.17

(a) Consider cases $n \leq m$ and $n \geq m$ separately.

(b) Use the previous part and that rank is the dimension of the span of the matrix's columns.

(c) Consider the matrix associated to T and used the rank-nullity theorem.

(d) Consider the matrix associated to T and used the rank-nullity theorem.

(e) Use the previous two parts.

3.6.1

(a) Equivalent by Definition 3.5.1.

(b) Not equivalent. $h \in \mathbb{R}^n$ and a vector divided by another vector is not defined, so this expression is meaningless.

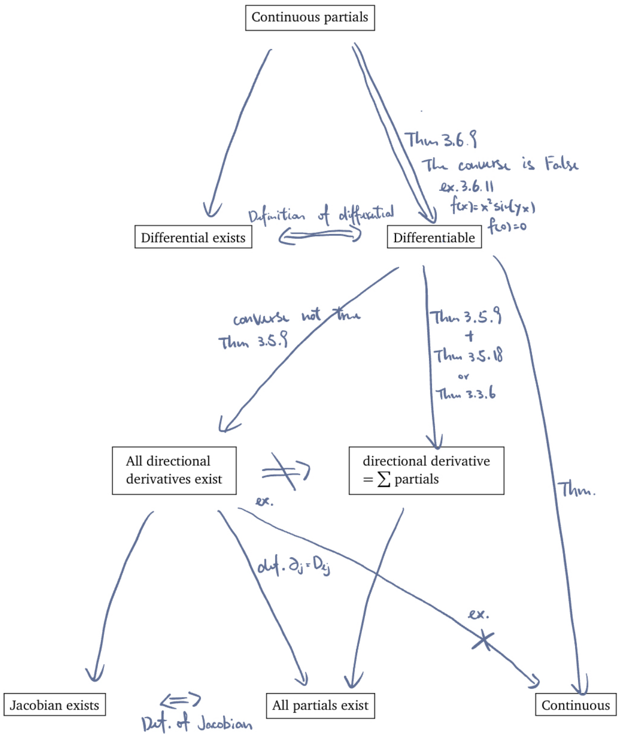
(c) Not equivalent, same reason as above.

(d) Equivalent

(e) Not equivalent, same reason as above.

3.6.2
(a) True by the definition of partial derivatives and Theorem 3.3.10.
(b) False, seeExample 3.3.8 for counterexample.
(c) False, the "only if" direction is true but the converse is not true. Check Example 3.3.9 for counterexample.
(d) True by Definition 3.5.1
(e) True by Theorem 3.6.9
(f) False, see Example 3.6.11 for counterexample

3.6.3



3.6.4 Since a is a limit point, you want to show $\lim_{x \to a} F(x) = F(a)$. Start with the expression $\lim_{x \to a} F(x)$ and begin manipulating it. Add and subtract stuff. Multiply and divide by stuff. Be careful with how you write this proof. It is easy to incorrectly write things in the wrong order. Did you accidentally assume your conclusion? Did you assume a limit or quantity exists without justifying it? What limit laws did you use? Did you check their assumptions?

3.6.5 Each partial derivative of F is a constant map. Therefore, ...

3.6.6 By linearity, you only need to prove the statement is true for monomials. Use induction on the degree of the monomial.

3.6.7
(a) You will need to break down the vector-valued limit into components. Let $g(t)$ be the expression appearing in the righthand limit on Line 5. Write $g = (g_1, \ldots, g_m)$ in components and state an inequality for each $g_i(t)$ on an interval of t around $0 \in \mathbb{R}$.

(b) The first equality uses that $F(\gamma(t))$ is equal to the leftmost quantity for $t \in (-\delta, \delta)$. This requires $F(\gamma(t))$ to be defined for such t which is guaranteed by Lines 7, 9, and 10. The second equality uses continuity of γ at 0 and of F at $\gamma(0) = a$ (per Line 8), as well as a limit law about composition of continuous functions. The third equality uses Line 10 again, and the fourth equality uses Line 7 again.

(c) Choose $\varepsilon = \delta/||v||$ for the radius of your open ball.

3.6.8

(a) You will need Remark 3.3.3, differentiability, and the definition of the Jacobian.

(b) Express v as a sum of standard basis vectors and try to prove the required result using what has already been shown.

3.6.9

(a) If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $a \in \mathbb{R}^n$ then f is differentiable at a.

(b) Your proof should start with Lemma II. Then you will need to apply Theorem D a total of m times, once for each coordinate function. Then you will apply Lemma I once.

(c) i) f is C^1 at a so all its partials exist at a implying $\nabla f(a)$ exists.
ii) To prove f is differentiable at a, you must choose a linear map. What is the linear map?
iii) $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\forall h \in \mathbb{R}^n, 0 < ||h|| < \delta \implies \left| \frac{f(a+h)-f(a)-\nabla f(a)^T h}{||h||} \right| < \varepsilon$

(d) If $\partial_j f$ has domain $A_j \subseteq \mathbb{R}^n$, then the equation in Line 6 should start with "$\forall x \in A_j$" instead of "$\forall x \in \mathbb{R}^n$. You can verify that these statements are equivalent because, by Line 5, a must be an interior point of A_j.

(e) i) By comparing limit definitions, you can verify that

$$g_i'(x) = \partial_i f(a_1 + h_1, \ldots, a_{i-1} + h_{i-1}, x, a_{i+1}, \ldots, a_n)$$

so g_i is differentiable at $x \in (a_i - \delta, a_i + \delta)$ if and only if $\partial_i f$ exists at $y = (a_1 + h_1, \ldots, a_{i-1} + h_{i-1}, x, a_{i+1}, \ldots, a_n)$. Based on Line 8, this means you need to prove y belongs to which set? To prove your assertion, you will need to estimate $||y - a||$, apply the triangle inequality, and notice that $||h(i)|| \leq ||h|| < \delta$.

ii) If h_i is positive, then it is applied on the interval $[a_i, a_i + h_i] \subseteq \mathbb{R}$. By the MVT, you need that g is continuous on $[a_i, a_i + h_i]$ and differentiable on $(a_i, a_i + h_i)$. Why does this follow from the conclusion in Line 12? (A similar argument applies when h_i is negative.)

(f) ii) $b_i = (a_1 + h_1, \ldots, a_{i-1} + h_{i-1}, c_i, a_{i+1}, \ldots, a_n) \in \mathbb{R}^n$. To justify $b_i \in B_\delta(a)$, estimate $||b_i - a||$, remember that c_i is between a_i and $a_i + h_i$, and $||h|| < \delta$.

(g) In no particular order, you need: the triangle inequality, $|h_i| \leq ||\mathbf{h}||$, "the definition of the gradient", "by equation (2) as $b_i \in B_\delta(a)$", and "by equation (3) as $f(a+h) - f(a) = \sum_{i=1}^n f(a + h(i)) - f(a + h(i))$".

3.6.10

(b) The total change in f from a to $a + h(3)$ can be broken up into a coordinate-by-coordinate change. In particular, the total change is the sum of its change from a to $a + h(1)$, its change from $a + h(1)$ to $a + h(2)$, and its change from $a + h(2)$ to $a + h(3)$. Each change is in the direction of a single coordinate.

(c) A) corresponds to a single sentence.
B) corresponds to two lines.

C) corresponds to the application of the single variable MVT.
D) corresponds to a short justification at the very end.
E) corresponds to a bunch of equations and inequalities.

3.6.12
(a) Use Section 3.5.3(b).
(b) Use Section 3.5.3(c) and the idea that a linear approximation at a should preserve a local property at a.
(c) The same as the previous part.

A.4. Derivative applications

4.1.1 The first and third equality follow from the theorem that the matrix of the differential is the Jacobian. The second equality follows from the chain rule.

4.1.2 Note that in the Theorem above we obtain the differential of the composition as a composition of the differentials of the two maps at $F(a) \in V$ and $a \in U$ respectively. This immediately rules out two out of five. The remaining three are correct.
(a) Incorrect, $DG(a)$ is evaluated at the wrong point.
(b) Correct. Recall that the matrix of the composition of two linear maps is nothing but the product of the corresponding matrices.
(c) Correct, it is the statement of Corollary 4.1.4.
(d) Incorrect, dG_a is evaluated at the wrong point.
(e) Correct, again by Corollary 4.1.4.

4.1.3
(a) Rational functions are C^1 on their domains. In particular, F is differentiable at $(0,1)$ and G is differentiable at $F(0,1) = (0,4,3)$
(b) $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$

4.1.4 The chain rule tree is not really a trick. It is just a nice way to visualize multiplying Jacobians.
(a) $\dfrac{\partial u}{\partial s} = \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial s} + \dfrac{\partial u}{\partial y}\dfrac{\partial y}{\partial s} + \dfrac{\partial u}{\partial z}\dfrac{\partial z}{\partial s}$. Similar for $\dfrac{\partial u}{\partial t}$.
(b) Repeat the method for each component. How many trees will you need?
(c) i) $\dfrac{\partial u}{\partial s} = -1/2$ and $\dfrac{\partial u}{\partial t} = 0$ when $(s,t) = (0,1)$
ii) $\dfrac{\partial v}{\partial s} = -1$ and $\dfrac{\partial v}{\partial t} = 7/2$ when $(s,t) = (0,1)$

4.1.5

(a) You cannot apply the chain rule. Are all of the functions given here differentiable?

(b) $(F \circ \gamma)'(0) = 1/2$

(c) Plot the graph of F in $\mathbb{R}^3$. What does (a) say about this graph? Then sketch the part of the graph corresponding to $F \circ \gamma$. It should trace out a curve on the graph. What does (b) say about this curve?

4.1.6

(a)

$$D(f \circ T)(r,\theta) = \left[\partial_1 f(r\cos\theta, r\sin\theta)\cos\theta + \partial_2 f(r\cos\theta, r\sin\theta)\sin\theta \quad -\partial_1 f(r\cos\theta, r\sin\theta) r\sin\theta + \partial_2 f(r\cos\theta, r\sin\theta) r\cos\theta\right]$$

(b)

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial u}{\partial y}\sin\theta$$
$$\frac{\partial u}{\partial \theta} = \frac{\partial u}{\partial x}(-r\sin\theta) + \frac{\partial u}{\partial y} r\cos\theta$$

4.1.7

(a)

$$D(f \circ g)(0) = \begin{bmatrix} 1 & 2 & -5 \\ 0 & 1 & -2 \\ -e^4 & 3e^4 & -5e^4 \end{bmatrix}$$

(b)

$$D(g \circ f)(0) = \begin{bmatrix} 6-4e^2 & 1+2e^2 \\ 2+2e^2 & 1-e^2 \end{bmatrix}$$

4.1.8

(a) $\gamma'(t) = (0, -4, 3)$

(b) $\frac{\partial f}{\partial t}(3\pi/4) = -27\pi$

4.1.9

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial u}{\partial y}\sin\theta$$
$$\frac{\partial u}{\partial \theta} = \frac{\partial u}{\partial x}(-r\sin\theta) + \frac{\partial u}{\partial y} r\cos\theta$$
$$\frac{\partial u}{\partial z} = \frac{\partial u}{\partial z}$$

4.1.10

$$\frac{\partial u}{\partial \rho} = \frac{\partial u}{\partial x}\cos\theta\sin\phi + \frac{\partial u}{\partial y}\sin\theta\sin\phi + \frac{\partial u}{\partial z}\cos\theta$$
$$\frac{\partial u}{\partial \theta} = \frac{\partial u}{\partial x}(-\rho\sin\theta\sin\phi) + \frac{\partial u}{\partial y}\rho\cos\theta\sin\phi$$
$$\frac{\partial u}{\partial \phi} = \frac{\partial u}{\partial x}\rho\cos\theta\cos\phi + \frac{\partial u}{\partial y}\rho\sin\theta\cos\phi + \frac{\partial u}{\partial z}(-\rho\sin\phi)$$

4.1.11

(a) Your picture should include a choice of γ and two vectors on S.

(b) Define a level set and differentiate $f \circ \gamma$ at a point in the set.

4.1.12 It may help to start with the case $n = 3, m = 2$ or some other case with a small number of variables. You will need Theorem 2.7.16 and Corollary 2.7.18.

4.1.13

(a) Write down the differentials dF_a and $dG_{F(a)}$ along with their domains and codomains. Then relate them to the differential $d(G \circ F)_a$.

(b) The linear approximation of $G \circ F : \mathbb{R}^n \to \mathbb{R}^k$ at a is given by $L : \mathbb{R}^n \to \mathbb{R}^k$ where $L(x) = G(F(a)) + dG_{F(a)}\big(dF_a(x-a)\big)$.

(c) Each arrow between the pictures represents a transformation, and the vectors in each picture are tangent vectors to the surface.

(d) Think of the temperature $T : \mathbb{R}^3 \to \mathbb{R}$ as fixed as you move along a path $\gamma : \mathbb{R} \to \mathbb{R}^3$ on the Earth's surface $S \subseteq \mathbb{R}^3$ so $\gamma(\mathbb{R}) \subseteq S$. At a given point on the path, there is your velocity at that point and there is the temperature gradient at that point. Combine these using the chain rule. What does that measure?

4.2.1 Carefully examine the statement of the mean value theorem. Remember the Jacobian of a real-valued function is the transpose of its gradient.

(a) False

(b) True

(c) True

(d) False

4.2.2

(a) You will need the Cauchy-Schwarz inequality on the dot product $\nabla f(c) \cdot (b-a)$.

(b) You must justify that the maximum exists using the extreme value theorem.

4.2.3

(a) The fatal mistake occurs in the third step. What does the quantity c depend on?

(b) If a and b are the endpoints then $F(b) - F(a) = (0,0)$ but the derivative of F is never zero. How can you tell this from the graph?

4.3.1

(a) Not equivalent. This is the definition of a global maximum.

(b) Not equivalent. This is always false when a is a boundary point of A. Can you see why?

(c) Equivalent.

(d) Not equivalent. This is is the definition of a local extremum, but it does not specify whether a is a local min or local max.

4.3.2

(a) Saddle point.

(b) Local minimum.

(c) Local maximum.

(d) Local minimum.

(e) Local minimum.

(f) Saddle point.

4.3.3

(a) A and C appear to be local maxima. B, E, G appear to be saddle points. F appears to be a local min. All points except D appear to be critical points.

(b) Note that D is the closed rectangle. There are 2 saddle points, 4 local maxima, and 2 (visible) local minima. It also appears that there might be one more local minima hidden on the back which we cannot label. There are only 3 critical points.

4.3.4 Only (b) is false.

4.3.5

(a) There is precisely one local minimum.

(b) There is precisely one local maximum.

(c) There are no critical points.

(d) There are infinitely many local minima.

(e) There is precisely one saddle point.

(f) There is precisely one local maximum.

4.3.6 $f(x) = x^3$.

4.3.7 $g(x, y) = x^2 - y^2$.

4.3.8

(a) $(2, 2)$.

(b) There are no critical points.

4.3.9 $(2^{-1/3}, 2^{-1/3})$. Note $(0, 0)$ is not a critical point.

4.3.10 $(0, 0, 0)$.

4.3.11 a is the only critical point since $\nabla F(a) = 0$.

4.3.12 a is the only critical point since $\nabla G(a) \neq 0$.

4.3.13
(a) Your lefthand figure should include

$$(a, f(a)) \qquad \{(a+te_j, f(a+te_j) \in \mathbb{R}^n \times \mathbb{R} : t \in (-\varepsilon, \varepsilon)\} \qquad \{(a+te_j, f(a)+\partial_j f(a)t) \in \mathbb{R}^n \times \mathbb{R} : t \in (-\varepsilon, \varepsilon)\}.$$

Your righthand figure should include

$$(0, g(0)) \qquad \{(t, g(t)) : t \in (-\varepsilon, \varepsilon)\} \qquad \{(t, g(0) + g'(0)t) : t \in (-\varepsilon, \varepsilon)\}$$

(b) If $\nabla f(a)$ does not exist, then the conclusion is already true and there is nothing prove.

(c) Line 3 uses that a is an interior point. To prove this claim, take an open ball with radius δ and choose ε in terms of δ.

(d) Line 6 holds by Line 1 and the definition of the gradient.

(e) Consider cases. Recall that g has a local maximum at $0 \in \mathbb{R}$ if and only if there exists $\delta > 0$ such that $g(t) \leq g(0)$ for $t \in (-\delta, \delta)$. You must choose δ using the assume that f has a local maximum at $a \in A$.

4.3.14 Try using the global extreme value theorem on a certain set. After applying the global EVT, remember to consider cases for where the local extremum can occur, i.e. on the boundary or the interior. Neither situation is complicated but you must address this.

4.3.15 Apply the definition of critical points, which leads to two cases. Are both cases possible?

4.4.1
(a) True

(b) True

(c) False

(d) False

(e) True

4.4.2 One of the first things to inquire about the given optimization problem is whether global EVT applies, so the first questions can be:

- Is f continuous on its domain?
- Is the domain S compact?

If both answers are positive, one can proceed by the method described in the textbook; otherwise, one has to use methods specific to the problem. In any case, local EVT can help eliminating those points of the interior which are not local extrema, so the following questions to inquire are:

- At which points of the interior S° is the function f differentiable?
- What is the set of critical points of f?

4.4.3
(a) There is a maximum and minimum on D but we cannot determine what they are with the given information.

(b) The maximum is 8 and the minimum is $-\pi$.

(c) There is a maximum and minimum on S and they lie on the boundary of S.

4.4.4

(a) Your graph can be generic, but make sure to label all critical points, the corners of S, and the extrema.

(b) The only critical point on the interior of S is $(3, -\frac{3}{2})$.

(c) This will require four separate parametric curves. Don't forget to check the boundary points of each of these curves.

(d) In the end you should have 7 points on which to check the value of g. The minimum is $-\frac{29}{4}$ and the maximum is 14.

4.4.5 Saddle point at (π, π); local maxima at $(\frac{\pi}{2}, \frac{\pi}{2})$ and $(\frac{3\pi}{2}, \frac{3\pi}{2})$; local minima at $(\frac{\pi}{2}, \frac{3\pi}{2})$ and $(\frac{3\pi}{2}, \frac{\pi}{2})$

4.4.6 Minimize the square of the distance. The shortest distance is $\sqrt{\frac{23}{3}}$.

4.4.7 The point $(\frac{83}{24}, \frac{83}{12}, \frac{415}{24})$ on the plane is closest to the point $(\frac{1}{2}, 1, \frac{81}{4})$ on the surface with distance $\sqrt{\frac{5041}{96}}$.

4.4.8 Use polar coordinates for the boundary. The maximum value is $\sqrt{2}$.

4.4.9 Use spherical coordinates for the boundary. The maximum value is $\sqrt{3}$.

4.4.10

(a) Only points on the interior of the domain have been checked.

(b) After we have parameterized the boundary, we must solve a brand new optimization problem for a new function on a new domain. In particular, we have to check both the interior points of the new domain and the boundary points of the new domain.

(c)
(a) The definition of critical points gives two possibilities. Only one is considered here.
(b) Lines 8 and 10.
(c) Lines 8 and 10.
(d) Your answer should involve lines 2, 8, and 9.

4.4.11

(a) How do we know that a (global) maximum exists?

(b) The local extreme value theorem only applies to the interior of S.

4.4.12 This requires a careful combination of the global and local extreme value theorems.

4.4.13

(a) Use the definition of $\lim_{||x||\to\infty} f(x) = \infty$ and both extreme value theorems.

(b) Apply the previous part to $-g$.

4.4.14

(a) $S = \{(x, y, z) \in \mathbb{R}^3 : x + y + z \leq 4500 \text{ and } x, y, z \geq 0\}$, $f(x, y, z) = xyz$.

(b) There are no critical points in the interior of S.

(c) This will require parameterizing 4 different surfaces and then 6 different curves. The argument can simplified by noting that our function takes a particularly simple form along some of these surfaces.

(d) The maximum is $(1500)^3$ mm^3.

4.4.15 The line $y = -\frac{5}{2}x + \frac{29}{6}$ is the line of best fit.

4.4.16 The optimal cargo bay is 4 meters long and 2 meters wide, potentially swapped. You will need to compute some limits to justify that this is the global minimum.

4.4.17 The maximum value is $\sqrt{n}$ but you cannot easily solve this without Lagrange multipliers, which will appear later. The lesson here is that parametrizing surfaces is really hard! Sorry for the wild goose chase.

4.5.1
(a) B

(b) A

(c) $q = p + v$

4.5.2
(a) B and C are tangent vectors.

(b) D and E are equal to the tangent space.

(c) C and E are the tangent plane. Make sure the set does not contain the origin.

4.5.3 (d) and (e) are equivalent to v being a tangent vector of S at p.

4.5.4

(a)
i) I is an open interval in $\mathbb{R}$ containing 0.
ii) $\gamma(0) = p = (2,3,6)$.
iii) $\gamma'(0) = v = (4,-1,10)$.
iv) $\operatorname{im}(\gamma) \subset S$.

(b)
i) I is an open interval in $\mathbb{R}$ containing 0.
ii) $x(0) = 2,\ \ y(0) = 3,\ \ z(0) = 6$.
iii) $x'(0) = 4,\ \ y'(0) = -1,\ \ z'(0) = 10$.
iv) $\forall t \in I,\ z(t) = x(t)y(t)$.

(c) Let $I = (-\sqrt{237}, \sqrt{237})$, define $x, y, z : I \to \mathbb{R}$ as

$$x(t) = 2 + 4t, \qquad y(t) = 3 - t, \qquad z(t) = x(t)y(t) = 6 + 10t - 4t^2. \tag{A.4.1}$$

Note that we have to satisfy condition iv), but S is a graph of the function $f(x,y) = xy$. So we can choose a straight line in the XY-plane and lift it to a path on the graph by explicitly defining the third coordinate as $z(t) := f(x(t), y(t))$.

(d) This follows by computation of derivatives at $t = 0$. The choice of the path is not unique.

(e) $n = 3, k = 2, V = \mathbb{R}^2, F(x,y) = xy$, and $a = (2,3)$

(f) $dF_{(2,3)}(w) = (3,2) \cdot w$

(g) $T_pS = \operatorname{span}\{(1,0,3),(0,1,2)\}$

(h) $p + T_pS = \{(x,y,z) \in \mathbb{R}^3 | 3x + 2y - z = 6\}$

4.5.5

(a) $T_pS = \operatorname{span}\{(1,0,0,2,-4),(0,1,0,0,-2),(0,0,1,-4,2)\}$

(b) $p + T_pS = \{(x,y,z,u,v) \in \mathbb{R}^5 : u = 2x - 4z - 5, v = -4x - 2y + 4\}$

4.5.6

1. $DF(1,1) = \begin{bmatrix} 3 & -\pi \\ -2 & 4 \end{bmatrix}$
2. $F(1.2, 0.8) \approx (2.6 + 0.2\pi, -2.2)$
3. $q = (2, 0, F(2,0))$

4.5.7 Use the definition of a tangent vector and consider the constant curve through p.

4.5.8 What is the relationship between the differential and the derivative of f?

4.5.9

(a) Use the relationship between the differential and the Jacobian.

(b) Show that two vectors are linearly independent.

(c) The vector $(x,y,z) - (a,b,f(a,b))$ must belong to T_pS. This occurs if and only if there exists scalars $\lambda, \mu \in \mathbb{R}$ such that

$$\begin{bmatrix} x - a \\ y - b \\ z - f(a,b) \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 0 \\ \partial_1 f(a,b) \end{bmatrix} + \mu \begin{bmatrix} 0 \\ 1 \\ \partial_2 f(a,b) \end{bmatrix}.$$

Solve for λ and μ in terms of the other quantities, then express z accordingly.

(d) Any scalar multiple of $(\frac{\partial f}{\partial x}(a,b), \frac{\partial f}{\partial y}(a,b), -1)$.

(e) Try to think of the tangent plane as the graph of a function, namely the function which approximates f at p.

4.5.10

(a) You must prove two things. First, the span is equal to the finite set of k vectors

(b) Use the linear map $G(x) = (x, F(x))$. What is the rank of this linear map?

4.5.11

(a) Both T_pS and $p + T_pS$ are lines.

(b) $T_pS = \operatorname{span}\{(1,-1)\}$

(c) $p + T_pS = \{(x,y) \in \mathbb{R}^2 : x + y = 2\}$

4.5.12

(a) Check that your sets have the correct dimensions.

(b) Use the graph of F to define the inverse.

4.6.1

(a) Only the first two have this property.

(b) None have this property.

(c) Only the first two have this property.

4.6.2 B, C, and E.

4.6.3

(a) Not a smooth curve. It has 1 point where it fails.

(b) A smooth curve.

(c) Not a smooth curve. It has 1 point where it fails.

(d) A smooth curve.

4.6.4

(a) Yes. To prove that, first note

$$U \cap S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1, y > 0\}.$$

Define

$$V = (-1, 1), \qquad f : V \to \mathbb{R}, \quad f(x) = \sqrt{1 - x^2}.$$

It is a differentiable function on V and $S \cap U = \{(x, f(x)) : x \in V\}$ is a graph of this function.

(b) Yes. Note that again $U \cap S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1, y > 0\}$, so the same function as in part 1 works.

(c) Yes. Take $V = (-0.5, 0.5)$ and $f : V \to \mathbb{R}$, $f(x) = \sqrt{1 - x^2}$.

(d) No. Here $S \cap U = S$ which is not a graph of any function. One can prove this by contradiction. Suppose that $S \cap U$ is a graph. Then, there are only two options for the choice of the argument of the function: either x is an argument and $y = f(x)$, or y is an argument and $x = g(y)$. The first fails the "vertical" line test, and the second fails the "horizontal" line test.

(e) Yes. Same as in part 3.

(f) No. Same idea as in part 4.

(g) No. Same idea as in part 4.

(h) No, because $p \notin S \cap U$.

4.6.5

(a) Yes. Let $p = (x, y, z)$ be a point of the sphere. At least one of the three coordinates must be nonzero. Prove that in some neighbourhood of p, the sphere is a graph of the function of the two other coordinates. When defining a function, pay attention to the domain and differentiability of your function.

(b) No, it is not a manifold at the origin. No matter which pair of coordinates you choose and no matter how small of a neighborhood U of the origin you consider, the double cone would fail the "vertical" line test. Hence, it cannot be presented as a graph of the function in some neighborhood of the origin.

(c) No. It is a manifold, but a 3-dimensional one.

(d) Yes. To prove it, start by identifying all points p on a torus, where you can show that $f(x, z)$ does the job in some neighbourhood of p (spoiler: almost everywhere but the two circles). Deal with the remaining points on the pair of circles using either $g(x, y)$ or $h(y, z)$.

(e) No. It is not a manifold at the origin. Unlike part 2, the single cone is indeed a graph of $f(x, y) = \sqrt{x^2 + y^2}$, but this function is not differentiable at $(x, y) = (0, 0)$.

(f) Yes. To prove it, start by identifying the points p of the helicoid for which you can find a small enough neighborhood $U = B_\varepsilon(p)$ s.t. the intersection of U with the helicoid is a graph of a function $f(x, y)$. This will cover all points of the helicoid but the vertical line. As for the points on the vertical line, you can prove that either $g(x, z)$ or $h(y, z)$ will do the job.

4.6.6 All choices work except for (d). In each case you have to start by identifying $S \cap U$ using the set-builder notations. Note that quite often you will get the same intersection $S \cap U$ for different choices of U. This means that you can reuse the same V and f for seemingly different parts of the problem.

4.6.7

1. $\varepsilon \in (0, \sqrt{2})$
2. None.

4.6.8

1. $\varepsilon \in (0, \sqrt{2})$
2. None.
3. None.

4.6.9 After finishing this problem, reflect on how this strategy emulates the proof of Corollary 4.6.19.

(a) There are many valid choices. One possible choice is $U = \mathbb{R} \times (0, \infty)$ and $f : (-2, 2) \to \mathbb{R}$ given by $f(x) = \sqrt{2 - x^2}$.

(b) Use a theorem on tangent spaces of graphs to find that $T_p(S') = \operatorname{span}\{(1, -1)\}$.

(c) Tangent spaces are local so you can conclude $T_pS = T_p(S')$. Explain why using the previous parts.

4.6.10

(a) There are many valid choices. One possible choice is $U = \mathbb{R}^2 \times (0, \infty)$ and $f : \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 3\} \to \mathbb{R}$ given by $f(x, y) = \sqrt{3 - x^2 - y^2}$.

(b) Use a theorem on tangent spaces of graphs to find that $T_p(S') = \operatorname{span}\{(1, 0, -1), (0, 1, -1)\}$.

(c) Tangent spaces are local so you can conclude $T_pS = T_p(S')$. Explain why using the previous parts.

4.6.11

(a) There are several correct choices. One possibility is to use Problem A.4.4 part 1.

(b) You cannot use either graph to prove that S is a 1-dimensional smooth manifold at $(1, 0)$ or at $(-1, 0)$.

(c) The union of graphs of C^1 maps is not necessarily a smooth manifold. Can you find a counterexample? Try a pair of lines in $\mathbb{R}^2$.

4.6.12

(a) The article 'a' instead of 'the' makes a big difference somewhere in Line 2.

(b) Your explanation should include Line 10, Line 11, and $S \cap U$.

(c) Your drawing should use the two ε-points to informally illustrate the idea that "S fails the horizontal line test".

4.6.13

(a) Try to swap S and S' in the theorem. What changes?

(b) Since U is open and $p \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. How can you use this ε with the definition of continuity of γ at 0?

4.6.14 You will need at least 2 open sets. Use half planes.

4.6.15 You will need at least 4 open sets. Use half planes.

4.6.16 You will need at least 6 open sets. Use half spaces.

4.6.17 Proceed by contradiction with some arbitrary open set and a function. You will need to show that it is not a graph of the function in 2 ways.

4.6.18 Proceed by contradiction with some arbitrary open set and a function. You will need to show that it is not a graph of the function in 3 ways.

4.6.19 See the proof of Lemma 4.5.17 in the textbook for an outline. Try $U = \mathbb{R}^n$ in the definition.

4.6.20 You only need to prove it is not a 1-dimensional smooth manifold at a single point. Your proof should proceed by contradiction involve some arbitrary open set and a function. The domain of your function will produce a contradiction.

A.5. Inverse and implicit functions

5.1.1

(a) $F|_U$ and $F|_V$ are diffeomorphisms.

(b) You can use U or V for this one.

(c) $(0, \infty)$

5.1.2

(a) No. Try the horizontal line test in a small open set containing a.

(b) Yes. Your two choices should involve two different open sets containing b.

(c) Yes. Your local inverses should be different from your local inverses for the previous part.

5.1.3

(a) None. F and G are not invertible and H is invertible but not C^1.

(b) Only F is a local diffeomorphism at 0.

(c) All three are local diffeomorphisms at 1.

5.1.4 False. Consider $n = 1$ and $F(x) = x^3$.

5.1.5 All four are necessarily false.

5.1.6 Linear maps are C^1. Which maps do you need to check satisfy this property?

5.1.7 This is a direct application of the definition of a local diffeomorphism.

5.1.8

(a) Line 1. What are the domain and range of arcsin? Note that "inverses" of trigonometric functions are inverse functions of specific restrictions of the trigonometric functions.

(b) Line 3. Note that $\arcsin(0) = 0$ which is not in U.

(c) A choice of G that works is $G(x) = -\arcsin x + \pi$.

5.1.9 You will need to consider an arbitrary open ball containing $\frac{\pi}{2}$ and show that the restriction of F to such a ball is not a diffeomorphism.

5.1.10

(a) To conclude that a bijective map $H = F^{-1}$ is a diffeomorphism you have to show that both H and H^{-1} are C^1.

(b) Line 5. $(G \circ F)^{-1} \neq G^{-1} \circ F^{-1}$.

5.1.11

1. Use Theorem 2.7.8 in two different ways.
2. Let $\{y(k)\}_k$ be a sequence in $F(S)$ such that $y(k) \to y$ for some $y \in \mathbb{R}^n$. For each k there exists $x(k) \in S$ such that $F(x(k)) = y(k)$. Now, take the inverse, take limits, and use that S is closed. If you do not apply the inverse, then your argument is missing something.

5.1.12

(a) $F|_A$ and $F|_C$ appear to be diffeomorphisms.

(b) $F|_A$ and $F|_C$.

(c) Are all points in U mapped to distinct points?

(d) You will need to show that every open set containing $(0, \theta_0)$ contains multiple points which are mapped to the same point by F.

5.1.13

(a) $S = (0, \infty) \times (-\frac{\pi}{2}, \frac{\pi}{2})$

(b) $(F|_U)^{-1}(x, y) = \left(-\sqrt{x^2 + y^2}, \arctan(\frac{y}{x})\right)$

(c) $S \cup U$

(d) $(F|_V)^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, \pi + \arctan(\frac{y}{x})\right)$

(e) $(F|_W)^{-1}(x, y) = \left(-\sqrt{x^2 + y^2}, \frac{\pi}{2} - \arctan(\frac{x}{y})\right)$

(f) $S \cup U \cup V \cup W$

5.1.14 You will need to consider an arbitrary open ball containing $(0, \theta_0, z_0)$. Which points in such a ball will be mapped to the same point by F?

5.1.15 Follow the same procedure in the previous question for the specified points.

5.2.1 Without calculating any determinants, you should be able to see that (c), (e), and (f) are invertible.

5.2.2 This is a direct application of the inverse function theorem and the previous question. Only F is not a local diffeomorphism at its given point.

5.2.3

(a) True.

(b) False. See Example 5.2.12.

5.2.4

(a) Explain why $DF(p) = \begin{bmatrix} 0 & \frac{\pi}{4} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$ is invertible. Then apply the inverse function theorem.

(b) For some open set $U \subseteq \mathbb{R}^2$ containing p, you should find that $DG(u,v) = \frac{1}{x\cos x + y\sin y}\begin{bmatrix} \sin y & x \\ \cos x & -y \end{bmatrix}$ for $(x,y) \in U$ and $(u,v) = F(x,y)$.

5.2.5

(a) Explain why $DF(p) = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ is invertible. Then apply the inverse function theorem.

(b) $DG(1,0) = \begin{bmatrix} -1/2 & 0 \\ 0 & 1 \end{bmatrix}$.

5.2.6

(a) Explain why $DF(p)$ is invertible. Then apply the inverse function theorem.

(b) Indeed, this computation is a huge mess. Remember it is only a bunch of row reduction and nothing more, so it is a *feasible* huge mess.

5.2.7 $DF(a,b) = \begin{bmatrix} 2a & 2b \\ b & a \end{bmatrix}$. What happens when $a = \pm b$?

5.2.8

1. You will need to show that $\det DF(a,b,c) = (a-b)(a-c)(b-c)$.
2. Compare your answer with WolframAlpha. Yuck.

5.2.9

(a) Cameron needs to be more careful with U.

(b) Alisa claims that the IFT provides an inverse. Is this true?

(c) Abbigael demonstrates a misunderstanding of the definition of a local diffeomorphism.

(d) Sarah does not know whether $G(y)$ is defined at $y = 3$. The IFT does not give an explicit open set, so Sarah cannot decide how close y must be to e for $G(y)$ to be defined.

5.2.10 When does the derivative vanish?

5.2.11 $DF(r,\theta,z) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$. What happens when $r = 0$?

5.2.12 $\det DF(\rho,\theta,\phi) = -\rho^2 \sin\phi$. What happens when $\rho = 0$ or when ϕ is an integer multiple of π?

5.2.13

(a) $DF(a)$

(b) dF_a

(c) $\partial_1 F(a), \dots, \partial_n F(a)$

5.2.14

(a) $L : \mathbb{R}^n \to \mathbb{R}^n$ is given by $L(y) = G(F(a)) + DG(F(a))(y - F(a))$ for $y \in \mathbb{R}^n$.

(b) Show that $(L \circ \ell)(x) = x$ and $(\ell \circ L)(x) = x$. Apply the inverse function theorem and the fact that G is a local inverse of F at a.

5.2.15

(a) The left plot should have (a, b) and the vectors e_1, e_2, which are transformed under F in the right plot.

(b) The left plot should have $(0, b)$ and the vectors e_1, e_2. Note the entire vertical line is mapped to $(0, 0)$.

(c) The blanks should be filled with "$e_1, \dots, e_n$", "$dF_a(e_1), \dots, dF_a(e_n)$", and "linearly independent", in that order.

5.3.1

(a) $x = 0$.

(b)
i) 16
ii) 0
iii) 0

5.3.2 Try to do this by looking at square submatrices of A or its reduced row echelon form.

(a) Yes.

(b) Yes.

(c) No.

5.3.3

(a) This occurs at C and E.

(b) This occurs at B and E.

5.3.4

(a) No. ϕ_A does not have domain $V = (-\infty, \infty)$.

(b) No. ϕ_B does not have codomain $W = (0, \infty)$ and also ϕ_B is not C^1 at $x = \pm 1$.

(c) None. The choice of V is too wide. To justify, proceed by contradiction: if such a map ϕ exists, consider the point $(x, y) = (2, \phi(2)) \in V \times W$ and show one side of the $\iff$ is necessarily true whereas the other side is necessarily false.

5.3.5 (a), (c), and (e) can lead to a correct proof.

5.3.6 (b), (c), (e), and (f) can lead to a correct proof.

5.3.7

(a) $C_1 \neq 0$

(b) $C_1 D_2 \neq D_1 C_2$

5.3.8 Try to do this by looking at submatrices of A or its reduced row echelon form. The answer is no to (a), yes to (b) and (c).

5.3.9

(a) The rank is limited by the number of columns and by the number of rows.

(b) The rank of a matrix is the dimension of the space spanned by its columns. Write out the equation $Ax = 0$ in terms of $x_1, \ldots, x_{n+k}$ and the columns of A.

5.3.10 Consider modifications of the sets given in 5.3.5(a).

5.3.11 Consider modifications of the sets given in 5.3.6(f).

5.3.12

(a) Take a look back at Definition 5.3.7. What things need to be checked? Where is Maryam using her particular choice of V and W?

(b) Fabian needs to be more careful when he tries to negate Definition 5.3.7. What is the correct negation?

(c) Semeon's argument suggests that x can be defined as a function of y near $(0,0)$ but that function is not C^1. Is this really the case?

5.3.13

(a) $V = (-1,1)$, $W = (0,\infty)$

(b) $V = (-1,1)$, $W = (-\infty,0)$

(c) You will need to show that two sets are equal (one of them being the graph of your chosen function). Try to use set notation carefully.

5.3.14

(a) One possible choice is $V = (0,\infty)$, $W = (-\infty,0)$.

(b) This proof is tricker than you may initially think. Make sure you take arbitrary V and W. A picture and a short calculation will help.

5.3.15 Use the sets defined in 5.3.6(f).

5.4.1 The answers to (a) and (c) are yes. For (b), the implicit function theorem does not apply and we cannot know definitively without further (difficult) exploration.

5.4.2 This boils down to checking invertibility of submatrices of $DF(p)$.

(a) Yes. The submatrix corresponding to the partial derivatives $\frac{\partial F}{\partial x_1}$ and $\frac{\partial F}{\partial x_2}$ is invertible.

(b) Yes. The submatrix corresponding to the partial derivatives $\frac{\partial F}{\partial x_1}$ and $\frac{\partial F}{\partial x_3}$ is invertible.

(c) Inconclusive since implicit function theorem provides only a sufficient condition, not a necessary one. The submatrix corresponding to the partial derivatives $\frac{\partial F}{\partial x_2}$ and $\frac{\partial F}{\partial x_4}$ is not invertible.

(d) Implicit function theorem never provides an explicit formula for f, g, or h. This can be a very difficult problem in general.

5.4.3

(a) These should correspond to the equation $F(x, y) = 0$, the variables $(x, y) \in \mathbb{R}^n \times \mathbb{R}^k$, the condition that $\frac{\partial F}{\partial y}(a, b)$ be an invertible matrix, the variables $x_1, \ldots, x_n$, and the variables $y_1, \ldots, y_k$, in that order.

(b) These should correspond to the variables $(x, y) \in \mathbb{R}^n \times \mathbb{R}^k$, the condition that $\frac{\partial F}{\partial y}(a, b)$ be an inverible matrix, the variables $y_1, \ldots, y_k$, and the variables $x_1, \ldots, x_n$, in that order.

5.4.4

(a) $C_1 \neq 0$.

(b) $(A_1, B_1, C_1, D_1) \neq (0, 0, 0, 0)$.

(c) $C_1 D_1 \neq C_2 D_1$.

(d) (A_1, B_1, C_1, D_1) and (A_2, B_2, C_2, D_2) are linearly independent.

5.4.5 The answer is yes to the last two questions.

5.4.6

(a) You must check that the 1×1 matrix $\left[\dfrac{\partial F}{\partial y}(2,2)\right]$ is invertible.

(b) For some open sets $V \subseteq R$ and $W \subseteq \mathbb{R}$ such that $(2, 2) \in V \times W$, the function $\phi : V \to W$ satisfies $\phi'(x) = x/\phi(x)$ for $x \in V$.

5.4.7

(a) You will need to show that $\frac{\partial F}{\partial y}(-1, -1)$ is nonzero.

(b) For some open sets $V \subseteq \mathbb{R}$ and $W \subseteq \mathbb{R}$ such that $(-1, -1) \in V \times W$, the function $\phi : V \to W$ satisfies

$$\forall x \in V, \quad \phi'(x) = -\frac{x + \pi \sin(\pi(x + y))}{y + \pi \sin(\pi(x + y))} \quad \text{where } y = \phi(x)$$

5.4.8

(a) You will need to show that $\frac{\partial F}{\partial z}(1/2, 0, 1)$ is nonzero.

(b) For some open sets $V \subseteq \mathbb{R}^2$ and $W \subseteq \mathbb{R}$ such that $(1/2, 0, 1) \in V \times W$, the function $\phi : V \to W$ satisfies

$$\forall (x, y) \in V, \quad \nabla\phi(x, y) = \left(\frac{z - 2yz^4 e^{2xyz}\sqrt{1 - x^2}}{2xyz^3 e^{2xyz}\sqrt{1 - x^2} - 2\arcsin(x)\sqrt{1 - x^2}}, \frac{-2xz^4 e^{2xyz}}{2xyz^3 e^{2xyz} + 2\arcsin(x)} \right)$$

where $z = \phi(x, y)$.

5.4.9

(a) You will need to show that the matrix $\frac{\partial(F_1, F_2, F_3)}{\partial(w, y, z)}$ is invertible when evaluated at p.

(b) For some open sets $V \subseteq R^2$ and $W \subseteq \mathbb{R}^3$ such that $(1, 1) \in V$ and $(0, -1, 1) \in W$, the function $\phi : V \to W$ satisfies

$$\forall (v, x) \in V, \quad D\phi(v, x) = -\begin{bmatrix} 2vw & v - z^3 & 2x^2 z - 3yz^2 \\ xy^2 + 2wx^2 & 2wxy - 2xyz & -xy^2 \\ -3w^2 z & vx & -w^3 \end{bmatrix}^{-1} \begin{bmatrix} w^2 + y & 2xz^2 \\ 0 & wy^2 + 2w^2 x - y^2 z \\ xy & vy \end{bmatrix}$$

where $(w, y, z) = \phi(v, x)$. If you have the above expression, then that is enough. Optionally, you can simplify this expression with a very messy calculation or via a computer algebra system.

5.4.10

(a) You will need to show that the 2×2 matrix $\frac{\partial(F_1,F_2)}{\partial(x,y)}$ is invertible when evaluated at $(w,x,y,z) = (3,-1,0,2)$.

(b) For some open sets $V \subseteq \mathbb{R}^2$ and $W \subseteq \mathbb{R}^2$ such that $(3,2) \in V$ and $(-1,0) \in W$, the function $\phi : V \to W$ satisfies

$$D\phi(x,z) = -\begin{bmatrix} -3x^2 & 3w^2z+4y \\ -2xy^2 & -2x^2y+3z^4 \end{bmatrix}^{-1} \begin{bmatrix} 6wyz & 3w^2y+1 \\ 3w^2 & 12yz^3 \end{bmatrix}$$

where $(x,y) = \phi(w,z)$. If you have the above expression, then that is enough. Optionally you can simplify this expression with a very messy calculation or via a computer algebra system.

5.4.11

(a) Can we apply the converse of IFT?

(b) The argument does not check that the point in question belongs to the zero set of the function.

(c) Does the IFT give a specific solution and domain?

5.4.12 z can be expressed in terms of the others.

5.4.13 What functions should you choose, and for which points?

5.4.14 Use the same strategy as the previous problem.

5.4.15 Choose $F(x) = ||x||^2 - 1$. What property of ∇F do you need to check?

5.4.16

(a) The answer is yes if $\dfrac{\partial f}{\partial y} \neq 0$.

(b) The answer is yes if $\dfrac{\partial g}{\partial x_i} \neq 0$ for some $i \in \{1,\ldots,n\}$.

(c) The IFT can be applied if $\dfrac{\partial h}{\partial y} \neq 0$.

(d) The answer is yes if $\dfrac{\partial F}{\partial y}(a,b) := \left[\dfrac{\partial F_i}{\partial y_j}(a,b)\right]_{i,j}$ is invertible.

(e) The answer is yes if $\left[\dfrac{\partial H_i}{\partial x_j}(a,b)\right]_{i,j}$ is invertible.

5.4.17 First, show that $DH(x,y) = \begin{bmatrix} I_n & 0_n \\ \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \end{bmatrix}$. Then apply the inverse function theorem to solve for y in terms of x. There are a lot of technical details in a well written proof. Be careful with your domains and codomains.

5.5.1 Only (b) is not a full rank matrix.

5.5.2

(a) The blanks should be filled by 137 and 100, in that order.

(b) i), iii), iv)

5.5.3

(a) True.

(b) False. The unit circle is a counterexample.

(c) True.

(d) False. You need another condition for your map.

5.5.4

(a) False.

(b) True.

(c) True. What happens if $\nabla f(p) = 0$?

(d) False.

5.5.5 S is a smooth manifold of dimension 2.

5.5.6

(a) Use Theorem 5.5.3.

(b) $\{(x, y, z) \in \mathbb{R}^3 | -x + 4z = 0\}$

(c) $\{(x, y, z) \in \mathbb{R}^3 | -x + 4z - 26 = 0\}$

5.5.7

(a) You will need to check that the gradient is nonvanishing on S.

(b) $\{(x, y, z) \in \mathbb{R}^3 | x + 2y + 3z = 14\}$

5.5.8

(a) Note that the rank of $DF(p)$ is 2. The dimension of S should be $4 - 2 = 2$ by Theorem 5.5.7.

(b) There are several valid answers since any matrix obtained from $DF(p)$ via elementary row operations would give equivalent answers. One possible answer is

$$T_pS = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -\frac{1}{2} \end{bmatrix} \right\}$$

To see that this equals $\{v \in \mathbb{R}^4 : DF(p)v = 0\}$, remember you can use the row reduced echelon form of $DF(p)$.

(c) Your set should be equivalent to

$$p + T_pS = \left\{ (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \\ x_3 - p_3 \\ x_4 - p_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

5.5.9

(a) Your function should be of the form $F : \mathbb{R}^4 \to \mathbb{R}^2$.

(b) You should compute $DF(1,0,0,1) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}$, perhaps with the rows swapped.

(c) What is the rank of the matrix in the previous part?

(d) $\{(-1,1,0,0),(0,0,2,-1)\}$ is a basis for T_pS.

5.5.10

(a) One good choice of function f would be $f(x,y,z) = x^2 + 2y^3 - 3xyz$. You should calculate $\nabla f(1,1,1) \neq 0$.

(b) $\{(x,y,z) \in \mathbb{R}^3 : -(x-1) + 3(y-1) - 3(z-1) = 0\}$

5.5.11 The theorem shows that (a), (c), and (d) are smooth manifolds. Only (b) is not a smooth manifold; it is a cone. However, the theorem cannot make this conclusion about (b). You would need to prove it by definition.

5.5.12

(a) Consider the function $f(x) = ||x||^2 - 1$. Try to show that $\nabla f(x) = 2x$.

(b) $p + T_p(S^{n-1}) = \{x$
$in\mathbb{R}^n : p \cdot x = 1\}$

5.5.13

(a) This should be a condition on the gradient of f. At which points should it be nonvanishing?

(b) Consider the function $f(x,y) = y - \phi(x)$.

(c) Choose your function $f : \mathbb{R}^2 \to \mathbb{R}$ so that the hyperbola is a level set of f. Why doesn't the gradient of f vanish on the hyperbola?

(d) Does Corollary A give a necessary and sufficient condition?

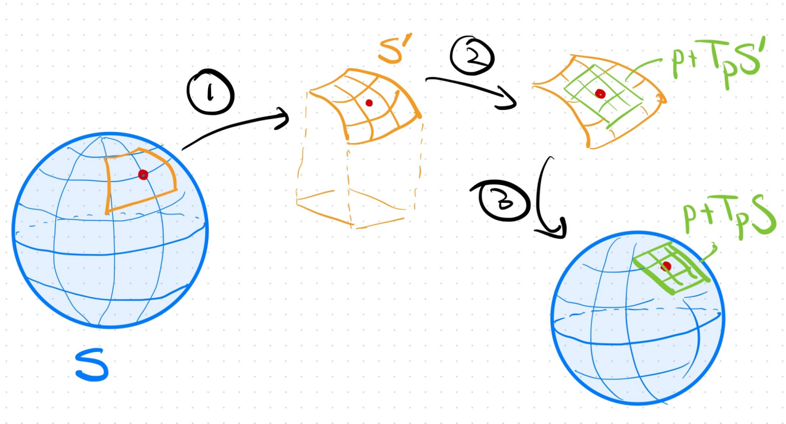
5.5.14 Consider the function $G(x,y,z) = z - f(x,y)$ and show that $DG(x,y,z) = \begin{bmatrix} -\partial_1 f(x,y) & -\partial_2 f(x,y) & 1 \end{bmatrix}$.

5.5.15 Generalize your argument from previous problem.

5.5.16 One good choice is $F(x,y,z) = (x^2 + y^2 + z^2, (x-1)^2 + (y-2)^2 + (z-3)^2)$. Then calculate the determinants of all 2×2 submatrices of $DF(x,y,z)$.

5.5.17 This proof is very short. Remember that the Jacobian is the transpose of the gradient in this $k = 1$ case.

5.5.18 The three stages should correspond to the three components of this diagram.



5.6.1 Only (a) is true.

5.6.2

(a) None are local extrema of f.

(b) Only b is a critical point of f.

(c) a and c are local extrema of f on S.

(d) a, b, and c are solutions to the Lagrange multipliers system.

5.6.3

(a) A is a local minimum, B, D are local maxima, and C is neither.

(b) P, Q, S

5.6.4

(a) Your labelled points should be roughly $(-0.5, 0.5)$ and $(0.5, -0.5)$. These are only approximate.

(b) ∇g should be parallel to ∇f at both points. One must point in the same direction as ∇f (and hence $\lambda > 0$ there), and one must point in the opposite direction as ∇f (and hence $\lambda < 0$ there). There are two valid ways to satisfy these requirements.

5.6.5

(a) Approximately 16000.

(b) Only $\nabla f(R)$ should point orthogonal to the constraint curve $g(x, y) = 6400$ and towards the northeast. Note $\nabla f(Q)$ points roughly north, and $\nabla f(P)$ points roughly east.

(c) Only R.

5.6.6

(a) The warmest is 47 degrees Celsius, and the coldest is approximately 41.7 degrees Celsius.

(b) The warmest occurs near $(4, 3)$ and the coldest occurs near $(5.8, 0)$.

(c) You will only find $(4, 3)$. To explain why, notice the domain $U \subseteq \mathbb{R}^2$ chosen in the method of Lagrange multipliers must be open.

5.6.7

(a) a should be below the x-axis and b above. Note $f(a) = 3$ and $f(b) = -3$.

(b) The tangent line is a horizontal line passing through a.

(c) The tangent line is a horizontal line passing through b.

(d) $\nabla f(a)$ and $\nabla f(b)$ both point down.

(e) Either $\nabla g(a)$ points up and $\nabla g(b)$ points down, or $\nabla g(a)$ points down and $\nabla g(b)$ points up. You need more information about g to decide.

For all optimization problems below, remember to justify your calculations.

5.6.8 The maximum is 16 and the minimum is -16.

5.6.9 The minimum value is $4\sqrt{2}$.

5.6.10 There is one local maximum and one local minimum with values $\pm 2\sqrt{21}$.

5.6.11 The maximum and minimum values are $\pm||a||$. You should be able to do this in "vector notation" without referring to coordinates.

5.6.12 This function has no global maximum on these constraints. Its global max is $4\sqrt{2}-6$.

5.6.13 You should find five solutions to the Lagrange system of equations resulting in a maximum value of $14-3\sqrt{2}$.

5.6.14

(a) Are all solutions to the Lagrange multiplier system necessarily maxima? Is the set S compact?

(b) They need to verify the existence of global extrema on S, and they need to ensure S is a smooth manifold. How can they do both of these things?

(c) They found the maximum on S, not the minimum. They must check the points $(2,0)$ and $(-2,0)$ separately. Notice the method of Lagrange multipliers requires an **open** set U to be chosen for the domains of f and g.

(d) A minimum (and maximum) on S must exist since the S is compact and f is continuous. They must check the points $(7/2,0)$ and $(0,7/3)$ separately as the method of Lagrange multipliers will always miss these points. Notice you must choose an **open** set U for the domains of f and g when applying Lagrange.

5.6.15 One way to set up this problem is to let (x,y) be the coordinates for one of the vertices of the rectangle where $(0,0)$ is the center of the circle. However you choose x and y, you should end up with the maximum perimeter being $4\sqrt{2}r$. It will require some careful justification to show that this is the maximum.

5.6.16

(a) You should optimize $C(K,L) = K+L$ over the domain $S = \{(K,L) \in \mathbb{R}^2 | P(K,L) = 20000\}$.

(b) The Lagrange multiplier system is solved by $K = \frac{4000}{4^{4/5}}$, $L = 4^{1/5} \cdot 4000$.

(c) Your picture should reflect the ideas of a lemma of the EVT for unbounded sets.

5.6.17 The maximum satisfaction is $\frac{3^{3/4}c}{4p_1^{1/4}p_2^{3/4}}$. Solving the system might take some messy algebra but it simplifies nicely after plugging in the values. To make a fully rigorous argument, you will need to check the points $(x,y) = (0, c/p_2)$ and $(c/p_1, 0)$ separately. Lagrange will miss these points.

5.6.18

(a) In your plot, $\nabla g(a)$ should be orthogonal to $a + T_a S$.

(b) Make sure to indicate the direction of motion of γ, which should be the same direction as v.

(c) h attains its maximum value $f(a)$ at 0, which should correspond to the point $(a, f(a))$ in your 3D plot.

5.7.1

- Identify possible extrema in the interior of S by looking for critical points.
- Identify possible extrema on the boundary of S, either by parametrizing the boundary or by using the method of Lagrange multipliers.
- Ensure a global extremum on S exists using the global extreme value theorem or limits.
- Compare values of the function at possible extrema to find the global maximum and minimum.

5.7.2

(a) Use the global extreme value theorem.

(b) Local extrema on the interior occur at critical points.

(c) Local extrema on the boundary occur at solutions to the Lagrange multiplier system.

(d) Have all possibilities been checked? If so, which has the largest value?

5.7.3

(a) Arthur's strategy would need at least 2 (and as many as 5) different optimization problems. He can start by using the spherical coordinates map $g : \mathbb{R}^2 \to \mathbb{R}^3$ and optimize $f \circ g$ on the 2-dimensional rectangle $[0, 2\pi] \times [0, \pi] \subseteq \mathbb{R}^2$. The interior can be handled with the local EVT. The boundary of the rectangle must again be parametrized with 4 pieces. This totals to 5 optimization problems. Alternatively, notice that the spherical coordinates map takes the boundary and maps the entire piece onto the curve $C = \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, y = 0, x \geq 0\}$. You can instead parametrize this curve and solve only 1 more optimization problem over $\mathbb{R}$ instead, which totals to 2 optimizations.

(b) Cameron's strategy would need only 1 step. Define $g : \mathbb{R}^3 \to \mathbb{R}$ by $g(x,y,z) = x^2 + y^2 + z^2$ so $S = g^{-1}(\{1\})$. You can apply Lagrange directly and find all solutions.

(c) Both will succeed but Cameron's strategy seems much more efficient.

5.7.4

(a) Amy's strategy would need at least 2 different optimization problems. She can start by using the graph map $g : \mathbb{R}^2 \to \mathbb{R}^3$ given by $g(x,y) = (x, y, \sqrt{1-x^2-y^2})$ and optimize $f \circ g$ on the disk $D = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \le 1\}$. The interior can be handled with the local EVT. The boundary of the disk must be parametrized with 1 piece. This totals to 2 optimization problems.

(b) Victoria's strategy would need at least 2 different optimization problems. There are many choices of constraint she can use. For example, she can use $g : U \to \mathbb{R}$ given by $g(x,y,z) = x^2+y^2+z^2$ with $U = \mathbb{R}^3$ or with $U = \mathbb{R}^2 \times (0,\infty)$. The logic of the argument changes depending on her choice, but both can be successful. Either way, she must also separately consider the circle $C = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2 = 1, z = 0\}$. Since Victoria wants to use Lagrange, she must introduce two constraints to optimize f over C, namely $h_1 : \mathbb{R}^3 \to \mathbb{R}$ given by $h_1(x,y,z) = x^2+y^2-1$ and $h_2 : \mathbb{R}^3 \to \mathbb{R}$ given by $h_2(x,y,z) = z$. This implies

$$C = \{(x,y,z) \in \mathbb{R}^3 : h_1(x,y,z) = 0, h_2(x,y,z) = 0\}$$

Victoria will be able to successfully apply Lagrange with these choices, but it is going to be a 5 variable system of nonlinear equations. Yuck.

(c) Both choices are pretty messy, so this will be subjective. Optimization gets messy but that's life for you.

5.7.5 This question is asking you to articulate the optimization strategy we've seen so far: handle the interior and boundary separately. The boundary is the union of 3 components that need to be handled separately: a disk, a hemisphere, and a circle where they intersect. They all have multiple strategies that can be used to optimize.

5.7.6

1. One approach is to first eliminate y, yielding:

$$x + 2 = 4\lambda^2 x$$
$$x^2 + 4\lambda^2 x^2 = 1$$

Combining these gives

$$x^2 + (x+2)x = 1$$

so $x = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}$. Of these, only $x = \frac{\sqrt{3}}{2} - \frac{1}{2}$ can occur on the circle. This gives $y = \pm\sqrt{\frac{\sqrt{3}}{2}}$.

2. $f(x,y)$ has no critical points on the interior of the disk, so it has no extrema on the open ball.

5.7.7 You're finding the minimum of $S(w,l) = \frac{\sqrt{3}}{2}w^2 + 3wl$ with the constraint that $\frac{\sqrt{3}}{4}w^2 l = 250$. Solving this gives $w = 10\text{cm}, l = 10/\sqrt{3}\text{cm}$.

5.7.8

(a) The Lagrange multiplier system in this case is:

$$1 = \frac{60\lambda}{x^{0.4}}$$
$$1 = \frac{30\lambda}{y^{0.5}}$$
$$100x^{0.6} + 60y^{0.5} = 10{,}000{,}000$$

Using a computer, we find that the only reasonable solution of this system, is approximately $y = 1,150,450$ computers. The corresponding value of x is approximately $x = 213,138,000$ phones, while the corresponding value of λ is $\lambda \approx 35.7$.

(b) The change in maximum profit if we increase our spending by \$1,000,000 is approximately $1,000,000\lambda$. So, we approximate that profit will increase by \$35.7 billion.

(c) Solve another Lagrange multiplier system with the new production budget.

5.7.9 The Lagrange system is quite nasty to solve (perhaps best done by a computer). You should find a maximum at $(1000, 600, 500)$.

5.7.10 You should find that the minimum value is $1/n$. It will not be easy to show that this is actually the minimum. One method is to first show that this is a minimum of the function on some compact subset and then show that the function is larger outside of this subset.

5.7.11
(a) Lines 9, 10, and 12 use the global EVT.

(b) Line 10 requires the local EVT.

(c) Line 9 requires the method of Lagrange multipliers. Note Line 4 sets up the system but no conclusion is made, i.e. the theorem itself is not yet applied.

(d) Your answer should involve lines 2, 9, and both EVTs.

5.7.12
1. Apply the global extreme value theorem.
2. For matrices B, C, recall that $(BC)^T = C^T B^T$ (provided that B and C are the right dimensions).
3. Note that $\lambda \nabla g(x) = 2\lambda x$.

5.7.13 Notice that the gradient of a linear function is never 0, unless that function is constantly 0. This means you only need to consider the boundary for any polygon.

(c) f necessarily attains its extrema on the vertices of S.

(d) What are you going to do if you're working in $\mathbb{R}^{1000000}$ with 500,000,000 constraints? Is the procedure we've been using to check for global extremum reasonable in this situation?

A.6. Approximations

6.1.1
(a) $f(-1, 2)$ is negative.

(b) i) zero ii) negative

(c) i) positive ii) negative iii) cannot be determined iv) cannot be determined

6.1.2
(a) $f(a)$ is positive

(b) i) negative ii) positive

(c) i) negative ii) negative iii) positive iv) positive

6.1.3 Pay attention to the size and direction of the arrows along vertical (or horizontal) lines.

$f_x(0,0) = f_y(0,0) = 0$ and $f_{xx}(0,0) > 0$ and $f_{yy}(0,0) < 0$

$g_x(0,0) = g_y(0,0) = 0$ and $g_{xx}(0,0) > 0$ and $g_{yy}(0,0) > 0$

6.1.4 For the mixed partials, it is easier to analyze (a) and (c) first. How does $\frac{\partial h}{\partial t}(x,5)$ vary with x? How does $\frac{\partial h}{\partial t}(1,t)$ vary with t?

(a) (a) negative (b) zero

(b) (a) zero (b) zero (c) positive (d) zero

(c) (a) positive (b) positive (c) zero (d) zero

(d) (a) zero (b) zero (c) positive (d) positive

6.1.5 $f_x = 2x + y^2$, $f_y = 2xy + 3y^2$,
$f_{xx} = 2$, $f_{yy} = 2x + 6y$, $f_{xy} = f_{yx} = 2y$

6.1.6 $f_x = ye^{xz}(xz+1)$, $f_y = xe^{xz}$, $f_z = x^2ye^{xz}$,
$f_{xx} = yze^{xz}(xz+2)$, $f_{yy} = 0$, $f_{zz} = x^3ye^{xz}$, $f_{xy} = e^{xz}(xz+1)$, $f_{xz} = xye^{xz}(xz+2)$, $f_{yz} = x^2e^{xz}$

6.1.7 $Hq = \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix}$

6.1.8 $Hq = 2A$

6.1.9 You should find that one of the mixed partials is 0 and the other is 1. Why does Clairaut's theorem not apply?

6.1.10 Apply Theorem 3.3.10 iteratively.

6.1.11 Same idea as the previous question.

6.1.12

(a) $f(0,0) = A$

(b) Try to differentiate P and then substitute $(x,y) = (0,0)$.

(c) If you plug in your previous result, then notice this statement is exactly the same as the definition of differentiability of f.

(d) Differentiate P twice and then substitute $(x,y) = (0,0)$.

(e) It may help to note that the limit relationship for the degree 0 polynomial P is $\lim_{(x,y)\to(0,0)} [f(x,y) - P(x,y)] = 0$.

(f) Use part (d).

6.1.13 Given a fixed point and fixed time, the change over time in the slope of the string at the fixed point is equal to the change over space in the vibration velocity of the string at the fixed time.

6.2.1 Only (b) and (c) are false.

6.2.2 $\partial_3\partial_1\partial_2 f(x,y,z) = \lim_{h\to 0} \dfrac{\partial_1\partial_2 f(x,y,z+h) - \partial_1\partial_2 f(x,y,z)}{h}$

6.2.3

(a) $g_x = 2x - y$, $g_y = -x + 6y$, $g_{xx} = 2$, $g_{yy} = 6$, and $g_{xy} = -1$. $\partial^\alpha g = 0$ for any $\alpha \in \mathbb{N}^2$ with $|\alpha| \geq 3$.

(b) $f_x = yze^{xz}$, $f_y = e^{xz}$, $f_z = xye^{xz}$
$f_{xx} = yz^2e^{xz}$, $f_{yy} = 0$, $f_{zz} = x^2ye^{xz}$, $f_{xy} = ze^{xz}$, $f_{xz} = (y + xyz)e^{xz}$, $f_{yz} = xe^{xz}$
$f_{xxx} = yz^3e^{xz}$, $f_{yyy} = 0$, $f_{zzz} = x^3ye^{xz}$, $f_{xxy} = z^2e^{xz}$, $f_{xxz} = (2yz + xyz^2)e^{xz}$, $f_{xzz} = (2xy + x^2yz)e^{xz}$, $f_{yzz} = x^2e^{xz}$, $f_{xyz} = (xz + 1)e^{xz}$, $f_{xyy} = f_{yyz} = f_{yyy} = 0$

6.2.4

(a) $\partial^{(1,1)} f = \frac{\partial}{\partial x}\frac{\partial}{\partial y} f = (2y + 2xy^3)e^{xy^2}$. The order is 2.

(b) $\partial^{(2,2,2)} f = 0$. The order is 6.

6.2.5 Figure out ways to use linearity and Clairaut's theorem to minimize work.

(a) If you take the derivative with respect to z three times you see that only the last term remains. Further taking an x derivative on that term will give 0.

(b) 12

(c) 12

6.2.6 $\dfrac{\partial^2 u}{\partial s^2}(0,0) = 2$ and $\dfrac{\partial^2 u}{\partial t^2}(0,0) = 2$ and $\dfrac{\partial^2 u}{\partial s \partial t}(0,0) = 3$. In general,

$$\begin{aligned}\frac{\partial^2 u}{\partial s^2} = {} & \frac{\partial^2 u}{\partial x^2}\Big(\frac{\partial x}{\partial s}\Big)^2 + \frac{\partial^2 u}{\partial y^2}\Big(\frac{\partial y}{\partial s}\Big)^2 + \frac{\partial^2 u}{\partial z^2}\Big(\frac{\partial z}{\partial s}\Big)^2 \\ & + 2\frac{\partial^2 u}{\partial x \partial y}\frac{\partial x}{\partial s}\frac{\partial y}{\partial s} + 2\frac{\partial^2 u}{\partial x \partial z}\frac{\partial x}{\partial s}\frac{\partial z}{\partial s} + 2\frac{\partial^2 u}{\partial y \partial z}\frac{\partial y}{\partial s}\frac{\partial z}{\partial s} \\ & + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2} + \frac{\partial u}{\partial z}\frac{\partial^2 z}{\partial s^2} \quad \text{and}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 u}{\partial s \partial t} = {} & \frac{\partial^2 u}{\partial x^2}\frac{\partial x}{\partial s}\frac{\partial x}{\partial t} + \frac{\partial^2 u}{\partial y^2}\frac{\partial y}{\partial s}\frac{\partial y}{\partial t} + \frac{\partial^2 u}{\partial z^2}\frac{\partial z}{\partial s}\frac{\partial z}{\partial t} + \\ & + \frac{\partial^2 u}{\partial x \partial y}\frac{\partial x}{\partial s}\frac{\partial y}{\partial t} + \frac{\partial^2 u}{\partial x \partial z}\frac{\partial x}{\partial s}\frac{\partial z}{\partial t} + \frac{\partial^2 u}{\partial y \partial z}\frac{\partial y}{\partial s}\frac{\partial z}{\partial t} \\ & + \frac{\partial^2 u}{\partial x \partial y}\frac{\partial x}{\partial t}\frac{\partial y}{\partial s} + \frac{\partial^2 u}{\partial x \partial z}\frac{\partial x}{\partial t}\frac{\partial z}{\partial s} + \frac{\partial^2 u}{\partial y \partial z}\frac{\partial y}{\partial t}\frac{\partial z}{\partial s} \\ & + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s \partial t} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s \partial t} + \frac{\partial u}{\partial z}\frac{\partial^2 z}{\partial s \partial t}\end{aligned}$$

6.2.7 Show that $\partial_2\partial_3\partial_1 f = \partial_2\partial_1\partial_3 f$ and $\partial_2\partial_1\partial_3 f = \partial_1\partial_2\partial_3 f$ using Clairaut's theorem for C^2 functions.

6.2.9

(a) For example, $p(x, y) = a_0 + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2$.

(b) There are three of them.

(c) There will be $k + 1$ of them.

(d) For example, $p(x,y) = a_0 + a_{100}x + a_{010}y + a_{001}z + a_{200}x^2 + a_{020}y^2 + a_{002}z^2 + a_{110}xy + a_{101}xz + a_{011}yz$.

(e) There will be six degree 2 terms.

(f) $\binom{N+3}{3} = \frac{(N+3)!}{3!N!}$ To see why, look up the "stars and bars" counting argument.

6.2.10 Try this one with $(a,b) = (0,0)$ first.

(a) Start with an arbitrary polynomial of degree 2 and calculate its coefficients using the above relations.

(b) Start with a polynomial $f(x,y) = \sum_{k=0}^{N} \sum_{i+j=k} a_{ij} x^i y^j$ and calculate its coefficients using the above relations.

(c) You should get

$$P_N(z) = \sum_{|\alpha| \le N} \frac{\partial^\alpha f(0,0)}{\alpha!} z^\alpha$$

where we sum over all multi-indices of degree less than or equal to N.

(d) $P_2(x,y) = y$

6.2.11

(a)

$$P_N(x,y) = \sum_{k=0}^{N} \sum_{i+j=k} \frac{1}{i!j!} \partial^{(i.j)} f(a,b)(x-a)^i (y-b)^j$$

(b)

$$P_N(x,y) = \sum_{k=0}^{N} \sum_{|\alpha|=k} \frac{1}{\alpha!} \partial^\alpha f(a,b)(x-a, y-b)^\alpha = \sum_{|\alpha| \le N} \frac{1}{\alpha!} \partial^\alpha f(a,b)(x-a, y-b)^\alpha$$

(c)

$$\begin{aligned} P_2(h_1,h_2) = & f(0,0) && (|\alpha| = 0 \text{ term}) \\ & + \frac{\partial f}{\partial x}(0,0)h_1 + \frac{\partial f}{\partial y}(0,0)h_2 && (|\alpha| = 1 \text{ terms}) \\ & + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0)h_1^2 + \frac{\partial^2 f}{\partial x \partial y}(0,0)h_1 h_2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0)h_2^2 && (|\alpha| = 2 \text{ terms}) \end{aligned}$$

(d) Note that

$$\frac{D_h^2 f(0,0)}{2!} = \frac{1}{2}\partial_1^2 f(0,0)h_1^2 + \frac{1}{2}\partial_2^2 f(0,0)h_2^2 + \partial_1^{(1,1)} f(0,0)h_1 h_2$$

6.3.1 Only (c) is false. Should there be any degree requirements for the polynomial P in (c)?

6.3.2

(a) $f(0.1,0,0.3) \approx 1.01$

(b) $f(0,0,0)=0$, $P_2(0,-1,0)=2$, $P_1(0,-1,0)=-2$

(c) $\frac{\partial f}{\partial x}(0,0,0)=1$ and $\frac{\partial^2 f}{\partial y^2}(0,0,0)=8$

(d) $\nabla f(0,0,0)=(1,2,3)$ and $Hf(0,0,0)=\begin{bmatrix} 2 & -3 & 0 \\ -3 & 8 & -7 \\ 0 & -7 & 0 \end{bmatrix}$

(e) iii) and v) could be equal to P_3.

6.3.3 (h) is necessarily false since P_2 has non-zero order 2 terms. There is not enough information to determine whether or not (g), since we do not know the 3rd Taylor polynomial. The rest are necessarily true.

6.3.4

(a) $P_2(x,y,z)=y+z+(x-1)y+(x-1)z$.

(b) $P_3(x,y)=1-xy^2$.

6.3.5 $||x-a||^M=||x-a||^N \cdot ||x-a||^{M-N}$ and $M-N \geq 1$.

6.3.6

(a) Line 3 is missing key details. If you correctly fill in the details, then you should need to make an argument involving a multivariable squeeze theorem.

(b) As in a previous problem, $P_2(x,y,z)=y+z+(x-1)y+(x-1)z$. Use a Taylor expansion for $\sin t$ at $t=0$.

6.3.7

(a) For the lefthand sum, each multi-index α consists of either a single 2 or two 1s and the rest are zeros. For the righthand double sum, notice that if $i=j$ then $x_ix_j=x_i^2$ appears only once, but if $i \neq j$ then x_ix_j appears twice (since $x_ix_j=x_jx_i$).

(b) What is $\displaystyle\sum_{\substack{\alpha\in\mathbb{N}^n \\ |\alpha|=1}} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha$?

(c) Remember that $Hf(a)$ is symmetric and $x-a$ is a vector.

(d) You now know that

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i-a_i)(x_j-a_j)=\frac{1}{2}(x-a)^T Hf(a)(x-a)$$

so what is the last thing to be shown in $P_2(x)$?

6.3.8

(a) $P_N(a+h)$ is the Nth order approximation of f at a and $||h||^N \varepsilon_{a,N}(h)$ is the error.

(b) It doesn't just tend to zero. It tends to zero faster than ______.

(c) After choosing $\varepsilon_{a,N}$, the proof is short. Be careful to justify every step and not violate any limit laws or write nonsensical expressions. Remember to define $\varepsilon_{a,N}$ at the origin separately.

6.3.9

(a) $\lim_{x \to a} \frac{f(x) - P_N(x)}{||x-a||^N} = 0$

(b) To get you started, the red point in the middle is $(a_1, a_2, f(a_1, a_2))$ and the blue point in the corner on the bottom surface can be labelled $(x, y, f(x, y))$.

(c) All four sentences are true. i), iii), and iv) relate to Taylor's theorem but only one is the most precise. ii) is not about Taylor's theorem because it does not refer to the limit $x \to a$. Instead, it refers to $N \to \infty$, which relates to whether a multivariable function is analytic, i.e. equal to its Taylor series. That is beyond the scope of this course.

(d) Your sentence should refer to the rate at which the errors decay and how this depends on the order of the approximation. Explain how you are observing this in the picture by describing specific surfaces.

6.4.1 Only (b) is true. Note that a critical point can only either be a local extremum or a saddle point.

6.4.2

(a) There is one positive eigenvalue and one negative eigenvalue

(b) There are two positive eigenvalues.

(c) There are two negative eigenvalues.

(d) There is one positive eigenvalue and one zero eigenvalue.

6.4.3 You should fill in the blanks with:

i) All eigenvalues are positive.
ii) All eigenvalues are nonnegative.
iii) All eigenvalues are negative.
iv) All eigenvalues are nonpositive.
v) There is at least one negative eigenvalue and at least one positive eigenvalue.

6.4.4

(a) The critical point is a saddle.

(b) The critical point is a local min.

(c) The critical point is a local max.

(d) The second derivative test cannot classify the critical point.

6.4.5

(a) For some $a > 0$, $q(x) = ax^2$ or $q(x) = -ax^2$ or $q(x) \equiv 0$ (the trivial form).

(b) $f(x) = x^3$ has a saddle point at 0 and has the trivial quadratic form $q(x) \equiv 0$ at 0.

(c) True. A saddle point will always have the trivial quadratic form $q(x) \equiv 0$.

6.4.6 The second derivative test is inconclusive.

6.4.7 a is a local min, b is a critical point for which the test is inconclusive, and c is a saddle point.

6.4.8

(a)
i) f has either a local min or a local max at a.
ii) f has a saddle point at a.
iii) The second derivative test is inconclusive.

(b) The second derivative test shows it is a saddle point.

6.4.9

(a) A should be a symmetric $n \times n$ matrix.

i) A has only positive eigenvalues.
ii) A has only negative eigenvalues.
iii) A has negative and positive eigenvalues but all are nonzero.
iv) A has at least one zero eigenvalue.

(b) Note case iv) for Corollary D does not appear in Corollary C. It is special to two dimensions.

i) $f_{xx}(a)$ and $f_{xx}(a)f_{yy}(a)-(f_{xy}(a))^2$
ii) $f_{xx}(a)$ and $f_{xx}(a)f_{yy}(a)-(f_{xy}(a))^2$
iii) $f_{xx}(a)f_{yy}(a)-(f_{xy}(a))^2$
iv) $f_{xx}(a)f_{yy}(a)-(f_{xy}(a))^2$

(c) $(-1,5/2)$ is a saddle point and $(3,-3/2)$ is a local minimum.

6.4.10 $(0,-1)$ is a saddle point. $(-2,-1)$ and $(2,-1)$ are local minima.

6.4.11 $(0,0,0)$ is the only critical point and the test is inconclusive.

6.4.12 a is the only critical point and it is a local maximum.

6.4.13

(a) The second derivative test only tells you local information.

(b) The second derivative test is inconclusive in this case.

6.4.14 If you know the second derivatives of the third Taylor polynomial, what can you say about the second derivatives of the function f itself? Justify your claim using theorems.

6.4.15

(a) I) corresponds to lines 2 and 3.
II) corresponds to lines 4 and 7.
III) corresponds to lines 5, 6 and 8.

(b) These are lines 4, 5, and 8.

6.5.1 All of them are true. (c) can be proved by induction. (e) is Taylor's theorem. (f) is Lemma 6.5.7.

6.5.3

(a) Your picture should show $D, \varphi(I), L$, and the endpoints of L. Be careful about which sets are open or closed.

(b) Your explanation should use that the endpoints of L are interior points of D.

(c) A proper justification should use the claim directly with $k = N+1$ as well as the fact that $\varphi(t) = a + th$ is continuous. It is not justified to simply say *"since φ is C^∞ and f is C^N, by the chain rule, g is C^N"*. This is not an immediate consequence. That is almost exactly what the proof of the line 4 claim shows.

(e) Apply the theorem with $h = g, J = I, a = 0$, and $x = 1$.

(f) Carefully study the statement of Lagrange's remainder theorem for 1 variable. Notice how you applied it for the interval $[0,1]$ inside I. What do you need to be true at the endpoints of $[0,1]$?

(h) Lines 10-11 correspond to the base case. Lines 12-14 correspond to the induction step. The inductive hypothesis is applied in line 12 but how exactly?

(i) Your justifications should include (in no particular order) something like "by Edwards II.2 Theorem 2.2", "by line 12", "since $\varphi(t) = a + th$", and "by the chain rule". It's okay if you have more detail.

6.5.4

(a) For the leftmost graph, the curve is the graph of g, i.e. the set of points $(t, y) \in \mathbb{R}^2$ such that $y = g(t)$. Finish labelling the rest. There is not too much to write on this one.

For the middle left graph, notice the surface is the graph of f, i.e. the set of points $(x, y, z) \in \mathbb{R}^3$ such that $z = f(x, y)$. Be careful to label the objects accordingly. The square beneath it is the set D. The centre red point in the square is a but the corresponding point on the surface should be $(a, f(a))$. Finish labelling the rest.

(b) For the middle right graph, it is not easy to know exactly where c lies so just draw it anywhere in the middle of the interval. The new purple curve should be written as $y = P(t)$ where $P(t)$ is a polynomial in t of degree N with coefficients $g^{(k)}(0)$ for $k = 0, 1, \ldots, N$. Label the rest, including the yellow line and its endpoints.

For the rightmost graph, label ξ in the square so it is consistent with where you put c in the left graph. Then draw $(\xi, f(\xi))$ on the original grey surface above. The new purple surface should be written as $z = Q(x, y)$ where Q is a polynomial in x, y with coefficients $\partial^{(i,j)} f(a)$ for $i + j \leq N$. Label the rest, including the yellow line and its endpoints.

(c) Toggle it so you only see the 1D animation and then the full 2D animation. Look at one surface at a time. Pay attention to the rate at which the error shrinks for each surface. Pretty neat, eh?

6.5.5

(a) $\partial^\alpha f$ is continuous on an open ball at a for every $\alpha \in \mathbb{R}^n$ with $|\alpha| = N$.

(b) Line 7 in the second last inequality.

(c) Start your proof by writing $\xi = a + \lambda(x - a)$ for some $\lambda \in [0,1]$.

(d) First, $|h^\alpha| = |h_1^{\alpha_1} \cdots h_n^{\alpha_n}| \leq ||h||^{\alpha_1 + \cdots + \alpha_n} = ||h||^{|\alpha|}$. Second, if $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ satisfies $|\alpha| = N$, then what are all possible values for α_1? Continue this idea. The estimate is overkill.

(e) Line 9 is modified by replacing "unique degree $\leq N$" with "smallest degree". Line 10 is modified by saying "Let P be the smallest degree polynomial satisfying this property. Such a polynomial exists and has degree $\leq N$ since P_N has degree $\leq N$ and satisfies this property."

A.7. Integrals

7.1.1

(a) This is a partition of R by Definition 7.1.8.

(b) Order is important! This is a partition of a different rectangle.

(c) This is a collection of rectangles associated to a partition, not a partition itself.

(d) Recall that all the values of the partition must be contained within the interval.

(e) Recall the definition of the trivial partition.

7.1.2

(a) The index set is $I = \{(1,1),(1,2),(2,1),(2,2)\}$.

(b) The associated subrectangles are indexed by I and are given by

$$R_{(1,1)} = [0,0.4]\times[2,3], \quad R_{(1,2)} = [0,0.4]\times[3,5], \quad R_{(2,1)} = [0.4,1]\times[2,3], \quad R_{(2,2)} = [0.4,1]\times[3,5].$$

7.1.3 (e) is the only one that is not a partition. By Definition 7.1.8, partitions of the rectangle in $\mathbb{R}^n$ come from n-tuples of partitions of intervals. See also Definition 7.1.5 for the $n = 2$ case.

(a,b,d) are regular partitions. The subrectangles created by a regular partition all have the same volume.

7.1.4

(a) True. Recall Theorem 7.1.10.

(b) True, by definition of the index set.

(c) False in general. This is only true if the partition is regular. Otherwise, the volumes of each subrectangle do not have to be equal.

(d) False. If the two subrectangles are adjacent, they share a boundary.

(e) True. Recall Theorem 7.1.10.

(f) False. Think back to (d). What if x is on the boundary of R_i?

7.1.5 (a,c,e,f) are refinements of P.
(b,c,d,e) are regular partitions of R.

7.1.6 Your sketch should look like the two partitions superposed. Your partition should have 9 subrectangles.

7.1.7 P_1, P_3, P_5, P_6 are refinements of P by Definition 7.1.15. The most convenient way to determine this is by verifying the inclusion of each of the two partitions of the one-dimensional interval.

P_2, P_3, P_4, P_5 are regular partitions because each of their subrectangles have the same volume.

7.1.8 $P''' = (\{1,3,5,7\},\{0,1,3,4\})$. The answer to the second question is yes. The union is assumed to be taken componentwise.

7.1.9
(a) True. Recall Definition 7.1.15.
(b) False. P and P' are not sets when $n \geq 2$.
(c) True. Recall Theorem 7.1.10(c).
(d) True.
(e) True. RecallTheorem 7.1.17.
(f) False.
(g) True. Recall Definition 7.1.19.

7.1.10
(a) $x_i = x_0 + i\Delta x$ and $\Delta x = \frac{b-a}{M}$
(b) $y_j = y_0 + j\Delta y$ and $\Delta y = \frac{d-c}{N}$
(c) MN and $\frac{(b-a)(d-c)}{MN}$

7.1.11
(a) $P_j = \{a_j, a_j + \frac{b_j - a_j}{N}, a_j + 2\frac{b_j - a_j}{N}, ..., a_j + (N-1)\frac{b_j - a_j}{N}, b_j\}$
(b) The number of indices is precisely n, the dimension of $\mathbb{R}^n$.
(c) Recall that the volume of a rectangle is the product of lengths of its sides.

7.1.12 There will be 2 telescoping sums.

7.1.13 Your geometric argument should use the observation that $\Delta x_i \Delta y_j = \text{area}(R_{ij})$. Your formal proof should start with something like

$$\sum_{i=1}^{M}\sum_{j=1}^{N} \Delta x_i \Delta y_j = \sum_{i=1}^{M}\Big(\Delta x_i \sum_{j=1}^{N} \Delta y_j\Big) = \cdots = \text{area}(R)$$

You will need to use that a sum is telescoping twice. Notice you have proved (7.1.6).

7.1.14 There will be n telescoping sums. To set up your calculation, you will need n indices which run from 1 to N to define your set, say $i_1, i_2, \ldots, i_n$. Start by trying to write down what an arbitrary rectangle in this set my look like, say $R_{i_1,\ldots,i_n}$. If you having trouble, try $n = 2$ to start.

7.1.15 $P = (\underbrace{Q, \ldots, Q}_{n \text{ times}})$ where $Q = \{\frac{i}{N} : i \in \{0, 1, \ldots, N\}\} = \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$ is a regular partition of $[0, 1] \subseteq \mathbb{R}$.

7.1.16
(a) $\sqrt{(b_1 - a_1)^2 + \cdots + (b_n - a_n)^2}$
(b) False. Assume one and prove the other does not hold.
(c) True. Your counterexample rectangle should be long and thin.
(d) $\frac{1}{N}\sqrt{(b_1 - a_1)^2 + \cdots + (b_n - a_n)^2}$

(e) Choose a regular partition as in the previous part and select N large enough in terms of $\delta, a_1, b_1, \ldots, b_n, a_n$.

7.1.17
(a) Explain why if $x \in [a, b]$ then there exists a minimum $i \in \{1, \ldots, k\}$ such that $x \leq x_i$. Then consider cases when $i = 1$ and when $i \geq 2$.
(b) Use properties of the interior with intersections.
(c) Use that $(A \cup B) \times C = (A \times C) \cup (B \times C)$ for sets A, B, C.
(d) Use properties of the interior with intersections and Cartesian products.

7.1.18
(a) This is the same as saying $\exists k \in I$ s.t. $i_k \neq j_k$
(b) Two of the three cases occur when $i_1 = j_1 \pm 1$. The third case is $|i_1 + 1| < j_1$
(c) Recall how interior behaves with respect to the inclusion of sets.
(d) See Lemma 2.1.11.

7.1.19 You can derive it from the transitivity of set inclusions.

7.1.20 This is false. A counterexample would be the partitions $P_1 = \{0, \frac{1}{2}, 1\}$ and $P_2 = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ of the interval $[0, 1]$.

7.1.21 Note that if rectangles R and R' satisfy $R' \subseteq R$, then the diameter of R' is smaller than or equal to the diameter of R.

7.2.1
(a) Lemma 7.2.8. Notice Lemma 7.2.6 only picks 1 partition.
(b) Lemma 7.2.7

7.2.2
(a) B and F represent an upper sum. Of the remaining four, one is a lower sum, one is neither an upper or a lower sum, and two are nonsense. Notice Δx_0 and Δy_0 are not defined.
(b) A corresponds to the Riemann sum.

7.2.3 (a) $L_p(f + g) \geq L_p(f) + L_p(g)$
(b) $L_p(\lambda f) = \lambda L_p(f)$ for any $\lambda > 0$
(c) $L_p(-f) = -U_p(f)$
(d) If $f \leq g$ on R, then Ł$_p(f) \leq L_p(g)$

7.2.4
(a) Your partition should have 4 subrectangles, each of which have different dimensions.
(b) $U_P(f) = 38.28$ and $L_P(f) = 21.72$. Take the partial derivatives of $f(x, y)$ to figure out the values of M_{ij} and m_{ij}.

7.2.5

(a) Your diagram should include at least the following notation: R_{ij}, M_{ij}, m_{ij}, x_0, x_{i-1}, x_i, x_N, y_0, y_{j-1}, y_j, y_N.

(b) Your answer may look different but remember there are equivalent ways to express the same sum.

$$U_{P_N}(f) = \sum_{i=1}^{N}\sum_{j=1}^{N}\Big(-\frac{i-1}{N}+6+\frac{9j}{N}\Big)\frac{3}{N^2} \qquad L_{P_N}(f) = \sum_{i=1}^{N}\sum_{j=1}^{N}\Big(-\frac{i}{N}+6+\frac{9(j-1)}{N}\Big)\frac{3}{N^2}$$

(c) $U_{P_N}(f) = 30+\frac{15}{N}$ and $L_{P_N}(f) = 30-\frac{15}{N}$. Recall $\displaystyle\sum_{i=1}^{N} i = \frac{N(N+1)}{2}$.

7.2.6

(a) $U_P(f) = 16$ for all partitions P of R.

(b) There are many valid choices. Without loss, you may assume $\varepsilon < 1$. One choice is to construct P_ε from the partition $\{0, \frac{\varepsilon}{3}, 2\}$ of the interval $[0,2]$ and the partition $\{3,7\}$ of the interval $[3,7]$.

7.2.7

(a) $S_P^*(f) = 145$

(b) Your argument should involve the partial derivatives of f. What are their signs? What does this tell you?

7.2.8 First choose a function representing the depth of the sand, then choose a partition of the sandbox. Calculate the upper or lower sums for your chosen function and partition.

7.2.9

(a) Consider the case $\lambda = -1$. Lemma 7.2.9(b) will help you fix the claim.

(b) Consider the functions $f(x) = x$ and $g(x) = -x$ on the trivial partition of $R = [0,1]$. Lemma 7.2.9(a) will help you fix the claim.

7.2.10

(a) You should include the notation R_{11}, R_{12}, R'_{11}, R'_{12}, R'_{21}, R'_{22}.

(b) Visually, you should see that the subrectangles of the partition P' fit inside those of P. Express this mathematically. We need the assumption that P' is a refinement of P. This is not necessarily the case if P' and P are arbitrary partitions.

(c) Note that $M'_{11} \le M_{11}$ since $R'_{11} \subseteq R_{11}$ and similarly for other suprema.

7.2.11 For any nonempty set $S \subseteq \mathbb{R}$, $\inf S \le \sup S$. You should also use the definitions of upper and lower sums.

7.2.12 For any set $S \subseteq \mathbb{R}$, $\sup(-S) = -\inf(S)$. You should also use the definitions of upper and lower sums.

7.2.13 The second claim is true. Use the definitions of upper sums and an arbitrary partition to prove it. For the other one, take the functions $f(x) = x$ and $g(x) = x+1$ and the trivial partition on the rectangle $R = [0,2]$.

7.2.14 Use the definition of Riemann sums and the linearity of sums.

7.3.1
(a) $\overline{I_R}f = \inf\{U_P(f) | P \text{ a partition of } R\} = \inf\{16\} = 16$
(b) $\underline{I_R}f = \sup\{L_P(f) | P \text{ a partition of } R\} \geq \sup\{L_{P_\varepsilon}(f) | 0 < \varepsilon < 1\} \geq 16$
(c) $16 \leq \underline{I_R}f \leq \overline{I_R}f = 16$

7.3.2
(a) True, what is the definition of $\underline{I_R}f$?
(b) False.
(c) True, by definition of supremum.
(d) False. Think of $f(x) = x^2$ with any arbitrary partition.

7.3.3
(a) There are many valid choices. Without loss, you may assume $\varepsilon < 1$. One choice is to construct P_ε from the partition $\{0, \frac{\varepsilon}{3}, 2\}$ of $[0,2]$ and the partition $\{3, 3+\frac{\varepsilon}{3}, 7\}$.
(b) $L_{P_\varepsilon}(f) = 16$ for all $\varepsilon > 0$ so $\underline{I_R}f \geq 16$. You must compute $L_P(f)$ for all partitions P to deduce $\underline{I_R}f = 16$.
(c) Yes. Your argument should use that $\underline{I_R}f \leq \overline{I_R}f$.

7.3.4
(b) Your sketch should include the following labels: R_{ij}, m_{ij}, M_{ij}, x_0, x_{i-1}, x_i, x_N, y_0, y_{i-1}, y_i, y_N
(c) $m_{ij} = 0$ if $1 \leq i \leq j \leq N$ and 1 otherwise. $M_{ij} = 0$ if $1 \leq i < j \leq N$ and 1 otherwise.
(d) $U_{P_N}(f) = 2 + \frac{2}{N}$ and $L_{P_N}(f) = 2 - \frac{2}{N}$
(e) Yes and $\displaystyle\int_R f\,dA = 2$. Your argument should use that $\underline{I_R}f \leq \overline{I_R}f$.

7.3.5 The equalities in line 3 cannot be deduced in that order. Note that $\overline{I_R}f = \inf\{U_P(f)\} \leq U_{P_N}(f) = 30 + \frac{15}{N}$ and $\underline{I_R}f = \sup\{L_P(f)\} \geq L_{P_N}(f) = 30 - \frac{15}{N}$

7.3.6
(a) Line 2 is flawed. What if a pair of partitions P' and P'' somehow satisfies $U_{P'}f < L_{P''}f$? Can you still ensure $\underline{I_R}f \leq \overline{I_R}f$? Of course, this inequality with upper and lower sums cannot happen by Lemma B, but that is not mentioned in the given argument.
(b) Line 2 is flawed. You should write out a careful proof of the fact that

$$\underline{I_R}f = \sup_{P'} L_{P'}(f) \leq \inf_{P''} U_{P''}(f) = \overline{I_R}f.$$

Use the definition of supremum and infimum involving ε.

7.3.7
(a) You will need to use two facts. First, $\overline{I_R}(f) - \underline{I_R}(f) \geq 0$. Second, for a given $x \in \mathbb{R}$, if $\forall \varepsilon > 0$, $0 \leq x < \varepsilon$, then $x = 0$.
(b) You will need to use a common refinement and the definition of f is integrable.

7.3.8

(a) The first error is in line 2. Should those be equal signs? Refer back to Lemma 7.2.9.

(b) First, prove for any partition P that $U_P(f+g) \le U_P(f) + U_P(g)$ and similarly for the lower integral. Second, prove for any two partitions P' and P'' we have $\overline{I_R}(f+g) \le U_{P'}(f) + U_{P''}(g)$ and similarly for the lower integral. Note, that you have to use common refinement of P' and P'' at some point of your proof of the second step. Third, take the infimum (or supremum) for one quantity at a time. Fourth, squeeze upper and lower integrals.

7.3.9

(a) Line 2 is wrong. Be careful about using the integral symbol on a function which may or may not be integrable.

(b) False. Consider Example 7.3.9.

7.3.10 Calculate the upper and lower sums for an arbitrary partition of R. What do you notice?

7.3.11 There are many possible proofs. One strategy is to compute one integral using the lower sums and the other with its upper sums. Use Lemmas 7.2.8 and 7.2.9.

7.3.12

(a) False. Consider Example 7.3.9.

(b) True. How do $\underline{I}_R f, \overline{I}_R f$ relate to $U_P(f)$ and $L_P(f)$ and to the integrability of f on R?

7.4.1 For each part, consider whether S is compact and f is continuous on S. Is S a subset of a bigger set where this is the case? Does the rate of change of the function grow uncontrollably on the set?

(a) Yes

(b) No

(c) No

(d) Yes

(e) Yes

7.4.2

(a) No

(b) Yes

(c) No

(d) No

(e) No

(f) Yes

(g) Yes

(h) Yes

7.4.3 You can use find a single counterexample that works for both "False" options.
(a) Nonsense. Can a function be continuous at a point?
(b) False.
(c) True. Follows by definition. See Lemma 7.4.6.
(d) True. Follows by definition.
(e) True. Follows from a theorem.
(f) False.

7.4.4 The "False" choices are not equivalent. What do these statements say instead?
(a) True.
(b) False.
(c) False.
(d) True.
(e) True.
(f) False.
(g) True.

7.4.5 $f(x)$ is a continuous function on a compact set R.

7.4.6 As usual, your proof should start by fixing an ε. Remember f' is bounded means that there is some M such that $|f'(x)| < M$ for all $x \in \mathbb{R}$. Consider taking $\delta = \frac{\varepsilon}{M}$.

7.4.7 Note that $x^3 - y^3 = (x-y)(x^2 + yx + y^2)$.

7.4.8 Choose x and y such that $\frac{1}{x} = (k + \frac{1}{2})\pi$ and $\frac{1}{y} = (k - \frac{1}{2})\pi$.

7.4.9 Consider the function $f(x) = x^2$.

7.4.10 Start from the defintion of uniform continuity. This will require the triangle inequality.

7.4.11
(a) Rectangles become intervals. Volumes become lengths. Norms become absolute values. No triangle inequality needed. The choice of N will look a bit simpler. Otherwise, not much else changes.
(b) The errors occur on Line 6 and Line 8. They come from sloppy use of convergence notation.

7.4.12
(a) Line 6 uses Line 3. The choice ensures $||P_N|| < \delta$
(b) $||P|| < \delta$
(c) $M_i - m_i$
(d) Uniform continuity should be the key part of your explanation.

7.5.1
(a) True by definition.
(b) True by the invariance of volume.
(c) True by definition.
(d) True by the topological definition of Jordan measurability.
(e) True by the topological definition of Jordan measurability.

7.5.2 There are many examples for each part. Only one possible example is shared.
(a) $A = \mathbb{R}^n$
(b) $B = [0,1]^n \cap \mathbb{Q}^n$
(c) $C = [0,1]^n \cap \mathbb{Q}^n$
(d) D can be a fat Cantor set. Sets can be really weird, huh?

7.5.3 Everything except (c) is equal to vol(D). Note (d,e) are simply notational convention.

7.5.4
(a) $L_P(\chi_S)$
(b) $U_P(\chi_S)$
(c) vol(R)
(d) $U_P(\chi_{\partial S})$ is bounded above by this quantity.

7.5.5 Ernie is correct. Remember that the volume of a rectangle was defined as $(d-c)(b-a)$ before the integral definition of volume was introduced.

7.5.6 You may have slightly different answers depending on your interpretation of the picture.
(a) Your answer should be near 18.
(b) Your answer should be near 45.
(c) The answer should be somewhere around 27.
(d) You would expect the difference to approach zero.

7.5.7 For each identity below, you should prove this by considering cases depending on where x belongs.
(a) $\chi_\varnothing$ is the constant zero function. $\chi_{\mathbb{R}^n}$ is the constant 1 function.
(b) Note $0 \leq \chi_A(x) \leq 1$ for all $x \in \mathbb{R}^n$.
(c) $\chi_{A^c}(x) = 1 - \chi_A(x)$ for all $x \in \mathbb{R}^n$.
(d) $\chi_A(x) \leq \chi_B(x)$ for all $x \in \mathbb{R}^n$.
(e) $\chi_{A\cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_{A\cap B}(x)$ for all $x \in \mathbb{R}^n$.

7.5.8 For example, directly calculate upper sums and lower sums for a regular partition of $R = [0,1]\times[0,1]$.

7.5.9

(a) $f(t) = (\cos t, \sin t)$

(b) $f(s,t) = (\cos s \sin t, \sin s \sin t, \cos t)$

7.5.10

(a) You will also need the topological definition of Jordan measurability. Remember $S \setminus S^o \subseteq \partial S$ and $S^o \subseteq S$. You will need to use upper and lower integrals and monotonicity to finish the proof.

(b) Linearity of integration. You have already proven the required integrals exist.

7.5.11

(a) Property I : $S^o \subseteq S \implies S^o$ is bounded.
Property II: $\partial(S^o) \subseteq \partial S$.
Property III: If a set is Jordan measurable with zero volume, any subset of this set is also Jordan measurable with zero volume.

(b) Your proof should follow quickly assuming the three properties. It is a good exercise to actually prove each of the three properties by following the definitions.

7.5.12 False. Think of your notoriously bad example.

7.5.13

1. You will need to verify and use that $\overline{S} \setminus S$ is Jordan measurable with zero volume. Follow the ideas from the previous problem on S^o.
2. Use that $\chi_S = \chi_{\overline{S}} + (\chi_{\overline{S}} - \chi_S)$. Use linearity and explain why the second integral is zero. Note it is not always true that $(\chi_{\overline{S}} - \chi_S) = \chi_{\partial S}$. Instead establish a related helpful inequality.

7.5.14

(a) Use that $\partial(S \cup T)$ and $\partial(S \cap T)$ are subsets of $\partial S \cup \partial T$. Squeeze their upper integrals to prove they are Jordan measurable and have zero volume.

(b) You will need linearity and an identity involving $\chi_{S\cup T}$, χ_S, χ_T and $\chi_{S\cap T}$.

7.5.15

(a) Proceed by induction on k and use the previous problem.

(b) Proceed by cases depending on whether $x \in S$ or $x \notin S$.

(c) Apply (a), (b), and linearity.

7.5.16

(a) By the definition of integrability, the integral is equal to the upper integral. By the definition of infimum, you can find an upper sum with ε of the infimum.

(b) Line 5. The invariance of rectangle volume has a crucial assumption when integrating over R'. Does R'_i satisfy this assumption? What about R_i?

(c) Line 5 uses linearity to justify that the function $g = \sum_{i \in I} M_i \chi_{R'_i}$ is integrable on R'. Hence, the upper integral of g on R' is equal to the integral of g on R'.

(d) Prove the identity with R'_i replaced by R_i. Then multiply both sides by $\chi_{R'}$.

7.5.17

(a) A. ∂S, B. is Jordan measurable and has zero volume, C. sufficiently refining the partition.

(b) A corresponds to Lines 5, 6, and the second equality in Line 7. B corresponds to Line 1 (and arguably Lines 3 and 4). C corresponds to Lines 4 and 7.

For convenience and the sake a brevity, we shall say a set has **zero Jordan measure** if the set is Jordan measurable with zero volume.

7.6.1

(a) Yes, this is a zero Jordan measure set.

(b) Yes

(c) No

(d) No

(e) No

(f) No

(g) No. What is the closure of this set?

(h) Yes

7.6.2 The following will be helpful:
Note that S has zero Jordan measure $\iff$ $\overline{S}$ has zero Jordan measure. Do you see why?
Draw a picture of S when possible.
What does the contrapositive of theorem 7.5.9 tell you?
Can you apply Sard's theorem?

(a) True

(b) False

(c) True

(d) True

(e) False

(f) False

(g) False

(h) True

(i) True

(j) True

7.6.3 For each set, is the set bounded and does its boundary have zero Jordan measure?

(a) Yes

(b) Yes

(c) Yes

(d) No, the set is not bounded.

(e) Yes

(f) Yes

(g) Yes

(h) No, the set is not bounded.

(i) No, the boundary of S does not have zero Jordan measure.

(j) Yes

7.6.4

(a) True

(b) True

(c) True

(d) False

(e) True

(f) False

7.6.5

(a) Take $f(t) = (\cos t, \sin t)$. What should you choose for the rectangle R? What should be your domain for f? Remember the domain of f must be open. Make sure to explicitly state that all conditions of the theorem are met.

(b) Use (a) and the topological definition of Jordan measurability.

(c) Take $f(s,t) = (\cos s \sin t, \sin s \sin t, \cos t)$. What should you choose for R? What should be your domain for f? Remember the domain of f must be open. Make sure to explicitly state that all conditions of the theorem are met.

(d) Use (c) and the topological definition of Jordan measurability.

7.6.6 Take $f(u,v) = (R\cos u, R\sin u, v)$. What should you choose for R? What should be your domain for f?

7.6.7 Take $F(x) = (x, f(x))$ for $x \in \mathbb{R}^n$. What should you choose for R? What should be your domain for F? You need to use a property about subsets of zero volume sets.

7.6.8 To show that S is bounded, proceed by definition. To show ∂S has Jordan measure 0, break up ∂S into two pieces and use Sard's theorem on each piece. It's always helpful to draw a picture. What is the union of zero volume sets?

7.6.9 Follow the same strategy as the previous question, except ∂S will need to broken up into 3 pieces.

7.6.10

(a) Your picture proof should include a number line with the three points. Each point should have a closed interval containing it. You should label the intervals and define them in terms of $\varepsilon > 0$.

(b) Make sure to fix $\varepsilon > 0$. For example, take $R_1 = [1 - \frac{\varepsilon}{6}, 1 + \frac{\varepsilon}{6}]$, $R_2 = [\frac{1}{2} - \frac{\varepsilon}{6}, \frac{1}{2} + \frac{\varepsilon}{6}]$, and $R_3 = [\frac{1}{3} - \frac{\varepsilon}{6}, \frac{1}{3} + \frac{\varepsilon}{6}]$. Show R_1, R_2, R_3 satisfy the rectangular covering theorem.

7.6.11

(a) Fix $\varepsilon > 0$. Take $R = [0,1] \times [0,1] \times [0,\varepsilon]$. It can be helpful to draw a picture.

(b) Take $f(u,v) = (u,v,0)$. Make sure to explicitly state that all the conditions for Sard's theorem are met.

7.6.12

(a) Note that the function $f(x,y) = 2x + 3y$ increases with increasing x and y. Use this fact and a regular partition of $[0,1] \times [0,2]$ to construct your rectangles.

(b) Take $f(s,t) = (s,t,2s+3t)$.

7.6.13 You can choose the same rectangles for S that are given by the fact that T is zero Jordan measure.

7.6.14 Use the rectangular covering theorem. Take rectangles $R_1, \ldots, R_n$ covering S such that $\sum_{i=1}^{n} \operatorname{vol}(R_i) \leq \frac{\varepsilon}{2}$ and $R_{n+1}, \ldots, R_{n+m}$ covering T such that $\sum_{i=n+1}^{n+m} \operatorname{vol}(R_i) \leq \frac{\varepsilon}{2}$.

7.6.15 Take the zero volume sets to be singletons.

7.6.16 Use the rectangular covering theorem and the fact that the closure of a closed set is itself and if $A \subset B$ then $\overline{A} \subset \overline{B}$.

7.6.17 It will be easier to show the contrapositive.

7.6.18

(a) You should draw the quartercircle with some rectangles, similar to example 7.6.11. In your diagram, label S, R_i, $\cos(\theta_i)$, $\cos(\theta_{i-1})$, $\sin(\theta_{i-1})$ and $\sin(\theta_i)$.

(b) Why is $\cos(\theta_{i-1}) > \cos(\theta_i)$ and $\sin(\theta_i) > \sin(\theta_{i-1})$?

(c) Note that $\theta_i - \theta_{i-1} = \frac{\pi}{2N}$. How does MVT help?

(d) Monotonicity is important here.

7.7.1

(a) Continuity on $B_1(0)$ is immediate. In order to justify that f is bounded one can proceed using the following trick: Consider restriction of f on a closed ball instead, this restriction $f|_{\overline{B}_1(0)}$ is a continuous function on a closed set, hence bounded by EVT. But then $f_{B_1(0)}$ must also be bounded and we can apply Theorem A to conclude that $f_{B_1(0)}$ is integrable.

(b) Again, the continuity is immediate. However, here $f|_{B_1(0)}$ is not bounded as $\lim_{(x,y,z)\to(1,0,0)} = +\infty$. Thus it is not integrable.

(c) $f|_S$ is continuous and bounded, so we can apply Theorem A to conclude that it is integrable.

(d) Here S is measurable of Jordan measure zero. The set of discontinuities of $f|_S$ is all of S, but S has measure zero so we can still apply Theorem A and conclude that f is integrable on S.

(e) Note that the set of discontinuities of f is precisely a closed unit cube, it is not of measure zero (it is Jordan measurable, but has nonzero volume), so we cannot apply Theorem A. Finally, to prove that f is not integrable, simply note that for all partitions, the lower sum is zero, while the upper sum equals to one for all partitions.

7.7.2 The second, third and fifth expressions are correct. Remember that when you move from D to a rectangle, you need to multiply g by χ_D and the rectangle must contain D.

7.7.3

(a) Your argument should be very short. The set of discontinuities is empty.

(b) Your argument should again be very short. The set of discontinuities is the boundary.

(c) Take $S = [0,1]^n \cap \mathbb{Q}^n$ and choose f similarly. Try to ensure that $\chi_S f$ is always zero.

7.7.4

(a) The proof that S is bounded follows by definition very quickly. The proof that ∂S has zero Jordan measure requires a parametrization of the circle $x^2+y^2 = 8$ with trigonometric functions. Your parametrization should not involve functions of the form $\pm\sqrt{8-x^2}$ because these will not satisfy the assumptions of Sard's theorem.

(b) D is a diagonal line segment, which has zero volume by Sard's theorem. Parametrize it using $f(t) = (t,t)$.

(c) Use Theorem A.

7.7.5 The set of discontinuities of f on S is an arc of a circle. To parametrize this, you'll need to find where the two circles intersect or show the entire set of discontinuities of f has zero Jordan measure.

7.7.6

(a) First note that $f|_S$ is bounded as S is a rectangle, so it is Jordan measurable. Start by determining the set D of discontinuities of $f|_S$, then sketch this set. Can you cover D by a pair of rectangles of an arbitrary small volume?

(b) Same idea as in the previous part.

(c) Theorem 7.7.17 is slightly preferable since S is already a rectangle.

7.7.7

(a) Show that $f|_S$ is continuous on its domain and that S is Jordan measurable.

(b) Here we have to use function $g = \chi_S f : [-1,1] \times [-1,1] \to \mathbb{R}$. Start by finding the set of discontinuities of g.

(c) There should be almost no difference between your proofs. You will need to verify the same sets have zero Jordan measure.

7.7.8

(a) What is the boundary of S?

(b) To show that this function is integrable, you'll need to look at upper and lower sums. Notice that f is 0 on $([2,3] \cap \mathbb{Q})^2$. Choose your partition so that it refines the partition $\{[-1,2] \times [-1,2], [-1,2] \times [2,3], [2,3] \times [-1,2], [2,3] \times [2,3]\}$. You can show that the upper and lower sums over the subrectangles contained in $[2,3] \times [2,3]$ must be 0.

7.7.9

(a) See Theorem 7.3.13, the linearity property for integration on rectangles.

(b) Remember to introduce an indicator function and a rectangle. You should cite the invariance of rectangles somewhere in your proof. Be careful to prove that $f + \lambda g$ is integrable on S before writing symbols like $\int_S (f + \lambda g) dV$ or $\int \chi_S (f + \lambda g) dV$.

7.7.10 Use that $g = f + (g - f)$. Be careful, don't forget to prove that g is integrable on S before writing symbols like $\int_S g dV$ or $\int \chi_S g dV$.

7.7.11

(a) Apply the rectangular result to $\chi_S f$ and $\chi_S g$. Remember to introduce indicator functions and a rectangle. You should cite the invariance of rectangles somewhere in your proof. Make sure to check the integrals exist.

(b) Introduce the set $T = \{x \in S : f(x) > g(x)\}$ and use Theorem 7.7.8.

7.7.12 Remember to introduce indicator functions and a rectangle. You should cite the invariance of rectangles somewhere in your proof. Use the fact that $\chi_{S' \cup S''} = \chi_{S'} + \chi_{S''} - \chi_{S' \cap S''}$ along with linearity on rectangles, and Theorem 7.7.8.

7.7.13

(a) Use the topological definition of Jordan measurability (Theorem 7.5.14).

(b) It uses the property that a subset of zero Jordan measure sets also has zero Jordan measure. Note that the set of discontinuities of $\chi_S f$ is not necessarily equal to $D \cup \partial S$.

(c) Be careful: in this case $D \cup \partial S$ is not equal to the set of discontinuities of $\chi_S f$.

7.7.14

(a) Lines 4, 5, 6 and 11.

(b) Lines 1, 2, and 10. Note Line 1 parallels the identity that $|\chi_S| \leq 1$ for any indicator function χ_S.

(c) Line 3 and 7. Line 3 crucially splits R into two pieces: one large piece on which f is continuous, and one small piece on which f is discontinuous.

(d) $U_{P''}(f) - L_{P''}(f)$ is the sum of these two sums, so we have shown that there exists a partition P'' with $U_{P''}(f) - L_{P''}(f) < \varepsilon$.

A.8. Integral applications

8.1.1

(a) This is equal to area(S) by definition.

(b) This is equal to area(S). Note $S \subseteq [0,2] \times [0,4]$.

(c) This integral is nonsense. What is the domain of f?

(d) This is equal to area(S) by 1D calculus.

(e) This is equal to area(S) by 1D calculus.

(f) This is equal to area(S). It is the same as 4.

8.1.2

(a) $T = \{(x, y) \in \mathbb{R}^2 : 1 \leq x \leq 4, x \leq y \leq x^2\}$

(b) $\text{area}(T) = \iint_T 1 dA = \int_1^4 (x^2 - x) dx$

8.1.3 Try to apply Theorem 8.1.3.

(a) A candidate is $T = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 4, 0 \leq z \leq y^2 + 1\}$ and $\text{vol}(T) = \iiint_T 1 dV$

(b) A candidate is $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$ and $\text{vol}(T) = \iint_D f\, dA$ with $f(x, y) = y^2 + 1$.

(c) Both are correct, but Alisa's does not follow from definition.

8.1.4

(a) $T = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 4, -\sqrt{4 - x^2 - y^2} \leq z \leq \sqrt{4 - x^2 - y^2}\}$

(b) $\text{vol}(T) = \iiint_T 1 dV = \iint_S \sqrt{4 - x^2 - y^2} - (-\sqrt{4 - x^2 - y^2}) dA$ where $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$

8.1.5

(a) Let $\varepsilon > 0$. Try to find a partition P such that

$$R_i \times [0, f(x_i^*) - \varepsilon] \subseteq T_i \subseteq R_i \times [0, f(x_i^*) + \varepsilon].$$

(b) Use your argument from the previous part.

(c) The first equality hard to prove. The second follows from a theorem in the previous chapter. What does integrability of f tell us about Riemann sums?

8.1.6 Consider any rectangle R which contains S. Start by showing that for every partition of R, the lower (rep. upper) sums of the suggested integral underestimate (resp. overestimate) the area of T. This will lead to an informal derivation. Look at Example 8.1.4 for more guidance on a formal proof.

8.1.7

(a) To prove that T is Jordan measurable always start by finding the boundary of T.

(b) This integral is precisely the definition of a volume of a Jordan measurable set.

(c) This is an application of Theorem 8.1.3.

8.1.8 Follow the same ideas as for the previous question.

8.1.9

(a) The statement is nearly identical with another continuous non-negative function $g : S \to [0, \infty)$. Your statement must assume that $g \leq f$ on S, and the set T must be defined slightly different. The volume is equal to $\int_S (f-g)dV$. Notice how this represents the difference of two volumes.

(b) The derivation is almost identical with very minor changes.

8.2.1

(a) You should calculate an average value of $\frac{13}{240}$.

(b) You should have one under-estimate and one over-estimate. Make sure to weight your sums appropriately.

(c) "Limit" here is used inclusively to refer to things like supremum and infimum, in addition to limits along natural numbers.

8.2.2

(a) Start by finding over-estimates for the average value of f on R_i. Then take the mean value of all of these N^n sample estimates. You should get $\frac{1}{N^n} \sum\limits_{i \in I_n} M_i$

(b) Same idea as above, but with under-estimates. You should get $\frac{1}{N^n} \sum\limits_{i \in I_n} m_i$

(c) You will need to use integrability of f and to do so you will have to relate your over-estimates and under-estimates to upper and lower sums.

8.2.3 You will need to use a variety of facts:

- That f has a maximum value M and a minimum value m on S.
- That there exist $x, y \in S$ with $f(x) = M$ and $f(y) = m$.
- That there is a curve in S that starts at x and ends at y.
- That $m\,\mathrm{vol}(S) \leq \int_S f\,dV \leq M\,\mathrm{vol}(S)$.

Notice we haven't justified any of these claims above. You will need to justify each of these claims before you use them.

8.2.4 First, explain why you can replace $B_\varepsilon(p)$ with $\overline{B_\varepsilon(p)}$. Second, use the integral mean value theorem to obtain point. Third, you will need to deal with the fact that the point obtained depends on $\varepsilon > 0$. Your proof must use continuity in the last step.

8.2.5 Start by showing that Riemann sums are bounded below (resp. above) by lower (resp. upper) sum corresponding to the same partition.

8.2.6 The derivation is almost identical to Example 8.2.1 except the expression N^n is replaced by $|I|$. Nothing else needs to change because as $N \to \infty$, the norm $||P_N||$ tends to zero.

8.2.7 The derivation will be quite similar to Example 8.2.1 but you must replace f by $\chi_S f$. The discontinuities introduced must be handled as they were for the volume under a graph in Example 8.1.4.

8.2.8 This will be almost identical to the previous exercise with mostly aesthetic changes. Most importantly, your discrete approximation of total value will be slightly different compared to average value. Remember you must scale things according to the volume of each subrectangle.

8.3.1 Tigger. What is the volume of a point? You could improve Roo's answer by taking the average density of S on a small ball centered at x.

8.3.2
(a) Think of objects made of the same material like any metal casting, glass. They would have almost uniform densities. Note that shape is allowed to be very irregular as long as the density function is constant inside S.

(b) Any example from the previous part will certainly work, but there are examples of objects with continuous non-uniform density. Earth's atmosphere is one of them.

(c) Practically any object made of different materials will give you an example. Say, an object made by perfect welding of two metals would have density function continuous everywhere but at the welding surface.

(d) No object in the known universe should behave so abhorrently.

8.3.3
(a) $\displaystyle\int_A \delta_A dV > \int_B \delta_B dV$

(b) $\displaystyle\int_A 1 dV = \int_B 1 dV$

(c) $\dfrac{\int_A \delta_A dV}{\int_A 1 dV} > \dfrac{\int_B \delta_B dV}{\int_B 1 dV}$

(d) The object B has uniform density.

(e) The mass density of A at a point is equal to the limit of average density of shrinking solid balls centered at the point.

8.3.4
(a) For one of these the centroid is to the right of the centre of mass and for the other the centroid is to the left of the centre of mass.

(b) For objects with constant density, the centre of mass is equal to the centroid.

(c) For objects with varying density, the denser parts shift the centre of mass.

8.3.5
(a) The integral gives the volume of the region $\{(x, y) \in \mathbb{R}^{n+1} : x \in S, 0 \le y \le f(x)\}$ provided f is non-negative on S.

(b) $\int_S 1 dV$

(c) $\frac{1}{\text{vol}(S)} \int_S f\, dV$ provided $\text{vol}(S) > 0$

(d) The mass of a solid S with mass density function f is given by $\int_S f\, dV$ provided f is non-negative on S.

8.3.6 First, note that the inequalities for partial derivatives are strong condition from which you can derive the global properties but not vice-versa. For the questions about the center of mass, think of your whole slice as a union of two subparts; suppose you know the masses and positions of the center of mass of the subparts, can you conclude something about position of the center of mass of the whole slice?

1. True
2. False
3. False
4. False
5. False
6. False
7. False
8. False

8.3.7

(a) We can approximate mass by assuming that average density of each piece is approximately equal to the density at the sample point. You should get 10.9.

(b) It could be either. You cannot decide with the given information.

8.3.8

(a) δ is continuous so there should be a value $x_i^* \in R_i$ which is equal to the average density of R_i.

(b) The mass m is an upper bound for each lower sum so by definition the supremum, $m \geq \underline{I_R}(f)$. The same holds for the other inequality.

(c) Assume S is Jordan measurable and δ is continuous on S (or possibly continuous on S aside from a set of zero volume). In the given argument, replace δ with $\chi_S \delta$ and almost everything else remains the same.

8.3.9

(a) Recall Problem A.8.7 part 1.

$$m = \sum_i \delta(x_i^*)\mathrm{vol}(R_i)$$

Under the assumption that the density of R_i is approximately uniform,

$$m = \sum_i \mathrm{mass}(R_i)$$

(b) For each subrectangle you can use the supremum of the density function to overestimate this rectangle's average density.

(c) Again, for each subrectangle the infinum of the density function would underestimate this rectangle's average density.

(d) Do you get a better estimate if you refine your partition? What does this mean for the partition's norm?

8.3.10

(a) The quantity represents the average density of some solid. Can you tell which solid?

(b) Informally speaking, you should assume that for a small cube centered at x, the density is approximately uniform. More formally, you must assume δ is continuous at x. These are reasonable assumptions because on a small scale, the density of an object doesn't change dramatically.

(c) This is not a rigorous definition since mass has not been defined.

8.3.11 Your argument should look very similar to Example 8.3.6. Be careful that you are summing vector quantities in many cases, so you must suitably interpret your expressions as vector quantities instead of scalar quantities.

(a) Your conjecture should be something like

$$\overline{x} = \frac{1}{\text{mass}(R)} \int_R x\delta(x,y)dV.$$

(b) Use Example 8.3.6 as inspiration. Start by approximating the mass of any small subrectangle R_i, assuming the density is close to uniform. Then calculate your discrete approximation by treating every subrectangle as if it were a point mass and using the formula for centre of mass.

(c) You may assume that R is a rectangle. Your limit should be as the norm of a sequence of partitions tend to zero. They do not necessarily need to be regular.

8.3.12 You should follow similar steps to Problem A.8.9 and take the appropriate limit at the end. You will need to approximate m_i at some stage of your argument.

8.3.13 Your argument should look like a combination of Example 8.3.6 and Example 8.2.1, except your partition will not necessarily be regular. For N objects with respective masses $m_1,\ldots,m_N$ and volumes $V_1,\ldots,V_N$, the average density is the total mass divided by the total volume. Use this principle for your discrete approximation.

8.3.14 By assumption, there exists a constant $\rho \in \mathbb{R}$ such that $\delta(x) = \rho$ for all $x \in S$. Use this constant in Definition 8.3.7.

8.4.1

(a) No, it is the set of all Jordan measurable subsets. See Theorem 8.4.3.

(b) No, we assign probability only to events $A \in \Sigma$, hence A has to be Jordan measurable. See Theorem 8.4.7

(c) Yes, see Theorem 8.4.7

(d) No, probability density $\phi(x)$ itself can be arbitrary large at certain x, only the integral of probability density over any measurable $A \subset \Omega$ must be less than one.

(e) No, this formula won't work when $\mathbb{P}(A \cap B) > 0$.

8.4.2

(a) Abed

(b) Pierce

(c) Left

(d) Keep in mind that $\mathbb{P}(\{(0,0)\}) = \int_{\{(0,0)\}} \phi\, dV$. A single point in R^2 has zero volume.

8.4.3 Note that $\phi(x)$ must be nonnegative on its domain and continuous everywhere but on a set of zero Jordan measure.

(a) Yes

(b) No, for $x = -1$, $\phi_2 < 0$

(c) Yes

(d) No, the set on which ϕ_4 is discontinuous does not have zero Jordan measure.

8.4.4 Diagram B corresponds to $\phi(x, y)$. How likely should it be that the grocery store sells a lot of pasta but very little tomato sauce? What kind of set describes this event?

8.4.5

(a) This event will always happen, so it has probability one.

(b) This occurs with probability zero.

(c) This occurs with probability zero because the edge of the square has zero volume.

(d) This occurs with probability zero.

(e) This occurs with probability zero because the horizontal axis has zero volume.

(f) This occurs with probability zero because the unit circle has zero volume.

(g) This has non-zero probability.

(h) This has non-zero probability.

(i) This is not an event as it is does not constitute a Jordan measurable set.

(j) This is not an event as it is does not constitute a Jordan measurable set.

8.4.6 To show that the set is Jordan measurable you will need to parameterize the boundary.

8.4.7

1. Use Example 7.6.9 or the rectangular covering theorem (Theorem 7.6.6).
2. You must explain why $\{x\}$ is an event in Σ. This follows by Definition 8.4.6.
3. Use Definition 8.4.6 and Theorem 7.7.8.

8.4.8

(a) The key property is monotonicity. Can you see how to apply it?

(b) You can use either monotonicity or additivity for your justification. Be careful to explain why all functions in question are integrable over the given sets.

8.4.9 Rewrite the set complement as an intersection.

8.4.10 You may find Lemma 7.5.18 and the ideas behind its proof helpful.

8.4.11

(a) You should have $\Omega = [-100, 100]^4$. Can you fill in the rest? Recall that the probability function is uniform.

(b) The event in question is $A = \{(a, b, c, d) \in \mathbb{R}^4 | ad = bc\}$.

(c) What should be the volume of A?

[0]Images created by Cindy Blois and used with permission.

8.4.12 Your pairwise disjoint events could be singletons for example.

A.9. Integration methods

9.1.1 I corresponds to region A and J corresponds to region B.

9.1.2
(a) $\int_0^2 \varphi(0.5, y)dy$
(b) $\int_{-1}^{0.5} \int_0^2 \varphi(x, y)dydx$

9.1.3
(a) Trueby Corollary 9.1.19.
(b) False, see Example 9.1.12.
(c) True by the definition of integral.
(d) False, see Example 9.1.14.
(e) False, see Example 9.1.14.
(f) False, see Example 9.1.16.

9.1.4
(a) This requires (C) to be true.
(b) This requires (A) and (C) to be true.
(c) This requires (A) and (C) to be true.
(d) This requires all three to be true.

9.1.5
(a) True
(b) True
(c) False, see Example 9.1.12.

9.1.6
(a) 2/3
(b) 4/3
(c) 2/3
(d) 4/3
Note that you can use Fubini's theorem for parts 3 and 4.

9.1.7

(a) The y-slices you calculate should be constant functions.

(b) The x-slices you calculate should be step functions. What can you say about their discontinuities? What can you deduce about their integrability?

(c) Theorem 7.7.4 will be helpful here.

(d) Have you checked all the required assumptions to apply Fubini's theorem?

(e) Again, check the relevant assumptions to apply Fubini's theorem.

(g) You should calculate $\displaystyle\iint_R \varphi\, dA = 4$.

9.1.8

(a) $\displaystyle\int_1^3 \int_{\cos y}^{\sin y} x \cos y\, dx\, dy$

(b) $\displaystyle\int_1^3 \int_{-1}^{1} \chi_S(x,y) x \cos y\, dx\, dy$ where $R = [-1,1] \times [1,3]$. There are several valid choices of R.

(c) For fixed $y \in [1,3]$, define the y-slice $f^y : [-1,1] \to \mathbb{R}$ by $f^y(x) = \chi_S(x,y)x\cos y$. The red line corresponds to the integral $\displaystyle\int_{-1}^{1} f^y(x)dx$.

(d) The fully blue shaded region corresponds to the integral $\displaystyle\int_1^3 F(y)dy = \int_1^3 \int_{-1}^{1} \chi_S(x,y)x\cos y\, dx\, dy$ where $F : [1,3] \to \mathbb{R}$ is given by $F(y) = \displaystyle\int_{-1}^{1} f^y(x)dx$.

(e) The 2-variable function is $f(x,y) = \chi_S(x,y)x\cos y$. The 1-variable functions are f^y for all $y \in [1,3]$. Your assumptions should not include any iterated double integrals.

(f) The 1-variable function is $F : [1,3] \to \mathbb{R}$ above. The double integral is $\displaystyle\int_1^3 \int_{-1}^{1} f(x,y)\,dx\,dy$ and it is also equal to $\displaystyle\int_R f\,dA$.

(g) For two variables, Fubini's theorem has two key assumptions: integrability of a 2-variable function on a rectangle and integrability of (1-variable) slices of that function on an interval. Cameron's reasoning only verifies one of these two key assumptions.

9.1.9 You can prove all of these properties by the definition of supremum.

(a) This follows from monotonicity of suprema:
If $g(x) \le h(x)$ for all $x \in [x_{i-1}, x_i]$, then $\displaystyle\sup_{x\in[x_{i-1},x_i]} g(x) \le \sup_{x\in[x_{i-1},x_i]} h(x)$.

(b) This follows from sublinearity of suprema:
$$\sup_{x\in[x_{i-1},x_i]} \big(g(x)+h(x)\big) \le \sup_{x\in[x_{i-1},x_i]} g(x) + \sup_{x\in[x_{i-1},x_i]} h(x)$$

(c) This follows from the property:
for $\lambda > 0$, $\displaystyle\sup_{x\in[x_{i-1},x_i]} (\lambda g(x)) = \lambda \sup_{x\in[x_{i-1},x_i]} g(x)$.

(d) This follows from the property:
$$\sup_{x\in[x_{i-1},x_i]} \Big(\sup_{y\in[y_{j-1},y_j]} f(x,y)\Big) \le \sup_{(x,y)\in[x_{i-1},x_i]\times[y_{j-1},y_j]} f(x,y).$$

This last one is trickier to prove. Show that any upper bound B of the righthand supremum is an upper bound of the supremum over y with x fixed. With one more step, you can conclude it is an upper bound of the iterated supremum.

9.2.1

(a) zero

(b) zero

(c) positive

(d) zero

(e) positive

(f) negative

9.2.2

(a) $A(x) = \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} f(x,y)dy$

(b) $\mathrm{vol(T)} = \int_{-2}^{2} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} f(x,y)\,dy\,dx$

9.2.3

(a)
A. The region is both x-simple and y-simple.
B. The region is both x-simple and y-simple.
C. The region is neither x-simple nor y-simple.
D. The region is both x-simple and y-simple.
E. The region is both x-simple and y-simple.
F. The region is x-simple and is not y-simple.

(b) All can be written as a finite union of x-simple regions or as a finite union of y-simple regions. Can you see what these unions are in each case?

9.2.4 Draw the region of integration and determine what symmetries it has.

(a) B, C, D

(b) B

(c) A, C

(d) B

(e) B

(f) D

9.2.5 The region is a solid triangle with vertices $(0,0), (1,0), (1,1)$. The value is $\frac{1}{2}(1-\frac{1}{e})$.

9.2.6 $\displaystyle\int_4^{16} \int_{(y-4)/4}^{\sqrt{y}/2} \frac{x-y}{x+y} dx dy$

9.2.7

(a) This region is the solid triangle with vertices $(0,0),(1,0)$, and $(1,1)$. Switching the order gives the double iterated integral $\int_0^1 \int_x^1 xy^2 dy dy$. You can calculate the integral using either order.

(b) The region is the right half of the unit disc centered at the origin. Switching the order gives the double iterated integral $\int_0^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-y^2}} y dy dx$.

9.2.8

(a) A rectangular prism. The integral is equal to LWH.

(b) A triangular prism. The integral is equal to 2.

(c) A half ball. The integral is equal to $\frac{2\pi R^3}{3}$.

9.2.9 You should calculate a value of $\frac{11}{12}$.

9.2.10 You should calculate a value of $\frac{1}{2}(e^{\pi^2}-1-\pi^2)$.

9.2.11 This is the volume of a cone. You should get 9π

9.2.12 You should calculate a value of 2.

9.2.13

(a) For what values of x do the functions $\sqrt{4-x^2}$ and $\sqrt{1-x^2}$ have real values?

(b) The best strategy is to draw the region of integration, if possible.

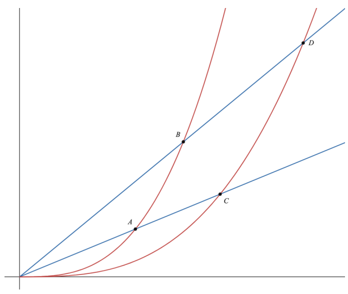
9.2.14

(a) S can be written as a union of 4 x-simple sets, but not fewer. Two of the sets look like $\{(x,y)\in\mathbb{R}^2 : -2\le x\le -1, -\sqrt{4-x^2}\le y\le\sqrt{4-x^2}\}$ and $\{(x,y)\in\mathbb{R}^2 : -1\le x\le 1, \sqrt{1-x^2}\le y\le\sqrt{4-x^2}\}$. The other two are very similar.

(b) Your four double integrals correspond precisely to the 4 sets from part (a). By symmetry, you can reduce this to only two double integrals but that is not necessary.

9.2.16

(a) Your sketch should look something like the following



(b) Here is one way you might express the mass:

$$\int_{\frac{1}{\sqrt{3}}}^{\sqrt{\frac{2}{3}}}\int_x^{3x^3}\delta(x,y)dydx+\int_{\sqrt{\frac{2}{3}}}^{1}\int_x^{2x}\delta(x,y)dydx+\int_1^{\sqrt{2}}\int_{x^3}^{2x}\delta(x,y)dydx$$

9.2.17
1. Note that x^2y^3 gives negative values for some values of x and y.
2. You should calculate a volume of $\frac{3}{2}$.

9.2.18 You should calculate a mass of $\frac{7}{3}$.

9.2.19 You should calculate an average value of $\frac{80}{7}$.

9.2.20 You should calculate a center of mass of $(\frac{3}{4}, \frac{3}{8})$.

9.2.21 You should split the integral up into four components and show that they pair up to cancel out.

9.3.1
(a) (C) $r = 1 + \cos 2\theta$
(b) (F) $r = 1 + 2\cos\theta$
(c) (A) $r = \theta, \theta > 0$
(d) (E) $r = \sin 3\theta$

9.3.2 Use this Desmos polar graphing calculator to check your answer.

9.3.3 In no particular order, you should get the following integrals:

$$\int_1^3 \int_0^2 1dydx \qquad \int_0^{\pi} \int_1^2 rdrd\theta$$

$$\int_0^{2\pi} \int_0^2 rdrd\theta \qquad \int_0^2 \int_{\frac{1}{2}x-1}^{-\frac{1}{2}x+3} 1dydx$$

9.3.4
(a) $0 \leq \theta \leq \pi/3$
(b) $\displaystyle\int_0^{\sin(3\theta)} r^2 \sin\theta dr$ for $0 \leq \theta \leq \pi/3$
(c) $\displaystyle\int_0^{\pi/3} \int_0^{\sin(3\theta)} r^2 \sin\theta dr d\theta$
(d) $0 \leq r \leq 1$
(e) $\displaystyle\int_{\arcsin(r)/3}^{(\pi-\arcsin(r))/3} r^2 \sin\theta d\theta$ for $0 \leq r \leq 1$
(f) $\displaystyle\int_0^1 \int_{\arcsin(r)/3}^{(\pi-\arcsin(r))/3} r^2 \sin\theta d\theta dr$
(g) $\frac{27\sqrt{3}}{640} \approx 0.073$

9.3.5 (a,b,c) and (f) are equal to the mass of P. Recall that you must include a stretch factor when using polar coordinates and that $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

9.3.6 Can you see why Chuck's answer is equal to Archie's?

9.3.7

(a) W is the set described by $3 \le r \le 5$ and $0 \le \theta \le \pi/3$. That is, $W = g([3,5] \times [0,\pi/3])$.

(b) You should find that $\int_W f\, dA = \int_0^{\frac{\pi}{3}} \int_3^5 1 dr d\theta = \frac{2\pi}{3}$.

9.3.8 256/105

9.3.9 $\pi \sin(9)$

9.3.10 $\frac{2-\sqrt{3}}{2}\pi$

9.3.11 $\frac{16\pi}{5}$

9.3.12 $\frac{2\pi}{3} - \sqrt{3}$

9.3.13 $\frac{3}{8}$

9.3.14 $\pi - 2\sqrt{3} + \frac{\sqrt{3}}{2}$ and $\pi + 3\sqrt{3}$

9.3.15 $\frac{5\pi}{3}$ and $(\frac{21}{20}, 0)$. Recall Definition 8.3.9 to find the centre of mass.

9.3.16 $\frac{81\pi}{2}$

9.3.17

(a) Theorem 9.3.4 is incorrectly applied. Remember that $g(r,0) = g(r,2\pi)$.

(b) Fubini's theorem is incorrectly applied. What kinds of sets does Fubini's theorem apply to?

(c) Apply Theorem A with an open set Ω. Then add an integral over a set of measure zero. Remember that this extra integral must be equal zero. Then, you should apply Fubini's theorem.

9.3.18

(b) Your informal explanation should include linear approximations and/or the definition of partial derivates. A formal justification involves the mean value theorem.

(c) You want $g(R)$ to be approximated by a rectangle. What do the partial derivatives represent geometrically?

9.4.1

(a) This is false. f can be non-integrable on a single slice, similarly to the 2-dimensional case.

(b) This is false. Again, the counterexample is similar to the 2-dimensional case.

9.4.2 For $dxdz$, the function $G(x,z) = \varphi(x,-2,z)$ must be integrable on $[1,5] \times [0,2]$ and for every $z \in [0,2]$, the z-slice $G^z(x) - \varphi(x,-2,z)$ must be integrable on $[1,5]$.

For $dzdx$, the function $G(x,z) = \varphi(x,-2,z)$ must be integrable on $[1,5] \times [0,2]$ and for every $x \in [1,5]$, the x-slice $G^x(z) - \varphi(x,-2,z)$ must be integrable on $[0,2]$.

9.4.3 $\int_e^f \int_c^d \int_a^b \varphi(x,y,z)dxdydz$, $\int_c^d \int_e^f \int_a^b \varphi(x,y,z)dxdzdy$, $\int_e^f \int_a^b \int_c^d \varphi(x,y,z)dydxdz$ $\int_a^b \int_e^f \int_c^d \varphi(x,y,z)dydzdx$, $\int_c^d \int_a^b \int_e^f \varphi(x,y,z)dzdxdy$

9.4.4 Let $R=[a,b]\times[c,d]\times[e,f]$ and let $\varphi : R \to \mathbb{R}$ be bounded. If

- For every $y \in [c,d], z \in [e,f]$, the (y,z)-slice $\varphi^{(y,z)}$ is integrable on $[a,b]$.
- For every $y \in [c,d]$, the y-slice φ^y is integrable on $[a,b]\times[e,f]$.
- φ is integrable on $R=[a,b]\times[c,d]\times[e,f]$.

Then the iterated triple integral

$$\int_c^d \int_e^f \int_a^b \varphi(x,y,z)dxdzdy$$

exists and is equal to the integral $\displaystyle\iiint_R \varphi dV$.

9.4.5

(a) Only B represents the mass density of the cheese string. Note that the cheese string is chosen by picking a value for x and z, while the value of y changes.

(b) Only E represents the mass density of the cheese sheet.

(c) B and C both represent the mass of the cheese block. Note that φ is continuous.

9.4.6

(a) Assumption (G). This is necessary since it is the definition of the single variable integral existing.

(b) Assumptions (D) and (G). This follows by the 2-variable Fubini's theorem; these assumptions are sufficient but not necessary for the double iterated integral to exist.

(c) Assumptions (A), (D), and (G). This follows by a 3-variable Fubini's theorem; these assumptions are sufficient but not necessary for the triple iterated integral to exist.

(d) Assumptions (A), (D), and (G).

9.4.7 You should calculate that this integral is equal to 54.

9.4.8 You should confirm that the volume is $\frac{4\pi}{3}$.

9.4.9 Prove the base case for $n=1$. Then show that $\int_0^1 x_1x_2^2\cdots x_n^n dx_n = x_1x_2\cdots x_{n-1}^{n-1}\int_0^1 x_n^n dx_n$.

9.4.10

(a) You should find (x,y)-slices that are step functions. What can you say about the Jordan Measure of their set of discontinuities?

(b) Try to draw the domain of the x-slices. What values does the x-slice take in different regions? What can you say about the Jordan Measure of its set of discontinuities?

(c) What is the set of discontinuities of φ?

(d) Have you checked all the necessary assumptions to apply Fubini's theorem?

(e) You should find that the integral is equal to 24.

9.4.11 Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(x,y,z) = \chi_S(x,y,z)ye^{xz}$.

(a) $R = [-2,2] \times [0,4] \times [0,4]$ is one possibility.

(b) For any $x \in [-2,2]$, $y \in [0,4]$, the animation corresponds to the integral $\int_0^4 f(x,y,z)dz$. The 1-variable function is the xy-slice of f and hence its domain should be $[0,4]$.

(c) For any $x \in [-2,2]$, by Fubini's theorem for double integrals, the animation corresponds to the iterated integral $\int_0^4 \int_0^4 f(x,y,z)dzdy$ and the integral $\int_{[0,4]\times[0,4]} f^x dA$ where f^x is the x-slice of f.

(d) By Fubini's theorem for triple integrals, the animation corresponds to the integral $\int_R f dV$ and the iterated integral $\int_{-2}^2 \int_0^4 \int_0^4 f(x,y,z)dzdydx$.

(e) $\int_{-2}^2 \int_0^4 \int_0^4 f(x,y,z)dzdydx$

(f) The functions should be f, f^x, and f^{xy} with domains R, $[0,4]\times[0,4]$ and $[0,4]$ respectively. The integrals should be $\int_R f dV$, $\int_{[0,4]\times[0,4]} f^x dA$, and $\int_0^4 f^{xy}(z)dz$. Notice there are no iterated integrals in the assumptions.

9.4.12 Remember you must deal with the discontinuities introduced by the indicator function. Define $f : R \to \mathbb{R}$ by $f(x,y,z) = \chi_S(x,y,z)ye^{xz}$. You will apply Theorem 7.7.4 and Sard's theorem (Theorem 7.6.13) many times. To check f is integrable on R, show S is Jordan measurable using Sard's theorem. To check f^x is integrable on $[0,4]\times[0,4]$, show the x-slice S^x of S is Jordan measurable in $\mathbb{R}^2$. To check f^{xy} is integrable on $[0,4]$, show the xy-slice S^{xy} of S is Jordan measurable (this is easy; it is a compact interval).

9.4.13 You will need to use Fubini's theorem (Theorem 9.4.19) and the linearity property of integration.

9.4.14

(a) Korra should define f^z and $f^{(y,z)}$ carefully.

(b) She is using $R = [a_1,b_1]\times[a_2,b_2]$, $\varphi = f$, and $[a,b] = [a_3,b_3]$. She must assume that f is integrable on $[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]$ and that for ever $z \in [a_3,b_3]$ the slice f^z is integrable on $[a_1,b_1]\times[a_2,b_2]$.

(c) She is using $R = [a_1,b_1]$, $\varphi = f$, and $[a,b] = [a_2,b_2]$. She must make the same assumptions as before and additionally that for ever $y \in [a_2,b_2]$ the slice $f^{(y,z)}$ is integrable on $[a_2,b_2]$.

9.4.15 There are $2^n - 1$ types of slices, $\binom{n}{k}$ types of slices fixing k coordinates, and $n!$ iterated integrals.

9.4.16 You will still need to assume that f is integrable and you will only be able to conclude that one iterated integral exists and is equal to the integral of f on R.

9.4.17

1. Your expression should be a double iterated integral. Can you see what this geometrically represents in the case $n = 3$?
2. If you're stuck, try doing this for $n = 2$ and $n = 3$ first. You will see a pattern.

4. Recall the big theorem for sequences.

9.5.1 (a), (f,g) and (j) are valid integrals. You don't want any variables left in your final expression, so the integrals with endpoints containing variables must be evaluated before the integral with the associated infinitesimal is evaluated.

9.5.2

(a) This is shape (C). To handle this, you will need to break the solid into multiple pieces and integrate over each piece separately. One way to write this volume is:

$$V = \int_0^1 \int_0^{1-x} \int_0^2 1 dz dy dx + \int_0^1 \int_{1-y}^2 \int_0^{3-x-y} 1 dz dx dy + \int_1^2 \int_0^{3-y} \int_0^{3-x-y} 1 dz dx dy$$

(b) This is shape (D). You can express the volume as:

$$V = \int_0^4 \int_{-\sqrt{z}}^{\sqrt{z}} \int_{-\sqrt{z-y^2}}^{\sqrt{z-y^2}} 1 dx dy dz$$

(c) This is shape (A). You can express the volume as:

$$V = \int_0^4 \int_{\frac{-z}{2}}^{\frac{z}{2}} \int_{-\sqrt{\frac{z^2}{4}-y^2}}^{\sqrt{\frac{z^2}{4}-y^2}}$$

(d) This is shape (B). You can express the volume as

$$V = \int_1^3 \int_0^2 \int_{-1}^1 1 dz dy dx$$

9.5.3

(a) $P = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 9\}$

(b) $\int_0^{9-x^2-y^2} f(x,y,z)dz$. Note that without the density function in the integral, you'd be evaluating the length of the vertical wire.

(c) $\int_{-3}^{3}\int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}}\int_0^{9-x^2-y^2} f(x,y,z)dzdydx$

(d) A z-slice of S can be expressed as $\{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 9-z\}$. This is non-empty for $0 \leq z \leq 9$. You should have a sketch of a disk of radius $\sqrt{9-z}$.

(e) $\int_{-\sqrt{9-z}}^{\sqrt{9-z}}\int_{-\sqrt{9-z-x^2}}^{\sqrt{9-z-x^2}} f(x,y,z)dydx$

(f) $\int_0^9\int_{-\sqrt{9-z}}^{\sqrt{9-z}}\int_{-\sqrt{9-z-x^2}}^{\sqrt{9-z-x^2}} f(x,y,z)dydxdz$

(g) A y-slice of S can be expressed as $\{(x,z) \in \mathbb{R}^2 : 0 \leq z \leq 9-x^2-y^2\}$. This is non-empty for $-3 \leq y \leq 3$.

(h) $\int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}}\int_0^{9-x^2-y^2} f(x,y,z)dzdx$

(i) $\int_{-3}^{3}\int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}}\int_0^{9-x^2-y^2} f(x,y,z)dzdxdy$

9.5.4

(a)

$$V = \int_0^5 \int_0^{2-\frac{2}{5}x} \int_0^{10-2x-5y} g(x,y,z)dzdydx$$

(b)

$$V = \int_0^2 \int_0^{10-5y} \int_0^{5-\frac{5}{2}y-\frac{1}{2}z} g(x,y,z)dxdzdy$$

9.5.5

(a) Notice that z is between $y+1$ and 8, that y is between $2x^2$ and $4-2x^2$, and x is between -1 and 1.

(b) Notice that there are 3 different typical z-slices. The middle z-slice needs to be divided into 3 regions in $\mathbb{R}^2$.

$$\begin{aligned} I = &\int_1^3 \int_{-\sqrt{z-1}}^{\sqrt{z-1}} \int_{2x^2}^{z-1} f(x,y,z)dydxdz + \int_3^5 \int_{-1}^{-\sqrt{5-z}} \int_{2x^2}^{4-2x^2} f(x,y,z)dydxdz \\ &+ \int_3^5 \int_{-\sqrt{5-z}}^{\sqrt{5-z}} \int_{2x^2}^{z-1} f(x,y,z)dydxdz + \int_3^5 \int_{\sqrt{5-z}}^{1} \int_{2x^2}^{4-2x^2} f(x,y,z)dydxdz \\ &+ \int_5^8 \int_{-1}^{1} \int_{2x^2}^{4-2x^2} f(x,y,z)dydxdz \end{aligned}$$

9.5.6 Your expressions should be equal to $\frac{2\pi}{5}(32-9\sqrt{3})$. Use WolframAlpha to check.

(a) One valid expression is:

$$\int_0^2 \int_{-\sqrt{4-z^2}}^{\sqrt{4-z^2}} \int_{-\sqrt{4-z^2-y^2}}^{\sqrt{4-z^2-y^2}} (x^2+y^2+z^2)dxdydz - \int_0^{\sqrt{3}} \int_{-\sqrt{3-z}}^{\sqrt{3-z}} \int_{-\sqrt{3-z^2-y^2}}^{\sqrt{3-z^2-y^2}} (x^2+y^2+z^2)dxdydz$$

(b) One valid expression is:

$$\int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{3-x^2}}^{\sqrt{3-x^2}} \int_{\sqrt{3-x^2-y^2}}^{\sqrt{4-x^2-y^2}} (x^2+y^2+z^2)dzdydx$$
$$+\int_{-\sqrt{4}}^{-\sqrt{3}} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_0^{\sqrt{4-x^2-y^2}} (x^2+y^2+z^2)dzdydx + \int_{\sqrt{3}}^{\sqrt{4}} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_0^{\sqrt{4-x^2-y^2}} (x^2+y^2+z^2)dzdydx$$
$$+\int_{-\sqrt{3}}^{\sqrt{3}} \int_{\sqrt{3-x^2}}^{\sqrt{4-x^2}} \int_0^{\sqrt{4-x^2-y^2}} (x^2+y^2+z^2)dzdydx + \int_{-\sqrt{3}}^{\sqrt{3}} \int_{-\sqrt{4-x^2}}^{-\sqrt{3-x^2}} \int_0^{\sqrt{4-x^2-y^2}} (x^2+y^2+z^2)dzdydx$$

You can use even symmetries to simplify these into only 3 triple integrals.

9.5.7 Check (using a computer or by hand) that all of your expressions are equal to 8π. For each of these, it is helpful to draw a diagram illustrating the slices, and labeling which equations correspond to each part of your diagram.

(c) You will need three separate integrals for this one.

9.5.8 $\int_0^3 \int_{-2}^2 \int_0^{\sqrt{4-y^2}} 5z^3 dzdydx = 128$

9.5.9 $\int_{-5}^5 \int_{-\sqrt{25-x^2}}^{\sqrt{25-x^2}} \int_{\sqrt{x^2+y^2}}^5 dzdydx$ and $\int_0^5 \int_{-z}^z \int_{-\sqrt{z^2-y^2}}^{\sqrt{z^2-y^2}} dxdydx$

9.5.10 $\int_0^2 \int_0^{2-x} \int_{5-3x-5y}^{5-2x-y} 1dzdydx$

9.5.11 $\int_0^3 \int_{-3+z}^{3-z} \int_{-3+z}^{3-z} 1dxdydz = 36$

9.5.12 $\int_{-12}^{12} \int_{-\sqrt{144-x^2}}^{\sqrt{144-x^2}} \int_5^{\sqrt{169-x^2-y^2}} dzdydx$

9.5.13 $(\overline{x},\overline{y},\overline{z}) = (\frac{16}{3\pi},\frac{16}{3\pi},\frac{3}{4})$

9.6.1

(a) (F)

(b) (D)

(c) (E)

(d) (B)

(e) (C)

(f) (A)

9.6.2 (a) (A) Cylinder
(b) (H) Half Cone
(c) (C) Solid between two cylinders
(d) (G) Cone hollowed by cylinder
(e) (B) Half cylinder
(f) (F) Cone with flat bottom and flat top

9.6.3 The valid expressions are parts 1, 2, 5, and 7.

9.6.4 Make sure you always include the stretch factor of $|r| = r$.
(a) $P = \{(r,\theta) \in \mathbb{R}^2 : 0 \le r \le 3, 0 \le \theta \le 2\pi\}$
(b) $\int_0^{9-r^2} f(r\cos\theta, r\sin\theta, z) r\,dz$
(c) $\int_0^{2\pi}\int_0^3\int_0^{9-r^2} f(r\cos\theta, r\sin\theta, z) r\,dz\,dr\,d\theta$
(d) $\{(r,\theta) \in \mathbb{R}^2 : 0 \le r \le \sqrt{9-z}, 0 \le \theta \le 2\pi\}$ for each fixed $0 \le z \le 9$.
(e) $\int_0^{\sqrt{9-z}}\int_0^{2\pi} f(r\cos\theta, r\sin\theta, z) r\,d\theta\,dr$
(f) $\int_0^9\int_0^{\sqrt{9-z}}\int_0^{2\pi} f(r\cos\theta, r\sin\theta, z) r\,d\theta\,dr\,dz$
(g) $\{(r,z) \in \mathbb{R}^2 : 0 \le r \le \sqrt{9-z}, 0 \le z \le 9\}$ for each fixed $0 \le \theta \le 2\pi$
(h) $\int_0^9\int_0^{\sqrt{9-z}} f(r\cos\theta, r\sin\theta, z) r\,dr\,dz$
(i) $\int_0^{2\pi}\int_0^9\int_0^{\sqrt{9-z}} f(r\cos\theta, r\sin\theta, z) r\,dr\,dz\,d\theta$
(j) $\{(\theta,z) \in \mathbb{R}^2 : 0 \le \theta \le 2\pi, 0 \le z \le 9-r^2\}$ for each fixed $0 \le r \le 3$
(k) $\int_0^{9-r^2}\int_0^{2\pi} f(r\cos\theta, r\sin\theta, z) r\,d\theta\,dz$
(l) $\int_0^3\int_0^{9-r^2}\int_0^{2\pi} f(r\cos\theta, r\sin\theta, z) r\,d\theta\,dz\,dr$

9.6.5 Make sure you always include the stretch factor of $|r| = r$ in every integral.
(a) $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 \le 4, z \ge 1\}$ and, in cylindrical form, $\{(r,\theta,z) \in \mathbb{R}^3 : r^2+z^2 \le 4, z \ge 1\}$
(b) $\int_0^{\sqrt{4-z^2}} f(r\cos\theta, r\sin\theta, z) r\,dr$ for each fixed $0 \le \theta \le 2\pi$ and $1 \le z \le 2$
(c) $\int_0^{2\pi}\int_0^{\sqrt{4-z^2}} f(r\cos\theta, r\sin\theta, z) r\,dr\,d\theta$ for each fixed $1 \le z \le 2$
(d) $\int_1^2\int_0^{2\pi}\int_0^{\sqrt{4-z^2}} f(r\cos\theta, r\sin\theta, z) r\,dr\,d\theta\,dz$
(e) $\int_1^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz$ for each fixed $0 \le \theta \le 2\pi$ and $0 \le r \le \sqrt{3}$
(f) $\int_0^{2\pi}\int_1^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz\,d\theta$ for each fixed $0 \le r \le \sqrt{3}$
(g) $\int_0^{\sqrt{3}}\int_0^{2\pi}\int_1^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz\,d\theta\,dr$
(h) $\int_1^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz$ for each fixed $0 \le \theta \le 2\pi$ and $0 \le r \le \sqrt{3}$
(i) $\int_0^{\sqrt{3}}\int_1^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz\,dr$ for each fixed $0 \le \theta \le 2\pi$
(j) $\int_0^{2\pi}\int_0^{\sqrt{3}}\int_0^{\sqrt{4-r^2}} f(r\cos\theta, r\sin\theta, z) r\,dz\,dr\,d\theta$
(k) Use the example $f(x,y,z) = 1$ to calculate the volume for your different orders. You should get the same value for all of them.

For the following questions, you are encouraged to plot regions on Desmos or Math3D to help you set up your integrals.

9.6.6 You can check by calculating them. You should get $\frac{4\pi R^3}{3}$ in all cases.

9.6.7 $0 \le r \le 10, 0 \le \theta \le \pi/3, 0 \le z \le 8$

9.6.8 $\int_0^2 \int_0^\pi \int_{r^2}^{2r} zr^2 \sin(\theta) dz d\theta dr$

9.6.9 Check your solution by computing the integrals by hand or using a calculator. You should get $\frac{64\pi}{3}(2-\sqrt{2})$

9.6.10 Check your solution by computing the integrals by hand or using a calculator. You should get $\frac{148\pi}{3}$

9.6.11 $\frac{32\pi}{3}$

9.6.12 For either part, you should find the mass to be $\frac{2\pi}{5}(32-9\sqrt{3})$.

(a) There are two ways to approach this. One involves a sum of triple integrals corresponding to different slice shapes:

$$\int_0^{\sqrt{3}} \int_0^{2\pi} \int_{\sqrt{3-z^2}}^{\sqrt{4-z^2}} (r^2+z^2) r dr d\theta dz + \int_{\sqrt{3}}^{2} \int_0^{2\pi} \int_0^{\sqrt{4-z^2}} (r^2+z^2) r dr d\theta dz$$

The other involves a difference of triple iterated integrals:

$$\int_0^{2} \int_0^{2\pi} \int_0^{4-z^2} (r^2+z^2) r dr d\theta dz - \int_0^{\sqrt{3}} \int_0^{2\pi} \int_0^{\sqrt{3-z^2}} (r^2+z^2) r dr d\theta dz$$

(b) It is easiest to use a projection onto the x-y plane.

$$\int_0^{2\pi} \int_0^{\sqrt{3}} \int_{\sqrt{3-r^2}}^{\sqrt{4-r^2}} (r^2+z^2) r dz dr d\theta + \int_0^{2\pi} \int_{\sqrt{3}}^{2} \int_0^{\sqrt{4-r^2}} (r^2+z^2) r dz dr d\theta$$

9.6.13

(a) Label $(r+\Delta r, \theta, z)$, $(r, \theta+\Delta\theta, z)$, and $(r, \theta, z+\Delta z)$ in R and label these three transformed points in $g(R)$.

(b) You should get three partial derivatives evaluated at (r, θ, z) by doing a linear approximation for each quantity.

(c) What geometric condition is satisfied by the edges of a rectangular prism? Try taking the dot products of each pair of partial derivatives. What do you notice? How does this matter for Line 4?

9.6.14 See (9.6.3) and the corresponding derivation on polar coordinates in Section 9.3.2 which has more details.

9.7.1
(a) (D)
(b) (F)
(c) (E)
(d) (C)
(e) (B)
(f) (A)

9.7.2
(a) (B) upper hemisphere
(b) (F) 1/8-sphere
(c) (A) solid between two spheres
(d) (D) right half sphere
(e) (E) quarter sphere
(f) (H) cone with spherical cap

9.7.3 Make sure you always include the stretch factor $\rho^2 \sin\phi$ for all integrals.
(a) $\{(x,y,z)\in\mathbb{R}^3 : x^2+y^2+z^2\le 4, z\ge 1\}$ and $\{\rho,\theta,\phi)\in\mathbb{R}^3 : \rho\le 2, \rho\cos\phi\ge 1, 0\le\theta\le 2\pi\}$
(b) $\int_{\sec\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho$ for each fixed θ and ϕ
(c) $\int_0^{2\pi}\int_{\sec\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho\, d\theta$ for each fixed ϕ
(d) $\int_0^{\frac{\pi}{3}}\int_0^{2\pi}\int_{\sec\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho\, d\theta\, d\phi$
(e) $\int_0^{2\pi} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\theta$ for each fixed ρ and ϕ
(f) $\int_0^{\arccos(\frac{1}{\rho})}\int_0^{2\pi} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\theta\, d\phi$ for each fixed ρ
(g) $\int_1^2\int_0^{\arccos(1/\rho)}\int_0^{2\pi} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\theta\, d\phi\, d\rho$
(h) $\int_{1/\cos\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho$ for each fixed θ and ϕ
(i) $\int_0^{\pi/3}\int_{1/\cos\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho\, d\phi$ for each fixed θ
(j) $\int_0^{2\pi}\int_0^{\pi/3}\int_{1/\cos\phi}^{2} f(\rho\sin\phi\cos\theta,\rho\sin\phi\sin\theta,\rho\cos\phi)\rho^2\sin\phi\, d\rho\, d\phi\, d\theta$

9.7.4 Note that your solution will depend on where you choose to locate the region in $\mathbb{R}^3$.
(a) $\int_0^{\pi}\int_0^4\int_0^r r\, dz\, dr\, d\theta$
(b) $\int_0^{\pi}\int_0^{\pi}\int_3^4 \rho^2\sin\phi\, d\rho\, d\phi\, d\theta$
(c) $\int_0^6\int_0^3\int_0^{2-\frac{2}{3}y} dz\, dy\, dx$
(d) $\int_0^{\pi/2}\int_0^5\int_0^4 r\, dr\, dz\, d\theta$

9.7.5 Don't forget the stretch factor in your triple iterated integral. $0\le\theta\le 2\pi, 0\le\phi\le\pi, 0\le\rho\le R$.

9.7.6 $\int_0^{2\pi}\int_0^{\pi}\int_3^4 \rho^2\sin(\phi)d\rho\, d\phi\, d\theta = \frac{148\pi}{3}$

9.7.7 $\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\int_{\sqrt{3}}^{2} \rho^4\sin(\phi)d\rho\, d\theta\, d\phi = \dfrac{2\pi}{5}(32-9\sqrt{3})$

9.7.8 $\int_0^{\frac{\pi}{4}}\int_0^{2\pi}\int_0^{4} \rho^2\sin(\phi)d\rho\, d\theta\, d\phi = \frac{64\pi}{3}(2-\sqrt{2})$

9.7.9

You can always express them with any coordinate system. The goal is for you to consider which one is most natural or simple. The best choices include: 4 of them are cylindrical and rectangular, 2 of them are spherical and cylindrical, 1 of them is spherical only and 1 of them is rectangular only.

9.7.10

(a) Label $(\rho+\Delta\rho,\theta,\phi)$, $(\rho,\theta+\Delta\theta,\phi)$, and $(\rho,\theta,\phi+\Delta\phi)$ in R and label their transforms in $g(R)$.

(b) You should get three partial derivatives evaluated at (ρ,θ,ϕ) by doing a linear approximation for each quantity.

(c) What geometric condition is satisfied by the edges of a rectangular prism? Try taking the dot products of each pair of partial derivatives. What do you notice? How does this matter for Line 4?

9.7.11 See Section 9.7.2 and follow the details in the argument of Section 9.3.2.

9.8.1

(a) $vol(T(\Omega)) = |\det(T)|\text{vol}(\Omega) = 0.6$

(b) $vol(g(\Omega)) \approx |\det \mathrm{Dg}(1,2)|vol(\Omega) = 0.6$

9.8.2 None of them.

(a) Only satisfies I and III.

(b) Ony satisfies III.

(c) Only satisfies II and III. By restricting the domain and codomain of g_C, you will be able to satisfy I as well.

9.8.3

(a) Only satisfies I and III.

(b) I, II, and III. Note $g_B:\mathbb{R}^2\to\mathbb{R}^2$ is given by $g_B(u,v)=\frac{1}{3}(u+v,-2u+v)$.

(c) I, II, and III. Note $g_C:\mathbb{R}^2\to\mathbb{R}^2$ is given by $g_C(u,v)=\frac{1}{3}(u+v,u-2v)$.

(d) I, II, and III. Note $g_B:\mathbb{R}^2\to\mathbb{R}^2$ is given by $g_B(u,v)=\frac{1}{3}(2u+3v,-4u+3v)$.

9.8.4

(a) Only satisfies I and III.

(b) Only satisfies III.

(c) Only satisfies II and III. In fact, h_C is not defined on its domain. Even if you replace its domain $\mathbb{R}^2$ with $\{(x,y)\in\mathbb{R}^2 : x\neq 0\}$, it will still fail I.

(d) I, II, and III. Note $g_D:(0,\infty)\to(0,\infty)$ is given by $g_D(u,v)=(u^{2/7}v^{1/7},u^{-1/7}v^{3/7})$.

9.8.5

(a) Your first step should be $\mathrm{Dg}\,(r,\theta,z)=\begin{bmatrix}\cos\theta & -r\sin\theta & 0\\ \sin\theta & r\cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}$

(b) Your first step should be $\mathrm{Dg}(\rho,\theta,\phi)=\begin{bmatrix}\sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta\\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta\\ \cos\phi & -\rho\sin\phi & 0\end{bmatrix}$

9.8.6 Note $g(u,v)=\left(u^{1/2}v^{-1/2},u^{3/2}v^{-1/2}\right)$. Many of the required details appear as part of the justification in Example 9.8.13.

9.8.7
(a) $\Omega = \{(u,v) \in \mathbb{R}^2 : 1 \le u \le 2, 1 \le v \le 3\}$

(b) $(x,y) = g(u,v) = \left(u^{-1/3}v^{1/3}, u^{1/3}v^{2/3}\right)$

(c) $\det\left[Dh(x,y)\right] = -3y/x^2$

(d) $\det\left[Dg(u,v)\right] = -1/(3u)$

(e) 32/9.

9.8.8 14/3. For this parallelepiped, $0 \le z - y \le 3$, $0 \le x - y \le 2$, $1 \le x + y + z \le$.
Apply a change of basis (in the linear algebra sense) as your change of variables. This should convert the parallelepiped into a cube. Use a property of inverses of Jacobians for quicker calculations.

9.8.9
1. Construct a linear map which sends the standard basis elements $(1,0)$ and $(0,1)$ to $(3,1)$ and $(-1,1)$. You will obtain the formulae $x = \frac{u+1}{4}$ and $y = 2v$
2. 4
3. 1/4
4. 8

9.8.10 The integral evaluates to 12. Follow the same strategy as the previous problem. Construct a linear map sending the standard basis to $(3,0,0),(1,0,1)$ and $(0,2,1)$.

9.8.11 Pick a change of variables to instead integrate over $1 \le u \le 4$ and $2 \le v \le 6$. The area is $\dfrac{1}{12}$.

9.8.12 Try to transform the region to the rectangle $1 \le u \le 2$ and $1 \le v \le 2$. The area is $\dfrac{\ln 2}{2}$.

9.8.13 Try elliptical coordinates $x = 3r\cos\theta$ and $y = 5r\sin\theta$. The volume is $\dfrac{255\pi}{2}$.

9.8.14 $\dfrac{\pi^2}{2}$

9.8.15 None of these answers are unique.
(a) Your choice of U should be the product of two open intervals and your choice of V should be the plane minus a ray.

(b) You could use your sets from the previous part times $\mathbb{R}$.

(c) Your choice of U should be the product of three open intervals. Your choice of V should be $\mathbb{R}^3$ minus a half-plane.

9.8.16 You should define your map g to be the polar coordinate transformation. It should be a diffeomorphism whose image is S so make sure you satisfy all the conditions. The lemma from the previous problem may be helpful.

9.8.17 Follow the outline of Example 9.8.13. There are many items you need to prove. A key first step will be to prove that $g : (0,\infty)^2 \to (0,\infty)^2$ given by $g(u,v) = \left(u^{-1/3}v^{1/3}, u^{1/3}v^{2/3}\right)$ is its inverse.

9.8.18 Remember that the volume of any region is defined to be the integral of the constant function 1 over that region.

9.8.19 Your change of variables should be $g(u,v,w)=(au,bv,cw)$. Why is this a diffeomorphism whose image is the ellipsoid?

9.8.20 Apply the change of variables theorem with $\Omega=[a,b]$.

9.8.21 Play with Desmos for a visual illustration when $g:\mathbb{R}^2\to\mathbb{R}^2$ is the polar coordinate transformation.

(c) $\ell(x,y)=g(a,b)+dg_{(a,b)}(x-a,y-b)$

(e) Your vectors should be labeled $\Delta a\cdot e_1$ and $\Delta b\cdot e_2$.

(f) The images should be labeled $\Delta a\frac{\partial g}{\partial x}(a,b)$ and $\Delta b\frac{\partial g}{\partial y}(a,b)$

(g) The area of $g(R)$ is approximately the area of $\ell(R)$ which is $|\det Dg(a,b)|\,\mathrm{area}(R)=|\det Dg(a,b)|\Delta a\Delta b$ by Theorem 9.8.1.

9.8.22

(a) The square R should have bottom left vertex (r,θ) and (going counterclockwise) vertices $(r+\Delta r,\theta),(r+\Delta r,\theta+\Delta\theta),(r,\theta+\Delta\theta)$.

The image $g(R)$ should have the bottom right corner labelled $g(r,\theta)$ and (going counterclockwise) corners $g(r+\Delta r,\theta),g(r+\Delta r,\theta+\Delta\theta),g(r,\theta+\Delta\theta)$

The image $\ell(R)$ should overlap and appear as a parallelogram. For more details, see this Desmos demo and play with the sliders.

9.8.23 For line 4, while F is continuous on Ω, does F have to be continuous on the rectangle R?

For line 6, this suffers from the error in line 4. In addition to that, how does the set $dg_{t_i}(R_i)$ compare to the set $g(R_i)$?

For line 8, is R' a rectangle? And are all the R_i' rectangles?

A.10. Improper integrals

10.1.1 Exactly two of the functions are not locally integrable. Remember integrable functions and continuous functions are both locally integrable.

10.1.2 Exactly two of the statements are true. All of these can be determined by a carefully reading of the definition of improper integral.

10.1.3 There are many valid answers. Draw pictures to help yourself come up with several.

(a) One such exhaustion would be $\Omega_k=[a+\frac{b-a}{2k},b-\frac{b-a}{2k}]$ for $k\geq 1$.

(b) One such exhaustion would be $\Omega_k=\overline{B_k(0)}$ for $k\geq 1$.

(c) One such exhaustion would be $\Omega_k=\{x\in\mathbb{R}^n:\frac{1}{k}\leq\|x\|\leq 2k\}$ for $k\geq 1$.

(d) One such exhaustion would be $\Omega_k=\{(x,y)\in\mathbb{R}^2:\frac{1}{k}\leq\|(x,y)\|\leq 2k, y\geq\frac{1}{k}|x|\}$ for $k\geq 1$. (Draw this!)

10.1.4
(a) This statement is true. *Hint*: Work straight from definitions.
(b) This statement is false. *Hint*: You can find a very silly counterexample using the zero function.

10.1.5 Use the extreme value theorem and your main theorem on integrability.

10.1.6 Use Cauchy-Schwarz and that the definition of Jordan measurability on subsets of Ω.

10.1.7 Exactly two of these are true.

10.1.8 Toph is correct.

10.2.1 Exactly three of the statements are true. Only one of (a) and (b) is true.

10.2.2 Remember to use the monotone convergence theorem in all cases. Note the worksheet was updated instructing you to *not* calculate the last example.
(a) converges to 2π
(b) diverges to ∞

10.2.3 Only one of the integrals converge. *Hint:* Change of variables.

10.2.4 This appears in the textbook.
(a) Converges for $p > 2$
(b) Converges for $p < 2$

10.2.5 Make sure you use monotone convergence theorem, check local integrability, and define an exhaustion by compact sets.

10.2.6 This converges only when $\beta > 0$. Can you prove this?

10.3.1 Four of the nine improper integrals converge.

Note: To justify the last integral diverges, you need: if $\alpha > 0$ and $\int_\Omega f\,dV$ diverges to infinity then so does $\int_\Omega \alpha f\,dV$. You may assume this even though we have not explicitly proved it for improper integrals.

10.3.2
(a) (B)
(b) (B), (C)

10.3.3 For each part, either all of the integrals converge or none of them do.

10.3.4 Make sure you check local integrability, use monotone convergence theorem, and define an exhaustion by compact sets. Spherical coordinates will be useful.

10.3.5

(a) The improper integral diverges.

(b) The improper integral converges. *Hint*: Give a simple upper bound for $\frac{x^2}{\|(x,y,z)\|^2}$.

10.3.6 Line 2 and line 3 are the lines missing important justifications. For line 3, there are two issues: does the limit exist? And why is it equal to the improper integral?

10.3.7 Note $|\sin(x)\cos(y)| \leq 1$.

10.3.8

(a) Line 1 is missing the critical justification. *Hint*: See example 2 in the readings.

(b) Compare with $1/||x||^3$.

10.3.9 The error is in the application of the basic comparison test. Can you see why?

10.3.10 This converges only when $\alpha > 0$. Use polar coordinates and a well chosen exhaustion to compute.

10.3.11 First apply monotone convergence theorem to show that this integral is independent of choice of exhaustion. Then try to formalize the argument in Example 9.3.8.

10.3.12 It converges $\beta > 0$, and diverges for $\beta \leq 0$. One method is to use the monotone convergence theorem and calculate directly with a specific choice of exhaustion. Another method is to follow the ideas in Example 10.3.2.

10.3.13 Follow the ideas in Example 10.3.3 and use the previous problem.

10.3.14 Use limit laws after applying monotone convergence theorem.

10.3.15 Note $e^{-w^2} \leq 1$ since $w^2 \geq 0$.

10.3.16 Use the integral triangle inequality.

10.3.17

(a) See this desmos demo.

(b) Two such inequalities are $f \leq f^+ \leq |f|$ and $-f \leq f^- \leq |f|$.

A.11. Integration on curves

11.1.1
(a) All of them except γ_B are parametrizations of S.

(b) All of them except γ_F are smooth.

(c) All of them except γ_E are regular.

(d) All of them except γ_E are simple.

(e) Only γ_E is closed.

11.1.2
1. All of these are parametrizations of S.
2. All of these are smooth.
3. All of these are regular.
4. All of these except γ_B are simple.
5. All of these except γ_B are closed.

11.1.3 (a,c) are piecewise curves only and (b,d) are both piecewise curves and curves.

11.1.4
(a) The map γ is not regular at $t = 0$.

(b) Note that C is a line segment and thus has a smooth simple regular parametrization.

11.1.5 The true statements follow from the definition or a theorem from topology. Each false statement can be disproven with one of the following counterexamples: half a lemniscate, an infinite line, the boundary of a semicircular disk.

(a) True

(b) True

(c) False

(d) True

(e) False

(f) False

11.1.6
(a) γ_1 is a reparametrization with the same orientation.

(b) γ_2 is not a reparametrization because its image is not the same.

(c) γ_3 is a reparametrization with the opposite orientation.

(d) γ_4 is not a reparametrization because it traverses the curve twice instead of once.

(e) γ_5 is a reparametrization with the same orientation.

11.1.7
(a) The flaw is in line 5. The theorem gives a sufficient condition, not a necessary one.

(b) The flaw is in line 7. Cameron's application of the theorem tells us nothing about the point $(1,0) = \gamma(0) = \gamma(2\pi)$.

11.1.8 You must check that γ is injective and C^1 with $\gamma' \neq 0$ on $(0,4\pi)$.

11.1.9 Take $\gamma(t) = tq + (1-t)p$.

11.1.10 First sketch C. Parametrize one piece using a smooth simple regular parametrization $\gamma_1 : [0,1] \to \mathbb{R}^2$ and the other piece with a smooth simple regular parametrization $\gamma_2 : [1,2] \to \mathbb{R}^2$. You can use slightly different domains but they must be "adjacent" time intervals. Then define $\gamma : [0,2] \to \mathbb{R}^2$ using γ_1 and γ_2 in a piecewise manner.

11.1.11 Try to graph this using desmos. You may find that it is easiest to break C up into four pieces, each of which may be described as the graph of some function. Using this function you will be able to come up with a parametrization. For a more fancy (hard) solution, try taking x and y to be rational functions of t.

11.1.12

(a) Use $\varphi = \mathrm{id}_{[a,b]}$, the identity map of $[a,b]$. Remember you have to check a few things but they are all straightforward to check.

(b) Write down the equation that defines a reparametrization and try to solve for $\phi(t)$. You will only need to look at the first coordinate to find ϕ but you will have to check the second coordinate to ensure that your choice of ϕ does give a reparametrization.

11.1.13 This proof is directly from the definitions. Your proof should include two key justifications: a composition of bijective maps is bijective and the (single-variable) chain rule.

11.1.14 You will need the inverse function theorem. There are many little details you will need to check.

11.1.15

(a) (i) the same orientation
(ii) the opposite orientation

(b) (i) the same orientation
(ii) the opposite orientation
(iii) the opposite orientation
(iv) the same orientation

(c) Apply the chain rule to the bijection you constructed in the previous problem.

11.2.1

(a) $\gamma(t)$

(b) $\|\gamma'(t)\|$ and $s'(t)$

(c) $\displaystyle\int_a^b \|\gamma'(t)\| dt$ and $s(b)$

(d) $\displaystyle\int_a^t \|\gamma'(u)\| du$ and $s(t)$

11.2.2

(a) It will help if you try to describe what each quantity represents physically or geometrically.

i) 30
ii) 10
iii) Cannot be computed with the given information.
iv) Cannot be computed with the given information.
v) 9
vi) Cannot be computed with the given information.

(b) For any $A \in \mathbb{R}$, $B = A + 30$. Note A is not uniquely specified though it is common to assume $A = 0$. You may assume $A = 0$ for the next part.

(c)
i) 30
ii) 2
iii) Cannot be computed with the given information.
iv) 4
v) 1
vi) Cannot be computed with the given information.

11.2.3

(a) ≈ 125.68

(b) ≈ 9.69

(c) ≈ 497.77

11.2.4

(a) $s(t) = \dfrac{(4+9t^2)^{3/2}}{27} - \dfrac{8}{27}$

(b) $\gamma_{\text{arc}}(s) = \gamma\left(\dfrac{1}{3}\left((27s+8)^{2/3}-4\right)^{1/2}\right)$ for $0 \le s \le \dfrac{1}{27}(2527625\sqrt{20221}-8)$.

The arclength parameter of γ tells you how distance along the curve is a function of time. If you want to reparametrize γ in terms of arclength, you must determine how time is a function of distance along the curve.

11.2.5 ≈ 2.489

11.2.6 $2\sqrt{2}\pi + 2\sqrt{2}\pi^2$

11.2.7 $\dfrac{9\pi^2\sqrt{17}}{8}$

11.2.8 At some point in your proof, you will have to address two cases: γ_1 and γ_2 have the same orientation or they have the opposite orientation. If you do not, you are probably missing something.

11.2.9

1. $\dfrac{\int_a^b f(\gamma(t))\|\gamma'(t)\|dt}{\int_a^b \|\gamma'(t)\|dt}$
2. Break the curve into segments of length ds. Then set-up the average a sum of values of f weighted by the length of the segments.
3. Use the integral mean value theorem.

11.2.10 Use chain rule and integration by substitution.

11.2.11 Use the fundamental theorem calculus in both directions.

11.2.12

(a) ≈ 4.76073

(b) If your path were parametrized by arclength, then you must always be traveling at unit speed. Your proof should use the contrapositive of this statement. What is your average speed?

11.2.13

(a) $\sum_{i=1}^{N} ||\gamma(t_i) - \gamma(t_{i-1})||$ where $t_i = a + \frac{(b-a)}{N} i$ for $0 \leq i \leq N$.

(b) First, does the limit necessarily exist? Second, why must it equal the supremum? More justification is needed for both of these statements.

11.2.14

(a)

1. ds
2. mass of the piece
3. The first blank should say "totalling all the pieces (or integrating)" and the second blank should say "$\int_C \rho \, ds$."

(b) Density is approximately uniform on small scales. Therefore, the physical principle that the mass of an object with *constant* density is the density multiplied by its volume tells us that the mass of the infinitesimal piece is equal to ρds.

11.2.15

(a) $\displaystyle\int_a^b ||\gamma'(t)|| dt$

(b) $\displaystyle\int_a^b f(\gamma(t)) ||\gamma'(t)|| dt$

(c) $\displaystyle\sup_{t \in [a,b]} f(\gamma(t))$

(d) $\displaystyle\frac{\int_a^b f(\gamma(t)) ||\gamma'(t)|| dt}{\int_a^b ||\gamma'(t)|| dt}$

11.2.16 $2kR^2$ for some positive unknown constant k. The expression kR^2 is also correct since k is an arbitrary constant.

11.3.1

(a) $\dfrac{\gamma'(t)}{\|\gamma'(t)\|}$ and $T(t)$

(b) $F(\gamma(t)) \cdot T(t)$

(c) $\|\gamma'(t)\|dt$

(d) $\displaystyle\int_a^b F(\gamma(t)) \cdot T(t)\|\gamma'(t)\|dt$ and $\displaystyle\int_a^b F(\gamma(t)) \cdot \gamma'(t)dt$

11.3.2 Three of them are positive, one is negative, and one is zero. To compare their relative sizes, you must at least consider the magnitude of the tangential force along the curve as well as the length of the curve. C_5 is zero because it is tangent to the vector field.

(a) C_2, C_3, C_4

(b) C_1

(c) C_1, C_5, C_3, C_2, C_4

11.3.3 Two of them are negative because the tangential contribution is always negative along the curve. One of them is positive because the tangential contribution is always positive along the curve. One of them is positive because the tangential contribution is more positive on one half than it is negative on the other half. Two of them are zero by symmetry.

(a) Negative

(b) Positive

(c) Zero

(d) Positive

(e) Zero

(f) Negative

11.3.4 124

11.3.5 $\dfrac{39}{2}$

11.3.6 $\dfrac{154}{3}$

11.3.7 $e^{16} - 1$

11.3.8

(a) All of your answers in the F column should be 48. The numbers in the G column are:

(a) 36
(b) $\dfrac{112}{3}$
(c) 32
(d) 38
(e) $40 - \dfrac{12\sqrt{3}}{\pi}$

(b) F is a gradient vector field, meaning F is a vector field which can be written as the gradient of a real valued function.

11.3.9 This proof follows from applying the chain rule to the definition of reparametrization.

11.3.10
(a) This proof is mostly about unpacking definitions. It is intuitively true but there is something to show using a parametrization of C and a (related) parametrization of $-C$.

(b) A good proof should appeal to the property of single variable integrals for swapping limits of integration. Note the notation used can be a bit confusing but you must rely on context. In this case, the letter T in the integral $\int_{-C} F \cdot T\,ds$ is actually the unit tangent vector of $-C$, whereas the letter T in the integral $\int_C F \cdot T\,ds$ represents the unit tangent vector of C. An equivalent unambiguous statement is:

$$\int_{-C} F \cdot d\gamma = -\int_C F \cdot d\gamma.$$

11.3.11 The proof is again about unpacking definitions and nothing more. Once you unpack everything, the key step will boil down to linearity of the single variable integral.

11.3.12 Use a parametrization $\gamma : [a, b] \to \mathbb{R}^n$ of $C_1 + C_2$ to define parametrizations γ_1 of C_1 and γ_2 of C_2. Start with the lefthand integral using γ. Break it up using a standard property of single-variable integrals. Then replace γ with γ_1 and γ_2 using the definitions.

11.3.13
(a) Negative

(b) Zero

(c) $-\dfrac{GMm}{40000}$

11.3.14 $\frac{-11}{2}$

11.3.15 0

11.4.1 The FTLI implies that all paths have work equal to $f(0,2) - f(2,0) = 0$.

11.4.2 (a,b,c) are all $2 - e - \dfrac{1}{e}$. (e,f,g) are all 0.

11.4.3 Your argument will need FTLI and the fact that the work done by F along C is zero.

11.4.4 (a), (c), and (d) are irrotational.

11.4.5 Three of these are conservative.
(a) $f(x, y) = xy + C$ with $C \in \mathbb{R}$.

(b) G is not conservative.

(c) $h(x, y) = xy + x + \sin(y) + C$ with $C \in \mathbb{R}$.

(d) $j(x, y, z) = xyz^2 + C$ with $C \in \mathbb{R}$.

11.4.6

(a) ≈ 1.58

(b) F is not conservative. Your argument should use FTLI.

11.4.7

(a) Joel's argument is incorrect. Lines 2 and 3 have the same error. When integrating equation (1) w.r.t. x, the other variable y is a parameter. Strictly speaking, you are solving infinitely many problems, one for each slice $y \in \mathbb{R}$, so the constant can be different in each case and depend on y. The corrected form of line 2 must read $f(x,y) = x^2 + 3xy + C_1(y)$. Similarly, the corrected form of line 3 will read $f(x,y) = 3xy + 5y + C_2(x)$. Note that corrected forms of the two lines are no longer contradict each other, so you cannot make a conclusion of line 4.

(b) Ellie's argument is correct and well-justified.

(c) Tess's argument is correct but not quite well-justified at Line 4. Find a result from Section 4.2 to justify this line. In addition, Tess's argument lacks sufficient work because she does not explain how she found her answer.

11.4.8

(a) Jasmine's argument is incorrect. Line 4 has the first error. Read Mulan's argument to see why Jasmine's argument is wrong.

(b) Mulan's argument is correct and well-justified, but she should be more clear in Line 4 about what exactly is the contradiction. $\phi'(y) = x + 2$ is a contradiction because ϕ must be a function of x only.

(c) Tiana's argument is correct and well-justified by Lemma 11.4.11.

11.4.9

(a) Your proof should use Clairaut's theorem. If not, you are missing something.

(b) See Example 11.4.14.

11.4.10 Explicitly write out the condition $\partial_i F_j = \partial_j F_i$ for $1 \le i < j \le n$ for $n = 2$ and $n = 3$. Your proof should be a short sequence of equivalences.

11.4.11

(a) Line 3 uses (C). Line 4 uses (D). Line 5 uses (B). Line 6 uses (A).

(b) γ is C^1 on (a,b), and f is C^1 on open set containing $\gamma((a,b))$

(c) $\nabla f(\gamma(t)) \cdot \gamma'(t)$ is continuous for $t \in (a,b)$ by the previous part's explanation.

(d) Asif should use more words and explain what he is doing! Lots of implication symbols are a poor way to write mathematics. It is better to write in full sentences, and to justify each step.

11.4.12 Show that $\nabla f = F$.

11.4.13

(a) $\displaystyle\int_0^{4\pi} -gm\frac{t-4\pi}{\pi^2}\,dt$ where g is the acceleration due to gravity and m is the mass of the ball.

(b) Yes. Its potential function is $-gmz$, where g is the acceleration due to gravity at sea level and m is the mass of the ball.

(c) $8gm$

11.4.14

(a) $F(\gamma(t)) = m\gamma''(t)$

(b) $K(x,y,z) = K(\gamma(t)) = \dfrac{m}{2}\|\gamma'(t)\|^2$

(c) $\displaystyle\int_C F\cdot d\gamma = K(\gamma(b)) - K(\gamma(a))$. The trick here is to note that $\dfrac{d\gamma'(t)\cdot\gamma'(t)}{dt} = 2\gamma''(t)\cdot\gamma'(t)$.

(d) $\displaystyle\int_C F\cdot d\gamma = P(\gamma(a)) - P(\gamma(b))$.

(e) Your argument should use your results from (c) and (d).

11.5.1 No, it is not conservative. Your explanation should be related to path independence.

11.5.2

(a) Theorem 11.5.2 (b) or (c)

(b) Theorem 11.5.2 (a)

(c) Not equivalent.

(d) Not equivalent. The curves must be closed.

(e) Theorem 11.5.2 (b) or (c)

11.5.3 A steady flow of lava and the gravitational force field of the Earth are examples of conservative vector fields.

11.5.4

(a) Irrotational by B); Not conservative/path-independent by C) or by D).

(b) Irrotational by B) or by D); Conservative/path-independent by D).

(c) Irrotational by B) or by D); Conservative/path-independent by B) or by D).

(d) No such vector field exists. Your explanation should be short. There are two valid explanations: one using B), C) and Theorem 11.5.2, and another using A), C), and Poincaré's lemma.

11.5.5 (a), (c), (e), and (g) are simply connected. Remember that a simply connected domain must be path connected.

11.5.6 (a,b,c,d,e) and (g) are simply connected domains.

11.5.7 Seven of these are irrotational. Of those, Poincare's lemma can only be directly applied to 4. Note however that more than 4 of these are conservative (in fact at least six are conservative).

(a) F is irrotational and you can apply Poincaré's lemma to conclude it is conservative. One potential function is $f(x,y,z) = (xyz)$.

(b) F is irrotational. Although you cannot apply Poincaré's lemma, F is still conservative. One potential function is $f(x,y,z) = \ln(xyz)$.

(c) F is not irrotational.

(d) F is irrotational and you can apply Poincaré's lemma to conclude it is conservative. One potential function is $f(x,y,z) = (xyz)$.

(e) F is irrotational and you can apply Poincaré's lemma to conclude it is conservative. One potential function is $f(x,y,z) = (xyz)$.

(f) F is irrotational but not necessarily conservative. The function $f(x,y,z) = (xyz)$ is defined on $\mathbb{R}^3 \setminus \operatorname{span}\{(0,0,1)\}$, satisfies $\operatorname{curl}(\nabla f) = (0,0,0)$, and is conservative. Additionally, the function $F(x,y,z) = (-\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2}, 0)$ is also defined on $\mathbb{R}^3 \setminus \operatorname{span}\{(0,0,1)\}$ and satisfies $\operatorname{curl}(F) = (0,0,0)$. However, F is not conservative by Theorem 11.5.2 since $\int_C F \cdot T\, ds \neq 0$ when $C = \{(x,y,0) \in \mathbb{R}^3 : x^2+y^2 = 1\}$.

(g) F is irrotational and you can apply Poincaré's lemma to conclude it is conservative. One potential function is $f(x,y,z) = (\frac{x^2}{2} + \frac{y^2}{2} + \frac{z^2}{2})$.

(h) F is not irrotational.

(i) F is irrotational. Although you cannot apply Poincaré's lemma, F is still conservative. One potential function is $f(x,y,z) = \ln(xyz)$.

(j) F is not irrotational.

11.5.8

(a) Your proof should use the FTLI.

(b) If you have a closed curve C, can you present it as a concatenation of C_1 and $-C_2$, for some pair of curves C_1, C_2 with the same start/end points?

11.5.9

(a) Line 2

(b) Line 5

(c) Lines 3 and 6

(d) Your justifications should include (in no particular order): " by line 6", "by line 7", "by the FTC", "by definition of the partial derivative".

(1) By definition of the partial derivative.
(2) By line 6.
(3) By line 7.
(4) By the FTC.

11.5.10

(a) In line 5 equation (5) on the right side, the kth partial derivative was moved from outside of the integral to inside of the sum.

(b) Unlike the previous proof, you do not need to check anything related to path-independence because the curve C_x is a specific explicitly defined curve depending only x (and the origin). As C_x lies inside U and F is continuous on U, the line integral is defined and hence $f(x)$ is (unambiguously) defined.

(d) One justification relates to swapping something. In no particular order, the other justifications include: "by the FTC", "since F is irrotational on U", "by the chain rule and product rule".

(5) In addition to (a), Line 5 uses the product rule for partial derivatives.
(6) Since F is irrotational on U.
(7) By the chain rule and product rule.
(8) By the FTC.

A.12. Fundamental theorems in 2D

12.1.1 All except one of these is true.

12.1.2

(a) Negative for F. Zero for G.

(b) Negative for both F and G.

(c) Negative for F and positive for G.

(d) Negative for F and positive for G.

12.1.3 These are the same vector fields (in a different order) from Exercise A.12.4.

(a) Irrotational

(b) Irrotational

(c) Not irrotational

(d) Irrotational

(e) Irrotational

(f) Not irrotational

(g) Irrotational

(h) Not irrotational

12.1.4 (a), (c), (d), (f), and (g) are irrotational.

12.1.5

(a) i) positive ii) negative iii) zero iv) negative

(b) i) 2 ii) -3 iii) 0 iv) -5

(c) positive

(d) $\oint_C (F \cdot T)ds = -\frac{16}{3}$. You should get the same answer as $\iint_R \operatorname{curl}(F)dA$. This follows by Green's theorem (see K2 module). Either way, it's good to have a second way to check things.

12.1.6

(a) Clockwise. You can check this from the picture or by calculating the curl which is $\operatorname{curl}(F)(-3,1) = -\frac{16}{121}$.

(b) $\oint_C (F \cdot T)ds = -\log\left(\frac{25}{7}\right) \approx 1.273$.

(c) Your answer should be the same as (b). The secret, as you shall see, is Green's theorem.

12.2.1 Exactly 4 of the regions are regular. Out of the remaining 4, exactly 2 regions are not compact and exactly 2 regions are compact but not equal to the closure of their interior.

(a) Regular region.

(b) This region is not compact.

(c) Regular region.

(d) This region is not compact.

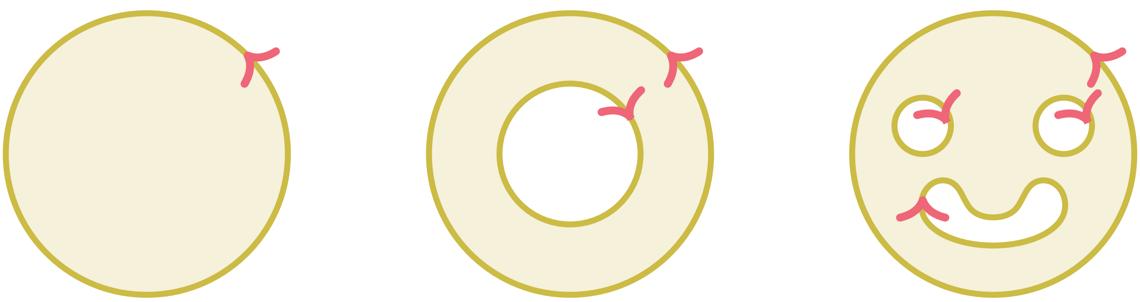
(e) Regular region.

(f) This region is not equal to the closure of its interior.

(g) This region is not equal to the closure of its interior.

(h) Regular region.

12.2.2



12.2.3

(a) Recall Definition 12.2.3 which says that positive orientation of each boundary component corresponds to the region staying on the left as you traverse the boundary. This gives
i) $\partial D = -C_1$
ii) $\partial(R \cup D) = C_2 \cup (-C_3)$
iii) $\partial R = C_1 \cup C_2 \cup (-C_3)$

(b) i) -23
ii) Not enough information
iii) Not enough information
iv) 43

(c) 10. To get the answer, apply Green's Theorem to $R \cup D$ and use part 1 ii).

(d) -6. To get the answer, apply Green's Theorem to R.

12.2.4 The work done is 8π. Notice the double integral can be evaluated without computation.

(a) One possible parametrization of C is $\gamma(t) = (2\cos t, 2\sin t)$ for $t \in [0, 2\pi]$. Using the aforementioned parametrization, $\oint_C F \cdot T \, ds = \int_0^{2\pi} (-2\sin t)^2 + (2\cos t)^2 dt = \int_0^{2\pi} 4dt = 8\pi$.

(b) Note that $\text{curl}(F) = 2$. Then if R is the region enclosed by C, by Greens theorem,
$\iint_R \text{curl}(F) dA = \iint_R 2dA = 2 \cdot \text{Area}(R) = 8\pi$

12.2.5

(b) $-16/3$

(c) $-104/3$

(d) $-88/3$

12.2.6 16

12.2.7 32

12.2.8 $2 - \dfrac{\pi^2}{2}$

12.2.9 One possible choice is $\left(-\dfrac{y}{2}, \dfrac{x}{2}\right)$. There are infinitely many correct answers.

12.2.10 30π

12.2.12 Line 4 needs more justifications. One approach uses the integral mean value theorem as well as continuity of curl(F) at p. Other approaches are more difficult.

12.2.13
(a) R_1, R_2, R_3, R_4.
(b) R_1, R_2, R_3.
(c) You have placed your heart R in a rectangle R' and divided it using a partition $\{R'_i\}_{i=1}^4$ of the rectangle. Each piece is given by $R_i = R \cap R'_i$ for $1 \le i \le 4$. No crying.
(d) Chop! Use vertical and horizontal lines only. This can be accomplished with only one cut. That single cut should be a vertical line which breaks the heart perfectly into two.
(e) Green's theorem has been proved in the case when the region is both x-simple and y-simple. To complete the proof, you must apply Green's theorem on each piece. If you sum all the line integrals, you will get miraculous cancellation and be left with the boundary of your heart. If you sum all the double integrals, you should get a double integral over your entire heart. Soul mended!

12.3.1 All of these are false.

12.3.2
(a) Positive for F and G
(b) Source for F and G
(c) Sink for F and source for G
(d) Source for F and source for G

12.3.3 These are the same vector fields (in a different order) from Exercise A.12.4.
(a) Sourceless
(b) Not sourceless
(c) Sourceless
(d) Sourceless
(e) Not sourceless
(f) Sourceless
(g) Sourceless
(h) Not sourceless

12.3.4
(a) $T = B = -D$
(b) $n = C = -A$
(c) Use $x(t) = \cos t$ and $y(t) = \sin t$. Evaluate at $t = 0$ to choose the correct formula. You know $n(0) = (1, 0)$ in this case by geometric considerations.

12.3.5
(a) Not sourceless
(b) Sourceless
(c) Sourceless
(d) Not sourceless
(e) Sourceless
(f) Sourceless
(g) Sourceless
(h) Not sourceless

12.3.6
(a) i) Positive ii) Negative iii) Positive
(b) Source
(c) $\oint_C (F \cdot n) ds = 16$. Your answer should be the same as $\iint_R \operatorname{div}(F) dA$. This follows by Green's theorem (see K3 module). Either way, it's good to have a second way to check things.

12.3.7
(a) Sink. You can check this from the picture or by calculating the divergence which is $\operatorname{div}(F)(2,2) = \frac{16}{81}$.
(b) $\oint_C (F \cdot n) ds \approx -0.297$.
(c) Your answer should be the same as (b). The secret is another version of Green's theorem.

12.3.8
(a) For the unit tangent, do some direct computation. For the unit normal, use Definition 12.3.1 and follow an argument similar to Example 12.3.4.
(b) Both sides are equal to a single variable integral dt. Start by expressing them in terms of P, Q, x, y, a, and b.
(c) Both sides are equal to a single variable integral dt. Start by expressing them in terms of P, Q, x, y, a, and b.

12.3.9 Here is the formal statement. Let $\gamma_1 : [a,b] \to \mathbb{R}^2$ and $\gamma_2 : [c,d] \to \mathbb{R}^2$ be parametrizations of an oriented curve C with unit normals n_1 and n_2 respectively. Let F be a vector field in $\mathbb{R}^2$ defined on C. The function $(F \circ \gamma_1)(t) \cdot n_1(t) ||\gamma_1'(t)||$ is integrable on $[a,b]$ if and only if $(F \circ \gamma_2)(t) \cdot n_2(t) ||\gamma_2'(t)||$ is integrable on $[c,d]$. If so,

$$\int_a^b F(\gamma_1(t)) \cdot n_1(t) ||\gamma_1'(t)|| dt = \int_c^d F(\gamma_2(t)) \cdot n_2(t) ||\gamma_2'(t)|| dt.$$

The proof is similar to line integrals (Theorem 11.3.1). Use the chain rule, change of variables, and the formula for the unit normal in components (Lemma 12.3.5).

12.4.1
(a) C
(b) A

12.4.2 Exactly three of the integrals equal outward flux and exactly three of them equal inward flux.

(a) Outward
(b) Inward
(c) Inward
(d) Outward
(e) Inward
(f) Outward

12.4.3
(a) The vector field is constant so both integrals should be zero.
(b) Both integrals would be zero by symmetry on opposite edges.
(c) Both integrals will be positive.
(d) The line integral is negative and the double integral is positive. The two integrals have opposite sign because C has negative orientation.

12.4.4
(a) From Lemma 12.3.5, you can see $n = \dfrac{(y'(t), -x'(t))}{||\gamma'(t)||} = \dfrac{(y'(t), -x'(t))}{||(x'(t), y'(t))||}$ but you should attempt to prove why this formula is correct based on Definition 12.3.1.
Here is how to obtain n for a specific example (the unit circle): Let $\gamma(t) = (x(t), y(t)) = (\cos t, \sin t)$. You know the tangent vector would be $\dfrac{\gamma'(t)}{||\gamma'(t)||} = (-\sin t, \cos t)$. So $n = (\cos t, \sin t)$ or $n = (-\cos t, -\sin t)$. The former option points outwards and so we choose that.
(b) Use $\gamma(a + b - t)$ for $a \leq t \leq b$.
(c) $n(t) = \dfrac{(-y'(a + b - t), x'(a + b - t))}{||\gamma'(a + b - t)||}$
Repeat (a) and be careful with the chain rule.

12.4.5
(b) 16.

12.4.6
(a) Orienting ∂R positively we get $\displaystyle\oint_{\partial R} F \cdot n ds = \int_{C_1} F \cdot n ds - \int_{C_2} F \cdot n ds.$
(b) -4
Hint: F is sourceless. Notice how using Green's theorem saved us a lot of work! Instead of parametrizing 4 different sides, determining their 4 unit normals, and computing 4 different integrals, you can instead parametrize one side, determine one unit normal, and compute 2 integrals.

12.4.7 0 for both parts

12.4.8 $\dfrac{1}{2}$

12.4.9 $F(x,y) = \left(\dfrac{x}{2}, \dfrac{y}{2}\right)$ is one valid choice. There are infinitely many. By Green's theorem, F must satisfy $\operatorname{div}(F) = 1$.

12.4.10 Once you apply Green's theorem you will be integrating the zero function.

12.4.12 Define a vector field G in terms of $F = (P, Q)$ such that $\operatorname{div} G = \operatorname{curl} F$.

12.4.13 The proof will be very similar but you will be computing $(F_1, 0) \cdot n$ instead of $(F_1, 0) \cdot T$, and similarly for F_2.

A.13. Integration on surfaces

13.1.1

(a) Yes. G is C^1 on $(0, 2\pi) \times (0, \pi)$ and its partials are bounded on this set, namely $||\partial_1 G|| \leq 1$ and $||\partial_2 G|| \leq 1$.

(b) Yes. $\partial_1 G(u, v)$ and $\partial_2 G(u, v)$ are linearly independent whenever $\sin v \neq 0$. If $(u, v) \in (0, 2\pi) \times (0, \pi)$ then $0 < v < \pi$ implies $\sin v \neq 0$, so the partials are always linearly independent on the interior of the domain.

(c) Yes. Note that G fails to be injective precisely on the boundary. You do not need to prove this, but you should briefly explain why geometrically.

13.1.2

(a) B and E are parametrizations of S.

(b) A, B, C, D, and E are smooth.

(c) A, B, D, E, and F are regular.

(d) A, B, D, and F are simple.

13.1.3

(a) G_A, G_B, and G_E are parametrizations of S.

(b) All of them are smooth.

(c) G_A, G_B, G_D and G_E are regular.

(d) G_A and G_B are simple.

13.1.4
(a) True, by definition.
(b) True, by definition.
(c) False.
(d) True as the continuous image of compact sets is compact.
(e) False; consider the half cone.
(f) False; consider an infinite plane.

13.1.5 The double cone and cube are piecewise surfaces, but not surfaces. The rest are surfaces (and hence also piecewise surfaces).

13.1.6 Consider the change of variables from spherical to cartesian coordinates. You should find that they have the same orientation.

13.1.7 Four out of five of these are reparametrizations of G. Of these, two have the same orientation.

(a) H_1 is a reparametrization with the same orientation.
You may verify this using $\varphi : [0,1]^2 \to [0,2] \times [0,3]$ where $\varphi(u,v) = (2u, 3v)$.
(b) H_2 is a reparametrization with the opposite orientation.
You may verify this using $\varphi : [0,3] \times [0,2] \to [0,2] \times [0,3]$ where $\varphi(u,v) = (v,u)$.
(c) H_3 is not a reparametrization.
(d) H_4 is a reparametrization with the opposite orientation.
You may verify this using $\varphi : [0,2] \times [0,3] \to [0,2] \times [0,3]$ where $\varphi(u,v) = (2-u, v)$.
(e) H_5 is a reparametrization with the same orientation.
You may verify this using $\varphi : [0,2] \times [0,3] \to [0,2] \times [0,3]$ where $\varphi(u,v) = (2-u, 3-v)$.

13.1.8 All three of your parametrizations should be simple and regular. Only the spherical parametrization will be smooth.

13.1.9 Two vectors $v, w \in \mathbb{R}^3$ are linearly independent if and only if the 3×2 matrix $[v|w]$ has an invertible 2×2 submatrix. This fact leads to a quick proof, but you can also proceed by definition of linear independence.

13.1.10 You only need to show the linear independence condition fails at one point in the interior. Both of these are very short justifications since any choice of point will be a valid choice.

13.1.11 There are many valid disproofs. For example, notice $G(0,1) = G(\pi,1)$.

13.1.12 The proof for G is straightforward but tedious. The disproof for H has many valid choices, such as $H(\pi,1) = H(3\pi,1)$.

13.1.14 Construct S as the union of a parametrized disk and a paraboloid.

13.1.16 Express S as the union of a cone and a portion of a hemisphere. Cylindrical coordinates will be useful.

13.1.17 There are six surfaces you need to check. It is enough to check one and state that the rest are similar, but you should provide a parametrization for each side.

13.1.18

(a) The error is in line 5. Note that the theorem gives a sufficient condition, not a necessary one.

(b) The flaw is in line 8. Theorem 13.1.16 only tells Fabian that S is a smooth manifold at interior points. Fabian must address the boundary separately.

13.1.19

(a) All of the required properties follow very quickly from your assumptions on G. Before careful to take into account the swapped variables.

(b) Define $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ by $\varphi(u, v) = (v, u) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$. The rest of the proof follows quite quickly with this choice.

13.1.20 This is a direct application of the definition of a reparametrization.

13.2.1

(a) False. Cross product outputs a vector and area is a scalar.

(b) True.

(c) True.

(d) True.

(e) False. $(\lambda a) \times (\lambda b) = \lambda^2 (a \times b)$

(f) True.

(g) False. $e_1 \times e_3 = -e_2$

13.2.2 (a), (d), (e), and (f) are equal to the surface area of S.

13.2.3 (c) and (e) are correct. Of the incorrect ones, (a) and (b) are nonsensical because f is a scalar function, (f) is a vector quantity, and is a mismatch of notation with iterated integrals.

13.2.4

(a) $\partial_1 P(s, t) = (1, 0, 0)$, $\partial_2 P(s, t) = (0, 1, 2t)$

(b) $\|\partial_1 P \times \partial_2 P\|(s, t) = \sqrt{1 + 4t^2}$

(c) Evaluating with an integral calculator, you should get approximately 5.9.

13.2.5 There are two different ways to parametrize the sphere. You can use the radius and the polar angle or you can use the height and the polar angle.

One possible parametrization is $G : U \to \mathbb{R}^3$ where $G(\theta, z) = (\sqrt{R^2 - z^2}\cos\theta, \sqrt{R^2 - z^2}\sin\theta, z)$ and $U = \{(\theta, z) \in \mathbb{R}^2 : 0 \le \theta \le 2\pi, -R \le z \le R\}$.

13.2.6 One possible parametrization is $G : U \to \mathbb{R}^3$ where $G(\theta, \phi) = R\cos\theta\sin\phi, R\sin\theta\sin\phi, R\cos\phi)$ and $U = \{(\theta, \phi) \in \mathbb{R}^2 : 0 \le \theta \le 2\pi, 0 \le \phi \le \pi\}$.

13.2.7

(a) $30\sqrt{741}$. You can estimate the surface area with a single parallelogram.

(b) $12\sqrt{741}+162\sqrt{3}$. To improve your estimate, use one parallelogram for $[1,3]\times[2,4]$ and another parallelogram for $[1,3]\times[4,7]$.

13.2.8 $3\pi\sqrt{2}$

13.2.9 The surface area is ≈ 6.62414. Try a variant of cylindrical coordinates. You can simplify some expressions with some trigonometric identities. Note the surface is actually two pieces and you can use symmetry.

13.2.10 18π

13.2.11 $\frac{5\sqrt{5}-1}{6}\approx 1.697$

13.2.12 This proof is just unpacking definitions in terms of parameterizations and double integrals.

13.2.13 Use the chain rule in Jacobian matrix form to express each partial of G as a linear combination of partials of H with scalars involving partials of φ. Then carefully apply properties of the cross product from Remark 13.2.3. The calculation will get messy, so persist!

13.2.14 Use 13.2.13 with the reparametrizing map φ. Then apply a change of variables (Theorem 9.8.2). Be careful about the direction you are applying change of variables since $G=H\circ\varphi$. Note the theorem's assumptions can be verified using the many properties of H and φ.

13.2.15

(a) $\displaystyle\iint_S \rho\,dS$

(b) Your justification should have three ingredients: a small infinitesimal piece of the surface, an estimae for the mass of this small piece, and adding up all the pieces. Your explanation should include the physical principal that the mass of a sheet of constant density is the density multiplied by surface area.

(c) $\displaystyle m=\int_0^{6\pi}\int_0^1 (s+1)^2\sqrt{4(s+1)^2+9}\,ds\,dt$

13.2.17 As always, the main challenge is parametrizing the surface S. Your parametrization should use x and a polar angle around the x-axis.

13.2.18 $\dfrac{k\pi(1+391\sqrt{17})}{60}$ where k is the constant of proportionality. The expression $\dfrac{k(1+391\sqrt{17})}{60}$ is also correct since k is an arbitrary constant.

13.3.1

(a) No, $A\times B$ points in the opposite direction as N.

(b) Yes, $A\times B$ points in the same direction as N.

(c) Mo, $A\times B=0$, so the question does not make any sense for this example.

(d) Yes, $A\times B$ points in the same direction as N.

13.3.2 Only one of these gives an orientation. One is not normal. One does not output unit vectors. One is not continuous.

(a) This figure shows an orientation of a plane.

(b) The vectors are not unit vectors.

(c) The vectors are not orthogonal.

(d) The unit normal is not continuous.

13.3.3 The sphere has no boundary and the tip of the cone is not a boundary point.

13.3.4

(a) Your picture should include a plot of $\mathbb{R}^2$. There should be a piece of the upper half plane mapping to the a piece of the upper hemisphere. The point $(1,0,0)$ should be the image of a point in $\mathbb{R}^2$ lying on the x-axis.

(b) Gina is correct, but hasn't explained the reasoning. If a point is on the relative boundary of a surface, then it must be come from a point on the topological boundary of the parametrization's domain. Hassan incorrectly believes the converse holds, but it does not. If a point is on the topological boundary of a parametrization's domain, then it may or may not map to the relative boundary of the surface. This aspect relates to the definition of simple (Definition 11.1.8).

13.3.5 Brita and Chang have better descriptions than Abed, but they make an error.

(a) No, the definition does not refer to parametrization's anywhere. Also, Abed's argument would incorrectly imply points like B are also on the relative boundary.

(b) No, the map φ must be invertible. What part of Brita's argument will fail this requirement?

(c) No, the map φ must be continuous and have a continuous inverse. What part of Chang's argument will fail this requirement?

(d) Only A.

13.3.6 The unit normal n of S is either equal to the parametrization's unit normal or -1 times the parametrization's unit normal.

(a) G is not a parametrization of the oriented surface S because $\frac{\partial_1 G \times \partial_2 G}{||\partial_1 G \times \partial_2 G||}$ points in the opposite direction as the unit normal n of S.

(b) H is not a parametrization of the oriented surface S.

(c) J is a parametrization of the oriented surface S.

(d) K is not a parametrization of the oriented surface S.

13.3.7 For (d) and (e), the relative boundary is empty. For each of the others, there are two components making up the relative boundary, requiring two parametrized curves. There are infinitely many possible parametrizations, the ones given below are simply one possibility.

(a) The two components can be parametrized with $\gamma_1(t) = (3\cos t, 3\sin t, 9)$, $0 \le t \le 2\pi$ and $\gamma_2(t) = (3\cos t, 3\sin t, 0)$, $0 \le t \le 2\pi$.

(b) The two components can be parametrized with $\gamma_1(t) = (3\cos t, 3\sin t, 3\cos t + 3\sin t)$, $\frac{\pi}{4} \le t \le \frac{3\pi}{4}$ and $\gamma_2(t) = (3\cos t, 3\sin t, 0)$, $0 \le t \le 2\pi$.

(c) The two components can be parametrized with $\gamma_1(t) = (\cos t + 1, \sin t, \sqrt{7 - 2\cos t})$, $0 \le t \le 2\pi$ and $\gamma_1(t) = (\cos t + 1, \sin t, -\sqrt{7 - 2\cos t})$, $0 \le t \le 2\pi$.

13.3.8 Parametrize the graph as $G(s,t) = (s,t,f(s,t))$ and then compute from there.

13.3.9 Once you have written everything down, the proof is quite short since the tricky calculations are in Exercise Section 13.2.3. You will need to use the fact that $\frac{x}{|x|} = 1$ if $x > 0$ and $= -1$ if $x < 0$.

13.3.10 Think about what happens to the roller coaster when you make one full lap around. Do you end up arriving the right way up or upside down? The Leviathan is an orientable surface and the Behemoth is not.

13.3.11

(a) Only regularity takes some time; the details can be quite messy but it all works out.

(b) ≈ 5.848

(c) Choose two limits towards the boundary of the domain $(0,2\pi) \times (-1,1)$ such that G maps to the same point, but the function n does not. How can you conclude from this outcome?

13.4.1

(a) Not a flux. The integrand is a vector-valued function.

(b) Not even a well-defined expression in our conventions. We have never interpreted dS as a vector, so one cannot take a dot product $F \cdot \mathrm{d}S$.

(c) Yes

(d) Not even a well-defined expression in our conventions. We have reserved dA to denote infinitesimal area of regions in $\mathbb{R}^2$, while $S \subset \mathbb{R}^3$ is a surface embedded in $\mathbb{R}^3$.

(e) Yes

(f) Yes

13.4.2

(a) F_1, F_2, F_4 are positive.

(b) F_3 is negative.

(c) $\displaystyle\int_S F_3 \cdot n \, dS < \int_S F_1 \cdot n \, dS < \int_S F_2 \cdot n \, dS < \int_S F_4 \cdot n \, dS.$

(d) You might expect $\displaystyle\int_S F_3 \cdot n \, dS < \int_S F_1 \cdot n \, dS < \int_S F_4 \cdot n \, dS < \int_S F_2 \cdot n \, dS.$

13.4.3

(a) Line 2. Technically speaking, a matrix cannot have vectors as entries but this notational shorthand for calculating cross products is perfectly fine to use.

(b) Line 4. Pierce forgot the factor $||(\partial_1 P \times \partial_2 P(s,t)||$ when replacing dS with $dsdt$. Alternatively, you can use the pseudo-identity $ndS = (\partial_1 P \times \partial_2 P(s,t)dsdt$ since the norms always cancel.

(c) Line 3. Pierce needs to explain why the unit normal points in the correct direction. There are two different ways to do so: plug in a point (e.g. $(s,t) = (0,0)$) or look at the sign of the z-component. The latter method only works since the unit normal is referred to as "upwards". The former method works in more diverse scenarios.

(d) $\frac{8}{3}$

13.4.4 14

13.4.5 $\frac{5}{4}$

13.4.6 72π

13.4.7

(a) Your answers in the G column should all be $-\pi$. The answers in the F column are

i) π
ii) $\frac{\pi}{2}$
iii) $\frac{13\pi}{3}$

(b) G is divergence-free.

13.4.8 This proof is just unpacking definitions. Introduce a parametrization of S. Use linearity of the double integral. All the required assumptions are satisfied by properties of your parametrization and of the vector fields F and G.

13.4.9 This statement may appear circular but it is not. There is a lot of little details to prove. Let $G : U \to \mathbb{R}^3$ be a parametrization of S. Express the unit normal n in terms of G. Now, define $H : V \to \mathbb{R}^3$ using the map G by swapping the variables. Check this map H is a reparametrization of G with opposite orientation. This shows H parametrizes $-S$. Compute its unit normal n' in terms of H and show that $n' = -n$. You will need a basic property of cross products from Remark 13.2.3.

13.4.10 Be careful that the variable n has two different meanings in each surface integral. Use Exercise Section 13.4.3 and its proof to clarify these interpretations. You will need to write each surface integral in terms of a parametrization.

13.4.11 The strategy is identical to Section 13.2.3. Use 13.2.13 with the reparametrizing map φ. Then apply a change of variables (Theorem 9.8.2). Be careful about the direction you are applying change of variables since $G = H \circ \varphi$. Note the theorem's assumptions can be verified using the many properties of H and φ. The only meaningful distinction is the orientation; this will matter at one step in your proof when writing $x = |x|$ for $x > 0$.

13.4.12
(a) The flux integral represents the flow of the water, measured in unit of volume per unit of time.
(b) The integrals should evaluate to the same number.
(c) Your answer should refer to the areas of the two surfaces as well as the velocity and angle of the stream.

A.14. Fundamental theorems in 3D

14.1.1 Exactly four statements are true and exactly two are false. The other two statements do not make sense.
(a) True
(b) True
(c) False
(d) True
(e) Nonsense
(f) False
(g) True
(h) Nonsense

14.1.2 (a,c,d) are sourceless.

14.1.3 The sphere, torus, and cube are closed. The rest are not.

14.1.4
(a) $\partial P = M$
(b) $\partial T = A - B + S$
(c) $\partial R = A - B + S - M$

14.1.5
(a) 0
(b) $\dfrac{5\pi}{2}$
(c) $4\pi(237^3 - 137^3)$

14.1.6 F is not sourceless. Every point (x, y, z) satisfying $x + y + z > 0$ is a source. Every point (x, y, z) satisfying $x + y + z < 0$ is a sink.

14.1.7 $\operatorname{div}(F) = 3yz^2 + 2y\sin z + 2xe^{2z}$. Not sourceless. It has both sources and sinks.

14.1.8 Any vector field with constant positive divergence will work. $F(x, y, z) = (2x, 0, y)$ is one example.

14.1.9 One such vector field is $F(x, y, z) = (0, 0, z^2)$.

14.1.10 $\frac{2}{3}$

14.1.11
1. You need to parametrize the 4 triangles that form the boundary. For example, a parametrization of the bottom face of the solid tetrahedron would be $\gamma : [0,1]^2 \to \mathbb{R}^3$ given by $\gamma(s,t) = (s(1-t), t, 0)$.
2. $\frac{1}{2}$

14.1.12 $\frac{3}{4}$

14.1.13 Fix a point $p \in \mathbb{R}^3$ and verify both sides hold at the point p. Expand the left hand side and use the product rule. Remember f is a scalar-valued function and F is a vector-valued function.

14.1.14
(a) Try to formulate this in 2-dimensions and then generalize the 2-dimensional statement. Informally this could be stated as "The normalized flux of F through shrinking rectangles centered at p tends to the divergence of F at p". You will need to define $R_\varepsilon(p) = \{p + x : x \in [-\varepsilon, \varepsilon]^3\}$ for $p \in \mathbb{R}^3$.

(b) Draw the normals of the different faces of the rectangle and try to relate them to the component functions. Again, parallel the 2-dimensional statement.

14.2.1 Exactly 3 expressions equal outward flux.
(a) Outward
(b) Inward
(c) Inward
(d) Outward
(e) Inward
(f) Outward

14.2.2
(a) $\iiint_P \operatorname{div}(F)dV.$

(b) $\iint_A F \cdot n ds = \iint_B F \cdot n ds + \iiint_T \operatorname{div}(F)dV - \iint_S F \cdot n ds.$

(c) One expression is a single surface integral. For the other two expressions, apply the divergence theorem to the solid T and also separately to the solid R. Here are four expressions:

$$\iint_A F \cdot n ds$$

$$\iiint_T \operatorname{div}(F)dV - \iint_S F \cdot n ds + \iint_B F \cdot n ds$$

$$\iiint_R \operatorname{div}(F)dV - \iint_S F \cdot n ds + \iint_B F \cdot n ds + \iint_M F \cdot n ds$$

$$\iiint_R \operatorname{div}(F)dV - \iint_S F \cdot n ds + \iint_B F \cdot n ds + \iiint_P \operatorname{div}(F)dV$$

14.2.3
(a) 0
(b) 3

14.2.4
(a) 2π
(b) $4\pi(237^3 - 137^3)$

14.2.5 $\frac{3}{4}$. Remember to relate the surfaces via the divergence theorem.

14.2.6 ≈ -381.7

14.2.7 $\frac{1}{2}$.

14.2.8 $\frac{5}{2}$.

14.2.9 The proof is analogous to the corresponding on lemma for the worksheet on Green's theorem and divergence in 2D.

14.2.10
(a) Compute div(curl(G)) and then apply Clairaut's theorem.

14.2.11 A good proof should use the divergence theorem, the mean value theorem for integrals, and the continuity of partial derivatives of F.

14.2.12 Express each side as the same sum of 6 iterated double integrals.

14.2.13 The first box should be $\oiint_{\partial R} (F \cdot n)\, dS$ and the second box should be $\iiint_R \operatorname{div}(F) dV$.

14.2.14 Take a look at the previous two questions. $F(x, y, z) = (x, 0, 0)$ is one possible vector field.

14.3.1 (d,e,g) are nonsense. The remaining 5 are true.

14.3.2
(a) i. counterclockwise viewed from above on the positive z-axis; ii. clockwise viewed from below on the negative z-axis.
(b) $(\operatorname{curl} F)(0, 0, 0) \cdot e_3$ and $(\operatorname{curl} F)(0, 0, 0) \cdot (-e_3)$
(c) Yes, if and only if $\{v_1, v_2, v_3\}$ is linearly independent.
(d) $n = \frac{(\operatorname{curl} F)(p)}{||(\operatorname{curl} F)(p)||}$ with magnitude $||(\operatorname{curl} F)(p)||$ provided $\operatorname{curl} F(p) \neq 0$

14.3.3 $a > 0, b > 0, c = 0$

14.3.4

(a) Sign of $(\operatorname{curl} H) \cdot e_3$ appears to be always negative.

(b) Sign of $(\operatorname{curl} H) \cdot e_2$ appears to be always zero. Notice the projection of H into a plane with normal e_2 appears to be constant everywhere.

(c) Sign of $(\operatorname{curl} H) \cdot e_1$ appears to be always zero. Notice the projection of H into a plane with normal e_1 appears to be zero everywhere.

14.3.5 (a,c,e,f) are curl free. For the graphs, remember to consider $(\operatorname{curl} F) \cdot e_i$ for each $i = 1, 2, 3$. In other words, does the vector field have some spin somewhere in one of the 3 standard basis directions?

14.3.6 clockwise at rate $1/\sqrt{2}$

14.3.7 Every entry in the table equals π.

14.3.8 The curl is always zero; it is irrotational.

14.3.9

1. $(2x^3yz + 6x^7y^5 - xy, -3x^2y^2z - 7x^6y^6 + y, yz - z)$
2. No, it is not curl-free.
3. It is spinning fastest around the unit vector $\left(\frac{7}{\sqrt{130}}, \frac{-9}{\sqrt{130}}, 0\right)$ at rate $\sqrt{130}$.

14.3.10 There are many valid choices. One example is $F(x, y, z) = (3z, 7x, 2y)$.

14.3.11 The calculation is purely by brute force. Compute each side and check that they are equal.

14.3.12

(a) For the first bullet, a valid parametrization would be $(x + \varepsilon \cos t, y + \varepsilon \sin t, z)$. The other two bullets are similar but be careful with the orientation; it matters which component has sine and which has cosine because of the righthand rule. For the second bullet, a valid parametrization would be $(x + \varepsilon \sin t, y, z + \varepsilon \cos t)$. For the third bullet, a valid parametrization would be $(x, y + \varepsilon \cos t, z + \varepsilon \sin t)$.

(b) For the e_3 component, your sketch should have the point p, the e_3 basis vector, and a small ε-disk centered at p lying in the plane orthogonal to e_3. The boundary of the disk must obey the righthand rule. The other 2 pictures are similar.

14.3.13

(a) In order, the three blanks can be filled in with phrases like "infinitesimal circulation in the direction of n", "$(\operatorname{curl} F) \cdot n dS$", and "Integrating over the entire surface S".

(b) $(\operatorname{curl} F) \cdot n$ represents the amount of local swirl on the surface (positive if counterclockwise; negative otherwise).

14.3.14 By direct calculation, both answers are zero. As you will see in the next module, this is a consequence of Stokes' theorem.

14.4.1 (f) and (g) cannot be oriented with the Stokes orientation because the curves do not give a compatible choice. Notice (b), (c), and (d); (e) and (h) are related; (f) and (g) are related. Remember as you walk along the surface near the boundary, the surface should always be on your left.

(a) demo

(b) demo

14.4.2 See this demo for a better look at the surfaces.

(a) One possible choice is $(4\cos t, 4\sin t, 0)$ for $0 \le t \le 2\pi$

(b) One possible choice is $(4\sin t, 0, 4\cos t)$ for $0 \le t \le 2\pi$. Be careful to make sure you parametrization obeys the righthand rule.

14.4.3

(a) $\displaystyle\iint_T (\operatorname{curl} F)\cdot n\,dS$

(b) $\displaystyle\iint_T (\operatorname{curl} F)\cdot n\,dS = \iint_R (\operatorname{curl} F)\cdot n\,dS$

(c) $\displaystyle\oint_A (F\cdot T)ds = -\iint_S (\operatorname{curl} F)\cdot n\,dS - \oint_B (F\cdot T)ds$

(d) $\displaystyle\oint_B (F\cdot T)ds = -\iint_S (\operatorname{curl} F)\cdot n\,dS - \oint_A (F\cdot T)ds = -\iint_{S\cup R} (\operatorname{curl} F)\cdot n\,dS$

(e) $\displaystyle\iint_R (\operatorname{curl} F)\cdot n\,dS = \oint_A (F\cdot T)\,ds = \iint_T (\operatorname{curl} F)\cdot n\,dS = -\iint_S (\operatorname{curl} F)\cdot n\,dS - \oint_B (F\cdot T)ds$

14.4.4 Both answers are zero, but for different reasons.

14.4.5 The first blank can be filled with "S is a closed surface" and the second blank can be filled with "F is irrotational".

14.4.6 $-\pi$

14.4.7 $\dfrac{5}{2}$

14.4.8 Choose one surface and calculate the total circulation. The value will be π. What do all three surfaces have in common? Use Stokes. (Notice this exercise is the same as Exercise Section 14.3.4 but now you have Stokes' theorem!)

14.4.9 For part (a), one possible parametrization is $\gamma(t) = (4\cos t, 4\sin t, 17)$ for $0 \le t \le 2\pi$. Your answers for both (b) and (d) should be the same. They are (coincidentally) both equal to zero. For part (c), one possible parametrization is $G(s,t) = (s, t, 81 - 4x^2 - 4y^2)$ for $x^2 + y^2 \le 16$ which gives the normal vector field $n(x,y,z) = (8x, 8y, 1)$.

14.4.10 Much easier than it looks! Use Lemma A from Exercise 14.4.5.

14.4.11 Both answers are (coincidentally) zero.

14.4.12 Similar to Exercise 14.2.11, a good proof should use Stokes' theorem, the mean value theorem for integrals, and the continuity of partial derivatives of F.

14.4.13 Let $R \subseteq \mathbb{R}^2$ be a regular region. Define $S = \{(x,y,z) \in \mathbb{R}^3 : (x,y) \in R, z = 0\}$ and show S is a surface parametrized by $G : R \to \mathbb{R}^3$ defined as $G(x,y) = (x,y,0)$. What is the relative boundary of S? Parametrize ∂S using a parametrization of ∂R. Orient S to keep the Stokes orientation. Now, finish the proof.

14.4.14 The first box should be $\oint_{\partial S} (F \cdot T)\, ds$ and the second box should be $\iint_S (\operatorname{curl} F) \cdot n\, dS$.

14.5.1
(a) True
(b) False
(c) False
(d) True
(e) Nonsense
(f) Nonsense

14.5.2
(a) False
(b) True
(c) True
(d) False
(e) False
(f) Nonsense
(g) False
(h) False

14.5.3
(a) $\iiint_X \operatorname{div}(F) = \iint_T F \cdot ndS + \iint_P F \cdot ndS + \iint_L F \cdot ndS - \iint_R F \cdot ndS$

You can rearrange the above identity to relate the downward flux through L to the downward flux through R. Note the downward flux through R is $-\iint_R F \cdot ndS$. You should get

$$\iint_L F \cdot ndS = -\iint_T F \cdot ndS - \iint_P F \cdot ndS + \iint_R F \cdot ndS + \iiint_X \operatorname{div}(F)$$

(b) $\oint_A F \cdot d\gamma = \iint_P (\operatorname{curl} F) \cdot ndS - \oint_B F \cdot d\gamma + \oint_C F \cdot d\gamma$

(c) $\oint_B F \cdot d\gamma = \iint_{P+T} (\operatorname{curl} F) \cdot ndS + \oint_C F \cdot d\gamma$

14.5.4 Both Abbigael and Alisa are correct. By moving the surface, D combined with P, as well as D combined with H both enclose regular regions where divergence theorem can be applied.

14.5.5 Both Amy and Sarah are correct. See this Math3D demo.

14.5.6

(a) Not a curl vector field. Your explanation should refer to the divergence of F.

(b) A curl vector field. Your explanation should refer to the divergence of G as well as the domain of G.

14.5.7 Yes, it is a curl vector field. There are many valid choices of G so that $F = \operatorname{curl} G$. One such choice is $G(x, y, z) = (-y, 0, -xyz + x^4)$.

14.5.8 $-\frac{\pi}{4}$; Note F was simplified to $F(x, y, z) = (x^2 + \cos(z^3), yz, 3xz)$

14.5.9 $-\frac{17}{24}$; Use the divergence theorem to move the flux integrals to the side of the tetrahedron lying in the plane $x + y + z = 1$.

14.5.10 16π

14.5.11 0; You cannot apply Stokes' theorem to the entire pentagonal surface since the z-axis lies inside it and F is not defined on the z-axis.

14.5.12 0; You can calculate the same quantity by using either Stokes to compute a line integral along the boundary or the divergence theorem to calculate a surface integral over an elliptical disk in the $z = 0$ plane.

14.5.13 -8π; You can apply Stokes and compute two line integrals over two circles.

14.5.14 0

14.5.15 One proof uses Stokes' theorem. The other proof uses the divergence theorem. Both proofs are short.

14.5.16 This is a quick application of the divergence theorem.

14.5.17

(d) Note the statement of Gauss' law was corrected so the point charge belongs to the interior of R. To prove the law, put a small ball around the point charge; you can do this because it belongs to the interior of R.

(e) It is an indicator function for whether or not the point charge belongs to the interior of R! Note it is not defined if the point charge lies on the boundary of R.

[1] *Annual Demographic Estimates: Subprovincial Areas, July 1, 2021*. 2022. URL: https://www150.statcan.gc.ca/n1/pub/91-214-x/91-214-x2022001-eng.htm.

[2] T. Bazett. *Calculus III: Multivariable Calculus*. YouTube. 2021. URL: https://www.youtube.com/playlist?list=PLHXZ9OQGMqxc_CvEy7xBKRQr6I214QJcd.

[3] T. Bazett. *Calculus IV: Vector Calculus*. YouTube. 2021. URL: https://www.youtube.com/playlist?list=PLHXZ9OQGMqxfW0GMqeUE1bLKaYor6kbHa.

[4] C. Chudzicki. *Math3D*. 2021. URL: https://www.math3d.org/.

[5] *Desmos*. 2021. URL: https://www.desmos.com/calculator.

[6] C.H. Edwards. *Advanced Calculus of Several Variables*. Dover Books on Mathematics. Dover Publications, 2012. ISBN: 9780486131955.

[7] G.B. Folland. *Advanced Calculus*. Featured Titles for Advanced Calculus Series. Prentice Hall, 2002. ISBN: 9780130652652.

[8] Larry J. Gerstein. *Introduction to Mathematical Structures and Proofs*. Springer New York, 2012. DOI: 10.1007/978-1-4614-4265-3. URL: https://doi.org/10.1007%2F978-1-4614-4265-3.

[9] A. Gracia-Saz. *MAT137*. YouTube. 2021. URL: https://www.youtube.com/channel/UCLzpR8AiHx9h_-yt2fAxd_A/featured.

[10] E. Herman and G. Strang. *Calculus*. Open Textbook Library v. 3. OpenStax, Rice University, 2016. ISBN: 9781938168079.

[11] T. Holden. *MAT237 lecture notes*. 2016. URL: http://home.tykenho.com/LectureNotes237.pdf.

[12] D. Hughes-Hallett et al. *Calculus: Single and Multivariable*. Wiley, 2017. ISBN: 9781119696551.

[13] R. Jerrard and Z. Wolske. *MAT237 online notes*. 2020. URL: https://www.math.toronto.edu/courses/mat237y1/20199/contents.html.

[14] Nathaniel Johnston. *Introduction to Linear and Matrix Algebra*. Springer International Publishing, 2021. ISBN: 978-3-030-52810-2. URL: https://www.springer.com/us/book/9783030528102.

[15] Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem*. Birkhäuser Boston, 2003. DOI: 10.1007/978-1-4612-0059-8. URL: https://doi.org/10.1007%2F978-1-4612-0059-8.

[16] G. Sanderson. *3Blue1Brown*. YouTube. 2021. URL: https://www.youtube.com/channel/UCYO_jab_esuFRV4b17AJtAw.

[17] Todd Schneider. *BallR: Interactive NBA Shot Charts*. 2020. URL: https://github.com/toddwschneider/ballr.

[18] T. Shifrin. *Multivariable mathematics: linear algebra, multivariable calculus, and manifolds*. 8th ed. ISBN 978=0-471-52638-4. John Wiley & Sons, 2005.

[19] Charles H. Stolze. "A history of the divergence theorem". In: *Historia Mathematica* 5.4 (1978), pp. 437–442. ISSN: 0315-0860. DOI: https://doi.org/10.1016/0315-0860(78)90212-4. URL: https://www.sciencedirect.com/science/article/pii/0315086078902124.

[20] *Wikimedia Commons*. 2022. URL: https://commons.wikimedia.org.

A

B

C

D

E

F

G

H

I

J

L

M

N

T

U

V